2-14-2014

Simulations of Chemical Catalysis

Gregory K. Smith

Gregory K. Smith
Candidate

Chemistry and Chemical Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Prof. Hua Guo, Chairperson

Prof. Martin Kirk

Prof. Debra Dunaway-Mariano

Prof. Susan Atlas

SIMULATIONS OF CHEMICAL CATALYSIS

by

GREGORY K. SMITH

B.S., Chemistry, Northeastern University, 1999

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

Chemistry

The University of New Mexico
Albuquerque, New Mexico

December, 2013

DEDICATION

To

My parents:
Anne and Edison Smith

ACKNOWLEDGEMENTS

First, a heartfelt thanks to my advisor, Prof. Hua Guo, for his support, guidance, and patience throughout this long process. From challenging me as a student, to guiding me in my first steps as a researcher, to expanding my exposure to many research areas, and finally to encouraging me as I muddled through the writing process, I appreciate it all deeply.

Thank you to my committee members, Prof. Martin Kirk, Prof. Debra Dunaway-Mariano, and Prof. Susan Atlas, for the many thoughtful comments which have greatly improved the final manuscript. I've learned an enormous amount from each of you. In particular, thank you Prof. Kirk for early encouragement to pursue graduate school and your many insights on molecular orbital theory and symmetry. Thank you Prof. Dunaway-Mariano for teaching me how to approach biological systems as a chemist, a subject area that now remains an enduring fascination. Thank you Prof. Atlas for teaching me the finer aspects of density function theory and the insightful and detailed feedback on the manuscript which has greatly improved the final result.

Thank you to the many graduate students, post-docs, and other professors that have helped me along the way. In particular, let me thank my lab mate and friend Mr. Ryan Johnson for the many stimulating discussions about chemistry, science, and life as we tried and try to work out the many tricky aspects of our research projects. Thank you Dr. Lin Sen, both for watching out for me in China, and your insight as we worked

through the beginning stages of the surface catalysis project. My introduction to the research group came from Dr. Dingguo Xu, as he patiently taught me about enzyme simulations as I made the transition from student to researcher. Also, thank you to Dr. Zhihong Ke, for teaching me the details of *ab initio* QM/MM simulations and for many fruitful discussions and collaboration on the SpvC project.

Thank you to the NSF, NIH, and ACS Petroleum Research Fund for providing the funding to pursue this research.

Thank you to Brian and Amanda, for being good friends as I went through this very long process.

Lastly, thank you to my parents, my brother, and my extended family for their understanding and encouragement as I pursued my passion in science.

Simulations of Chemical Catalysis

by

Gregory K. Smith

B.S., Chemistry, Northeastern University
PhD, Chemistry, University of New Mexico

Abstract

This dissertation contains simulations of chemical catalysis in both biological and heterogeneous contexts. A mixture of classical, quantum, and hybrid techniques are applied to explore the energy profiles and compare possible chemical mechanisms both within the context of human and bacterial enzymes, as well as exploring surface reactions on a metal catalyst. A brief summary of each project follows.

Project 1 – Bacterial Enzyme SpvC

The newly discovered SpvC effector protein from *Salmonella typhimurium* interferes with the host immune response by dephosphorylating mitogen-activated protein kinases (MAPKs) with a β-elimination mechanism. The dynamics of the enzyme substrate complex of the SpvC effector is investigated with a 3.2 ns molecular dynamics simulation, which reveals that the phosphorylated peptide substrate is tightly held in the active site by a hydrogen bond network and the lysine general base is positioned for the abstraction of the alpha hydrogen. The catalysis is further modeled with density functional theory (DFT) in a truncated active-site model at the B3LYP/6-31 G(d,p) level of theory. The truncated model suggested the reaction proceeds via a single transition

state. After including the enzyme environment in *ab initio* QM/MM studies, it was found to proceed via an E1cB-like pathway, in which the carbanion intermediate is stabilized by an enzyme oxyanion hole provided by Lys104 and Tyr158 of SpvC.

Project 2 – Human Enzyme CDK2

Phosphorylation reactions catalyzed by kinases and phosphatases play an indispensable role in cellular signaling, and their malfunctioning is implicated in many diseases. *Ab initio* quantum mechanical/molecular mechanical studies are reported for the phosphoryl transfer reaction catalyzed by a cyclin-dependent kinase, CDK2. Our results suggest that an active-site Asp residue, rather than ATP as previously proposed, serves as the general base to activate the Ser nucleophile. The corresponding transition state features a dissociative, metaphosphate-like structure, stabilized by the Mg(II) ion and several hydrogen bonds. The calculated free-energy barrier is consistent with experimental values.

Project 3 – Bacterial Enzyme Anthrax Lethal Factor

In this dissertation, we report a hybrid quantum mechanical and molecular mechanical study of the catalysis of anthrax lethal factor, an important first step in designing inhibitors to help treat this powerful bacterial toxin. The calculations suggest that the zinc peptidase uses the same general base-general acid mechanism as in thermolysin and carboxypeptidase A, in which a zinc-bound water is activated by Glu687 to nucleophilically attack the scissile carbonyl carbon in the substrate. The catalysis is

aided by an oxyanion hole formed by the zinc ion and the side chain of Tyr728, which provide stabilization for the fractionally charged carbonyl oxygen.

Project 4 – Methanol Steam Reforming on PdZn alloy

Recent experiments suggested that PdZn alloy on ZnO support is a very active and selective catalyst for methanol steam reforming (MSR). Plane-wave density functional theory calculations were carried out on the initial steps of MSR on both PdZn and ZnO surfaces. Our calculations indicate that the dissociation of both methanol and water is highly activated on flat surfaces of PdZn such as (111) and (100), while the dissociation barriers can be lowered significantly by surface defects, represented here by the (221), (110), and (321) faces of PdZn. The corresponding processes on the polar Zn-terminated ZnO(0001) surfaces are found to have low or null barriers. Implications of these results for both MSR and low temperature mechanisms are discussed.

Table of Contents

Chapter 1

Introduction

Simulations of Chemical Catalysis

"The first principle is that you must not fool yourself, and you are the easiest person to fool."

-Richard Feynman, 1964, Galileo Symposium in Italy

1.1 Catalysis in the modern world

One of the key pillars of modern society is chemical catalysis. Whether feeding the world through nitrogen fixation, producing refined fuels and materials, or treating complex diseases in biology, understanding how catalysts function is essential to building or improving new catalysts or inhibiting dysfunctional ones.

Biological catalysis, largely through the action of enzymes, makes the chemistry of life possible from the smallest single cell bacteria up to the complexity of a human being. The chemical reactions necessary for metabolism, growth, locomotion, replication, and the complex networks of controls within cells depend on catalysts to create products from reactants and create energy gradients needed to do work of all kinds. These enzymes are largely the result of natural selection over many millennia, refining functions, or creating new ones by duplicating and modifying old templates, keeping what works, and discarding what does not in the crucible of the natural world.

Language, technology, and scientific thinking have allowed us to adapt our societies beyond natural selection. Incremental advances by prior generations create new technology and tools, which in turn make greater advances possible. One of the most important advances for our current civilization is certainly the knowledge and application of chemical catalysis. From early examples such as the fermentation of sugars into alcohol, to present day chemical industries such as agriculture, energy, and chemical synthesis, to modern examples of harnessing biological catalysis through biochemical and genetic methods, catalysis underlies much of modern civilization.

1.2 What is catalysis?

Strictly speaking, a catalyst is an entity that which speeds up a chemical reaction from reactants to products, without itself being consumed or irreversibly changed by the process. Increasing the rate of a reaction can be accomplished in two ways:

1. Lowering the energy difference between reactants and the transition state.
2. Increasing the chance for reactants to approach the transition state.

Each of these categories encompasses a wide array of phenomena that will be touched upon in subsequent sections.

1.3 Conceptions of catalysis throughout history

Though use of catalysts (both incidental and purposeful) has a long history, recorded, systematic study of catalytic reactions began in the 1800s. The term “catalysis” was first used by Berzelius in 1835[1] and early conceptions centered on the idea of contact processes, whereby certain reactions are made possible or greatly increased in velocity by bringing the reactants in contact with some material. The mechanism by which this occurred was not known, but some early theories had elements of truth. For example, Faraday proposed that the action of a catalyst is to bind the reactants simultaneously, which has some explanatory power in terms of attempt probability but further developments were still to come regarding the concept of reaction barriers.[2]

Modern conceptions of catalysis largely began to emerge with the pioneering work by Jacobus Henricus van 't Hoff on developing the theory of chemical equilibrium and kinetics. Equation 1.1 was first written down by van’t Hoff in a general empirical form,

but it was Svante Arrhenius who recognized its importance in describing rates of chemical reactions in the 1880s.

His key insight was noting that the temperature dependence of typical chemical reactions was too large to be reconciled solely with regards to the translational energy of the reactants.[3] This led to the idea of activation energy, or an energy threshold that must be overcome to go from products to reactants.

$$k = A * e^{-E_a/RT} \qquad (1.1)$$

Here A is the prefactor, which generally describes the probability of a successful collision. E_a, also shown in Figure 1.1, refers to the reaction barrier or the difference between the reactant and transition state energies. R is the gas constant and T is temperature. When $E_a > RT$, this factor dominates due to the exponential dependence. When $E_a \leq RT$, the prefactor dominates.

More advanced treatments of chemical kinetics came throughout the 1920s through the work of Lindemann, Hinshelwood, Rice, Ramsperger, Kassel and others, culminating in what is today called Hinshelwood-RRK theory.[4,5] In the 1930s, transition state theory (TST) was developed independently by both Henry Eyring and by Meredith Evans and Michael Polanyi[6], and is still in widespread use today. The key feature of TST is the concept of a *transition state* or *activated-complex*, which is in equilibrium with the reactants. If we take two colliding reactants, A and B, there is a point of maximum energy that must be overcome before the product complex is formed. This transition state is indicated by the $[A \cdots B]^{\ddagger}$ notation in Equation 1.2.

$$A + B \rightleftharpoons [A \cdots B]^{\ddagger} \rightarrow AB \qquad (1.2)$$

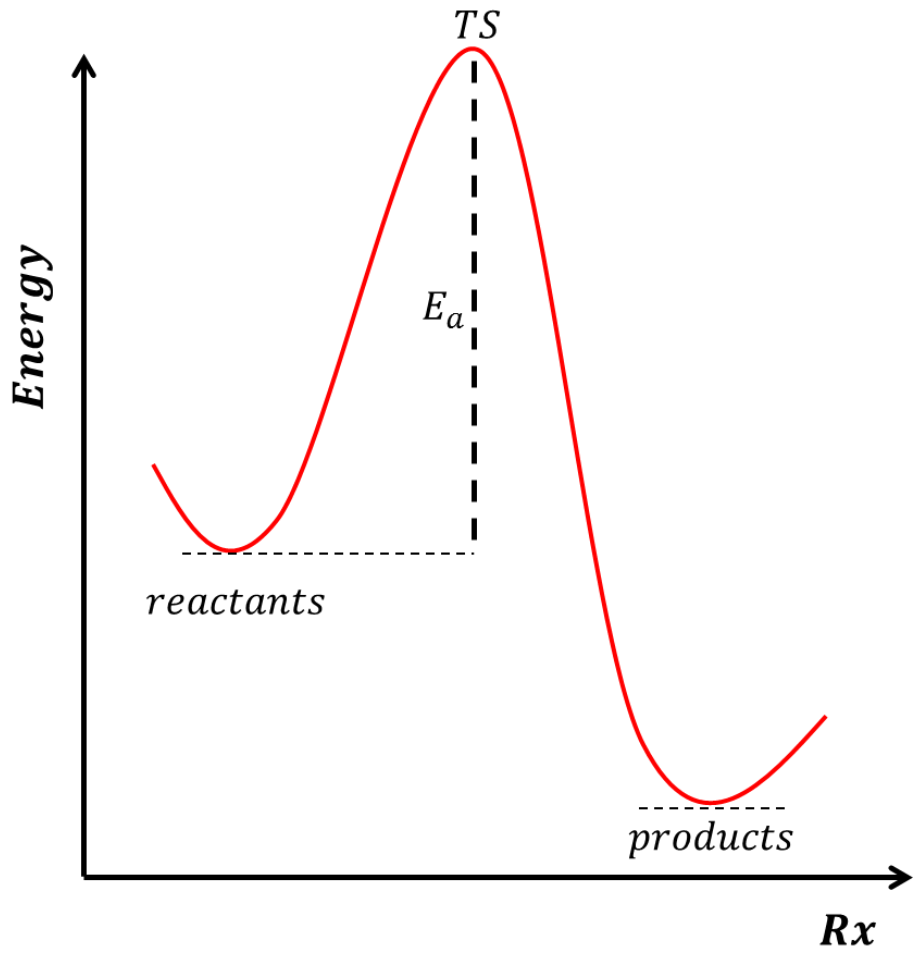


Figure 1.1 – Reaction profile showing reaction barrier $\boldsymbol{E_a}$ as the difference in energy between reactants and the transition state along a generic reaction coordinate Rx.

Using a combination of equilibrium expressions, their related partition functions, and the Maxwell-Boltzmann distribution, the rate equation can be expressed in terms of the free energy barrier.[3]

$$k = \frac{k_B T}{hc}\, e^{-\Delta G^{\ddagger}/RT} \qquad (1.3)$$

Where k_B is Boltzmann's constant, h is Planck's constant, and c is standard state concentration (defined as 1.0 Molar). The free energy barrier, $\Delta G^{\ddagger}$, combines entropic and enthalpic contributions to the reaction barrier, but these can be separated out to obtain a form similar to the Arrhenius equation.[3]

$$k = \frac{k_B T}{hc}\, e^{\Delta S^{\ddagger}/R}\, e^{-\Delta H^{\ddagger}/RT} \qquad (1.4)$$

Where E_a and the pre-exponential factor A can be written as:

$$E_a = \Delta H^{\ddagger} + 2RT \qquad (1.5)$$

$$A = \frac{e^2 k_B T}{hc}\, e^{\Delta S^{\ddagger}/R} \qquad (1.6)$$

Interestingly, it was later shown by Hinshelwood that Eyring's conception of transition-state theory is essentially equivalent to the Hinshelwood-RRK theory in the high pressure limit, where an equilibrium distribution of reactant and transition states becomes a reasonable assumption.[7]

In general, all catalysts either increase A or decrease E_a, and many catalysts do both. How these two goals are achieved can vary widely. The most obvious case is the lowering of the transition-state energy directly through interaction with the catalyst. However, since E_a is a difference in energy, an equivalent effect can be achieved by destabilizing or raising the energy of the reactant state, which is shown in Figure 1.2.

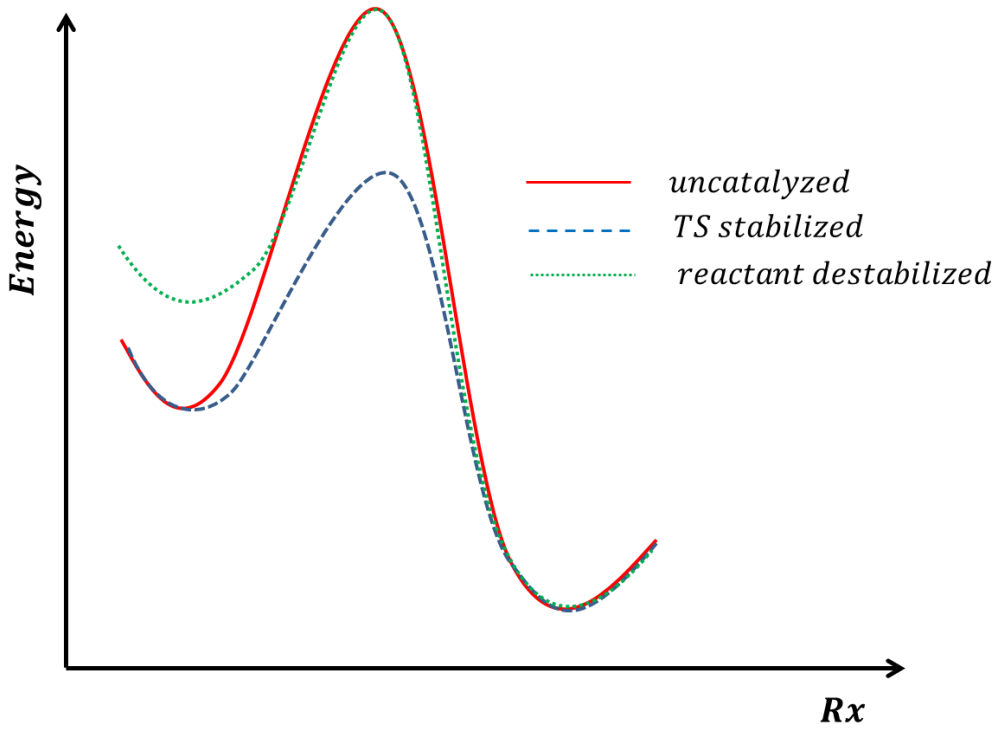


Figure 1.2 – Transition state stabilization and reactant destabilization approaches to catalysis

To explore transition state stabilization, consider the following decomposition example.

$$A - B \rightleftharpoons \left[A^{\delta+} \cdots B^{\delta-}\right]^{\ddagger} \rightarrow A^{+} + B^{-} \qquad (1.7)$$

If the neutral reactant has energy E, the transition state would have energy $E + \Delta E$ to account for the work done to separate the charge density against their Coulombic attraction.

If one introduces counter ions in the catalyst that are perfectly complementary with the transition state, the charge separation is partially screened by the nearby counter ions, resulting in a lower ΔE and hence lower reaction barrier. An example is shown in Figure

1.3, where Zn^{2+} ions are used by an enzyme catalyst to complement a reaction transition state and resulting intermediate.

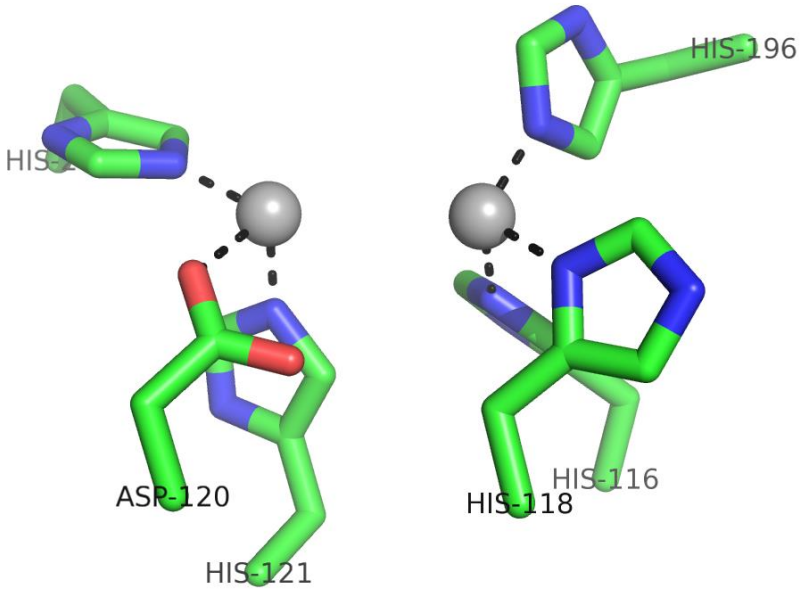


Figure 1.3 – Beta lactamase active site with 2 Zn^{2+} ions in grey, held in place by protein residues (carbon colored green). This enzyme is used by bacteria to cleave the C-N lactam bond in penicillin. The intermediate structure produced in this reaction consists of a nitrogen anion, which is stabilized by the positively charged zincs.

The prefactor in Equation 1.1 can be modified by a catalyst as well, with the classic example being one the earliest models of catalysis, that the catalyst simply binds the reactants to the catalyst surface. For a bimolecular reaction, a catalyst could simply have binding sites for A and B to increase their propensity for interacting over that of random collisions in solution.

$$A + B \rightarrow [A \cdots B]^{\ddagger} \rightarrow A - B \qquad (1.8)$$

Even if the binding sites are not adjacent, the slowing down of reactant species as they search the surface results in a greater probability that they will interact with each other when they meet.

One could achieve the same effect by lowering the temperature of the system to allow a greater interaction cross-section as well, but this would also change the E_a/RT ratio of the exponential term, and so would only be an effective strategy for reactions with zero or near-zero barriers. Indeed, for barrier-less reactions, inverse temperature dependence is a distinctive signature.

1.4 Free energy is the key quantity in chemical reactions

Most modern discussions of catalysis center on the concept of free energy, which combines the enthalpic and entropic terms.

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T \Delta S^{\ddagger} \qquad (1.9)$$

For small molecules, one can obtain the free energy of a particular conformation with a high degree of precision by calculating the internal energy using quantum mechanics and estimating the entropy using partition functions from statistical mechanics.

However, for complex systems like enzymes, no simple analytical result is possible since the system is too large and contains too many degrees of freedom to describe via quantum mechanical methods. Several approximations become necessary for these large systems.

One approximation is to partition the system into quantum and classical regions, where the bond breaking and bond forming events occur within the quantum region, and the electrostatic effects and protein motions are treated classically. Another approximation is to average out the enzyme degrees of freedom by sampling along a reaction path using a bias potential for a particular reaction coordinate, while letting the other degrees of freedom fluctuate. This method allows calculation of a potential of mean force, which is equivalent to the free energy profile along the reaction coordinate. [8] More details on reaction coordinates, bias potentials and the potential of mean force can be found in Chapter 2.

1.5 Categories of catalysis

Catalysis can be broadly defined into three categories: *homogeneous*, *heterogeneous*, and *biological*. These categories are not always distinct; each domain shares certain characteristics with the others and hybrid catalytic approaches are becoming increasingly common. This work will focus largely on biological and heterogeneous catalysis, with homogeneous catalysis serving to contrast specific aspects that are unique to these regimes.

1.5.1 Homogeneous catalysis

Homogeneous catalysis requires that the catalyst and reactants be in the same phase (solid/liquid/gas). The catalyst and reactant encounter each other and interact, either through the formation of chemical bonds or Coulombic forces, promote formation of products, and the catalyst is then released or regenerated. Homogeneous catalysis typically involves small organic molecules which interact at a particular chemical

functional group, at a certain atom pair. This method of catalysis is largely the realm of chemists and can be very efficient and controllable while requiring only mild conditions, but separation of the catalyst from the product can be problematic.

1.5.2 Heterogeneous catalysis

Heterogeneous catalysis occurs at the interface between two phases (most commonly gas/solid or liquid/solid systems). The interaction depends heavily on the nature of the interface and in the case of solid materials can be quite variable as even highly ordered crystals show defects on their surfaces. As a result, heterogeneous catalysts are typically less specific and controllable, often leading to the formation of a reaction network of competing chemical pathways. This method of catalysis is generally not as efficient or selective as homogeneous catalysis and therefore often requires elevated temperatures and pressures.

Heterogeneous catalysts have several advantages, however; the primary one being the separation of product is usually straight forward since the catalyst is a different phase. They are also ideally suited to continuous flow operation, rather than single batch. As can be imagined, this type of catalysis is largely the domain of the chemical engineer or material scientist, either on a research setting or on an industrial scale.

1.5.2.1 Surface catalysis and the materials/pressure gap

Traditionally, studies to characterize catalytic surfaces and their energetics are largely done in high vacuum conditions on low defect crystal surfaces. Conversely, working catalysts can often be highly disordered or amorphous and operate at moderately high

temperature and pressure conditions. Several advances in analysis techniques are allowing experimentalists to bridge this gap and look at catalytic systems under more realistic reaction conditions, but much remains to be improved in this area in both the experimental and modeling domains.

1.5.2.2 Sabatier's principle and volcano plots

Paul Sabatier observed that there seemed to be a relationship between the binding energy of a reactant and how effectively a catalysis performs.[9] The general principle is there is an optimum balance between tight binding of the reactant to the catalyst and lowering of the activation energy. If the reactant is bound too weakly, the collision frequency between reactants will be low, spending only a short time at the active site before releasing. If the reactants are bound too tightly, the collision frequency will be high, but the now the reactants will have a steeper hill to climb in terms of energy to get to the transition state.

Experimentally, this trend can be visualized by plotting substrate-catalyst binding energy vs. the temperature at 50% conversion, as shown in Figure 1.4. These plots are called *volcano plots* due to their distinction shape, and were first introduced by Balandin.[10] This trend appears to be very general in nature and similar plots can be constructed for many reaction families.

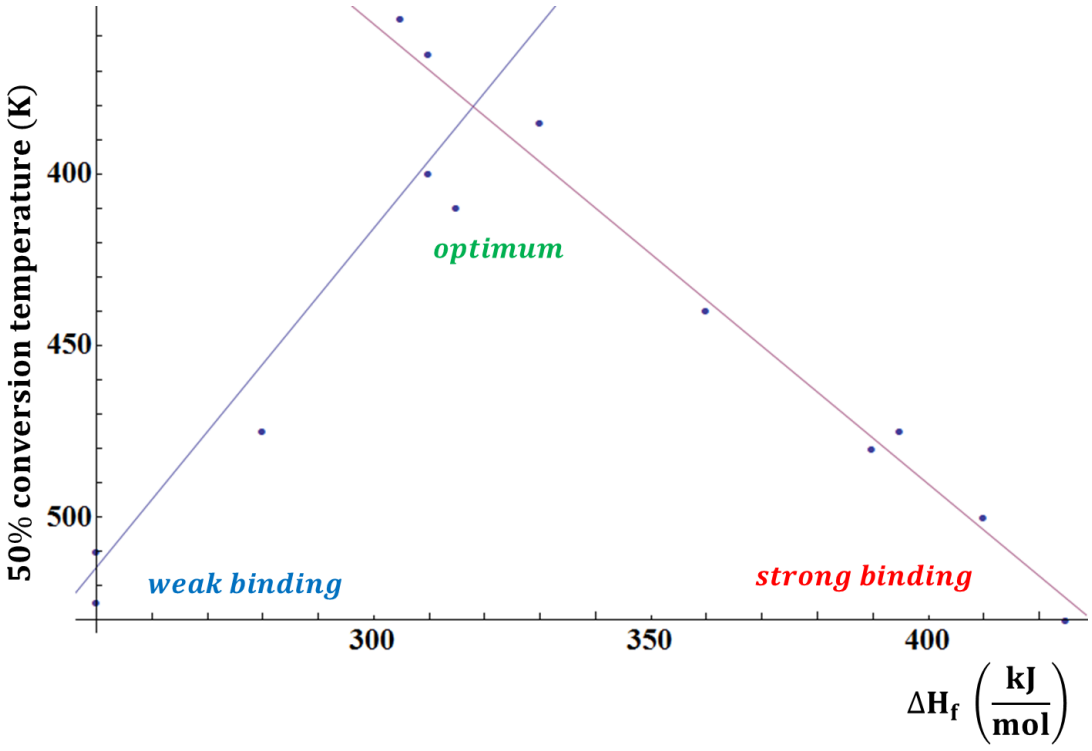


Figure 1.4 – A volcano plot illustrating Sabatier's principle. The x-axis shows average binding energy of substrate to metal surface; the y-axis is the temperature at which 50% conversion is met, and the data points represent a variety of different metal catalysts.

1.5.2.3 Bell-Evans-Polanyi (BEP) relation

Another general relationship across reaction families can be found between the difference in activation energy and the difference in enthalpy. These BEP relations are expressed as follows:

$$E_a = E_0 + \alpha\, \Delta H$$

or equivalently, (1.10)

$$\Delta\, E_a = \alpha\, \Delta\, H$$

Using this relationship, if one knows the difference in reaction enthalpy between two similar reactions, one can predict the change in activation energy. The factor α is a fractional value that specifies how far along the transition state is with respect to the reaction coordinate. For example, enthalpy differences would have a larger effect for a late transition state (product-like), and a smaller effect for an early transition state (reactant-like). For example, consider a reaction family like the hydrogenation of unsaturated hydrocarbons. Once ΔH and ΔE_a is known for a few reaction family members, reasonable ΔE_a predictions can be made for unexplored reactions by measuring ΔH only.

1.5.3 Biological catalysis

Biological catalysis is dominated by enzymes, though RNA and small molecules have catalytic roles as well. Enzymes are simply proteins that perform catalysis, and they are macromolecular polymers consisting of amino-acid building blocks. Though enzymes are non-branched linear polymers, upon synthesis they usually fold into a specific three-dimensional shape or a limited range of conformers. Due to the diversity of amino acid types, sequence, and sequence length, a large variety of shapes are possible.

The typical environment for enzymes is the cellular cytoplasm, where many enzymes and reactant molecules are all competing in a complex reaction network. Though many enzymes work in solution just as well as *in vivo*, many are considerably less robust stripped from structural scaffolds, and the synthesis and garbage collection machinery that exist in the cell. As can be imagined, enzymes have traditionally been the domain of biochemists, though their study and application diffuses to all areas of science.

Enzymes function very generally in three steps:

1. Binding of reactants at the active site
2. Increasing the rate of reactants to products
3. Releasing the products to recover the original enzyme state

Though step two is often the primary focus in enzyme studies, many enzymes can be rate limited by steps 1 and 3. However, if the catalytic step is rate limiting, enzymes display what is known as *saturation* or Michaelis-Menten kinetics.

1.5.3.1 Michaelis-Menten kinetics[11]

Saturation kinetics typical of enzymes is described by the formulism developed by Michaelis and Menten in 1913.[12] First, we begin with the general description of both enzyme-substrate binding followed by the catalytic step to produce products.

$$E + S \underset{\mathbf{k_{-1}}}{\overset{\mathbf{k_1}}{\rightleftharpoons}} ES \overset{\mathbf{k_2}}{\rightarrow} E + P \qquad (1.11)$$

Here E refers to the enzyme, S the substrate, ES the enzyme-substrate complex, and P the product. The rate equation for the formation of product, which depends on the catalytic rate constant k_2 and the concentration of the enzyme-substrate complex is shown below.

$$\frac{d[P]}{dt} = k_2[ES] \qquad (1.12)$$

The first major assumption is that $\mathrm{k}_1 \gg \mathrm{k}_2$ so that the catalytic step is rate limiting or serves as a bottleneck in the overall process. The second major assumption is that the concentration of the substrate is much greater than the concentration of the enzyme, $[S] \gg [E]$, which leads to the steady state approximation.

$$\frac{d[ES]}{dt} = 0 \qquad (1.13)$$

In other words, after a short ramp up period, the concentration of the ES complex becomes a constant. If the overall rate for ES is zero, this also implies that the microscopic rates both creating and consuming ES must be equal.

$$k_1[E][S] = (k_{-1} + k_2)[ES] \qquad (1.14)$$

The enzyme concentration $[E]$ can be written as the difference between an initial concentration and the enzyme-substrate complex concentration, or $[E] = [E]_0 - [ES]$. This allows us to write equation 1.14 as follows.

$$k_1[E]_0[S] - k_1[ES][S] = (k_{-1} + k_2)[ES] \qquad (1.15)$$

Solving for $[ES]$ and multiplying $\frac{1}{k_1}$ in both numerator and denominator yields:

$$[ES] = \frac{[E]_0[S]}{\left(\frac{k_{-1} + k_2}{k_1} + [S]\right)} \qquad (1.16)$$

The ratio of rate constants in the denominator is known as the Michaelis constant K_m, with units of concentration, and it describes how well the enzyme and substrate bind

together. Inserting K_m and substituting equation 1.16 into the product rate equation 1.12 yields the Michaelis-Menten rate equation.

$$\frac{d[P]}{dt} = k_2[ES] = \frac{k_2\,[E]_0[S]}{(K_m + [S])} \qquad (1.17)$$

This is usually rewritten in standard form by defining $v \equiv \frac{d[P]}{dt}$, and $v_{max} \equiv k_2[E]_0$. The reasoning is that the maximum rate is found in the limit of complete enzyme saturation or $[E]_0 \approx [ES]$. Equation 1.17 then becomes:

$$v = \frac{v_{max}[S]}{K_m + [S]} \qquad (1.18)$$

Figure 1.5 illustrates the rate as a function of the substrate concentration, showing the overall behavior of saturation kinetics.

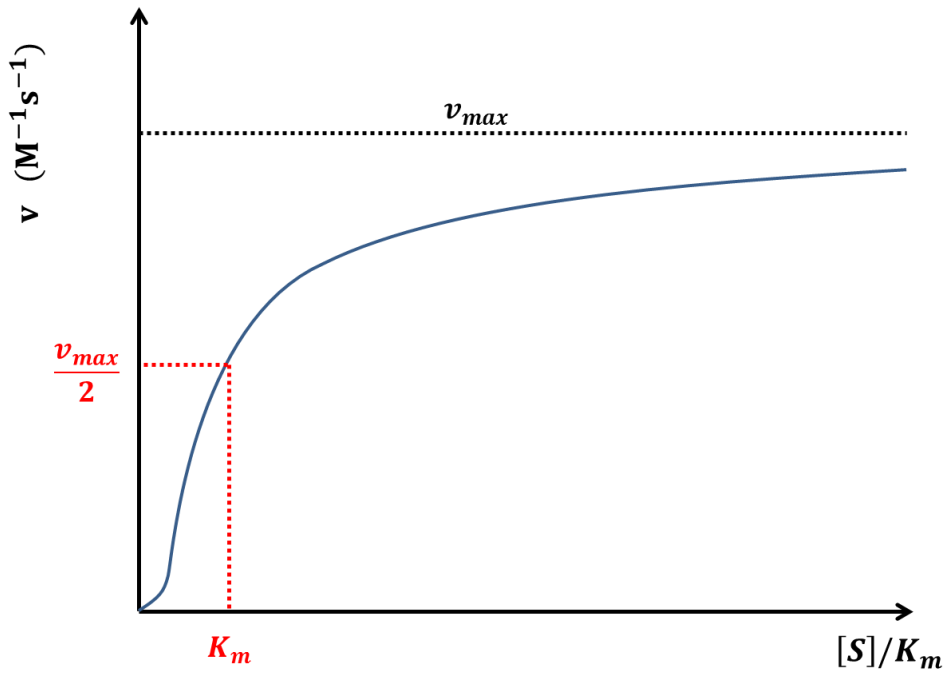


Figure 1.5 – A typical rate profile for Michaelis-Menton kinetics, showing concentration in units of K_m versus overall rate.

Figure 1.5 shows two limiting cases. First, when $[S] \gg K_m$, the rate approaches v_{max} asymptotically. Second, when the substrate concentration equals K_m, the rate equals $\frac{v_{max}}{2}$. These parameters can be found experimentally with the use of Lineweaver-Burk plots[13], where the rate equation is cast in reciprocal form.

$$\frac{1}{v} = \left(\frac{K_m}{v_{max}}\right)\frac{1}{[S]} + \frac{1}{v_{max}} \qquad (1.19)$$

Plotting the inverse substrate concentration versus the inverse rate gives a linear equation with slope $\frac{K_m}{v_{max}}$ and intercept $\frac{1}{v_{max}}$. This also enables the calculation of the catalytic rate constant once v_{max} is known, since $v_{max} = k_2[E]_0$.

The ratio of the catalytic rate constant (k_2 or the more common label, k_{cat}) to the binding constant is a common measure of enzyme efficiency. $\frac{k_{cat}}{K_m}$ ratios on the order of 10^7 or greater are considered to be kinetically perfect in that the enzymes are limited not by conversion to products, but rather by the rate of diffusion of substrate to the active site of the enzyme.

One example of a diffusion-controlled enzyme is carbonic anhydrase, which converts carbonic acid to water and carbon dioxide. While this reaction is fairly slow in solution, it proceeds rapidly in the presence of the enzyme. Carbonic anhydrase achieves this speed-up through several different effects. One factor is the presence of a Zn^{2+} ion in the active site, which provides a strong electrostatic interaction with hydroxide ion, serving as a counter ion. In addition to the electrostatic effects with zinc, the enzyme

also employs general acid catalysis and shape complementarity to achieve its catalytic properties.

Which factors are dominant in explaining enzyme activity is a matter of ongoing debate. The theories of how enzymes function are numerous, but a brief overview is given below.

1.5.3.2 Theories of enzyme catalysis

An early attempt at explaining enzyme catalysis was the lock and key model of Emil Fischer, who imagined a perfect shape complementarity between enzyme and substrate, which sought to give an explanation of the incredible specificity of many enzymes. This simple model was expanded upon by introducing the concept of strain in a rigid fashion by Haldane[14] and a more flexible induced fit model by Koshland.[15] The release of this strain by transitioning to products served as a possible mechanism of enzyme action.

Pauling introduced the concept of complementary binding of the transition state.[16] He posited that the active site of an enzyme is arranged both in shape and charge to complement the transition state structure of a reaction, binding preferentially to this structure, and thereby lowering the energy of activation. This is still widely accepted as largely correct, though many have formalized this idea more rigorously.

In particular, the concept of preferential binding of the transition state has been expanded by Ariel Warshel and colleagues in terms of electrostatic stabilization.[17,18] The basic concept is that the enzyme environment as a whole is pre-organized both to bind

with the transition state preferentially, thus reducing the activation energy, and also bind the reactants and products in a way to shorten the reaction path taken (and thus the reorganization energy). Warshel has also pioneered the use of computational methods to quantify these effects.

Several have proposed the importance of reactant orientation, including Koshland with the concept of orbital steering[19] and Bruice with the concept of near-attack conformers (NAC).[20] Entropic considerations including proximity, confinement, and entropy traps have also been proposed.[21] Firestone and Christensen posited that the enzymes main role is to minimize translational movement while favoring vibrational motion, including the vibrational mode associated with the transition state.[22]

Though the majority of the above processes are based on lowering the activation energy, another line of thought suggests that protein dynamics are a significant factor in catalytic action. Recent reviews by Benkovic and Hammes-Schiffer focus on coupled protein motions.[23,24] This has mirrored an increase in the use of NMR techniques, which makes it possible to look at the dynamics of enzymes in solution, enabling correlation with computation models. Several recent experiments have correlated protein conformer fluctuations with substrate turnover.[25-27] Other roles for dynamics include flexible loops that can open to admit substrate and then close to exclude solvent from the active site.

Along similar lines, some authors have proposed that many enzymes act mechanically, and that the limiting factor is not determined by activation energies, but is dominated

by the prefactor term in the rate equation. That is, the enzyme is essentially facilitating a successful collision between reactants due to the binding sites and natural protein dynamics, causing the enzyme to function as a molecular machine. For early treatments see Williams[28] and Moss[29], and a recent in-depth treatment by Swiegers.[5]

Though debate is ongoing about what factors are most important, it is clear that different families of enzymes employ different approaches to catalysis. For example, enzymes with metal ions in their active sites tend to support the electrostatic view of enzyme catalysis, due to the large electrostatic effect introduced by these metals. On the other hand, enzymes with flexible loops or important vibrational breathing modes tend to favor the dynamics view of enzyme catalysis.

1.5.4 Hybrid catalysis (Mimetics and Nano Catalysis)

Given the unique advantages of each type of catalysis, a natural question is: can the different types be combined so as to realize the benefits of all while mitigating the drawbacks for a specific application?

In the realm of organic and inorganic chemistry, the synthesis of bio-mimetics remain an active area of research. A bio-mimetic is generally a moderately sized molecule, synthesizable by organic or self-assembly techniques, that mimics the three-dimensional shape and function of biological macromolecules such as enzymes. This shape specificity can be accomplished directly by rigid molecular segments such as aromatic rings that adopt a limited range of conformers. Another approach is to use metal ions to serve as an electrostatic anchor to surrounding organic ligands, such as one would

see in metallocenes. These approaches try to combine the advantages of biological and homogeneous catalysis: efficiency, selectivity, and ease of synthesis.

Another set of approaches can be loosely classified under the realm of nanoscience, where attempts at unifying the advantages of heterogeneous catalysis with the homogeneous and biological domains remain active areas of research. Hetereogenous catalysts are fairly inefficient since the catalysis happens only at the interface, making the large proportion of the bulk completely inactive for catalysis. Creation of nanoparticles, often through simple chemical techniques, partially solves this problem by increasing the surface to volume ratio. However, quantum size effects can start to play a major role, with good or bad results depending on what reactivity is desired.

1.6 Projects contained in this work.

This dissertation contains four main projects, three involving biological catalysis, and one involving heterogeneous catalysis. The first entails a classical and QM/MM molecular dynamics study of a bacterial enzyme found in *Salmonella*, with possible implications for new anti-bacterial approaches involving the targeting of virulence factors. The second project uses an *ab initio* quantum/classical approach to study reaction mechanisms of an enzyme, CDK2, which is involved in the human cell cycle and thus has implications for cancer therapies. The third uses a semi-empirical quantum/classical approach to study anthrax lethal factor, with implications for treatment of inhalation anthrax infections, where antibiotics are not sufficient to deal with the circulating toxins that lead to rapid patient death. The last project uses a fully quantum approach to study surface reactions on a PdZn alloy, a catalyst used in

methanol steam reforming to generate hydrogen gas from methanol. This has implications for alternative point of use energy technologies. The projects are briefly outlined in the following sections.

1.6.1 Spvc – bacterial protein effector – classical and quantum mechanical study

Protein effectors regulate other proteins, switching them on, off, or changing their rates. Protein effectors can be part of the normal functioning of cellular processes or they can be foreign, released by pathogens to interfere with these processes.

The host response to bacterial infection involves a complex, highly conserved signal transduction pathway that consists of multiple protein kinases.[30] This pathway can be triggered by various stresses within the cell, including oxidative stresses and infection by pathogens. The cascade of protein activation is collectively known as the MAPK (mitogen activated protein kinases) signaling pathway.

By injecting protein effectors that disrupt the signal transduction cascade of a typical immune response, these pathogens can escape detection. Specifically, they dephosphorylate pMAPKs (the activated phosphorylated form of MAPK), preventing subsequent histone phosphorylation and transcription of pro-inflammatory chemokines that attract leukocytes to the infection site.[31]

Salmonella is a Gram negative, rod shaped motile bacterium that consists of over 2500 individual serovars, but the vast majority of cases in humans involve serogroup *S. enterica*. *Salmonella* causes salmonellosis, otherwise known as enteric or typhoid fever (*S. typhi*) as well a large number of cases of acute gastroenteritis, including a recent

2008 outbreak in the United States. Children and the immune-compromised are particularly vulnerable.[32]

This project is detailed in Chapter 5, but the main results are as follows.

1. Evaluated the enzyme-substrate complex using molecular dynamics to identify key residues including the general base and nearby polarization residues.
2. Determined approximate reaction barriers using a small molecule model to mimic the active site. In the context of the small model, the reaction was concerted.
3. QM/MM simulations were used to more thoroughly explored the reaction mechanism. With the addition of the enzyme environment, an intermediate appeared.

1.6.2 CDK2 – human cell cycle – a QM/MM study

Human Cyclin-dependent Kinase 2 (CDK2) is one of many protein kinases that regulate biological pathways through a process called *signal transduction*. This generally occurs through the transfer of a phosphate moiety, often originating from ATP, to a protein residue with a hydroxyl group, including serine, threonine, and tyrosine. This phosphorylation is shown schematically below.

$$ROH + PO_3 \rightarrow RPO_4 + H \qquad (1.20)$$

These signal transduction pathways are often complex, involving cascades of phosphoryl transfers from one protein to another, with many feedback mechanisms that can stop a signal at any point in the chain. The end of the chain generally involves phosphorylation

of a particular histone, which loosens the DNA strands in this region and allows for translation of these segments. In this way, cells can turn on or off various cellular processes, in response to stress (such as MAPKS) or simply in the normal lifecycle of the cell. CDK2 has been implicated in the cell cycle, including apoptosis (programmed cell death) and appears to have a specific role in meiosis (production of gamete cells). Regulation of this enzyme may have potential treatments in regards to cancer and infertility.[33]

There is some debate regarding the molecular steps involved in the transfer of the phosphate to the hydroxyl substrate. There is some support for a substrate-assisted mechanism, where the phosphate catalyzes its own transfer; however there is growing evidence in related enzyme family members that an aspartate group is playing the role of general base to facilitate the transfer.

This project is detailed in Chapter 6 but the main results are as follows:

1. Used an *ab initio* QM/MM MD approach to evaluate competing reaction paths, which showed strong evidence that the general base pathway is dominant over the alternative substrate-assisted pathway.
2. Used umbrella sampling to allow the system to fluctuate along the reaction path and determine the potential of mean force. This showed a general base pathway, with a barrier consistent with experiment.

1.6.3 Anthrax Lethal Factor – a semiempirical QM/MM study.

Anthrax is one of the oldest diseases known to man, but it has recently gained new attention due to several high profile incidents where an aerosolized version was used as a bioterrorism agent. While the more common cutaneous anthrax is curable, inhalation anthrax is far more dangerous, with significant mortality even with early treatment.[34] The bacterial infection can be treated effectively with antibiotics, but treatments are needed to deal with the virulence factors released into the circulatory system. The primary virulence factors are three polypeptides: protective antigen (PA), edema factor (EF), and lethal factor (LF), with PA serving to facilitate entry of either EF or LF into the cytosol of the cell.[35] Lethal factor is the more potent of the two, as injection of LF and PA lead to similar disease progression in animal studies as the bacterial infection.[36]

Though LF has been the subject of much empirical study, there had been no previous theoretical study of its catalysis due partly to the difficulty in modeling large active sites and metal ions in an efficient manner.

This project is detailed in Chapter 7 but the main results are as follows.

1. Investigated anthrax lethal factor catalysis using a DFT-based semi-empical QM/MM scheme (SCC-DFTB) required to properly treat the active site zinc metal ion. Evaluation of the Michaelis complex showed that the substrate is tightly held by both electrostatic and hydrophobic interactions.
2. The potential of mean force for the reaction was evaluated in a 1 ns simulation of two models: the wild type and a mutant where a key tyrosine residue is

changed to a phenylalanine. The mutant had a 4.5 kcal/mol higher barrier, indicating a key role for the tyrosine residue.

3. Proposed a mechanism of nucleophilic substitution of a glutamate general base activated water to a backbone carbonyl carbon, with the zinc ion and a neighboring tyrosine acting as an oxyanion hole to stabilize the charged carbonyl intermediate, followed by C-N bond breaking upon protonation by the same glutamate that did the proton abstraction.

1.6.4 Methanol Steam Reforming – Plane-wave DFT study

Proton exchange membrane (PEM) fuel cells are an efficient, clean, and robust approach to generating electricity for mobile applications, but storage of hydrogen (H_2) required for their operation remains a significant challenge. As a result, there has been much interest in circumventing hydrogen storage by on-board hydrogen generation, and one of the leading candidates for that is methanol steam reforming (MSR)[37]:

$$CH_3OH + H_2O \rightarrow 3H_2 + CO_2 \qquad (1.21)$$

Despite its activity and selectivity, the traditional Cu/ZnO catalyst used in this reaction has many drawbacks for mobile applications, including pyrophoricity, rapid degradation, and agglomeration through sintering.[37] Recent interest has centered on a Pd/ZnO catalyst, first identified by Iwasa and colleagues.[38] The new catalyst shows similar activity and selectivity to Cu/ZnO, but with vastly improved thermal stability.

Since the dehydrogenation of methoxy has been assumed to be the rate-limiting process,[39] there has been no study on the initial dissociation steps on PdZn surfaces. In

this work, we investigate the initial dissociation steps of both methanol and water and show that they are highly activated on flat PdZn surfaces.

This project is detailed in Chapter 8 but the main results are as follows.

1. Used planewave density functional theory (DFT) to model a representative selection of surface sites on a PdZn alloy with ZnO support. Energies and geometries of all relevant species were evaluated on clean, stepped, and kink surfaces.
2. Used a nudged elastic band approach to calculate minimum energy paths and barrier heights for the initial steps in MSR and showed that the initial O-H breaking reactions for both methanol and water are highly activated.

1.7 Publications

The project chapters in this work are based on collaborative work found in several publications, which are listed below. All results and figures from these publications are used with permission.

Chapter 5

Smith GK, Ke Z, Hengge AC, Xu D, Xie D, Guo H. Active-site dynamics of SpvC virulence factor from *Salmonella typhimurium* and density functional theory study of phosphothreonine lyase catalysis. *The Journal of Physical Chemistry B*. 2009; 113(46):15327-33.

Ke Z, **Smith GK**, Zhang Y, Guo H. Molecular mechanism for eliminylation, a newly discovered post-translational modification. *Journal of the American Chemical Society* 2011; 133(29):11103-05.

Chapter 6

Smith GK, Ke Z, Hengge AC, Guo H. Insights into the phosphoryl transfer mechanism of cyclin-dependent protein kinases from ab-initio QM/MM free-energy Studies. *Journal of Physical Chemistry B*. 2011; 115(46):13713-22.

Chapter 7

Smith CR, **Smith GK**, Yang Z, Xu D, Guo H. Quantum mechanical/molecular mechanical study of anthrax lethal factor catalysis. *Theoretical Chemistry Accounts*. 2010; 128(1):83-90.

Chapter 8

Smith GK, Lin S, Lai W, Datye A, Xie D, Guo H. Initial steps in methanol steam reforming on PdZn and ZnO surfaces: Density functional theory studies. *Surface Science*. 2011;605(7-8):750-759.

1.8 References

1 Berzelius, J. J. *Årsberättelsen om framsteg i fysik och kemi [Annual report on progress in physics and chemistry]*. p. 245 (Royal Swedish Academy of Sciences, 1835).

2 Robertson, B. A. J. B. & Thnard, L. J. The Early History of Catalysis. *Pans*, 64-69 (1975).

3 McQuarrie, D. A. & Simon, J. D. *Physical Chemistry - A Molecular Approach*. Ch. 28-29 (University Science Books, 1997).

4 Hinshelwood, C. N. *Kinetics of Chemical Change*. 4th edn, (Oxford University Press, 1962).

5 Swiegers, G. F. *Mechanical Catalysis*. (John Wiley & Sons, Inc, 2008).

6 Laidler, K. & King, C. Development of transition-state theory. *Journal of Physical Chemistry* **87** (1983).

7 Hinshelwood, C. N. The transition state method in chemical kinetics. *Journal of the Chemical Society*, 635-641 (1937).

8 Chandler, D. *Introduction to Modern Statistical Mechanics*. (Oxford University Press, 1987).

9 Sebatier, P. Hydrogénations et déshydrogénations par catalyse *Berichte der Deutschen Chemischen Gesellschaft* **44**, 1984-2001 (1911).

10 Balandin, A. Modern State of the Multiplet Theory of Heterogeneous Catalysis. *Advances in Catalysis* **19** (1969).

11 Voet, D. & Voet, J. G. *Biochemistry*. p472-494 (John Wiley & Sons, Inc., 2004).

12 Menten, L. & Michaelis, M. L. Die Kinetik der Invertinwirkung ("The Kinetics of Invertase Activity"). *Biochemische Zeitschrift* **49**, 333-369 (1913).

13 Lineweaver, H. & Burk, D. The determination of enzyme dissociation constants. *Journal of the American Chemical Society* **56**, 658-666 (1934).

14 Haldane, J. B. S. *Enzymes*. (Longmans, Green and Co., 1930).

15 Koshland, D. E. Applications of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 98-104 (1958).

16 Pauling, L. Molecular Architecture and Biological Reactions. *Chemical and Engineering News* **24**, 1375-1377 (1946).

17 Warshel, A. Energetics of enzyme catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 5250-5254 (1978).

18 Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chemical Reviews* **106**, 3210-3235 (2006).

19 Koshland, D. E. & Storm, D. R. A Source for the Special Catalytic Power of Enzymes: Orbital Steering. *Proceedings of the National Academy of Sciences of the United States of America* **66**, 445-452 (1970).

20 Bruice, T. C. A view at the millennium: the efficiency of enzymatic catalysis. *Accounts of Chemical Research* **35**, 139 (2002).

21 Page, M. I. & Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 1678 (1971).

22 Firestone, R. A. & Christensen, B. G. Vibrational activation I. A source for the catalytic power of enzymes. *Tetrahedron Letters* **389** (1973).

23 Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196-1202 (2003).

24 Hammes-Schiffer, S. & Benkovic, S. J. Relating protein motion to catalysis. *Annual Review of Biochemistry* **75**, 519 (2006).

25 Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295** (2002).

26 Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438** (2005).

27 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-A resolution. *Proceedings of the National Academy of Sciences of the United States of America* **100** (2003).

28 Williams, R. J. P. Are enzymes mechanical devices? *Trends in Biochemical Sciences* **18** (1993).

29 Moss, D. W. *Enzymes*. (Oliver and Boyd, 1968).

30 Dong, C., Davis, R. J. & Flavell, R. a. MAP kinases in the immune response. *Annual Review of Immunology* **20**, 55-72 (2002).

31 Arbibe, L. *et al.* An injected bacterial effector targets chromatin access for transcription factor NF-κB to alter transcription of host genes involved in immune responses. *Nature Immunology* **8**, 47-56 (2006).

32 Wallis, T. S. & Galyov, E. E. Molecular basis of Salmonella-induced enteritis. *Molecular Microbiology* **36** (2000).

33 Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews. Cancer* **9**, 153-166 (2009).

34 Dixon, T. C., Meselson, M., Guillemin, J. & Hanna, P. C. Anthrax. *New England Journal of Medicine* **341**, 815-826 (1999).

35 Ascenzi, P. *et al.* Anthrax toxin: a tripartite lethal combination. *FEBS Letters* **531**, 384-388 (2002).

36 Bossi, P. *et al.* Bioterrorism: management of major biological agents. *Cellular and Molecular Life Sciences* **63**, 2196-2212 (2006).

37 Palo, D. R., Dagle, R. A. & Holladay, J. D. Methanol steam reforming for hydrogen production. *Chemical Reviews* **107**, 3992 (2007).

38 Iwasa, N., Yamamoto, O., Akazawa, T., Ohyama, S. & Takazawa, N. Dehydrogenation of methanol to methyl formate over palladium zinc-oxide catalysts. *Chemical Communications*, 1322-1323 (1991).

39 Neyman, K. M. *et al.* Microscopic models of PdZn alloy catalysts: structure and reactivity in methanol decomposition. *Physical Chemistry Chemical Physics* **9**, 3470-3482 (2007).

Chapter 2

Classical Methods

Molecular Mechanics and Molecular Dynamics

"By convention there is color, by convention sweetness, by convention bitterness, but in reality there are atoms and the void"

-Democritus

2.1 Introduction

While modern computing and more sophisticated quantum methods have allowed researchers to apply quantum mechanics to systems ranging from 100-1000 atoms, many systems remain impenetrable to quantum techniques. Complex systems with thousands of atoms like enzymes, membranes, organelles, and even up to individual cells far exceed practical limits for a full quantum analysis, and so must be described with approximate methods. Molecular Mechanics (MM) is one of the simplest atomic scale approximations applied to these large systems and has been shown to be reliable for describing average properties of many systems near equilibrium.

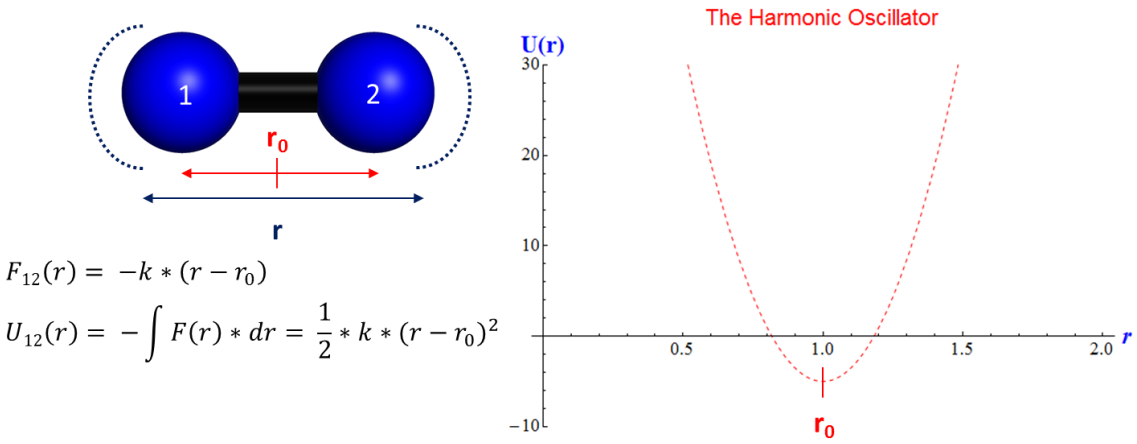
This chapter will describe the underlying principles on which molecular mechanics is based, followed by a discussion of its applications for studying molecular systems, and will conclude with how these principles intersect with individual project chapters in this dissertation.

2.2 Approximating Bonded Interactions[1]

A major constraint on the motion of any atom in a simulation is its bonded partners. Most MM schemes describe interactions up to 3 bonds away, and parameterize additional contributions. We start with the simplest case, the two body interactions or a single chemical bond.

2.2.1 Two Body Interactions

The harmonic approximation is one of the simplest descriptions of the chemical bond. The chemical bond is represented by a ball-and-spring model with equilibrium position r_0, current position r, and spring constant k describing the stiffness of the bond. The restoring force due to displacement is described by Hooke's law, and the potential energy function is derived as shown below.



$$F_{12}(r) = -k * (r - r_0)$$

$$U_{12}(r) = -\int F(r) * dr = \frac{1}{2} * k * (r - r_0)^2$$

Figure 2.1 – Force and potential terms for the two body bonded interaction with harmonic potential shown in the right panel.

Values of k and r_0 are bond parameters, and depend on the nature of the two atoms involved. For a diatomic molecule, this simple description describes the nature of the bond quite well for values of r near r_0, but breaks down near small and large values of r. An example of a more realistic potential, known as the Morse potential, is shown below in Figure 2.2 to contrast with the harmonic model. The two potentials overlap near the equilibrium position, but diverge at the two extrema, with the Morse potential flattening out near the dissociation limit and rising sharply at short distances.

For a harmonic potential to give reasonable results, the system must be near equilibrium. High pressure and high temperature simulations and those involving bond breaking will give artificial results and require a more complicated model.

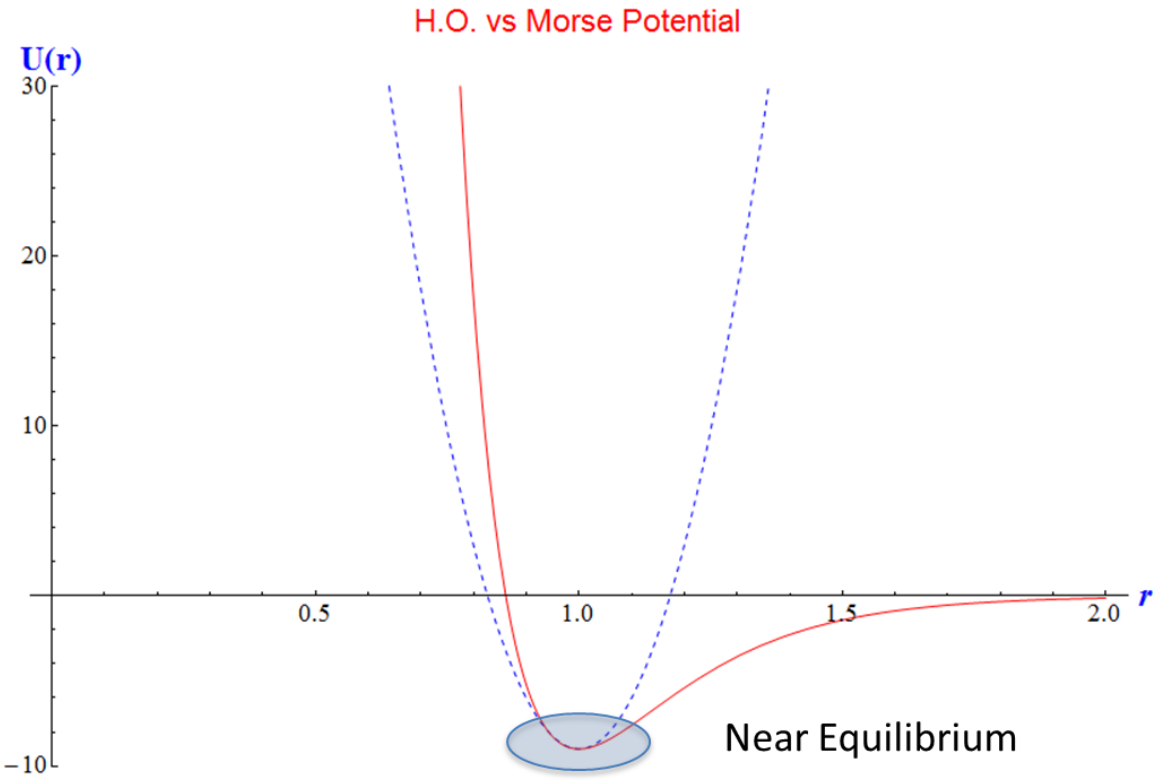
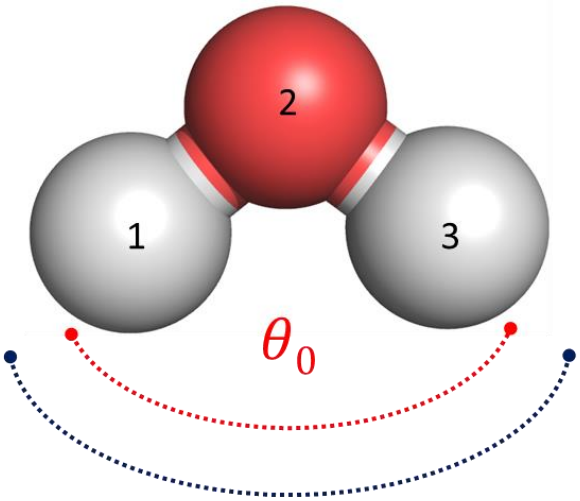


Figure 2.2 – Morse potential compared with the harmonic potential.

2.2.2 Three Body Interactions

If an additional atom is added to the molecule, each atom pair will contain a two body term, and now a 3-body harmonic term is needed based on the angle between the three atoms. As in the two body case, the terms k_θ and θ_0 are bond parameters that much be determined for each trio of atoms you wish to describe. Taking water as an example, the two body and three body terms are shown below.



$$U_{12}(r) = \frac{1}{2} * k * (r - r_0)^2$$

$$U_{23}(r) = \frac{1}{2} * k * (r - r_0)^2$$

$$U_{123}(\theta) = \frac{1}{2} * k_\theta * (\theta - \theta_0)^2$$

Figure 2.3 – The two 2-body terms and single 3-body term for a water molecule.

2.2.3 Four Body Interactions

If a fourth atom is added to our molecule, one must now consider 4 body interactions, commonly called dihedrals or torsions. The typical example involves protein backbone dihedral angles ϕ, ψ, and ω. The torsion potential for the $C_\beta - C_\alpha - C - O$ atom group is shown below. The torsion potential contains parameters to control the amplitude V, a periodicity parameter n, a phase or offset angle ξ, and the variable ψ describing the torsion angle. The functional form is chosen to allow hindered bond rotation with peaks and valleys depending on the bond constituents.

$$U_{1234}(\psi) = V * [\, 1 + \cos(n\psi \, - \, \xi) \,]$$

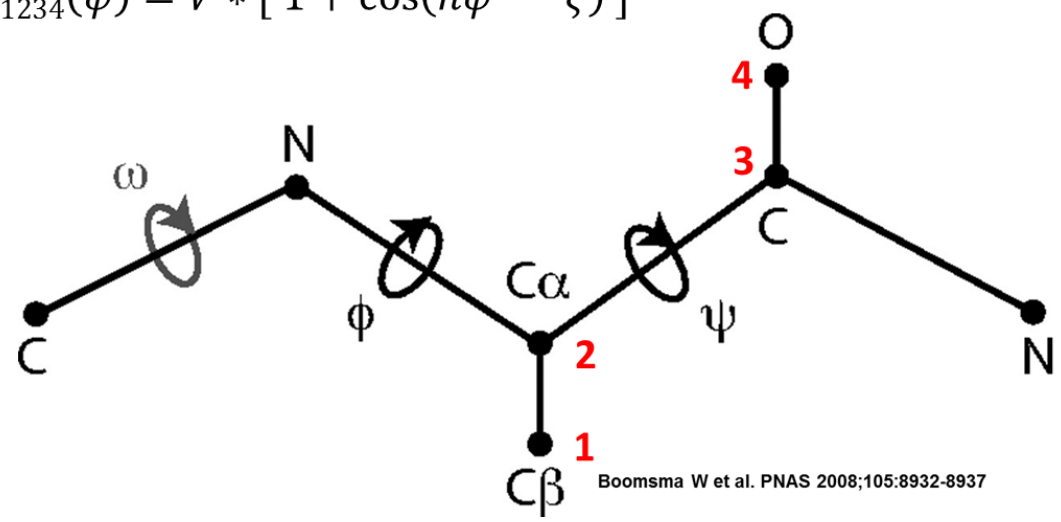


Figure 2.4 – Four body interactions critical for describing protein backbones. Figure adapted from Boomsma W et al. PNAS 2008; 105:8932-8937. Copyright 2008 National Academy of Sciences.

One could extend this treatment to 5-body and higher interactions, but in practice these interactions are small and get lumped into the 4-body term during the determination of parameters.

In practice these parameters are developed through a mix of high level quantum calculations and experimental data performed on a diverse set of organic molecules. If the system contains unique atom types or is significantly different than the training set in some way, it is generally necessary to develop custom parameters.

2.3 Non-bonded or Through Space Interactions

In addition to bonded interactions, one must consider that atoms near one another will interact both attractively and repulsively depending on their charge distributions and distances between atoms. The first element is a simple point charge model described by the Coulombic potential.

$$U_{12} = \frac{1}{4\pi\varepsilon\varepsilon_o} * \frac{q_1 q_2}{r} \tag{2.1}$$

The partial charges q_1 and q_2 are atom and environment specific and are determined from quantum calculations of many small molecules.

This static point charge model is woefully incomplete however for two major reasons. First, atoms are not point charges. They have a volume and consist of a very small positively charged nucleus surrounded by an electron cloud. Two nearby atom electron clouds of equal formal charge will repel, and if pushed together to nuclear distance, they will repel very strongly. The second reason is that charge distributions in atoms are not static or uniform in nature. Even in molecules without permanent dipoles, instantaneous fluctuations in atom charge density can lead to induced dipole interactions. These instantaneous fluctuations lead to what is known as London dispersion forces which fall off rapidly with distance. An empirical formula known as the Lennard-Jones (L-J) potential is shown below which addresses these two deficiencies.

$$U_{12} = D_e * \left[\left(\frac{r_e}{r}\right)^{12} - 2\left(\frac{r_e}{r}\right)^{6}\right] \tag{2.2}$$

Where D_e and r_e are fitted parameters that represent the dissociation energy and equilibrium distance respectively. One can see that when $r << r_e$, the 12th order term dominates the expression leading to a very rapidly rising energy to model the repulsion at short distances. When $r = r_e$, the expression reduces down to $-D_e$, or the depth of the well. When $r >> r_e$, the 12th order term vanishes quickly and the 6th order term dominates before quickly flattening towards 0, the point of dissociation. Figure 2.4 shows how both the Coulombic and L-J potentials interact. The relative contributions of both effects have been chosen ad-hoc to show general trends, rather than to represent a typical case.

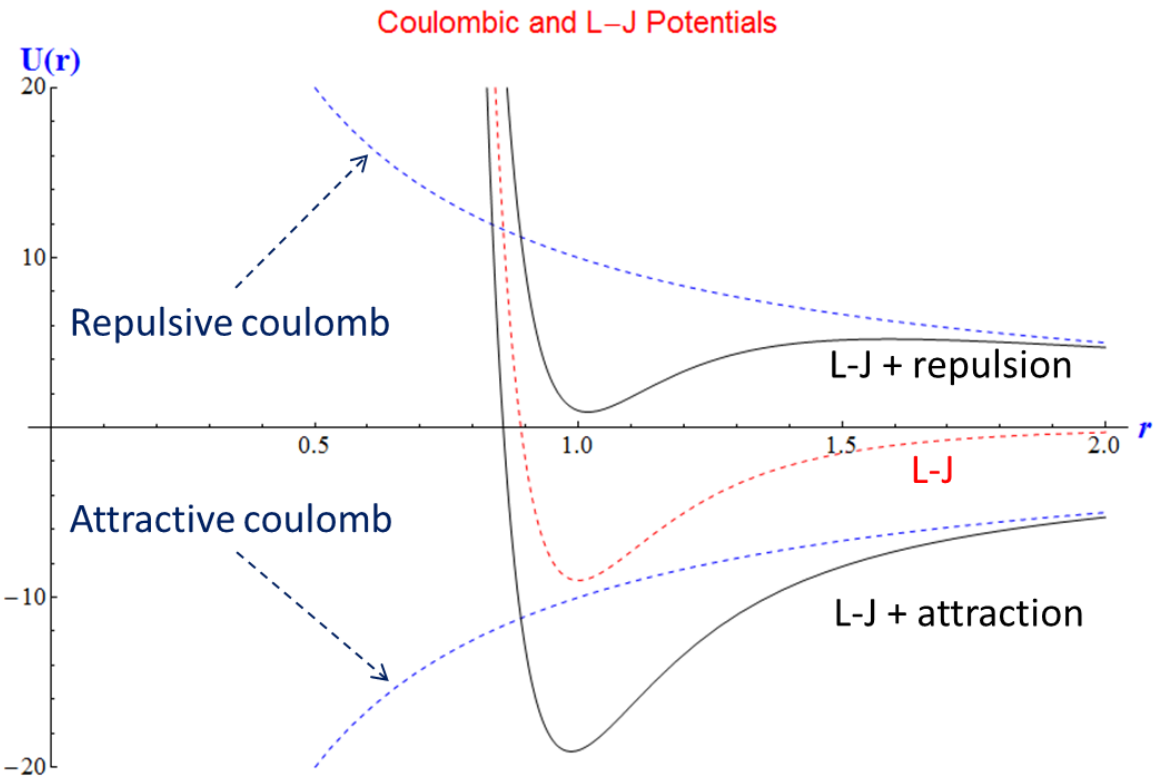


Figure 2.4 – Interaction of Lennard-Jones and Coulomb potentials

In practice, even using these simple models to describe through space interactions, calculating every pairwise interaction for systems with thousands of atoms can become expensive, so cutoffs are employed. The appropriate cutoff will depend on nature of your system, particularly the dielectric constant of the solvent. For highly polar solvents like water, a typical cutoff is on the order of 12-18 Å. Due to the charge screening offered by the solvent and surrounding residues, the pair wise interactions between atoms outside these cutoff distances should be negligible.

One last note should be made about a potentially significant effect that is not accounted for by the Coulombic or L-J terms, and that is polarization, the ability to induce a stable dipole-dipole interaction by perturbing the charge density of nearby atoms. Soft atoms like sulfur that are highly polarizable are particularly sensitive to this effect. The cumulative effect can be significant when dealing with large molecules interacting with one another, like one finds with many protein/protein binding events. Including polarization is an area of active research and many modern force fields do allow polarization effects to be included, but at increased computational cost.

2.4 The AMBER and CHARMM Force Fields

Now that the basic elements of molecular mechanics force fields have been explained, it's time to put these concepts together. From the perspective of a single atom, one can calculate the total potential felt by that atom by summing all the bonded and non-bonded interactions.

$$U_i = \sum U_{bonded} + \sum U_{non-bonded} \tag{2.3}$$

In describing the individual bonded and non-bonded terms, each contained 2 or more parameters. These included force constants, equilibrium positions, phase angles, and partial charges. The collection of all the parameters needed to fully describe a system along with the functional form of the potential is referred to as a force field. Many different force fields exist, with some suited specifically to biological systems like proteins, lipids, and DNA.

The two major force fields used in this work are the AMBER[2] and CHARMM[3] force fields. Both AMBER and CHARMM can refer to a set of parameters (the force-field) and to a suite of programs for running molecular simulations using these force fields.

The general form of the AMBER potential is as follows:

$$\begin{aligned} U_i^{AMBER} = & \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\psi - \xi)] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ D_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \end{aligned} \tag{2.4}$$

You can see this contains all the major features discussed previously, including bonded and non-bonded terms. The CHARMM force field adds two additional terms:

$$U_i^{CHARMM} = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\psi - \xi)] + \sum_{impropers} \frac{1}{2} k_\omega (\omega - \omega_0)^2 + \sum_{Urey-Bradley} \frac{1}{2} k_u (u - u_0)^2 + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ D_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (2.5)$$

The additional terms relate to maintaining planarity along protein backbones (impropers) and an additional harmonic term relating the distance between the 1st and 3rd atom in a 3 atom set referred to as the Urey-Bradley term. In general terms, both force fields produce reliable results for a wide variety of biological systems near equilibrium.

2.5 Solvent Models

Many systems exist in the condensed phase, surrounded by solvent and other solutes. For example, one of the major driving forces controlling protein folding and stability is the entropic cost of organizing water molecules around non-polar residues. The folded protein avoids this cost by desolvating the protein interior and allowing non-polar residues to interact with themselves, rather than water. While these effects cannot be neglected, their inclusion can be expensive and several techniques have been developed to include solvent while still allowing reasonable computational costs. There are two main approaches to treating solvent effects, explicit and implicit solvent models.

2.5.1 Explicit Solvent Models

A naïve way of simulating solvent is to simply immerse the solute in a bath of solvent molecules, which are treated using the same force field parameters as the solute. However, this approach has several disadvantages. First, the cost to simulate hundreds of thousands of solvent molecules is not trivial, and second, water is notoriously hard to simulate properly using standard force fields. The first problem can be addressed through various ways of dealing with the system boundary to mimic the effect of extended solvent while only simulating a small drop explicitly. The second problem led to the development of various simplified water models that are parameterized specifically to reproduce water's unique properties.

2.5.1.1 Periodic Boundary Conditions

The basic idea behind periodic boundary conditions (PBC) is fairly simple and derives from the periodicity of crystalline materials. Its use to describe extended solvent is an approximation. If one defines a standard cube and tracks a single particle as it approaches the edge of this cube from the inside. As the particle passes through the cube face on one side, it immediately appears from the opposite face with identical velocity. From an individual particles perspective, there is no cube, just an infinite expanse of space through which to travel. A two dimensional example is shown in Figure 2.5.

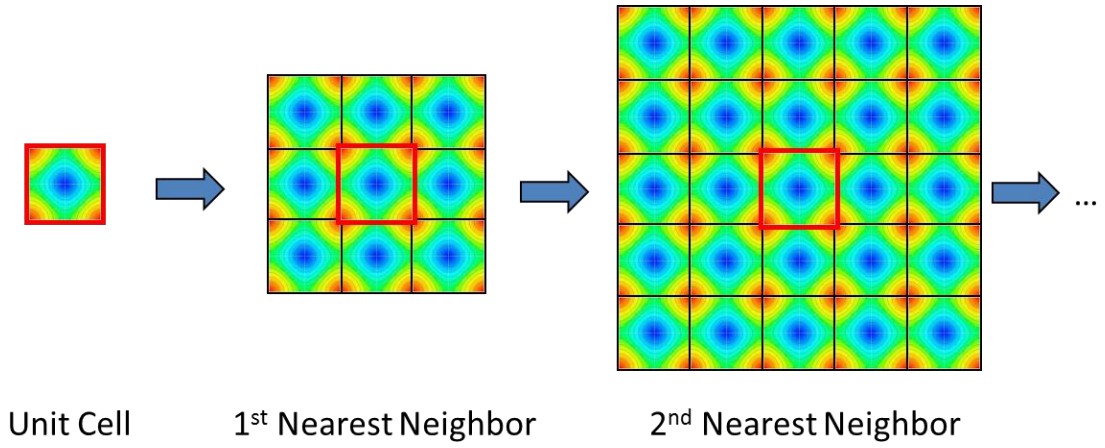


Figure 2.5 – A single unit cell can represent an extended surface using periodic boundary conditions

This creates the illusion of extended solvent, even though only the motions of the particles within the single cube are calculated explicitly. However, to calculate a correct potential on the unit cell, contributions from each of the images must be summed up. Nearby images can be included explicitly however since the surface spans an infinite extent, one must account for long range interactions. A rigorous way to calculate the electrostatic effects of these images is called the Ewald Sum.[4,5]

2.5.1.2 Stochastic Boundary Conditions

Stochastic boundary conditions[6] work by adding a spherical buffer region around the system which mimics the average effect of an extended solvent region. This often employs Langevin dynamics, which emphasizes macroscopic properties like viscosity to give a friction force (γ) and a random or stochastic force ($\vec{R}(t)$) to simulate the random kicks the system will get from the solvent shell at a given temperature.

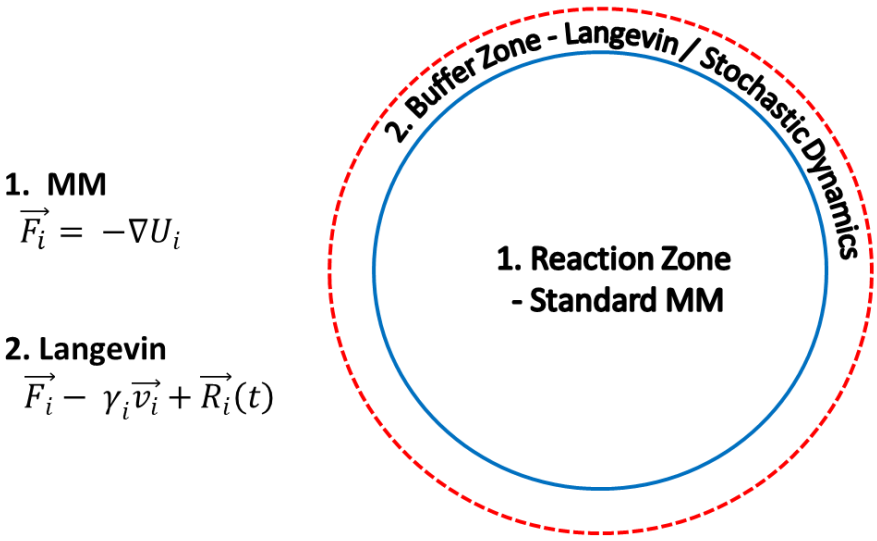


Figure 2.6 – Spherical boundary conditions introduces a Langevin term at the boundary to represent extended solvent.

This Langevin behavior is scaled up gradually at the boundary so that the simulation is subject to the average properties of the solvent while only explicitly simulating a small spherical drop.

2.5.1.3 TIP3P and other water models

Explicit water models generally differ by the number of interaction sites, their rigidity, and whether they allow for polarization effects. As might be expected, each improvement in accuracy carries an associated computational cost.

The TIP3P[7] model is fairly simple, consisting of a rigid water structure with 3 interaction sites. Since there are no internal degrees of freedom, these are fairly easy to calculate and still produce acceptable properties in molecular dynamics simulations. One reason for the continuing popularity of TIP3P is that the AMBER and CHARMM force fields were

developed using this model, so many of the drawbacks of TIP3P are compensated by the individual force field parameterizations.

2.5.2 Implicit Solvent Models

Implicit solvent models treat the solvent as a continuum or mean field, characterized by the dielectric constant of the solvent and the amount of solute exposed to solvent. Though this can be treated rigorously using the Poisson-Boltzmann equation, difficulty in solving this equation efficiently has led to the development of approximations, the most popular of which is the Generalized Born Model and its variants.[8,9]

2.6 Newtonian Mechanics and Minimization Methods

Given a molecular system with well-defined atomic coordinates and a set of force field parameters, one can calculate the potential felt by each atom in any configuration. With the potential one can calculate the force.

$$F_i = -\nabla U_i \tag{2.6}$$

If the net force on an atom is known, the atomic coordinates can be updated in the direction of the force, or in other words, in the direction of lower energy on the potential surface. For simple systems like a single water molecular, one can simply iterate every possible coordinate and develop a full potential energy map along the two bond coordinates r_1 and r_2 and the angle θ. The trouble with large systems is you have 3N degrees of freedom for each atom and the potential energy surface is unlikely to increase or decrease monotonically, but rather consists of peaks and valleys along a

multidimensional surface. The way to a true minimum may be blocked by multiple local minima surrounded by high energy regions. These surfaces are also very difficult to visualize unless you can identify coordinates of particular importance or develop composite coordinates related to the phenomena you are concerned with. Figure 2.7 shows a surface displayed along a constrained coordinate space and shows multiple minima and minimal energy paths between them.

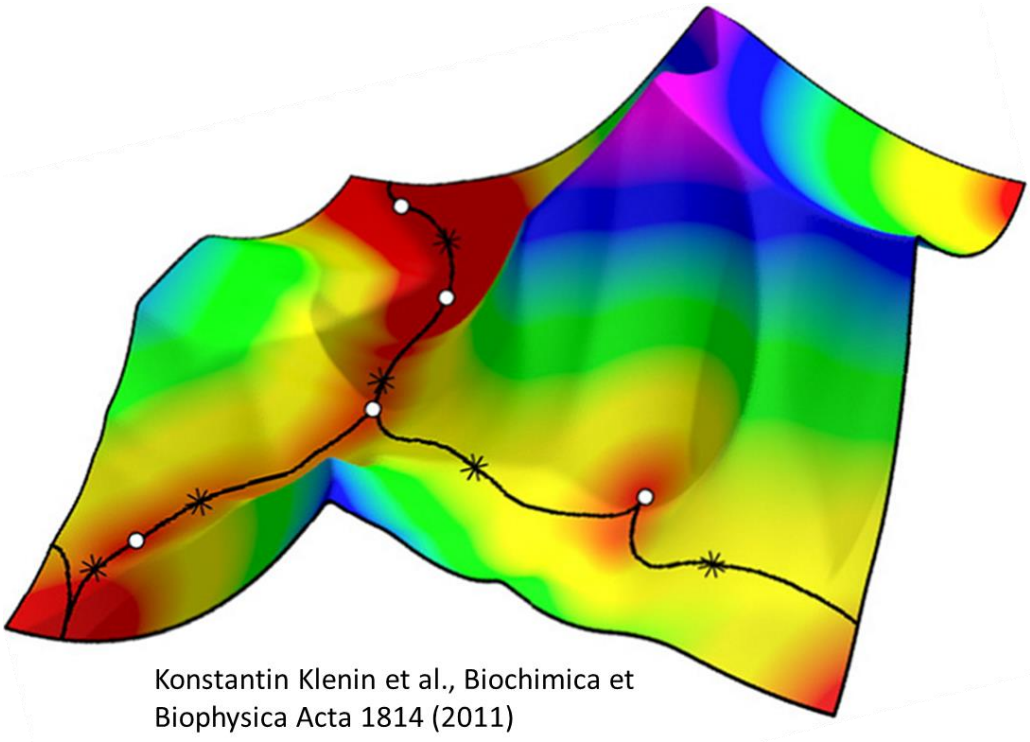


Figure 2.7 – An example potential energy surface showing minimal energy pathways connecting stationary points.

Reprinted from Klenin K, Strodel B, Wales DJ, and Wenzel W. Modelling proteins: Conformational sampling and reconstruction of folding kinetics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2011; 1814(8). Copyright 2011 with permission from Elsevier.

Depending on the start point for this surface, following the force to a minimum can end up in any number of local minima on the surface. Since the shape of the surface a priori is not known, one of the key problems in molecular simulations is how to sample these often very complex potential energy surfaces.

The first problem is given an initial starting configuration, how is a local minimum found? This is collectively referred to as minimization and there are many different approaches to this problem, but this discussion will only expound on two methods, steepest decent (SD) and Newton-Raphson (NR).

2.6.1 Steepest Descent

Steepest descent simply updates coordinates in the direction of the energy gradient over some step size. Each new position x_{n+1} depends on the initial position x_n and the force in that direction scaled by some reasonable fraction of the system size α.

$$x_{n+1} = x_n + \alpha * -\frac{\partial U(x,y,z)}{\partial x}$$
$$y_{n+1} = y_n + \alpha * -\frac{\partial U(x,y,z)}{\partial y} \qquad (2.7)$$
$$z_{n+1} = z_n + \alpha * -\frac{\partial U(x,y,z)}{\partial z}$$

This process is then repeated till some energy or force criteria stops changing. Steepest descent works well when you are far away from the minimum, but can be slow to converge as you approach the minimum, especially if your step size is too large.

2.6.2 Newton-Raphson (NR) and more advanced minimization schemes

The Newton-Raphson technique leverages the first derivative of whatever function you are trying to minimize in a fairly simple way. Consider the following figure.

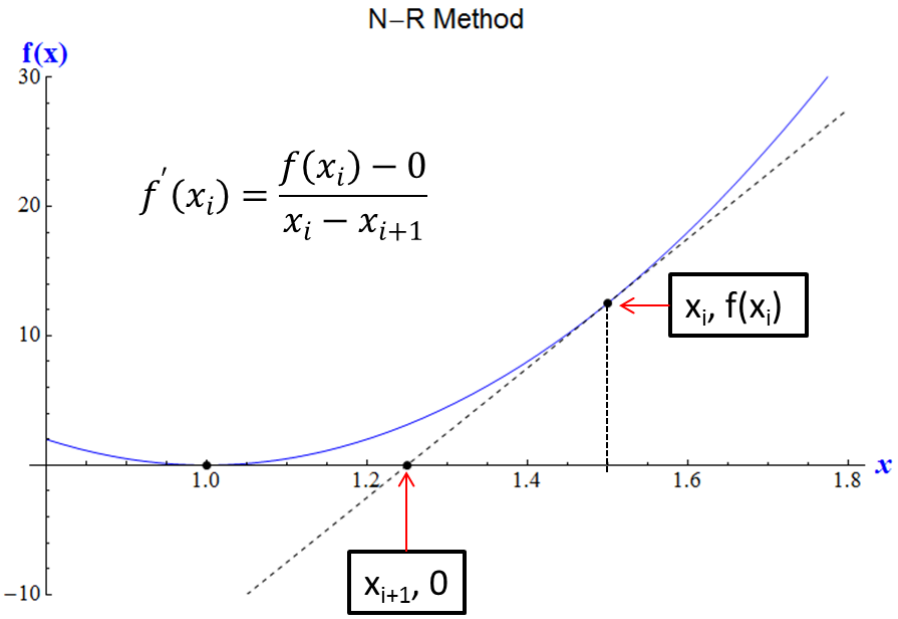


Figure 2.8 – Schematic depiction of the Newton-Raphson method

Here the minimum of the harmonic function is obvious by visual inspection. However, if this shape was not known in advance, one can only find minima by testing individual points in an iterative way (which is exactly the case for potential energy surfaces). The NR method starts with an initial guess x_i and evaluates the function at the point $f(x_i)$. One then draws a tangent line at this point representing the first derivative of the function $f'(x_i)$. This tangent line intersects the x-axis at some new point x_{i+1}. This creates a triangle with which one can define the first derivative $f'(x_i)$ as a simple rise over run.

$$f'(x_i) = \frac{f(x_i) - 0}{x_i - x_{i+1}} \tag{2.8}$$

Rearranging,

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{2.9}$$

For a potential energy function, the first derivative (or the force) is 0 at a minimum. This means it takes an initial guess and the first and second derivatives of the potential energy to iterate Newton-Raphson.

$$x_{i+1} = x_i - \frac{\partial U(x_i)/\partial x}{\partial^2 U(x_i)/\partial x^2} \tag{2.10}$$

This can be iterated until the energy and force are converged. In general, N-R converges much faster than steepest decent when near a minimum, but can be unreliable when far away from that minimum. Many minimization schemes use a combination of the techniques as a result. Steepest decent for initial minimization, then NR is used to converge quickly to the final minimum. AMBER and CHARMM both utilize variants of these methods for minimization, in addition to using more advanced optimizers like the Conjugate Gradient technique.

2.7 Molecular Dynamics[1]

While minimization techniques operate on the potential energy alone, molecular dynamics adds the kinetic energy back in and allows the system to fluctuate over time. This is based on simple Newtonian mechanics, but dealing with things like the

appropriate time step, method of propagation for trajectories, assignment of initial velocities and temperature control each require some explanation.

2.7.1 Propagation Algorithms

The general problem of propagating trajectories is given some initial position and velocity, what are the new positions and velocities at time $t + \Delta t$? Since the time step is not infinitely small, there will be some error introduced over each step of a trajectory, and many different schemes have been developed to reduce this error while minimizing the storage requirements and calculation time needed for each step. A few examples are discussed below.

2.7.1.1 The Verlet Algorithm[10]

Given an initial position $x(t)$ and velocity $v(t)$, the simple Verlet scheme updates $x(t)$ over a positive and negative timestep using the 2nd order kinematics equation.

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{F_x}{2m}\Delta t^2$$
$$x(t - \Delta t) = x(t) - v(t)\Delta t + \frac{F_x}{2m}\Delta t^2 \quad (2.11)$$

Combining these equations one can remove the velocity term, leading to the following iterator:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{F_x}{m}\Delta t^2 \quad (2.12)$$

Velocities are computed using a simple mean value formula:

$$v(t + \Delta t) = \frac{x(t) + x(t + \Delta t)}{2\Delta t} \quad (2.13)$$

These two formulas are then iterated to generate a trajectory. The major error in this scheme is in the velocity calculation which leads to a global error that is proportional to Δt.

2.7.1.2 The Velocity Verlet Algorithm[11]

Given an initial position $x(t)$ and velocity $v(t)$, the velocity Verlet algorithm starts similarly.

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{F_x(t)}{2m}\Delta t^2 \quad (2.14)$$

With this new position, one can calculate a new force $F_x(t + \Delta t)$ to update the velocity.

$$v(t + \Delta t) = v(t) + \frac{F_x(t) + F_x(t + \Delta t)}{2m}\Delta t \quad (2.15)$$

These two formulas are then iterated. Due to the improved velocity calculation, the global error is reduced to Δt^2.

2.7.1.3 The Beeman Algorithm[12]

The Beeman algorithm is more complex and uses a predictor-corrector scheme to propagate trajectories. This algorithm also requires more starting information: $x(t)$, $v(t)$ and $x(t - \Delta t)$. The Beeman algorithm usually employs one iteration of Velocity Verlet to generate $x(t - \Delta t)$ and $x(t)$ before starting the main loop.

First start with the prediction step:

$$x^P(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{6}\left[\frac{4F_x(t)}{m} - \frac{F_x(t-\Delta t)}{m}\right]\Delta t^2 \tag{2.16}$$

Then apply a correction based on the force evaluation from the prediction.

$$\begin{aligned} x(t+\Delta t) &= x^P(t) + v(t)\Delta t + \frac{1}{6}\left[\frac{F_x^P(t+\Delta t)}{m} + \frac{2F_x(t)}{m}\right]\Delta t^2 \\ v(t+\Delta t) &= v(t) + \frac{1}{6}\left[\frac{2F_x(t+\Delta t)}{m} + \frac{5F_x(t)}{m} - \frac{F_x(t-\Delta t)}{m}\right]\Delta t^2 \end{aligned} \tag{2.17}$$

These three steps are then iterated to propagate the trajectory. While the Beeman Algorithm increases the storage requirements and computational cost for each iteration, the tradeoff is the error is now proportional to Δt^3 so that a larger time step can be used.

2.7.2 Choosing a Time Step

The general rule is the time step should be an order of magnitude lower than the characteristic time scale of any motions you wish to study, though this restriction can be relaxed for higher order propagators. The fastest motions in most molecular systems are hydrogen stretching frequencies with heavy atoms (X-H) which fall in the femtosecond regime (10^{-15} s), whereas heavy atom stretching falls into the 10-20 fs regime. If one restricts the time step to ~10% of the characteristic motion time that would give a time step of 0.1 fs for X-H motions and 1-2 fs for studying heavy atoms. Since hydrogen motion in macromolecules is often not all that interesting, a technique has been developed to restrict hydrogen motion relative to their bonded atoms. The original formulation of this technique, called SHAKE[13], applies an internal constraint to the motion of X-H pairs by use of Lagrange multipliers. As a result, a longer time step (1-

2 fs) can be used in hydrogen containing systems even with simple propagators such as Verlet.

2.7.3 Initial Velocities

In most molecular dynamics simulations, one starts with some known set of three dimensional initial coordinates, often from x-ray crystallography or a similar technique. These structures generally do not have any velocity information and yet each of the algorithms discussed previously require knowing an initial velocity. Assuming 0 Kelvin conditions, one can just start the atoms with a velocity of 0 and use the force evaluations to generate $v(t + \Delta t)$. However, for studying systems at more relevant temperatures, assigning realistic initial velocities is needed. These initial velocities can be chosen from the Maxwell distribution.

$$f(s) = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} s^2 e^{-E_K/kT} \tag{2.18}$$

Here, f(s) refers to the fraction of atoms at a particular speed, k is the Boltzmann constant, T is the temperature, and E_K refers to kinetic energy. One can include the potential as well to create the Maxwell-Boltzmann distribution function. For a given temperature, random numbers are drawn from this distribution to assign initial velocities to your molecular system. For computer algorithms, this will change based on the starting seed, so using a new random seed each simulation is important to show convergence and that the results are not just a one-off event.

2.7.4 Ensembles and Thermostats

Simulation conditions generally fall into three statistical ensembles, the Microcanonical ensemble (NVE), the Canonical ensemble (NVT), and the Isobaric-Isothermal ensemble (NPT). Each is characterized by which variables are held constant. The Canonical ensemble, which is the most commonly used ensemble in this work, holds the number of particles (N), volume (V), and the temperature (T) constant. The bench top analogy to the canonical ensemble is a sealed test tube with a heating/cooling element to regulate the temperature.

While controlling N and V are fairly simple matters, controlling temperature is significantly more complex in MD simulations, since while Newtonian dynamics is energy conserving, it is not temperature (Kinetic Energy) conserving. The various methods of temperature regulation are collectively known as thermostats, and some of the more commonly used methods include velocity scaling as proposed by Andersen, and heat bath approaches developed by Nosé-Hoover and Berendsen. This brief discussion will primarily focus on the latter two approaches.

2.7.4.1 The Nosé-Hoover Thermostat[14,15]

The basic approach of Nosé-Hoover is the addition of new degrees of freedom in the form of a heat bath term in the system Hamiltonian.

$$\mathcal{H}(\vec{q},\vec{p},p_s,s) = \sum_i \frac{p_i^2}{2m\,s^2} + \frac{1}{2}\sum_i U(q_i) + \frac{p_s^2}{2Q} + gkT\ln(s) \qquad (2.19)$$

By coupling the heat bath parameters with the system kinetic energy, energy can be exchanged to allow temperature regulation to occur. The benefit of Nosé-Hoover is it formally samples the Canonical ensemble for the real system.

2.7.4.1 The Berendsen Thermostat[16]

The Berendsen approach is similar in concept, but simpler in implementation. The real system is coupled to a bath that exhanges energy at a rate based on the difference between the current temperature and the reference temperature.

$$\frac{dT(t)}{dt} = \frac{(\,T_0 - T(t)\,)}{\tau} \tag{2.20}$$

The parameter τ can be adjusted to control the coupling of the bath and the system. While the Berendersen thermostat does not formally yield the Canonical ensemble, for large systems the differences are negligible, so it is often employed due to its efficiency.

2.7.5 Hydrogen atoms and protonation states

X-ray structures, the starting point for many molecular simulations, do not contain information on hydrogen atom positions, and so these must be added back in before simulating. For small molecules with no titratable sites, this is usually a simple matter of adding hydrogen atoms to complete the valency of any heavy atoms. For macromolecules like proteins, not only are there titratable sites which may be in any number of possible protonation states, the polymer chain folds back on itself so blind hydrogen placement can generate significant steric clashes and resulting high energy structures.

For hydrogen placement, the CHARMM and AMBER suites both have routines that do a reasonable job in this area, as long as the user performs minimizations with all heavy atoms constrained to allow placed hydrogens to re-adjust based on their local environments.

Assigning protonation states of titratable groups is a harder problem, though in practice several methods do exist to assist user choices. First, it should be noted that in real systems, a dynamic ensemble of protonation states exist, rather than a single choice. To completely model this reality would result in a very large increase in computational cost. In practice, the user attempts to select a one or more of the dominant protonation states only and neglects the minor configurations. If a titratable group has a pKa either far above or far below the desired experimental conditions, the choice is generally clear. Protein residues like Histidine, with a pKa of 6.0, require a more careful analysis. At least 4 pieces of information should be considered for these hard to determine protonation states: experimental data, local environment, rule-based prediction tools, and physics based prediction tools.

First, if the group in question is involved in chemical catalysis, activity versus pH profiles can often shed light on the dominant protonation state of a group. Second, careful visual examination of surrounding residues can often make certain protonation states less likely due to steric or charge clashes, or more likely due to complementary residues that are nearby. Third, rule based tools such as PropKa[17] can often perform a similar role as visual inspection, but with less internal bias. Lastly, physics based tools that

solve the Poisson-Boltzmann equation can also make predictions of protonation states, with the major drawback being they are not as robust and can be quite sensitive to the starting configuration of your system.

2.7.6 RMSD

Since the major approximations in molecular dynamics are based on systems near equilibrium, some measure or indicator of when the system is relatively stable and hence near equilibrium is needed. The most commonly used measure is called the Root Mean Square Deviation (RMSD).

$$RMSD\ (\vec{q},\vec{r}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\ q_i - r_i)^2} \tag{2.21}$$

Here, q refers to the current coordinates of the system, r refers to the coordinates of the same system at some reference point (often t=0), and N refers to the total number of atoms. The profile of RMSD over time indicates both the absolute value of this deviation, and more importantly, the fluctuation, which is a prime indicator of having reached equilibrium. A representative profile is shown below in Figure 2.9, indicating key regions.

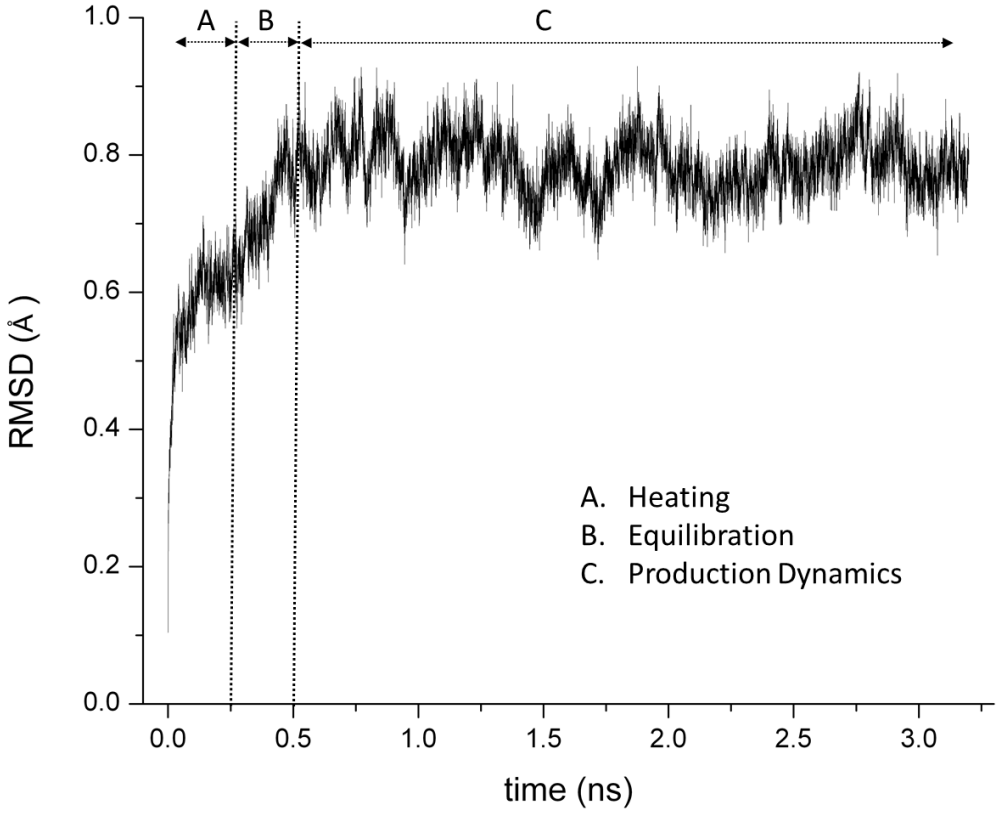


Figure 2.9 – An example molecular dynamics run showing the system RMSD stabilizing.

2.7.7 Heating, Equilibration and Production Dynamics

For complex systems, though it is possible to assign initial velocities from the Maxwell distribution for any desired temperature, in practice it is better to only use this strategy at low temperatures and then gradually heat the system to whatever temperature you desire. Typically you start the simulation at a very low temperature (0-50K) and over the course of 200-500 ps, gradually increase the thermostat to the desired conditions. Evaluation of the RMSD can indicate if heating is too slow or two fast. Once heating has concluded, the system is allowed to fluctuate until the RMSD starts to flatten out,

indicating the system has reached equilibrium. The trajectory data after this point can be used for analysis.

2.8 Average Properties from Trajectories

Once the system has reached equilibrium, properties can be calculated from the time average over the course of the simulation. This is simply expressed as follows:

$$A_{AVG} = \lim_{\tau \to \infty} \frac{\int_0^\tau A(\vec{p}, \vec{q}) dt}{\tau} \tag{2.8}$$

Here, A_{AVG} is a property of the system that can be calculated based on the system configuration $\vec{q}$ and momentum $\vec{p}$. The total time τ is shown in the infinite limit, but in practice the simulation is run until a desired property reaches convergence. For example, let's say one wants to know the average distance between an enzyme catalytic residue and a substrate molecule. Here, A would just be the sum of that distance calculated over each step divided by the total number of steps. Any property that can be calculated from the configuration and momentum can be measured from the simulation in this fashion.

A brief word is warranted on why this generally works. If a system is ergodic, then given enough time, a system trajectory will traverse all of the available low energy configurations, just as if you had arranged an ensemble of every possible configuration. In this way, the ensemble average, that is the sum of all possible configurations weighted by a Boltzmann factor, will be functionally equivalent to a time average from a long simulation. Since it is very difficult to know if a system shows ergodic behavior

over the time course of a simulation a priori, it should be noted that this time and ensemble average equivalency is not always true and counter-examples can be found. One example would be two wells separated by a very large barrier as shown in Figure 2.10. If the trajectory starts in one of the wells, though it may be possible to get a random kick to move to the other well, this is very unlikely to occur in the time scale of a reasonable simulation. Enhanced sampling methods like Replica Exchange Molecular Dynamics[18,19] or Simulated Annealing[20] can help address these issues.

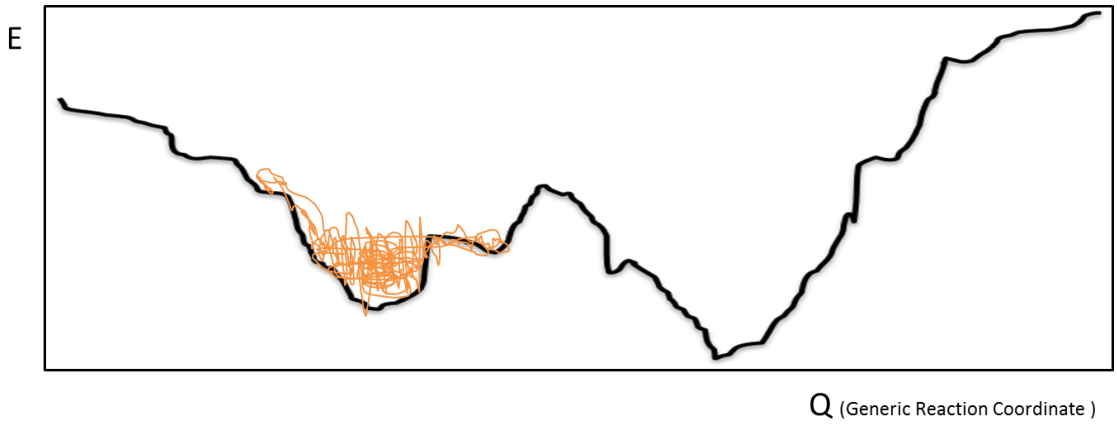


Figure 2.10 – A dynamics run can stay limited to a single well depending on the shape of the potential.

In addition to average properties, often times the dynamical behavior of a property is of more general interest. Figure 2.11 shows several distances inside an enzyme active site over the course of a simulation.

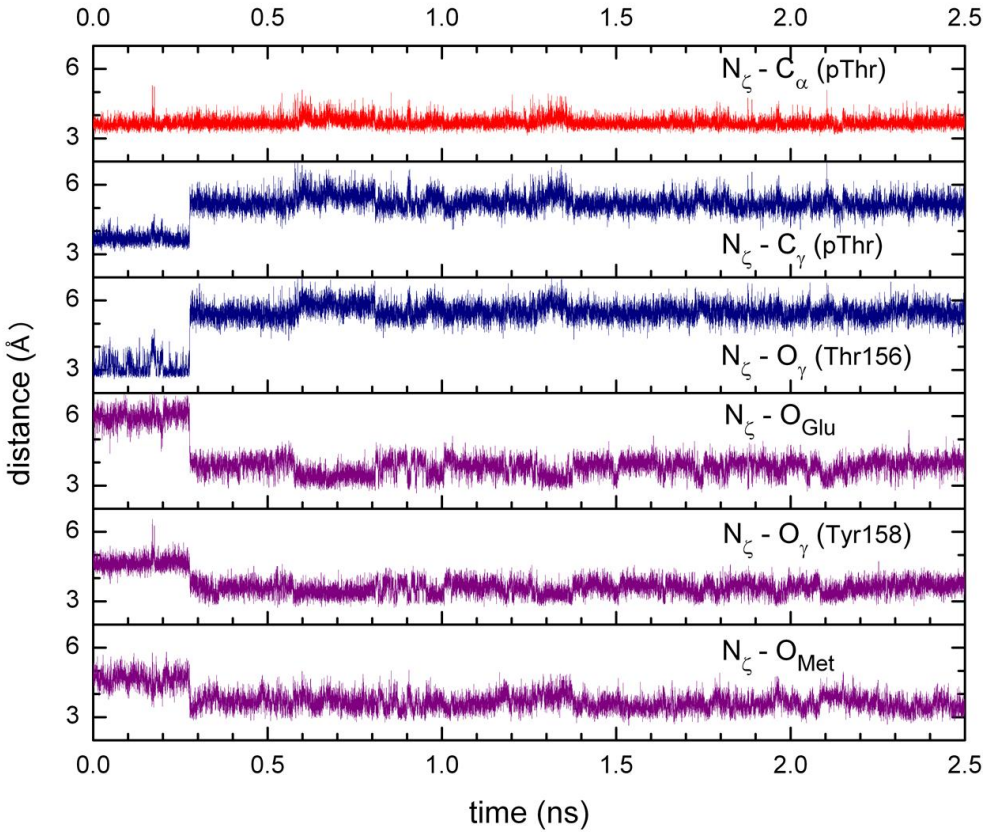


Figure 2.11 – Atom pair distances over the course of a molecular dynamics simulation.[21]

This data shows a conformational change happening in the active site of a bacterial enzyme. The 1st distance is unchanged during this conformational change, while the 2nd and 3rd increase in total distance, and the 4th through 6th decrease in total distance. The time course allows us to see what movements are correlated, anti-correlated and unaffected during a conformational change.

2.9 Classical Methods used in this Work

The classical methods outlined in this chapter were used most heavily in the study of large systems, particularly enzyme complexes. Many of the techniques describe here were also used in a hybrid fashion with quantum mechanical methods, which is the subject of the next Chapter.

Chapter 5 details the study of a bacterial enzyme called Spvc, in which the CHARMM force field was used to study the stability of the enzyme-substrate complex. Using the velocity verlet integrator and the Nose-Hoover thermostat, 2.5 nanoseconds of dynamics was collected and used for analysis. More technical details can be found in the methods section of Chapter 5.

Chapter 6 details the study of a human enzyme called CDK2, where the AMBER force field was used for initial system minimization and dynamics, before switching to a hybrid QM/MM technique that partitions the system into a reaction zone (described by Quantum Mechanics) and a buffer zone (described by Molecular Mechanics). This allowed the study of reaction paths and free energy profiles of reactions involved in this enzyme. More technical details can be found in the next Chapter on hybrid methods and in the methods section of chapter 6.

Chapter 7 details a study involving anthrax Lethal Factor (LF), which is an enzyme responsible for much of the damage done when anthrax bacteria infects the lymphatic system. This project used a hybrid QM/MM technique, with the CHARMM force field to describe the classical region, and a semi-empirical quantum method called SCC-DFTB to

describe the reaction zone. More technical details can be found in the next Chapter on hybrid methods and in the methods section of Chapter 5.

2.10 References

1 Leach, A. R. *Molecular Modeling*. (Prentice Hall, 2001).

2 Case, D. A. *et al.* The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26**, 1668-1688 (2005).

3 Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187-217 (1983).

4 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an NlogN method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089-10092 (1993).

5 Ewald, P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annals of Physics* **64**, 253 (1921).

6 Brunger, A., Brooks III, C. L. & Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* **105**, 495 (1984).

7 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* **79**, 926 (1983).

8 Feig, M. & III, C. L. B. Recent advances in the development and applications of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology* **14**, 217-224 (2004).

9 Bashford, D. & Case, D. A. Generalized Born models of micormolecular solvation effects. *Annual Review of Physical Chemistry* **51**, 129 (2000).

10 Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-103 (1967).

11 Swope, W. C. & Andersen, H. C. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *Journal of Chemical Physics* **76**, 637 (1982).

12 Beeman, D. SOME MULTISTEP METHODS FOR USE IN MOLECULAR-DYNAMICS CALCULATIONS. *Journal of Computational Physics* **20**, 130 (1976).

13 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).

14 Nose, S. A unified formulation of the constant temperature molecular-dynamics methods. *Journal of Chemical Physics* **81**, 511-519 (1984).

15 Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31**, 1695-1697 (1985).

16 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **81**, 3684-3690 (1984).

17 Bas, D. C., Rogers, D. M. & Jensen, J. H. Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes. *Proteins* **73**, 765-783 (2008).

18 Swendsen, R. & Wang, J. Replica Monte Carlo simulation of spin glasses. *Physical Review Letters* **57**, 2607-2609 (1986).

19 Zhou, R. Replica exchange molecular dynamics method for protein folding simulation. *Methods in molecular biology (Clifton, N.J.)* **350**, 205-223 (2007).

20 Kirkpatrick, S. & Gelatt, C. D. Optimization by Simulated Annealing. *Science* **220**, 671-680 (1983).

21 Smith, G. K. *et al.* Active-site dynamics of SpvC virulence factor from Salmonella Typhimurium and density functional theory study of phosphothreonine lyase catalysis. *Journal of Physical Chemistry B* **113**, 15327 (2009).

Chapter 3

Quantum Methods
Small Models and Hybrid Techniques

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."

-Paul Dirac, 1929, Proceedings of the Royal Society of London

3.1 Introduction

The previous chapter detailed how to model large systems from an atomistic point of view, using classical mechanics to approximate the motions of systems near equilibrium. One significant drawback to the molecular mechanics paradigm is it does not allow for describing the breaking and forming of chemical bonds. To describe bonding we need to model the behavior of electrons and nuclei, which is the natural domain of quantum mechanics. Though computational methods and computing power have improved dramatically since Dirac, his quote on the complexity of solving the wave equation for large systems still applies. In his time large systems referred to heavy atoms and medium size molecules, where today the limit of pure quantum calculations is typically a few thousand electrons when employing a Hartree-Fock (HF) or Density Functional Theory (DFT) approach.

So given these limitations, how does one simulate a reaction occurring in the active site of a large biomolecule like a protein? Since this is still too challenging for pure quantum approaches, some approximate methods must be employed. One option is two separate simulations, the first of which describes the whole system using classical methods as described in the previous chapter, and a separate quantum model that contains the reaction site and the first shell of atoms surrounding where the chemistry is occurring, allowing a reasonable calculation time. Though each simulation provides useful information, the simulations are not coupled and so any reaction phenomena that depend on the coupling between the reaction site and the surrounding system will be captured poorly with this approach. One way to address this problem is to use a

hybrid simulation, where quantum mechanics is used at the reaction site and the surrounding system is simulated classically, but the systems are coupled together.

This chapter will begin with a very brief introduction to quantum mechanics, with an emphasis on the approximations used in computation. Next, the practical details of setting up a small model system for pure quantum studies will be explored, which will be followed by a discussion of hybrid or QM/MM schemes. Particular emphasis will be placed on the methods used in this work.

3.2 Quantum Mechanics

The most important equation in quantum chemistry is the Schrödinger equation, formulated by Erwin Schrödinger in 1926.

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\psi(x,t) = i\hbar\frac{\partial\psi(x,t)}{\partial t} \tag{3.1}$$

Schrödinger drew on De Broglie's matter wave hypothesis and sought to find a differential equation similar in form to the classical wave equation. This equation is essentially an expression of energy conservation, and states that the configuration of the system $\psi(x,t)$ evolves over time depending on all the kinetic $-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}$ and potential $V(x)$ energy terms operating on the system. The biggest difference from the classical world is our system configuration is now defined by a superposition of waves rather than discrete point particles. It turns out this small difference has far ranging consequences.

For atoms and molecules, chemists are generally concerned with stationary states, described by the Time Independent Schrödinger Equation (TISE).

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\psi(x) = E\,\psi(x) \tag{3.2}$$

This is often shortened to simply,

$$\mathcal{H}\,\psi = E\,\psi \tag{3.3}$$

In the nomenclature of quantum mechanics this is an eigen-equation, where $\mathcal{H}$ is the energy operator (usually called the Hamiltonian) which represents a measurement, ψ is the eigenfunction of that operator, and E is the energy eigenvalue or observable.

Much like a vibrating string, where allowed frequencies can be expressed by whole number ratios, this equation can be satisfied by a series of wavefunctions that differ only by an integer value called a quantum number, where you have a quantum number for each degree of freedom in the system.

$$\mathcal{H}\,\psi_n = E_n\,\psi_n \tag{3.4}$$

If one knows the exact wavefunction and Hamiltonian for our system, every energy measurement should produce one of these energy eigenvalues. Which eigenvalue observed is not deterministic, but each can be assigned probabilities, so often the more useful quantity is the expectation value, which gives the average energy of the system.

$$\langle E\rangle = \frac{\int \psi^*\,\mathcal{H}\,\psi\,d\tau}{\int \psi^*\psi\,d\tau} \tag{3.5}$$

From this energy one can calculate atomic forces and refine equilibrium structures.

3.2.1 One Electron Atoms and the Born-Oppenheimer Approximation

The Hydrogen atom is a prototypical system containing one electron and a single proton nucleus. The Hamiltonian for this system is given below.

$$\mathcal{H} = -\frac{\hbar^2}{2m_p}\nabla_p{}^2 - \frac{\hbar^2}{2m_e}\nabla_e{}^2 - \frac{1}{4\pi\varepsilon_o}\frac{e^2}{r} \tag{3.6}$$

This consists of the respective kinetic terms for both proton and electron and the Coulombic attraction term describing the potential. Here, $\hbar$ is the reduced Planck's constant, m_p and m_e are the masses of proton and electron, e is charge on the electron, ε_0 is the vacuum permittivity, and ∇^2 represents the Laplacian or second derivative operator in spherical coordinates (r, θ, φ).

One important approximation in chemical simulations is separating the motion of nuclei from the electron distributions around them. This is not unreasonable since the nucleus is far more massive than electrons, so their characteristic velocities are widely different, allowing us to treat nuclear motion parametrically, where the nuclear coordinates are treated as fixed, and the electronic wavefunction is solved on this fixed nuclear framework. This also allows separating each contribution, giving a set of nuclear coordinates R and the electronic coordinates r, and forms the basis of the Born Oppenheimer approximation.

$$\Psi_{total}(r,R) = \phi_{nuclear}(R)\,\psi_{electronic}(r)$$

$$\mathcal{H}_{total} = \mathcal{H}_{nuclear} + \mathcal{H}_{electronic} \tag{3.7}$$

$$E_{total} = E_{nuclear} + E_{electronic}$$

For hydrogenic atoms, the general form of the electronic wavefunction is known in spherical coordinates.

$$\psi_{n,m,l}(r,\theta,\phi) = R_l^n\,(r)\;\;\Upsilon_l^m\,(\theta,\phi) \tag{3.8}$$

Here, $R_l^n\,(r)$ is the radial distribution function describing how the probability amplitude falls off away from the nucleus, and $\Upsilon_l^m\,(\theta,\phi)$ is the spherical harmonic, describing the angular dependence of a wavefunction under a spherical constraint.

3.2.2 Multi Electron Atoms and Molecules

Helium contains two electrons and so requires introducing a new term in the Hamiltonian, namely the electron-electron repulsion term.

$$\mathcal{H}_{elec} = -\frac{\hbar^2}{2m_e}\left[{\nabla_1}^2 + {\nabla_2}^2\right] - \frac{Z}{4\pi\varepsilon_o}\left[\frac{e^2}{r_1} + \frac{e^2}{r_2}\right] + \frac{1}{4\pi\varepsilon_o}\frac{e^2}{r_{12}} \tag{3.9}$$

The solution to this three body problem can no longer be written in closed analytical form and so a number of further approximations become needed. The first of which is the Hartree approximation, which assumes that the total wavefunction can be written as a product of the individual electron wavefunctions.

$$\psi(r_1, r_2, \ldots, r_N) = \varphi_1(1)\varphi_2(2)\ldots\varphi_N(N) \tag{3.10}$$

The Hartree approximation must be modified to include electron spin states α and β, and linear combinations of these Hartree products are constructed to meet the antisymmetry constraint of all fermions (also known as Pauli's exclusion principle), that their wavefunction reverses sign under exchange.

$$\psi(1,2) = -\psi(2,1) \tag{3.11}$$

The Slater determinant is a compact way of generating antisymmetric wavefunctions for multielectron systems. Here each function χ_i is a spin orbital consisting of the product of spatial and spin components.

$$\psi(1,2,\ldots,N) = \frac{1}{\sqrt{N}}\begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} \tag{3.12}$$

Electrons ↓

Spin Orbitals →

The progression of spin orbitals follows the usual Aufbau ordering for multielectron atoms.

$$\begin{aligned} \chi_1(1) &= 1s(1)\alpha(1) \\ \chi_2(1) &= 1s(1)\beta(1) \\ \chi_3(1) &= 2s(1)\alpha(1) \\ &etc. \end{aligned} \tag{3.13}$$

One functional form of χ_i is known as Slater Type Orbitals or STOs. STOs are used to represent the overall wavefunction of an atom and finding a solution to the Schrödinger equation generally consists of optimizing the exponential coefficients.

As atoms come together to form molecules, the atomic orbitals superimpose to give molecular orbitals, or patterns of constructive (bonding) and destructive (anti-bonding) interference. These molecular orbitals are generally approximated by a linear combination of atomic orbitals.

$$\phi_i = \sum_{a=AOs} C_{ai}\, \chi_a \tag{3.14}$$

The coefficients C_{ai} are guessed initially and then optimized until the energy is minimized to some set tolerance.

3.2.3 The Self Consistent Field (SCF) method of Hartree and Fock

The Hartree-Fock method recasts the many electron wavefunction as a set of 1-electron equations.

$$\mathcal{F}_i\, \phi_i = \varepsilon_i\, \phi_i \tag{3.15}$$

Where the Fock operator ($\mathcal{F}$) for an individual electron is as follows:

$$\mathcal{F}_i = \mathcal{H}^{core}(i) + \sum_{j=1}^{N} \big(J_j(i) - K_j(i)\big) \tag{3.16}$$

Where $\mathcal{H}^{core}(i)$ describes the electron kinetic energy and Coulombic interaction with the nucleus, $J_j(i)$ is the Coulombic term describing electron-electron repulsion, and $K_j(i)$ is the exchange operator describing a purely quantum mechanical effect involving the nature of fermions (indistinguishable and antisymmetric). One of the key assumptions in the HF approach is that the remaining electrons are treated as an average effect or mean field.

The Hartree-Fock approach completely neglects electron correlation, which can be significant, so the lowest possible energy obtainable with the HF method is always higher than the true ground state energy. The HF framework has served as a jumping off point for a large variety of so called post Hartree Fock methods that seek to include electron correlation, including perturbative approaches such as Møller–Plesset (MP), and Configuration Interaction (CI) approaches which are in principle exact.

3.2.4 Roothaan-Hall Formulation

Roothan and Hall developed a matrix formulation of the HF equations for closed shell molecules that is particularly suited for modern computations. They expressed the HF equations as a generalized matrix formula.

$$\boldsymbol{FC} = \boldsymbol{SCE} \tag{3.17}$$

Where $\boldsymbol{F}$ represents the Fock matrix, $\boldsymbol{C}$ the coefficients of the 1 electron wavefunctions, $\boldsymbol{S}$ the overlap matrix, and $\boldsymbol{E}$ the 1 electron energies. The Roothaan-Hall equations can be solved by matrix diagonalization using an iterative procedure and as a benefit, the basis functions can be either Slater type orbitals (STOs) or the less computationally expensive Gaussian type orbitals (GTOs).

3.2.5 Density Functional Theory (DFT)

The origins of Density Functional Theory trace back to Hohenberg and Kohn's seminal paper in 1964, which reformulated the ground state electronic structure problem in terms of the electron density, a scalar field.[1] This was quickly extended by Kohn and Sham in 1965, who reframed the N-electron interacting system into a series of N

coupled single particle equations for the non-interacting system.[2] This can be written in the time-independent formulation,

$$\left[-\frac{\nabla^2}{2}+\ v_{eff}(\boldsymbol{r})\right]\psi_i(\boldsymbol{r})=\ \varepsilon_i\psi_i(\boldsymbol{r}) \tag{3.18}$$

where

$$v_{eff}(\boldsymbol{r})=v_{ext}(\boldsymbol{r})+\ v_{hartree}(\boldsymbol{r})+\ v_{xc}[\rho;\boldsymbol{r}] \tag{3.19}$$

The first two terms are known exactly, where v_{ext} describes electron-nucleus interactions, and $v_{hartree}$ is the classical electrostatic potential describing coulombic electron-electron interaction. The last term, v_{xc}, is the exchange correlation potential, and describing this interaction is the core problem in density functional theory. This term is often recast in terms of the energy.

$$v_{xc}[\rho;\boldsymbol{r}]=\frac{\delta E_{xc}}{\delta\rho} \tag{3.20}$$

If the universal functional E_{xc} was known exactly, density functional theory would be essentially a solved problem. However, though much is known about the behavior of this functional in various limiting cases, no exact form has yet been discovered, and instead it is modeled by many varying approaches towards approximation.

DFT would remain largely an academic curiosity till the 1990s when improved functionals and computational methods led to an explosion of interest in DFT for modeling chemical problems.

3.2.6 Functionals

The first choice in using DFT is choosing a functional. One of the most used for modeling organic molecules is the B3LYP functional, developed by Becke, Lee, Parr and Yang. [3,4] B3LYP is what's known as a hybrid functional, incorporating a percentage of Hartree-Fock style exact exchange, and a mix of Local Density Approximation (LDA) and Generalized Gradient Approximations (GGA) approaches for the remainder of the exchange behavior and a degree of correlation effects. Most importantly, the behavior of B3LYP is well known for a variety of molecule types, including organic molecules and transition metals. Many other functionals exist with improved accuracy and behavior in varying situations. When quantitative accuracy is of utmost importance, the functional should be chosen carefully and several comparisons of functional performance can be found in these reviews. [5,6]

3.2.7 Basis Sets

Since the exact wavefunction for complicated chemical systems are not known a priori, a basis set consisting of simple functions (generally based on atomic orbitals) is used to represent the wavefunction. Though Slater type orbitals[7] are the most faithful representation of the atomic orbitals, their computational cost is high due to the cusp near the origin, making accurate integration prohibitively expensive. A solution is to use a linear combination of Gaussian type orbitals (GTOs) [8] to represent each STO. Though this increases the number of basis functions, each GTO is finite valued at the origin, making integration far simpler and as such the overall computational cost is greatly reduced.

The split valence double zeta basis of Pople and colleagues is commonly used in this work, starting at a minimum level of 6-31G* for most calculations. This implies 6 GTOs to represent each core STO orbital, and a linear combination of 3 contracted GTOs and 1 diffuse GTO per valence STO orbital. The star indicates additional d-like polarization functions added for each heavy atom to improve anisotropy. Higher level basis sets are often used to show convergence, particularly in cases where anisotropy is important such as hydrogen bonding, or polarization is significant, such as with soft atoms like sulfur.

3.2.8 Implicit Solvent Models

Condensed phase systems incur the additional complication of the solvent environment. For small model QM calculations, including explicit solvent (several layers of water molecules) would add significantly to the computational cost. Instead, the most common approach for small model QM calculations is to use a continuum solvent model, where different environments are modeled based on their average dielectric behavior. This is admittedly an oversimplification, but by contrasting calculated results in several environments (gas phase, water, and protein); one can get an estimate of the average error.

The most common model used with QM and DFT calculations is the polarized continuum model (PCM)[9] where the dielectric constant is chosen for each calculation and the influence of electrostatic, dispersion, and cavitation effects are included in the overall calculation.

3.3 Small Model Systems

In constructing a small model to represent a complex system like an enzyme active site, the first approximation starts with all the residues and molecules directly involved in any chemical transformations. One would then expand to include as much of the first shell of residues as possible, giving priority to residues that interact strongly at the point of reaction. Except in rare cases where the protein backbone is taking part in the reaction chemistry, individual residues have their backbone linkages severed and capped with terminal hydrogen. An example is shown below from the bacterial enzyme Spvc, which cleaves a phosphate group from a peptide strand. Each of the arginine residues has been modeled as a guandinium ion, the histidine as an imidazole, and the lysine residue as a methyl amine. Enough of the ligand protein backbone has been included to account for the major interactions, and capped with methyl groups on each side.

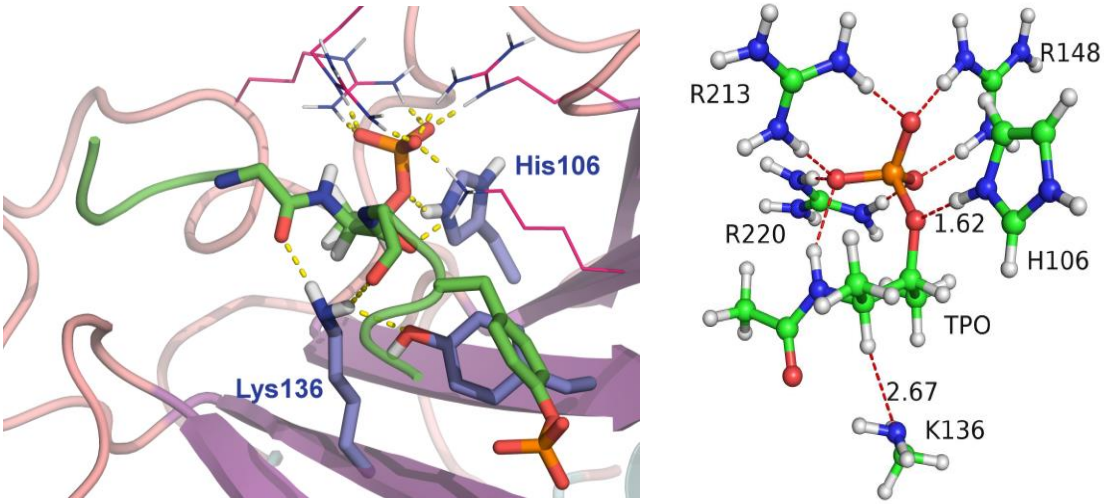


Figure 3.1 – The active site of bacterial enzyme Spvc (left panel) is studied by a truncated model (right panel) system including a portion of the substrate and nearby residues. The system is now small enough to model using quantum methods.

The appropriate model is a tradeoff between capturing all the major interactions while keeping the total electron count at a reasonable level for computation. In practice, you start with the simplest model possible and gradually add residues till a stable complex is found. Convergence of behavior is a reasonable stopping point if atom count does not force your hand earlier. That said, this is admittedly a crude model, since all long range effects are jettisoned, which may be significant. As such, the small model behavior should be considered as exploratory only and any mechanistic conclusions should be carefully weighed. This can still be quite useful for exploring possible reaction mechanisms however, as enzymes typically cannot overcome a solution barrier height greater than about 30 kcal/mol, so truncated models can often rule out unrealistic mechanistic pathways.

3.3.1 Transition State Search

Once stable reactant and product states are found, the next step is finding the transition state linking the two. Many small model QM programs such as Guassian 03 [10] can search for transition states between two minima. Transition states are confirmed in two ways. First, a frequency calculation is performed on the transition state, which should contain a single imaginary frequency, characteristic of a first-order saddle point. Second, the transition state is linked with product and reactant states through the intrinsic reaction coordinate (IRC) method [11] which follows the characteristic vibrational mode to both the reactant and product complexes. If the IRC does not connect smoothly to reactant and product states, the endpoints can be minimized to produce

new reactant and product minima and the search can be repeated till a true path is found.

After identifying reactant, transition state, and product stationary points, the barrier height and thermochemistry can be mapped out. Barrier heights in gas phase and with water and protein environments are calculated with the PCM model to estimate the effect of the solvent. Zero point energies are also reported (no additional calculation needed since vibrational frequencies of the stationary points are used to verify transition states and minima).

3.3.2 Reaction Barriers and Kinetics

Rate constants can be calculated from the reaction barrier information using transition-state theory.[12] Likewise, experimental results can be converted to free energy barriers.

$$k = \frac{k_b T}{hc}\, e^{-\Delta G^{\ddagger}/RT} \tag{3.21}$$

Here, k is the rate, k_b is the Boltzmann constant, T the temperature, h Planck's constant, c the standard state concentration, R the gas constant, and $\Delta G^{\ddagger}$ the free energy reaction barrier. Comparison of the gas phase calculation with the different solvent environments and the experiment values (both in solution for chemical analogues and enzyme results) can shed light on enzyme mechanisms. Large disparities between any of these numbers can indicate the main function of the enzyme or even suggest that the enzyme must progress through a different route than the solution analogue. Given the lack of protein environment, exact quantitative agreement should not be expected through anything but fortuitous cancelation of errors.

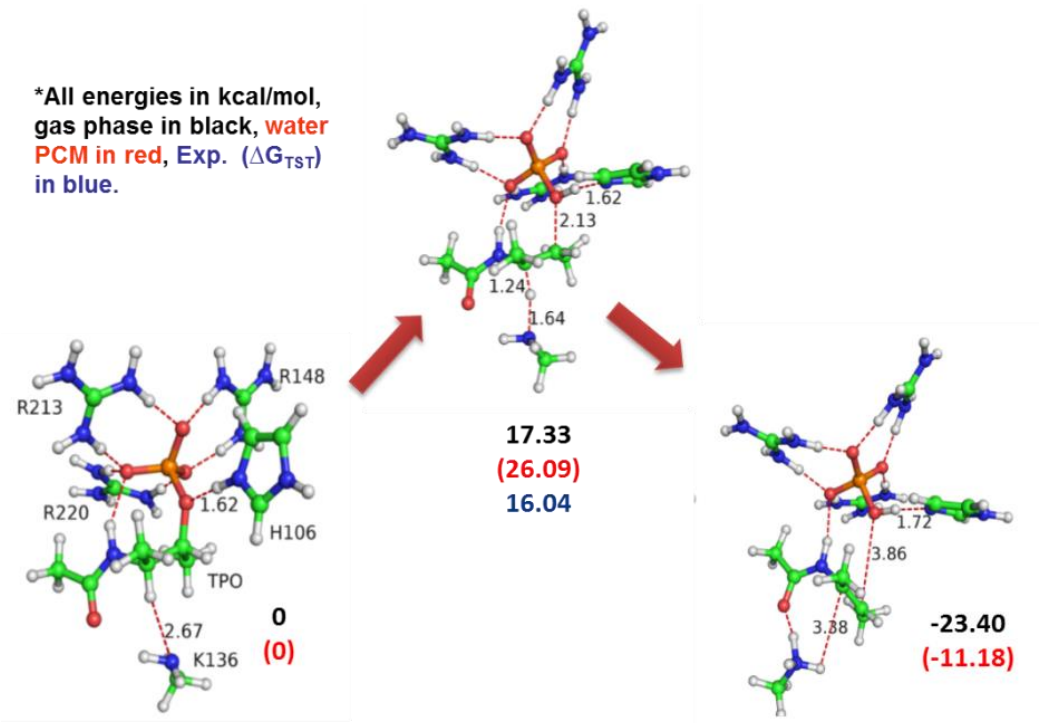


Figure 3.2 – Stationary states for C-O cleavage of a phosphate group on the substrate in the truncated model.

3.3.3 Kinetic Isotope Effects

Kinetic isotope effects (KIEs) are useful experimental tools in figuring out reaction mechanisms, so calculating the expected KIEs for our stationary points can be another way of validating our reaction mechanism. Here we use the Bigeleisen-Meyer method [13] as implemented in the ISOEFF[14] computer program. This method calculates the harmonic frequencies of the reactant and transition state to predict the effect of heavy isotopes on the reaction rate.

3.4 Quantum Mechanics / Molecular Mechanics

The major drawbacks of the truncated model approach are the lack of coupling with the extended environment and the lack of a bath to dissipate energy. This can be largely

remedied by using a hybrid approach, where the catalytic site is modeled quantum mechanically while the remainder of the protein and a large portion of the solvent is modeled using molecular mechanics. This approach is known as QM/MM[15] and has been utilized to study large systems that would be unfeasible to study solely with quantum mechanics. [16-23]

3.4.1 System partitioning

The first step is partitioning your system into QM and MM portions. The approach here is similar to the truncated model considerations. At minimum you include the residues or ligands directly involved in chemical transformations, and then add in surrounding residues that interact strongly with these residues. The QM region is generally expanded until the system stops changing or shows convergence. An example is shown below of a partitioning scheme for the enzyme active site in the human enzyme CDK2. The portion in red indicates the QM system, the portion in blue the MM system, and the atoms in green represent the interface between the two. The remainder of the protein and solvent (not shown) is also modeled using MM.

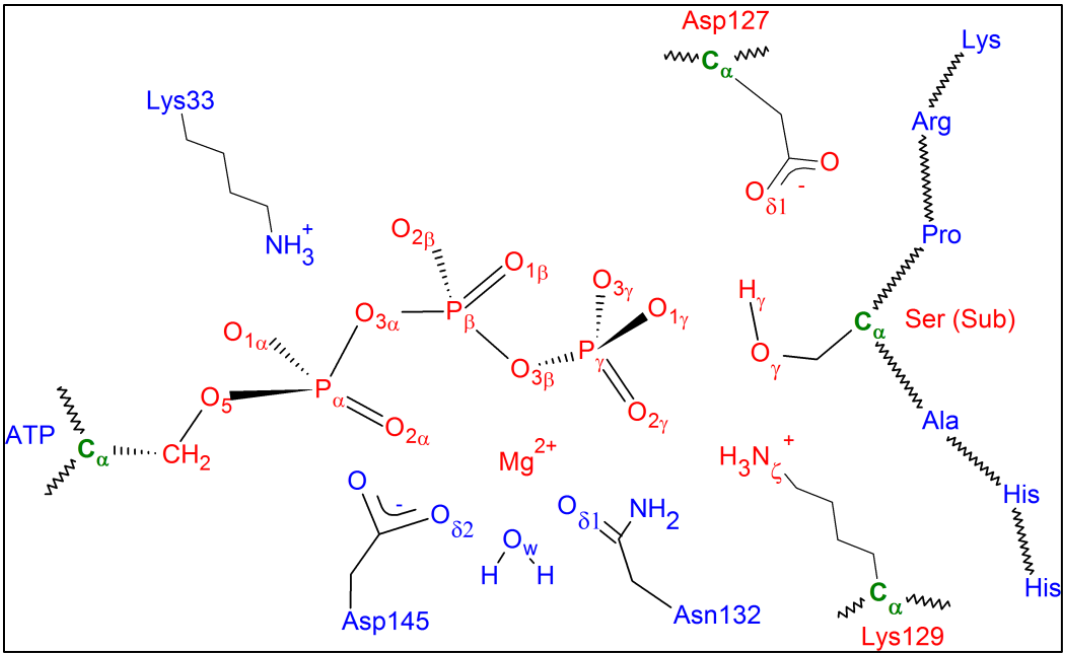


Figure 3.3 – Partitioning an enzyme active site into quantum (in red) and classical (in blue) regions.

With these three regions, you can write the Hamiltonian of the system generally as follows:

$$H = H_{MM} + H_{QM} + H_{\mathrm{QM/MM}}^{vdW} + H_{\mathrm{QM/MM}}^{ES} + H_{\mathrm{QM/MM}}^{bonded} \tag{3.22}$$

The MM subsystem is calculated with a choice of force field as detailed in Chapter 2. Force fields used in this work include the CHARMM[24] and AMBER[25] force fields. The QM subsystem can be calculated using a quantum method of choice, which can include HF, post HF, DFT, or semi-empirical methods depending on the size of the QM system and available computational resources.

How to handle the interaction between the subsystems is the difficult aspect of QM/MM calculations and many differing schemes are employed. The least

controversial is van der Waals interactions between MM and QM atoms, which are typically modeled by the Lennard-Jones potential.

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.23}$$

The parameters σ and ε are taken from the MM forcefield and calculated over a finite cutoff distance, commonly on the order of 12-14 Å. The latter two terms, $H^{ES}_{\mathrm{QM/MM}}$ and $H^{bonded}_{\mathrm{QM/MM}}$ are handled in a variety of ways and require more explanation.

3.4.2 $H^{ES}_{\mathrm{QM/MM}}$ - Mechanical versus Electrostatic Embedding

Electrostatic interactions are commonly handled in two ways. The first is known as mechanical embedding and is a point charge scheme, where MM atom charges comes from the force field employed and the QM atom charges come from a charge partitioning scheme such as Mulliken or Bader analysis. This additional electrostatic term is then added to the total Hamiltonian.

$$V_i(r) = -\frac{q_{MM}\, q_{QM}}{4\pi\epsilon_0\, r} \tag{3.24}$$

This is the basis of the ONION[26] scheme found in the Gaussian 03 program. Here the QM/MM through space interaction is treated essentially as a perturbation, without any strong coupling.

Another approach is known as electrostatic embedding, which differs in that the MM atoms are included in the QM Hamiltonian as point charges to allow polarization of the QM subsystem. This polarization is solely one way however, as the MM subsystem still

sees the QM atoms as partitioned point charges. Mutual polarization would require a polarizable force field, which are areas of active research in this field.

3.4.3 $H_{\mathrm{QM/MM}}^{bonded}$ - Boundary Atoms

For a small molecule, one can choose to include it in the MM or QM system and no boundary atoms need to be addressed. However, for long polymers like proteins, including the whole protein in the QM system is generally unfeasible computationally, and so a cut must be made along the chain separating MM from QM subsystems. No perfect solution exists for handling bonded atoms at the interface of the QM and MM subsystems, but there are varying approaches towards approximation.

The first issue is choice of atom and bonding pattern for the interface, and most approaches are parameterized for use with sp^3 carbon-carbon bonds. For a protein residue, this is generally the alpha and beta carbons. The simplest approach is to include a phantom hydrogen atom that exists solely in the QM subsystem to saturate the beta carbon. The connection between MM and QM system is maintained by the MM force field, keeping the alpha carbon tethered to the QM region. This is referred to as the "link-atom" approach [27,28] and is generally reasonable as long the link atom is not very close to the reaction center.

More sophisticated approaches exist as well including the generalized hybrid orbital (GHO)[29] and the pseudo-bond (PB)[30] approaches. The pseudo-bond approach is similar in philosophy to the link atom approach in that the QM system is capped by a phantom atom that only exists in the QM system, however the pseudo bond atom is carefully

chosen to mimic the behavior of a bonded sp^3 carbon atom, not a hydrogen as in the link atom approach. As a result, there tends to be more coherence between the MM C alpha position and the pseudo atom in a simulation. An additional benefit of the pseudo bond approach is that parameterizations exist for breaks along the protein backbone at the carbonyl carbon or the amide nitrogen as well.[31]

3.4.4 Simulation Boundaries – Periodic Cells and Stochastic Boundary Conditions

As detailed in the previous chapter, to include the average effect of extended solvent regions surrounding our protein, an additional region must be included in the simulation protocol.

The first method is to use periodic boundary conditions[32], constructing a box around our protein system and filling the box with explicit water molecules. While coordinates are only calculated for one of these boxes, when calculating electrostatic interactions the neighboring images are included. A rigorous way to sum up these long range electrostatic interactions is known as the particle mesh Ewald (PME) summation.[33,34]

Another approach is to include a spherical boundary, where the effects of solvent are modeled as an average effect, or buffer region. Stochastic boundary conditions [35,36] is one such spherical method, where the effect of extended solvent is modeled by Langevin dynamics, including a random impulse term (hence the name stochastic) and frictional forces.

3.4.5 Reaction Coordinate Driving

Given the complexity of the potential energy surface in a given protein active site, finding the minimum energy path connecting reactant and product states can pose a serious challenge, especially if the exact mechanism is not known. One way of addressing this is the reaction coordinate driving method. [37] In this approach, a reaction coordinate is chosen, generally starting with the inclusion of bond breaking and forming events. This reaction coordinate is then iterated and a minimized set of structures is found along the coordinate. In essence, bond breaking and bond forming events can be forced by using a strong harmonic potential centered on the desired reaction coordinate. By examining the energy profile along this path, the feasibility of a certain reaction route can be evaluated. Figure 3.4 shows an example potential mapped out iteratively by evaluating the energy at each point.

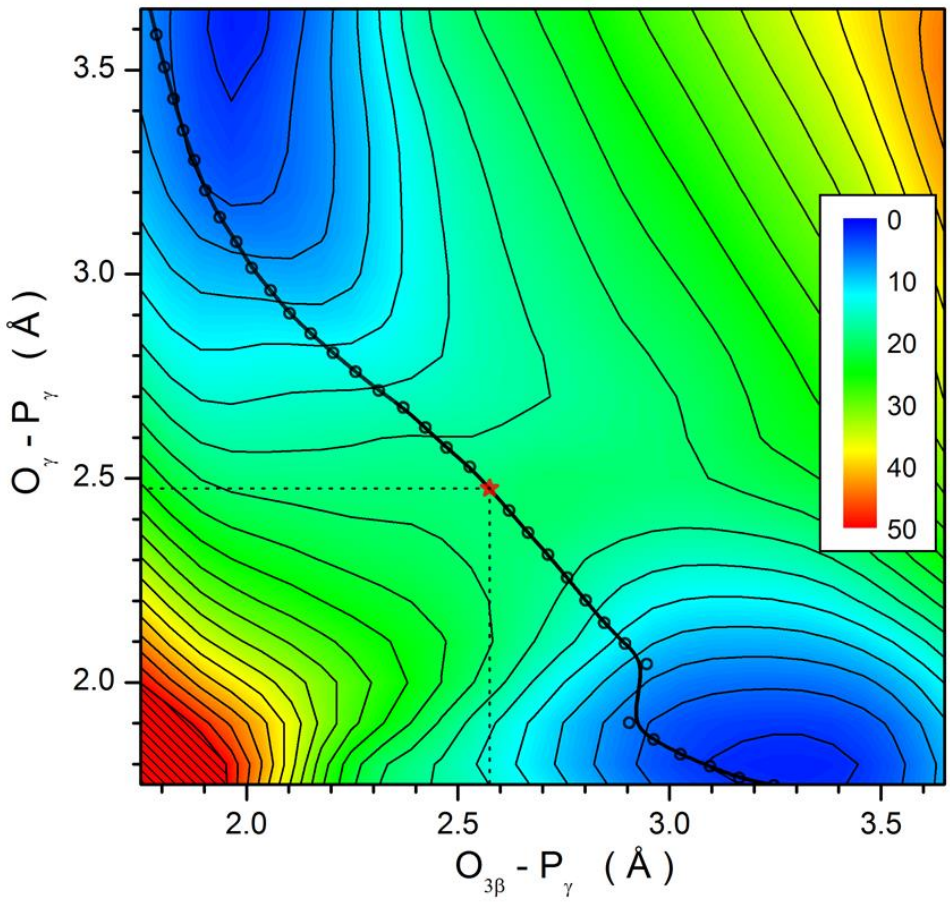


Figure 3.4 – A two dimensional potential energy mapping along bond breaking and bond forming coordinates for a phosphate transfer reaction. The minimum energy path and transition state is shown.

3.4.6 QM/MM MD – unconstrained dynamics and sampling the potential of mean force.

While reaction paths explore the local potential energy surface along a coordinate of interest, it is the free energy that determines reaction rates. Adding in the entropic contributions for a large system is not straight forward. One way of including this is to

allow the system to fluctuate using molecular dynamics. Thermal fluctuations allow the system to surmount local minima and sample the local potential energy surface. In this fashion one can sample the reactant and product states and get an estimate of free energy at these stationary states.

Reaction paths with a significant barrier run up against a sampling problem. Barrier crossing is a very rare event, and if one runs an unconstrained dynamics trajectory at 300 K for a few nanoseconds, the reactant or product states will be sampled heavily with little or no sampling near the transition states.

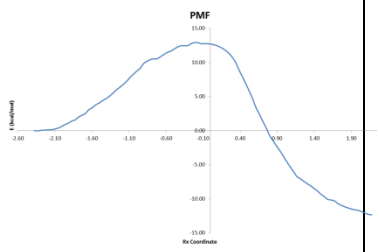
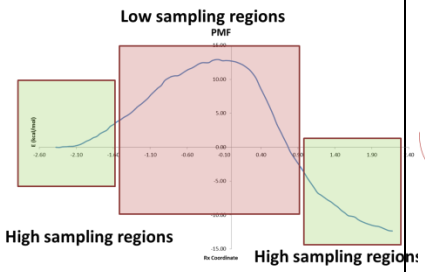
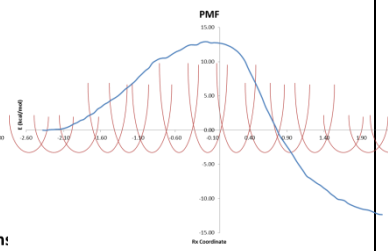
Unbiased Free Energy	Unbiased Sampling	Umbrella Sampling
		

Figure 3.5 – The sampling problem for unbiased molecular dynamics simulations and one solution using bias potentials.

To get around the sampling problem for high free energy points along the reaction coordinate, a bias potential is added to "flatten" out the effective potential. Figure 3.5 indicates this approach. The left panel shows an unbiased free energy. The middle panel shows the resulting sampling along our reaction coordinate, with high sampling in low energy regions and very little sampling in high energy regions. The right panel

shows the process of umbrella sampling [38], where a small bias potential is added at each point along the reaction coordinate, with weak bias potentials in low energy regions, and strong bias potentials in high energy regions.

The Weighted Histogram Analysis Method (WHAM)[39] addresses how each simulation must be weighted by its bias potential to arrive at a reasonable estimate of the unbiased free curve.

WHAM was developed using Statistical Mechanical arguments relating free energy to probability, and its fundamental equations are given below with a brief explanation.

$$P(x) = \frac{\sum_{i=1}^{N} n_i(x)}{\sum_{i=1}^{N} n_i \mathrm{e}^{\frac{F_i - U_{b,i}(x)}{kT}}} \tag{3.25}$$

$$F_i = -kT \ln\left[\sum_{bins} P(x) \mathrm{e}^{\frac{-U_{b,i}(x)}{kT}}\right] \tag{3.26}$$

In the above equations, the unknowns include $P(x)$, an estimate of unbiased probability, and F_i, an estimate of the unbiased free energy. The known quantities include x as the reaction coordinate, U_b as the bias potential at x, N is the number of simulations (also equal to the number of bins containing a unique bias potential), and n_i are the number of counts in any defined histogram bin.

The conceptual link between these two equations is worth noting. For example, in equation 3.25, if the bias potential is perfectly chosen to cancel the free energy at a particular point ($F_i - U_{b,i} = 0$), then the equation is reduced to the probability of a flat surface (equal sampling everywhere within the bin). Also notice that equation 3.25 and

3.26 are coupled, as P and F_i appear in both. This allows solving these two equations iteratively by making an initial guess and refining till self-consistency is reached. In practice the bias potential within each bin is adjusted iteratively until every bin meets a minimum threshold for sampling. In the limit of an infinite number of bins and perfectly adjusted bias potential, one recovers the exact free energy. In practice one merely must reach convergence with respect to the number of bins and bias potential adjustments.

The result is what is known as the potential of mean force or PMF[40] where the free energy is evaluated along the reaction coordinate, with all other degrees of freedom included as an average effect. The potential which matches the average force along the reaction path is the PMF.

3.4.7 Semi empirical QM/MM and SCC-DFTB

Since molecular dynamics on a QM/MM potential can be quite computationally expensive, it is common to employ semi-empirical methods to calculate the quantum subsystem. These are generally approximations to the HF approach, where a portion of the matrix integrals are either neglected or approximated with the use of empirical parameters. There are many semi-empirical methods, which differ widely in accuracy, speed, and how general their area of application. Two general purpose methods that have seen fairly wide application include AM1[41] and the PM3[42] methods.

While most semi-empirical methods derive from the HF approach, one semi-empirical method used in this work is based on DFT. Called self-consistent charge density

functional tight binding (SCC-DFTB)[43], this method is structured as a tabulated reference density modified by a density fluctuation term.

$$E_{SCC-DFTB} = \sum_{i,uv} c_u^i \, c_v^j \, \mathcal{H}_{uv}^0 + \frac{1}{2}\sum_{\alpha\beta} \Delta q_\alpha \, \Delta q_\beta \, \gamma_{\alpha\beta} + E_{core-repulsion} \quad (3.27)$$

The first term represents the reference density and is tabulated based on DFT calculations for the non-interacting system over a range of distances. This consists of the Hamiltonian matrix element $\mathcal{H}_{\alpha\beta}^0$ and the coefficients of the two body interactions c_u^i and c_v^j. The second term represents the density fluctuation, where Δq_α and Δq_β are calculated by Mulliken charge analysis and represent atomic charge fluctuations from a reference charge, while $\gamma_{\alpha\beta}$ is related to chemical hardness or softness of an atom. The third term represents pairwise repulsions at the reference density and is tabulated over a range of distances and connected using spline functions.

A key assumption of the method is tight binding, so the results tend be less reliable near the dissociation limit, leading to large errors in reaction barriers. Equilibrium geometries and energies compare well with results from ab-initio DFT but with computational cost cut by 2-3 orders of magnitude. Of particular interest is that SCC-DFTB has been applied successfully to zinc containing enzymes.[44,45]

3.4.8 *Ab initio* QM/MM MD

Recently, the convergence of increasing computer power and more efficient methods has made using *ab initio* QM in the quantum subsystem feasible both for evaluating reaction paths and more recently for running dynamics and calculating the free energy

of reaction. One recent approach utilizes the pseudobond method[46] of linking QM and MM subsystems and calculates QM/MM potential using an efficient ab-initio DFT method.[47]

Many QM/MM approaches fully optimize the QM subsystem and then fully optimize the MM subsystem to calculate each step of dynamics. This approach can sometimes lead to oscillations and difficulty in reaching convergence. Zhang and coworkers have developed a method leveraging the comparative efficiency of the MM calculation by fully minimizing the MM subsystem for each electronic step. The QM subsystem is then iterated with the MM environment fixed. The result is the QM subsystem generally converges faster since the surrounding environment is fully adjusted.

3.5 QM and QM/MM methods used in this work

Chapter 5 details the study of a bacterial enzyme called Spvc, in which two separate simulations were performed; A classical simulation using force fields and a truncated model of the active site complex using DFT and implicit solvent.

Chapter 6 details the study of a human enzyme called CDK2, where the AMBER force field was used for initial system minimization and dynamics, before switching to an ab-initio QM/MM method to model the active site chemistry. This allowed the study of reaction paths and free energy profiles of reactions involved in this enzyme.

Chapter 7 details a study involving anthrax Lethal Factor (LF), which is an enzyme responsible for much of the damage done when anthrax bacteria infects the lymphatic

system. This project used a hybrid QM/MM technique, using the CHARMM force field to describe the classical region, and the semi-empirical quantum method called SCC-DFTB to describe the reaction zone.

3.6 References

1 Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical Review* **136**, B864 (1964).

2 Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **140**, A1133 (1965).

3 Becke, A. D. Density-functional thermochemistry.3. The role of exact exchange. *Journal of Chemical Physics* **98**, 5648-5652 (1993).

4 Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **37**, 785-789 (1988).

5 Perdew, J. P. & Burke, K. Comparison Shopping for a Gradient-Corrected Density Functional. *International Journal of Quantum Chemistry* **57**, 309-319 (1994).

6 Sousa, S. F., Fernandes, P. A. & Ramos, M. J. General performance of density functionals. *Journal of Physical Chemistry A* **111**, 10439-10452 (2007).

7 Slater, J. C. *Physical Review* **36** (1930).

8 W. J. Hehre, R. F. Stewart & Pople, J. A. Self Consistent Molecular Orbital Methods. I. Use of Gaussian Expansions of Slater Type Atomic Orbitals *Journal of Chemical Physics* **51**, 8 (1969).

9 Tomasi, J. & Persico, M. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chemical Reviews* **94**, 2027-2094 (1994).

10 Gaussian 03 (Gaussian, Inc., Wallingford, CT, 2004).

11 Gonzalez, C. & Schlegel, H. B. An improved algorithm for reaction path following. *Journal of Chemical Physics* **90**, 2154 (1989).

12 McQuarrie, D. A. & Simon, J. D. *Physical Chemistry - A Molecular Approach*. (University Science Books, 1997).

13 Bigeleisen, J. & Mayer, M. G. Calculation of equilibrium constants for isototpic exchange reactions. *Journal of Chemical Physics* **15**, 261 (1947).

14 Anisimov, V. & Paneth, P. ISOEFF98. A program for studies of isotope effects using Hessian modifications. *Journal of Mathematical Chemistry* **26**, 75 (1999).

15 Warshel, A. & Levitt, M. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103**, 227-249 (1976).

16 Aqvist, J. & Warshel, A. Simulation of enzyme reactions using valence-bond force-fields and other hybrid quantum-classical approaches. *Chemical Reviews* **93**, 2523 (1993).

17 Gao, J. in *Reviews in Computational Chemistry* Vol. 7 (eds K. B. Lipkowitz & D. B. Boyd) 119-185 (VCH, 1996).

18 Warshel, A. Computer simulations of enzyme catalysis: methods, progress, and insights. *Annual Review of Biophysics and Biomolecular Structure* **32**, 425-443 (2003).

19 Zhang, Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts* **116**, 43-50 (2006).

20 Riccardi, D. *et al.* Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *Journal of Physical Chemistry B* **110**, 6458-6469 (2006).

21 Lin, H. & Truhlar, D. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **117**, 185-199 (2007).

22 Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie (International ed. in English)* **48**, 1198-1229 (2009).

23 Ranaghan, K. E. & Mulholland, A. J. Investigation of enzyme-catalyzed reactions iwth combined quantum mechanical/molecular mechanical (QM/MM) methods. *International Reviews in Physical Chemistry* **29**, 65-133 (2010).

24 Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187-217 (1983).

25 Case, D. A. *et al.* Amber 10. *University of California, San Francisco* (2008).

26 Svensson, M. *et al.* ONION, a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactionps and Pt(P(t-Bu)3)2 + H2 oxidative addition. *Journal of Physical Chemistry* **100**, 19357-19363 (1996).

27 Singh, U. C. & Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3Cl + Cl^-$ exchange reaction and gas-phase protonation of polyethers. *Journal of Computational Chemistry* **7**, 718-730 (1986).

28 Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **11**, 700-733 (1990).

29 Gao, J., Amara, P., Alhambra, C. & Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *Journal of Physical Chemistry* **A102**, 4714-4721 (1998).

30 Zhang, Y., Lee, T. & Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *Journal of Chemical Physics* **110**, 46-54 (1999).

31 Zhang, Y. Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods. *Journal of Chemical Physics* **122**, 24114 (2005).

32 Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids*. (Oxford University, 1986).

33 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an NlogN method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089-10092 (1993).

34 Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577 (1995).

35 Brooks III, C. L. & Karplus, M. Deformabal stochastic boundaries in molecular dynamics. *Journal of Chemical Physics* **79**, 6312-6325 (1983).

36 Brooks III, C. L. & Karplus, M. Solvent effects on protein motion and protein effects on solvent motion. *Journal of Molecular Biology* **208**, 159-181 (1989).

37 Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry* **53**, 291-318 (2002).

38 Torrie, G. M. & Valleau, J. P. Non-physical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).

39 Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free energy calculations on biomolecules. 1. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).

40 Chandler, D. *Introduction to Modern Statistical Mechanics*. (Oxford University Press, 1987).

41 Dewar, M. J. S., Zoebisch, E. G., Healy, E. & Stewart, J. J. P. The development and use of quantum-mechanical molecular models. 76. AM1 - A new general-purpose quantum mechanical molecular model. *Journal of the American Chemical Society* **107**, 3902-3909 (1985).

42 Stewart, J. J. P. Optimization of parameters for semi-empirical methods. 1. method. *Journal of Computational Chemistry* **10**, 209-220 (1989).

43 Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review* **B58**, 7260-7268 (1998).

44 Elstner, M. *et al.* Modeling zinc in biomolecules with the self consistent charge density functional tight binding (SCC-DFTB) method: Applications to structure and energetic analysis. *Journal of Computational Chemistry* **24**, 565-581 (2003).

45 Amin, E. A. & Truhlar, D. G. Zn coordination chemistry: Development of benchmark suites for geometries, dipole moments, and bond dissociation energies and their use to test and validate density functionals and molecular orbital theory. *Journal of Chemical Theory and Computation* **4**, 75-85 (2008).

46 Zhang, Y., Lee, T.-S. & Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *The Journal of Chemical Physics* **110**, 46, doi:10.1063/1.478083 (1999).

47 Zhang, Y., Liu, H. & Yang, W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *The Journal of Chemical Physics* **112**, 3483 (2000).

Chapter 4

Surface Methods

Planewave DFT

"The best that most of us can hope to achieve in physics is simply to misunderstand at a deeper level."

-Wolfgang Pauli

4.1 Introduction[1]

Crystalline solids can be described as a repeating unit cell that extends in all directions. This periodic arrangement of atoms gives rise to a periodic potential energy surface and so the resulting electronic wavefunction in a crystalline solid is periodic as well. The wavefunction is best described as a three dimensional standing wave across the entire lattice.

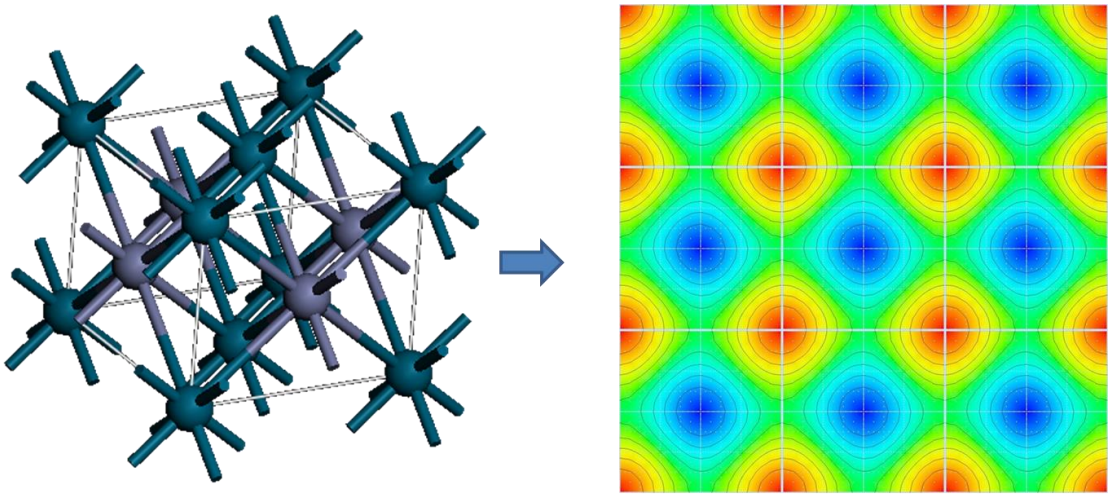


Figure 4.1 – Periodic unit cell for a crystalline solid (left) and a representative slice of the periodic potential setup over several unit cells.

Using a DFT framework, this periodicity can be exploited to calculate geometries and energies for extended lattices and surfaces by only explicitly calculating atomic positions in a single unit cell, and then solving for the energy self-consistently by summing up the contributions of the identical neighboring images.

For discrete systems, the limiting factor was the total of number of electrons treated quantum mechanically, with a practical limit on the order of a few thousand electrons.

For a crystalline solid, a similar limit can be imposed for the number of electrons in a single unit cell; however the resulting geometries and energies are representative of a lattice of infinite extent.

This Chapter will briefly introduce some of the theoretical foundations for plane wave DFT, including Bloch's theorem, plane waves, and some of the approximations used for practical calculations such as the use of pseudo-potentials to represent core electrons. Next is a discussion of the building of representative surfaces, followed by a description of the calculation techniques used in this work.

4.2 Bloch's Theorem

Bloch's theorem states that in systems with a periodic potential, the electronic wavefunction can be written as a product of a plane wave and a periodic function.

$$\psi_n^k(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\, u_n^k(\boldsymbol{r}) \tag{4.1}$$

Here $\boldsymbol{k}$ is the wave vector; this conserved value is sometimes called the crystal momentum and is analogous (but not identical) to the momentum of a free electron. The wavefunction is defined over the position vector $\boldsymbol{r}$, and is multiplied by a periodic Bloch function $u_n^k(r)$. This function must match the periodicity of the potential so that the following conditions are met:

$$\begin{aligned} \psi_n^k(r+a) &= \psi_n^k(r) \\ u_n^k(r+a) &= u_n^k(r) \end{aligned} \tag{4.2}$$

The vector $\boldsymbol{a}$ represents a translation to the equivalent point in a neighboring unit cell.

Each value of $\boldsymbol{k}$ represents an allowed state within the lattice, and a particular energy eigenvalue. Unlike in discrete atoms and molecules where these states have discrete values, in crystalline solids, these states are smeared into bands, a lower energy valence band and a higher energy conduction band, which in metals overlap. The total energy of the system must be summed over possible $\boldsymbol{k}$ values. Luckily the allowed $\boldsymbol{k}$ values are periodic as well and are only uniquely defined within what is known as the first Brilliouin zone, which is the subject of the next section.

4.2.1 Reciprocal space and the first Brilliouin zone

There are 14 unique basic unit cells, called Bravais lattices, that are characterized by 3 lattice vectors a, b, c and characteristic angles α, β, and γ. Though these lattice vectors can be expressed in real space, they are often expressed in what is called reciprocal space for electronic structure calculations, defined by the reciprocal lattice vectors.

$$a^* = \frac{b \times c}{a \cdot b \times c}$$
$$b^* = \frac{a \times c}{b \cdot a \times c} \qquad (4.3)$$
$$c^* = \frac{a \times b}{c \cdot a \times b}$$

These reciprocal lattice vectors have units of $\frac{1}{length}$ and just like the real space lattice vectors a basic unit cell can be constructed. The basic unit cell is called the 1st Brillouin zone and while defining the lattice in this fashion may seem cumbersome, it has

mathematical properties that are compatible with both plane waves and the requirements of Bloch's theorem.

The wave vector k has units of $\frac{1}{length}$ and so is naturally represented as a vector in the reciprocal lattice. Further, the periodic Bloch function can now be expressed as a Fourier series expansion of plane waves in the lattice.

4.2.2 Sampling $\boldsymbol{k}$-points

Though we can limit our $\boldsymbol{k}$ vector to the 1st Brillouin zone, $\boldsymbol{k}$ is a continuous variable and so the system must be discretized in some fashion into what are called k-points. In general, the density of the k-point grid is varied based on the accuracy required for whatever property being calculated. The atomic density of the system itself can help guide an initial guess at an appropriate k-point sampling, with compact lattices requiring many k-points and sparse lattices requiring only a few k-points. This is due to a broadening of the overall bandwidth as atomic wavefunctions are superimposed. Like in molecular orbital theory, the closer nuclei are brought together, the larger the splitting of bonding and anti-bonding orbital components.

Depending on the unit cell, the required density of k-points may not be equal in all directions and there is no strict requirement that they be isotropic. This is particularly relevant when discussing surfaces as a periodic unit cell is constructed which may be quite compact in the x-y plane (surface dimension) and require a dense set of k-points, while being quite sparse in the z-direction due to the inclusion of a vacuum region.

A common method of choosing representative k-point was proposed by Monkhorst and Pack. [2] Here the k-point sampling is listed as a trio of numbers such as 7x7x1 which describes the k-point density along the a^*, b^*, and c^* reciprocal lattice vectors. Higher numbers represent higher density of the sampling along this direction. All surface calculations reported in this work have either 5x5x1 or 7x7x1 k-points.

4.2.3 Plane-wave Basis Set and Pseudopotentials

Plane-waves are particularly well suited as a basis set for a metallic crystalline lattice for several reasons. First, since the valence electrons in metals are highly delocalized across the entire lattice, a plane wave description fits this situation far better and more efficiently than localized orbital functions. Second, plane waves are in principle a complete basis set, and improve monotonically as you increase the energy cutoff, making convergence testing straight-forward and predictable. Lastly, plane-waves can be combined with psuedopotentials to reduce the computational complexity even further.

As a valence electron approaches the nucleus and interacts more strongly with the core electrons; both kinetic and potential energy increases in magnitude, and the number of nodes in the wavefunction increases significantly. To model this nodal structure would require an increasing number of high energy plane waves, but this region of the wave-function is largely uninteresting in terms of chemical bonding or physical properties. One solution is the use of pseudopotentials in the core region, which discards the nodal structure while preserving the norm or overall magnitude and smoothly transitions into

the true wavefunction at large distance r from the nuclei. These psuedopotentials are determined in all-electron atomic calculations and then used as a frozen core potential on top of which outer regions are allowed to vary. The end result is the energy cutoff can be significantly lower, requiring less plane-waves to represent the system of interest, while still capturing the essential elements of that system. An additional advantage of psuedopotentials is that they are often constructed to include scalar relativistic effects, which can play a significant role in heavy atoms and would otherwise require far more intensive calculations to model.

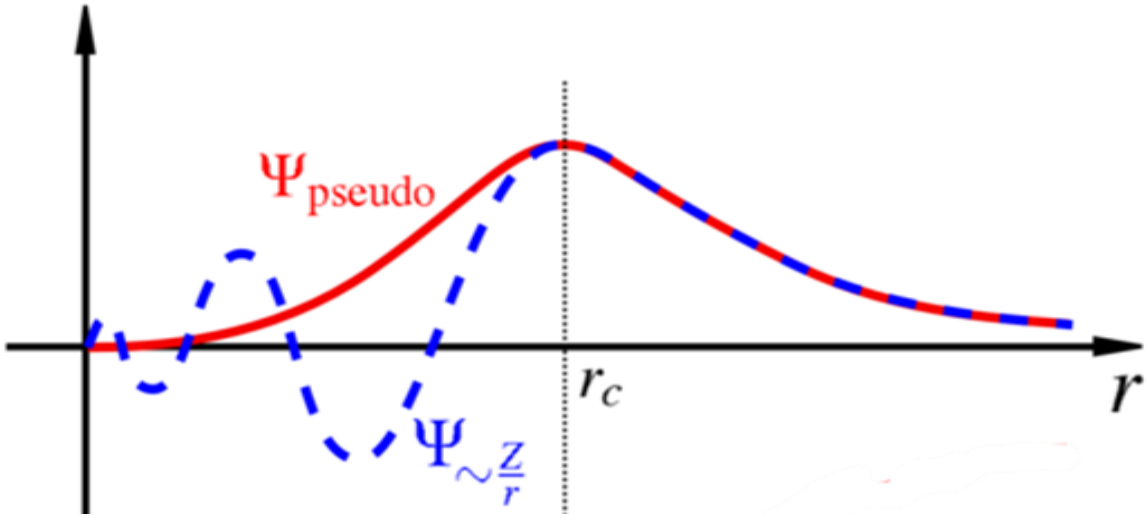


Figure 4.2 – The true wavefunction (blue, dashed) is approximated by a psuedopotential (red, solid) in the core region. This image is adapted from a public domain image found at http://en.wikipedia.org/wiki/File:Sketch_Pseudopotentials.png, created by user Wolfram Quester.

4.3 Surfaces

While the bulk is the dominant phase in all but highly dispersed nanoparticles, chemical catalysis usually occurs on the surface. This surface may be quite complex, consisting of

various crystal facets, described by their three Miller indices: h, k, and l. These indices define a vector emanating from the origin of the unit cell, and a plane which is orthogonal to this vector defines the surface cut. For example, a PdZn alloy unit cell is shown below and the 111 surface cut is shown in dotted line.

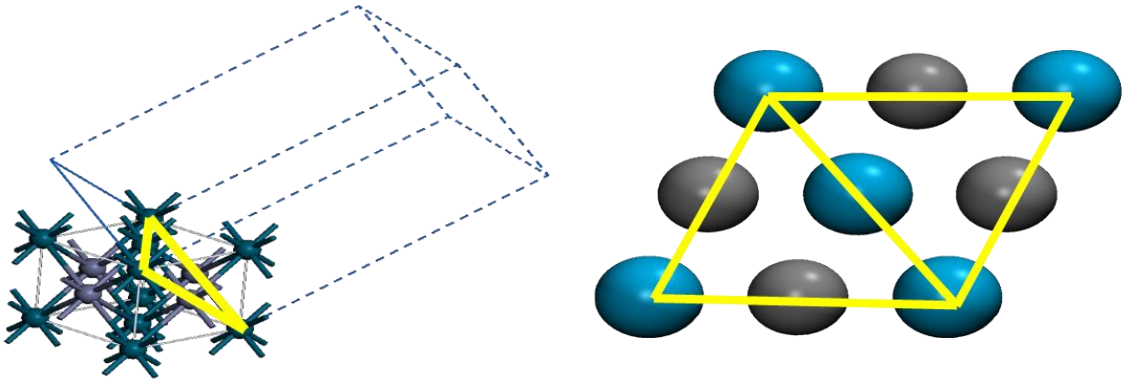


Figure 4.3 – Left panel shows a PdZn unit cell with a plane cutting orthogonal to the 111 miller indice. Right panel shows the resulting 111 surface from the top down perspective.

To describe an actual catalyst, a sampling of surface sites is chosen as representative of the real system. This is first informed by experimental results from X-ray diffraction (XRD) spectroscopy, which reveals the predominant surface sites in a catalytic sample. For example in PdZn alloy, the flat 111 and 100 surfaces dominate, each producing a distinguishable spectral peak. Any modeling of this catalyst must include these two surfaces. The flat surfaces are not the whole story however, as a catalytic site can be both rare and active, necessitating additional sampling.

These additional surfaces may be flat, stepped, kinked or consist of irregular features such as vacancies, adatoms and islands. These types of surfaces can be modeled by various surface cuts as shown below, using both PdZn and the catalytic support, ZnO.

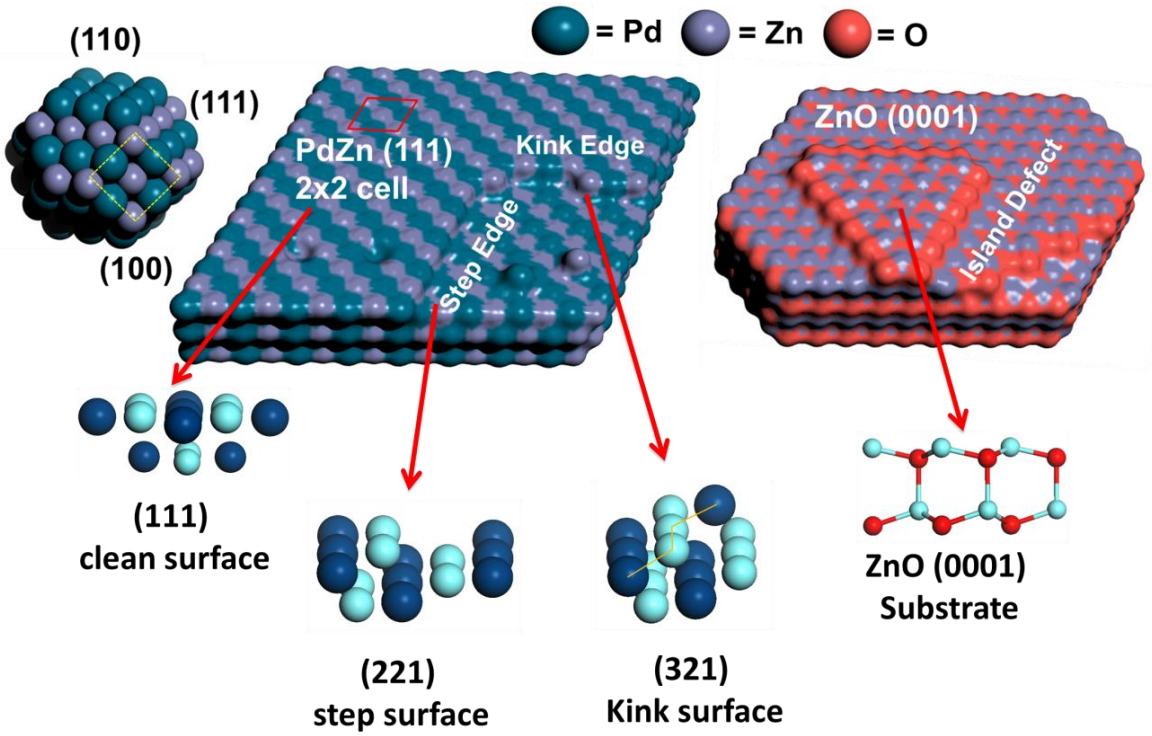


Figure 4.4 – Left side shows an example PdZn alloy surface with many defect features. Several of these defects can be modeled by a particular surface cut. Right side shows the catalytic support, ZnO, showing an island defect.

Very high energy surfaces can usually be ignored since they are unlikely to survive any surface reorganization. Though there are a huge variety of different surfaces one can include, generally surfaces with similar bond saturations will show some similarities in behavior. For example, 111 and 100 surfaces, both being flat with identical bond saturations, show very similar electronic effects on catalytic activity and selectivity.

They differ geometrically however, with the 100 surface being more open, which can affect properties like diffusion.

4.4 Some Preliminary Considerations for Building a Plane-wave DFT model

The first choice is the software package used to do electronic structure calculations. Many different options are available that differ in speed and usability. All calculations done in this work were performed using version 4.6 of the Vienna ab-initio simulation package (VASP)[3-5], using the supplied projected augmented-wave (PAW) psuedopotentials. [6,7]

The second major choice before building surface models is to decide on a functional for the DFT model. This is always a tradeoff between accuracy, computational costs, and transferability. Two common functionals used in surface modeling is PW91[8], PBE[9], and their various derivatives. Each of these functionals builds on the Local Spin Density Approximation (LSDA) by including information on the local change in density, sometimes referred to as the generalized gradient approximation or GGA. This work uses the PW91 functional, which has been shown to be well behaved for systems explored in this work. [10] Some care should be taken in the choice of functional, as this choice is difficult to change later without losing a basis of comparison between models.

The third major choice is the energy cutoff for the plane-wave basis set, and this largely depends on the parameterization of the pseudopotential per individual atoms. In this work, a cutoff of 400 eV was used for all PdZn models and a cutoff of 440 eV was used for all ZnO models. Convergence testing of higher cutoff energies showed little change.

4.5 Model Building

The main benchmark in all model building is convergence testing. Once a property stops changing as the size of the basis functions increases, the model is converged in this particular context. A balance between computation time and accuracy is usually found.

4.5.1 Building a surface

Building a surface starts from optimizing the bulk lattice parameters. A periodic unit cell containing the surface cut then is constructed by considering several aspects. First, the size of the unit cell in the x-y plane must be determined (2x2 is the smallest unit cell possible for PdZn, the next smallest is 4x4 to maintain a periodic unit cell). Second, the number of subsurface layers needed to approximate the interface between surface and the bulk phases is tested. Third, in order to model reactions on the surface while still maintaining a periodic cell, a large vacuum spacing is required so that any chemical species on the surface does not see significant interaction with the next unit cell along the z-axis. Lastly, a decision needs to be made on whether to allow the surface atoms to move or keep them fixed at an optimized bulk geometry. A compromise solution is to fix some of the lower subsurface layers at the bulk parameters while allowing the top 1-2 layers to relax.

4.5.2 Specific models used in this work.

Bulk optimization of 1:1 PdZn yielded lattice parameters of $a = b = 4.139$ Å, $c = 3.378$ Å, in good agreement with previously reported results[11]. Slab models for the 111 and 100 surfaces consisted of five layers of a 2×2 unit cell, with the top two layers

allowed to relax in all calculations. To simulate steps and kinks, the 221, 110 and 321 surfaces were used. For the 221 and 321 surfaces, five stepped layers were included with the uppermost three layers relaxed, while for the 110 surface, a total of five layers was used with the top two layers relaxed. The vacuum spacing between slabs was 14 Å. These calculations were performed with a 7×7×1 Monkhorst–Pack k-point mesh.

For the ZnO surface calculations, a similar protocol was used. The optimization of the hexagonal wurtzite ZnO crystal yielded lattice parameters of $a = b = 3.281$ Å , $c = 5.256$ Å, in good agreement with previous work [12]. The model of the defect-free polar ZnO(0001) surface consists of five double layers for a total of twenty oxygen and twenty zinc ions within a 2 × 2 unit cell. The two top layers were fully relaxed during the calculations. The large number of layers is necessary because of the polar nature of the surface. In addition, the leading errors due to the artificial dipole generated by the slab model were corrected using the methods introduced in Refs. [13,14] The defect ZnO(0001) surface was modeled with a larger 3 × 3 unit cell which contains eight double layers plus a four-atom island on the top. In this case, both the island and two layers underneath were relaxed. These calculations were performed with a 5×5x1 Monkhorst–Pack k-point mesh. The vacuum space is 18 Å for the defect-free surface and 16 Å for the defect surface.

4.6 Calculations

Bulk optimization proceeds by starting near the experimental lattice parameters and alternating ionic and electronic steps until convergence is reached. This step is done at

a k-point mesh of 11x11x11. Surface cuts are then minimized without a substrate, holding the fixed layers at the optimized bulk positions and letting the top layer relax.

Adsorption energies of chemical species on the surface are calculated as follows:

$$E_{adsorp} = E_{complex} - (E_{surface} + E_{free}) \quad (4.4)$$

Here, $E_{complex}$ consists of the surface plus substrate, $E_{surface}$ is the surface alone, and E_{free} consists of the substrate alone in a large box. All calculations are optimized to an energy tolerance of $10^{-4}\ eV$ and force tolerance of $0.05\ \frac{eV}{\text{Å}}$. Absorption energies are calculated starting at a wide variety of unique and stable surface sites.

To model chemical reactions, stationary points are found for reactant and product states on the surface. These points are connected using the climbing image nudged elastic band method.[15,16] This method connects the reactant states and product states by a series of interpolated images, each connected by a harmonic constraint. This method finds a minimum energy path between stationary points and the image highest in energy is freed from harmonic constraints and driven up the potential until the transition state is located. All stationary points are confirmed by normal mode analysis, with minima containing only positive frequencies and the transition states containing one imaginary frequency corresponding to the expected reaction coordinate. Normal frequencies are found by the finite difference method using an atomic displacement of $0.02\ \text{Å}$. These frequencies are also used to calculate zero-point energy (ZPE) corrections to the stationary points.

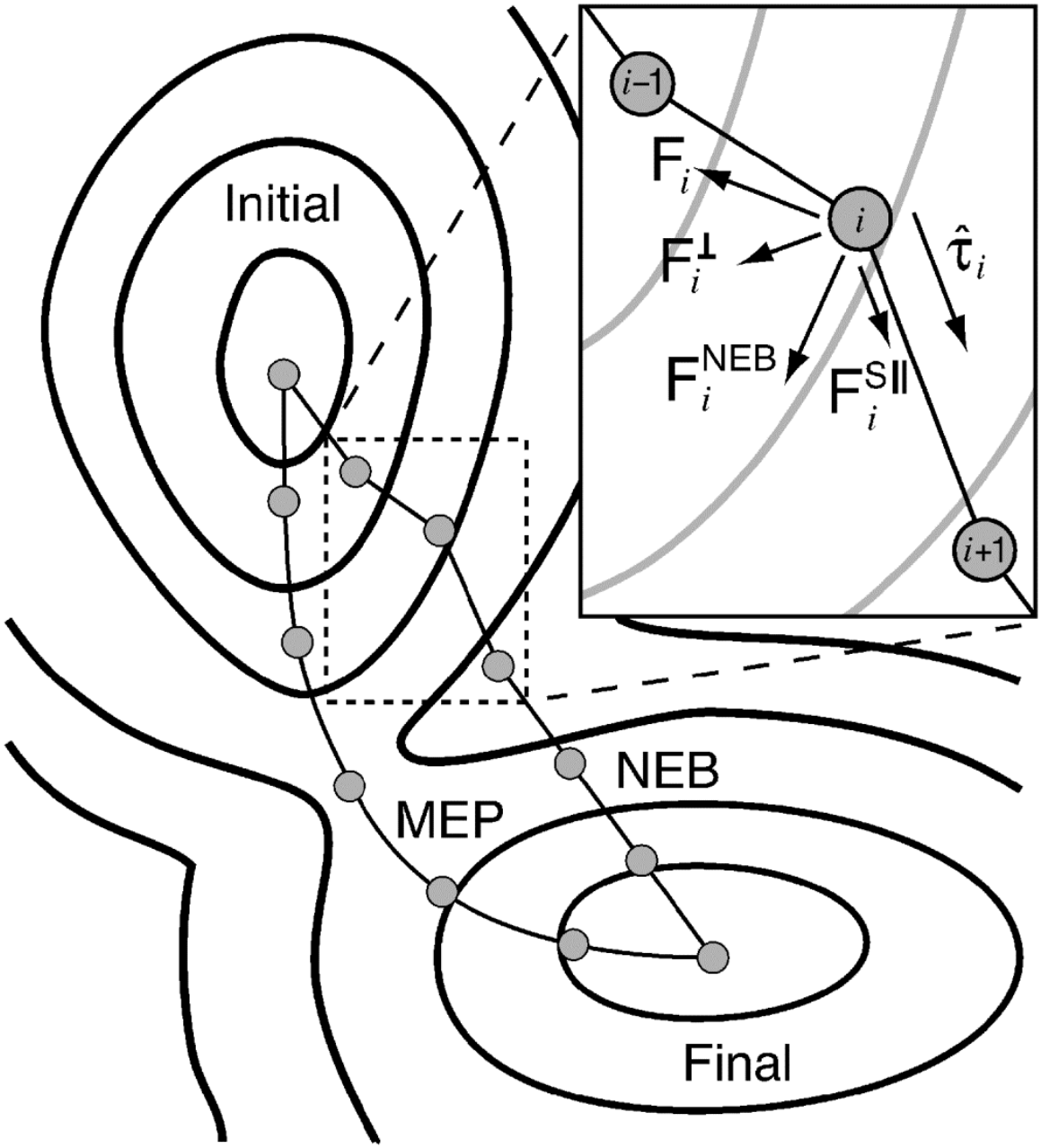


Figure 4.5 – The climbing image nudged elastic band method (CI-NEB). Several images are connected between initial and final states with harmonic potentials. Each image follows the constrained potential to the minimum energy path.

4.7 Surface methods used In this work.

Chapter 8 examines the initial steps in the MSR reaction on PdZn and ZnO surfaces using the plane-wave DFT. Stationary states and reaction barriers are calculated for O-H breaking steps in methanol and water on a variety of surfaces.

4.8 References

1 Leach, A. R. *Molecular Modeling*. (Prentice Hall, 2001).

2 Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Physical Review* **B13**, 5188 (1976).

3 Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using plane wave basis set. *Physical Review B* **54**, 11169 (1996).

4 Kresse, G. & Furthmuller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using plane wave basis set. *Computational Material Science* **6**, 15 (1996).

5 Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**, 558 (1993).

6 Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**, 1758 (1999).

7 Blochl, P. Project augmented-wave method. *Physical Review B* **50**, 17953 (1994).

8 Perdew, J. P. *et al.* Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **46**, 6671 (1992).

9 Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **77**, 3865 (1996).

10 Neyman, K. M. *et al.* Microscopic models of PdZn alloy catalysts: structure and reactivity in methanol decomposition. *Physical Chemistry Chemical Physics* **9**, 3470-3482 (2007).

11 Chen, Z.-X., Neyman, K., Gordienko, A. & Rösch, N. Surface structure and stability of PdZn and PtZn alloys:Density-functional slab model studies. *Physical Review B* **68**, 1-8 (2003).

12 Meyer, B. & Marx, D. Density-functional study of the structure and stability of ZnO surfaces. *Physical Review B* **67**, 035403 (2003).

13 Makov, G. & Payne, M. C. Periodic boundary conditions in ab initio calculations. *Physical Review B* **51**, 4014 (1995).

14 Neugebauer, J. & Scheffler, M. Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111). *Physical Review B* **46**, 16067 (1992).

15 Jonsson, H., Mills, G. & Jacobsen, K. W. in *Classical and Quantum Dynamics in Condensed Phase Simulations* (eds B. J. Berne, G. Ciccotti, & D. F. Coker) (World Scientific, 1998).

16 Henkelman, G., Uberuaga, B. P. & Jonsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics* **113**, 9901-9904 (2000).

Chapter 5

SpvC virulence factor from *Salmonella Typhimurium*

A classical MD, small model DFT, and QM/MM study of phosphothreonine lyase catalysis.

"If you do not change direction, you may end up where you are heading."

– Lao Tzu

This chapter is based on collaborative work originally found in the following two publications:

Adapted with permission from **Smith GK**, Ke Z, Hengge AC, Xu D, Xie D, Guo H. Active-site dynamics of SpvC virulence factor from Salmonella typhimurium and density functional theory study of phosphothreonine lyase catalysis. *The Journal of Physical Chemistry B*. 2009;113(46):15327-33. Copyright 2009 American Chemical Society.

Adapted with permission from Ke Z, **Smith GK**, Yingkai Z, Guo H. Molecular mechanism for eliminylation, a newly discovered post-translational Modification. *Journal of the American Chemical Society.* 2011; 133(29):11103-05. Copyright 2011 American Chemical Society.

5.1 Introduction

Cellular response to bacterial infection in both plants and animals involves a complex, highly conserved signal transduction cascade, collectively known as the mitogen-activated protein kinase (MAPK) pathway.[1] Many Gram-negative bacteria, such as *Shigella* and *Salmonella*, subvert this and other cellular pathways within the host cell by injection of bacterial enzymes called protein effectors using a Type III Secretion System (TTSS).[2,3] Figure 5.1 illustrates the basic scheme for subverting the immune response and these injected effectors are major components of bacterial virulence. As a result, they represent potential targets for indirect narrow spectrum antibiotics, a promising alternative to dealing with the growing problem of antibacterial resistance.[4,5]

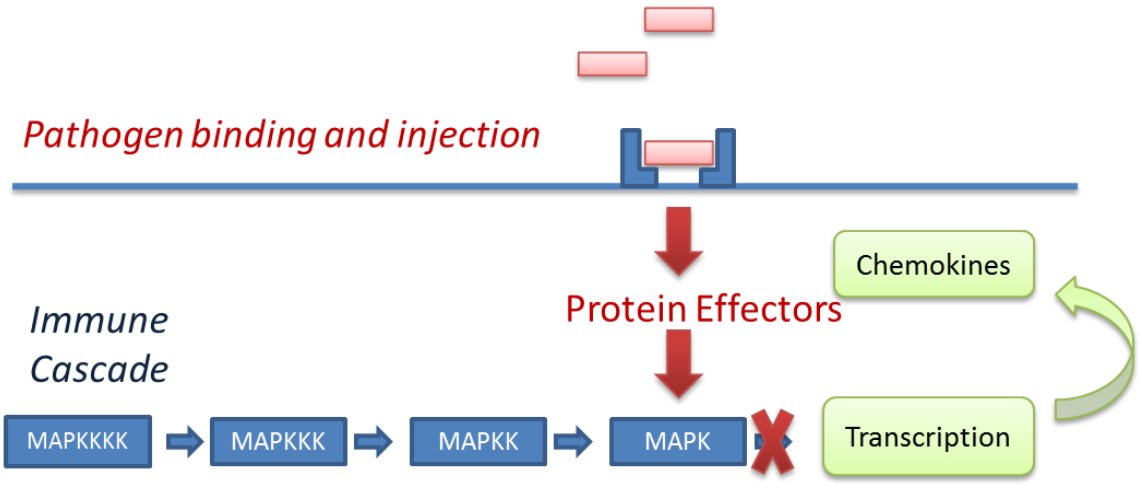


Figure 5.1 – The injected protein effectors disrupt the immune response by disrupting the signaling cascade within the cell. The effector studied in this work targets MAPK.

A subset of these effectors catalyze dephosphorylation of MAPKs into their inactive form, with examples including the acetylases found in *Yersinia pestis* and *Vibrio*

parahemeolyticus,[6,7] and the recently characterized phosphothreonine lyase family found in *Shigella*,[8] *Salmonella*,[9] and the plant pathogen *Pseudomonas syringae*.[10] The protein effector OspF found in *Shigella*, for example, dephosphorylates MAPKs such as Erk2 (extracellular signal-regulated kinase 2) and p38 in vivo, preventing histone phosphorylation and repressing activation of a subset of NF-κB responsive genes and subsequent production of proinflammatory cytokines.[11]

In contrast to the more commonly seen phosphatase catalysis which cleaves the O-P bond, the mechanism of OspF has been shown to proceed by β-elimination of phosphothreonine, breaking the $C_\beta - O_\gamma$ bond and forming a $C_\beta = C_\alpha$ double bond.[8,9] The resulting Michael acceptor product can then react further in an irreversible fashion, permanently inactivating the MAPK. This mechanism has been confirmed in another enzyme of the so-called OspF family, namely SpvC from *Salmonella typhimurium*, showing similar substrate specificity and evidence for production of the same product.[8,12] *Pseudomonas syringae* HopA1, a plant pathogenic effector and member of the same enzyme family has shown inactivation of MAPKs using the same mechanism.[10]

The β-elimination of phosphate from phosphorylated threonine and serine residues in proteins by chemical means under basic conditions is well known. In an early example, phosphoserine-containing peptides were converted to their S-ethylcysteine counterparts before subjecting them to amino acid sequencing by Edman degradation.[13] More recently, the resulting dehydroalaninyl and β-methyldehydroalaninyl residues have been

derivatized by Michael addition to attach affinity tags,[14] or to produce adducts with enhanced properties facilitating observation by fluorescence[15] or matrix-assisted laser desorption/ionization (MALDI).[16-18] In contrast to the numerous applications, evidence regarding the mechanism of the elimination reaction is scarce. The elimination in dilute hydroxide is catalyzed by group II metal ions, in the order Ba > Sr > Ca,[19] and the reaction is accelerated by the addition of DMSO or ethanol.[15]

In contrast to the chemical process, the enzymatic β-elimination of phosphate from phosphoproteins is a novel reaction and constitutes a new means for the dephosphorylation of proteins. Of known enzymatic reactions, the most analogous is that catalyzed by threonine dehydratase, which differs in the use of a pyridoxal phosphate (PLP) cofactor.[20] The details of the reaction have not been elucidated, but the mechanism is assumed to be E1cB.[21] Another PLP-dependent elimination reaction is catalyzed by threonine synthase, in which a Schiff base intermediate is formed between PLP and the substrate amine.[22] Members of the OspF family utilize neither PLP nor any other cofactors, offering a further testament to the novelty of these enzymes.

The impact of some aforementioned bacteria on public health is far reaching. *Shigella*, the cause of bacillary dysentery, is responsible for approximately 1.1 million deaths per year in the developing world.[23] Deadly *Salmonella* outbreaks are increasingly common in recent years. The rapid emergence of resistant strains compounds the problem.

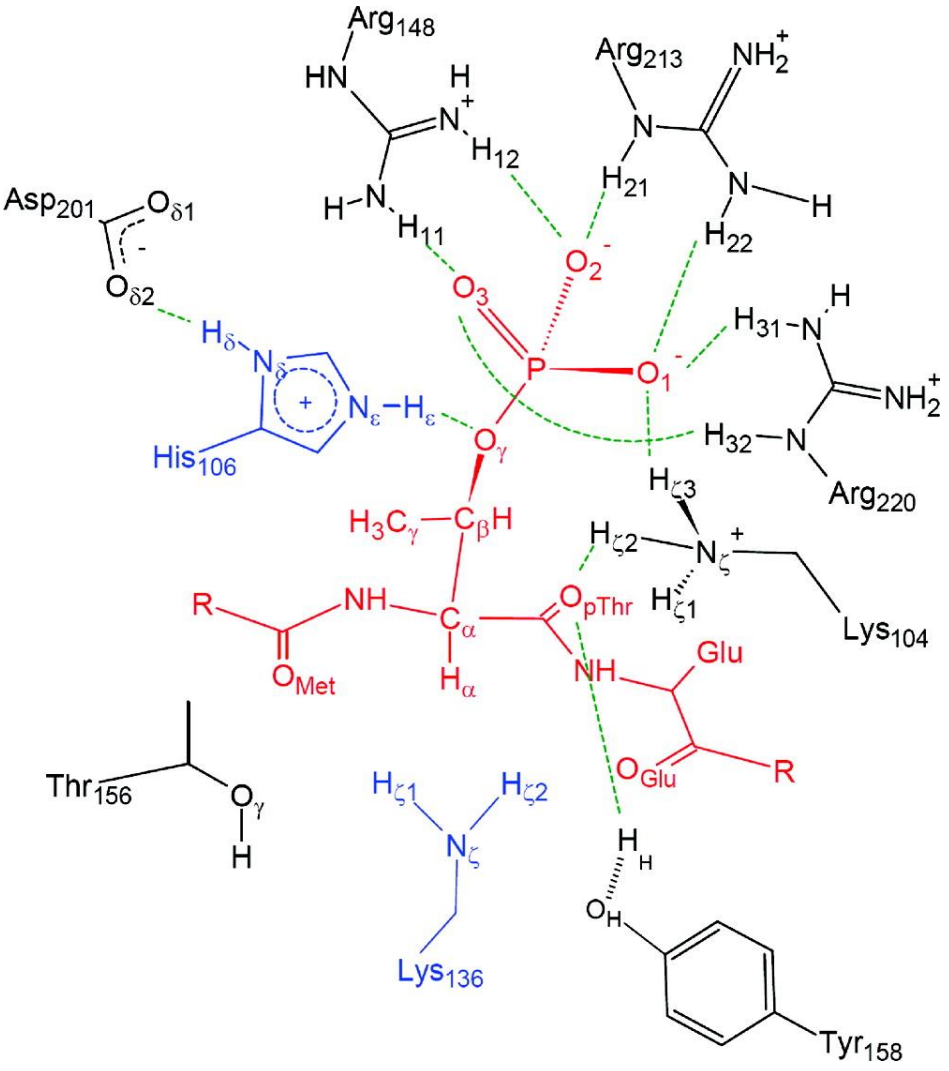
Interestingly, no eukaryotic phosphothreonine lyase has been identified, making this enzyme an attractive narrow spectrum antibiotic target. The inhibition of the pathogenic effectors differs from the many existing antibiotic paradigms and may elicit less resistance.[4,5] Understanding the mechanistic details of this family of enzymes can help guide drug discovery and may eventually lead to new antibiotics.

The crystal structure of SpvC has recently been reported by two groups in the apo form (protein only) and also for inactive mutants co-crystalized with phosphopetides.[9,24] These structures show a unique α/β fold likened to a cupped hand. Substrate specificity is determined by the pThr-X-pTyr motif of MAPKs, with pTyr resting in a surface groove and forming hydrogen bonds with Lys160, Lys134, and Arg80. The phosphate group of pThr is engulfed by a positive pocket formed by Arg220, Arg213, Arg148, and Lys104. A loop containing Arg220 undergoes a conformational change upon substrate binding, closing the positive pocket surrounding the phosphate, as shown by comparison of apo and substrate-bound structures. The residues involved in chemical transformation have been proposed based on mutagenesis and pH kinetic studies to be Lys136 and His106.[9] The proposed mechanism entails Lys136 acting as a general base, abstracting a proton from C_α of phosphothreonine, with His106 acting as a general acid to promote $C_\beta - O_\gamma$ bond breaking and the formation of a $C_\beta = C_\alpha$ double bond, resulting in the β-methyldehydroalanine product. The role of Lys136 as a general base requires the deprotonated form, which is presumably stabilized by

hydrogen bonding with Tyr158. His106's role as a general acid requires the protonated form, stabilized by Asp201, and is consistent with the reported pH rate profiles,[9] but more work is needed to definitively confirm its role in catalysis. Furthermore, mechanistic information reported to date gives no insight as to whether the enzymatic elimination is concerted or stepwise. The active-site arrangement is depicted in Scheme 5.1 and a summary of point mutation effects is given in Table 5.1.

Enzyme (SpvC)	V_{max} (μmol/min/mg)	K_m (μM)	k_{cat} (s^{-1})
WT	22.72 ± 0.38	52.2 ± 2.0	10.60 ± 0.18
R213Q	ND	ND	ND
R220Q	<0.1		
K104R	<0.1		
Y158F	1.11 ± 0.14	50.4 ± 7.7	0.52 ± 0.06
K136R	<0.1		
K136A	ND	ND	ND
H106N	ND	ND	ND
H106K	3.04 ± 0.08	49.9 ± 5.6	1.42 ± 0.04
D201N	0.48 ± 0.01	56.6 ± 3.0	0.22 ± 0

Table 5.1 – Summary of point mutation effects on the catalysis of SpvC on the Erk2 phosphopeptide.



Scheme 5.1 – Arrangement of Active Site Residues and Their Interaction with the substrate.

Definitive mechanistic action and microscopic details are often difficult to establish with experiment alone and hence theoretical studies are highly desired. In this Chapter, molecular dynamics (MD) studies of the Michaelis complex of SpvC with a phosphopeptide strand from the activation loop of Erk5 are presented, which shed light on the active-site arrangement. Furthermore, density functional theory (DFT) study of

the catalysis with a truncated active-site model are also presented, which establishes the feasibility of the proposed reaction mechanism. To guide future experimental exploration, the kinetic isotope effects based on the truncated active-site model have been determined. To understand the influence of the protein environment, a QM/MM study is carried out here, which sheds light on the catalysis in the protein environment.

This Chapter is organized as follows. In Sec. 5.2, theoretical methods used in MD, DFT, and QM/MM calculations are explained. The results are presented in Sec. 5.3 and discussed. The final section (Sec. 5.4) concludes the discussion.

5.2 Methods

The methods used in this project are described in a condensed fashion below. Foundational information and additional details on the techniques used can be found in Chapter 2 (Classical Methods) and Chapter 3 (Quantum Methods).

5.2.1 Molecular Dynamics Simulations

The starting geometry for the SpvC enzyme-substrate complex was largely based on crystal structure 2Q8Y, which consisted of an inactive K136A mutant complexed with a 13-mer phosphopetide derived from Erk5. Erk5 is closely related to the natural substrate Erk2, containing an identical pT-E-pY motif, and has been shown to have similar binding characteristics, though with a lower catalytic activity. Our substrate

model consisted of the nine amino acids that were resolved in the X-ray structure ($_{215}$YFM-pT-E-pY-VA$_{223}$). To restore the enzyme to the WT form, Lys136 was mutated back from Ala by using the mutagenesis function found in software package PyMOL (http://www.pymol.org), selecting a starting conformation that minimized steric clashes with surrounding residues and the substrate. Due to their proximity to the substrate, unresolved residues Ser96 and Gln97 were also restored by using coordinates from a mutant SpvC crystal structure (PDB identifier 2Z8P).[24] This proceeded by global structural alignment of the two crystal structures, followed by an additional alignment along residues 94-99 before transferring atom coordinates for the missing residues. The patched and immediately surrounding residues were then subjected to a short energy minimization to relieve strain and bad contacts. Unresolved N-terminal residues 1-26 were not restored, though their impact is expected to be minimal being far from the active site and substrate.

Nearly all crystal waters were retained, except those which clashed with the restored Lys136. Protonation states of all titratable residues were carefully evaluated by their local environment and hydrogen bonding networks. This included putative catalytic residues His106 in the protonated form, and Lys136 in the deprotonated form, based on the pH profile of the enzyme.[9] Hydrogen atoms were then added using the HBUILD function in the CHARMM suite (www.charmm.org). The energy of the complex was then briefly minimized, gradually releasing harmonic constraints applied to backbone

and side chain of the residues, to partially optimize the structure and remove bad contacts.

The complex was then solvated by repeated addition of a randomly rotated, pre-equilibrated 25 Å sphere of TIP3P waters,[25] with the origin of the sphere centered at the alpha carbon of the substrate phosphothreonine. Any waters that fell within 2.8 Å of a heavy atom were removed, before allowing the remainder to relax by 30ps of solvent molecular dynamics with protein and substrate fixed.

To model extended solvent effects, stochastic boundary conditions[26] were used to partition the system into three spherical regions. Atoms within a 22 Å radius from the origin are defined as the reaction zone, atoms between 22 and 25 Å as the buffer zone, and all atoms outside of 25 Å, referred to as the reservoir, are deleted from the system. Atoms in the reaction zone are governed by Newtonian dynamics on a classical potential, as provided by the CHARMM all-atom force field.[27] The buffer zone adds in a Langevin term, gradually scaled up at the boundary and consisting of frictional and random forces to represent the effect of extended solvent.

All minimizations and MD simulations were performed using the CHARMM suite of programs.[28] Non-bonded interactions were handled with an 8 Å atom-based cutoff. The SHAKE[29] algorithm was used to constrain bonds involving hydrogen. A time step of 1.0 fs was used for all dynamics simulations. Final structural refinement proceeded by

a short period of steepest-descent minimization, followed by Adopted Basis Newton-Raphson (ABNR) minimization till a gradient tolerance of 0.01 kcal mol^{-1} $Å^{-1}$ was reached. Dynamics simulations consisted of a 200ps heating window to bring the system to a final temperature of 300K, followed by a 500ps equilibration, and 2.5ns of data collection.

5.2.2 Truncated Active-site Models

To investigate the reaction mechanism of the SpvC phosphothreonine lyase, we have developed a truncated active-site model based on the Michaelis complex developed based on the X-ray structure, as described in the previous section. It contains the side-chains of several key active-site residues (His106, Lys136, Arg148, Arg213, Arg220), which have been demonstrated by mutagenesis experiments to be of critical importance.[8,9] To reduce computational costs, the arginine and lysine residues were approximated by guanidinium cations and methylamine, respectively. The substrate was truncated at the two adjacent C_α-C bonds and then was capped with methyl groups, which led to a system of 72 atoms with a total charge of +2. Admittedly, the truncated active-site model does not have all the key interactions present in the enzyme active site. Consequently, the calculation results are not expected to be quantitative. However, they may be sufficient provide some mechanistic insights into the catalytic reaction.

We are primarily interested in the stationary points along the reaction path. All the

DFT calculations were carried out using the Gaussian 03 suite of quantum chemistry programs.[30] All structures were minimized using the B3LYP functional and the standard 6-31+G(d,p) basis set. Frequency calculations at the same level have been carried out to confirm the nature of these stationary points and to obtain their zero-point energies (ZPE). Intrinsic reaction coordinate (IRC) calculations[31] were also performed to establish connectivity between the stationary points. To examine the dielectric effects from the solvent or the protein surrounding, more accurate single point energies were calculated with the larger 6-311+G(d,p) basis set and the polarized continuum model (PCM).[32] To calculate the kinetic isotope effects (KIEs), we have used the Bigeleisen-Mayer method[33] implemented in ISOEFF,[34] in which the KIEs were calculated from the harmonic frequencies of both the reactant complex and transition states. Neither anharmonicity nor tunneling correction was included.

5.2.3 QM/MM model

The QM/MM model started from the same modified structure used in part 5.2.1 for the natural enzyme. An additional mutant structure Y158F was constructed by replacing the hydroxyl group from Tyr158 with a hydrogen atom. To prepare the structures for QM/MM calculations, initial setup and minimization was done using the AMBER[35] molecular dynamics package. The models were solvated in a periodic rectangular box with dimensions $67 \times 69 \times 66\ \text{Å}^3$, leaving a minimum distance of 20 Å between adjacent protein images. This was followed by solvent minimization and dynamics with

the heavy atoms constrained by a harmonic potential of $50\ \mathrm{kcal}\cdot\mathrm{mol}^{-1}\cdot\mathrm{Å}^{-2}$. Minimization and dynamics were alternated while gradually relaxing the constraint until the system was completely unrestrained. The MD simulation employed the AMBER99SB force field[36,37] and the TIP3P water model.[25] The particle mesh Ewald method[38,39] was used to treat long range electrostatic interactions and the SHAKE[29] algorithm applied to all hydrogen covalent bonds. A 2.0 ns unrestrained MD simulation with a time step of 1 fs completed the system preparation for QM/MM studies.

A snapshot was taken from the MD simulation to use as a basis for the QM/MM models. Spherical boundary conditions were imposed, removing all atoms outside of a 27 Å radius from the C_α of the phosphorylated threonine. Atoms between 27 Å and 20 Å were kept fixed. The QM region was defined as phosphothreonine and the two adjacent backbone peptide bonds, His106, and Lys136. All other residues were described classically using the AMBER99SB force field[36,37]. The QM region was calculated using the B3LYP functional with the 6-31G(d) basis set. The QM-MM boundary atoms were treated with the psuedobond[40,41] approach. All QM/MM calculations are performed with a modified version of QChem[42] and TINKER programs[43].

After minimization of the ES complexes of both the wild type and Y158F mutant, reaction paths were explored using the reaction coordinate driving (RCD) method[44]. The 2-D potential energy surface was determined along the α-hydrogen abstraction coordinate and the $\mathrm{C}_\beta - \mathrm{O}_\gamma$ bond breaking coordinate. Many reaction coordinates

were tested, and the coordinate $\mathrm{RC} = \mathrm{r}_{\mathrm{C\beta-O\gamma}} - \mathrm{r}_{\mathrm{H\alpha-N\zeta}}$ was chosen based on the smoothness and height of the energy profile, and used as the basis for potential of mean force (PMF) calculations. The PMF was calculated using umbrella sampling[45] and the weighted histogram analysis method[46].

5.3 Results and Discussion

5.3.1 Active-site arrangement

Figure 5.2 illustrates the binding site of SpvC. The phosphopeptide substrate is bound along a shallow groove on the surface of the SpvC enzyme with the anionic phosphoryl group of pThr encircled by three positively charged Arg residues (Arg148, Arg213 and Arg220) and Lys104. The strong electrostatic interaction at this site, along with that in another binding site for pTyr two residues downstream, provide two anchoring points for the substrate and is thus likely responsible for the observed substrate specificity.

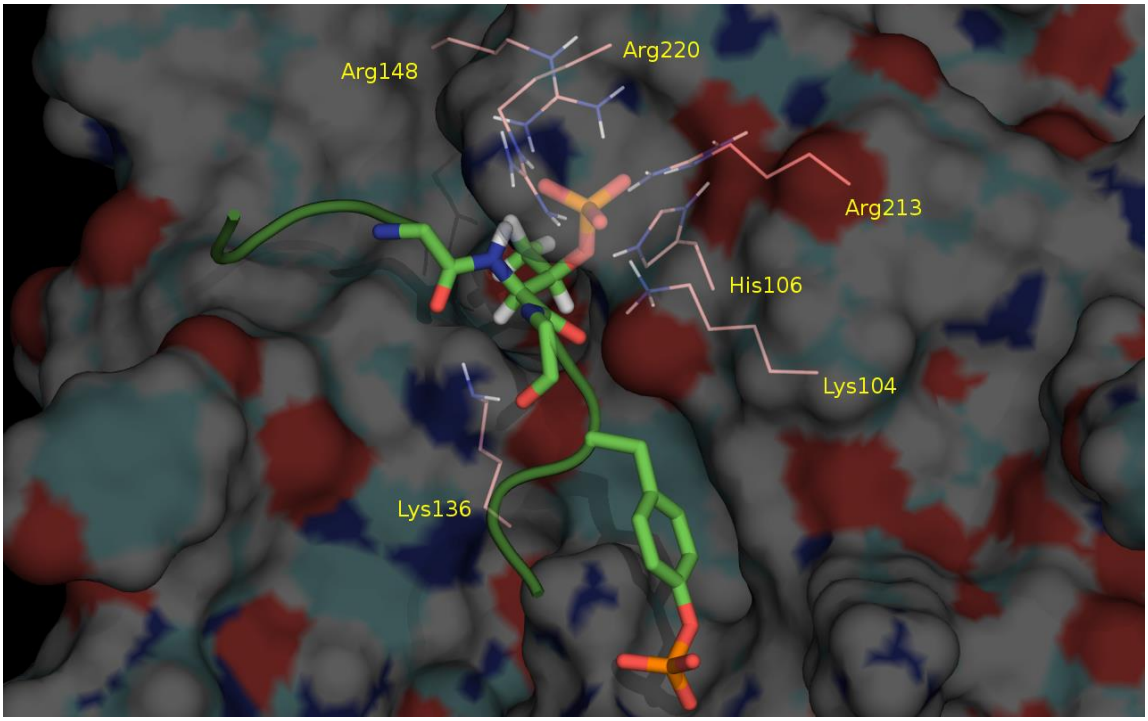


Figure 5.2 - Binding pocket for the phosphorylated peptide (green) on the surface of SpvC. The phosphoryl group of pThr is surrounded by a crown of positive residues such as Arg148, Arg213, and Arg220 as well as Lys104.

Figure 5.3 displays the minimized geometry of the active site of the Michaelis complex. In particular, the N_ζ atom of Lys136 is close to the H_α atom of pThr for proton abstraction. The Lys136 base is loosely held in place by two hydrogen bonds between its two hydrogen atoms, $H_{\delta 1}$ and $H_{\delta 2}$, and the backbone oxygen atoms of substrate residues Met (O_{Met}) and Glu (O_{Glu}), as well the enzyme hydroxyl oxygen of Tyr158 (O_H). In addition, the H_ε atom of His106 forms a hydrogen bond with the bridging oxygen O_γ of pThr, setting the stage for protonation of the phosphate leaving group, while its H_δ is stabilized by $O_{\delta 2}$ of Asp201 via a hydrogen bond.

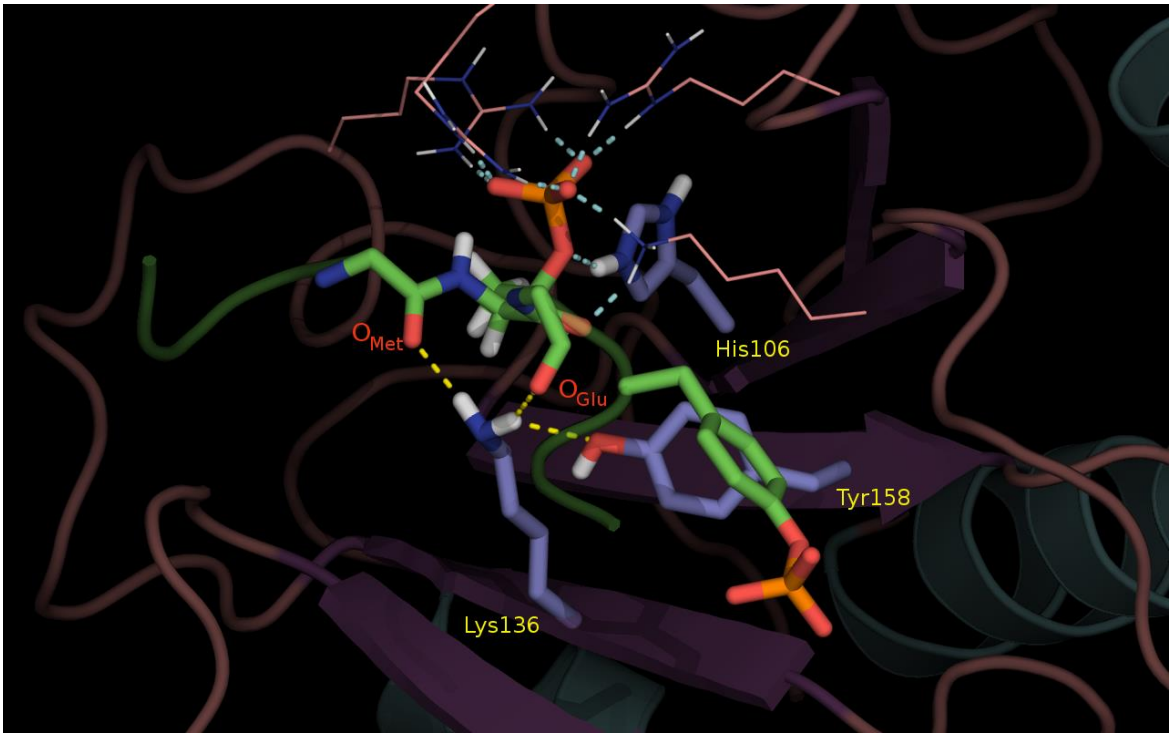


Figure 5.3 - Active-site arrangement of key residues in the Michaelis complex. The hydrogen bonds are illustrated in dashed lines.

The MD simulation began from the minimized geometry and consisted of 700 ps of heating and equilibration, at which point the RMSD stabilized (see Fig. 5.4), and the final 2.5 ns of simulation data were used to calculate bond distances and their fluctuations. Selected key distances are summarized in Table 5.2.

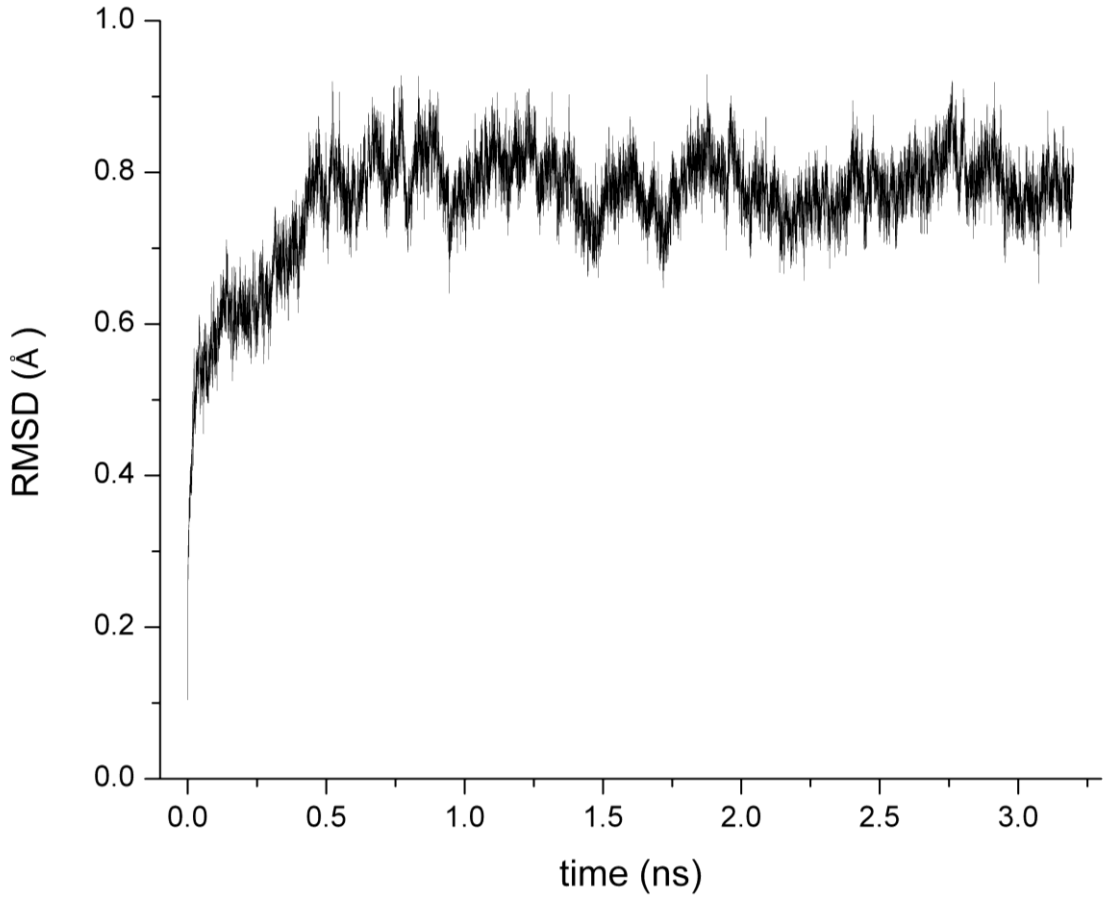


Figure 5.4 - RMSD of the 3.2 ns MD simulation.

We note that the positive pocket surrounding the phosphoryl group formed by Arg148, Arg213, Arg220, and Lys104, was quite stable during the simulation, as evidenced by the corresponding short hydrogen bond distances in Table 5.2. The dynamics as shown in Fig. 5.5 also suggests that the hydrogen bond between the H_{ε} atom of His106 and the bridging oxygen (O_{γ}) of pThr is reasonably stable, although larger fluctuations are present.

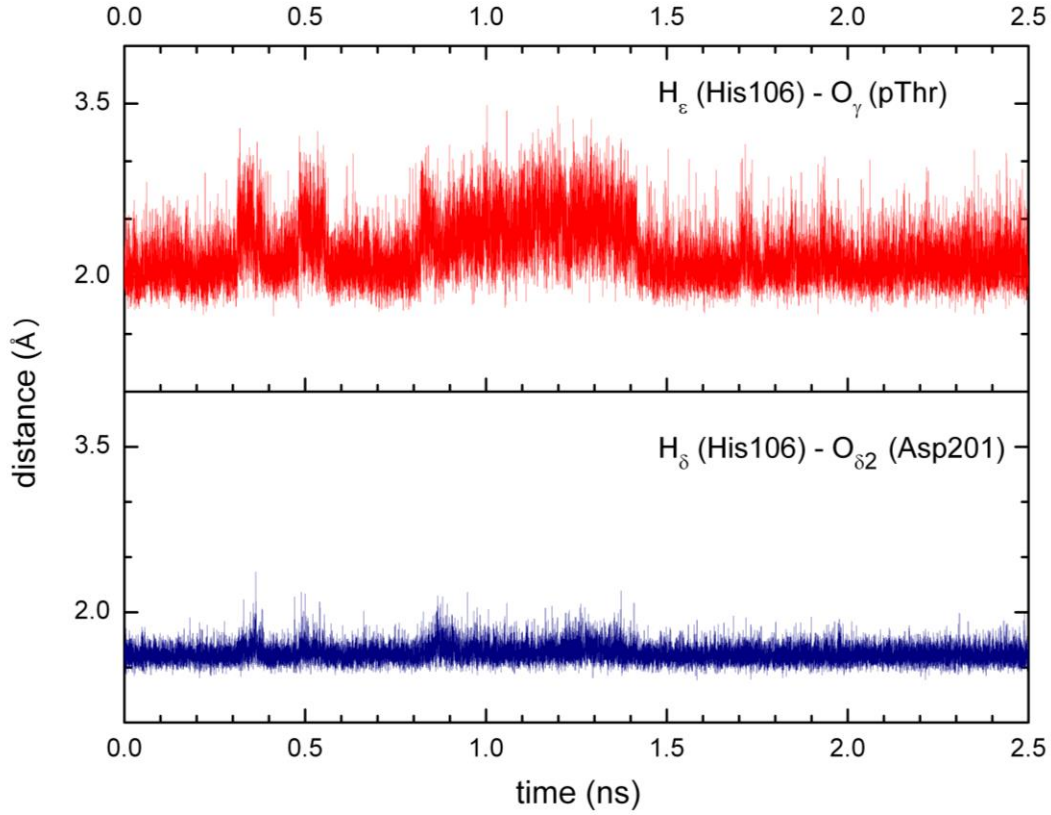


Figure 5.5 - Fluctuation of selected key distances involving the His106 residue.

The basicity of the His106 side chain is enhanced by its interaction with Asp201, evidenced by a strong hydrogen bond between H_δ of His106 and $O_{\delta 2}$ of Asp201, also shown in Fig. 5.5. This is vital if His106 is to serve as a general acid and transfer H_ε to the leaving phosphate group. More mechanistic aspects are discussed below, but this model supports His106's important role in both substrate binding and catalysis.

On the other hand, two possible conformations for Lys136 were observed as illustrated in Figs. 5.6 and 5.7. The primary conformation corresponds closely to the minimized structure shown in Fig. 5.3.

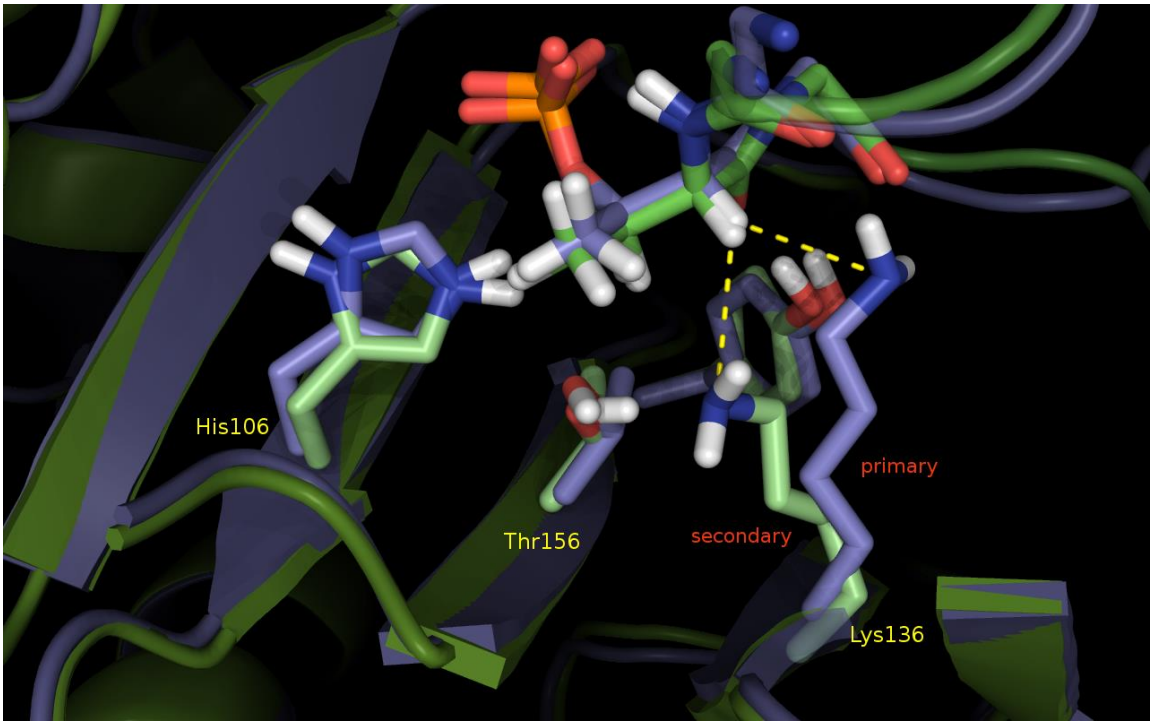


Figure 5.6 – Overlay of snapshots representing the two conformations observed in the MD simulation. The distance between N_ζ and H_α, as indicated by dashed lines, remains largely unchanged in the two conformations.

The secondary conformation occurs before 0.25 ns where Lys136 loses hydrogen bonds with the backbone carbonyl oxygen atoms O_{Tyr} and O_{Glu}, swinging under the peptide substrate to associate with the hydroxyl oxygen of Thr156 (O_H). This is clearly reflected in the corresponding distances shown in Fig. 5.7. Interestingly, the $N_\zeta - H_\alpha$ distance is seen in Fig. 5.7 to stay relatively constant during this switch, making proton abstraction from the alpha carbon a possibility in either conformation. Lys136 is also partially exposed to solvent in our model, but the solvation might be replaced by tighter protein-protein interaction between SpvC and its MAPK substrate.

Table 5.2 – Selected key distances and their average fluctuations in the Michaelis complex obtained from the MD simulation.

Residue	Bond	distance (Å)	
		primary	*secondary*
Lys136			
	$N_\zeta - H_\alpha(pThr)$	2.81 ± 0.26	2.78 ± 0.23
	$N_\zeta - C_\alpha(pThr)$	3.69 ± 0.23	3.64 ± 0.20
	$N_\zeta - C_\gamma(pThr)$	5.24 ± 0.35	3.69 ± 0.23
	$N_\zeta - O_{MET}$	3.63 ± 0.34	4.65 ± 0.35
	$N_\zeta - O_{GLU}$	3.79 ± 0.38	6.00 ± 0.35
	$N_\zeta - O_H(Tyr158)$	3.57 ± 0.31	4.66 ± 0.26
	$N_\zeta - O_\gamma(Thr156)$	5.53 ± 0.35	3.12 ± 0.35
His106			
	$H_\varepsilon - O_\gamma(pThr)$	2.20 ± 0.26	-
	$H_\delta - O_{\delta2}(Asp201)$	1.62 ± 0.08	-
Arg148			
	$H_{11} - O_3(pThr)$	1.65 ± 0.08	-
	$H_{12} - O_2(pThr)$	1.67 ± 0.09	-
Arg213			
	$H_{21} - O_2(pThr)$	1.69 ± 0.09	-
	$H_{22} - O_1(pThr)$	1.66 ± 0.09	-
Arg220			
	$H_{31} - O_1(pThr)$	1.71 ± 0.10	-
	$H_{32} - O_3(pThr)$	1.71 ± 0.10	-
Lys104			
	$N_\zeta - O_1(pThr)$	2.63 ± 0.07	-
	$N_\zeta - O_\gamma(pThr)$	3.27 ± 0.17	-
	$N_\gamma - O_{pThr}$	2.81 ± 0.12	-
	$H_{\zeta3} - O_1(pThr)$	1.61 ± 0.08	-
	$H_{\zeta3} - O_\gamma(pThr)$	2.73 ± 0.19	-
	$H_{\zeta2} - O_{pThr}$	1.86 ± 0.15	
Tyr158			
	$H_H - O_{pThr}$	2.06 ± 0.26	-

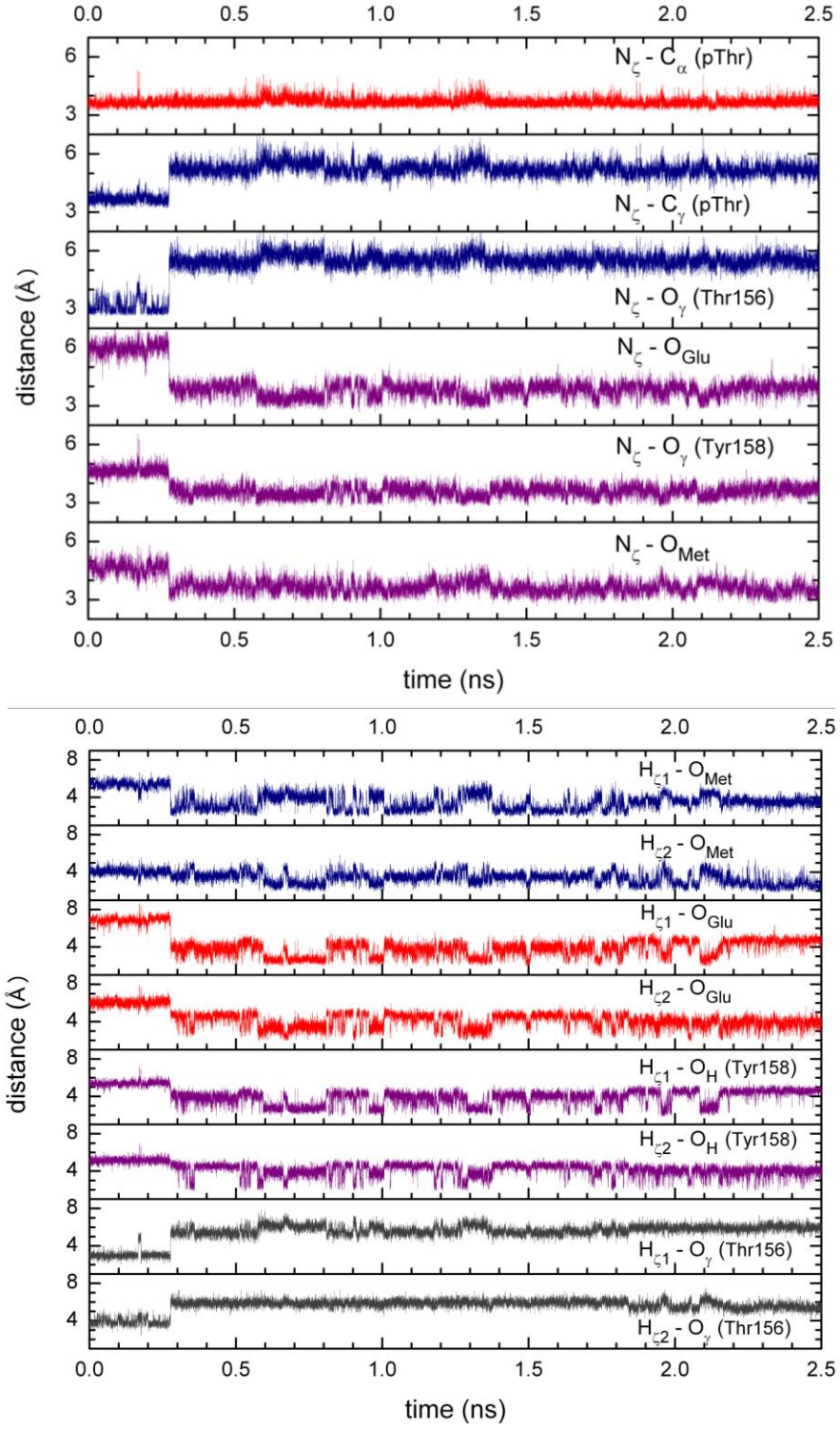


Figure 5.7 – Fluctuation of selected key distances involving N_ζ (upper panel) and $H_{\zeta1}/H_{\zeta2}$ (lower panel) of Lys136.

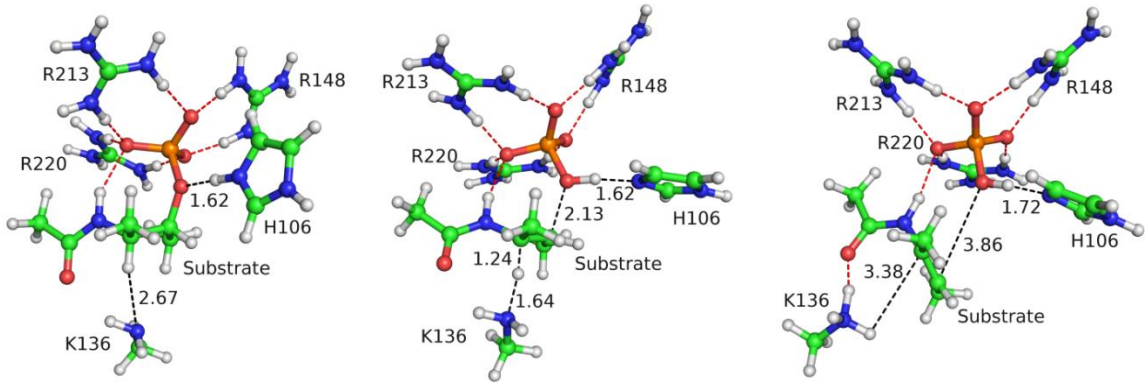
An important issue in the proposed catalytic mechanism is the acidity of the H_α atom, as no Schiff base is involved in the phosphothreonine lyase reaction. Instead, it has been proposed that the enzyme utilizes active-site hydrogen bonds to reduce the pK_a of H_α in pThr.[24] Indeed, our MD simulations revealed a very strong hydrogen bond between the amine group of Lys104 and the carbonyl oxygen of substrate pThr, as evidenced by the $H_{\zeta 2} - O_{pThr}$ distance of 1.86 ± 0.15 Å. This is consistent with the experimental observation that mutation of Lys104 inactivates the enzyme.[9,24] The pThr backbone oxygen makes an additional hydrogen bond interaction with the hydroxyl group of Tyr158 ($r(H_H - O_{pThr})$= 2.06 ± 0.26 Å). Finally, the alpha hydrogen may be further acidified by hydrogen bonds between $H_{\zeta 1}/H_{\zeta 2}$ of Lys136 and the backbone oxygen (O_{Met}) of the adjacent Met residue, as shown in Fig. 5.7. It is conceivable that these electrostatic interactions facilitate the unusual proton abstraction by the nucleophilic Lys136.

5.3.2 Truncated active-site model

The optimized structure of the reactant complex (RC) in the truncated active-site model is given in Fig. 5.8, and the key geometric parameters are labeled in the figure. The hydrogen bond between the imidazole H_ε of His106 and the bridge oxygen O_γ of pThr is 1.62 Å, which is in reasonably good agreement with the crystallographic distance between N_δ and O_γ of 2.90 Å, and shorter than the MD distance of 2.22±0.26 Å. The distance between the methylamine nitrogen (N_ζ) and the proton (H_α) that has been

proposed to transfer from the substrate to Lys136 during the reaction, is 2.67 Å, comparable to the MD distance of 2.81 ± 0.26 Å. In addition, the phosphate group is tightly bound with the guanidinium groups representing the positively charged arginines, closely resembling the X-ray structure and the MD results discussed above. Thus, this structure was used as the initial structure for subsequent optimizations.

Only one transition state (TS) was found, and it features a concerted reaction mechanism. As shown in Fig. 5.8, the $C_\beta - O_\gamma$ bond at TS is largely broken, with a bond length of 2.13 Å, and the proton is almost transferred from N_δ of the imidazole to O_γ, which results in the dianion of inorganic phosphate. In the meantime, the $C_\alpha - H_\alpha$ bond is elongated to 1.24 Å, while the distance between H_α and N_ζ of the methylamine decreases to 1.64 Å, signaling the proton in flight.



RC TS PC

Figure 5.8 – Structures of the stationary points along the β-elimination reaction path obtained at the B3LYP/6-31+G(d,p) level of theory. The hydrogen bonds are highlighted by red dashed lines, while the reaction coordinates are labeled in black dashed lines.

The reaction coordinate, which has an imaginary frequency of 338*i* cm^{-1}, is a combination of proton transfer and C-O bond stretch. The predicted kinetic isotope effects (KIEs) are listed in Table 5.3 and they reflect the involvement of various atoms in the reaction path. Note that the hydrogen KIEs may be underestimated because no tunneling is taken into consideration.

Table 5.3 - Calculated kinetic isotope effects from the truncated active-site model

Residue	Atom	KIE
pThr		
	H_α	3.119
	C_α	1.006
	C_β	1.031
	O_γ	1.016
His106		
	H_ε	1.184

The truncated active-site model shows a concerted or E2 mechanism, which is one of three possible mechanistic pathways shown in Figure 5.9. In the E1cB mechanism, the deprotonation takes place first to form a carbanion intermediate, which subsequently eliminates the phosphate leaving group. The E1 mechanism, on the other hand, creates a carbocation first due to elimination, followed by deprotonation. Transition states for the E1cB and E1 mechanisms were not found in the truncated model, however they cannot be ruled out since many nearby residues that could possibly stabilize a carbanion or carbocation intermediate are not included in this model.

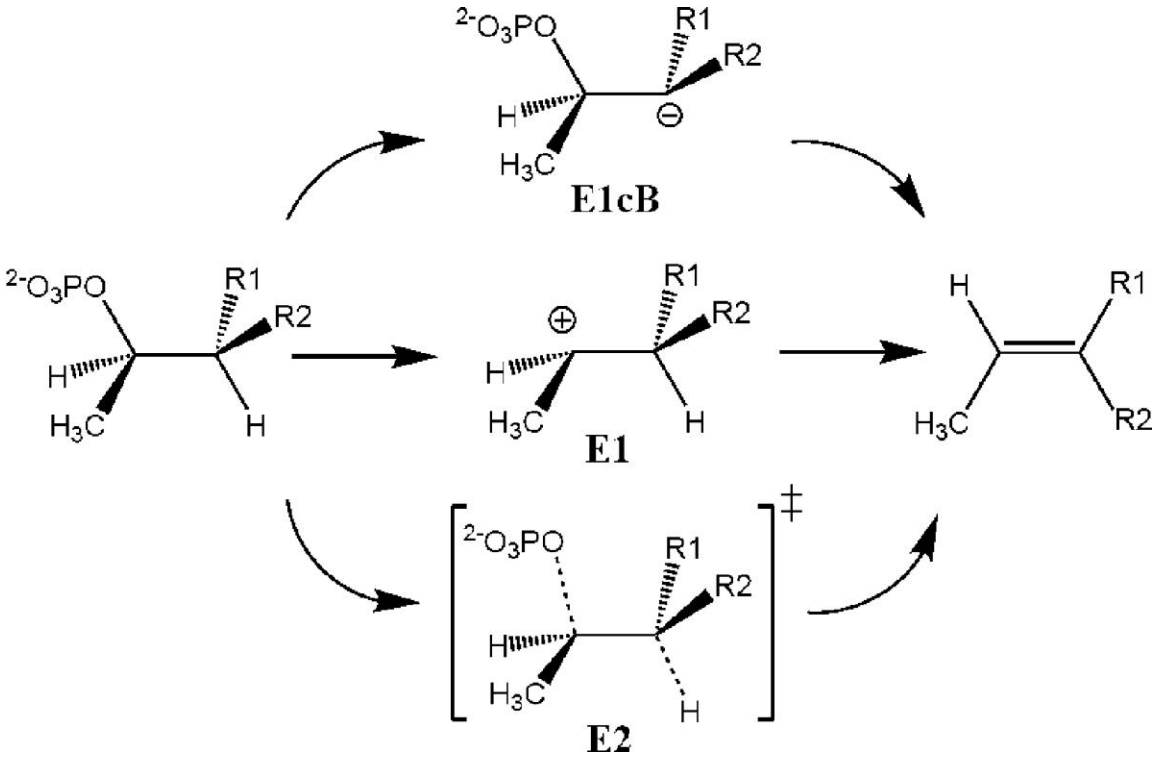


Figure 5.9 – Three possible reaction mechansims for β-elimination

The intrinsic reaction coordinate (IRC) method was used to yield the product complex (PC) shown in Fig. 5.8, in which the scissile C-O bond is clearly cleaved and the proton transfers completed. As shown in Table 5.4, the free energy barrier for this reaction is 17.33 kcal/mol in the gas phase, which is close to the experimental barrier heights suggested by k_{cat} for several substrate/enzyme combinations.

It should be noted that the backbone carbonyl group is not polarized by the hydrogen bond to Lys104 or Tyr158 in the truncated active-site model, which might contribute to the barrier.

Table 5.4 - Free energies (kcal/mol) of the RC, TS and PC for the truncated active-site model at the B3LYP/6-31+G(d,p) level of theory. The dielectric effects of water and protein are included with PCM by using the dielectric constants of 80 and 5, respectively. The experimental barrier heights for several substrate/enzyme combinations[9] are also listed for comparison.

	RC	**TS**	**PC**
ΔG_{gas}	0.00	17.33	-23.40
$\Delta G_{protein}$	0.00	24.34	-13.22
ΔG_{water}	0.00	26.09	-11.18
$\Delta G_{exp}\ (Erk5/OspF)$	0.00	18.24	-
$\Delta G_{exp}\ (Erk2/OspF)$	0.00	17.65	-
$\Delta G_{exp}\ (Erk5/SpvC)$	0.00	16.04	-

The barrier increases somewhat when solvation contributions obtained using the PCM model are added in, indicating strong solvent effects. Although several key residues have been included in the truncated active-site model, it might still be insufficient to correctly represent the enzymatic environment. Indeed, QM/MM calculations show a different mechanism and highlight the importance of two of the surrounding residues (Tyr158 and Lys104). Details follow in the next section.

5.3.3 QM/MM

The inclusion of the protein environment revealed a different mechanism than the small model system. In contrast to the concerted mechanism (E1) suggested by the truncated model, the QM/MM model showed an E1cB mechanism with a carbanion

intermediate after proton abstraction, followed by $C_\beta - O_\gamma$ breaking. The two dimensional potential energy surface was mapped out by a grid of points and is shown in Figure 5.10.

This difference is attributed to the inclusion of residues Lys104 and Tyr158 which both form hydrogen bonds with the carbonyl oxygen along the protein backbone adjacent to the carbanion intermediate. In the presence of these residues, the carbanion can be stabilized by an enolate resonance structure.

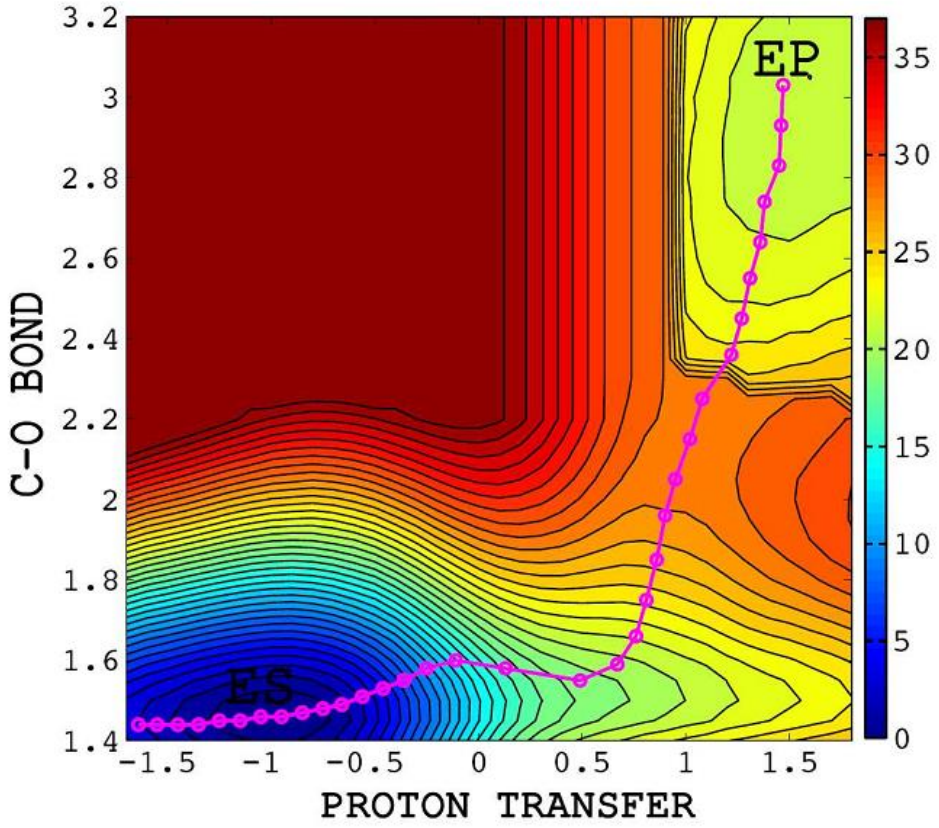


Figure 5.10 – 2D potential energy surface showing proton abstraction ($H_\alpha - C_\alpha$) shown on the x-axis, versus C-O bond breaking ($C_\beta - O_\gamma$) on the y-axis. The pink line shows a single reaction path from the combined coordinate, overlaid on the grid.

The mutant structure, Y158F, reinforces this interpretation, as without the tyrosine residue to stabilize the carbanion, the reaction barriers increase by several kcal/mol and the shallow well intermediate disappears, replaced by a shoulder. Though the reaction would still be considered as E1cb, it is further along the continuum of mechanisms towards concerted. It may be the case that if Lys104 and Tyr158 were included in the truncated model, this may be enough to shift the reaction to the E1cb mechanism.

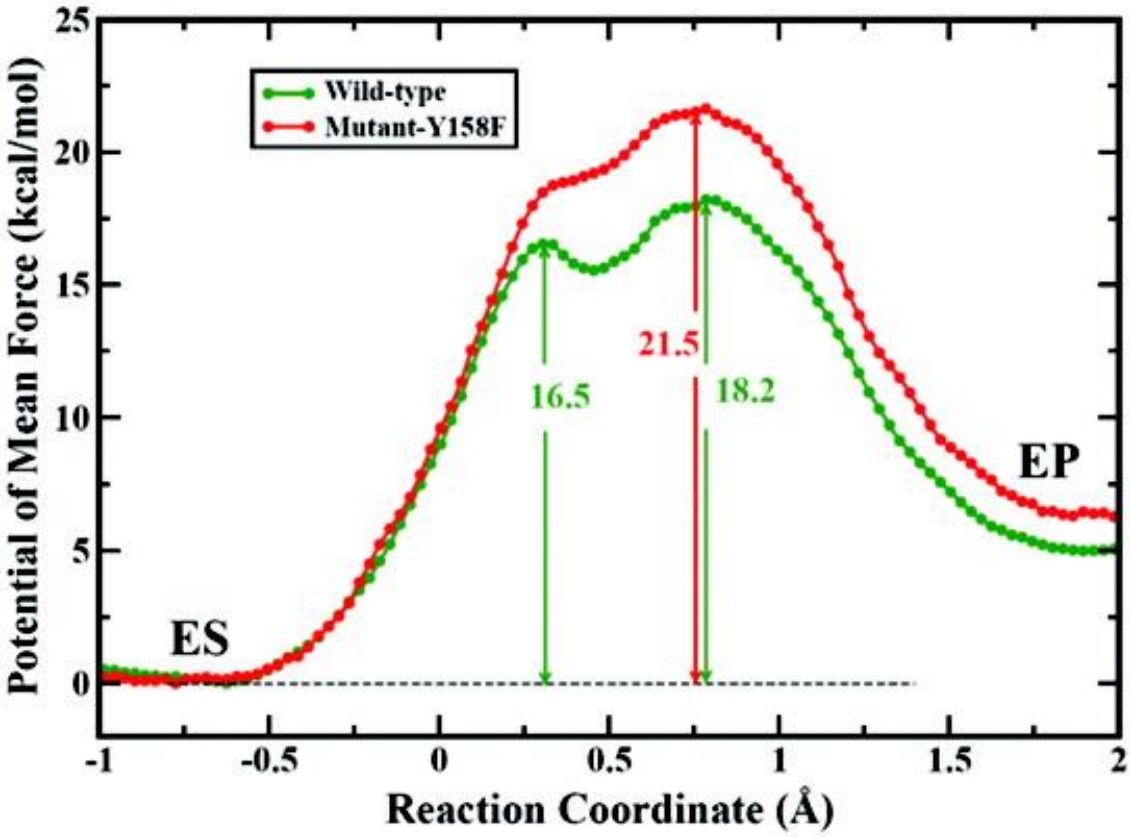


Figure 5.11 - PMFs for the wildtype and Y158F mutant from QM/MM calculations. $\left(\text{Reaction Coordinate} = \mathrm{r_{C\beta-O\gamma}} - \mathrm{r_{H\alpha-N\zeta}}\right)$.

The PMF for the wildtype, shown in Fig. 5.11, shows two transition states, with the proton abstraction barrier at 16.5 kcal/mol, leading to the shallow well of the carbanion

intermediate, and followed by the rate limiting step involving $C_\beta - O_\gamma$ breaking at 18.2 kcal/mol. This compares reasonably well with the experimental value of 16.0 kcal/mol estimated from the value of 10.60 s^{-1} for k_{cat}.[9]

The mutant PMF, which removes Tyr158 from the active site, shows a single transition state, where the proton abstraction proceeds first and is represented by a shoulder, and the $C_\beta - O_\gamma$ breaking transition state shows a barrier of 21.5 kcal/mol. This is consistent with experimental data for a Y158F mutant which retains only 5% of the catalytic activity of the wild type.[9]

Mulliken charge analysis shows that the charge on the adjacent carbonyl oxygen increases from -0.77 (ES) to -0.83 (INT) and hydrogen bond distances decrease from 1.95 ± 0.19 Å to 1.83 ± 0.12 Å for Lys104 and 1.81 ± 0.16 Å to 1.79 ± 0.12 Å for Tyr158.

Table 5.5 – Key distances (Å) for stationary points along the reaction path.

	ES	TS-I	INT	TS-II	EP
H_α(pT)-N_ξ(K136)	2.09±0.09	1.21±0.07	1.06±0.05	1.04±0.04	1.03±0.04
H_α(pT)-C_α(pT)	1.12±0.05	1.62±0.16	2.39±0.71	2.92±0.35	3.28±0.27
C_β(pT)-O_γ(pT)	1.45±0.06	1.51±0.07	1.51±0.06	1.84±0.06	2.93±0.12
H_ε(H106)-O_γ(pT)	2.33±0.40	2.01±0.25	1.93±0.20	1.77±0.16	1.12±0.18
H_ε(H106)-N_ε(H106)	1.03±0.05	1.03±0.04	1.03±0.04	1.06±0.05	1.53±0.25
C_α(pT)-C_β(pT)	1.55±0.07	1.52±0.06	1.51±0.06	1.42±0.05	1.34±0.06
C_α(pT)-C(pT)	1.53±0.07	1.47±0.06	1.46±0.07	1.44±0.05	1.49±0.06
C_α(pT)-N(pT)	1.46±0.06	1.47±0.06	1.46±0.05	1.45±0.05	1.43±0.06
C(pT)-O(pT)	1.24±0.05	1.26±0.05	1.26±0.04	1.27±0.04	1.25±0.05
H(Y158)-O(pT)	1.81±0.16	1.75±0.15	1.79±0.15	1.77±0.15	1.77±0.15
H_2(K104)-O(pT)	1.95±0.19	1.86±0.14	1.83±0.12	1.85±0.12	1.87±0.12

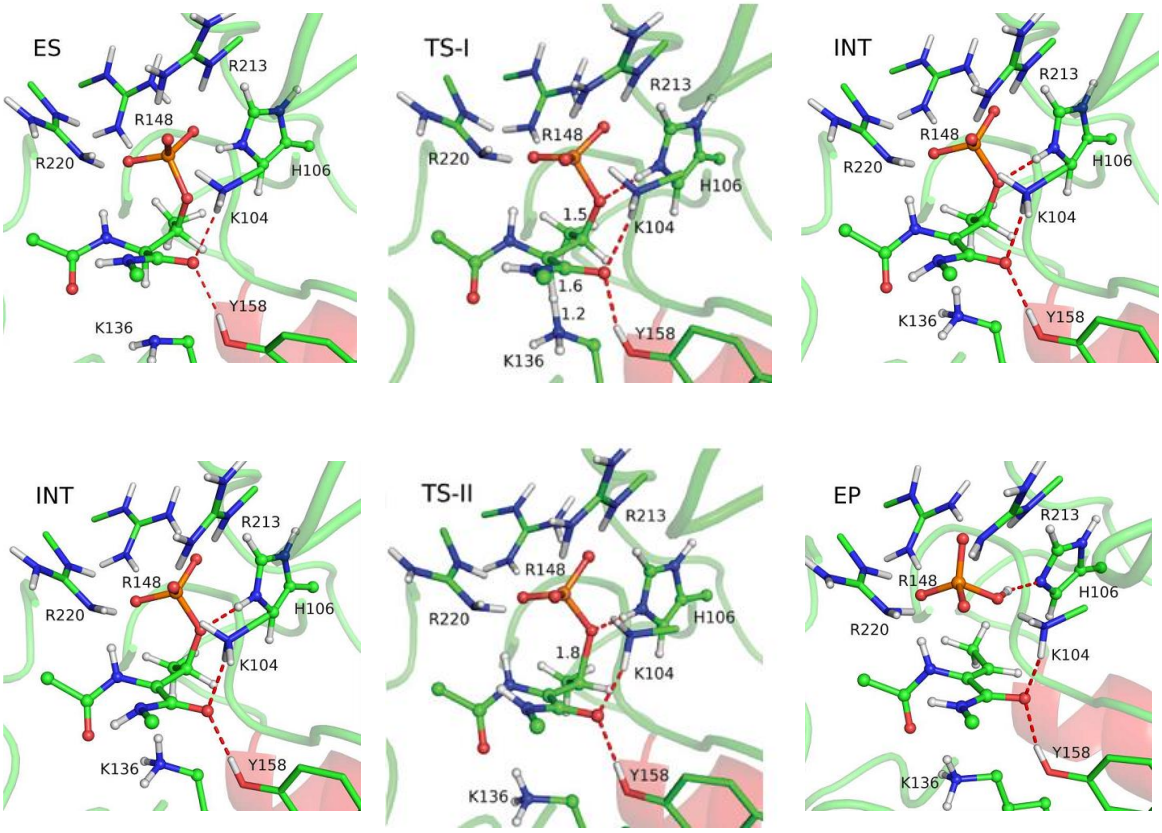


Figure 5.12 – Stationary states from the QM/MM model.

5.4 Conclusions

The OspF family of protein effectors differs significantly from the more well-known phosphatases in that they cleave the C-O bond instead of the O-P bond. It has been argued that there are important structural differences in the active sites of the two types of enzymes. In phosphothreonine lyases, such as SpvC, the phosphate moiety of the substrate is completely insulated from solvent. As a result, no hydrolysis is possible. On the other hand, water molecules are allowed to access the active site and they can thus be activated to hydrolyze the phosphorylated intermediate. The

completely different dephosphorylation mechanisms are associated with different transition states. As such, commonly used transition-state analogs for phosphatases, such as vanadate, are not effective in inhibiting phosphothreonine lyases.[12]

The MD simulations of the Michaelis complex of SpvC strongly support the notion that Lys136 is the base, based on its proximity to the acidic alpha hydrogen. This observation is consistent with the observation that mutation of the conserved lysine at this critical position (136 in SpvC and 134 in OspF) completely abolishes the catalytic activity of the effector.[8,24] In addition, His106 is shown in our MD simulations to donate a hydrogen bond to the bridge oxygen, thus setting the stage for the protonation of the phosphate leaving group. This is also consistent with mutagenesis experiments.[8,24] Finally, there exists a network of hydrogen bonds that polarize backbone carbonyl oxygen atoms, which may contribute to the lowered $\mathrm{pK_a}$ of $\mathrm{H_\alpha}$.

These structural determinants for the enzymatic reaction are confirmed by the DFT reaction path, which features a concerted transition state in which the proton abstraction and C-O bond cleavage occur concurrently. The QM/MM model showed the importance of residues Lys104 and Tyr158 in stabilizing a carbanion intermediate, resulting in an E1cB mechanism. To facilitate future comparison with experimental investigations, we have also computed kinetic isotope effects (KIEs) for several key atoms.

To summarize, the computational studies reported here provide detailed information on the interaction of active-site residues with the substrate and important insights into the novel β-elimination mechanism operating in the OspF family of effectors, thus laying the foundation for future studies of this potentially important drug target.

5.5 References

1 Dong, C., Davis, R. J. & Flavell, R. A. MAP kinases in the immune response. *Annual Review of Immunology* **20**, 55-72 (2002).

2 Cornelis, G. R. The type III secretion injectisome. *Nature Reviews Microbiology* **4**, 811-825 (2006).

3 Galan, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**, 567-573 (2006).

4 Clatworthy, A. E., Pierson, E. & Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nature Chemical Biology* **3**, 541-548 (2007).

5 Escaich, S. Antivirulence as a new antibacterial approach for chemotherapy. *Current Opinion in Chemical Biology* **12**, 400-408 (2008).

6 Murkherjee, S., Hao, Y.-H. & Orth, K. A newly discovered post-translational modification - the acetylation of serine and threonine residues. *Trends in Biochemical Sciences* **32**, 209-216 (2007).

7 Trosky, J. E. *et al.* VopA inhibits ATP binding by acetylating the catalytic loop of MAPK kinases. *The Journal of Biological Chemistry* **282**, 34299-34305 (2007).

8 Li, H. *et al.* The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **315**, 1000-1003 (2007).

9 Zhu, Y. *et al.* Structural insights into the enzymatic mechanism of the pathogenic MAPK phosphothreonine lyase. *Molecular cell* **28**, 899-913 (2007).

10 Zhang, J. *et al.* A *Psuedomonas syringae* effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell Host & Microbe* **1**, 175-185 (2007).

11 Arbibe, L. *et al.* An injected bacterial effector targets chromatin access for transcription factor NF-kappaB to alter transcription of host genes involved in immune responses. *Nature Immunology* **8**, 47-56 (2007).

12 Mazurkiewicz, P. *et al.* SpvC is a *Salmonella* effector with phosphothreonine lyase activity on host mitogen-activated protein kinases. *Molecular Microbiology* **67**, 1371-1383 (2008).

13 Meyer, H. E., Hoffmann-Posorske, E., Korte, H. & Heilmeyer, J., H. L. M. G. . Sequence analysis of phosphoserine-containing peptides: Modification for picomolar sensitivity. *FEBS Letters* **204**, 61 (1986).

14 McLachlin, D. T. & Chait, B. T. Improved b-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Analytical Chemistry* **75**, 6826-6836 (2003).

15 Mattila, K., Siltainsuu, J., Balaspiri, L., Ora, M. & Loennberg, H. Derivatization of phosphopeptides with mercapto- and amino-functionalized conjugate groups by phosphate elimination and subsequent Michael addition. *Organic & Biomolecular Chemistry* **3**, 3039-3044 (2005).

16 Klemm, C., Schroeder, S., Glueckmann, M., Beyermann, M. & Krause, E. Derivatization of phosphorylated peptides with S- and N-nucleophiles for enhanced ionization efficiency in matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **18**, 2697-2705 (2004).

17 Thaler, F. *et al.* A new approach to phosphoserine and phosphothreonine analysis in peptides and proteins: chemical modification, enrichment via solid-phase reversible binding, and analysis by mass spectrometry. *Analytical and Bioanalytical Chemistry* **376**, 366-373 (2003).

18 Bennett, K. L., Stensballe, A., Podtelejnikov, A. V., Moniatte, M. & Jensen, O. N. Phosphopeptide detection and sequencing by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry. *Journal of Mass Spectrometry* **37**, 179-190 (2002).

19 Byford, M. F. Rapid and selective modification of phosphoserine residues catalysed by Ba2+ ions for their detection during peptide microsequencing. *Biochemical Journal* **280**, 261-265 (1991).

20 Chargaff, E. & Sprinson, D. B. Studies on the mechanism of deamination of serine and threonine in biological systems. *The Journal of Biological Chemistry* **151**, 273-280 (1943).

21 Crout, D. H. *et al.* Stereochemistry of the conversions of L-threonine and D-threonine into 2-oxobutanoate by the L-threonine and D-threonine dehydratases of Serratia marcescens. *European Journal of Biochemistry* **106**, 97-105 (1980).

22 Garrido-Franco, M. *et al.* Structure and function of threonine synthase from yeast. *The Journal of Biological Chemistry* **277**, 12396-12405 (2002).

23 Kotloff, K. L. *et al.* Global burden of *Shigella*: Implications for vaccine development and implementation of control strategies. *Bulletin of the World*

Health Organization **77**, 651-666 (1999).

24 Chen, L. *et al.* Structural basis for the catalytic mechanism of phosphothreonine lyase. *Nature Structural & Molecular Biology* **15**, 101-102 (2008).

25 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **79**, 926-935 (1983).

26 Brooks III, C. L. & Karplus, M. Solvent effects on protein motion and protein effects on solvent motion. *Journal of Molecular Biology* **208**, 159-181 (1989).

27 MacKerell Jr., A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).

28 Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187-217 (1983).

29 Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).

30 Gaussian 03, Revision A.1 (Gaussian, Inc., Pittsburgh, PA, 2003).

31 Gonzalez, C. & Schlegel, H. B. An improved algorithm for reaction path following. *Journal of Chemical Physics* **90**, 2154 (1989).

32 Tomasi, J. & Persico, M. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chemical Reviews* **94**, 2027-2094 (1994).

33 Bigeleisen, J. & Mayer, M. G. Calculation of equilibrium constants for isototpic exchange reactions. *Journal of Chemical Physics* **15**, 261 (1947).

34 Anisimov, V. & Paneth, P. ISOEFF98. A program for studies of isotope effects using Hessian modifications. *Journal of Mathematical Chemistry* **26**, 75 (1999).

35 Amber 10 (University of California, San Francisco, 2008).

36 Cornell, W. D. *et al.* A second generation force field for the simulation of protein, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197 (1995).

37 Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).

38 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an NlogN method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089-10092 (1993).

39 Essmann, U. *et al.* A smooth particle mesh Ewald method. *Journal of Chemical Physics* **103**, 8577-8593 (1995).

40 Zhang, Y. Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods. *The Journal of Chemical Physics* **122**, 024114 (2005).

41 Zhang, Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts* **116** (2006).

42 QCHEM www.q-chem.com (2006).

43 TINKER, Software Tools for Molecular Design (dasher.wustl.edu/ffe/) (2004).

44 Zhang, Y., Liu, H. & Yang, W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *The Journal of Chemical Physics* **112**, 3483 (2000).

45 Torrie, G. M. & Valleau, J. P. Non-physical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).

46 Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free energy calculations on biomolecules. 1. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).

Chapter 6

Insights into the Phosphoryl Transfer Mechanism of Cyclin-dependent Protein Kinases

Ab Initio QM/MM Free-energy Studies

"Any living cell carries with it the experience of a billion years of experimentation by its ancestors."

— Max Ludwig Henning Delbrück

This chapter is based on collaborative work originally found in the following publication:

Adapted with permission from **Smith GK**, Ke Z, Hengge AC, Guo H. Insights into the Phosphoryl Transfer Mechanism of Cyclin-Dependent Protein Kinases from ab Initio QM/MM Free-Energy Studies. The Journal of Physical Chemistry B. 2011; 115(46): 13713–13722.

6.1 Introduction

Cellular signaling is largely controlled by post-translational modification of protein residues, such as phosphorylation catalyzed by protein kinases and dephosphorylation by protein phosphatases.[1] Mutations of these enzymes have been implicated in many disease states, including cancer.[2] As a result, there is strong interest in finding effective inhibitors for these key enzymes.[3-5] Significant effort has been devoted to structural and mechanistic studies.[6,7]

There are more than five hundred protein kinases encoded in the human genome, representing one of the largest protein families.[8] Despite sequence and structural diversities, many share nearly conserved catalytic domains.[9] We are concerned here with an important class of protein kinases, namely cyclin-dependent kinases (CDKs), which regulate the cell cycle in eukaryotes by a complex signal transduction network.[10,11] Of the 11 known human CDKs, CDK1 (also known as CDC2) is the most essential to the proper functioning of the cell cycle and is also the evolutionary analog to the primitive CDC28 first studied in yeast.[11] Mice knockout studies have shown that a loss of a single CDK can lead to lethality, such as with CDK5, or a host of developmental defects, such as CDK2, CDK4, CDK6, and CDK11.[12] The activity and specificity of these CDKs depend on the binding of cyclin subunits, including cyclin classes A, B, D, and E that are differentially expressed over the course of the cell cycle.[13] The precise regulation of the cell cycle is nominally controlled by the rise and fall of different cyclin levels and the activity of their CDK complexes, but many additional positive and negative feedbacks

exist as well, including natural inhibitor families INK4 and Cip/Kip, as well as genetic and metabolic checkpoints.[14]

In this work, we focus on the catalytic mechanism of CDK2, which is perhaps the most extensively studied interphase CDK. With over 200 published X-ray structures in the Protein Data Bank (PDB) since the pioneering work of De Bondt et al.,[15] it has served as a basis for much CDK research. CDK2 is generally classified as a serine/threonine protein kinase (SPK), and often more specifically as a proline-directed kinase with regards to substrate recognition, with the phosphorylation site (P) adjacent to a proline residue (P+1) in the direction of the C-terminus. Many substrates have been identified for CDK2, including pRb, p107, p53, and others that satisfy the consensus sequence from the P to P+3 position of **(S/T)**-P-X-(R/K), where X is any residue.[14] For full functioning, CDK2 must bind with its cyclin partners A or E, a molecule of ATP, and at least one Mg(II) ion. In addition, it must have a conserved threonine residue (Thr160) on the activation loop phosphorylated by a separate enzyme called CDK Activating Kinase (CAK).[16] The overall structure of a CDK2-Cyclin-ATP-Mg-substrate complex[17] is shown in Figure 6.1.

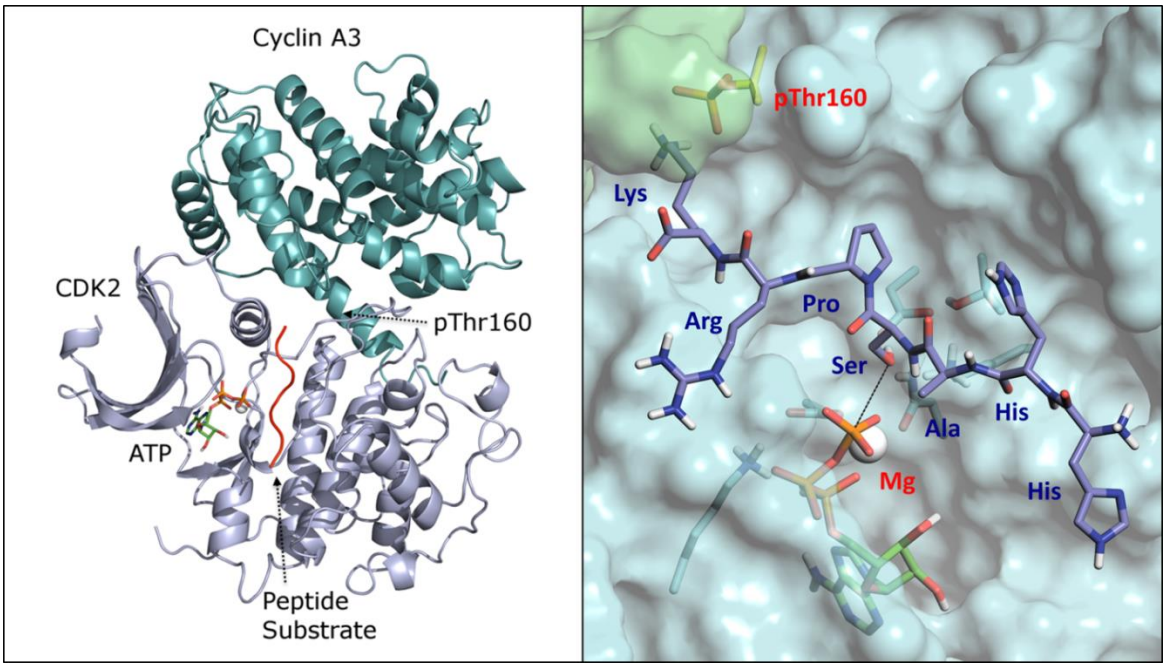


Figure 6.1 – Overall protein fold of the activated CDK complex with cyclin A3, Mg-ATP, and the peptide substrate[17] on the left panel. A close up picture of substrate binding is shown in the right panel.

The chemical step catalyzed by CDK2 is the transfer of a phosphoryl group from the γ position of ATP to the hydroxyl group of a Ser or Thr residue of the substrate. For such phosphoryl transfer reactions, three limiting-case mechanisms are possible.[18,19] The dissociative ($D_N + A_N$) mechanism envisions an initial dissociation of a free metaphosphate followed by addition. The associative ($A_N + D_N$) mechanism, on the other hand, features a pentacovalent phosphorane intermediate, flanked by two transition states. Finally, the concerted (A_ND_N) mechanism is characterized by a single transition state as a middle point for the cleavage and formation of P-O bonds. Within the concerted mechanism, the transition state could be either associative, featuring a pentacovalent phosphorane-like species, or dissociative, featuring a metaphosphate-like

species. For uncatalyzed phosphoryl transfer reactions involving phosphate monoester dianions, such as ATP, it is widely believed that the preferred pathway involves a concerted mechanism with a dissociative transition state.[20] (For an interesting and alternative perspective, see Ref. [21].) For the corresponding enzymatic reactions, on the other hand, the mechanism is not well established. There is thus a strong desire to uncover whether enzymatic reactions use the same mechanism as uncatalyzed reactions and how they are catalyzed.

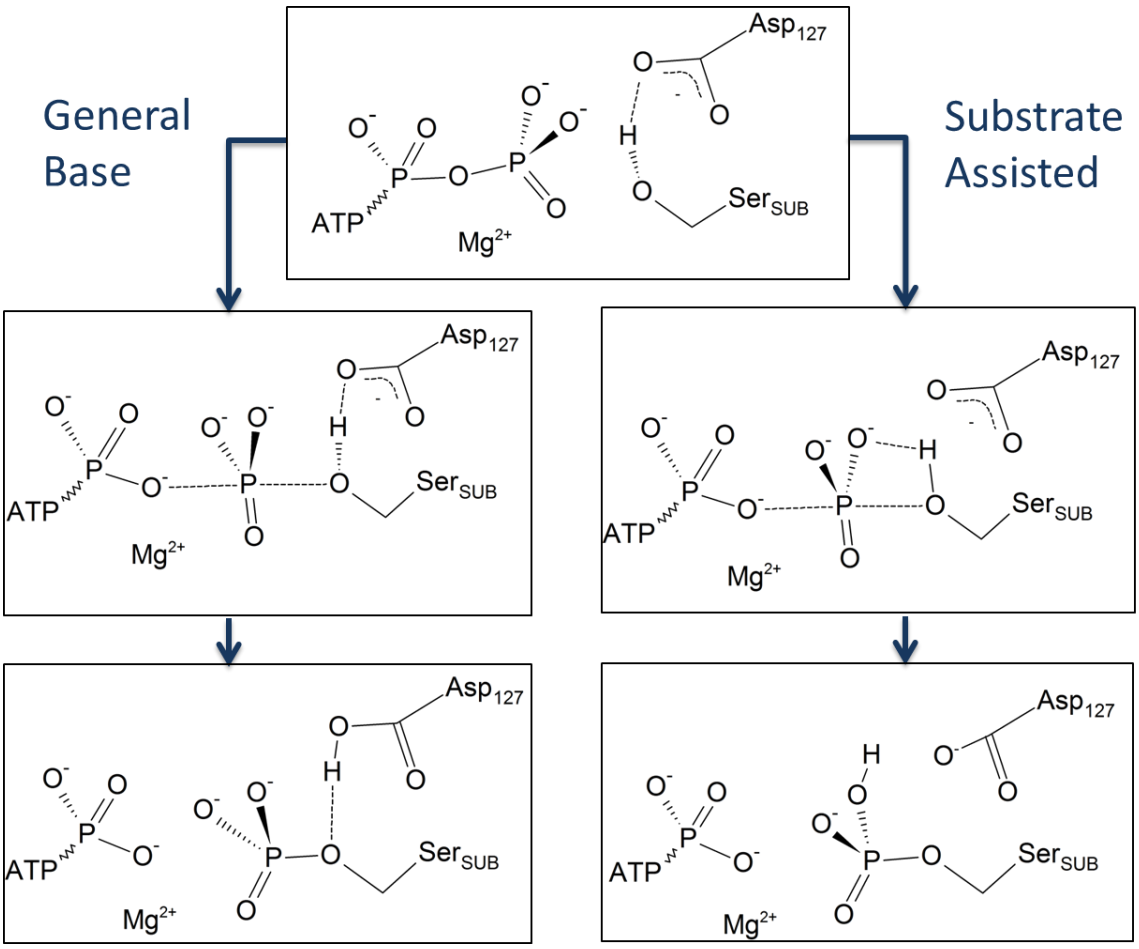
The active site of the activated CDK2-substrate complex[17] shown in Figure 6.1 consists of a shallow grove for substrate binding, an ATP hydrophobic binding pocket that aligns the γ-phosphate with the help of a single magnesium ion, and a triad of residues: Asp127, Lys129, and Thr165 which bracket the substrate serine oxygen on three sides, forming a hydrogen bond network and helping align the serine with the phosphate group to be transferred. The Mg(II) ion binds with conserved Asn132 and Asp145 residues, which help align and polarize the γ-phosphate. Very recently, a new X-ray structure of a CDK2/cyclin complex with a transition-state mimic (MgF_3^-) has been reported.[22] Interestingly, two Mg(II) ions have been found in the active site, although the second Mg(II) is thought to bind in a transient fashion during turnover.

Despite its importance, kinetic studies of CDK2 catalysis are surprisingly sparse, especially when compared with the other kinase prototype PKA (protein kinase A or cAMP-dependent protein kinase). The first kinetic study by Hogopian *et al.* indicated that the phosphorylation step is fast (k_3 = 22 s^{-1}), but overall rate is limited by product

release (k_4 = 11 s^{-1}).[23] The former rate constant was later revised to 35 s^{-1}.[24] Another kinetic investigation found the reaction has a random anticooperative mechanism, but did not report the rate constant for the chemical step.[25] Subsequently, there have also been kinetic studies focusing on the effect of cyclin partners E1 and A2,[26] a comparative study of CDK1/CDK2 partnered with cyclin A1 and A2,[27] and more recently on the effect of a mutation in the PSTAIRE helix on cyclin binding.[28] In addition, mechanistic studies have been reported for CDK5,[29,30] which shares ~60% sequence similarity with CDK2 and is likely to have the same catalytic mechanism. To the best of our knowledge, no systematic mutagenesis studies of active-site residues of any CDKs have so far been reported.

While these kinetic studies reported K_M and k_{cat} values, the molecular details of the catalysis are still unclear. For example, is the mechanism dissociative, associative, or concerted? What is the character of the transition state? What is the identity of the general base needed to activate the serine nucleophile? What is the role played by the Mg(II) cofactor and several active-site residues such as Lys33 and Lys129? To answer these questions, several theoretical models have been developed. An earlier density functional theory (DFT) study with a truncated active-site model,[31] and a subsequent QM/MM model[32] both suggested a ATP-assisted mechanism, in which the nucleophilic serine is deprotonated by the γ-phosphate of ATP, while the nearby Asp127 residue helps to align the substrate oxygen but not participating in proton transfer. In addition, the transition state was found to have a distinct associative character. However, this proposal is at odds with the mechanism proposed for other protein kinases, most

notably the extensively characterized PKA,[33-37] where the general base has been identified as a conserved active-site Asp. In CDK2, this conserved Asp residue is Asp127, and its potential role as the general base has also been proposed before.[17] In addition, the associative transition state obtained by these theoretical models is inconsistent with recent X-ray structures of CDK2/cyclin bound with transition-state analogs (nitrate and MgF_3^-), which suggest a dissociative transition state.[22,38] The Asp127 residue was not included in the QM region in the QM/MM simulations of De Vivo *et al.*,[32] thus precluding its capacity to serve as the general base, though the authors did not fully rule out this possibility.



Scheme 6.1 – Two proposed mechanisms for the phosphoryl transfer reaction catalyzed by CDK2.

In this Chapter the two mechanistic proposals shown in Scheme 6.1 are reexamined using an *ab initio* QM/MM method, and the free-energy is calculated for the proposed reaction scheme in order to take into account the fluctuation in the context of the protein environment. The results clearly support a concerted mechanism in which Asp127 serves as the general base. Furthermore, a single dissociative, metaphosphate-like, transition state was identified. The roles played by various frontline residues and

the Mg(II) cofactor are also discussed. This work is organized as follows: Section 6.2 discusses methods and details of models for determining the reaction pathways. Section 6.3 presents the free-energy profile and geometric/charge data. Section 6.4 discusses the comparison of these results with previous theoretical and experimental data and implications in the general mechanism of protein kinases. The final section (6.5) concludes the discussion.

6.2 Methods

The methods used in this project are described in a condensed fashion below. Foundational information and additional details on the techniques used can be found in Chapter 2 (Classical Methods) and Chapter 3 (Quantum Methods).

6.2.1 Model

Due to the large size of the enzymatic system, it is impractical to use conventional quantum chemistry methods to study the catalyzed reaction. Here, we treat the system within the QM/MM framework,[39,40] which divides the system into two parts. The QM region is treated here with DFT, while the MM region by force fields. Such an approach has been widely and successfully used in studying enzymatic reactions.[41-44]

The starting geometry for CDK2 was largely based on the crystal structure 1QMZ,[17] which was also used in the earlier QM/MM study of De Vivo *et al.*[32] This composite structure contains CDK2 with phosphorylated Thr160, bound with cyclin A3, ATP, a single Mg(II) cofactor, and a peptide substrate. The peptide substrate has the resolved

sequence HHA**S**PRK containing the **(S/T)**-P-X-(R/K) motif required for substrate recognition. Residues Arg297 and Leu298 on the N-terminus of CDK2 and residue Glu174 on the C-terminus of Cyclin A3 were restored using PyMOL (www.pymol.org), and allowed to relax with a short minimization. After careful consideration of hydrogen bond networks, histidine residues were assigned as HID120 and HID122 in CDK2, HID361 and HID424 in Cyclin A3, and all other histidines were assigned as HIE.

6.2.2 MD

The AMBER 10 simulation package[45] was used for initial setup, addition of hydrogens, solvation, neutralization, initial minimization of the structure, and molecular dynamics (MD). Force field parameters included the AMBER99SB force field for the proteins,[46,47] ATP parameters developed by Meagher *et al.*,[48] and pThr parameters developed by Homeyer *et al.*[49] The system was then solvated with a rectangular box of TIP3P waters[50] with dimensions of 88 x 95 x 106 $Å^3$, allowing for a 10 Å buffer on all sides between protein atoms and the periodic boundary. Two Na^+ ions were added to neutralize the box and gave a total system size of 74,938 atoms. Periodic boundary conditions were imposed with the long-range electrostatic interactions treated with the particle-mesh Ewald (PME) method.[51,52] Van der Waals interactions were calculated with an 8 Å cutoff. In MD simulations, hydrogens were restrained by the SHAKE algorithm,[53] and the time step was 1 fs.

After hydrogen atoms were allowed to relax with all heavy atoms fixed, a constrained minimization procedure was employed using harmonic restraining potentials with force

constants ranging from 50 kcal·mol^{-1}·Å^{-2} to 5 kcal·mol^{-1}·Å^{-2} on the protein and substrate. The structure was slowly relaxed by alternating MD and minimization as the constraint was scaled down, ending with a 2000 step unconstrained minimization. To equilibrate the refined structure, a 10 kcal·mol^{-1}·Å^{-2} constraint remained on the active-site region (Lys33, Asp127, Lys129, Asn132, Asp145, Thr165, ATP, Mg, and substrate) as the system was heated to 300 K over 100 ps. Dynamics was then continued for 100 ps at this constraint, and a further 200 ps where the active site constraint was reduced to 5 kcal·mol^{-1}·Å^{-2}. The final snapshot became the basis for the subsequent QM/MM studies.

6.2.3 QM/MM

In our investigation of the phosphorylation step, we used the pseudo-bond *ab initio* QM/MM approach,[54-57] which has been used to successfully investigate several enzymes,[58-66] including a protein kinase (PKA).[35] The QM/MM model was based on the last snapshot of the classical MD simulation by including all atoms within a 27 Å radius centered on the P_γ atom of ATP, resulting in 13,533 atoms. Atoms greater than 20 Å away from P_γ of ATP were held fixed in order to reduce computational costs.

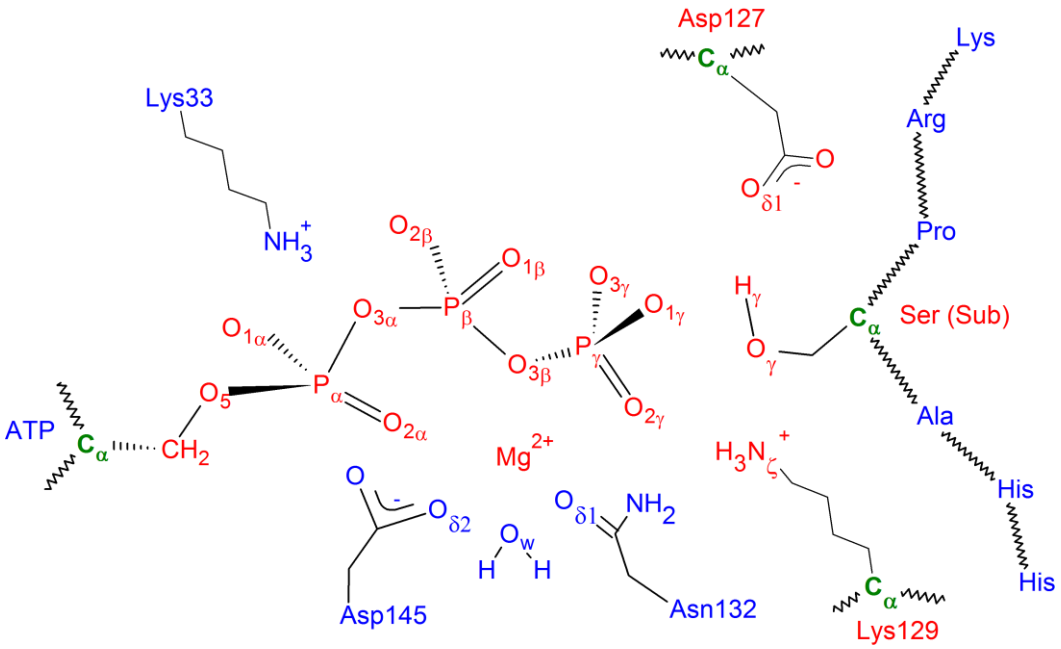


Figure 6.2 – Model showing several key active-site residues and QM subsystem (in red) with boundary atoms in green.

As shown in Figure 6.2, the QM subsystem, which consists of the Mg(II) ion, the triphosphate portion of ATP, and the side chains of Asp127, Lys129, and the substrate serine, has 48 atoms; they are treated at the B3LYP/6-31G(d) level of theory. An *ab initio* treatment of the QM system is important because of the involvement of the d orbitals in P.[67,68] On the other hand, the MM subsystem was modeled with the same force fields described above for the initial minimization/MD. All calculations were done using a modified version of the QChem[69] and TINKER[70] program. Cutoff radii were 12 Å for van der Waals interactions and 18 Å for electrostatic interactions between MM atoms, and no cutoff was imposed for electrostatic interactions between QM and MM atoms.

It should perhaps be pointed out that the model developed here is by no means perfect, due to limitations of computational resources. One area of possible errors is the treatment of the Mg(II) coordination sphere. While Mg(II) and some ligands are in the QM region, others are in the MM region, which might over-polarize the substrates. Fortunately, we expect the errors to be reasonably small as the ligand-metal interactions are dominated by electrostatic interactions.

The QM/MM model was then minimized and reaction paths were explored using the reaction coordinate driving (RCD) method.[55] Several reaction coordinates have been used to explore the reaction mechanism, which are discussed further in Section 6.3. Along each reaction path, the ESP charges of all quantum atoms were calculated. To calculate the potential of mean force (PMF), we have used the following reaction coordinate, $\mathrm{RC_I} = \mathrm{d}\left(\mathrm{O_{3\beta}} - \mathrm{P_\gamma}\right) - \mathrm{d}\left(\mathrm{O_\gamma} - \mathrm{P_\gamma}\right) - \mathrm{d}(\mathrm{O_{\delta 1}} - \mathrm{H_\gamma})$. (The definition of atomic labels can be found in Figure 6.2.) To sample high energy regions, the umbrella sampling method[71] was used. Specifically, the reaction path was partitioned into 22 windows, each with a harmonic bias potential with a force constant between 40 and 120 kcal·mol^{-1}·Å^{-2} adjusted iteratively depending on the resulting sampling histogram. The MD trajectories were propagated using the Beeman algorithm[72] at a time step of 1 fs, with the Berendsen thermostat[73] to maintain the system at 300 K. During the MD simulations, the MM subsystem was first allowed to move for 500 ps with the QM subsystem fixed. This is followed by MD for the entire system for 30 ps using the QM/MM potential. The resulting probability distributions were then converted to the PMF by use of the weighted histogram analysis method (WHAM).[74]

In addition to the PMF calculations, we have also carried out perturbation calculations to evaluate the influence of surrounding residues on the reactant complex and transition state.[75] To this end, electrostatic (ES) and van der Waals (vdW) contributions of individual MM residues on the QM subsystem were calculated at the two critical points. The individual residue contribution was calculated as: $\Delta E_i = E(vdW+ES)_i^{TS} - E(vdW+ES)_i^{RC}$, in which the superscripts TS and RC denote the reactant complex and transition state, respectively. A negative value indicates that the interaction of that particular residue is favorable for catalysis, while a positive value is unfavorable.

6.3 Results

As shown in Table 6.1, the serine nucleophile and the transferring phosphoryl group are well aligned in the reactant complex (RC), with an $O_\gamma - P_\gamma$ distance of 3.39 Å. This can be compared with the X-ray structure, in which the corresponding distance is 3.68 Å.[17] The $O_\gamma - O_{3\beta}$ distance of 5.09 Å is larger than 4.9 Å, which was suggested by Mildvan to be the onset for the dissociative concerted mechanism.[76] This in-line attack configuration is held in place by a number of interactions. For example, the serine nucleophile has its hydroxyl group donating a hydrogen bond to the carboxylate of Asp127 ($d(H_\gamma - O_{\delta 1})$=1.63 Å, $d(O_\gamma - O_{\delta 1})$=2.63 Å) and accepts a hydrogen bond from Lys129 ($d(H_{\zeta 3} - O_\gamma)$=1.72 Å, $d(N_\zeta - O_\gamma)$=2.64 Å). These distances between hydrogen bonded heavy atoms are also consistent with those in the X-ray structure (2.69 Å and 2.55 Å).[17] In addition, both Lys129 and Asp127 are hydrogen-bonded with Thr165, forming a relatively strong hydrogen bond network.

On the other hand, the ATP is stabilized by a number of electrostatic interactions with surrounding moieties. For instance, the Mg(II) ion coordinates with $O_{2\gamma}$, $O_{3\beta}$, $O_{2\alpha}$ of ATP, in addition to the side chains of Asp145, Asn132, and a water, in an octahedral coordination sphere. Furthermore, ATP is hydrogen bonded to Lys33, as evidenced by the $H_{\zeta 3} - O_{1\alpha}$ distance of 1.88 Å. In addition, a number of solvent water molecules are found to hydrogen bond with non-bridging oxygens of the triphosphate group. These interactions help in aligning the reactant for the in-line near attack configuration. Finally, the phosphorylated Thr160 of CDK2 forms a strong interaction with a crown of CDK2 arginine residues Arg50, Arg126, Arg150, and the lysine residue of the peptide substrate. Two of these arginine residues, Arg50 and Arg150, form hydrogen bonds with the backbone of Cyclin A3. The distances in our model are largely consistent with those reported in the X-ray structure,[17] as shown in Table 6.1.

Table 6.1 – Key interatomic distances for the phosphorylation reaction at three stationary points along the reaction path in Model I. If no residue is specified, the atom label refers to ATP/ADP.

Bond	Inter-atomic distance (Å)			
	Expt.[19]	RC	TS	PC
$P_\gamma - O_\gamma$ (Ser)	3.68	3.39	2.46	1.74
$P_\gamma - O_{3\beta}$	1.60	1.81	2.54	3.24
O_γ (Ser) $- H_\gamma$(Ser)	-	1.01	1.04	1.83
H_γ(Ser) $- O_{\delta1}$(Asp127)	-	1.63	1.58	1.00
H_γ(Ser) $- O_{2\gamma}$	-	3.11	2.73	2.80
$Mg - O_{2\gamma}$	2.25	2.00	2.08	2.08
$Mg - O_{3\beta}$	2.12	2.20	2.05	2.01
$Mg - O_{2\alpha}$	2.07	2.04	2.04	2.12
$Mg - O_{WAT}$	2.14	2.10	2.13	2.19
$Mg - O_{\delta1}$(Asn132)	2.00	1.94	1.99	2.02
$Mg - O_{\delta2}$(Asp145)	1.91	1.87	1.87	1.88
N_ζ(Lys33) $- O_{1\alpha}$	3.00	2.90	2.91	2.90
N_ζ(Lys33) $- O_{3\alpha}$	3.19	2.87	2.84	2.84
N_ζ(Lys33) $- O_{\delta1}$(Asp145)	3.85	2.81	2.82	2.82
N_ζ(Lys129) $- O_{2\gamma}$	3.58	4.17	4.00	3.94
N_ζ(Lys129) $- O_\gamma$(Ser)	2.55	2.64	2.72	2.80
N_ζ(Lys129) $- O_{\gamma1}$(Thr165)	2.69	2.85	2.86	2.89
$O_{\delta2}$(Asp127) $- O_{\gamma1}$(Thr165)	2.90	2.74	2.70	2.81
$N_{\delta2}$(Asn132) $- O_{2\gamma}$	3.66	3.40	3.17	2.91

The only sizable difference between our model of the Michaelis complex and the X-ray structure is the distance between N_ζ(Lys33) and $O_{\delta1}$(Asp145). The model distance is 2.81 Å, significantly shorter than that in the X-ray structure (3.85 Å).[17] The origin of the discrepancy is not clear, but we hypothesize that it might have something to do with the possibility of a disordered second Mg(II) ion, which would neutralize some of the positive charge in the sidechain of Asp145, thus weakening the hydrogen bond with Lys33. Indeed, the Mg(II) concentration used in soaking the crystal was quite high, and a

recent kinetic study of CDK5 has clearly demonstrated the necessity of a second Mg(II) in activating the catalysis.[30] This second Mg(II) ion is now confirmed by a recent X-ray structure of CDK2/Cyclin with a transition-state analog (MgF_3^-), but it was suggested that the binding of the second Mg(II) is transient.[22]

To explore the reaction mechanism, several reaction coordinates have been tested. The general base pathway was tested using the combined coordinate, $\mathrm{RC_I} = \mathrm{d}(\mathrm{O_{3\beta}} - \mathrm{P_\gamma}) - \mathrm{d}(\mathrm{O_\gamma} - \mathrm{P_\gamma}) - \mathrm{d}(\mathrm{O_{\delta 1}} - \mathrm{H_\gamma})$. Models II and III used the P-O bond distances to probe the associative and dissociative mechanisms: $\mathrm{RC_{II}} = \mathrm{d}(\mathrm{O_{3\beta}} - \mathrm{P_\gamma})$ and $\mathrm{RC_{III}} = \mathrm{d}(\mathrm{O_\gamma} - \mathrm{P_\gamma})$. Models IV and V test the substrate-assisted mechanism by using the following reaction coordinates: $\mathrm{RC_{IV}} = \mathrm{d}(\mathrm{O_{2\gamma}} - \mathrm{H_\gamma})$ and $\mathrm{RC_V} = \mathrm{d}(\mathrm{O_{3\gamma}} - \mathrm{H_\gamma})$, which pushes the Ser hydrogen ($\mathrm{H_\gamma}$) to either the Mg-coordinated phosphate $\mathrm{O_{2\gamma}}$ or the non-Mg-coordinated $\mathrm{O_{3\gamma}}$. The minimal energy paths calculated using these reaction coordinates and key distances are displayed in Fig. 6.3.

It is clear from Fig. 6.3 that energy increases monotonically if the proton is forced to transfer to a phosphate oxygen in Models IV and V. The non-productive reaction paths thus discount the substrate-mediated mechanism. The associative and dissociative mechanisms can also be ruled out as the $\mathrm{P} - \mathrm{O}$ bond lengths change in tandem in Models I, II, and III, strongly suggesting a concerted mechanism. Neither a free metaphosphate nor a pentacovalent phosphorane intermediate was found. Instead, only one transition state exists. As the system approaches the transition state in Model I, the $\mathrm{P_\gamma} - \mathrm{O_\gamma}$ distance is reduced from 3.39 to 2.46 Å, while the $\mathrm{P_\gamma} - \mathrm{O_{3\beta}}$ distance

increases from 1.81 to 2.54 Å. The $O_{\gamma} - O_{3\beta}$ distance of 4.9 Å at the transition state is consistent with that (5.5 – 5.7 Å) found in the X-ray structure of CDK2 complex with a transition-state analogue (nitrate).[38] Furthermore, the two $P - O$ distances are also in good agreement with the more recent transition-state structure, with the corresponding distances of 2.8 and 2.5 Å.[22] Interestingly, the proton transfer lags behind the phosphoryl transfer.

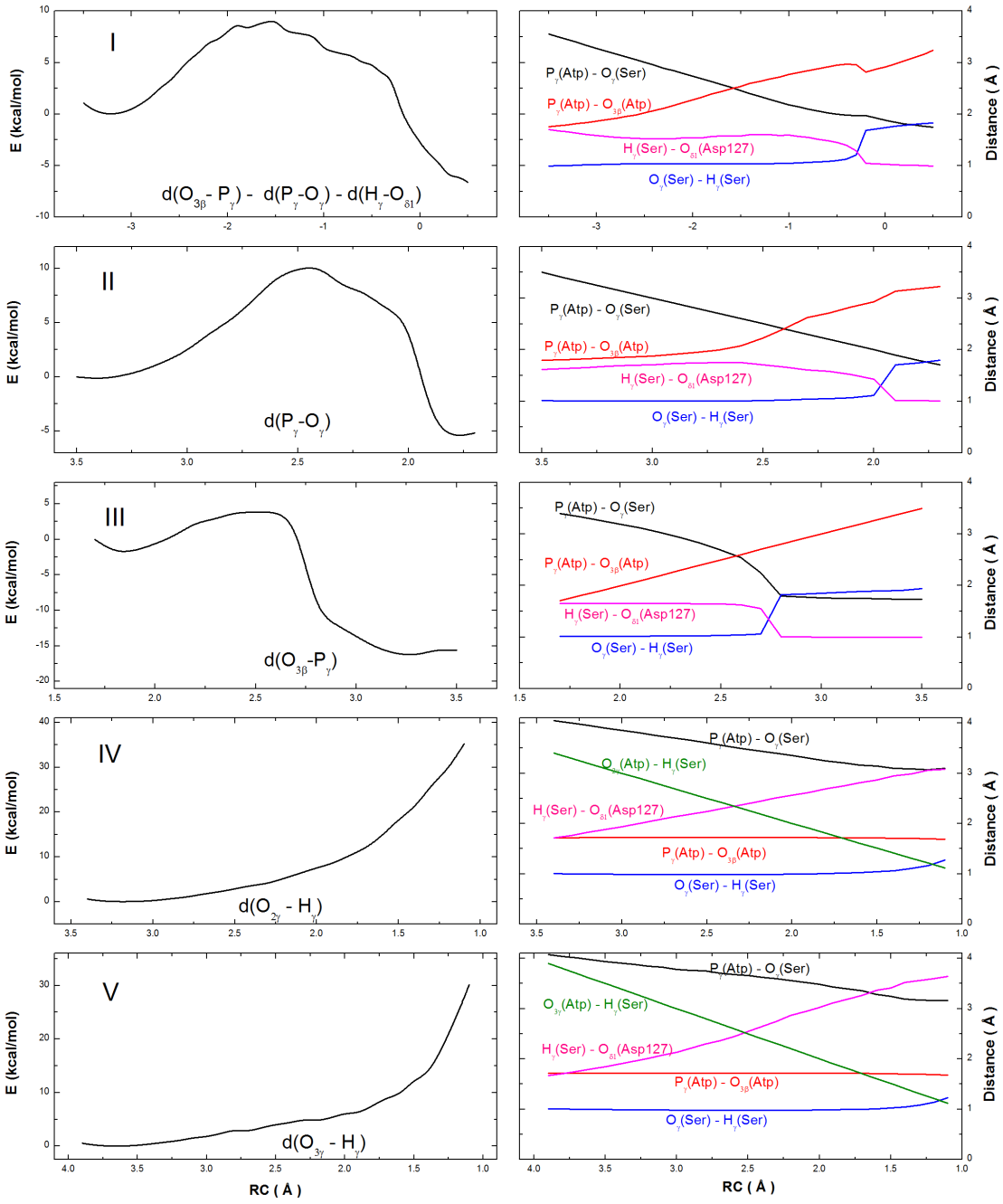


Figure 6.3 – Reaction paths for five models (I-V) with different reaction coordinates are displayed in the left panels. Variations of several key interatomic distances with the reaction coordinate are given in the right panels.

To further demonstrate the concerted nature of the reaction path, we have mapped out using RC_{II} and RC_{III} a two-dimensional minimal energy surface, as shown in Fig. 6.4. The reaction path proceeds essentially as a diagonal cut of the map from one minimum to another. The dissociative character of the transition state can also be estimated using Pauling's formula: D(*n*) = D(1) – 0.6log*n*, as suggested by Mildvan.[76] Here, D(1) = 1.73 Å for the $\mathrm{P-O}$ bond, and D(*n*) is defined as the average of the two $\mathrm{P-O}$ distances (2.5 Å) at the transition state. The fractional bond number (*n*) is thus 0.05, which gives a dissociative character of 95%. It is clear from the figure that the dissociative character of the concerted transition state is predetermined by the large $\mathrm{O-O}$ distance in the Michaelis complex.

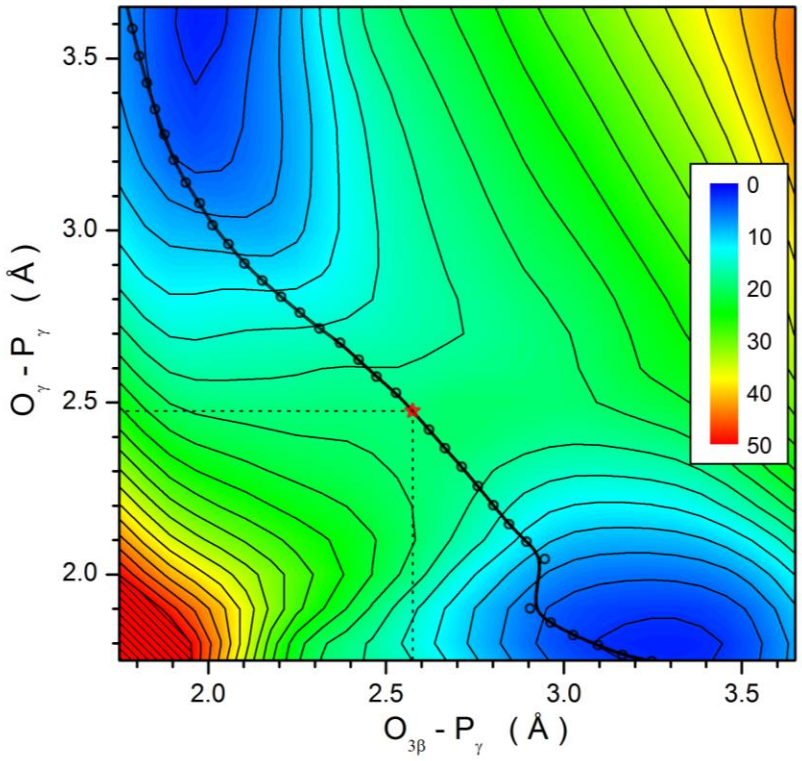


Figure 6.4 – Two-dimensional minimal energy surface in RC_{II} and RC_{III} defined in text. The minimal energy path and the transition state are shown.

Based on the aforementioned reaction path calculations and our test calculations of various combined coordinates, RC_I was selected as our choice for analysis and construction of the PMF. The sole transition state (TS) in this model features a trigonal planar metaphosphate-like phosphoryl group between the leaving group ($O_{3\beta}$) and nucleophile (O_γ) with large $P-O$ distances. Interestingly, the proton transfer from the nucleophilic Ser OH group to the Asp127 is delayed: the $O_\gamma - H_\gamma$ distances increases from 1.01 Å at RC to 1.04 Å at TS, and finally to 1.83 Å at the product complex (PC). On the other hand, the $H_\gamma - O_{\delta 1}$ distances varies from 1.63 Å at RC, to 1.58 Å at TS, and finally to 1.00 Å at PC. The late proton transfer at the TS, which was also observed in earlier QM/MM studies of PKA,[33,35,37] supports the dissociative character of the transition state, as the $P_\gamma - O_\gamma$ bond formation, which is facilitated by proton transfer to Asp127, has not progressed significantly at the transition state.

Mulliken charges on several key atoms are listed in Table 6.2. The most important change occurs for $O_{3\beta}$, which is converted from a bridging oxygen to a non-bridging one in the course of the reaction. As a result, its charge changes from -0.67 at RC to -0.92 at TS, and finally to -0.99 at PC. The increased charge leads to a stronger coordination with the Mg(II) ion, as evidenced by the corresponding distance changing from 2.20 Å at RC, to 2.05 Å at TS, and finally to 2.01 Å at PC. In the meantime, the bonds between Mg(II) and $O_{2\gamma}$ weakens somewhat, indicated by slight elongation of the corresponding bond distance in Table 6.1. These changes are presumably in response to the strengthened $Mg - O_{3\beta}$ bond. On the other hand, the bonding of Mg(II) with Asp145 and Asn132 changes little during the course of the reaction.

Table 6.2 – Key atomic charges at three stationary points along the reaction path of Model I.

Atom	RC	TS	PC
P_{γ}	1.18	1.17	1.30
$O_{1\gamma}$	-0.80	-0.73	-0.86
$O_{2\gamma}$	-0.98	-0.88	-0.98
$O_{3\gamma}$	-0.83	-0.74	-0.92
P_{β}	1.17	1.26	1.31
$O_{1\beta}$	-0.86	-0.92	-0.95
$O_{2\beta}$	-0.82	-0.88	-0.90
$O_{3\beta}$	-0.67	-0.92	-0.99
P_{α}	1.22	1.24	1.23
$O_{1\alpha}$	-0.82	-0.84	-0.86
$O_{2\alpha}$	-0.76	-0.78	-0.78
$O_{3\alpha}$	-0.52	-0.56	-0.57
Mg	1.68	1.69	1.69
$O_{\gamma}(\mathrm{Ser})$	0.03	-0.13	-0.14
$H_{\gamma}(\mathrm{Ser})$	0.36	0.44	0.37
$O_{\delta 1}(\mathrm{Asp127})$	-0.85	-0.86	-0.63
$N_{\zeta}(\mathrm{Lys129})$	-0.59	-0.79	-0.86

As shown in Figure 6.5, the three non-bridging oxygens of the transferring phosphoryl group are stabilized during the transition state. Specifically, $O_{1\gamma}$ forms hydrogen bonds with 4 solvent waters, $O_{2\gamma}$ bonds primarily with Mg(II), and $O_{3\gamma}$ forms hydrogen bonds with 3 solvent waters. As a result, the total charge of the phosphoryl group progresses from -1.42 (RC), to -1.18 (TS), and then back to -1.47 (PC). The Mg(II) ion clearly helps stabilize the transition state, as shown by the charge on its associated phosphate oxygen, $O_{2\gamma}$, which changes from -0.98 (RC), to -0.88 (TS), to -0.98 (PC). This charge stabilization is nearly equivalent to that of $O_{1\gamma}$, which is complexed by four waters. It is

interesting to note that there is no proton transfer between the ammonium group of Lys129 and other moieties during the reaction, despite the fact that the side chain is included in the QM region. It appears that the role of Lys129 is a structural one, mainly responsible for positioning the reactants in the near-attack configuration. Similar observations have been made for PKA.[35]

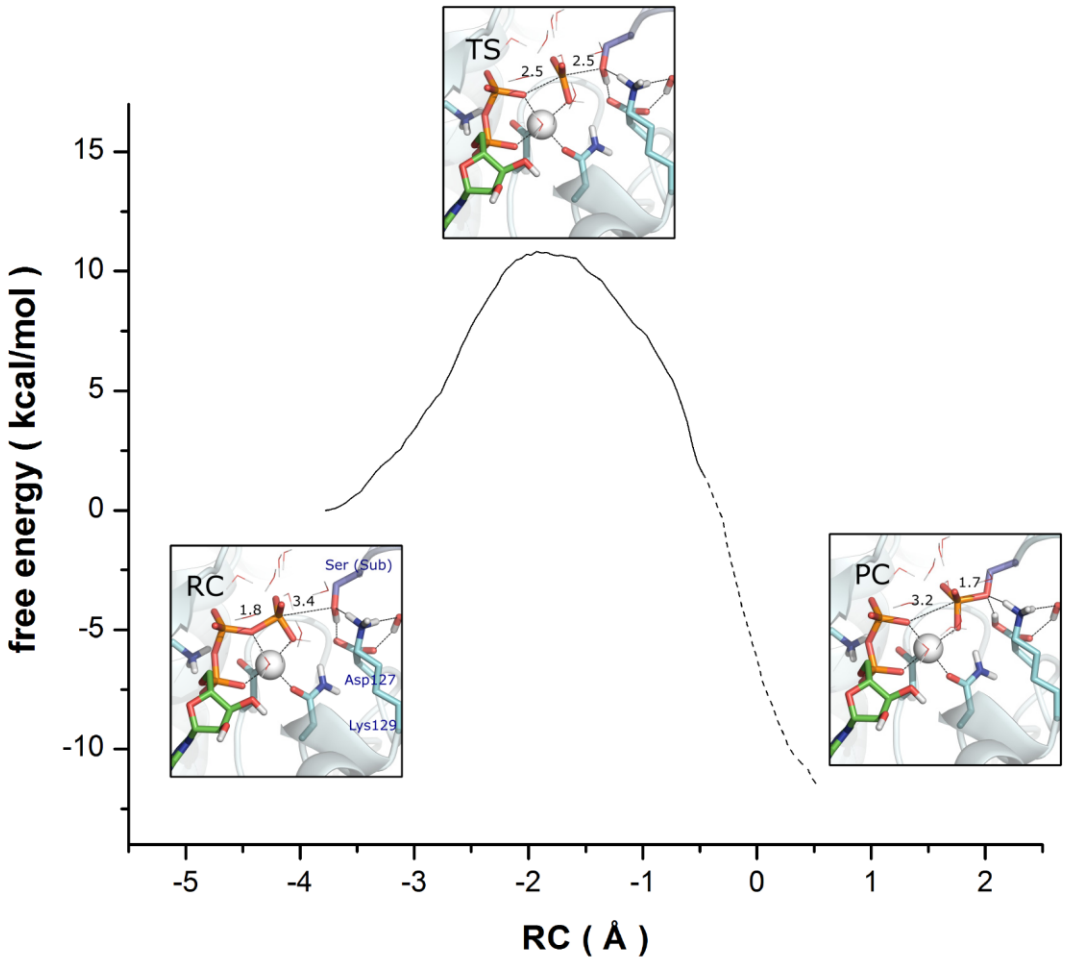


Figure 6.5 – Calculated free-energy profile for the phosphoryl transfer reaction catalyzed by CDK2 using RC_I as the reaction coordinate. The dashed line indicates the region where the PMF convergence is less satisfactory. The configurations of the RC, TS, and PC complexes obtained in the reaction coordinate calculations are also shown with the hydrogen bonds and Mg coordination bonds indicated via dashed lines.

The product complex (PC) features a phosphorylated serine residue in the substrate, with O_γ hydrogen bonded with the protonated Asp127. The phosphoryl group in the phosphorylated serine residue bonds with Mg(II) ion, and forms hydrogen bonds with Asn132 and six waters. On the other hand, the ADP have two non-bridging oxygens ($O_{2\alpha}$ and $O_{3\beta}$) chelating the Mg(II) ion, and form hydrogen bonds with Lys33, and a number of water molecules.

As shown in Figure 6.5, the calculated free-energy profile (PMF) gives a barrier height of 10.8 kcal/mol. To test convergence, the difference in the barrier height between 10 ps and 30 ps sampling was found to be less than 1.0 kcal/mol, although the PC region (indicated in dashed line) is less converged. This barrier height is significantly lower than that found using the associative pathway, 23.7 ± 4.5 kcal/mol, as reported by De Vivo et al. [32]

To further understand the roles played by other non-frontier residues in the catalysis, a perturbation analysis has been performed along the reaction path. As shown in Figure 6.6, Lys33 has the largest effect. Since it helps to stabilize ATP by forming a hydrogen bond with $O_{1\alpha}$, it is not difficult to understand that it preferentially stabilizes TS, where the charges of the ATP oxygen atoms increase relative to RC. It is known that its counterpart in PKA (Lys72) plays the same role, as demonstrated by mutagenesis[77] and computational studies.[35] On the other hand, the Asn132 and Asp145 residues have positive values, which stem presumably from their coordination with the Mg(II) cofactor, as seen in PKA.[35] It should however be note that the magnitudes indicated in

Figure 6 can only be considered qualitative, because significant enzyme reorganization might occur for mutants.

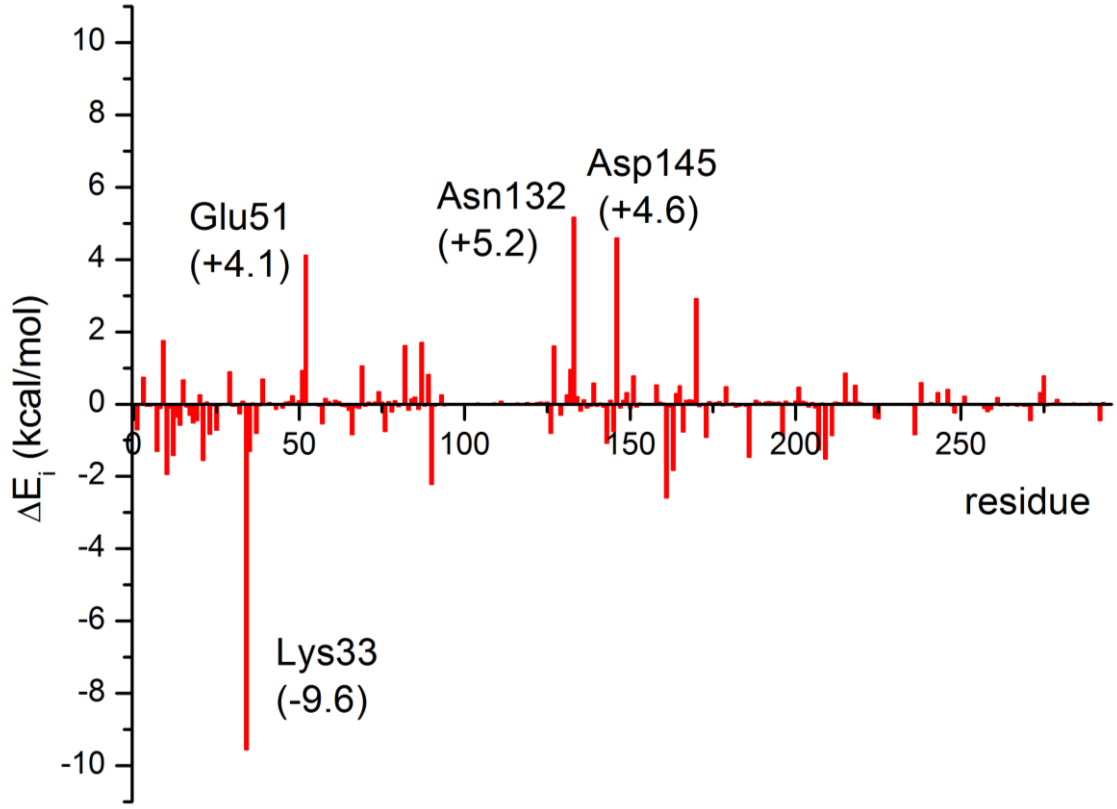


Figure 6.6 – Results of the charge perturbation analysis of the non-frontline residues in CDK2. A negative or positive value indicates that the charge of the residue increases or decreases the barrier, respectively.

6.4 Discussion

The results of our ab initio QM/MM free-energy simulations suggest that the CDK2 catalysis proceeds with a concerted phosphoryl transfer mechanism with Asp127 serving as the general base. The dissociative transition state observed in our calculation is consistent with two X-ray structures of activated CDK2/Cyclin A transition-state anaglogs.[22,38] For example, one X-ray structure indicates that the nitrate moiety binds

between ADP and the phosphorylizable serine, but lacks stabilization from positively charged residues, consistent with observation that nitrate is not an effective inhibitor of CDK2.[38] On the other hand, these authors found that orthovanadate, a mimic for an associative transition state, did not bind the CDK2 complex and did not inhibit the catalysis. These observations led them to conclude that the phosphorylation reaction catalyzed by CDK2 follows a mostly concerted mechanism with a dissociative transition state, which is also the conclusion derived from our QM/MM calculations reported here. In the more recent X-ray structure,[22] a transition-state analog, MgF_3^-, was found between ADP and the Ser nucleophile, with distances of 2.8 and 2.5 Å. This is again consistent with a dissociative transition state.

As mentioned earlier, most protein kinases share the same catalytic domain with highly conserved active-site residues despite sequence and structure diversities. In addition, the active-site residues in the activated form of protein kinases are arranged in a similar fashion, suggestive of a common catalytic mechanism.[6] In Fig. 6.7, the active site of CDK2 is overlaid with that of PKA[78] to illustrate the striking similarities in the catalytic scaffold of the two kinases. In the more extensively studied PKA, there was also a controversy on the role of the Asp residue (Asp166) that is hydrogen bonded with the serine nucleophile.[79] Despite mutagenesis data that its replacement by Ala reduced k_{cat} by approximately three orders of magnitude,[77] some experimental data have cast doubt on its role as the general base.[80] Theoretically, it was suggested based on a semi-empirical model that Asp plays a structural, rather than catalytic, role in the PKA catalysis.[81,82] However, it was later argued by several authors using more accurate ab

initio models that Asp indeed serves as the general base.[33,35,37] Our results reported here also support a general base mechanism for Asp127 for CDK2, consistent with the proposed mechanism for PKA,[33,35,37] and other protein kinases.[83]

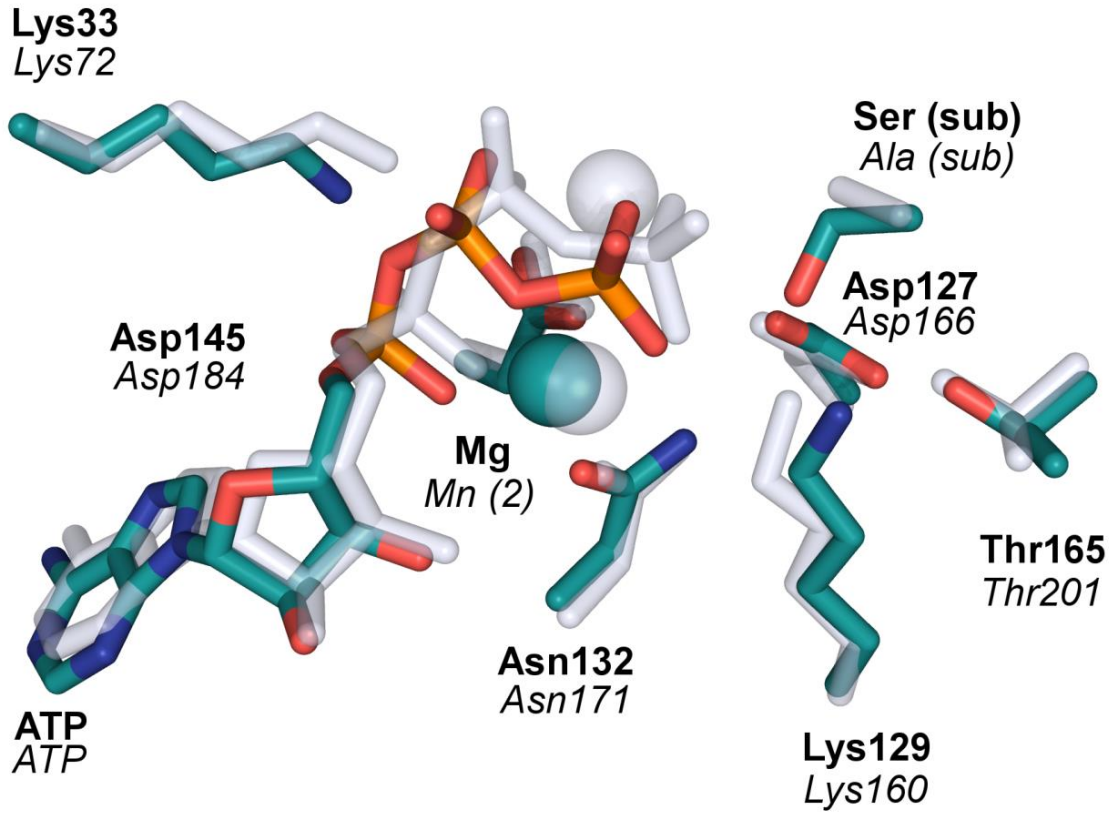


Figure 6.7 – Comparison of the active-site arrangement of CDK2 (green)[17] and PKA (gray).[78] Note that the nucleophilic Ser in the PKA structure was mutated to Ala, and the two Mg(II) ions were replaced by Mn(II) ions.

The calculated barrier height for the phosphorylation step (10.8 kcal/mol) is somewhat lower than that derived from the kinetic data (15.3 kcal/mol from k_3=35 s^{-1}).[24] (k_{cat} for CDK2 is much smaller due to the fact that the reaction, like that catalyzed by PKA, is limited by product release.[23,24]) However, the substrate used in the kinetic studies of

Hagopian et al. has the PK**T**PKKAKKL sequence,[23,24] which differs significantly from the one used in our study (HHA**S**PRK), including the identity of the nucleophile (**T** vs. **S**). Although there has been no kinetic data on the differences between serine and threonine, a recent theoretical study suggested that the former has a lower barrier for the phosphoryl transfer reaction by approximately 3 kcal/mol.[83] This difference would bring the theoretical and experimental data to a much better agreement. On the other hand, the work of De Vivo *et al.* reported a barrier of ~24 kcal/mol,[32] which is in less favorable agreement with the experimental data.

The role of Asp127 as the general base is also supported by a recent study of proton inventory of the CDK5/p35 enzyme, which suggests a general base with a pK_a of 6.1.[30] In the same study, a solvent kinetic isotope effect (SKIE) of 2.0 was observed for k_{cat} at high Mg(II) ion concentrations, which has been interpreted as the evidence that a single proton transfer is coupled with the transition state. Interestingly, at low Mg(II) ion concentrations, which are probably more relevant to our model, no SKIE was observed. However, this could be due to the fact that the product release is slower than the chemical step.[23,24] In any event, it should be pointed out that our calculations did not consider the potential involvement of a second Mg(II) ion in the active site of CDK2, for which a binding pocket exists. The involvement of the second Mg(II) ion is known for many other protein kinases, such as PKA (see Figure 6.7).[79] Interestingly, a recent X-ray structure of CDK2/Cyclin A complexed with a transition-state analog (MgF_3^-) did find two Mg(II) ions in the active site, but the second Mg(II) is believed to be transitory during turnover.[22] MD studies by the same authors have indicated that the binding of

the second Mg(II) cofactor seems to rigidify the complex, which might in turn helps the catalysis. In addition, the kinetic study of CDK5 showed that k_{cat} increases significantly with the Mg(II) concentration, suggesting a role of the second Mg(II) ion in the catalysis.[30]

It appears that both experimental and theoretical evidence exists in support of the role of Asp127 as the general base in CDK2 catalysis. The different conclusion reached by De Vivo *et al.*,[32] that Asp127 plays a structural role and the activation of the serine nucleophile is due to proton transfer to ATP, may be explained by the exclusion of Asp127 from the QM region, which precludes its catalytic role. The earlier DFT cluster model of the same group[31] did not consider the enzyme/solvent environment. The different conclusions reached by the two QM/MM studies underscore the importance of model building.

The mechanistic proposal suggested by De Vivo *et al.*[32] can be considered as a reincarnation of an earlier proposal for non-enzymatic phosphorylation reactions involving phosphate monoester dianions in solution, in which the transferring phosphoryl group is supposed to serve as the general base to deprotonate the nucleophile.[84] However, this substrate-assisted mechanism was later found to be inconsistent with experimental data.[20] This point is interesting in that our results, along with an increasing body of evidence,[33-37,83] seem to suggest that protein kinases employ the same mechanism as non-enzymatic phosphoryl transfer reactions for phosphate monoester dianions. The important determinants of this catalytic machinery include an

active site that aligns the reactants for in-line attack, one or more Mg(II) ions that help to bind the ATP and to stabilize transition state, and an elaborate hydrogen-bond network that stabilizes the charges in the transition state. In addition, there is increasing evidence from *ab initio* QM/MM studies that several other enzyme-catalyzed phosphoryl transfer reactions involving phosphate monoesters also favors the dissociative transition state.[85-87]

6.5 Conclusions

We report extensive *ab initio* QM/MM studies of the phosphoryl transfer mechanism in the CDK2 catalyzed phosphorylation of a peptide serine residue. In contrast to an earlier QM/MM study which assigned ATP as the general base,[32] our results suggest a catalytic role for the conserved Asp127, which activates the serine nucleophile via proton transfer. Our model has also found that that the catalyzed phosphoryl transfer reaction has a concerted mechanism, featuring a single dissociative metaphosphate-like transition state. The role of the Mg(II) ion is primarily to provide stability of the charges developed along the reaction path. This is helped by several hydrogen bonds provided by some conserved residues in the active site. The catalytic mechanism suggested by our QM/MM calculations is consistent with that determined for a more extensively studied protein kinase, namely PKA. This and other mechanistic studies of protein kinases thus suggest a common catalytic mechanism featuring a dissociative phosphoryl transfer mechanism assisted by a conserved Asp general base.

6.6 References

1 Hunter, T. Signaling -2000 and beyond. *Cell* **100**, 113-127 (2000).

2 Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672-1676 (1996).

3 Cohen, P. Protein kinases - the major drug target of the twenty-first century? *Nature Reviews* **1**, 309-315 (2002).

4 Noble, M. E. M., Endicott, J. A. & Johnson, L. N. Protein kinase inhibitors: Insights into drug design from structure. *Science* **303**, 1800-1804 (2004).

5 Blaskovich, M. A. T. Drug discovery and protein tyrosine phosphatases. *Current Medicinal Chemistry* **16**, 2095-2176 (2009).

6 Adams, J. A. Kinetic and catalytic mechanisms of protein kinases. *Chemical Reviews* **101**, 2271-2290 (2001).

7 Zhang, Z.-Y. Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Accounts of Chemical Research* **36**, 385-392 (2003).

8 Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1916 (2002).

9 Taylor, S. S. & Radzio-Andzelm, E. Three protein kinase structures define a common motif. *Structure* **2**, 345-355 (1994).

10 Harper, J. W. & Adams, P. D. Cyclin-dependent kinases. *Chemical Reviews* **101**, 2511-2526 (2001).

11 Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews* **9**, 153-166 (2009).

12 Satyanarayana, A. & Kaldis, P. Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene* **28**, 2925-2939 (2009).

13 Bloom, J. & Cross, F. R. Multiple levels of cyclin specificity in cell-cycle control. *Nature Reviews* **8**, 149-160 (2007).

14 Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences* **30**, 630-641 (2005).

15 De Bondt, H. L. *et al.* Crystal structure of cyclin-dependent kinase 2 *Nature* **363**, 595-602 (1993).

16 Brown, N. R. *et al.* Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *The Journal of Biological Chemistry* **274**, 8746-8756 (1999).

17 Brown, N. R., Noble, M. E. M., Johnson, L. N. & Endicott, J. A. The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nature Cell Biology* **1**, 438-443 (1999).

18 Knowles, J. R. Enzyme-catalyzed phosphoryl transfer reactions. *Annual Review of Biochemistry* **49**, 877-919 (1980).

19 Hengge, A. C. in *Comprehensive Biological Catalysis* Vol. 1 (ed M. Sinnot) 517-542 (Academic Press, 1998).

20 Admiraal, S. J. & Herschlag, D. The substrate-assisted general base catalysis model for phosphate monoester hydrolysis: Evaluation using reactivity comparisons. *Journal of the American Chemical Society* **122**, 2145-2148 (2000).

21 Klahn, M., Rosta, E. & Warshel, A. On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins. *Journal of the American Chemical Society* **128**, 15310–15323 (2006).

22 Bao, Z. Q., Jacobsen, D. M. & Young, M. A. Brief bound to activate: Transient binding of a second catalytic magnesium activate the structure and dynamics of CDK2 kinase for catalysis. *Structure* **19**, 675-690 (2011).

23 Hagopian, J. C. *et al.* Kinetic basis for activation of CDK2/Cyclin A by phosphorylation. *The Journal of Biological Chemistry* **276**, 275-280 (2001).

24 Stevenson, L. M., Deal, M. S., Hagopian, J. C. & Lew, J. Activation mechanism of DCK2: Role of cyclin binding versus phosphorylation. *Biochemistry* **8528-8534** (2002).

25 Clare, P. M. *et al.* The cyclin-dependent kinases cdk2 and cdk5 act by a random, anticooperative kinetic mechanism *The Journal of Biological Chemistry* **276**, 48292-48299 (2001).

26 Elphick, L. M. *et al.* A quantitative comparison of wild-type and gatekeeper mutant Cdk2 for chemical genetic studies with ATP analogues. *ChemBioChem* **10**, 1519–1526 (2009).

27 Joshi, A. R., Jobanputra, V., Lele, K. M. & Wolgemuth, D. J. Distinct properties of cyclin-dependent kinase complexes containing cyclin A1 and cyclin A2. *Biochemical and Biophysical Research Communications* **378**, 595-599 (2009).

28 Child, E. S. *et al.* A cancer-derived mutation in the PSTAIRE helix of cyclin-dependent kinase 2 alters the stability of cycline binding. *Biochimica et Biophysica Acta* **1803**, 858-864 (2010).

29 Liu, M. *et al.* Kinetic studies of cdk5/p25 kinase: Phosphorylation ot tau and complex inhibition by two prototype inhibitors. *Biochemistry* **47**, 8367-8377 (2008).

30 Liu, M., Girma, E., Glicksman, M. A. & Stein, R. Kinetic mechanistic studies of cdk5/p25-catalyzed H1P phosphorylation: Metal effect and solvent kinetic isotope effect. *Biochemistry* **49**, 4921-4929 (2010).

31 Cavalli, A., De Vivo, M. & Recanatini, M. Density functional study of the enzymatic reaction catalyzed by a cyclin-dependent kinase. *Chemical Communications*, 1308-1309 (2003).

32 De Vivo, M., Cavalli, A., Carloni, P. & Recanatini, M. Computational study of the phosphoryl transfer catalyzed by a cyclin-dependent kinase. *Chemistry - A European Journal* **13**, 8437-84444 (2007).

33 Valiev, M., Kawai, R., Adams, J. A. & Weare, J. H. The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: First-principles calculations. *Journal of the American Chemical Society* **125**, 9926-9927 (2003).

34 Diaz, N. & Field, M. J. Insights into the phosphoryl-trasnfer mechanism of cAMP-dependent protein kinase from quantum chemical calculations and molecualr dynamics simulations. *Journal of the American Chemical Society* **126**, 529-542 (2004).

35 Cheng, Y., Zhang, Y. & McCammon, J. A. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: An ab initio QM/MM study. *Journal of the American Chemical Society* **127**, 1553-1562 (2005).

36 Henkelman, G., LeBute, M. X., Tung, C.-S., Fenimore, P. W. & McMahon, B. H. Conformational dependence of a protein kinase phosphate transfer reaction. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15347-15351 (2005).

37 Valiev, M., Yang, J., Adams, J. A., Taylor, S. S. & Weare, J. H. Phosphorylation reaction in cAPK protein kinase - Free energy quantum mechanical/molecular mechanics simulations. *Journal of Physical Chemistry B* **111**, 13455-13464 (2007).

38 Cook, A. *et al.* Structural studies on phospho-CDK2/Cyclin A bound to nitrate, a transition state analogue: Implications for the protein kinase mechanism. *Biochemistry* **41**, 7301-7311 (2002).

39 Warshel, A. & Levitt, M. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103**, 227-249 (1976).

40 Gao, J. in *Reviews in Computational Chemistry* Vol. 7 (eds K. B. Lipkowitz & D. B. Boyd) 119-185 (VCH, 1996).

41 Gao, J. *et al.* Mechanisms and free energies of enzymatic reactions. *Chemical Reviews* **106**, 3188-3209 (2006).

42 Hu, H. & Yang, W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annual Review of Physical Chemistry* **59**, 573-601 (2008).

43 Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie (International ed. in English)* **48**, 1198-1229 (2009).

44 Ranaghan, K. E. & Mulholland, A. J. Investigation of enzyme-catalyzed reactions iwth combined quantum mechanical/molecular mechanical (QM/MM) methods. *International Reviews in Physical Chemistry* **29**, 65-133 (2010).

45 Case, D. A. *et al.* Amber 10. *University of California, San Francisco* (2008).

46 Cornell, W. D. *et al.* A second generation force field for the simulation of protein, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197 (1995).

47 Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).

48 Meagher, K. L., Redman, L. T. & Carlson, H. Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry* **24**, 1016-1025 (2003).

49 Homeyer, N., Horn, A. H. C., Lanig, H. & Sticht, H. AMBER force field parameters for phosphorylated amino acids in different protonation states: Phosphoserine,

phosphothreonine, phosphotyrosine and phosphohistidine. *Journal of Molecular Modeling* **12**, 281-289 (2006).

50 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **79**, 926-935 (1983).

51 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an NlogN method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089-10092 (1993).

52 Essmann, U. *et al.* A smooth particle mesh Ewald method. *Journal of Chemical Physics* **103**, 8577-8593 (1995).

53 Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).

54 Zhang, Y., Lee, T. & Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *Journal of Chemical Physics* **110**, 46-54 (1999).

55 Zhang, Y., Liu, H. & Yang, W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *Journal of Chemical Physics* **112**, 3483-3492 (2000).

56 Zhang, Y. Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods. *Journal of Chemical Physics* **122**, 24114 (2005).

57 Zhang, Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts* **116**, 43-50 (2006).

58 Wang, S., Hu, P. & Zhang, Y. Ab initio quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: The case of histone lysine methyltransferase SET7/9. *Journal of Physical Chemistry B* **111**, 3758-3764 (2007).

59 Hu, P., Wang, S. & Zhang, Y. How do SET-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by ab initio QM/MM molecular dynamics simulations. *Journal of the American Chemical Society* **130**, 3806-3813 (2008).

60 Hu, P., Wang, S. & Zhang, Y. Highly dissociative and concerted mechanism for the nicotinamide cleavage reaction in Sir2Tm enzyme suggested by ab initio QM/MM molecular dynamics simulations. *Journal of the American Chemical Society* **130**, 16721-16728 (2008).

61 Ke, Z., Wang, S., Xie, D. & Zhang, Y. Born-Oppenheimer ab initio QM/MM molecular dynamics simulations of the hydrolysis reaction catalyzed by protein arginine deiminase 4. *Journal of Physical Chemistry B* **113**, 16705–16710 (2009).

62 Ke, Z. *et al.* Active site cysteine is protonated in the PAD4 Michaelis complex: Evidence from Born-Oppenheimer ab initio QM/MM molecular dynamics simulations. *Journal of Physical Chemistry B* **113**, 12750–12758 (2009).

63 Wu, R. B., Wang, S. L., Zhou, N. J., Cao, Z. X. & Zhang, Y. K. A proton-shuttle reaction mechanism for histone deacetylase 8 and the catalytic role of metal ions. *Journal of the American Chemical Society* **132**, 9471-9479 (2010).

64 Zhou, Y. Z., Wang, S. L. & Zhang, Y. K. Catalytic reaction mechanism of acetylcholinesterase determined by Born-Oppenheimer ab initio QM/MM molecular dynamics simulations. *Journal of Physical Chemistry B* **114**, 8817-8825 (2010).

65 Zhou, Y. Z. & Zhang, Y. K. Serine protease acylation proceeds with a subtle re-orientation of the histidine ring at the tetrahedral intermediate. *Chemical Communications* **47**, 1577-1579 (2011).

66 Ke, Z. H., Guo, H., Xie, D., Wang, S. L. & Zhang, Y. K. Ab initio QM/MM free-energy studies of arginine deiminase catalysis: The protonation state of the Cys nucleophile. *Journal of Physical Chemistry B* **115**, 3725-3733 (2011).

67 Thiel, W. & Voityuk, A. A. Extension of MNDO to d orbitals: Parameters and results for the second-row elements and for the zinc group. *Journal of Physical Chemistry* **100**, 616-626 (1996).

68 Range, K., Lopez, C. S., Moser, A. & York, D. M. Multilevel and density function electronic structure calculations of proton affinities and gas phase basicities involved in biological phosphoryl transfer. *Journal of Physical Chemistry A* **110**, 791-797 (2006).

69 QChem (www.q-chem.com) (2006).

70 TINKER, Software Tools for Molecular Design (dasher.wustl.edu/ffe/) (2004).

71 Torrie, G. M. & Valleau, J. P. Non-physical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).

72 Beeman, D. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics* **20**, 130-139 (1976).

73 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **81**, 3684-3690 (1984).

74 Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free energy calculations on biomolecules. 1. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).

75 Bash, P. A. *et al.* Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* **30**, 5826-5832 (1991).

76 Mildvan, A. S. Mechanisms of signaling and related enzymes. *Proteins* **29**, 401-416 (1997).

77 Gibbs, C. S. & Zoller, M. J. Rational scanning mutagenesis of a protein kinase identifies functional regions involved in catalysis and substrate interactions. . *The Journal of Biological Chemistry* **266**, 8923-8931 (1991).

78 Zheng, J. *et al.* 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallographica Section D* **49**, 362-365 (1993).

79 Madhusudan, P. A., Xuong, N.-H. & Taylor, S. S. Crystal structure of a transition state mimic of the catalytic subunit of cAMP-dependent protein kinase *Nature Structural & Molecular Biology* **9**, 273-277 (2002).

80 Zhou, J. & Adams, J. A. Is there a catalytic base in the active site of cAMP-dependent protein kinase? *Biochemistry* **36**, 2977-2984 (1997).

81 Hart, J. C., Hillier, I. H., Burton, N. A. & Sheppard, D. W. An alternative role for the conserved Asp residue in phosphoryl transfer reactions. *Journal of the American Chemical Society* **120**, 13535-13536 (1998).

82 Hart, J. C., Sheppard, D. W., Hillier, I. H. & Burton, N. A. What is the mechanism of phosphoryl transfer in protein kinases? A hydbrid quantum

mechanical/molecular mechanical study. *Chemical Communications*, 79-80 (1999).

83 Turjanski, A. G., Hummer, G. & Gutkind, J. S. How mitogen-activated protein kinases recognize and phosphorylate their targets: a QM/MM study. *Journal of the American Chemical Society* **131**, 6141-6148 (2009).

84 Aqvist, J., Kolmodin, K., Florian, J. & Warshel, A. Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experiemntal data? *Chemical Biology* **6**, R71-R80 (1999).

85 De Vivo, M. *et al.* Proton shuttles and phosphatase activity in soluble epoxide hydrolase. *Journal of the American Chemical Society* **129**, 387-394 (2006).

86 Grigorenko, B. L., Nemukhin, A. V., Shadrina, M. S., Topol, I. A. & Burt, S. K. Mechanisms of guanosine triphosphate hydrolysis by Ras and Ras-GAP proteins as rationalized by ab initio QM/MM simulations. *Proteins* **66**, 456-466 (2007).

87 Grigorenko, B. L. *et al.* Mechanisms of the myosin catalyzed hydrolysis of ATP as rationalized by molecular modeling. . *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7057-7061 (2007).

Chapter 7

Quantum mechanical/molecular mechanical study of anthrax lethal factor catalysis

“What you see is that the most outstanding feature of life's history is a constant domination by bacteria.”

— Stephen Jay Gould

This chapter is based on collaborative work originally found in the following publication:

Adapted from Smith CR, **Smith GK**, Zhenxiao Y, Dingguo X, Guo H. Quantum mechanical / molecular mechanical study of anthrax lethal factor catalysis. *Theoretical Chemistry Accounts*. 2011; 128(1): 83-90. Copyright 2011 Springer-Verlag.

7.1 Introduction

Anthrax is an infection of humans and animals caused by the Gram-positive bacterium *Bacillus Anthracis*. Although anthrax is one of the oldest diseases known to man, it has recently attracted renewed attention because of its potential as an agent for bioterrorism and biological warfare.[1] This bacterium has the ability to form spores, which can lay dormant for decades even under unfavorable conditions. Infections occur when the bacterial spores enter the host via lacerations, ingestion, or inhalation. While the most common cutaneous anthrax is curable, systemic infection via inhalation is significantly more dangerous, leading to host death within days without treatment, and significant mortality rates even with early intervention.[2] The inhaled spores germinate in alveolar macrophages and are carried to the lymph nodes where they multiply and enter the bloodstream in mass. The vegetative bacteria then release several virulence factors, causing massive toxemia which leads to eventual host death. Due to the circulation of virulence factors, antibiotics are only effective if started early, and any late stage treatments must address the circulating toxins as well. The difficulties in diagnosis coupled with the short time window available to take effective action make anthrax a potent threat.

The primary virulence factors of *B. anthracis* include a protective capsule that inhibits phagocytosis and three polypeptides: protective antigen (PA), edema factor (EF), and lethal factor (LF).[3] The peptides work in concert as two binary toxins, as PA binds to surface anthrax toxin receptors (ATR), and facilitates entry of either EF or LF into the

cytosol. EF is a calcium and calmodulin-dependent adenylate cyclase that raises cAMP concentrations within the cell, leading to edema. The more potent of the two is LF, a zinc dependent protease that targets mitogen-activated protein kinase kinase (MAP2K). Its central role in mortality has been demonstrated in animal studies which show injection of lethal toxin (LF+PA) leads to similar progression as the bacterial infection.[1]

The LF is a four domain ~90 kDa protease with a zinc cofactor and consensus zinc-binding motif HEXXH.[4] This highly specific metalloprotease cleaves the amide bond in MAP2Ks near their N termini.[5-10] Although the exact mechanism of the resulting cell death is still unclear, it is understood that the cleavage of MAP2K disrupts downstream processes which are responsible for activating host immunological and inflammatory responses.[11] Given its central role in anthrax infections, LF offers an important target for designing therapeutic agents.[12] Although some successes have been reported,[13-18] it is apparent that further advancement will benefit from better understanding of the substrate binding and catalysis of LF.

There have been a number of kinetic studies of the LF catalysis,[19,20] which reported K_m and k_{cat} values for various substrates. X-ray structures of LF have also been reported in its apo form[21] and complexed with substrates[14,21] and inhibitors.[13,15-17] Apart from the hallmark HEXXH motif, the catalytic domain (IV) of LF has no sequence homology with any other known proteins, but its backbone scaffold resembles closely that of thermolysin.[21] Figure 7.1 shows the X-ray structure of LF complexed with a peptide substrate (LF20).[14]

It is now well-established that the zinc cofactor in the catalytic domain is coordinated at the bottom of the active-site cleft by three protein ligands: His686, His690, Glu735, and a water molecule.[21]

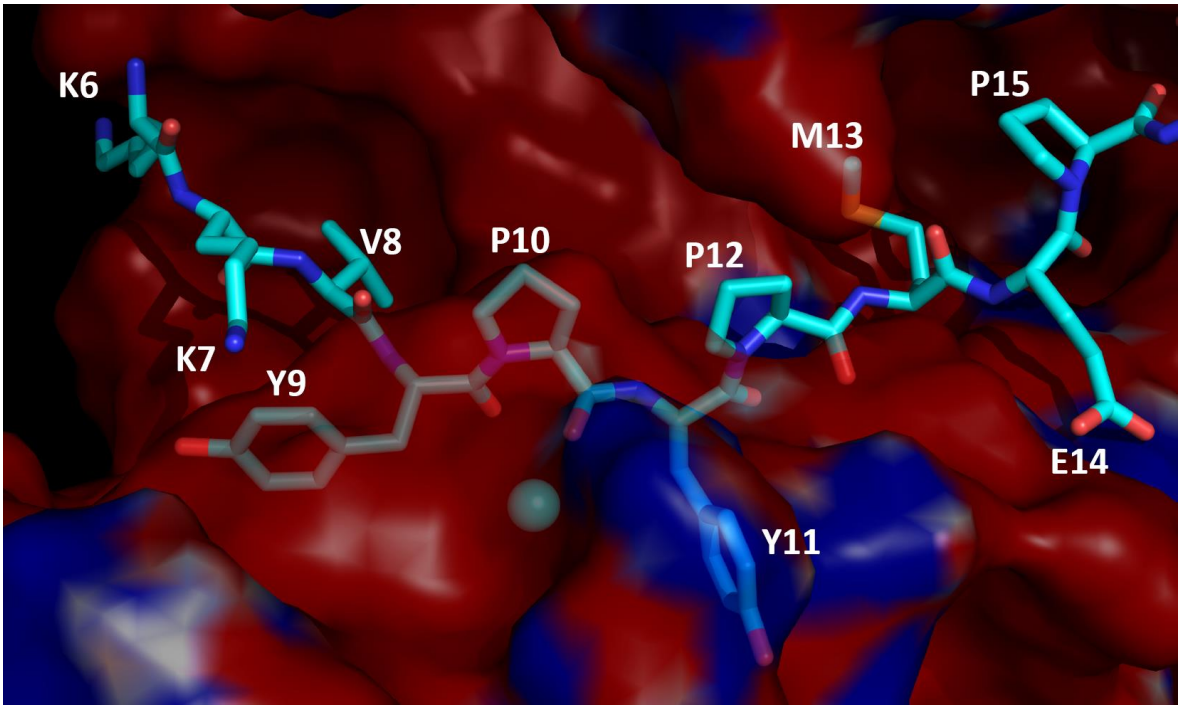
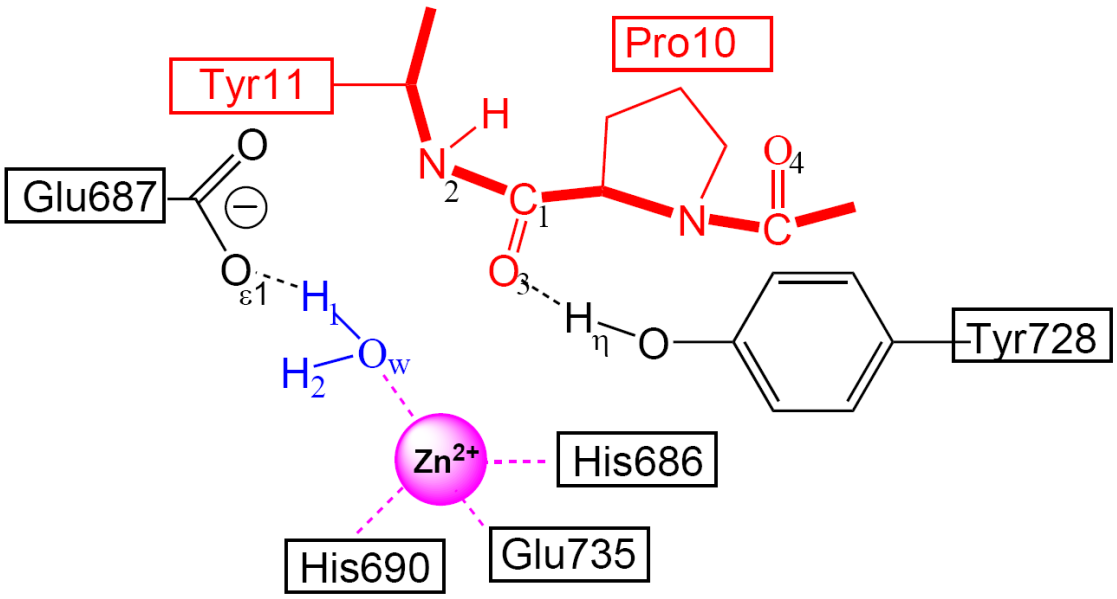


Figure 7.1 – The X-ray structure of anthrax lethal factor complexed with the LF20 peptide (PDB code 1PWW).[14] The protein surface is colored by the solvent accessible potential map, where blue represents areas of negative potential, red positive, and white neutral. LF20 is shown in stick representation, where carbon is light blue, nitrogen dark blue, and oxygen is red.

Mutations of zinc-binding residues completely abolish the activity of LF.[4,22,23] Near the metal center, there are two important second-shell residues: Glu687 and Tyr728 (see Scheme 7.1).

The LF active site is quite analogous to the much more extensively studied thermolysin[24] and carboxypetidase A[25] and a similar catalytic mechanism has been proposed.[21] This entails Glu687 serving as the general base to activate a zinc-bound water, which attacks the scissile amide bond. Mutation of Glu687 to Cys has been found to completely inactivate LF.[4,22]



Scheme 7.1 – Arrangement of the active site of the anthrax lethal factor and atom definition. The substrate is coded red while the nucleophilic water is coded blue.

The role of the nearby Tyr728 is more controversial. It was initially suggested based on the crystallographic structure of the apo LF that it may serve as a general acid to protonate the amine leaving group.[21] Later, Tonello *et al.* found that its mutation to Phe results in the complete abolishment of activity and these authors proposed that the Tyr residue might be involved in stabilizing the leaving amide group of the substrate.[23]

However, it is also possible that it serves as part of the oxyanion hole to stabilize the negative charge developed at the carbonyl oxygen atom,[12] a role played by His231 in thermolysin and Arg127 in carboxypeptidase A.

There have been few computational investigations of LF, and none has investigated its catalysis. Johnson *et al.* presented in 2006 a structure-activity relationship (SAR) study of several inhibitors of LF.[26] More recently, Hong *et al.* studied the Michaelis complex using density functional theory (DFT) and molecular dynamics (MD).[27] Convincing evidence was shown that the zinc bound nucleophile is a neutral water, rather than hydroxide, and it forms a hydrogen bond with the putative general base Glu687.[27] The interaction of LF with several MAPK/ERK and MAP2K substrates was also investigated recently by Dalkas *et al.*,[28] using docking and MD simulations. Both MD simulations have relied on force fields in describing the Zn(II)-ligand binding. Here, we report a theoretical investigation on the catalysis of LF. In particular, we studied the Michaelis complex as well as catalysis of the wildtype and Y728F mutant using a quantum mechanical/molecular mechanical (QM/MM) approach. These studies allowed a better understanding of the LF catalysis.

7.2 Methods

The methods used in this project are described in a condensed fashion below. Foundational information and additional details on the techniques used can be found in Chapter 2 (Classical Methods) and Chapter 3 (Quantum Methods).

7.2.1 Model

The starting point of the simulation was selected from an enzyme-substrate complex structure (PDB code 1PWW),[14] which is a point mutation (E687C) of LF complexed with a optimized synthetic substrate (LF20). This structure was chosen because it represented the ES complex that showed the least amount of deformation after mutation and bound the substrate in what appeared to be a productive conformation. Note that the substrate in this LF complex is not a member of the MAP2K family. Rather, it is an analog with a consensus sequence $M_1L_2A_3R_4R_5K_6K_7V_8Y_9$-$P_{10}Y_{11}P_{12}M_{13}E_{14}P_{15}T_{16}I_{17}A_{18}E_{19}G_{20}$-amide, in which the amide bond between Y_9 and P_{10} represents the cleavage site. In the X-ray structure, only the residues between K_6 and E_{14} in this substrate were resolved, and no attempt was made in our simulations to extend it to its full length. The 1PWW structure was also used by Hong *et al.* in their recent MD simulations of LF.[27] The earlier structure reported by Pannifer *et al.*[21] was also examined but not used due to the fact that the substrate was bound in a non-productive conformation and a ~180^o turn of the peptide chain is required to bring it to the correct binding mode. Such non-productive binding of substrates is known to be kinetically viable.[19]

The X-ray structure was modified by recovering the Glu687 side chain *in silico*. A water molecule was then added to the active site close to the metal ion. Hydrogen atoms were added using the HBUILD utility in CHARMM,[29] and the titratable residues in the enzyme were assigned the appropriate ionization states at pH=7. In particular, the histidine residues took the following ionization states: HSD (H on $\mathrm{N}_{\delta 1}$) for residues 42, 117, 197,

229, 424, 645, 654, 669, 686, 690, and 745, while HSE (H on $\mathrm{N}_{\delta 2}$) for residues 35, 91, 115, 277, 280, 309, 588, and 749. The resulting structure was then solvated by a pre-equilibrated sphere of TIP3P waters[30] with 25 Å radius centered at the zinc ion, followed by a 10 ps MD simulation with all protein and substrate atoms fixed. This process was repeated twice with randomly rotated water spheres to ensure uniform solvation.

To reduce computational costs, stochastic boundary conditions were employed.[31] To this end, atoms outside the 25 Å radius were removed while atoms in the buffer zone (22 Å < *r* < 25 Å) were subjected to harmonic restraining potentials. In the inner reaction region (*r* < 22 Å), the motion of atoms is governed by the QM/MM potential. A group-based switching scheme was used for non-bonded interactions.[32] The MD simulations feature Newtonian dynamics for atoms in the reaction zone, augmented by Langevin dynamics of atoms in the buffer region at 300 K.

7.2.2 QM/MM

The combined quantum mechanical and molecular mechanical (QM/MM) approach[33] has been extensively applied to study enzymatic systems.[34-39] Such a method has the advantage that a very large system such as an enzyme can be investigated with manageable computation costs. The basic idea of the QM/MM scheme is to divide the system into two parts: the smaller QM region where the chemical bond breaking and forming take place is treated with quantum mechanics, while the surrounding region is described by a MM force field.

In this work, the QM region was characterized by the self-consistent charge-density functional tight binding (SCC-DFTB) model[40] and the MM region by the CHARMM all atom force field.[41] SCC-DFTB is an approximate DFT method and it is much more efficient than ab initio QM/MM approaches. The efficiency is essential for metallo-enzymes because the active site is often much too large for an accurate ab initio QM/MM free-energy simulation. The SCC-DFTB model has been extensively tested for enzyme systems[37,42-44] including those with zinc cofactors.[45] Its accuracy is comparable to the commonly used AM1 and PM3.[43,44] The QM/MM approach based on the SCC-DFTB method has been shown to give a reasonable description of several zinc enzymes including carbonic anhydrase[46,47] β-lactamases,[48-52] cytidine deaminase,[53] and carboxypeptidase A.[54] On the other hand, we also note that SCC-DFTB is not appropriate for all situations, such as was reported for the ligand binding energy of Zn(II).[55] However, complete dissociation is not relevant here and we expect the impact of this imperfection to be minimal for near equilibrium properties.

All simulations were performed with CHARMM (version 32a2) with a SCC-DFTB interface.[42] The QM region includes the side chains of His686, His690, Glu687, Glu735, and Tyr728; all atoms of Pro_{10} and the backbone atoms of Tyr_9 and Tyr_{11} from the substrate; the catalytic water, and the zinc ion. At the boundary, the link atom (QQ) approach[56] was used. All mutants were treated in a similar fashion.

In the QM/MM MD simulation of the Michaelis complex, the starting system was brought to 300 K in 50 ps, followed by 150 ps of equilibration. The final 800 ps MD

trajectory was used for analysis. The MD trajectories were integrated with a 1.0 fs time interval, and the SHAKE algorithm[57] was used to maintain all covalent bonds involving hydrogen atoms.

Minimal energy paths were calculated using adiabatic mapping along two putative reaction coordinates. For the nucleophilic addition (NA) of the water nucleophile, the reaction coordinate is given by the distance between the water oxygen (O_w) and the substrate carbonyl carbon (C_1): $r_1 = d_{O_w \cdots C_1}$. For the elimination (E) of the leaving group, the corresponding reaction coordinate is given by a combination of N_2-H_1 and C_1-N_2 distances: $r_2 = d_{N_2 \cdots H_1} - d_{C_1 \cdots N_2}$.

7.2.3 PMF

The minimum energy structures along a putative reaction coordinate were used as the initial points for calculating the potential of mean force (PMF). The PMF calculations used umbrella sampling[58] with harmonic constraints of 100-400 kcal/mol·Å^2. At least twenty windows were used for the NA and E steps. In each window, a 50 ps equilibration simulation was first performed to bring the system to 300 K. The distribution function in the reaction coordinate was then collected in the subsequent 100 ps. The final PMF was obtained using the weighted histogram analysis method (WHAM).[59]

7.3 Results

The 1.0 ns MD simulation indicated that the enzyme-substrate (ES) complex is reasonably stable, as evidenced by the RMSD (root mean square deviation) shown in Fig. 7.2.

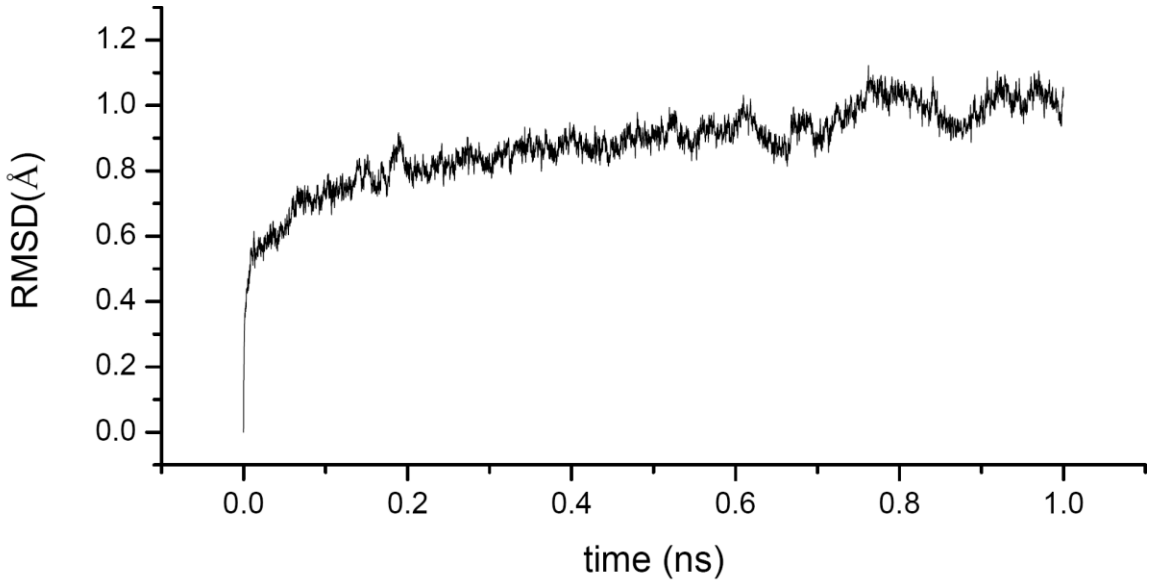


Figure 7.2 – RMSD for backbone atoms obtained in the QM/MM MD simulation of the ES complex.

The averaged values of several key internuclear distances are given in Table 7.1. In the ES complex, the substrate is tightly held by both electrostatic and hydrophobic interactions. As previously noted,[14] two upstream residues at the P4 and P5 positions (Lys_7 and Lys_6) of the substrate are locked in strong electrostatic interaction with negatively charged residues of LF including Asp647 and Glu733. Also note the P2 (Tyr_9) and P1' (Tyr_{11}) residues have their phenyl rings inserting into hydrophobic cavities, as shown in Fig. 7.1.

In the active site of the ES complex, the water serves as the fourth ligand to Zn(II), in addition to His686, His690, and Glu735. The O_w-Zn distance of 2.03±0.06 Å, typical for a neutral oxygen ligand, is consistent with the experimental value of 2.10±0.1 Å.[21] As in X-ray structures,[14,21] the carboxylate-Zn coordination for Glu735 is monodentate. However, multiple rotations of the carboxylate have been observed in the MD simulation, resulting in the switching of the coordination oxygen. Throughout the simulation, nonetheless, the coordination maintains monodentate, as evidenced by the averaged O-Zn distances for the zinc-bound oxygen of 2.07±0.07 Å. The zinc-bound water molecule is strongly hydrogen bonded with the carboxylate side chain of Glu687, and it forms an occasional hydrogen bond with the backbone carbonyl oxygen of $Tyr_9(O_4)$ as well.

Table 7.1 – Some key distances obtained from QM/MM MD simulations of the ES complex of the wildtype LF and from QM/MM reaction path calculations.

Distance (Å)	**MD**	**reaction path**				
	ES	**ES**	**TS1**	**INT**	**TS2**	**EP**
$C_1 - O_w$	2.73 ± 0.22	2.90	1.79	1.50	1.42	1.29
$O_w - H_1$	1.10	1.07	1.28	1.66	2.21	2.35
$O_{\delta 2} - H_1$	1.37 ± 0.07	1.43	1.15	1.02	1.08	1.91
$O_{\delta 1} - H_2$	2.58 ± 0.25	2.57	2.73	2.92	2.34	1.13
$C_1 - O_3$	1.24 ± 0.02	1.23	1.33	1.39	1.35	1.30
$C_1 - N_2$	1.38 ± 0.02	1.38	1.43	1.47	1.79	2.24
$N_2 - H_1$	3.25 ± 0.27	3.67	3.27	3.39	1.59	1.04
$O_4 - H_2$	3.10 ± 0.63	1.88	1.78	1.72	1.83	3.05
$O_3 - H_\eta$	1.97 ± 0.17	1.90	1.85	1.81	1.92	2.05
$O_3 - Zn$	3.59 ± 0.35	2.92	2.12	2.07	2.11	2.11
$O_w - Zn$	2.03 ± 0.06	2.03	2.88	2.89	2.85	2.86
$N(H690) - Zn$	2.03 ± 0.06	2.03	2.03	2.05	2.02	2.00
$N(H686) - Zn$	2.02 ± 0.06	2.04	2.05	2.07	2.02	2.01
$O(E735) - Zn$	2.07 ± 0.07	2.04	2.07	2.07	2.02	2.02

As a result, the nucleophile is in a perfect near attack position, with an average distance of 2.73±0.22 Å between O_w and the scissile carbonyl carbon (C_1). In addition, the scissile carbonyl oxygen (O_3) of the substrate forms a hydrogen bond with the

hydroxyl hydrogen of the Tyr728 side chain, as evidenced by the $O_3 - H_\eta$ distance of 1.97±0.17 Å. On the other hand, we found no evidence that the zinc-bound water interacts with the Tyr728, as suggested by the apo enzyme structure.[21] Interestingly, the carbonyl oxygen O_3 is not directly coordinated with Zn(II) and the distance between the two is 3.59±0.35 Å. The configuration of the active site described above is similar to the recent MD simulation of LF by Hong *et al.*[27]

The catalyzed reaction is initiated by the attack of the water nucleophile at the scissile carbonyl carbon (C_1), which leads to the first transition state (TS1), as shown by the PMF in Fig. 7.3. This nucleophilic addition (NA) barrier features a concerted addition of the water oxygen (O_w) to the carbonyl carbon (C_1) and the transfer of the water proton (H_1) to a carboxylate oxygen ($O_{\varepsilon 1}$) of the general base (Glu687), as suggested by the relevant internuclear distances listed in Table 7.1. The resulting tetrahedral intermediate (TI) is characterized by a sp^3 central carbon and protonated carboxylate side chain of Glu687. Due to the negative charge developed on the scissile carbonyl oxygen (O_3) as a result of the OH^- addition, it replaces the water molecule as the fourth ligand of the zinc ion.

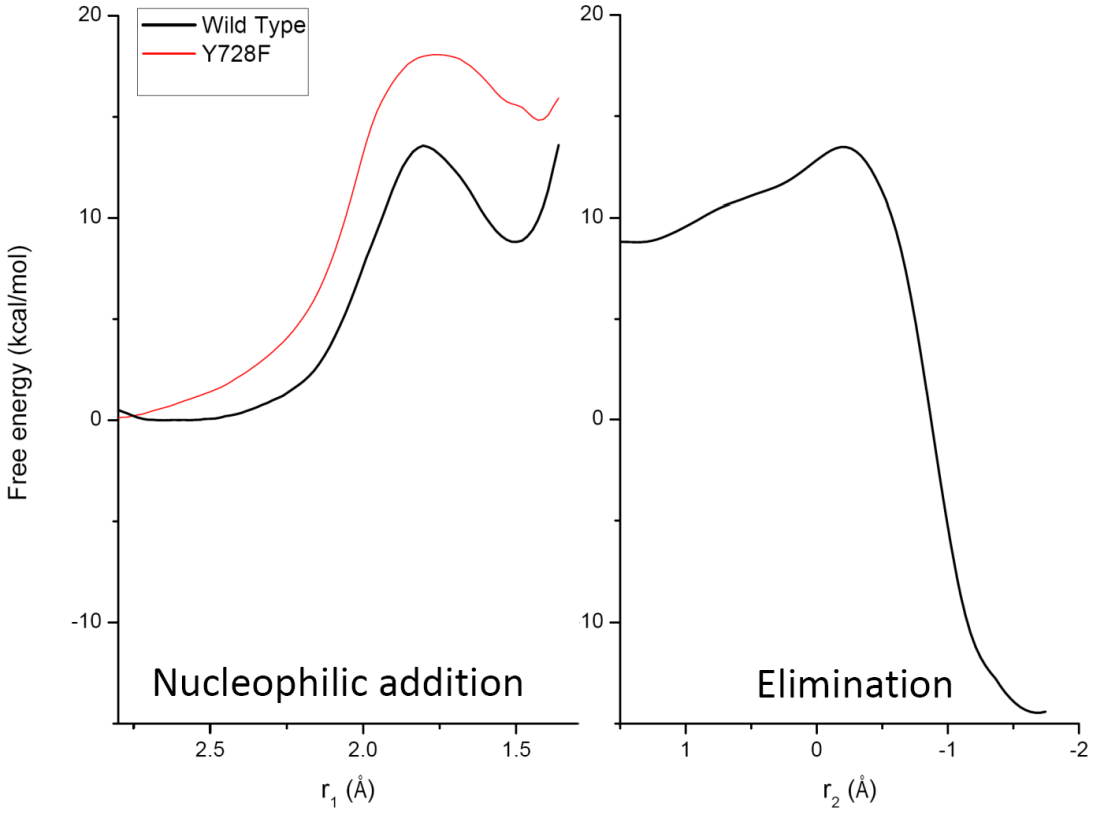


Figure 7.3 – Potentials of mean force (PMFs) for both the nucleophilic addition (NA) and elimination (E) steps of the hydrolysis reaction catalyzed by the wildtype and Y728F mutant of the lethal factor. For the nucleophilic addition (NA) of the water nucleophile, the reaction coordinate is given by the distance between the water oxygen (O_w) and the substrate carbonyl carbon (C_1): $r_1 = d_{O_w \cdots C_1}$. For the elimination (E) of the leaving group, the corresponding reaction coordinate is given by a combination of N_2-H_1 and C_1-N_2 distances: $r_2 = d_{N_2 \cdots H_1} - d_{C_1 \cdots N_2}$.

As Fig. 7.3 shows, the PMF indicates that the TI is metastable and eventually collapses to the enzyme-product (EP) complex via the second transition state (TS2). This elimination (E) barrier features the elongation of the C_1-N_2 bond concomitant with the transfer of the proton from the Glu687 carboxylic acid to the leaving group nitrogen (N_2). The final EP complex contains the two cleaved peptide fragments, each bound to the active site

by hydrogen bonds and augmented by Zn coordination. These two fragments eventually depart the active site to be replaced by solvent water. The general base Glu687 then recovers its ionization state and becomes ready for the next catalytic cycle. Throughout the reaction, the Zn coordination with the three protein ligands is maintained. Snapshots of the stationary points are displayed in Fig. 7.4 and the proposed catalytic mechanism is illustrated in Scheme 7.2.

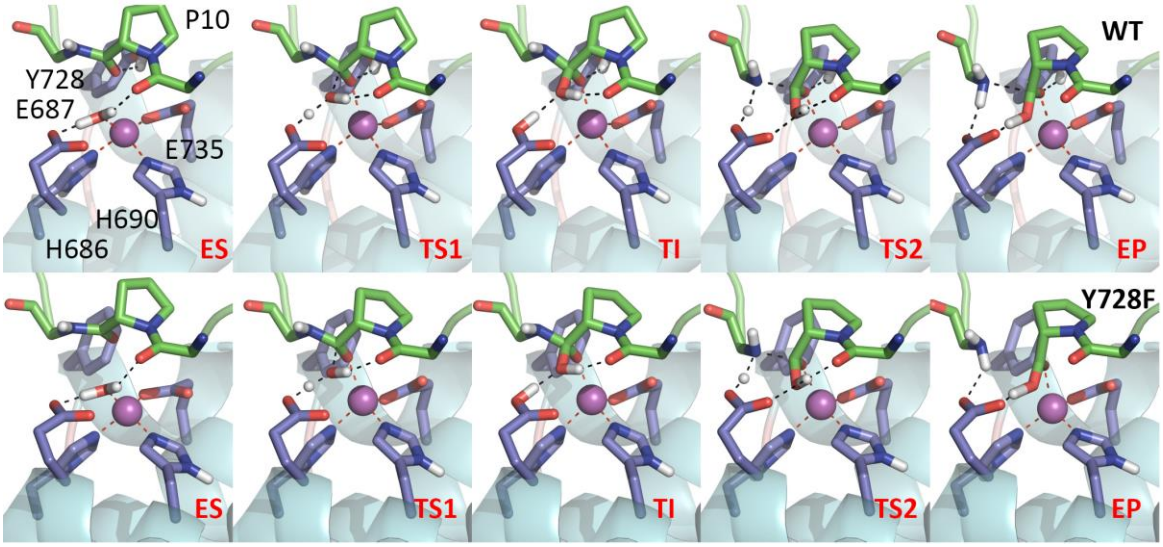
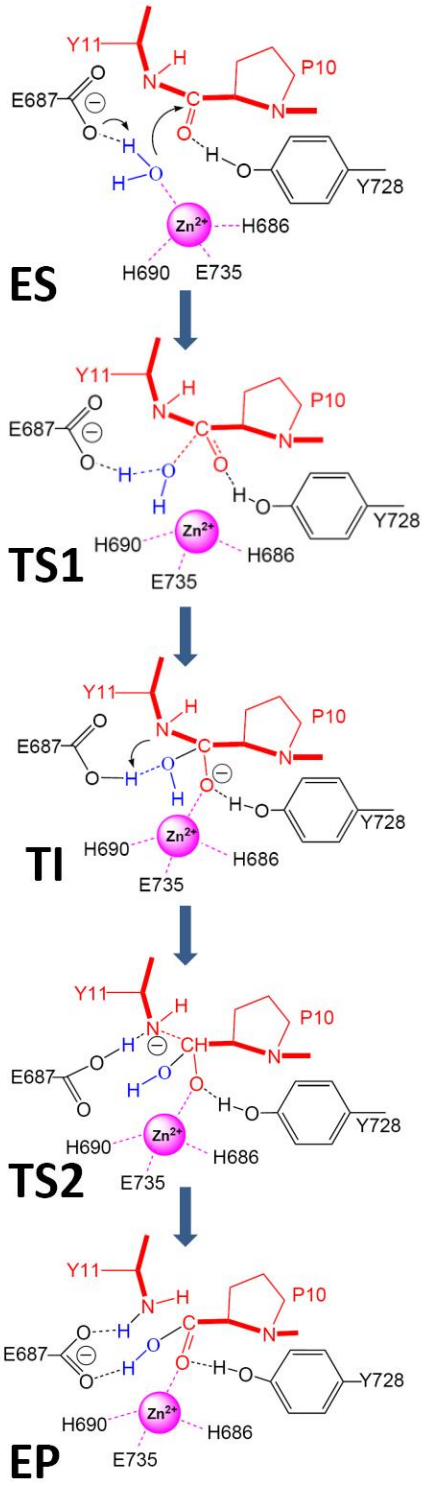


Figure 7.4 – Snapshots of the five stationary points along the reaction path for the wildtype and Y728F mutant of the lethal factor. The red dashed lines indicate the ligand-metal bonds, while the black dashed lines represent either hydrogen bonds or partial bonds. The reaction mechanism is also sketched.



Scheme 7.2 – Proposed catalytic mechanism of anthrax lethal factor based on the QM/MM calculations.

For the wildtype (WT), calculated free energy barriers for the NA and E steps are 13.6 and 13.5 kcal/mol, respectively. Unfortunately, no kinetic data is available for the substrate LF20. However, the k_{cat} value (3.4 s^{-1}) has been measured for LF15,[14] which is closely related to LP20. The corresponding barrier height is 16.7 kcal/mol, as calculated using transition-state theory. It appears that the calculated barrier height underestimates the experimental value, which is consistent with past experiences with SCC-DFBT based QM/MM methods.[42,51,60] We also note in passing that the barrier obtained in the reaction path for the first NA step is 25 kcal/mol. The lowering of the barrier by 11.4 kcal/mol underscores the relaxation of the system allowed by the room temperature MD simulation.

To gain more insights into the catalysis, we have also calculated the PMFs for the Y728F mutant, which has been studied experimentally.[23] The mutation removes the hydroxyl group in the side chain and thus the hydrogen bond with the carbonyl oxygen (O_3). The NA PMFs for the mutant is compared in Fig. 7.3 with that of the WT enzyme. It is clearly seen from the figure that the removal of the hydrogen bond with the scissile carbonyl oxygen (O_3) in the Y728F mutant leads to a significant increase of the barrier. The calculated barrier height of 18.1 kcal/mol for the Y728F mutant is 4.5 kcal/mol higher than the WT, which translates to approximately four orders of magnitude reduction in the rate constant. This dramatic increase of barrier underscores the importance of this residue in the catalysis as a part of the oxyanion hole, which along with the Zn ion stabilizes the negative charge on the carbonyl oxygen (O_3). This calculation result is

consistent with the experimental observation that this point mutation abolished the activity of LF.[23]

7.4 Discussion

Despite different folds, the LF active site is remarkably similar to that of thermolysin (TLN)[24] and carboxypeptidase A (CPA).[25] Like in LF, the two extensively studied peptidases have a single zinc cofactor, ligated by two His and one Glu residue. An active site Glu residue, Glu143 in TLN and Glu270 in CPA, serves as the general base in activating a zinc-bound water nucleophile. In addition, an active site residue provides hydrogen bonding to the scissile carbonyl oxygen. In TLN, it is His231, while in CPA it is Arg127. Site-directed mutagenesis studies indicated that mutation of this residue resulted in significant reduction of the catalytic activity.[61,62] Based on theoretical studies,[54,63] this residue contributes to both substrate binding as well as catalysis, serving as part of the oxyanion hole. This strategy is of course not restricted to these two well-studied zinc hydrolyases.[64] In histone deacetylase, for example, a second shell Tyr residue is also found to stabilize the transition state.[65] Similarly, a Tyr residue was recently found to play an important catalytic role in dizinc lactonases.[66]

Our PMF results reported here indicate that Tyr728 in LF serves the same role as His231 in TLN and Arg127 in CPA. Namely, it helps to bind the substrate and serves as part of the oxyanion hole to stabilize the negative charge developed in the carbonyl oxygen. These conclusions are at odds with the original proposal of Pannifer *et al.*,[21] who suggested that Tyr728 serves as a proton donor to the nitrogen leaving group, and

with that of Tonello et al,[23] who conjectured it as a stabilization for the leaving amino group. Our assigned role for Tyr728 is consistent with the mode of action in similar enzymes such as thermolysin and carboxypeptidase A, and with site-directed mutagenesis experiments on this residue.[23]

Our simulations indicate that the catalytic role of the sole zinc cofactor in LF is to activate the water nucleophile and to provide stabilization for transition state via interaction with the negatively charged carbonyl oxygen. An interesting observation from our simulation is the weak interaction of the scissile carbonyl oxygen with Zn(II) in the ES complex, which suggests that the metal co-factor is not involved in the polarization of the substrate. These conclusions are in good agreement with the recent simulation of the thermolysin[63] and carboxypeptidase A.[54] As pointed out by Christiansen and Lipscomb,[25] direct zinc coordination by the peptide carbonyl leads to a non-productive binding mode, which displaces the nucleophilic water. Our results provide further supporting evidence for this argument.

7.5 Conclusions

The lethal factor is a key target for drug development specific to anthrax attack. In this work, we report a hybrid quantum mechanical/molecular mechanical study on the catalysis of the enzyme. The QM/MM approach provided for the first time a complete and detailed picture of the mode of action. Our results suggest that the catalyzed reaction follows a typical nucleophilic substitution reaction mechanism. The nucleophilic addition of the zinc-activated water to the scissile carbonyl carbon leads to

a barrier featuring concurrent proton transfer to the general base Glu687. The resulting tetrahedral intermediate is stabilized by zinc coordination of the fractionally charged carbonyl oxygen and its hydrogen bond interaction with the hydroxyl group of Tyr728. The protonation of the leaving group nitrogen by the protonated carboxylate side chain of Glu687 eventually leads to the cleavage of the peptide bond.

Our model suggests that the Tyr728 residue, along with Zn(II), serves as the oxyanion hole in stabilizing the transition state, rather than protonating or stabilizing the nitrogen leaving group as suggested by previous authors. Its contribution was assessed by the Y728F mutant, which indicates a significant increase of the barrier height. On the other hand, the zinc cofactor is found to play no role in polarizing the scissile carbonyl. These mechanistic details are consistent with the established mode of action in several zinc peptidases such as thermolysin and carboxypeptidase A, underscoring the power of convergent evolution. These similarities derive apparently from the analogous active-site arrangement in these enzymes.

7.6 References

1 Bossi, P. *et al.* Bioterrorism: management of major biological agents. *Cellular and Molecular Life Sciences* **63**, 2196–2212 (2006).

2 Dixon, T. C., Meselson, M., Guillemin, J. & Hanna, P. C. Anthrax. *New England Journal of Medicine* **341**, 815-826 (1999).

3 Ascenzi, P. *et al.* Anthrax toxin: a tripartite lethal combination. *FEBS Letters* **531**, 384-388 (2002).

4 Klimpel, K. R., Arora, N. & Leppla, S. H. Anthrax toxin lethal factor contains a zinc metalloprotease consensus sequence which is required for letal toxin activity. *Molecular Microbiology* **13**, 1093-1100 (1994).

5 Duesbery, N. S. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734-737 (1998).

6 Vitale, G. *et al.* Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine-threonine phosphorylation of MAPKs in cutured macrophases. *Biochemical and Biophysical Research Communications* **248**, 706-711 (1998).

7 Pellecchia, M., Guidi-Rontani, C., Vitale, G., Mock, M. & Montecucco, C. Anthrax lethal factor cleaves MKK3 in macrophages and inhibits the LPS/IFNγ-induced release of NO and TNFα. . *FEBS Letters* **462**, 199-204 (1999).

8 Vitale, G., Bernardi, L., Napolitani, G., Mock, M. & Montecucco, C. Susceptibility of mitrogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor. *Biochemical Journal* **352**, 739-745 (2000).

9 Park, J. M., Greten, F. R., Li, Z.-W. & Karin, M. Macrophase apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048-2051 (2002).

10 Chopra, A. P., Boone, S. A., Liang, X. & Duesbery, N. S. Anthrax lethal factor proteolysis and inactivation of MAPK kinase. *Proceedings of the National Academy of Sciences of the United States of America* **278**, 9402-9406 (2003).

11 Turk, B. E. Manipulation of host signalling pathways by anthrax toxins. *Biochemical Journal* **402**, 405-417 (2007).

12 Turk, B. E. Discovery and development of anthrax lethal factor metalloproteinase inhibitors. *Current Pharmaceutical Biotechnology* **9**, 24-33 (2008).

13 Tonello, F., Seveso, M., Marin, O., Mock, M. & Montecucco, C. Screening inhibitors of anthrax lethal factor. *Nature* **418**, 386 (2002).

14 Turk, B. E. *et al.* The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nature Structural & Molecular Biology* **11**, 60-66 (2004).

15 Panchal, R. G. *et al.* Identification of small molecule inhibitors of anthrax lethal factor. *Nature Structural & Molecular Biology* **11**, 67-72 (2004).

16 Shoop, W. L. *et al.* Anthrax lethal factor inhibition. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7958-7963 (2005).

17 Forino, M. *et al.* Efficient synthetic inhibitors of anthrax lethal factor. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9499-9504 (2005).

18 Agrawal, A. *et al.* Thioamide hydroxypyrothiones supersede amide hydroxypyrothiones in potency against anthrax lethal factor. *Journal of Medicinal Chemistry* **52**, 1063-1074 (2009).

19 Tonello, F., Ascenzi, P. & Montecucco, C. The metalloproteolytic activity of the anthrax lethal factor is substrate-inhibited. *Proceedings of the National Academy of Sciences of the United States of America* **278**, 40075-40078 (2003).

20 Zakharova, M. Y. *et al.* Substrate recognition of anthrax lethal factor examined by combinatorial and pre-steady-state kinetic approaches. *The Journal of Biological Chemistry* **284**, 17902-17913 (2009).

21 Pannifer, A. D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414**, 229-233 (2001).

22 Hammond, S. E. & Hanna, P. C. Lethal factor active-site mutations affect catalytic avctivity in vitro. *Infection and Immunity* **66**, 2374-2378 (1998).

23 Tonello, F., Naletto, L., Romanello, V., Dal Molin, F. & Montecucco, C. Tyrosine-728 and gluamic acid-735 are essential for the metalloproteolytic activity of the lethal factor of Bacillus antracis. *Biochemical and Biophysical Research Communications* **313**, 496-502 (2004).

24 Matthews, B. W. Structural basis of the action of thermolysin and related zinc peptidases. *Accounts of Chemical Research* **21**, 333-340 (1988).

25 Christianson, D. W. & Lipscomb, W. N. Carboxypeptidase A. *Accounts of Chemical Research* **22**, 62-69 (1989).

26 Johnson, S. L. *et al.* Anthrax lethal factor protease inhibitors: Synthesis, SAR, and structure-based QSAR studies. *Journal of Medicinal Chemistry* **49**, 27-30 (2006).

27 Hong, R., Magistrato, A. & Carloni, P. Anthrax lethal factor investigated by molecular simulations. *Journal of Chemical Theory and Computation* **4**, 1745 (2008).

28 Dalkas, G. A., Papakyriakou, A., Vlamis-Gardikas, A. & Spyroulias, G. A. Insights into the anthrax lethal factor–substrate interaction and selectivity using docking and molecular dynamics simulations. *Protein Science* **18**, 1774-1785 (2009).

29 Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187-217 (1983).

30 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **79**, 926-935 (1983).

31 Brooks III, C. L. & Karplus, M. Deformabal stochastic boundaries in molecular dynamics. *Journal of Chemical Physics* **79**, 6312-6325 (1983).

32 Steinbach, P. J. & Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulations. *Journal of Computational Chemistry* **15**, 667 (1994).

33 Warshel, A. & Levitt, M. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103**, 227-249 (1976).

34 Gao, J. in *Reviews in Computational Chemistry* Vol. 7 (eds K. B. Lipkowitz & D. B. Boyd) 119-185 (VCH, 1996).

35 Monard, G. & Merz Jr., K. M. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Accounts of Chemical Research* **32**, 904-911 (1999).

36 Zhang, Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts* **116**, 43-50 (2006).

37 Riccardi, D. *et al.* Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *Journal of Physical Chemistry B* **110**, 6458-6469 (2006).

38 Hu, H. & Yang, W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annual Review of Physical Chemistry* **59**, 573-601 (2008).

39 Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie (International ed. in English)* **48**, 1198-1229 (2009).

40 Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review* **B58**, 7260-7268 (1998).

41 MacKerell Jr., A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).

42 Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T. & Karplus, M. A QM/MM implementation of the self consistent charge density functional tight binding (SCC-DFTB) method. *Journal of Physical Chemistry B* **105**, 569-585 (2001).

43 Sattelmeyer, K. W., Tirado-Rives, J. & Jorgensen, W. L. Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methos for organic molecules. *Journal of Physical Chemistry A* **110**, 13551-13559 (2006).

44 Otte, N., Scholten, M. & Thiel, W. Looking at self-consistent-charge density functional tight binding from a semiempirical perspective. *Journal of Physical Chemistry A* **111**, 5751-5755 (2007).

45 Elstner, M. *et al.* Modeling zinc in biomolecules with the self consistent charge density functional tight binding (SCC-DFTB) method: Applications to structure and energetic analysis. *Journal of Computational Chemistry* **24**, 565-581 (2003).

46 Riccardi, D. & Cui, Q. pKa analysis for the zinc-bound water in human carbonic anhydrase II: benchmark for "multi-scale" QM/MM simulations and mechanistic implications. *Journal of Physical Chemistry A* **111**, 5703-5711 (2007).

47 Riccardi, D., Konig, P., Guo, H. & Cui, Q. Proton transfer in carbonic anhydrase is controlled by electrostatics rather than the orientation of the acceptor. *Biochemistry* **47**, 2369-2378 (2008).

48 Xu, D., Zhou, Y., Xie, D. & Guo, H. Antibiotic binding to monozinc CphA b-lactamase from Aeromonas hydropila: quantum mechanical/molecular mechanical and density functional theory studies. *Journal of Medicinal Chemistry* **48**, 6679-6689 (2005).

49 Xu, D., Xie, D. & Guo, H. Catalytic mechanism of class B2 metallo-b-lactamase. *The Journal of Biological Chemistry* **281**, 8740-8747 (2006).

50 Xu, D., Guo, H. & Cui, Q. Antibiotic binding to dizinc b-lactamase L1 from Stenotrophomonas maltophilia: SCC-DFTB/CHARMM and DFT studies. *Journal of Physical Chemistry A* **111**, 5630-5636 (2007).

51 Xu, D., Guo, H. & Cui, Q. Antibiotic deactivation by dizinc b-lactamase: mechanistic insights from QM/MM and DFT studies. *Journal of the American Chemical Society* **129**, 10814 (2007).

52 Wang, C. & Guo, H. Inhibitor binding by metallo-b-lactamase IMP-1 from Pseudomonas aeruginosa: Quantum mechanical/molecular mechanical simulations. *Journal of Physical Chemistry B* **111**, 9986-9992 (2007).

53 Guo, H., Rao, N., Xu, Q. & Guo, H. Origin of tight binding of a near-perfect transition-state analogue by cytidine deaminase: Implications for enzyme catalysis. *Journal of the American Chemical Society* **127**, 3191-3197 (2005).

54 Xu, D. & Guo, H. Quantum mechanical/molecular mechanical and density functional theory studies of a prototypical zinc peptidase (carboxypeptidase A) suggest a general acid-general base mechanism. *Journal of the American Chemical Society* **131**, 9780-9788 (2009).

55 Amin, E. A. & Truhlar, D. G. Zn coordination chemistry: Development of benchmark suites for geometries, dipole moments, and bond dissociation energies and their use to test and validate density functionals and molecular orbital theory. *Journal of Chemical Theory and Computation* **4**, 75-85 (2008).

56 Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **11**, 700-733 (1990).

57 Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).

58 Torrie, G. M. & Valleau, J. P. Non-physical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).

59 Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free energy calculations on biomolecules. 1. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).

60 Zhang, X., Zhang, X. & Bruice, T. C. A definitive mechanism for chorismate mutase. *Biochemistry* **44**, 10443-10448 (2005).

61 Phillips, M. A., Fletterick, R. & Rutter, W. J. Arginine 127 stablilizes the transition state in carboxypeptidase. *The Journal of Biological Chemistry* **265**, 20692-20698 (1990).

62 Beaumont, A. *et al.* The role of Histine 231 in thermolysin-like enzymes. *The Journal of Biological Chemistry* **270**, 16803-16808 (1995).

63 Blumberger, J., Lamoureux, G. & Klein, M. L. Peptide hydrolysis in thermolysin: Ab initio QM/MM investigation of the Glu143-assisted water addition mechanism. *Journal of Chemical Theory and Computation* **3**, 1837-1850 (2007).

64 Lipscomb, W. M. & Strater, N. Recent advances in zinc enzymology. *Chemical Reviews* **96**, 2375-2433 (1996).

65 Corminboeuf, C., Hu, P., Tuckerman, M. E. & Zhang, Y. Unexpected deacetylation mechanism suggested by a density functional theory QM/MM sudy of histone-deacetylase-like protein. *Journal of the American Chemical Society* **128**, 4530-4531 (2006).

66 Momb, J. *et al.* Mechanism of the quorum-quenching lactonase (AiiA) from Bacillus thuringiensis: 2. Substrate modeling and active site mutations. *Biochemistry* **47**, 7715-7725 (2008).

Chapter 8

Initial steps in methanol steam reforming on PdZn and ZnO surfaces

Plane-wave DFT studies

"God made the bulk; surfaces were invented by the devil."

-Wolfgang Pauli

This chapter is based on collaborative work originally found in the following publication:

Adapted with permission from **Smith GK**, Sen L, Wenzhen L, Abhaya D, Daiqian X, Hua G. Initial steps in methanol steam reforming on PdZn and ZnO surfaces: Density functional theory studies. *Surface Science*. 2011; 605(7-8): 750-759.

8.1 Introduction

Proton exchange membrane (PEM) fuel cells are an efficient, clean, and robust device for generating electricity for mobile applications, but storage of hydrogen (H_2) required for their operation remains a significant challenge.[1,2] As a result, there has been much interest in circumventing hydrogen storage by on-board hydrogen generation, and one of the leading candidates for that is methanol steam reforming (MSR):[3]

$$CH_3OH + H_2O \xrightarrow{Catalyst} 3H_2 + CO_2$$

Methanol has many intrinsic advantages as a hydrogen carrier, including a high H/C ratio, low sulfur content, and storage requirements comparable to those for existing liquid fuels.[3,4] A summary of fuel sources and schematic of a PEM fuel cell is given in Fig. 8.1.

species	formula	H/C ratio	density (g/L)
hydrogen	H_2	-	0.09
hydrogen (2000 psi)	H_2	-	10
methane	CH_4	4	0.72
methanol	CH_3OH	4	792
ethanol	CH_3CH_2OH	3	789

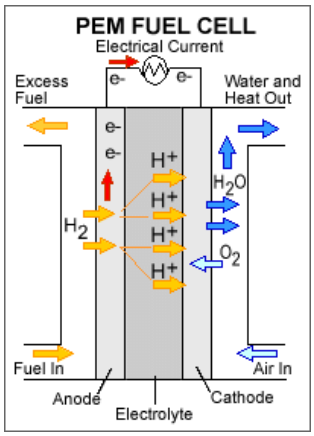


Figure 8.1. (Left Panel) H/C ratios and densities across different hydrogen storage or hydrogen carrier possibilities. (Right Panel) Schematic image showing the main parts of a PEM fuel cell.

Public domain image is courtesy of DOE's Office of Energy Efficiency and Renewable Energy, and the original can be found here. http://www1.eere.energy.gov/hydrogenandfuelcells/fuelcells/fc_types.html

The MSR reaction has traditionally used a Cu/ZnO catalyst,[5] which selectively produces CO_2 rather than CO; an important aspect as PEM fuel cell catalyst anodes[6], which typically consist of platinum supported by high surface area carbon, are poisoned by CO.[5] Despite its activity and selectivity, the Cu/ZnO catalyst has many drawbacks for mobile applications, including pyrophoricity, rapid degradation, and agglomeration through sintering.[3,5,7]

Recent interest has centered on a Pd/ZnO catalyst, first identified by Iwasa and colleagues.[8-12] The new catalyst shows similar activity and selectivity to Cu/ZnO, but with vastly improved thermal stability, a key component for mobile applications. This result was surprising as Pd is known to selectively produce CO and H_2 under steam reforming conditions.[13] The difference has been attributed to PdZn alloy, which forms when heating Pd/ZnO under reductive conditions.[9] PdZn alloy possesses a similar valence density of states to that of Cu, as shown by density functional theory (DFT) calculations,[14] and by XPS valence band spectra.[15]

This promising Pd/ZnO catalyst has stimulated many recent experimental studies.[16-28] For example, Karim *et al.* have advanced a low-temperature method to make PdZn particles with controlled sizes.[20] On the other hand, Vohs and coworkers have focused on the structure, composition, and catalysis of PdZn formed by modifying the Pd(111) surface with Zn under ultra-high vacuum conditions. The PdZn alloy was shown to form ordered structures on Pd(111), as evidenced by LEED patterns.[21,25] TPD experiments of CH_3OH on this PdZn surface indicated two peaks near 140 and 170 K, which have been

assigned to desorption of physisorbed CH_3OH and recombinative desorption from adsorbed CH_3O and H, respectively.[23] However, it was found that the incorporation of Zn to Pd(111) significantly increases the barrier for the dehydrogenation of CH_3O and CH_2O,[23] which eventually leads to the formation of CO. This observation provides a plausible explanation of the selectivity of the PdZn catalyst towards CO_2. More recently, Rameshan *et al.* presented evidence that a near-surface PdZn alloy, rather than a single layer of PdZn, is responsible for the selectivity in MSR towards CO_2.[28]

The role of the ZnO support in MSR has also been investigated. ZnO has two types of surfaces. The non-polar surfaces, such as ($00\bar{1}0$) consisting of rows of slightly tilted ZnO "dimers", are electrostatically stable,[29] but non-reactive towards methanol dissociation.[30] There are two types of polar surfaces: the (0001) surface is terminated by Zn, while the ($000\bar{1}$) surface is terminated by O. Earlier experimental studies have indicated that methanol dissociates readily on the Zn-terminated (0001) surface, but not on the O-terminated ($000\bar{1}$) surface.[30,31] The former, which will be investigated here, was found recently to possess single-layer island/pit formation with excess oxygen ions,[32,33] which presumably is responsible for its stabilization.[34] When Pd is deposited on ZnO surfaces, it forms the PdZn alloy by incorporating Zn from the substrate and aggregates at step edges of ZnO, which prevents the oxidation of methanol.[22] It was also found that the Pd adsorbed (0001) and ($000\bar{1}$) surfaces of ZnO behave differently when it comes to the dehydrogenation of CH_3OH. The former is significantly more active, which has been attributed to the synergism between the metal and support.[27]

Despite these advances, the mechanism of MSR is still not established. It has been proposed that the reaction starts with the dissociation of both CH_3OH and H_2O on PdZn, leading to the formation of methoxy (CH_3O), which then proceeds to form formaldehyde (CH_2O) before reacting further with adsorbed OH to produce the H_2 and CO_2 products.[11] Additional competing processes may be occurring on the surface as well, including methanol decomposition ($CH_3OH \rightarrow 2H_2 + CO$), partial oxidation ($CH_3OH + ½\ O_2 \rightarrow 2H_2 + CO_2$) and the water-gas shift reaction ($CO + H_2O \rightarrow H_2 + CO_2$).[5]

Definitive mechanistic action is often difficult to establish through experiment alone and so theoretical studies have become increasingly valuable. An elegant series of theoretical studies on PdZn catalyzed methanol decomposition has been performed by Rösch's group,[21,35-40] using a plane-wave DFT method. Their studies showed that the decomposition pathway of adsorbed CH_3O to CO has significantly higher activation barriers on Cu(111) and PdZn(111) than on Pd(111),[40] which was confirmed by a recent high-vacuum study of Jeroro and Vohs.[23] Since the dehydrogenation of methoxy has been assumed to be the rate-limiting process[11], there has been no study on the initial dissociation steps on PdZn surfaces. In this work, we investigate the initial dissociation steps of both methanol and water and show that they are highly activated on flat PdZn surfaces.

It is well known that some heterogeneous catalytic reactions are promoted by unsaturated active sites on defect surfaces.[41,42] To explore this possibility, we have

examined the dissociation reactions on several stepped PdZn surfaces. A collection of possible surface sites is shown below in Fig. 8.2.

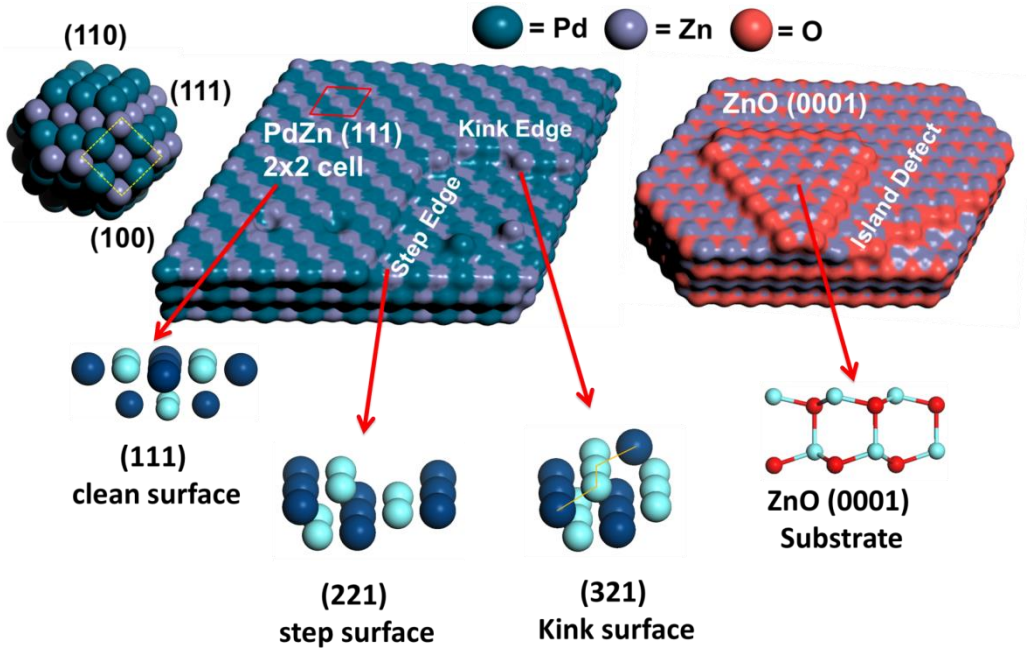


Figure 8.2 – Surface sites explored in this work for PdZn and ZnO.

As discussed below, the unsaturated sites on stepped PdZn surfaces significantly lower the dissociation barriers, which might be responsible for the observed low-temperature dissociation of CH_3OH on PdZn.[23] In addition, calculations have also been performed on the ZnO(0001) surface. The support is known to play an important role in some heterogeneous catalytic reactions,[43,44] and the ZnO(0001) surface has been shown to be highly reactive towards methanol decomposition.[30,31] Indeed, the dissociation of both CH_3OH and H_2O is found to be much more facile on the ZnO surface than on PdZn surfaces. The results reported here have important implications for the reaction mechanism of MSR. This Chapter is organized as follows. The theoretical method is discussed in Sec. 8.2 The results are presented in Sec. 8.3. A discussion on the role of the

initial dissociation steps in MSR is given in Sec. 8.4. Finally, a summary is given in Sec. 8.5.

8.2 Theory

The methods used in this project are described in a condensed fashion below. Additional details on the techniques used here can be found in Chapter 4 (Plane-wave DFT).

All DFT calculations were performed using the Vienna ab initio simulation package (VASP).[45-47] Energies and geometries were calculated using the Perdew-Wang 91 (PW91)[48] GGA approximation[49] (GGA) of the exchange-correlation functional. The core electrons were treated using the projector augmented-wave (PAW) method.[50,51] The Brillouin zone was sampled using either a 7×7×1 or 5×5×1 Monkhorst-Pack *k*-point grid[52] with Methfessel-Paxton smearing[53] of 0.1 eV. The plane wave basis cut off was 400 eV for PdZn and 440 eV for ZnO.

1:1 PdZn alloy adopts a AuCu ($L1_0$-type) configuration in space group P4/mmm. Bulk optimization yielded lattice parameters of a=b=4.139 Å, c=3.378 Å, in good agreement with previously reported results.[14] Slab models for the (111) and (100) surfaces consisted of five layers of a 2×2 unit cell, with the top two layers allowed to relax in all calculations. (The small unit cell used here is motivated by computational efficiency, and it might result in errors for low coverage adsorption.) To simulate steps and kinks, the (221), (110) and (321) surfaces were used. For the (221) and (321) surfaces, five stepped layers were included with the uppermost three layers relaxed, while for the (110)

surface, a total of five layers was used with the top two layers relaxed. The vacuum spacing between slabs was 14 Å.

For the ZnO surface calculations, a similar protocol was used. The optimization of the hexagonal wurtzite ZnO crystal yielded lattice parameters of a=b=3.281 Å, c=5.256 Å, in good agreement with previous work.[29] The model of the defect-free polar ZnO(0001) surface consists of five double layers for a total of twenty oxygen and twenty zinc ions within a 2×2 unit cell. The two top layers were fully relaxed during the calculations. The large number of layers is necessary because of the polar nature of the surface. In addition, the leading errors due to the artificial dipole generated by the slab model were corrected using the methods introduced in Refs.[54] and[55]. The defect ZnO(0001) surface was modeled with a larger 3×3 unit cell which contains eight double layers plus a four-atom island on the top. In this case, both the island and two layers underneath were relaxed. These calculations were performed with a 5×5×1 Monkhorst-Pack k-point mesh. The vacuum space is 18 Å for the defect-free surface and 16 Å for the defect surface.

The adsorption energy was calculated as follows: E_{ad} = E(adsorbate+surface) – E(free molecule) – E(free surface). Transition states were calculated using the climbing image nudged elastic band (CI-NEB) method[56,57] with energy (10^{-4} eV) and force (0.05 eV/Å) convergence criteria. Stationary and transition points were confirmed by normal mode analysis using a finite differencing method with an atomic displacement of 0.02 Å. The normal frequencies were used to compute vibrational zero-point energy (ZPE)

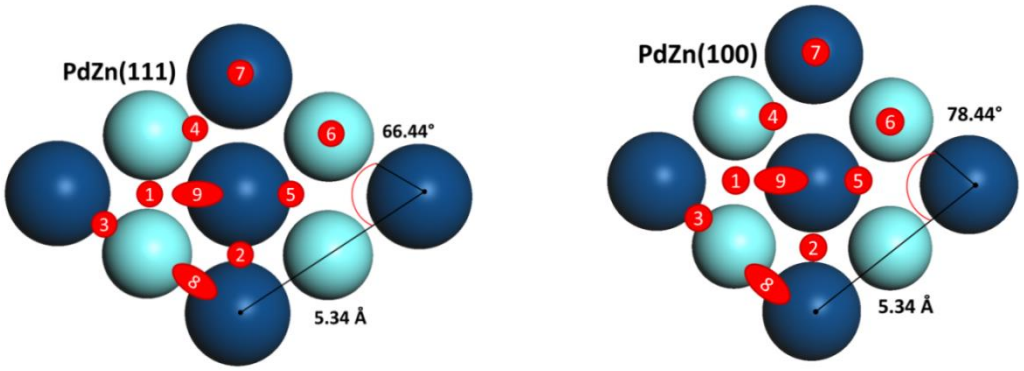
corrections. All reported energies are ZPE corrected with the exception of Tables 8.1 and 8.2.

8.3 Results

8.3.1 Adsorptions on PdZn (111) and (100) surfaces

Extensive testing of convergence has been carried out on the (111) and (100) faces of PdZn, which represent the most prevalent defect-free surfaces of the alloy.[14] Definitions of the surface sites on (111) and (100) surfaces are given in Figure 8.3.

Figure 8.3 – Surface sites for the (111) and (100) surfaces.



1. B^{Zn} 2. B^{Pd} 3. B^{PdZn} 4. H^{ZnPd2} 5. H^{PdZn2} 6. T^{Zn} 7. T^{Pd} 8. tB^{PdZn} 9. tB^{Zn}

To verify our models, adsorption energies and geometries were calculated for several species involved in the initial dissociation of CH_3OH and H_2O. Only minor differences were observed compared to previous studies.[40] For example, our adsorption energies for CH_3O and OH at PdZn(111) are −2.29 and −3.05 eV, which can be compared with the

literature values of −2.30 and −3.14 eV.[40] These differences are expected as our systems have one additional atomic layer and a denser *k*-point mesh (7×7×1). Figure 1 defines all the investigated sites for these surfaces. The energetics (no ZPE correction) and geometric parameters for major adsorption conformations are collected in Table 8.1 and 8.2.

8.3.1.1 CH_3OH

CH_3OH shows weak adsorption at all surfaces sites. On PdZn(111), the strongest adsorption is at a zinc top site (T^{Zn}) with the C-O bond deviating by 47.7^o from surface normal, and an adsorption energy of −0.32 eV. The next most stable conformation was a bridge site (B^{PdZn}) with the C-O bond closer to surface normal at 7.0^o. The more open surface PdZn(100) showed similar adsorption energies, with the strongest adsorption at the B^{PdZn} site with a C-O bond angled 30.9^o from surface normal and an adsorption energy of −0.21 eV. These adsorption energies compare reasonably well with that derived using the Redhead formula[58] from a recent TPD experiment (0.37 eV).[23]

8.3.1.2 H_2O

Similar to methanol, H_2O also weakly adsorbs on these PdZn surfaces. The strongest adsorptions on PdZn(111) are at the B^{PdZn} and T^{Zn} sites, with adsorption energies of −0.25 and −0.24 eV, respectively. In contrast to methanol, water tends to assume a roughly parallel orientation with the surface, with moderate asymmetries. For example, the O-H bonds at the B^{PdZn} site with respect to surface normal were 114.7^o and 75.7^o,

respectively. Water adsorption on PdZn(100) was strongest at the T^{Zn} and H^{ZnPd2} sites, with energies of −0.23 and −0.19 eV, respectively.

Table 8.1. Adsorption energies (eV) and geometric parameters of the most stable orientations on PdZn (111). ZPE correction is not included.

PdZn (111)									
surface parameters			**distances (Å)**					**angles**	
Species	**Site**	E_{ad} **(eV)**	**O-Zn / H-Zn**	**O-Pd / C-Pd**	**H-Pd**	**O-C**	**O-H**	**∠COH/ ∠HOH**	**∠n(CO)/ ∠n(OH)**
CH_3OH	T^{Zn}	−0.32	2.17	3.07, 3.44	2.40, 3.03	1.44	0.99	109.6	47.7
	B^{PdZn}	−0.18	2.48	3.06	2.75, 2.72	1.44	0.99	111.7	7.0
CH_3O	H^{PdZn2}	−2.30	2.05, 2.05	3.05, 2.71	-	1.43	-	-	8.5
	B^{Zn}	−2.28	2.05, 2.05	3.02, 2.88	-	1.43	-	-	2.0
CH_3	T^{Pd}	−1.38	-	2.15	-	-	-	-	-
CH_2O	tB^{PdZn}	−0.21	2.01	2.26	2.67, 2.72	1.30	-	-	97.4
	tB^{Zn}	−0.13	3.26, 3.17	3.18	3.30, 3.24	1.23			106.5
CH_2OH	tB^{PdZn}	−1.44	2.28	2.13	-	1.47	0.98	109.2	81.8
	tB^{Zn}	−1.36	2.63, 2.64	2.12	-	1.46	0.98	110.5	84.6
H_2O	B^{PdZn}	−0.25	2.48	3.23	2.73, 3.17	-	0.98, 0.99	103.4	114.7, 75.7
	T^{Zn}	−0.24	2.32	-	3.32, 3.39	-	0.98, 0.98	104.4	99.3, 75.2
OH	B^{Zn}	−3.14	2.05, 2.01	2.97, 2.86	-	-	0.98	-	15.2
	B^{PdZn}	−2.93	1.96	2.30	-	-	0.98	-	48.3
H	H^{ZnPd2}	−2.50	1.56	-	1.88, 1.88	-	-	-	-
	B^{Pd}	−2.45	2.25, 2.40	-	1.70, 1.70	-	-	-	-

Table 8.2 – Adsorption energies (eV) and geometric parameters of the most stable orientations on PdZn (100). ZPE correction is not included.

PdZn (100)									
Species	**Site**	E_{ad} **(eV)**	**O-Zn / H-Zn**	**O-Pd / C-Pd**	**H-Pd**	**O-C**	**O-H**	**∠COH/ ∠HOH**	**∠n(CO)/ ∠n(OH)**
CH_3OH	B^{PdZn}	−0.21	2.66	3.4	2.69	1.44	0.98	109.1	30.9
	T^{Zn}	−0.19	2.39	3.16, 3.66	2.45, 3.58	1.44	0.99	110.7	20.0
CH_3O	B^{Zn}	−2.45	2.08, 2.09	2.65, 2.63	-	1.44	-	-	0.1
	H^{PdZn2}	−2.45	2.11, 2.13	2.35, 3.15	-	1.44	-	-	18.3
CH_3	T^{Pd}	−1.50	-	2.15	-	-	-	-	-
H_2O	T^{Zn}	−0.23	2.36	-	3.05, 3.05	-	0.98, 0.98	104.6	87.5, 88.3
	H^{ZnPd2}	−0.19	2.53	-	2.77, 3.79	-	0.98, 0.98	104.9	70.6, 102.2
CH_2O	tB^{Zn}	−0.29	2.47, 2.80	2.24	2.74, 2.74	1.29	-	-	105.9
	tB^{PdZn}	−0.27	2.08	2.24	2.59, 2.72	1.30	-	-	95.8
CH_2OH	tB^{PdZn}	−1.54	2.24	2.13	-	1.48	0.98	110.3	85.6
	tB^{Zn}	−1.44	2.82, 2.81	2.13	-	1.46	0.98	109.4	82.8
OH	B^{Zn}	−3.35	2.08, 2.08	2.63, 2.63	-	-	0.98	-	0
	H^{PdZn2}	−3.35	2.10, 2.10	2.37, 3.09	-	-	0.97	-	16.1
H	H^{ZnPd2}	−2.46	2.03, 2.79	-	1.85, 1.85	-	-	-	-
	B^{Pd}	−2.45	2.42, 2.42	-	1.82, 1.82	-	-	-	-

8.3.1.3 Other species

The adsorption of CH_3O, CH_2OH, CH_2O, CH_3, OH, and H species was also investigated. CH_3O and OH both interact strongly with the PdZn(111) and PdZn(100) surfaces, preferring Zn-rich 3-fold (H^{PdZn2}) and 2-fold (B^{Zn}) sites. The preference of CH_3O to interact with as many Zn sites as possible has been noted in a previous DFT study.[36] Methyl preferentially adsorbs at the T^{Pd} site on both surfaces, though weaker binding

sites were observed at B^{Pd} and H^{ZnPd2}. Hydrogen adsorbs strongly at Pd-rich 3-fold (H^{ZnPd2}) and 2-fold (B^{Pd}) sites on both surfaces. A weaker but stable hydrogen adsorption was also observed on T^{Pd} site with energies of −2.15 eV and −2.22 eV for the (111) and (100) faces of PdZn. Hydroxy methyl (CH_2OH) interacts strongly with both surfaces, adopting a flat conformation with the methylene carbon near a Pd and the oxygen near either one Zn atom along the B^{PdZn} site (refered to as tB^{PdZn}) or bisecting two Zn atoms in a B^{Zn} conformation (refered to as tB^{Zn}). The hydroxyl hydrogen orients away from the surface in both conformations. Formaldehyde (CH_2O) adsorbs weakly in a flat conformation on both surfaces in the tB^{PdZn} and tB^{Zn} conformations, similar to hydroxyl methyl.

8.3.2 Reaction paths on PdZn (111) and (100) surfaces

To explore the dissociation of selected adsorbates, reaction paths were determined using the CI-NEB method, which connect the initial states (IS) to the final states (FS) via the transition states (TS). Reaction barrier information can be found in Table 8.3.

The forward and reverse barriers and the corresponding thermodynamic parameters are listed by referencing against the lowest reactant (IS*) and product (FS*) energies, respectively. The lowest reactant states were located by an exhaustive search of all adsorption sites. On the other hand, the lowest energy product states were obtained by optimizing the co-adsorption of product species at adsorption sites that were found to be strongest in isolation. If co-adsorption at the strongest sites was not viable for both species due to unit cell constraints, the second strongest adsorption site was used. In

these cases, the difference in isolation between strongest and second strongest adsorption was 0.03 eV or less. These parameters were used to make conclusions on the several possible (O-H, C-H, and C-O) dissociation routes for CH_3OH. The diffusion barriers between these states are not reported, but are assumed to be lower than bond breaking.

8.3.2.1 CH_3OH bond scissions

On the (111) surface, the O-H scission of CH_3OH starts with an initial state at H^{PdZn2} and proceeds to two bridge sites, CH_3O at B^{Zn} and H at B^{Pd}. This path showed a ZPE-corrected barrier of 0.88 eV and ΔE of 0.23 eV. On the (100) surface, the path for O-H scission traveled from a 2-fold bridge (B^{PdZn}) initial state to 3-fold hollow sites for CH_3O (H^{PdZn2}) and H (H^{ZnPd2}). This path showed a reaction barrier of 0.73 eV and ΔE of 0.15 eV, similar to those on PdZn(111). Stationary point geometries for these reaction paths are displayed in Figure 8.5.

The O-C and C-H scission processes were also explored. The O-C scission on the (111) surface proceeded from a T^{Zn} initial state to a top site for CH_3 (T^{Pd}) and a bridge site for OH (B^{Zn}), with a 1.39 eV barrier. On the (100) surface, an initial state at tB^{Zn} proceeded to a top site for CH_3 (T^{Pd}) and a bridge site for OH (B^{Zn}) with a 1.54 eV barrier. The C-H scission showed a barrier of 1.21 eV on PdZn(111) and 1.05 eV on PdZn(100), but the product states were significantly more endothermic. Based on these results, we conclude that the O-H scission is the dominant pathway for initial dissociation of CH_3OH on clean surfaces.

8.3.2.2 CH_3O C-H scission

The presumed rate-limiting step, namely the dehydrogenation of methoxy, was also explored. On the (111) surface, methoxy had an initial state of B^{Zn}, which proceeded to final states of tB^{Zn} for formaldehyde and B^{Pd} for hydrogen. The ZPE-corrected reaction barrier was 0.89 eV, which is only 0.01 eV higher than that for the O-H scission of methanol on the same surface. On the (100) surface, methoxy had an initial state of H^{PdZn2}, which proceeded to final states of tB^{PdZn} for formaldehyde and an adjacent H^{PdZn2} site for hydrogen. The reaction has a barrier of 1.06 eV, which is 0.33 eV higher than that for O-H scission of methanol on the same surface. These barrier heights compare reasonably well with those reported in earlier theoretical studies:[35] 0.96 eV (111) and 0.93 eV (100). Our calculated ΔE of 0.64 eV (111) and 0.78 eV (100) are also consistent with the literature values of 0.63 eV (111) and 0.92 eV (100).[35]

8.3.2.3 H_2O O-H scission

As shown in Table 8.3 and Figure 8.7, H_2O has similar energetics for O-H scission as CH_3OH. The dissociation reaction proceeds from an initial adsorption at the T^{Zn} site on the (111) surface to 2 bridge sites for the OH (B^{Zn}) and H (B^{Pd}) products. On the (100) surface, the initial state was a bridge site (B^{Zn}), and products adsorb at the B^{Zn} (OH) and H^{PdZn2} (H) sites. The dissociation barriers are 0.92 eV on the (111) surface and 0.74 eV on the (100) surface, with respective ΔE of 0.21 eV and 0.24 eV. The geometries of the stationary states are displayed in Figure 8.6.

Table 8.3 – Reaction barriers (eV) for various pathways on the all the explored surfaces. Barriers and thermodynamic parameters are referenced against the lowest energy reactant and product states.

Pathway	Surface	Forward Barrier	Reverse Barrier	ΔE
$CH_3OH^* \rightarrow CH_3O^* + H^*$	PdZn (111)	0.88	0.65	0.23
	PdZn (100)	0.73	0.59	0.15
	PdZn (221)^{Zn}	0.59	0.97	−0.37
	PdZn (110)^{Zn}	0.57	0.81	−0.24
	PdZn (321)	0.54	0.94	−0.40
	ZnO (0001)	0.39	0.77	−0.38
$CH_3OH^* \rightarrow CH_3^* + OH^*$	PdZn (111)	1.39	-	-
	PdZn (100)	1.54	-	-
$CH_3OH^* \rightarrow CH_2OH^* + H^*$	PdZn (111)	1.21	-	-
	PdZn (100)	1.05	-	-
$CH_3O^* \rightarrow CH_2O^* + H^*$	PdZn (111)	0.89	0.25	0.64
	PdZn (100)	1.06	0.28	0.78
$H_2O^* \rightarrow OH^* + H^*$	PdZn (111)	0.92	0.71	0.21
	PdZn (100)	0.74	0.51	0.24
	PdZn (221)^{Zn}	0.61	1.03	−0.42
	ZnO (0001)	0.42	0.85	−0.43

8.3.3 Reaction paths on defect PdZn surfaces (221), (110), and (321)

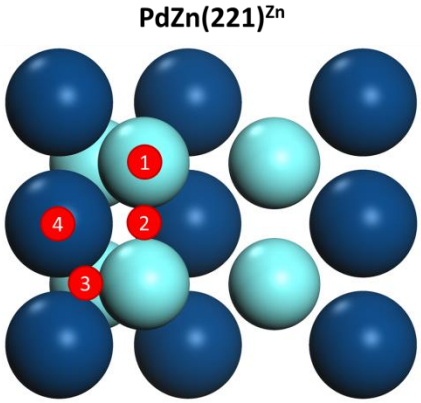
To understand the impact of defect sites on the dissociation processes, we have investigated these reactions on non-flat surfaces of PdZn. We emphasize here that the surface characteristics of the real catalyst might have quite different defects from the ones investigated here. In fact, the composition and morphology of the active form of the catalyst are still not clearly known. Thus, the studies reported here are meant to provide a sampling of various possible undercoordinated sites and their impact on the chemical processes. Specifically, step defects are modeled by the (221) and (110) faces of PdZn, while a kink defect by the (321) face.

In terms of notation, step edge sites are denoted by the letter S (ST^{Zn}, SB^{Zn}, etc.) for (221) and (110) surfaces. Similarly, kink edges are denoted by the letter K for the (321) surface to distinguish between the defect site and the terrace directly adjacent to it. Each of these surfaces is distinguished by unsaturated Pd or Zn atoms with lower bond coordination at the defect interface. The most stable adsorption sites for isolated species are reported in Table 8.4 without ZPE correction.

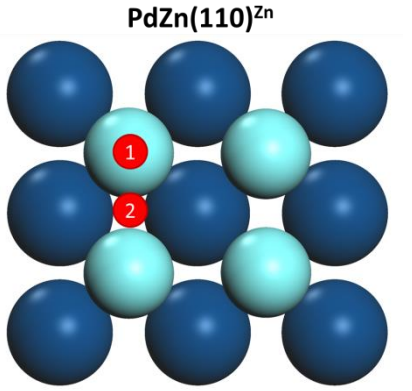
Since the C-O and C-H scissions of methanol have been found to have significantly higher barriers on the flat PdZn surfaces, we will focus here only on the O-H scission on the step and kink surfaces for both CH_3OH and H_2O. Our convergence testing showed a less dense *k*-point mesh of (5×5×1) was needed to produce similar accuracy on the defect surfaces. Reaction energies with ZPE correction are given in Table 8.3 and in

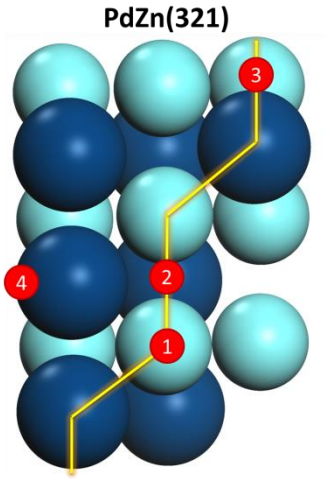
Figure 8.7. The structural details of IS, TS, and FS for representative reaction paths are shown in Figure 8.5 for CH_3OH and Figure 8.6 for H_2O.

Figure 8.4 – Surface sites for the $(221)^{Zn}$, $(110)^{Zn}$, (321) and ZnO(0001) surfaces.

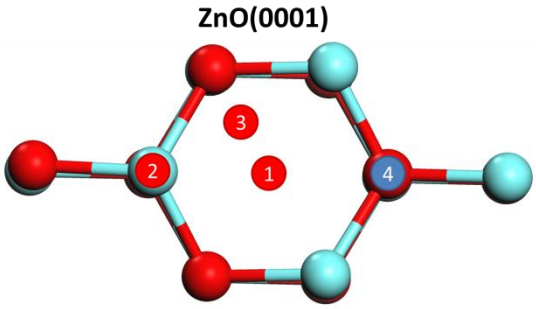


1. ST^{Zn} 2. SB^{Zn} 3. H^{ZnPd2} 4. T^{Pd}





1. KT^{Zn} 2. KB^{Zn} 3. KB^{Pd} 4. B^{Pd}



1. fcc 2. T^{Zn} 3. B^{Zn} 4. hcp

Table 8.4 – Adsorption energies (eV) of the most stable geometries on defect PdZn and ZnO surfaces. ZPE correction is not included.

Species	PdZn(221)^{Zn}		PdZn(110)^{Zn}		PdZn(321)		ZnO(0001)	
	Site	E_{ad} (eV)	Site	E_{ad} (eV)	Site	E_{ad} (eV)	Site	E_{ad} (eV)
CH_3OH	ST^{Zn}	−0.43	ST^{Zn}	−0.35	KT^{Zn}	−0.44	T^{Zn}	−0.40
CH_3O	SB^{Zn}	−2.79	SB^{Zn}	−3.05	KB^{Zn}	−2.72	fcc	−3.37
H_2O	ST^{Zn}	−0.34	ST^{Zn}	−0.28	KT^{Zn}	−0.35	T^{Zn}	−0.37
OH	SB^{Zn}	−3.73	SB^{Zn}	−3.95	KB^{Zn}	−3.66	fcc	−4.37
H	H^{ZnPd2}	−2.39	SB^{Zn}	−2.20	KB^{Pd}	−2.59	T^{Zn}	−2.74

8.3.3.1 PdZn(221)

The (221) surface can be constructed with both Zn and Pd atoms at the step edge, named (221)^{Zn} and (221)^{Pd} respectively. On the (221)^{Zn} surface, methanol adsorbs weakly at the ST^{Zn} site with an adsorption energy of −0.43 eV. An initial state for the dissociation of CH_3OH was found at a stepped Zn bridge site (SB^{Zn}) that proceeded to H at a top site (T^{Pd}) with the CH_3O remaining at initial bridge site (SB^{Zn}). This reaction had a barrier of 0.59 eV and ΔE of −0.37 eV, both significantly lower than those on the (111) and (100) surfaces. The (221)^{Pd} surface has also been studied, but it had a methanol O-H scission barrier of at least 0.75 eV, making its activity similar to the clean surfaces. Unsaturated Pd atoms at the step site do not appear to significantly lower the barrier for this reaction, in contrast to Zn.

Similar to methanol, water adsorbs weakly at the ST^{Zn} site with an adsorption energy of −0.34 eV. The dissociation of H_2O follows essentially the same path as CH_3OH, while the barrier (0.61 eV) and exothermicity (−0.42 eV) also showed significant decrease compared with flat surfaces.

8.3.3.2 PdZn(110)

The (110) surface exists in Zn and Pd top forms, named $(110)^{Zn}$ and $(110)^{Pd}$, respectively. On the $(110)^{Zn}$ surface, methanol adsorbs weakly at the STZn site (−0.35 eV). The reaction path proceeded from a single Zn bridge site (SB^{Zn}) to adjacent Zn bridge sites for CH_3O and H. In the transition state, the hydroxyl hydrogen does interact with the second layer Pd atoms, coming within 1.89 Å. The barrier for this reaction was 0.57 eV with ΔE of −0.24 eV.

On the $(110)^{Pd}$ surface (data not shown), CH_3OH proceeded from a single Pd bridge site (SB^{Pd}) to adjacent Pd bridge sites for CH_3O and H. This reaction had a barrier of at least 0.63 eV, and further suggests undercoordinated Pd does not stabilize these initial reaction steps as well as Zn.

8.3.3.3 PdZn(321)

The PdZn(321) surface adopts a saw tooth pattern with alternating Pd and Zn atoms at the vertices. Methanol and water both adsorb weakly at the KT^{Zn} site, with adsorption energies of −0.44 and −0.35 eV, respectively. Based on our experience with the (221) and (110) surfaces which showed unsaturated Pd atoms do not significantly lower

reaction barriers compared to Zn, we focused on Zn rich sites for exploring kink reactions paths. The (321) surface proved quite facile for methanol O-H scission, proceeding from an initial kink edge Zn bridge (KB^{Zn}) to a final state with CH_3O at the same Zn bridge and H moving to a sloped bridge site (B^{Pd}). This reaction had a barrier of 0.54 eV and ΔE of −0.40 eV.

Figure 8.5 – Geometries and key distances for initial, transition, and final states in methanol dehydrogenation on the PdZn (111), PdZn (221)^{Zn}, PdZn (321), and ZnO (0001) surfaces. Atom colors are Pd (dark blue), Zn (light blue), O (red), C (grey), and H (white).

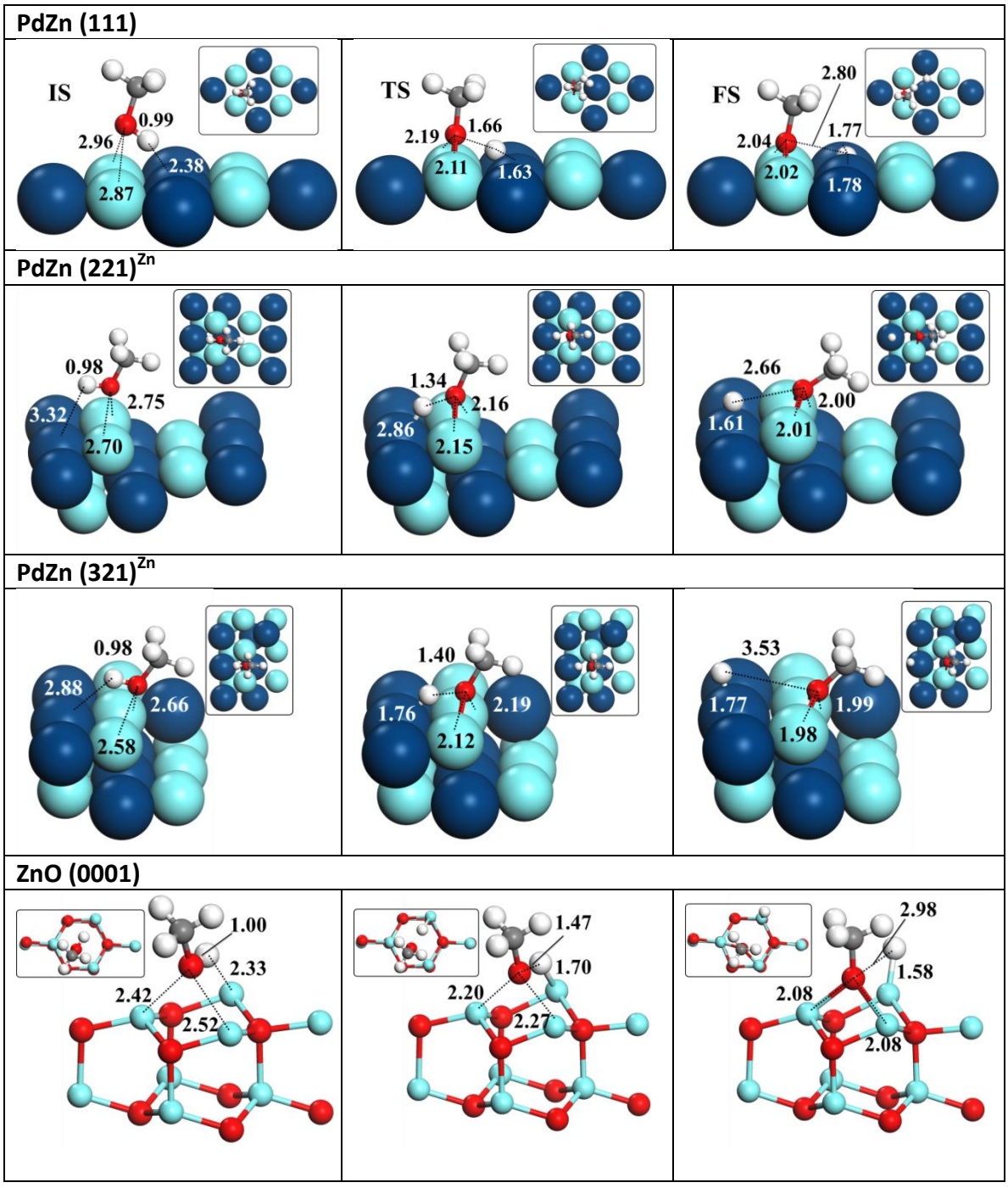


Figure 8.6 – Geometries and key distances for initial, transition, and final states in water dehydrogenation on the PdZn (111), PdZn (221)^{Zn}, and ZnO (0001) surfaces. Atom colors are Pd (dark blue), Zn (light blue), O (red), C (grey), and H (white).

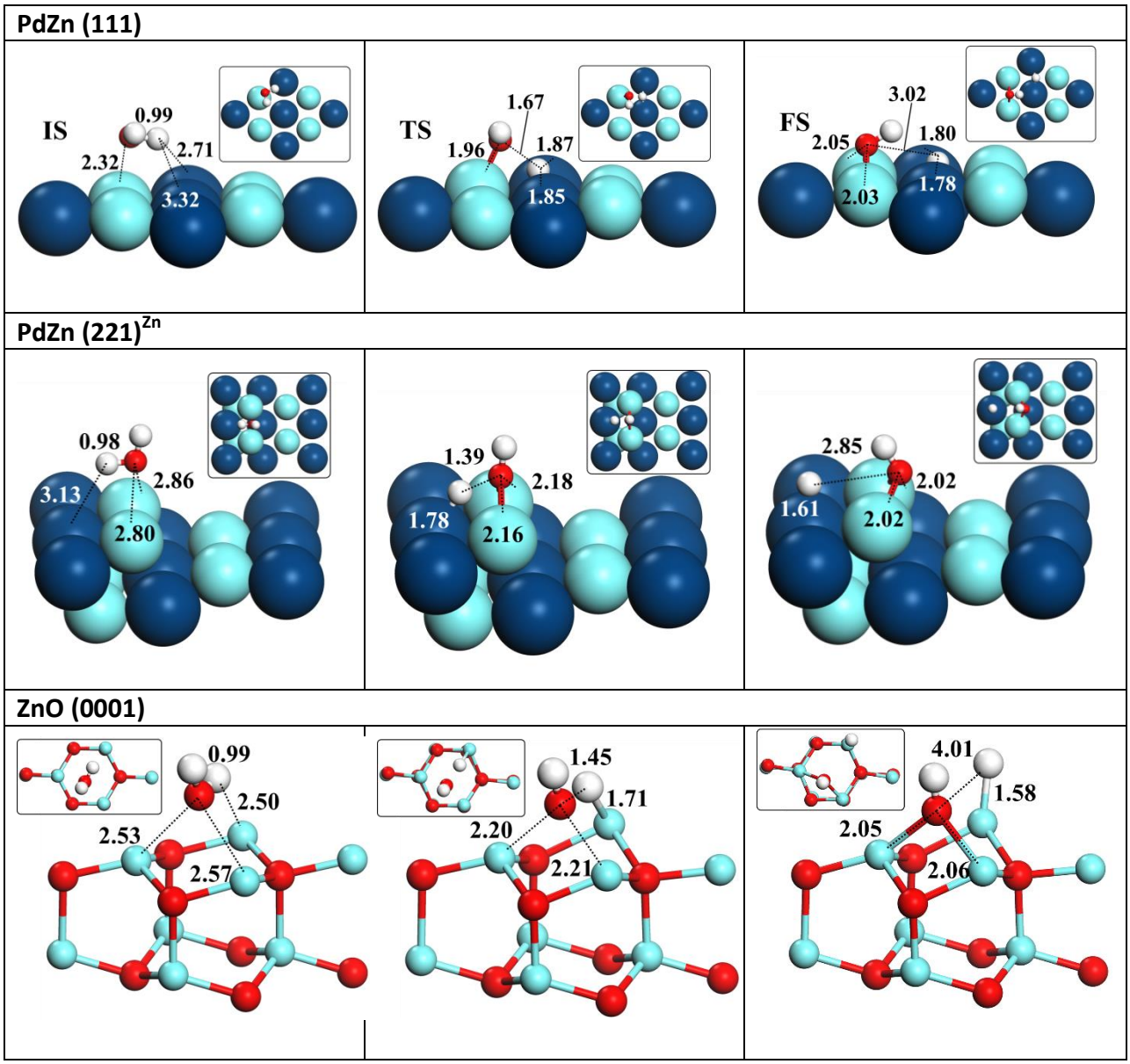
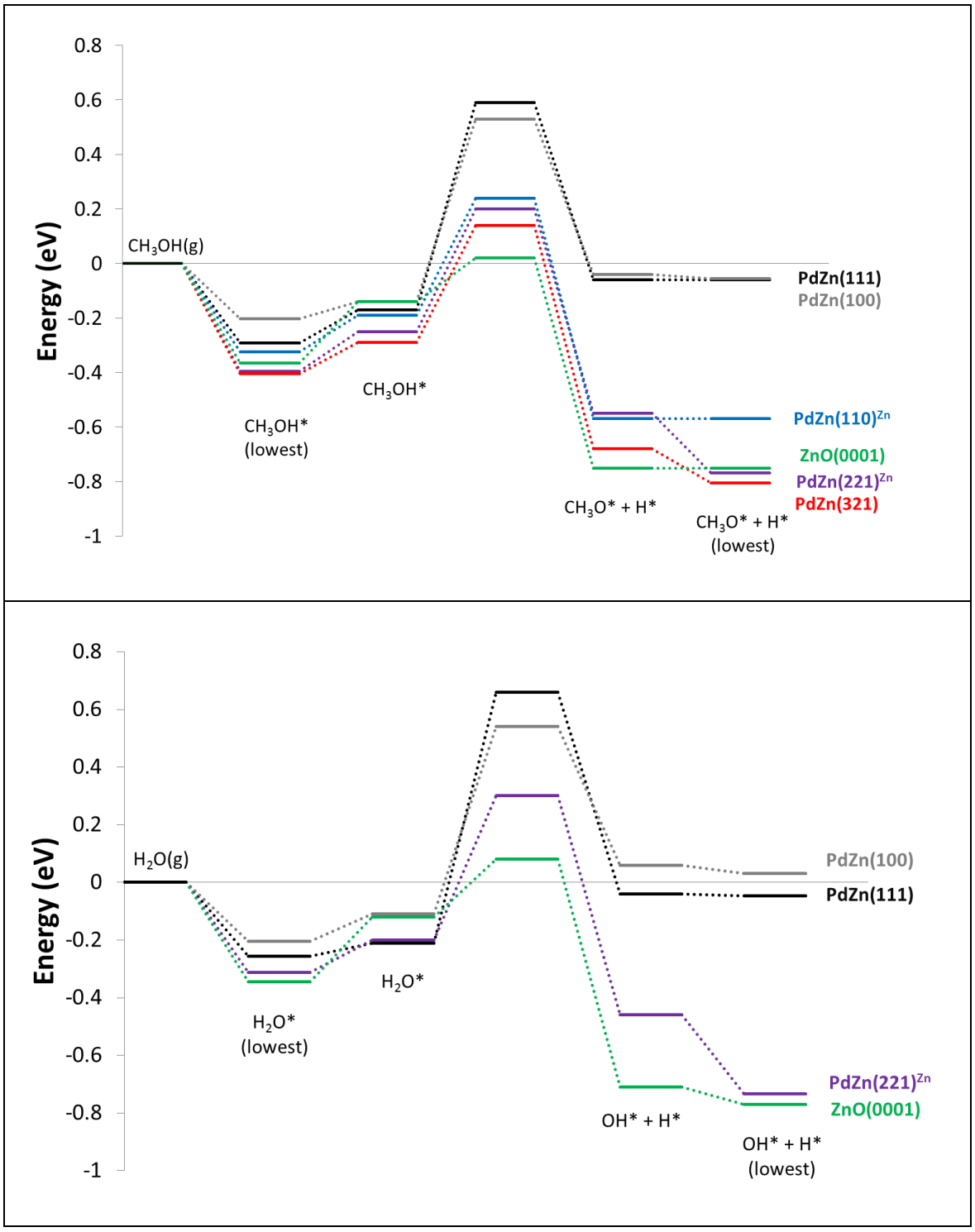


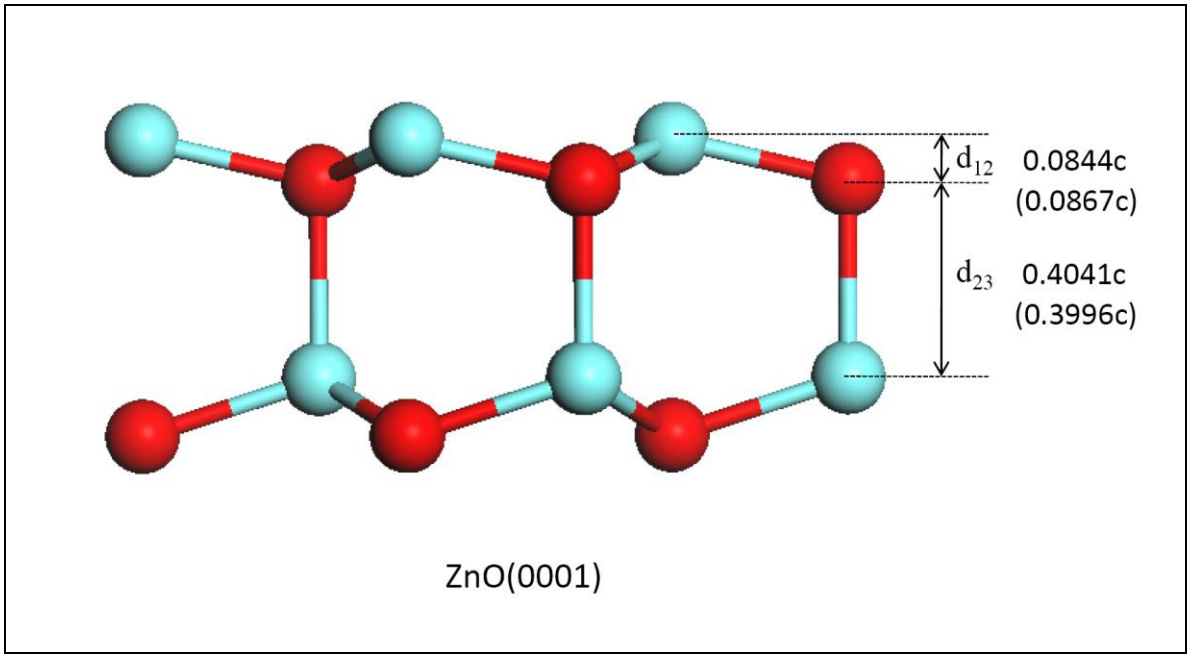
Figure 8.7 – Potential energy surfaces for methanol (top panel) and water (bottom panel) O-H scission on PdZn (100), PdZn (111), PdZn (110)^{Pd}, PdZn (110)^{Zn}, PdZn (221)^{Zn}, PdZn (321) and ZnO (0001). All ZPE corrected energies (eV) are referenced against the sum of energies for methanol in the gas phase and either clean PdZn or clean ZnO surfaces.



8.3.4 Reaction paths on ZnO(0001) surfaces

Here, we will focus on the reactive (0001) surface of ZnO. It is not straightforward to simulate oxide surfaces using the slab model, particularly for polar surfaces. In our work, many more atomic layers than PdZn were needed. Extensive convergence tests have been carried out on the free surfaces. As shown in Figure 8.8, the polar ZnO(0001) surface terminated with Zn atoms undergoes significant reconstruction. The structural parameters for the surface obtained in our calculations are compared in the figure with the benchmark work of Meyer and Marx,[29] and the agreement is quite satisfactory.

Figure 8.8 – Optimized surface structure of defect-free ZnO(0001). Atom colors are Zn (light blue) and O (red). The numbers in parentheses are from Ref. [29] and *c* (5.291 Å) is one of the lattice constants.



The strongest adsorption energies for species on ZnO(0001) are listed without ZPE correction in Table 8.4, and the energetics along the reaction paths are given with ZPE correction in Table 8.3.

8.3.4.1 CH_3OH O-H scission on ZnO(0001)

On the defect-free polar ZnO(0001) surface, methanol adsorbs weakly either on a Zn top site or a hollow (hcp or fcc) site, with its oxygen lone pair pointing to the surface, tending to maximize the interaction with Zn atoms in the top layer. The adsorption energy is −0.40 eV at the T^{Zn} site. On the other hand, CH_3O and H adsorb strongly with the surface. The strongest adsorption site is fcc for CH_3O (−3.37 eV) and T^{Zn} for H (−2.74 eV). These adsorption energies are significantly larger than those on flat PdZn surfaces.

The reaction path between the methanol and its dissociation products (Fig 3. and Table 3) yielded a barrier of 0.39 eV, which is significantly lower than those found on PdZn surfaces. The dissociation is also thermodynamically favored, with ΔE of −0.38 eV. These results are consistent with the experimental observations that the ZnO(0001) is highly reactive for methanol decomposition.[30,31] The stationary point geometries are shown in Figure 8.5.

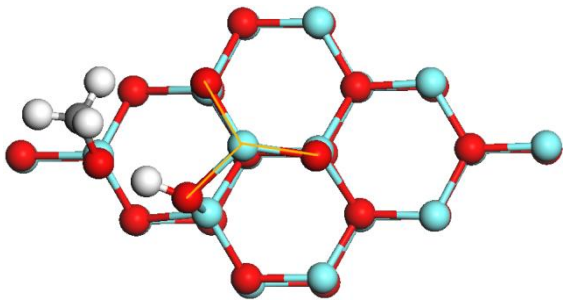
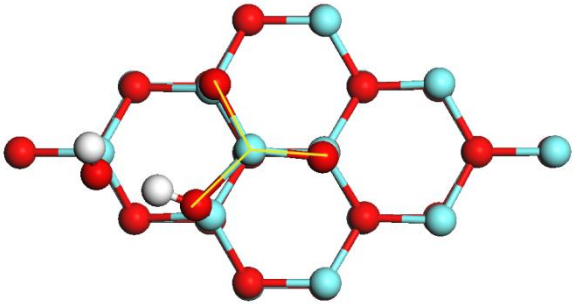
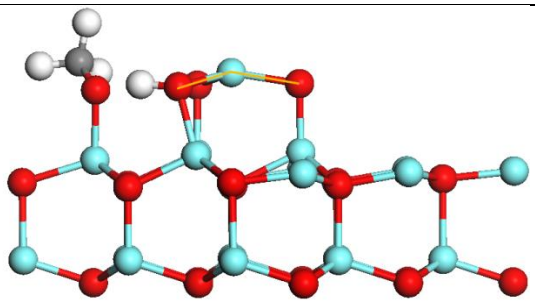
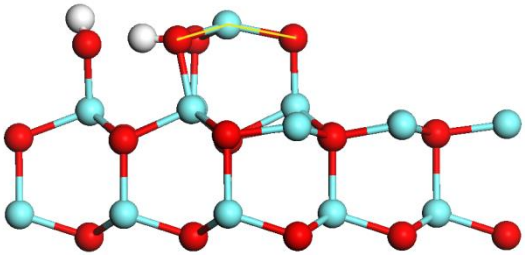
8.3.4.2 H_2O O-H scission on Zn(0001)

A similar path was explored for water (Fig. 8.6 and Table 8.3) on ZnO(0001). Water O-H scission proceeded from an initial fcc hollow site to a top site for hydrogen and the same fcc hollow for hydroxyl. This path showed a barrier of 0.42 eV and ΔE of −0.43 eV. Like CH_3OH, water adsorbs weakly on this surface (E_{ad} = −0.37 eV at the T^{Zn} site). However, OH has an adsorption energy of −4.37 eV at the fcc site, which is again much larger than those on flat PdZn surfaces. The stationary point geometries are shown in Figure 8.6.

8.3.4.3 Island defects

A recent STM study of the ZnO(0001) surface has revealed significant island/pit formation.[32,33] These structures provide abundant oxygen step-edge sites, which are only three-fold coordinated. It is believed that these triangularly shaped single-layer-deep defects stabilize the surface by removal of some Zn ions.[34] Thus, the defect-free Zn(0001) model should only be considered as a representation of the terraces of the real surface. Defects with abundant unsaturated oxygen ions have been proposed[32] to promote dissociation of methanol. To model methanol dissociation at these defect sites, we have performed calculations using a model ZnO(0001) surface with a ZnO_3^{4-} island. The optimization of an adsorbed methanol near the step-edge of the island resulted in spontaneous dissociation. The H atom transfers to the step-edge oxygen without a barrier, but retains a hydrogen bond with the methoxy oxygen, as shown in Fig. 8.9. Similarly, a dissociation of H_2O near the island was also found to be spontaneous.

Figure 8.9 – Optimized structures of CH_3OH (left panels) and H_2O (right panels) adsorbed on a ZnO(0001) island defect surface. Atom colors are Zn (light blue), O (red), C (grey), and H (white). The island defect is traced in yellow lines.

CH_3OH on defect ZnO(0001) surface	**H_2O on defect ZnO(0001) surface**
	
	

8.4 Discussion

The dissociation of methanol and water represents the initial steps in MSR. While insufficient to yield the complete picture of the mechanism, these initial processes shed valuable light onto the role of the catalyst. One finding in this work is that the dissociation barriers for both adsorbed CH_3OH and H_2O are quite high on defect-free PdZn surfaces. Our results are consistent with previous theoretical studies of methanol decomposition on other transition metal surfaces, where a substantial barrier is the norm. For example, the calculated barrier for methanol O-H scission is 0.68 eV on

Cu(110),[59] 0.81 eV on Pt(111),[60] and 0.75 eV on Pd(111).[61] Similarly, the calculated barrier for water O-H cleavage is 1.36 eV on Cu(111),[62] and 0.88 eV on Pt(111).[63] Hence, it seems that it is firmly established that the cleavage of the O-H bond on flat defect-free transition metal surfaces requires substantial activation. In addition, the cleavage of the C-O and C-H bonds is even more unfavorable on flat PdZn surfaces, which essentially excludes them from playing a significant role in MSR.

Our results indicate that the energy barrier for the initial dissociation of methanol (0.88 eV) is comparable to that of methoxy dehydrogenation (0.89 eV) on PdZn(111). However, the latter has a much higher barrier (1.06 eV vs. 0.73 eV) on PdZn(100).

Interestingly, it has been reported that CH_3OH decomposes to CH_3O on PdZn surfaces at low temperatures (~200 K), although further decomposition to CO is blocked.[23] Given the high barriers calculated on clean surfaces, it is likely that highly active defects are playing a major role in the low temperature decomposition process. Indeed, it is well established that the defect sites are responsible for the facile dissociation of CH_3OH on Pt surfaces at low temperatures,[64,65] despite a significant barrier found on defect-free surfaces.[60] Here, we have investigated the corresponding barriers for CH_3OH dissociation on the defective (221), (110), and (321) surfaces of PdZn. In general, our results indicate that the unsaturated Zn sites on these surfaces help to lower both the dissociation barriers and product energy. Thus, the dissociation processes become more favored kinetically and thermodynamically than that on the flat surfaces. We also note in passing that the (221) surface was also shown to lower the dehydrogenation barrier for CH_3O on PdZn.[37] Although the types of defects investigated

here are by no means exhaustive, our calculations lend support to the idea that the dissociation of methanol on the PdZn surface observed in the recent TPD experiment[23] is due to defect sites.

Interestingly, the TPD data of Jeroro and Vohs on PdZn reveal two distinct desorption peaks near 140 and 170 K.[23] As discussed above, the narrow peak near 140 K gives an adsorption energy of 0.37 eV, which is consistent with the 0.32 eV value for physisorbed CH_3OH on PdZn(111) obtained here. Our results on the dissociation of methanol on defective PdZn surfaces are also consistent with the assignment of the broad desorption peak at 170 K to the recombination of adsorbed CH_3O and H. The corresponding reverse barriers range from 0.59 to 0.99 eV, which are not inconsistent with the desorption temperature of 170 K (~0.45 eV). Again, due to the large variation of barrier heights for the recombination, it is difficult to identify the precise pathways. On the other hand, the broad peak in the experimental TPD spectrum could be interpreted as an indication that more than one pathway may be involved in the recombinative desorption.

It should be pointed out that the impact of the defect sites on dissociation may not be significant for MSR, which typically proceeds at relatively high temperatures (~500 K). There, the available thermal energy would allow dissociation on defect-free as well as the minority defect surfaces.

The role of the ZnO support in the dissociation of both CH_3OH and H_2O is an interesting issue. There is a growing body of evidence that oxide supports often have an active role to play in heterogeneous catalysis, particularly at the metal/oxide interface.

For instance, CeO_x and TiO_x nanoparticles have been found to promote the water-gas shift reaction on Au(111).[66] Similarly, TiO_2 in the Au/TiO_2 catalyst has recently been found to catalyze a key step in converting nitro groups to amino groups in several multifunctional organic molecules.[43] In methanol synthesis, which is the reverse process of MSR, the ZnO support in the Cu/ZnO catalyst is known to play an indispensible role, although the exact mechanism is still under debate.[44] Finally, the recent work by Vohs and coworkers has found strong synergy between the PdZn alloy and the ZnO(0001) support for methanol decomposition.[27]

Our results on both the terrace and step-edge of the ZnO(0001) surface suggest facile dissociation of both CH_3OH and H_2O. They confirmed the experimental observations,[30,31] and offered a microscopic picture of the dissociation mechanism. For example, it is clear that the dissociation on the defect-free ZnO(0001) surface does not follow the conventional Brønsted acid-base argument,[31] as both the methoxy and hydrogen product of the dissociation prefer the terminal zinc sites. This can be readily understood as the oxygen atoms on this surface are all saturated and could not make new bonds without breaking old ones.[67] On the other hand, the ready transfer of the methanol hydroxyl hydrogen to a step-edge oxygen on the defect ZnO(0001) surface is clearly due to the unsaturated nature of the surface oxygen ion, which is consistent with the Brønsted acid-base model. Our results provide strong support to the proposal of Dulub *et al.*[32] on the dissociation of CH_3OH on the ZnO(0001) surface, which highlighted the role played by the oxygen ions at step-edges. The predominant role of ZnO(0001) defects in CH_3OH dissociation was also supported by a recent experimental study, in

which 0.01 ML of Pt was found to severely attenuate the activity of ZnO(0001) at low temperatures.[68]

The facile dissociation of CH_3OH and H_2O on the ZnO surface suggested by our calculations, combined with the high dissociation barriers on PdZn surfaces, raises an interesting possibility that the reaction might occur at the interface between the oxide support and the catalytic metal. The presence of ZnO may help the initial dissociation of CH_3OH and H_2O, while the metal might catalyze further reactions including the rate-limiting dehydrogenation of CH_3O. This proposal, which needs be verified by further studies, assigns important roles for both the metal and the support, which might not perform the catalysis alone. Indeed, there is some supporting evidence for such synergism. For instance, a recent study on methanol decomposition revealed that Pd supported by ZnO(0001) is far more active than Pd alone or Pd/ZnO($10\bar{1}0$).[27] Dependence of MSR activity on ZnO morphology was also noted for the Pd/ZnO catalyst.[20] In addition, a recent theoretical study indicated a synergistic effect in water dissociation at the interface between Cu and ZrO_2.[69]

8.5 Conclusions

To summarize, we have carried out extensive plane-wave DFT calculations on the dissociation reactions of both CH_3OH and H_2O on PdZn and ZnO surfaces. These processes represent the initial steps of methanol steam reformation. It is found that the dissociation of both species is highly activated on defect-free PdZn surfaces. On the other hand, defect sites lower the barrier significantly and may play a role at low temperatures, while remaining less relevant at MSR conditions. Additionally, our data

suggests a possible role for the oxide support, as shown by low dissociation barriers on ZnO(0001) surface and null barriers at island defects.

8.6 References

1 Schlapbach, L. & Zuttel, A. Hydrogen-storage materials for mobile applications. *Nature* **414**, 353-358 (2001).

2 Eberle, U., Felderhoff, M. & Schuth, F. Chemical and physical solutions for hydrogen storage. *Angewandte Chemie International Edition* **48**, 6608-6630 (2009).

3 Trimm, D. L. & Onsan, Z. I. Onboard fuel conversion for hydrogen-fuel-cell-driven vehicles. *Catalysis Today* **43**, 31 (2001).

4 Olah, G. A. After oil and gas: methanol economy. *Catalysis Letters* **93**, 1 (2004).

5 Palo, D. R., Dagle, R. A. & Holladay, J. D. Methanol steam reforming for hydrogen production. *Chemical Reviews* **107**, 3992 (2007).

6 Litster, S. & McLean, G. PEM fuel cell electrodes. *Journal of Power Sources* **130**, 61-76 (2004).

7 Twigg, M. V. & Spencer, M. S. Deactivation of copper metal catalysts for methanol decomposition, methanol steam reforming and methanol synthesis. *Topics in Catalysis* **22**, 191 (2003).

8 Iwasa, N., Yamamoto, O., Akazawa, T., Ohyama, S. & Takazawa, N. Dehydrogenation of methanol to methyl formate over palladium zinc-oxide catalysts. *Chemical Communications*, 1322-1323 (1991).

9 Iwasa, N., Masuda, S., Ogawa, N. & Takezawa, N. Steam reforming of methanol over Pd/ZnO: Effect of the formation of PdZn alloys upon the reaction. *Applied Catalysis A: General* **125**, 145 (1995).

10 Iwasa, N., Mayanagi, T., Wataru, N. & Takewasa, T. Effect of Zn addition to supported Pd catalysts in the steam reforming of methanol. *Applied Catalysis A: General* **248**, 153 (2003).

11 Iwasa, N. & Takezawa, N. New supported Pd and Pt alloy catalysts for steam reforming and dehydrogenation of methanol. *Topics in Catalysis* **22**, 215 (2003).

12 Iwasa, N., Yoshikawa, M., Nomura, W. & Arai, M. Transformation of methanol in the presence of steam and oxygen over ZnO-supported transition metal catalysts under stream reforming conditions. *Applied Catalysis A: General* **292**, 215-222 (2005).

13 Takezawa, N. & Iwasa, N. Steam reforming and dehydrogenation of methanol: Difference in the catalytic functions of copper and group VIII metals. *Catalysis Today* **36**, 45 (1997).

14 Chen, Z.-X., Neyman, K. M., Gordienko, A. B. & Rosch, N. Surface structure and stability of PdZn and PtZn alloys: Density-functional slab model studies. *Physical Review B* **68**, 1-8 (2003).

15 Tsai, A. P., Kameoka, S. & Ishii, Y. PdZn = Cu: Can an intermetallic compound replace an element? *Journal of the Physical Society of Japan* **73**, 3270-3273 (2004).

16 Cubeiro, M. L. & Fierro, J. L. G. Selective production of hydrogen by partial oxidation of methanol over ZnO-supported palladium catalysts. *Journal of Catalysis* **179**, 150 (1998).

17 Chin, Y.-H., Dagle, R. A., Hu, J., Dohnalkova, A. C. & Wang, Y. Steam reforming of methanol over highly active Pd/ZnO catalyst. *Catalysis Today* **77**, 79 (2002).

18 Ranganathan, E. S., Bej, S. K. & Thompson, L. T. Methanol steam reforming over Pd/ZnO and Pd/CeO2 catalysts. *Applied Catalysis A: General* **289**, 153-162 (2005).

19 Pfeifer, P., Schubert, K., Liauw, M. A. & Emig, G. PdZn catalysts prepared by washcoating microstructured reactors. *Applied Catalysis A: General* **270**, 165 (2004).

20 Karim, A., Conant, T. & Datye, A. The role of PdZn alloy formation and particle size on the selectivity for steam reforming of methanol. *Journal of Catalysis* **243**, 420 (2006).

21 Bayer, A. *et al.* Electronic properties of thin Zn layers on Pd(111) during growth and alloying. *Surface Science* **600**, 78 (2006).

22 Bera, P. & Vohs, J. M. Reaction of CH_3OH on Pd/ZnO(0001) and PdZn/ZnO(0001) model catalysts. *Journal of Physical Chemistry C* **111**, 7049-7057 (2007).

23 Jeroro, E. & Vohs, J. M. Zn modification of the reactivity of Pd(111) toward methanol and formaldehyde. *Journal of the American Chemical Society* **130**, 10199-10207 (2008).

24 Lebarbier, V. *et al.* CO/FTIR spectroscopic characterization of Pd/ZnO/Al2O3 catalysts for methanol steam reforming. *Catalysis Letters* **122**, 223-227 (2008).

25 Jeroro, E., Lebarbier, V. M., Datye, A., Wang, Y. & Vohs, J. M. Interaction of CO with surface PdZn alloys. *Surface Science* **601**, 5546-5554 (2007).

26 Jeroro, E. & Vohs, J. M. Exploring the role of Zn in PdZn reforming catalysts: absorption and reaction of ethanol and acetaldehyde on two-dimensional PdZn alloys. *Journal of Physical Chemistry C* **113**, 1484-1494 (2009).

27 Hyman, M. P., Lebarbier, V. M., Wang, Y., Datye, A. K. & Vohs, J. M. A Comparison of the Reactivity of Pd Supported on ZnO(10$\bar{1}$0) and ZnO(0001). *The Journal of Physical Chemistry C* **113**, 7251-7259 (2009).

28 Rameshan, C. *et al.* Subsurface-controlled CO2 selectivity of PdZn near surface alloys in H2 generation by methanol steam reforming. *Angewandte Chemie (International ed. in English)* **49**, 3224-3227 (2010).

29 Meyer, B. & Marx, D. Density-functional study of the structure and stability of ZnO surfaces. *Physical Review B* **67**, 035403 (2003).

30 Cheng, W. H., Akhter, S. & Kung, H. H. Structure sensitivity in methanol decomposition on ZnO single-crystal surfaces. *Journal of Catalysis* **82**, 341-350 (1983).

31 Vohs, J. M. & Barteau, M. A. Coversion of methanol, formaldhyde, and formic acide on the polar faces of zinc oxide. *Surface Science* **176**, 91-114 (1986).

32 Dulub, O., Boatner, L. & Diebold, U. STM study of the geometric and electronic structure of ZnO(0001)-Zn, (0001bar)-O, (101bar0), and (112bar0) surfaces. *Surface Science* **519**, 201-217 (2002).

33 Dulub, O., Diebold, U. & Kresse, G. Novel Stabilization Mechanism on Polar Surfaces: ZnO(0001)-Zn. *Physical Review Letters* **90**, 1-4 (2003).

34 Kresse, G., Dulub, O. & Diebold, U. Competing stabilization mechanisms for the polar ZnO(0001)-Zn surface. *Physical Review B* **68**, 245409 (2003).

35 Chen, Z.-X., Lim, K. H., Neyman, K. M. & Rosch, N. Density functional study of methoxide decomposition on PdZn(100). *Physical Chemistry Chemical Physics* **6**, 4499 (2004).

36 Chen, Z.-X., Neyman, K. M., Lim, K. H. & Rosch, N. CH_3O decomposition on PdZn(111), Pd(111), and Cu(111). A theoretical study. *Langmuir* **20**, 8068 (2004).

37 Chen, Z.-X., Lim, K. H., Neyman, K. M. & Rosch, N. Effect of steps on the decomposition of CH_3O at PdZn alloy surfaces. *Journal of Physical Chemistry B* **109**, 4568 (2005).

38 Lim, K. H., Chen, Z.-X., Neyman, K. M. & Rosch, N. Comparative theoretical study of formaldehyde decomposition on PdZn, Cu, and Pd surfaces. *Journal of Physical Chemistry B* **110**, 14890 (2006).

39 Lim, K. H., Moskaleva, L. V. & Rosch, N. Surface composition of materials used as catalysts for methanol steam reforming: A theoretical study. *ChemPhysChem* **7**, 1802 (2006).

40 Neyman, K. M. *et al.* Microscopic models of PdZn alloy catalysts: structure and reactivity in methanol decomposition. *Physical Chemistry Chemical Physics* **9**, 3470 (2007).

41 Zambelli, T., Wintterlin, J., Trost, J. & Ertl, G. Identification of the "active sites" of a surface-catalyzed reaction. *Science* **273**, 1688 (1996).

42 Dahl, S. *et al.* Role of steps in N_2 activation on Ru(0001). *Physical Review Letters* **83**, 1841 (1999).

43 Li, S.-C. & Dideberg, U. Reactivity of TiO2 rutile and anatase surfaces towards nitroaromatics. *Journal of the American Chemical Society* **132**, 64-66 (2010).

44 Spencer, M. S. The role of zinc oxide in Cu/ZnO catalysts for methanol synthesis and the water-gas shift reaction. *Topics in Catalysis* **8**, 259-266 (1999).

45 Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**, 558 (1993).

46 Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using plane wave basis set. *Physical Review B* **54**, 11169 (1996).

47 Kresse, G. & Furthmuller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using plane wave basis set. *Computational Material Science* **6**, 15 (1996).

48 Perdew, J. P. *et al.* Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **46**, 6671 (1992).

49 Greeley, J., Norskov, J. K. & Mavrikakis, M. Electronic structure and catalysis on metal surfaces. *Annual Review of Physical Chemistry* **53**, 319 (2002).

50 Blochl, P. Project augmented-wave method. *Physical Review B* **50**, 17953 (1994).

51 Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**, 1758 (1999).

52 Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Physical Review* **B13**, 5188 (1976).

53 Methfessel, M. & Paxton, A. T. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B* **40**, 3616 (1989).

54 Neugebauer, J. & Scheffler, M. Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111). *Physical Review B* **46**, 16067 (1992).

55 Makov, G. & Payne, M. C. Periodic boundary conditions in ab initio calculations. *Physical Review B* **51**, 4014 (1995).

56 Jonsson, H., Mills, G. & Jacobsen, K. W. in *Classical and Quantum Dynamics in Condensed Phase Simulations* (eds B. J. Berne, G. Ciccotti, & D. F. Coker) (World Scientific, 1998).

57 Henkelman, G., Uberuaga, B. P. & Jonsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics* **113**, 9901-9904 (2000).

58 Redhead, P. A. Thermal desorption of gases. *Vacuum* **12**, 203-211 (1962).

59 Sakong, S. & Gross, A. Total oxidation of methanol on Cu(110): A density functional theory study. *Journal of Physical Chemistry A* **111**, 8814 (2007).

60 Greeley, J. & Mavrikakis, M. A first-principles study of methanol decomposition on Pt(111). *Journal of the American Chemical Society* **124**, 7193 (2002).

61 Schennach, R., Eichler, A. & Rendulic, K. D. Adsorption and desorption of methanol on Pd(111) and on a Pd/V surface alloy. *Journal of Physical Chemistry B* **107**, 2552 (2003).

62 Gokhale, A. A., Dumesic, J. A. & Mavrikakis, M. On the mechanism of low-temperature water gas shift reaction on copper. *Journal of the American Chemical Society* **130**, 1402 (2008).

63 Grabow, L. C., Gokhale, A. A., Evans, S. T., Dumesic, J. A. & Mavrikakis, M. Mechanism of the water gas shift reaction on Pt: First principles, experiments, and microkinetic modeling. *Journal of Physical Chemistry C* **112**, 4608 (2008).

64 Gibson, K. D. & Dubois, L. H. Step effects in the thermal decomposition of methanol on Pt(111). *Surface Science* **233**, 59 (1990).

65 Diekhoner, L., Butler, D. A., Baurichter, A. & Luntz, A. C. Parallel pathways in methanol decomposition on Pt(111). *Surface Science* **409**, 384-391 (1998).

66 Rodriguez, J. A. *et al.* Activity of CeOx and TiOx nanoparticles grown on Au(111) in the water-gas shift reaction. *Science* **318**, 1757-1760 (2007).

67 Halevi, B. & Vohs, J. M. Reactions of CH_3SH and $(CH_3)_2S_2$ on the (0001) and (000) surfaces of ZnO. *Journal of Physical Chemistry B* **109**, 23976 (2005).

68 Grant, A. W. *et al.* Methanol decomposition on Pt/ZnO(0001)-Zn model catalysts. *Journal of Physical Chemistry B* **105**, 9273 (2001).

69 Tang, Q.-L. & Liu, Z.-P. Identification of the active Cu phase in the water-gas shift reaction over Cu/ZrO2 from first principles. *Journal of Physical Chemistry C* **114**, 8423 (2010).

Conclusions

Project 1 – Bacterial Enzyme SpvC

1. Classical MD study of the enzyme active-site showed key residues involving in substrate binding.
2. The truncated model showed a concerted mechanism.
3. Inclusion of enzyme environment (especially Lys104 and Tyr158) showed the truncated model was incomplete. The QM/MM model showed a step-wise E1cB mechanism with a barrier of 18.2 kcal/mol compared to 21.5 kcal/mol for the Y158F mutant.

Project 2 – Human Enzyme CDK2

1. *Ab-initio* QMMM/MD evaluation of competing reaction paths showed strong evidence that the general base pathway is dominant over the substrate assisted pathway.
2. Construction of the potential of mean force along the reaction path showed a general base pathway, with a barrier of 10.8 kcal/mol, consistent with experimental results.

Project 3 – Bacterial Enzyme Anthrax Lethal Factor

1. Semi-empirical QM/MM (SCC-DFTB) analysis of the Michaelis complex showed the substrate is tightly held by both electrostatic and hydrophobic interactions.
2. Construction of the potential of mean force along reaction paths showed the key role played by Tyr728 as the barrier height rose by 4.5 kcal/mol when this residue was mutated to a phenylalanine residue.
3. A mechanism was proposed for nucleophilic substitution on the backbone carbonyl carbon by a water molecule activated by a nearby glutamate residue. The zinc ion and a neighboring tyrosine act as an oxyanion hole to stabilize the charged carbonyl intermediate and C-N bond breaking was stabilized by subsequent protonation by the same glutamate that did the initial proton abstraction from water.

Project 4 – Methanol Steam Reforming on PdZn alloy

1. Plane-wave DFT studies of MSR – model building for both PdZn alloy and ZnO support and subsequent binding energies showed the variety of possible surface sites available for reactant and intermediate binding.
2. Initial O-H breaking is highly activated on flat surfaces, and progressively less so at Zn rich defect sites depending on the degree of saturation.