COMPUTATIONAL CHEMISTRY: INVESTIGATIONS OF PROTEIN-PROTEIN INTERACTIONS AND POST-TRANSLATIONAL MODIFICATIONS TO PEPTIDES

By

Michael R. Jones

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Chemistry—Doctor of Philosophy

2017

ABSTRACT

COMPUTATIONAL CHEMISTRY: INVESTIGATIONS OF PROTEIN-PROTEIN INTERACTIONS AND POST-TRANSLATIONAL MODIFICATIONS TO PEPTIDES

By

Michael R. Jones

Computational chemistry plays a vital role in understanding chemical and physical processes and has been useful in advancing the understanding of reactions in biology. Improper signaling of the nuclear factor- κ B (NF- κ B) pathway plays a critical role in many inflammatory disease states, including cancer, stroke, and viral infections. Aberrant regulation of this pathway happens upon the signal-induced degradation of the inhibitor of κ B (I κ B) proteins. The activation of I κ B kinase (IKK) subunit β (IKK β) or NF- κ B Inducing Kinase (NIK), initiates this cascade of events. Understanding the structure-property relationships associated with IKK β and NIK is essential for the development of prevention strategies. Although the signaling pathways are known, how the molecular mechanisms respond to changes in the intracellular microenvironment (i.e., pH, ionic strength, temperature) remains elusive.

In this dissertation, computer simulation and modeling techniques were used investigate two protein kinases complexed with either small molecule activators or inhibitors in the active, inactive, and mutant states to correlate structure-property and structure-function relationships as a function of intracellular ionic strength. Additionally, radical-induced protein fragmentation pathways, as a result of reactions with reactive oxygen species, were investigated to yield insight into the thermodynamic preference of the fragmentation mechanisms. Analyses of the relationship between structure-activity and conformationalactivity indicate that the protein-protein interactions and the binding of small molecules are sensitive to changes in the ionic strength and that there are several factors that influence the selectivity of peptide backbone cleavage. As there are many computational approaches for predicting physical and chemical properties, several methods were considered for the predictions of protein-protein dissociation, protein backbone fragmentation, and partition coefficients of drug-like molecules.

This dissertation is dedicated to family.

ACKNOWLEDGEMENTS

I am truly grateful for all of those who have been supportive during my academic career. Firstly, I would like to thank my research advisor and mentor, Professor Angela K. Wilson for her unremitting support and guidance. As an advisor, Professor Wilson consistently pushed me to be my best (propel through challenges), encouraged me to attend conferences and scientific meetings, apply for awards and seize opportunities for research enhancement and professional development. As a mentor, she fosters creativity in problem solving and I am very thankful for her patience through this journey.

I acknowledge the training and support from:

- Professor Maria C. Nagan for teaching me to be fearless of caveats in scientific research;
 Professor Thomas R. Cundari for promoting out-of-the-box thinking and his support on several projects;
- Professor H. David Wohlers for inspiring me on innovative teaching;
- Dr. Bernard R. Brooks, for training in software development and the opportunity to experience the National Institute of Health;
- Dr. Frank Pickard, III and Dr. Andy Simmonett for teaching me better practices in programming; the National Institutes of Health Graduate Partnership Program for providing financial support throughout my dissertation research;
- Reata Pharmaceuticals for their support and Dr. Mike Visnick for continued collaboration;
- the Ronald E. McNair Scholars Program at Truman State University, for the academic counseling and continuing support.

I thank my committee members at Michigan State University Professor Kenneth M. Merz, Jr., Professor Piotr Piecuch, and Professor Gary J. Blanchard for their time and guidance. Additionally, I would like to acknowledge my former committee members at the University of North Texas for their mentoring and encouragement: Professor Thomas R. Cundari, Professor Martin Schwartz, and Professor Weston T. Borden.

The Wilson Research Group, collective for both current and past members, is characterized by a team of diverse thinkers from different backgrounds with a variety of talents that act as a family, sharing knowledge and promoting an environment of unconditional positive regard. I am very grateful for the Wilson Research Group for limitless discussion and feedback, advice, and exposure to amazing foods. Thank you Dr. Zainab H. A. Alsunaidi and Dr. Jiaqi Wang for sharing workspace and sharing your lunches with me; Dr. Kameron Jorgensen, Dr. Marie Laury, Dr. Andrew Mahler Dr. Cong Liu, Dr. Michael Drummond, Dr. George Schoendorff, Dr. Deborah Penchoff, and Dr. Inga Ulusoy for helping me troubleshoot through problems. Many thanks to Lucas Aebersold, Thomas Diaz, and Yigitcan Eken for great research discussions and for sharing pizza with me during the late evenings in lab.

I extend my gratitude to my friends who filled in the roles of family, for their support and encouragement through my academic and professional pursuits: Michael D., Edward A., Erica S., Stephanie M., Kelsey W., Jessi B., Thiky V., Alexis M., Darius T., Shawn G., Chelsea C. Gabino M., Gail W., Jesseca S., Sherard L. Darrell M., Jose D., and Matthew C. I am grateful for my fraternity, Sigma Lambda Beta International Fraternity Incorporated, for providing an extended network of support. Special thanks to Anna Penchenina for unconditional love, motivation during the long nights of studying at the coffee shops, and making sure I stayed healthy and focused during my doctoral candidacy exams.

Little Brother, Grandmother, and Intermediate Relatives, thank you for being there.

Lastly, Mom and Dad, I thank you for everything. Although you were not able to see me through the completion, I carried your teachings and trainings throughout. Because of you, I am.

TABLE OF CONTENTS

LIST OF TABLES	X
LIST OF FIGURES	xi
CHAPTER 1 Introduction	1
CHAPTER 2 Theory and Methods in Molecular Modeling	5
2.1 Equations of Motion	5
2.2 Molecular Dynamics	7
2.2.1 Considerations	10
2.2.2 Limitations	13
2.3 Quantum Mechanics	14
2.3.1 <i>Ab initio</i> Methods	16
2.3.2 Density Functional Theory	18
2.3.3 Basis Sets	19
REFERENCES	21
CHAPTER 3 Molecular Dynamics Studies of the Protein–Protein Interactions in Inhibitor KB Kinase-ß	of 29
3.1 Introduction	
3.2 Computational Methods	32
3.3 Results	37
3.3.1 Docking is Aesthetic	37
3.3.2 Dimer Stability of the Crystal System	40
3.3.3 Coarse-Grained Simulations	41
3.3.4 Atomistic Simulation of Dimers and Monomer	42
3.3.5 Hydrogen Bonding Between Dimer Interfaces	45
3.3.6 Binding Free Energies	49
3.4 Discussion	52
3.5 Conclusion	54
REFERENCES	56
CHAPTER 4 Impact of Intracellular Ionic Strength on Dimer Binding in the NF-kB Induc Kinase	ing 64
4.1 Introduction	64
4.2 Material and Methods	68
4.3 Results and Discussion	70
4.3.1 Structural Fluctuation	70
4.3.2 Changes in Solvent Accessibility	73
4.3.3 Hydrogen-Bonding Analysis/Dimer vs Monomer	75
4.3.4 Aggregation Propensity Sensitive to Ion Buffer	77
4.3.5 Comparisons Between the Dimer Binding Energies	78
4.4 Conclusions	79
REFERENCES	81

CHAPTE	R 5 Selectivity in ROS-Induced Peptide Backbone Bond Cleavage	
5.1	Introduction	
5.2	Methodology	
5.3	Results and Discussion	
5.3.1	Pathway Favorability	
5.3.2	Role of Structure on Bond Strength and Site Reactivity	
5.3.3	Insights into Pathway Preferences	
5.3.4	Evaluation of Methods	
5.4	Conclusion	
REFERE	NCES	
CHAPTE	R 6 Partition Coefficients for the SAMPL5 Challenge using Transfer Fr	ee Energies
6.1	Introduction	108
6.2	Methods	111
6.3	Results and Discussion	113
6.4	Conclusion	121
REFERE	NCES	

CHAPTER 7	CONCLUDING REMARKS	129
REFERENCES		133

LIST OF TABLES

Table 3.1 Summary of modeled systems.	
Table 3.2 Binding cavities near the activation segment.	
Table 3.3 Summary of dimer assembly energetics, determined with PISA	
Table 3.4 Averaged thermodynamic data for the 10 ns atomistic simulation	
Table 3.5 Hydrogen bonding occupancies for the AB dimer interface	
Table 3.6 Hydrogen bonding occupancies for the AD dimer interface	
Table 3.7 Binding free energies of the IKKβ dimer. ^a	50
Table 4.1 Average thermodynamic parameters from simulations	
Table 4.2 Summary of the dimer interface of NIK.	74
Table 4.3 Hydrogen bonding analysis of the NIK dimer interface. ^a	76
Table 4.4 Binding free energies of the NIK dimer. ^a	79
Table 5.1 Calculated reaction enthalpies for Pathway [A] and Pathway [B] at the CCS of theory.	SD(T) level
Table 5.2 Calculated reaction enthalpies for Pathway [A] and Pathway [B]	
Table 5.3 Calculated reaction barriers for reactions [A1] and [B1].	
Table 6.1 Comparison of predicted logP with experimental logD.	114
Table 6.2 Overview of the results submitted to the SAMPL5 challenge	121

LIST OF FIGURES

Figure 3.1 Representation of the models. The crystal structure was solved with 8 identical molecules (a) forming two "dimer of dimers" tetrameric complex (b). Although a dimer is defined as AB form (d), the alternative dimer AD (c) has significance for structure stability.
Figure 3.2 Characterization of ATP/Mg ²⁺ docking cavity. The ATP ligand is positioned in the mouth of the activation loop (Top). The adenine head of ATP is pointed towards Asp166 and the phosphate tail is outside of the pocket. The binding pocket is outlined and described by colored circles (purple and green) representing surrounding amino acids (Bottom); the different outlines on the purple circles contrast the different types of polar side chains. Blue and green arrows indicate sidechain and backbone acceptor-donors, respectively. The blue spheres surrounding the phosphate tail represent the pocket exposure, whereas the lighter-blue spheres highlight the receptor's exposure
Figure 3.3 Visualization of the predicted cavities. All alpha-spheres are shown for the four sites (Top). The red and white spheres indicated hydrophilic and hydrophobic cavities, respectively. Wireframe representations of the four sites are shown individually: Site 2 (middle left); Site 3 (middle right); Site 6 (bottom left); and Site 10 (bottom right)
Figure 3.4 RMSD plots for the different assemblies using coarse-graining MD. Frames were plotted for every 1 ps for a total of 50,000 frames (50 ns)
Figure 3.5 The RMSD from the starting structure for the 10 ns simulation. The different RMSDs of the monomer (a-b), AB dimer (c-d) and AD dimer (e-f) are colored accordingly (blue, WT; red, S177/181E; green, WT+ATP; purple;S177/181pS). Plots A, C, and E represent the RMSD of the entire structure whereas the plots B, D, and F represent only the activation segment of the principle monomeric chain. Frames were plotted for every 2 ps for a total of 5,000 frames (10 ns)
Figure 3.6 Comparison of the phosphorylated AB Dimer. Shown are the starting structure (top) and the final structure (bottom) after 10 ns
Figure 3.7 Comparison of binding free energy trends with the HCT, OBC1, and OBC2 models.
Figure 4.1 Root mean square deviations (RMSDs) obtained from the 215 ns simulation. (A) The RMSDs were calculated from the starting structure of the inactive (WT) and phosphomimetic mutant (S549D) states in a neutral solution (color coded in blue and green, respectively) and a 100 mM NaCl solution ([Na+]; color coded in red and purple, respectively). For comparison, plots B and C highlight the trends between the inactive and the active states, whereas the impact of the ionic environment is highlighted in plots D and E. The darker colored lines represent the moving average per 100 frames

 Figure 4.2 Root mean square fluctuations (RMSF) of each protomer of NIK. The RMSFs over the 215 ns were calculated per residue for each protomer (332-675) for the modeled inactive states, WT and WT [Na+] (represented in blue and red, respectively), and active states, S549D and S549D [Na+] (represented in green and purple, respectively)73
Figure 4.3 Solvent residing between the dimer interface
Figure 4.4 Changes in the compactness of the crystal structure
Figure 4.5 Binding free energy trends with the HCT and OBC Generalized Born models 79
 Figure 5.1 Proposed reaction pathways. The diamide (pathway [A]) and α-amidation (pathway [B]) pathways were modeled in this study. The different sites of radicalization C₁, C₂, and C₃ represent Cα1, Cα2 and Cα3, respectively.
Figure 5.2 Modeled backbone conformations. (1) Represents the β -strand or "sheet-like" component of a β -pleated-sheet. (2) The β -turn structure. In both conformations, the first α -carbon has an alkoxyl radical in place of a hydrogen at C α 1
Figure 6.1 Structure of the 53 compounds investigated in the SAMPL5 challenge. The 13 molecules represent (left) correspond to the Batch 0 subset
Figure 6.2 Subset of molecules (Batch0) LogP: effect of increasing basis sets. The predicted logP with respect to increasing basis set size (DZ, TZ, QZ) are represented by the red box, green triangle, and purple 'X', respectively

CHAPTER 1 Introduction

Theoretical and computational chemistry play a vital role in understanding chemical and physical processes at both qualitative and quantitative levels and has proven useful towards advancements in research areas within chemistry, physics, materials science, biology, and medicine. In terms of chemical physics and physical chemistry, advances in computing technology have enabled theoretical predictions to yield insight about molecular properties and reactions, connecting microscopic and macroscopic phenomena that take place across various time scales. The development of next-generation simulation and modeling techniques have had a noteworthy influence on advancing the understanding of biological phenomena; applications of these theoretical approaches are often essential for resolving or uncovering phenomena not readily observable or explorable by traditional experimental techniques.

A strength in computational modeling stems from the ability to commence investigations from either *ab initio* (from the beginning) or *a posteriori* (from the latter)—a bottom-up or topdown approach—allowing theoretical predictions to provide useful insight, especially in an absence of experimental knowledge. The foundations of theoretical chemistry lie within the motions of atoms and molecules and are described using quantum, classical, and statistical mechanics. Methods in quantum chemistry are used to obtain comprehensive detail about structure, whereas methods in classical and statistical mechanics are employed to examine and relate structural dynamics to function or the macroscopic observation. Despite the advancements in high performance computing, computational chemistry methods face limitations as the complexity of the system of study increases as this requires more computing resources, hence there are many methods that have advantages over the other at the expense of adding additional approximations.

For the study of biological species, where complex processes happen at the cellular level, computation can play an important role. Mechanisms of the cell are a result of a series of molecular recognition events of proteins with other molecules—a many body problem. From a mesoscopic perspective, the structure, function, and activity of proteins rely on the local environment and must respond to perturbations to maintain equilibrium. Cellular dysfunctions occur when there is a fault within intracellular mechanisms, which may happen as a result of homeostatic imbalances, external signals, or errors within the genetic code. It is essential to understand the faults from not only a macroscopic perspective, but also a microscopic perspective, and, it is here where computation can provide important insight.

Critical to the onset and progression of disease states, it is essential to understand not only the origin of the fault, but also the intermediates involved within the progression, like a chemical reaction, as this allows the potential to discover or design strategies to control the course of the disease. As electrostatic interactions mediate protein recognition of other molecules, including RNA, DNA, ligands and other proteins, it is essential to discern how the structure and function of key mechanisms, components, and proteins within the inflammatory pathway respond to perturbations.

In this dissertation, theoretical methods, discussed in Chapter 2, were employed to examine post-translational modifications to proteins and how these alter protein structure and function. NF- κ B (nuclear factor kappa-light-chain-enhancer of activated B cells) proteins are involved in many cellular and organismal processes, including immune and inflammatory responses, developmental processes, cellular growth, and apoptosis. In most cells, NF- κ B is a latent, inactive complex in the

2

cytoplasm, but can rapidly enter the nucleus and activate gene expression via an abnormal signal cascade once it receives any number of extracellular signals. Many cancers are induced by NF- κ B signaling, including breast cancer, leukemia, lung cancer, pancreatic cancer, lymphoma, colon cancer, prostate cancer, and ovarian cancer. Thus, control of NF- κ B activity is critical, and NF-B activating protein kinases are of great interest as a target for cancer therapeutics.

In this work, two key enzymes associated with the canonical and non-canonical pathway of the NF- κ B pathway are investigated. Activation of the inhibitor of κ B kinase subunit β (IKK β) oligomer initiates a cascade that results in the translocation of NF- κ B transcription factors activating the canonical pathway. Dimerization of IKK β is required for its activation. In Chapter 4, coarse-grained and atomistic molecular dynamics simulations were employed to investigate the conformation-activity and structure-activity relationships within the oligomer assembly of IKK β that are impacted upon activation, mutation, and binding of ATP. The non-canonical pathway is dependent upon the activation of NF- κ B Inducing Kinase (NIK). While the activation of the canonical pathway is straightforward as a series of post-translational modifications, the activation of the NIK differs as it is dependent on the enzyme stability. In Chapter 5, molecular dynamics simulations were employed to determine how the ionic environment alters the structure-property relationships and aid in the stabilization of NIK.

In cancer cells, high levels of reactive oxygen species (ROS) are prevalent and can result in post-translational mechanisms of protein oxidation that yield selective side-chain and backbone modifications including abstractions, donations, additions, substitutions, and fragmentation. From a molecular perspective, prediction of the area of a protein most likely to be oxidized is of interest as this may shed insight into mechanisms that follow. In Chapter 6, to characterize the selectivity of radical-mediated fragmentation, quantum mechanical investigations using *ab initio* and density functional methods were employed to evaluate site, conformation, and fragmentation pathway trends of peptide models.

In Chapter 7, physiochemical properties of small drug-like molecules were predicted using a series of *ab initio* approaches to assess the reliability of different methods that may serve in lieu of when experimental data is unavailable. All of the research is summarized in Chapter 8 and future directions and interests are highlighted.

CHAPTER 2 Theory and Methods in Molecular Modeling

2.1 Equations of Motion

Chemistry is the study of structure, composition, and properties of molecules and the transformations they undergo. Understanding the underlying mechanisms that drive these transformations lies within the concept of mechanics, specifically the motions of electrons and nuclei. The properties of a many-body system can be expressed as a function of the position and the rate of change of displacement.

In classical mechanics, the motion of a system of *N* particles can be described by Newton's laws of motion (2.1). $^{1-3}$

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}_1, \dots, \mathbf{r}_N, \dot{\mathbf{r}}_1, \dots, \dot{\mathbf{r}}_N, t)$$
(2.1)

For a general particle *i* of mass *m*, the force **F** depends on the position $\mathbf{r}_i = (x_i, y_i, z_i)$, velocity $(\dot{\mathbf{r}})$, and time (t).^{*} This relationship of Newton's second law is also expressed as the change in momentum (**p**) with respect to time (t).

$$\mathbf{F}_{i} = \frac{d}{dt} m_{i} \dot{\mathbf{r}}_{i} = \frac{d}{dt} \mathbf{p}_{i}$$

$$\mathbf{F}_{i} = \mathbf{F}_{i}^{\text{ext}} + \sum_{i \neq j, i=1}^{N} \mathbf{F}_{ji}$$
(2.2)

For a system of N particles, the motion of particle i is influenced by external forces outside the system and internal forces that arise from the particle interacting with every other particle in the

^{*}Vector quantities are represented in the boldface notation. Total time derivatives are expressed using Newton's notation, where a dot is placed over the dependent variable ($\dot{x} = \frac{dx}{dt}, \ddot{x} = \frac{d^2x}{dt^2}$).

system. Forces in action may do work and transfer energy. For isolated systems in which there are no external forces ($\mathbf{F}_i^{\text{ext}} = 0$), the momentum and the total energy (*E*) are conserved. For conservative forces, the change in energy of motion, or kinetic energy (*T*), is defined as

$$\frac{d}{dt}T = \sum_{i=1}^{N} \mathbf{F}_{i} \cdot \dot{\mathbf{r}}_{i}$$
$$\mathbf{F}_{i} = -\nabla_{i}V$$
$$V = V(\mathbf{r}_{1}, \dots, \mathbf{r}_{N})$$
(2.3)

where the forces are derived from the gradient of a potential energy function V.

In an alternative formulation, Newton's laws can be expressed in the Lagrangian formula,

$$L = T(\mathbf{q}, \dot{\mathbf{q}}, t) - V(\mathbf{q}, \dot{\mathbf{q}}, t)$$
(2.4)

where the Lagrangian (*L*) is defined as the difference between the kinetic energy and potential energy and is expressed in generalized coordinates \mathbf{q} and generalized velocity ($\dot{\mathbf{q}}$). Lagrangian mechanics differ from Newtonian mechanics because it uses generalized coordinates

Hamilton's equations (2.5) replaces the generalized velocity with generalized momenta (\mathbf{p}), in which the Hamiltonian (H) represents the total energy.

$$H(\mathbf{q}, \mathbf{p}, t) \equiv \sum_{l=1}^{f} \mathbf{p}_{l} \mathbf{v}_{l} (\mathbf{q}, \mathbf{p}, t) - L(\mathbf{q}, \mathbf{v}(\mathbf{q}, \mathbf{p}, t), t)$$

$$(2.5)$$

These equations of motions are useful as they can be used in quantum mechanics, thermodynamics, and statistical mechanics.

2.2 Molecular Dynamics

Molecular dynamics (MD) simulations and the atomic level detail that they provide have enhanced our understanding of natural phenomena, allowing for enabling predictions about a system of interacting particles as time evolves. While serving as a connection between microscopic and macroscopic observations, MD simulations also act as a bridge between theory and experiment by establishing a connection between structure and dynamics.^{4–6} Using theory to help interpret and guide experiment, MD simulations are widely used in the study of hard and soft condensed matter physics, chemical reactions, and biological systems. For biological systems, MD simulations not only can provide insight into intermolecular interactions and structure, but can additionally examine properties not readily discernable by traditional experimental or other structural biology methods.^{7–9}

Classical MD simulations are deterministic and employ Newton's equations of motion (2.6) to propagate the positions and velocities of particles under the influence of a given, predefined potential (V) where **F** is the force on a particle *i* of mass *m* at configuration **r** with the acceleration ($\ddot{\mathbf{r}}$) and is the gradient of the potential.⁴

$$\mathbf{F}_{i} = m_{i} \ddot{\mathbf{r}}_{i} = m_{i} \frac{d^{2} \mathbf{r}_{i}}{dt^{2}} = -\nabla \mathbf{V}(\mathbf{r}_{i})$$
(2.6)

Dynamics of the system are determined by integrating Newton's motion equation as a function of time, *t*, and is done so numerically (as there is no analytical solution) with a time stepping algorithm, such as the Verlet,¹⁰ Leap-Frog,¹¹ Velocity-Verlet,¹² or Beeman's¹³ algorithms that differ by how the velocities are assigned. Trajectories of particle *i* at time *t* can be determined using a Taylor series expansion for a finite time step (δt) of the forward and backward steps (2.7).

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^{2} + \cdots$$
$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^{2} - \cdots$$
(2.7)

Adding the two expansions together yields the Verlet algorithm (2.8), which only uses the position and acceleration to predict the next step. Since it requires a previous step's position, the velocity is not directly involved within the expression, which is needed for the kinetic energy. The velocity can be estimated by the difference between the forward and backward steps.

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \ddot{\mathbf{r}}(t)\delta t^{2} + \cdots$$
$$\dot{\mathbf{r}}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t}$$
(2.8)

The Leap-Frog algorithm (2.9) is a modification of the Verlet algorithm, where the velocities are predicted at each half step independently of the positions.

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}\left(t+\frac{1}{2}\delta t\right)\delta t$$
$$\dot{\mathbf{r}}\left(t+\frac{1}{2}\delta t\right) = \dot{\mathbf{r}}\left(t-\frac{1}{2}\delta t\right) + \ddot{\mathbf{r}}(t)\delta t$$
(2.9)

The Velocity-Verlet algorithm (2.10) is a more precise numerical algorithm than the Leap-Frog algorithm, as within its formulation the acceleration, velocities, and positions are predicted simultaneously.

$$\dot{\mathbf{r}}\left(t+\frac{1}{2}\delta t\right) = \dot{\mathbf{r}}(t) + \frac{1}{2}\left(\ddot{\mathbf{r}}(t) + \ddot{\mathbf{r}}(t+\delta t)\right)\delta t$$
(2.10)

The Beeman's algorithm is similar to the Velocity-Verlet algorithm, but has a more complex expression for the Taylor expansion that is advantageous for infinitesimal time steps (see Ref 13).

The motions and interactions between particles are modeled using an empirical potential energy function (2.11), better known as a force field, which is represented as the sum of bonding and nonbonding energies.^{14–16}

$$V(\mathbf{r_1}, \dots, \mathbf{r_N}) = \sum_{bonds} K_i^b (b_i - b_{eq})^2 + \sum_{angles} K_i^\theta (\theta_i - \theta_{eq})^2 + \sum_{torsions} K_i^\phi (1 + \cos(\eta\phi_i - \delta_i))$$
$$+ \sum_{impropers} K_i^\phi (\varphi_i - \varphi_{eq})^2 + \underbrace{\sum_{i} \sum_{j \neq i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}\right)}_{nonbond} + \sum_{i} \sum_{j \neq i} \frac{q_i q_j}{\varepsilon r_{ij}}$$

(2.11)

The first four terms of Eq. 2.11 represent the energy of bond stretching (*b*), angle bending (Θ) , proper (ϕ) and improper torsion (ϕ). The potential utilizes the force constants, K_b , K_{Θ} , K_{ϕ} , and K_{ϕ} , for the bonds, angles, proper and improper torsions between atoms, respectively, exhibited at a configuration and reference-equilibrium position to characterize the potential. Van der Waals interactions between particles *i* and *j* are described by the Lennard-Jones potential separated by a distance *r*, in which *A* and *B* are constants that control the interatomic distances (well-depth) for the repulsive and attractive terms, respectively. The electrostatic interactions are defined by partial charges, *q*, of the two particles in a medium, ε , described by the effective dielectric constant, ε .

Despite not being able to account for excited electronic states, quantum effects, or chemical bond breakage breaking and formation, classical MD simulations give insight into many biological problems of both chemical and physical interests. For example, biological macromolecules such as proteins, nucleic acids, and lipids are naturally dynamic systems in which structure, function, and activity are modulated by intrinsic forces.

While internal motions can be observed using spectroscopic methods like infrared spectroscopy and nuclear magnetic resonance spectroscopy, such investigations can become expensive and time intensive for complex systems. MD simulations clarify time-dependent attributes and dynamic processes, allowing for an in-depth understanding and explanation of biophysical behaviors and chemical properties, such as the free energy contribution to structural rearrangements.^{17–19}

MD studies can aid in defining the role of the motifs in protein-protein interactions and protein-ligand recognition, which are capable of many different types of interactions including van der Waals interactions through aliphatic chains, cation- π interactions with aromatics, hydrogen bonding and purely electrostatic interactions with solvent.²⁰ For instance, solvent effects, structure, stacking interactions between aromatic residues and the strength of ion binding can all be correlated to biological activity to elucidate mechanisms in protein recognition of other molecules.

2.2.1 Considerations

There are many parameters needed to carry out a simulation. As improper set up of a simulation can generate trajectories that have no basis in reality, hence much detail should be taken into account when employing MD simulations.²¹

For biomolecular systems, many force fields have been developed.^{16,22} Popular force fields, such as OPLS,²³ GROMOS,²⁴ CHARMM,²⁵ and AMBER,²⁶ are parameterized based on a mixture of empirical data and quantum mechanical calculations. Each force field differs by how a parameter is defined per atom (atom-type). While harmonic potentials for bond stretching and angle bending are typically fitted to experimental vibrational data (normal mode frequencies), the

force fields differ by how the in-plane and out-of-plane torsion and non-bonded terms have been parameterized and the data sets to which the parameters have been fitted.

Partial atomic charges, although not a physically measurable property, have a critical impact on interaction energies, hence effective partial charges impact the bonding terms within the force field.²⁷ For example, partial atomic charges in the CHARMM force field are parameterized based upon Mulliken population analysis while the AMBER force field uses the Restrained Electrostatic Potential (RESP)²⁸ approach.

The environment of the solute is critical in the determination of the properties and reactivity of biomolecular systems.^{29,30} For MD simulations, the solvent may be treated with an implicit water model, such as the Generalized Born model,^{31,32} or explicitly using rigid or flexible water molecules.³³ Using an explicit water solvent is advantageous because the solute can relax and interact with the solvent. Many empirical water models have been developed over the years to reproduce bulk water properties,^{34,35} including the TIP3P³⁶ (transferable intermolecular potential 3 point), several variants of TIP4P^{37–43} and TIP5P,^{36,44} the SPC⁴⁵ (simple point charge) and the SPC/E⁴⁶ (extended simple point charge) water models.

Periodic boundary conditions (PBCs) are used to represent the simulation as a continuous system.²¹ Under PBCs, if an atom were to leave the unit cell in one direction, its image would reenter the cell on the other side. Imposing PBCs avoids surface artifacts that would arise from the atoms' unrealistic interactions with the unit cell surface.

Statistical mechanics is fundamental to MD simulations as it allows for macroscopic properties (i.e. free energy) to be determined from properties of microstates.^{47,48} For a given microstate (N) characterized at position \mathbf{r} , momentum \mathbf{p} , and time t, the instantaneous value of

property A will fluctuate over time as a result of the interacting particles (2.12). In experiment, the value reflects an averaged value of the property over the time of the measurement.

With the assumption that all accessible microstates are attainable and equally probable, the calculated value of *A* approaches the true average value of the property as time evolves. The average value \overline{A} (2.13) is expressed as a time average over infinite time. For a finite time period, the average value is expressed as the ensemble average (2.14), in which the probability ρ of the state is proportional to the total energy (E), where Boltzmann's constant (k_B) and temperature (T).

$$A(\mathbf{p}^N(t),\mathbf{r}^N(t))$$

(2.12)

$$\bar{A} = \lim_{\tau \to \infty} \frac{1}{t} \int_{t=0}^{\tau} A(\mathbf{p}^{N}(t), \mathbf{r}^{N}(t)) dt$$

$$(2.13)$$

$$\langle A \rangle = \iint d\mathbf{p}^{N} d\mathbf{r}^{N} A(\mathbf{p}^{N}, \mathbf{r}^{N}) \rho(\mathbf{p}^{N}, \mathbf{r}^{N})$$

$$\rho(\mathbf{p}^{N}, \mathbf{r}^{N}) \propto e^{-\frac{E(\mathbf{p}^{N}, \mathbf{r}^{N})}{k_{\mathrm{B}}T}}$$

(2.14)

Fundamental internal motions, such as molecular vibrations, rotations, and torsions occur in the femtosecond, picosecond, and nanosecond timescales, respectively, while larger motions such as allosteric transitions and large-scale conformational changes, are observed on much larger time scales.⁴⁹ In order to observe the molecular vibrational modes, the time step used in the numerical integration scheme must be smaller than the period of the fastest oscillation. Constraint algorithms, such as SETTLE,⁵⁰ SHAKE,⁵¹ EEM,⁵² SHAPE,⁵³ and LINCS,⁵⁴ are used to freeze high frequency motions that have minimal effects on the overall dynamics of the system by placing algebraic constraints on bonds or angles to an equilibrium distance, allowing for larger time steps to be used.

For example, the SHAKE algorithm⁵¹ can be computationally beneficial when a simulation of a large system where a molecule is surrounded by bulk solvent (e.g. explicit water) because bond stretching motions (e.g. O-H stretching) can be constrained. SHAKE is applied to the Verlet algorithm by defining pairwise bond distance constraints between atoms and is used to calculate the force constants iteratively, allowing the use of a larger time step.

2.2.2 Limitations

A major limitation to the computational modeling of the chemistry of large biological systems using atomistic MD simulations is the inadequate sampling of conformational states that are computationally feasible. In order to circumvent the time-scale limitation of atomistic MD simulations, alternative potentials and protocols, such as coarse-grained and continuum modeling, have been developed to offer a reduced representation that allows for more conformational state sampling than atomistic simulations.^{55–57} For example, in a coarse-grained model for a protein, the amino acid lysine can be reduced from 24 atoms to three pseudo atoms that represent the polar backbone, the apolar hydrocarbons, and the charged amino group. Together, utilizing both atomistic and coarse-grained simulations techniques, many rare events can be uncovered.

2.3 Quantum Mechanics

While classical mechanics is used to study large systems, containing many microscopic particles, quantum mechanics explains the motion of electrons.⁵⁸ In quantum mechanics, the state of a system is described by a continuous wave function (Ψ) and can be used with an operator (\hat{O}) to determine the expectation value (EO) which can correspond to physical observables, such as position, momentum, or energy.

$$\langle EO \rangle = \frac{\int \Psi^* \hat{O} \Psi \, d\tau}{\int \Psi^* \Psi \, d\tau}$$
(2.15)

The quantum mechanical Hamiltonian operator (2.16) is similar to the classical Hamiltonian, where \hat{T} is the kinetic energy operator that is dependent on the momentum and \hat{V} is the potential energy operator that is dependent on a position. Within the kinetic energy operator, the momentum operator \hat{p} is adapted to wave mechanics, where *i* is an imaginary unit equal to $\sqrt{-1}$, \hbar is the reduced Planck constant, and *m* is mass.

$$H = T + V$$
$$\hat{T} = \frac{(\hat{p})^2}{2m} = \frac{-\hbar^2}{2m} \nabla^2 \qquad \hat{p} = i\hbar \frac{\partial}{\partial t}$$

(2.16)

The non-relativistic time-dependent Schrödinger equation (2.17a) describes the dynamics of a system given by the wave function, however for many applications in chemistry, the non-relativistic time-independent Schrödinger equation is used for obtaining the total energy of chemical systems (2.17b).

$$-\frac{\hbar}{i}\frac{\partial}{\partial t}\Psi(r,R,t) = \widehat{H}\Psi(r,R,t)$$

$$\widehat{H}\Psi(r,R) = E\Psi(r,R)$$
(2.17a)

(2.17b)

The non-relativistic molecular Hamiltonian is expressed as the sum of kinetic energy and potential energy operators for the electrons (*N*) and the nuclei (*M*), shown in eq. (2.18) where m_A is the mass of nucleus A, Z_A is the atomic number of nucleus A, and *R* and *r* represent the position of the nuclei (*A* and *B*) and electrons (*i* and *j*), respectively. This expression includes the kinetic energy of the electrons (\hat{T}_e) and nuclei (\hat{T}_n), the columbic attraction between the nuclei and electrons (\hat{V}_{ne}), and the repulsion between nuclei (\hat{V}_{nn}) and electrons (\hat{V}_{ee}).

$$\widehat{H} = \underbrace{-\frac{1}{2} \sum_{i=1}^{N} \nabla_{i}^{2}}_{\widehat{r}_{e}} - \underbrace{\frac{1}{2} \sum_{A=1}^{M} \frac{\nabla_{A}^{2}}{m_{A}}}_{\widehat{r}_{n}} - \underbrace{\sum_{A=1}^{M} \sum_{i=1}^{N} \frac{Z_{A}}{|R_{A} - r_{i}|}}_{\widehat{V}_{ne}} + \underbrace{\sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_{i} - r_{j}|}}_{\widehat{V}_{ee}} + \underbrace{\sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_{A}Z_{B}}{|R_{A} - R_{B}|}}_{\widehat{V}_{nn}}$$
(2.18)

As the masses of the nuclei are much larger than the electrons and therefore the nuclei move much slower than the electrons, the motion of the nuclei and the electrons can be separated. This is the Born-Oppenheimer approximation and allows for the separation of the nuclear and electronic components for the wave function, Ψ , as shown in eq. (2.19).

$$\Psi = \psi_n \cdot \psi_e \tag{2.19}$$

Within the Born-Oppenheimer approximation, the molecular Hamiltonian can be simplified to the electronic Hamiltonian (\hat{H}_{el}) where \hat{T}_n is zero and \hat{V}_{nn} is treated as a constant (2.20).⁵⁹ The electronic wave function includes all possible states for electrons for a nucleus.

$$\widehat{H}_{el} = \underbrace{-\frac{1}{2} \sum_{i=1}^{N} \nabla_{i}^{2}}_{\widehat{T}_{e}} - \underbrace{\sum_{A=1}^{M} \sum_{i=1}^{N} \frac{Z_{A}}{|R_{A} - r_{i}|}}_{\widehat{V}_{ne}} + \underbrace{\sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_{i} - r_{j}|}}_{\widehat{V}_{ee}}$$

(2.20)

The Pauli-Exclusion Principle expresses that no two electrons can occupy the same quantum state, meaning that the total wave function must be antisymmetric with respect to exchange of the two electrons. This is accounted for by a linear combination of the possibilities for *N* electrons with respect to the spin (x_N) and spatial relations (χ_N) and is expressed with a Slater determinant as shown in eq. (2.21).

$$\Psi = \chi_{1} (x_{1})\chi_{2}(x_{2}) - \chi_{1}(x_{2})\chi_{2}(x_{1})$$

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_{1}(x_{1}) & \chi_{2}(x_{1}) & \cdots & \chi_{N}(x_{1}) \\ \chi_{1}(x_{2}) & \chi_{2}(x_{2}) & \cdots & \chi_{N}(x_{2}) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{1}(x_{N}) & \chi_{2}(x_{N}) & \cdots & \chi_{N}(x_{N}) \end{vmatrix}$$
(2.21)

2.3.1 *Ab initio* Methods

The Hartree-Fock (HF) approximation⁶⁰ is one of the most commonly utilized electronic structure methods for solving the electronic Schrödinger equation by using a mean field approximation for the two electron term \hat{V}_{ee} in which each electron is treated as interacting with an effective average potential of the other electrons rather than an individual electron. The Fock operator (\hat{f}) is composed of the sum of the one electron terms (2.22), in which the first term represents the kinetic energy of the electron, the second term represents the electron-nuclear attraction, and the third term represents the HF potential, which approximates the two electron term of the Schrödinger equation.

$$\hat{f}_{i} = -\frac{1}{2}\nabla_{i}^{2} - \sum_{j}^{N} \frac{Z_{A}}{|R_{A} - r_{i}|} + v^{HF}(i)$$
(2.22)

Although this method accounts for most of the electronic energy of a system, the HF approximation does not fully describe the wave function because it neglects electron-electron correlation effects for electrons of the opposite spin, thus cannot describe dynamic correlation (the instantaneous repulsion of moving electrons) or static correlation (nearly degenerate electron configurations)—which is significant for understanding chemical phenomena.⁶ Post-Hartree-Fock methods have been developed to recover electron correlation using a single slater determinant, including configuration interaction (CI),^{6.59} Møller-Plesset (MP*n*) perturbation theory (e.g. MP2 and MP4),⁶¹⁻⁶⁵ and coupled cluster (CC)⁶⁶⁻⁶⁹ methods.

Full CI includes all possible configurations or excitations of electrons (2.23), where a_i is the coefficient corresponding to unique single (S), double (D), triple (T), quadruple (Q), and *i*-th level excitations, respectively, that are added to the HF reference (Ψ_0).

$$\Psi_{\rm CI} = a_0 \Phi_{\rm HF} + \sum_{\rm S} a_{\rm S} \Phi_{\rm S} + \sum_{\rm D} a_{\rm D} \Phi_{\rm D} + \sum_{\rm T} a_{\rm T} \Phi_{\rm T} + \sum_{\rm Q} a_{\rm Q} \Phi_{\rm Q} + \dots = \sum_{i=0}^{N} a_i \Phi_i$$
(2.23)

When truncated, CI is still variational but is no longer size extensive, or linearly scaled with the electron count and the theory is no longer exact. In CC methods, the wave function uses an exponential form ansatz that contains connected excitation operators (\hat{T}) to approximate the true wave function, in which each Nth excitation includes previous excitations (2.24).⁷⁰ By formulation, CC methods are size extensive, but no longer variational. Similar to CI methods, increasing the number of excitations in CC yields (e.g. CCS, CCSD, CCSDT, etc.) approach the

exact energy. While CI and CC are rigorous methods for recovering electron-electron correlation, they are both computationally expensive.

$$\Psi_{CC} = e^{T} \Phi_{0}$$

$$\hat{T} = \hat{T}_{1} + \hat{T}_{2} + \hat{T}_{3} + \hat{T}_{4} + \dots + \hat{T}_{N}$$
(2.24)

A popular CC method is CCSD(T), which includes single (S) and double (D) excitations and perturbative triples (T).

MP*n* methods use a reference, unperturbed Hamiltonian (\hat{H}_0), which is the sum of the oneelectron Fock operators, and a perturbation (\hat{V}) on the reference Hamiltonian to recover electron correlation, where λ is the expansion parameter that determines the strength of the perturbation (2.25).

$$\widehat{H} = \widehat{H}_0 + \lambda \widehat{V}$$
(2.25)

2.3.2 Density Functional Theory

Density functional theory (DFT) is an alternative to wave function-based methods that relates the electron density (ρ) to the ground state energy of a system through the Hohenberg-Kohn and Kohn-Sham theorems.^{71–73}

$$E_{DFT}[\rho] = T_s[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho]$$

$$\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2$$
(2.26)

In Kohn-Sham theory, (2.26), $T_s[\rho]$ is the kinetic energy for the non-interacting reference system, $E_{ne}[\rho]$ is the columbic nuclear-electron attraction energy, $J[\rho]$ is the electron-electron repulsion energy, and $E_{xc}[\rho]$ is the remaining energy termed the exchange-correlation functional. The exact functional form of the exchange-correlation term is unknown. DFT functionals have been developed that attempt to approximate this term, such as local density approximation (LDA), generalized gradient approximation (GGA), meta-GGA, and hybrid-GGA.^{74,75} The LDA functionals are only dependent on the stationary density whereas the GGA functionals are based on the change in electron density. Meta-GGA functionals include corrections for kinetic energy density. Hybrid functionals incorporate a mixture of exact exchange from HF with other DFT exchange-correlation approximations. A popular DFT functional used in this work is B3LYP, which is a hybrid-GGA functional optimized to contain 20% of HF exchange.^{76–78} Although popular, B3LYP is not the best functional. There are many DFT functionals which can outperform one another, however the superiority is often dependent upon the problem of interest.^{79,80} For example, the B3LYP functional is good for predicting the structure of organic molecules.

2.3.3 Basis Sets

Electronic structure methods use basis sets, which are mathematical representations used to describe the wave function. The Slater-type orbital (STO) is one of the earlier basis functions to model atomic orbitals similar to the hydrogen atom (2.27), where *N* is a normalization constant, $Y_{l,m}$ represents the angular function, *n* is the principle quantum number, and ζ is exponential parameter that determines radial distance *r*. The spherical harmonic is defined by the orbital angular momentum (*l*) and magnetic quantum number (*m*).

$$\psi(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{(n-1)}e^{-\zeta r}$$
(2.27)

However, STOs are difficult to evaluate numerically. Gaussian-type orbitals (GTO), on the other hand, have a more advantageous mathematical form (2.28), however GTOs poorly represent the

atomic orbital. Rather, linear combinations of multiple GTOs are taken to better approximate STOs. A solution was to combine GTOs to better approximate STOs, which in general, improve by the addition of basis functions.

$$\psi(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{(2n-2-l)}e^{-\zeta r^2}$$
(2.28)

There are many different families of atomic orbital basis sets. Two of the more popular families are the correlation consistent sets of Dunning and co-workers^{81–85} and the Pople-style sets.^{86–89} For the Pople-style basis sets, there are many different levels of complexity, beginning with a minimal basis such as STO-3G⁸⁶ and increasing to split valence sets such as 3-21G⁸⁷ and 6-31G.⁸⁸ Additional basis functions, such as diffuse and polarization functions can be added to the basis set to give a more complete representation of the orbital space.

The correlation consistent basis sets were designed to systematically recover correlation energy with the addition of basis functions. These basis sets are noted correlation consistent, polarized, valence, n- ζ level, or cc-pVnZ, where n is equal to double- ζ (DZ), triple- ζ (TZ), quadruple- ζ (QZ), etc. An advantage of the correlation consistent basis sets is that they can be used to extrapolate to the complete basis limit. REFERENCES

REFERENCES

- (1) Goldstein, H.; Poole, C. P.; Safko, J. L. *Classical Mechanics*, 3. ed.; Addison Wesley: San Francisco, NJ, 2002.
- (2) Chaichian, M.; Merches, I.; Tureanu, A. *Mechanics an Intensive Course*; Springer: Berlin; New York, 2012.
- (3) Fowles, G. R.; Cassiday, G. L. *Analytical Mechanics*, 7th ed.; Thomson Brooks/Cole: Belmont, CA, 2005.
- (4) Haile, J. Molecular Dynamics Simulation: Elementary Methods. **1992**, 512.
- (5) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models, 2nd Ed.*, 2nd ed.; John Wiley and Sons: West Sussex, England, 2004.
- (6) Jensen, F. *Introduction to Computational Chemistry, 2nd Ed.*; John Wiley & Sons, Ltd: West Sussex, England, 2007.
- van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. a; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. a; Yu, H. B. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chem. Int. Ed. Engl.* 2006, *45* (25), 4064–4092.
- (8) van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. Molecular Mechanics in Biology: From Structure to Function, Taking Account of Solvation. *Annu. Rev. Biophys. Biomol. Struct.* 1994, 23, 847–863.
- (9) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106* (5), 1589–1615.
- (10) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98–103.
- (11) Cuendet, M. A.; van Gunsteren, W. F. On the Calculation of Velocity-Dependent Properties in Molecular Dynamics Simulations Using the Leapfrog Integration Algorithm. *J. Chem. Phys.* 2007, 127 (18), 184102.
- (12) Swope, W. C. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76* (1), 637.
- (13) Beeman, D. Some Multistep Methods for Use in Molecular Dynamics Calculations. J. Comput. Phys. 1976, 20 (2), 130–139.
- (14) Wang, J.; Hou, T.; Xu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput. Aided. Drug Des.* **2006**, *2* (3), 287–306.

- (15) Ponder, J. W.; Case, D. a. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66* (Protein Simulations), 27–85.
- (16) González, M. A. Force Fields and Molecular Dynamics Simulations. *École thématique la Société Française la Neutron.* **2011**, *12*, 169–200.
- (17) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652.
- (18) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **1990**, *347* (6294), 631–639.
- (19) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci.* U. S. A. 2005, 102 (19), 6679–6685.
- (20) 3rd, T. E. C.; Young, M. A. Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers* **2000**, *56* (4), 232–256.
- (21) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2002**, *12* (2), 190–196.
- (22) Mackerell, A. D. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25* (13), 1584–1604.
- (23) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc. 1996, 118 (45), 11225–11236.
- (24) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. J. Phys. Chem. A **1999**, 103 (19), 3596–3607.
- (25) MacKerell, a D.; Banavali, N.; Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2001**, *56* (4), 257–265.
- (26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. a. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
- (27) Meister, J.; Schwarz, W. H. E. Principal Components of Ionicity. J. Phys. Chem. 1994, 98 (33), 8245–8252.
- (28) Wang, J.; Cieplak, P.; Kollman, P. a. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? J. Comput. Chem. 2000, 21 (12), 1049–1074.
- (29) Davis, M. E.; McCammon, J. A. Electrostatics in Biomolecular Structure and Dynamics.

Chem. Rev. **1990**, *90* (3), 509–521.

- (30) Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108* (1), 74–108.
- (31) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.* **1996**, *100* (5), 1578–1599.
- (32) Calimet, N.; Schaefer, M.; Simonson, T. Protein Molecular Dynamics with the Generalized born/ACE Solvent Model. *Proteins Struct. Funct. Genet.* 2001, 45 (2), 144– 158.
- (33) Chen, J.; Brooks, C. L.; Khandogin, J. Recent Advances in Implicit Solvent-Based Methods for Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 140–148.
- (34) Wallqvist, A.; Mountain, R. D. Molecular Models of Water: Derivation and Description; Lipkowitz, K. B., Boyd, D. B., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1999; Vol. 13, pp 183–247.
- (35) Ouyang, J. F.; Bettens, R. P. A. Modelling Water: A Lifetime Enigma. *Chim. Int. J. Chem.* 2015, 69 (3), 104–111.
- (36) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. J. Chem. Phys. 2000, 112 (20), 8910.
- (37) Jorgensen, W. L.; Madura, J. D. Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water. *Mol. Phys.* **1985**, *56* (6), 1381–1392.
- (38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.
 Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983, 79 (2), 926.
- (39) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. J. Chem. Phys. 2004, 120 (20), 9665–9678.
- (40) Rick, S. W. Simulations of Ice and Liquid Water over a Range of Temperatures Using the Fluctuating Charge Model. *J. Chem. Phys.* **2001**, *114* (5), 2276–2283.
- (41) Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A Potential Model for the Study of Ices and Amorphous Water: TIP4P/Ice. J. Chem. Phys. 2005, 122 (23), 234511.
- (42) González, M. A.; Abascal, J. L. F. A Flexible Model for Water Based on TIP4P/2005. *J. Chem. Phys.* **2011**, *135* (22), 224516.
- (43) Fuentes-Azcatl, R.; Barbosa, M. C. Thermodynamic and Dynamic Anomalous Behavior in the TIP4P/ε Water Model. *Phys. A Stat. Mech. its Appl.* 2016, 444, 86–94.
- (44) Rick, S. W. A Reoptimization of the Five-Site Water Potential (TIP5P) for Use with Ewald Sums. J. Chem. Phys. 2004, 120 (13), 6085–6093.
- (45) Berendsen, H. J. .; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Springer Netherlands, 1981; Vol. 14, pp 331–342.
- (46) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
- (47) Leach, A. R. Molecular Modelling: Principles and Applications. 2001.
- (48) Allen, M.; Tildesley, D. Computer Simulation of Liquids. 1987.
- (49) McCammon, J. A. Protein Dynamics. *Reports Prog. Phys.* 1984, 47 (1), 1.
- (50) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.
- (51) Ryckaert, J.-P. J.; Ciccotti, G.; Berendsen, H. H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. J. Comput. Phys. 1977, 23 (3), 327–341.
- (52) Edberg, R.; Evans, D. J.; Morriss, G. P. Constrained Molecular Dynamics: Simulations of Liquid Alkanes with a New Algorithm. *J. Chem. Phys.* **1986**, 84 (12), 6933–6939.
- (53) Tao, P.; Wu, X.; Brooks, B. R. Maintain Rigid Structures in Verlet Based Cartesian Molecular Dynamics Simulations. *J. Chem. Phys.* **2012**, *137* (13), 134110.
- (54) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. J. Comput. Chem. 1997, 18 (12), 1463– 1472.
- (55) Tozzini, V. Coarse-Grained Models for Proteins. *Curr. Opin. Struct. Biol.* **2005**, *15* (2), 144–150.
- (56) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On Developing Coarse-Grained Models for Biomolecular Simulation: A Review. *Phys. Chem. Chem. Phys.* 2012, 14 (36), 12423– 12430.
- (57) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annu. Rev. Phys. Chem.* 2011, 62, 41–64.
- (58) Levine, I. N. Quantum Chemistry, 5th ed.; Prentice Hall: Upper Saddle River, N.J, 2000.
- (59) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: Mineola, New York, 1996.

- (60) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Phys.* **1930**, *61* (1–2), 126–148.
- (61) Cremer, D. Møller-Plesset Perturbation Theory: From Small Molecule Methods to Methods for Thousands of Atoms. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2011, 1 (4), 509–530.
- (62) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46* (7), 618–622.
- (63) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153* (6), 503–506.
- (64) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. Direct MP2 Gradient Method. *Chem. Phys. Lett.* **1990**, *166* (Journal Article), 275–280.
- (65) Raghavachari, K.; Pople, J. A.; Replogie, E. S.; Head-Gordon, M. Fifth Order Moller-Plesset Perturbation Theory : Comparison of Exististing Correiation Methods and Implementation of New Methods Correct to Fifth Order. J. Phys. Chem. 1990, 94 (14), 5579.
- (66) Bishop, R. F. An Overview of Coupled Cluster Theory and Its Applications in Physics. *Theor. Chim. Acta* **1991**, *80* (2–3), 95–148.
- (67) Scuseria, G. E.; Schaefer, H. F. Is Coupled Cluster Singles and Doubles (CCSD) More Computationally Intensive than Quadratic Configuration Interaction (QCISD)? J. Chem. Phys. 1989, 90 (7), 3700–3703.
- (68) Bartlett, R. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, 359–401.
- (69) Bartlett, R. J.; Purvis, G. D. Many-Body Perturbation Theory, Coupled-Pair Many-Electron Theory, and the Importance of Quadruple Excitations for the Correlation Problem. *Int. J. Quantum Chem.* **1978**, *14* (5), 561–581.
- (70) Bartlett, R.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79* (1), 291–352.
- (71) Parr, R. G. Density Functional Theory. Annu. Rev. Phys. Chem. 1983, 34 (1), 631–656.
- (72) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. Phys. Rev. 1964, 136, B864.
- (73) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*. 1965, pp A1133–A1138.
- (74) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. *AIP Conf. Proc.* **2001**, *577* (May), 1–20.

- (75) Peverati, R.; Truhlar, D. G. Quest for a Universal Density Functional: The Accuracy of Density Functionals across a Broad Spectrum of Databases in Chemistry and Physics. *Philos. Trans. A. Math. Phys. Eng. Sci.* **2014**, *372* (2011), 20120476.
- (76) Gill, P.; Johnson, B.; Pople, J.; Frisch, M. The Performance of the Becke—Lee—Yang—Parr (B—LYP) Density Functional Theory with Various Basis Sets. *Chem. Phys. Lett.* 1992, *197* (4–5), 499–505.
- (77) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. J. *Chem. Phys.* **1993**, *98*, 5648–5652.
- (78) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789.
- (79) Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* 2011, 13 (14), 6670–6688.
- (80) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J. Chem. Theory Comput.* **2008**, *4* (2), 297–306.
- (81) Jr, T. H. D.; Jr., T. H. D.; Dunning, T. H. J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. J. Chem. Phys. 1989, 90, 1007–1023.
- (82) Dunning, T. H. J.; Peterson, K. A.; Wilson, A. K.; Jr, T. H. D.; Jr., T. H. D. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum through Argon Revisted. J. Chem. Phys. 2001, 114 (21), 9244–9253.
- (83) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. J. Chem. Phys. 1993, 98 (2), 1358.
- (84) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. V. Core-Valence Basis Sets for Boron through Neon. J. Chem. Phys. 1995, 103 (11), 4572.
- (85) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. J. Chem. Phys. 1994, 100 (4), 2975.
- (86) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansion of Slater-Type Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657.
- (87) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbitals Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Moelcular Orbital Studies of Organic Molecules. J. Chem. Phys. 1972, 56, 2257.

- (88) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. J. Am. Chem. Soc. 1980, 102 (3), 939–947.
- (89) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. J. Chem. Phys. 1980, 72 (1), 650.

CHAPTER 3 Molecular Dynamics Studies of the Protein–Protein Interactions in Inhibitor of κB Kinase-β[†]

3.1 Introduction

Impeding activation of the nuclear factor (NF)- κ B pathway has been a popular aim for designing new anti-inflammatory therapeutics for many disease states, including cancer, stroke, and viral infections.¹ Directly involved in regulating immune responses to a variety of stimuli, the NF- κ B complex is comprised of an assembly of five transcription activators² related by the presence of a distinctive Rel homology domain (RHD):^{3,4} RelA (p65), RelB , c-Rel, p50/p105 (NF- κ B1), and p52/p100 (NF- κ B2).⁵⁻⁹ Separated into two classes by C-terminal domains, the NF- κ B subfamily proteins, p50/p105 and p52/p100, are not always transcription activators unless they form dimers with the Rel subfamily proteins, RelA, RelB, and c-Rel.¹⁰⁻¹⁷ The RHD, responsible for dimerization and binding to DNA, contains the nuclear localization sequence (NLS)¹⁸⁻²⁰ and the binding sites for inhibitors.²¹⁻²³

Although there are various ways to signal activation of the NF- κ B pathway, the inhibitory subunits I κ B (I κ B α , I κ B β , and I κ B ϵ)⁵ are responsible for sequestering the inactive state of NF- κ B dimers in the cytoplasm by acting as a steric block against NLS functionality.²⁴ Liberation of NF- κ B complex happens upon signaled-induced degradation of the I κ B complex thus mediating gene transcription henceforth inhibition of I κ B activation is a promising target for anti-inflammatory drug development.²⁴⁻²⁹ Enzymes responsible for this cascade of events have been linked to the

[†] This chapter is presented in its entirety from: Jones, M. R; Liu, C.; Wilson, A. K. "Molecular Dynamics Studies of the Protein–Protein Interactions in Inhibitor of κ B Kinase- β " J. Chem. Inf. Mod. **2014** *54* (2), 562-572.

catalytic subunits of (I κ B kinase) IKK,³⁰⁻³³ specifically activation of the IKK α and IKK β isoforms. Dual phosphorylation of the IKK α (at Ser176 and Ser180) and IKK β (at Ser177 and Ser181) catalytic subunits regulate aberrant signaling pathways by which the activated subunits, both capable of undergoing autophosphorylation, phosphorylate I κ B α at Ser 32 and Ser26 and I κ B β at Ser19 and Ser23.

Both IKK α and IKK β subunits, having a 52% sequence identity, contain a kinase domain (KD), leucine zippers (LZ), and helix-loop-helix (HLH) motifs that are capable of forming both homo and heterodimers.³⁴ Studies in vitro have shown that dimerization mediated by the LZ motifs regulate the activity and activation of KD by mutation of LZ and HLH motifs, although the HLH motif does not affect dimerization, it has been shown to contribute to KD activity.³⁵ Serine to alanine mutations in IKK^β prevent IKK activation, whereas in IKK^α, autophosphorylation is deactivated but has no effect on IKK activity highlighting IKKB as an attractive target. Additionally, in the wild-type conformation of IKK β , the position of the activation loop containing the Ser177 and Ser181 hampers the docking of substrates into the pocket whereas the activated conformation is more accessible as compared to the cyclic AMP-dependent kinase (cAPK).³⁶ This is mainly because the phosphorylation of the activation loop residues causes conformational change of the loop and even surrounding residues. This is an important phenomenon that suggests conformational change induced by these interactions play an important role in bioactivity of the protein and medicinal chemistry. The activation loop also contains Cys179 within the motif that is known to be important for the phosphorylation of both IKKβ and IKKα. Studies have shown that modifications to the cysteine suppress activity of the enzyme and activation of the NF-KB pathway.³⁷⁻⁴⁰

With IKK β being a key enzyme in activating the NF- κ B pathway, the mechanism for inhibiting activation of IKK β has remained elusive due to a lack of a tertiary structural insight. The newly determined crystal structure of Xenopus laevis IKK β (xIKK β), having a gapless 74% sequence identity of human gene (hIKK β), was reported as a "dimer of dimers" in which dimerization was reasoned critical for activation, rather than the activity for each homogenous protomer.⁴¹ Each protomer consists of 647 amino acids composed to form a kinase domain (KD) (residues 16-307), an ubiquitin-like domain (ULD) (residues 310-394), and scaffold/dimerization domain (SDD) (410-666).

The presented structure of the kinase domain in xIKK β does not resemble the active conformation upon comparison of the activation segments in available protein kinase crystal structures.⁴² The activation segments were defined by a primary sequence beginning adjacent to a β sheet at a conserved DFG (Asp-Phe-Gly) triad and extends to a conserved APE (Ala-Pro-Glu) motif near an alpha-helix. Although the random coils share a consistent fold between the two motifs, the activation loop is diverse in both sequence and conformation unique to its protein-protein interactions and regulation. The activation segment in the xIKK β structure begins at Asp166 and ends at Glu192, where Ser177-Ser181 (SXXXS) contain residues shown critical for activation. While the KD contains the mitogen-activated protein (MAP) kinase⁴³⁻⁴⁶ for serine phosphorylation, the SDD regulates activation via maintaining dimerization. Mutagenesis studies showed that the wild-type of IKK β was a dimer whereas the mutant was a monomer with a weaker affinity for dimerization and upon mutation of Leu654 and Trp655 in the SDD, dimerization was lost.⁴¹

Previous computational and experimental work with IKK β and similar Ser/Thr/Tyr kinases has revealed that the allosteric adenosine triphosphate (ATP) binding cavity is near the hinge

connecting the N and C lobes of the kinase domain.^{41,42,47-51} Comparative homology modeling of the monomer of IKK β suggested that the KD was flexible and allosteric ATP binding conformational changes were not easily observed using 80 ns simulations without the crystal structure.^{47,52}

In this study, molecular dynamics (MD) simulations of IKK β are presented offering a greater insight on how conformational changes and protein-protein interactions within a multimeric assembly are influenced by distal modifications in the interacting subunits of IKK β . Using a multiscale approach, coarse-grained modeling offers a reduced representation useful for predicting large-scale conformational changes whereas atomistic simulations provide a more descriptive insight on activity; together this approach helps discern dimerization disruption.

3.2 Computational Methods

System Preparation. The initial coordinates for eight phosphomimetic mutant (S177E/S181E) molecules of xIKKβ were obtained from the Protein Data Bank Database⁵³ (PDBID: 3QA8).⁴¹ Combinations of the eight molecules were constructed to model multiple sets of tetramers, trimers, and dimers while using a monomer as a control (Table 3.1; Figure 3.1). Missing atoms on the side chains of Val79, Asp90, and Asp291 were added using the Molecular Operating Environment v.2011.10 (MOE).⁵⁴ Coarse-grained models (all models shown in Table 3.1) were protonated using MOE and minimized using the AMBER99 force field.^{55,56} Models for the atomistic simulations (AB and AD dimers and monomer) were prepared using the Leap module of AmberTools12.^{57,58} The Site Finder application in MOE was used to investigate potential sites for ligand docking. Contact sites and Propensity for Ligand Binding (PLB) indices⁵⁹ were used to compare both allosteric and non-allosteric ATP binding sites. ATP coordinated with Mg²⁺ ions were extracted from a crystal structure of the Vgrg protein (PDBID:4DTH)⁶⁰ and docked using

MOE with the induced fit method positioning the adenine head towards the hinge (DFG motif) and tail near the phosphoacceptor serines, using the same starting position for all models (Figure 3.2). Ligand parameters were retrieved from the AMBER parameter database⁶¹ and were set up using the GAFF.⁶² Parameters for phosphoserine⁶³ mutants (S177/181pS) were retrieved from Leap.

 Table 3.1 Summary of modeled systems.

Model	Chain
Tetramer	ABCD, DEFG, BCEH
Trimer	ABC, BCE
Dimer	AB, AD, BE
Monomer	A



Figure 3.1 Representation of the models. The crystal structure was solved with 8 identical molecules (a) forming two "dimer of dimers" tetrameric complex (b). Although a dimer is defined as AB form (d), the alternative dimer AD (c) has significance for structure stability.



Figure 3.2 Characterization of ATP/Mg²⁺ docking cavity. The ATP ligand is positioned in the mouth of the activation loop (Top). The adenine head of ATP is pointed towards Asp166 and the phosphate tail is outside of the pocket. The binding pocket is outlined and described by colored circles (purple and green) representing surrounding amino acids (Bottom); the different outlines on the purple circles contrast the different types of polar side chains. Blue and green arrows indicate sidechain and backbone acceptor-donors, respectively. The blue spheres surrounding the phosphate tail represent the pocket exposure, whereas the lighter-blue spheres highlight the receptor's exposure.

Simulation Protocol. In order to observe large, dynamical conformational changes, coarse-grained simulations were carried out for 50 ns at 300 K using CafeMol 2.0⁶⁴ with the AICG2 model accounting for nonlocal intra-chain interaction (electrostatic, repulsion, and hydrophobic interactions). Atomistic simulations were performed for 10 ns at 300 K and 1 atm using AMBER 11 in the presence of 100 mM of sodium chloride⁶⁵ and 10 Å of SPC/E⁶⁶ water beyond the solute in a cuboid box under the Amber99SB⁶⁷ force field. Prior to simulation, each solvated system was gradually relaxed under the NVT ensemble by six minimization procedures (500 cycles of steepest descent minimization followed by 500 cycles of conjugate gradient minimization) with decreasing restraints on the protein of 500.0, 200.0, 20.0, 10.0, and 5.0 kcal/mol-Å² per procedure to no restrains and then heated to 300 K over 100 ps. The baths were maintained using the isotropic position scaling protocol and the Langevin thermostat with the SHAKE algorithm.

Trajectory Analysis. Trajectory analysis was performed using Cafemol and AMBER 11 for coarse-grained and atomistic simulations, respectively, and visualized in Chimera⁶⁸ and VMD.⁶⁹

Simulation Stability. Simulation parameters were extracted from the trajectories and plotted to access atomistic simulation stability. RMSDs from the starting structure were performed using the PTRAJ module of Amber suite to investigate flexibility and conformational changes.

Hydrogen Bonding Analysis. Hydrogen bonding analysis was performed with the CPPTRAJ module of AmberTools, where a hydrogen bond was defined by a cutoff distance of 3.5 Å and a donor-hydrogen-acceptor angle of $180^{\circ}\pm60^{\circ}$.

Energetic Analysis. Free energy calculations were carried out using the MMPBSA.py⁷⁰ module with Amber Tools and Amber 11. The electrostatic solvation energy was calculated for

every ps over a span of 10 ns using three GB models: the pairwise GB models of Hawkins and coworkers (GB-HCT)⁷¹⁻⁷³ two modified models developed by Onufriev and coworkers (GB-OBC1 and GB-OBC2)⁷⁴, setting dielectric constants to 1 and 80 for the interior and exterior of the molecule, respectively. Because relative free energy trends were of interest, solute entropy was neglected.⁷⁵ Additionally, the PISA (protein interfaces, surfaces, and assemblies) procedure⁷⁶ was used to predict the free energy of formation and dissociation of the two dimeric assemblies studied.

3.3 Results

3.3.1 Docking is Aesthetic

Potential binding sites for ligands were investigated using the Site Finder application in MOE to observe the accessible binding sites from the static tetrameric assembly (A, B, C and D stand for the four subunits of IKK β , respectively). Approximately 46 accessible cavities were found in a single protomer; only four of the predicted sites have contacts with the activation segment and are listed in Table 3.2. Sites 2-3, 6 and 10 have pockets containing 103-270 contact atoms and a wide range of PLB scores. PLB indices are ranked on concavity in which the highest PLB index would be considered to be the most probable ligand binding site.⁵⁹ The largest pocket is shown in Site 2 (Figure 3.3). It is stationed at the bottom of the kinase domain in a cavity that includes the APE motif as contact atoms. Deeper into the pocket of the activation loop, Sites 3 and 6 both reach towards the hinge of the N and C terminus and include the DFG triad. Both sites are within a reasonable proximity (\leq 5.0 Å) of the SXXXS sequence. Site 10 is also in the hinge region and corresponds to an allosteric binding region. The only adjacent contact atom pertinent to the activation segment is Asp166 of the DFG triad. These accessible pockets within the structure led rise to the proposal of a non-allosteric docking pose for ATP.

	α-	Contact	Amino	PLB	
Site	Spheres	Atoms	Acids	Index	Residues
2	118	270	171	2.37	LEU186, LEU189, ALA190, PRO191,
					GLU192, LEU193, TRP206, PRO221,
					PHE222, ASN225, GLN227, GLN230,
					TRP231, HIS232, GLY233, LYS234, VAL235,
					ILE243, VAL244, VAL245, TYR246, ASP247,
					ASP248, LEU249, THR250, VAL253,
					PHE255, SER256, SER257, LEU277,
					GLN278, LEU281, MET282, TRP283
3	87	200	122	1.53	GLY22, THR23, GLY24, GLY25, ARG144,
					ASP145, LEU146, LYS147, GLU149,
					ASP166, LEU167, TYR169, THR180,
					PHE182, VAL183, GLY184, THR185,
					LEU186, GLN187, TYR188, LEU189,
					GLU192, LEU194, GLU195, TYR199,
					SER207
6	35	115	76	0.64	GLU61, ILE64, MET65, LYS66, LEU68,
					ASN69, VAL73, VAL74, SER75, MET96,
					TYR135, LEU136, ILE141, HIS143, ILE164,
					ILE165, LEU167, GLY168, ALA170
10	40	103	68	0.32	LEU21, GLY22, VAL29, TYR98, CYS99,
					GLU100, GLY101, GLY102, ASP103,
					LYS106, GLU149, ASN150, VAL152,
					ILE165, ASP166

Table 3.2 Binding cavities near the activation segment.

Column 2 characterizes the number of alpha spheres identified within the cavity. Quantities in columns 3 and 4 denote the number of contact atoms and amino acids surrounding the cavity. The PLB index ranks the concavity in an increasing order.



Figure 3.3 Visualization of the predicted cavities. All alpha-spheres are shown for the four sites (Top). The red and white spheres indicated hydrophilic and hydrophobic cavities, respectively. Wireframe representations of the four sites are shown individually: Site 2 (middle left); Site 3 (middle right); Site 6 (bottom left); and Site 10 (bottom right).

3.3.2 Dimer Stability of the Crystal System

Using the static conformations of the initial ligand-free dimer models, theoretical predictions on the dimer assemblies were calculated using PISA. The free energy of binding, ΔG^{int} , is used to describe the free energy of interface formation between subunits, whereas the free energy of change upon dissociation, ΔG^{diss} , corresponds to the difference between the associated and dissociated structures.^{76, 77} In these approximations, a negative ΔG^{diss} value indicates that the structure is thermodynamically unstable and would readily dissociate and a positive value suggests a more stable assembly in which an external component would need to be added to induce dissociation. The results shown in Table 3.3 Summary of dimer assembly energetics, determined with PISA.indicate that the AB dimer is more tightly bound than the AD dimer for the WT and S177/188E models. In contrast, the S177/181pS model has a weaker binding free energy in the AB dimer and a stronger binding affinity in the AD dimer. Additionally, both the WT and S177/188E models have larger ΔG^{diss} values than the S177/181pS model in the AB dimer but smaller values in the AD dimer. This indicates that the S177/181pS model is less thermodynamically stable in the AB dimer and more thermodynamically stable in the AD dimer than the WT and S177/188E models. Although predictions for the WT agree with experimental observations, identical values for the WT and S177/181E models were predicted for both the AB and AD dimers highlighting an inherent weakness in the parameterization of the PISA procedure.⁷⁶ This method may not always be useful for predicting the conformational stability for proteinprotein interactions using static structures.

			AB dime	er				AD dim	er	
Model	A _{BSA}	ΔG^{int}	ΔG^{diss}	$T\Delta S^{\text{diss}}$	$\mathbf{P}_{\mathbf{v}}$	A _{BSA}	ΔG^{int}	ΔG^{diss}	$T\Delta S^{\text{diss}}$	P_{v}
WT	1614.2	-21.9	9.5	15.7	0.061	440.0	1.3	-14.9	15.6	0.801
S177/188E	1614.2	-21.9	9.5	15.7	0.061	440.0	1.3	-14.9	15.6	0.801
S177/188pS	1783.7	-14.9	1.5	17.2	0.207	615.8	-3.0	-12.1	17.1	0.491

Table 3.3 Summary of dimer assembly energetics, determined with PISA.

The buried surface between the engaged interfaces (ABSA) is reported in Å².The P-value (P_v) indicates the probability of finding a more hydrophobic interface, eg. Lower Pv values indicate fewer hydrophobic sites. The dissociation free energies (ΔG^{diss}) and binding free energies (ΔG^{int}) are reported in kcal/mol.

3.3.3 Coarse-Grained Simulations

All coarse-grained models were simulated for 50 ns to gain qualitative insight about the significance of simulating multiple protomers that may contribute to the conformational stability of the xIKK^β assembly. The root-mean-square deviation (RMSD) calculated from the starting structure was compared against the multiple models shown in Table 3.1 and scaled to fit a monomer (Figure 3.4). Although calculating RMSD from the starting structure is one way to determine the stability of the structure, the dynamics in flexible regions can yield misleading RMSDs. Each model exhibits continuous increases in the deviation indicating sustained conformational changes. In these simulations, the continuous structural changes occurred in the kinase domain. Qualitatively, throughout the simulation, the activation loop of the kinase domain periodically opened and closed suggesting that its structural flexibility may play a role in regulating molecular traffic in the activation loop. In terms of comparing conformational stability among the different models, it is apparent within the first 25 ns that there is conformational favorability, specifically fewer fluctuations in the assembly, for models containing multiple protomers. The monomer was the least stable system in these simulations because it began to denature within the first 25 ns of the simulation, thus RMSD after 25 ns were ignored. Although coarse-grained timescales cannot be directly compared with atomistic time scales, this

demonstrates the influence that the assembly of a complex has on the mobility of individual protomers on short time scales.



ABCD — ABC — AB — A — BCEH — BCE — BE — DEFG Figure 3.4 RMSD plots for the different assemblies using coarse-graining MD. Frames were plotted for every 1 ps for a total of 50,000 frames (50 ns).

3.3.4 Atomistic Simulation of Dimers and Monomer

To characterize the significance and relationship between kinase domain activity and dimerization, atomistic simulations were carried out for 10 ns for two sets of dimers while the monomer was observed as a control. Simulation stability was characterized by monitoring thermodynamic parameters (Table 3.4) from the starting and average structures. Average total

energies, densities, and temperatures were compared amongst the different morphs of the monomer and dimers. Properties of each monomer, AB dimer and AD dimer were comparable among the WT, S177/181E, WT+ATP and S177/181pS models, indicating that each monomer and dimers structures, respectively, maintained the defined simulation parameters. The RMSD values for the twelve starting structures shown in Figure 3.5 reveal variations of conformational flexibility throughout the 10 ns simulations. For the monomer, the WT has much more structural fluctuations than the mutant and ATP bound models. For the dimers, the WT, WT+ATP and S177/181pS models have more comparable RMSD values than the S177/181E model. In the WT+ATP model, the flexibility of the activation loop is minimized due to electrostatic interactions with the non-allosteric bound ATP. From this data, it is indicative that the flexibility of the activation segment in the kinase domain contributes little towards the larger fluctuations observed in the full structure of the dimers indicating that the observed structural changes occur in other parts of the complexes.

	Model	Etot	σ	rho	σ	Т	σ
	Monomer	-371875.0	398.6	1.0362	0.0008	300.0	0.9
WT	Dimer:AB	-448514.1	528.9	1.0530	0.0009	300.0	0.8
	Dimer:AD	-880796.6	651.6	1.0222	0.0005	300.0	0.6
	Monomer	-371885.3	400.0	1.0362	0.0008	300.0	0.9
S177/188E	Dimer:AB	-460262.3	531.5	1.0513	0.0009	300.0	0.8
	Dimer:AD	-880880.4	652.1	1.0221	0.0005	300.0	0.6
	Monomer	-373339.9	394.6	1.0364	0.0007	300.0	0.9
WT-ATP	Dimer:AB	-448774.7	524.0	1.0528	0.0010	300.0	0.8
	Dimer:AD	-881975.2	659.6	1.0226	0.0005	300.0	0.6
	Monomer	-372369.3	396.6	1.0363	0.0007	300.0	0.9
S177/181pS	Dimer:AB	-447724.1	519.1	1.0521	0.0009	300.0	0.8
	Dimer:AD	-881267.0	657.1	1.0221	0.0005	300.0	0.6

Table 3.4 Averaged thermodynamic data for the 10 ns atomistic simulation.

The total energy (E_{tot}), density (rho) and temperature (T) per system are reported in units of kcal/mol, (g/mL) and Kelvin, respectively next to the standard deviation (σ).



Figure 3.5 The RMSD from the starting structure for the 10 ns simulation. The different RMSDs of the monomer (a-b), AB dimer (c-d) and AD dimer (e-f) are colored accordingly (blue, WT; red, S177/181E; green, WT+ATP; purple;S177/181pS). Plots A, C, and E represent the RMSD of the entire structure whereas the plots B, D, and F represent only the activation segment of the principle monomeric chain. Frames were plotted for every 2 ps for a total of 5,000 frames (10 ns).

3.3.5 Hydrogen Bonding Between Dimer Interfaces

Structure-based mutagenesis and ultracentrifugation studies investigating SDD mediated dimerization established that the wild-type is a dimer, whereas upon mutation of Leu654, Trp655, and Leu658 dimerization failed. To quantify the loss in dimerization upon activation of the complex, a hydrogen bond analysis was performed. Hydrogen bonding patterns in the AB dimer includes series of both charged and uncharged amino acids including Glu, Asp, Gln, Ser, and Leu. Activation by both phosphomimetic mutation and phosphorylation lead to a decrease in hydrogen bond occupancy throughout the simulation. Specifically, Glu576, Ser489, Asp493, Gln478, Glu565 and Leu658 of one protomer maintain hydrogen bonds with the opposing protomer 50% to 100% of the simulation for the wild-type; however these hydrogen bond occupancies decrease in the activated species (Table 3.5). Trends for the AD dimer (Table 3.6), however, do not follow a similar trend, which is due to the distances between the dimer interface. Occupancies for the WT+ATP models failed to follow a trend readily comparable with the other models; however, it is clear that the docking of ATP has the ability to disrupt SDD mediated dimerization.

(AB) WT			((AB) S177/181E			
Y	X-H	0	Ŋ	ζ	X-H	0	
Glu 576	Arg 573	99	Glu	576	Arg 573	71	
Ser 489	Gln 651	96	Asp	569	Arg 572	68	
Glu 576	Arg 572	89	Ser	489	Gln 651	57	
Tyr 497	Gln 651	81	Asp	493	Gln 651	53	
Asp 546	Arg 460	75	Asp	493	Gln 652	50	
Asp 493	Gln 652	63	Ser	489	Gln 652	45	
Gln 478	Gln 478	55	Glu	19	Arg 592	39	
Glu 565	Arg 438	52	Phe	485	Phe 485	29	
Leu 658	Leu 658	49	Gln	110	Gln 566	28	
Trp 655	Gln 500	39	Glu	502	Arg 666	25	
Arg 573	Arg 572	36	Lys	482	Phe 485	22	
Glu 464	Thr 463	36	Glu	19	Asn 588	21	
Trp 655	Tyr 497	34	Gln	548	Gln 548	21	
Lys 659	Gly 504	29	Gly	504	Trp 655	21	
Arg 573	Arg 573	28	Ala	481	Lys 482	19	
Glu 648	Ser 489	27	Phe	503	Arg 666	18	
Pro 578	Glu 576	26	Glu	19	Arg 419	17	
Lys 467	Lys 467	26	Trp	655	Lys 496	17	
Asp 569	Arg 572	26	Gln	548	Arg 452	16	
Cys 662	Phe 503	25	Glu	648	Phe 485	16	
Glu 502	Ser 653	24	Ser	619	Phe 503	15	
Gln 500	Trp 655	23	Lys	482	Lys 482	15	
Lys 482	Phe 485	18	Gln	110	Leu 570	15	
Ala 481	Lys 482	18	Gln	478	Gln 478	14	
Lys 482	Ala 481	17	Gly	504	Lys 659	13	

Table 3.5 Hydrogen bonding occupancies for the AB dimer interface.

Table 3.5 (cont'd)

(AB) WT-ATP			(AB) \$	(AB) S177/181pS		
Y	X-H	0	Y	X-H	0	
Arg 650	Gln 651	84	Glu 565	Arg 568	70	
Asp 546	Arg 460	71	Lys 482	Lys 482	59	
Trp 655	Tyr 497	50	Lys 482	Phe 485	53	
Gln 478	Gln 478	36	Asp 493	Gln 652	45	
Ser 489	Glu 648	31	Asp 493	Trp 655	36	
Phe 485	Phe 485	29	Phe 485	Phe 485	33	
Phe 485	Gln 647	26	Gln 651	Gln 651	31	
Gln 500	Lys 659	26	Glu 648	Ser 488	30	
Phe 486	Gln 651	22	Phe 485	Lys 482	29	
Ala 481	Lys 482	22	Trp 655	Tyr 497	28	
Trp 268	Gln 651	22	Phe 503	Cys 652	28	
Asp 484	Lys 482	22	Ala 481	Lys 482	27	
Ala 481	Gln 478	19	Phe 485	Phe 486	25	
Lys 482	Phe 485	19	Gln 478	Gln 478	23	
Lys 482	Lys 482	16	Ser 357	Gln 548	23	
Leu 654	Trp 655	16	Trp 655	Asp 493	21	
Phe 485	Lys 482	16	Asp 493	Gln 651	20	
Phe 486	Phe 485	14	Ser 489	Gln 651	20	
Phe 485	Phe 486	14	Asp 569	Arg 568	19	
Gln 652	Lys 496	14	Phe 485	Glu 648	18	
Gln 500	Trp 655	13	Ala 481	Gln 478	17	
Ile 490	Gln 651	12	Lys 467	Lys 467	16	
Ser 489	Gln 651	11	Arg 549	Ser 357	16	
Phe 485	Glu 648	11	Glu 16	Arg 419	16	
			Leu 658	Leu 658	15	
			Lys 659	Tyr 497	14	
			Ser 489	Glu 648	13	
			Phe 503	Arg 666	13	
			Gln 500	Lys 659	13	
			Arg 573	Arg 572	13	

(AD) WT			(AD) S177/181E			
Y	X-H	0	Y X-H	0		
Asn 225	Leu 223	97	Thr 250 Trp 226	78		
Leu 193	Thr 250	84	Asp 536 Arg 579	55		
Thr 250	Trp 226	82	His 232 Trp 231	39		
His 232	Val 229	38	His 232 His 232	39		
Asn 225	Phe 222	29	Pro 224 Asn 225	35		
Pro 228	Asp 248	29	Arg 577 Arg 582	35		
Gly 251	Trp 226	27	Glu 49 Arg 592	23		
His 232	Trp 231	22	Pro 224 Pro 228	20		
Pro 228	Asp 247	19	Thr 250 Gln 227	19		
Trp 226	Tyr 423	18	Arg 236 Pro 228	19		
Trp 226	Phe 255	17	Gln 230 Val 229	18		
Asn 225	His 425	16	Thr 424 Asn 225	18		
Pro 224	Pro 224	15	Arg 579 Arg 582	17		
Asn 225	Pro 224	14	Val 183 Thr 422	17		
Leu 384	Trp 226	13	Asp 248 Pro 228	14		
Val 183	Leu 421	12	Glu 49 Arg 419	13		
Phe 255	Trp 226	11	Arg 582 Arg 582	13		
Thr 250	Gln 227	10	Asn 225 Val 229	12		
			Leu 186 Thr 250	10		

Table 3.6 Hydrogen bonding occupancies for the AD dimer interface.

(AD) WT-ATP			 (AD) S177/181pS		
Y	X-H	0	 Y	X-H	0
Asp 536	Arg 582	59	Asp 536	Arg 579	88
Asp 536	Arg 579	49	Pro 228	Thr 250	53
Asp 248	Trp 226	25	Pro 224	Asn 225	45
Pro 417	Gln 48	21	Leu 249	Leu 194	40
Asn 225	Trp 226	20	Asp248	Leu 193	34
His 232	Trp 226	16	Pro 578	Arg 579	29
Trp 226	Tyr 423	15	Thr 250	Trp 226	28
Leu 384	Ser 181	14	Pro 228	His 232	28
Trp 231	Trp 231	12	Trp 226	His 425	27
Gln 227	Thr 250	10	Pro 228	Asp 248	25
Leu 193	Leu 249	10	Ser 257	Pro 228	17
Trp 231	His 232	9	Thr 250	Gln 227	15
Pro 578	Arg 579	7	Leu 249	Glu 195	13
			Trp 226	Tyr 423	13
			Thr 250	Leu 194	12
			Sep 181	Lys 418	9
			His 232	Val 229	9

Table 3.6 (cont'd)

3.3.6 Binding Free Energies

Calculating the average binding free energy using GB approximations can provide relative trends of free energies of association and dissociation. All binding free energies are shown in Table 3.7. Using the pairwise de-screening approach (HCT), the dimerization of the AB model is less favorable upon mutation of the S177/181. Although the phosphomimetic mutations show a weaker binding affinity trend than the phosphoserine model, the phosphoserine dimer does begin dissociating (Figure 3.6). In contrast, the AD model has a stronger binding affinity upon active rendering with similar exaggerations from the phosphomimetic mutation. This relative trend visualized in Figure 3.7 supports the favorability for dimerization in the wild type but also proposes an affinity shift towards the AD dimer upon activation. Observing the OBC models, the trend is consistent, excluding the phosphomimetic mutant with the OBC2 model. For the AB dimer

containing ATP, only the HCT model was used to calculate ΔG^{bind} between the dimer interfaces to investigate how ATP binding affects dimerization. The results show that ATP, as expected, prompts dissociation.

	HCT	OBC1	OBC2
AB	-180.7	-88.6	-50.3
AD	-74.6	-41.6	-25.8
AB	-153.7	-65.5	-30.4
AD	-93.9	-51.0	-35.0
AB	-168.5	-82.4	-39.5
AD	-69.5	-28.2	-13.4
AB	-122.0		
	AB AD AB AD AB AD AB	HCTAB-180.7AD-74.6AB-153.7AD-93.9AB-168.5AD-69.5AB-122.0	HCTOBC1AB-180.7-88.6AD-74.6-41.6AB-153.7-65.5AD-93.9-51.0AB-168.5-82.4AD-69.5-28.2AB-122.0-

Table 3.7 Binding free energies of the IKK β dimer.^a

^aAll energies are in kcal/mol.



Figure 3.6 Comparison of the phosphorylated AB Dimer. Shown are the starting structure (top) and the final structure (bottom) after 10 ns.



Figure 3.7 Comparison of binding free energy trends with the HCT, OBC1, and OBC2 models.

3.4 Discussion

Different structure-guided approaches to drug discovery have been developed over the years to better guide rational design.⁷⁸⁻⁸¹ MD simulations are widely used to elucidate structure-activity relationships, however this approach has its own weaknesses due to inadequate conformational sampling which can be due to high computational costs and inherent weaknesses within the force-field parameterizations.^{79,82,83} Focusing on the limitation of conformational sampling, coarse-graining techniques have been designed to provide reduced representations that permit investigations of various conformational states of large biomolecular systems and assemblies with longer time scales^{84,85} that provide a foundation for multiscale modeling approaches. Typically, molecular dynamics simulations of protein-ligand recognition are studied using non-covalently bound ligands to a single subunit or domain of a protein, often neglecting the protein's contribution to the quaternary structure. Although these free, unbound ligands are unable to induce a change in effective force constants or break bonds with point-charged based force

fields, they do induce sterics which can make changes in the local tertiary structure. Despite the limits to conformational sampling, it is important to study the multiple protomers or subunits of quaternary structures when investigating the impact of ligand binding or point mutations to avoid misleading details describing the recognition events, especially those of systems containing homo-multimeric subunits.

As shown in this work, it is important to consider the full structure of IKK β for the study of the effects of distal modifications. The assembly of the inactive xIKK β is described as a symmetric stationary assembly whereas upon activation, the quaternary structure dissociates from the central configuration. From the coarse-grained simulation, the large-scale motions of the kinase domain are pertinent to the kinase activity, in that the activation loop symbolizes a pair of lips that opens and closes repeatedly during the MD simulation. This 'open' and 'closed' conformation resembles the backbone flexibility experimentally shown in the cAMP-dependent protein kinase (cAPK or PKA) catalytic subunit.^{86,87} A central location for potential conformational dynamics is at the hinge, thus opening of the hinge makes the possibility for ligand entrance in the lips of the KD possible in both conformations. Due to the orientation of the dimer-dimer interface, the activation loop was unhindered and accessible. The accessibility of the hinge through the activation loop in the crystal structure is tight, but accessible for flat structures, such as ATP. The polar functional groups of the residues in the pocket of the activation loop cause great polarity of the pocket, which favors the formation of hydrogen bonds. Non-covalent docking of ATP in the binding pocket made an impact on the dimerization. From this docked ATP position, the role of CYS179 modulating ATP binding and activation can be described as a side chain that aids in holding ATP in the pocket.

Since the publication of the xIKK β crystal structure, two structures have been solved for the active conformation of hIKKβ.^{88,89} Both studies confirm that the overall modular arrangement of the different domains between hIKKB and xIKKB are consistent. Additionally, both of the crystal structures confirm that the active assembly of IKKβ resembles a pair of opening shears; this is in agreement with our simulation data. Despite the subtle differences of the opened and closed conformations, the activation segment of the KD regulates activity and dynamics. Modifications to the distal activation loop made a significant impact to the dimerization within a short simulation time. Although the atomistic simulation time was not long enough to observe the opened and the closed conformations, we saw that the pocket from the KD began to open throughout the 10 ns simulation across all models. The phosphomimetic and phosphoserine mutations formed hydrogen bonds and salt interactions with the solvent and nearby side-chains, however, the occupancies of these interactions were not maintained which suggests an entropic flexibility of the KD. Previous investigations of IkB kinase activity showed that IKKB activity can be expressed with phosphomimetic mutations.³² In agreement with experiment, the effects of imploring phosphomimetic mutations rather than actual phosphorylated amino acids were observed in these simulations. Although the opened conformation resembles the active kinase, the inactive KD shows similar activity.

3.5 Conclusion

In this study, molecular dynamics simulations were employed to investigate conformational stability for modeling multimeric species and to decipher whether dimerization is interrupted upon activation of IKK β . Coarse-grained MD simulations showed that simulating more that one protomer of the multimeric assembly provided more stability for the assembly of IKK β .

Interactions at the dimer-dimer interface between the SDD and KD promote a conformational stability for the inactive state of IKK β . Atomistic simulations of the two dimer models confirmed that each protomer mediates electrostatic interactions that are responsible for activity of IKK β . Dimerization of IKK β is disrupted upon activation of Ser177 and Ser181. Rather than inducing local conformational changes within the KD observable within the 10 ns simulation, binding of ATP induces conformational changes that disrupt dimerization. Additionally, phosphomimetic mutations can adequately express the active state of IKK β .

Regarding the molecular modeling of IKK β activity, we highlight the importance of considering both conformation-activity and structure-activity relationships. Considering how protein-protein interactions within assemblies are affected when docking ligands or modifying protein side-chains is critical for understanding structure-activity relationships, even in sites distal from the protein-protein interface. Significant structure-activity relationships can be underestimated by neglecting protein-protein interactions.

REFERENCES

REFERENCES

- (1) Baldwin, A. S., Jr. Series Introduction: The transcription factor NF-κB and human disease. *J. Clin. Invest* **2001**, *107*, 3-6.
- (2) de Martin, R.; Schmid, J. A.; Hofer-Warbinek, R. The NF-kappaB/Rel family of transcription factors in oncogenic transformation and apoptosis. *Mutat. Res.* **1999**, *437*, 231-243.
- (3) Ghosh, S.; May, M. J.; Kopp, E. B. NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.* **1998**, *16*, 225-260.
- (4) Sullivan, J. C.; Kalaitzidis, D.; Gilmore, T. D.; Finnerty, J. R. Rel homology domaincontaining transcription factors in the cnidarian Nematostella vectensis. *Dev. Genes Evol.* **2007**, *217*, 63-72.
- (5) Baldwin, A. S. Control of oncogenesis and cancer therapy resistance by the transcription factor NF-kappaB. *J. Clin. Invest.* **2001**, *107*, 241-246.
- (6) Ghosh, S.; Karin, M. Missing pieces in the NF-kappaB puzzle. *Cell* **2002**, *109 Suppl*, S81-96.
- (7) Gilmore, T. D.; Garbati, M. R. Inhibition of NF-kappaB signaling as a strategy in disease therapy. *Curr. Top. Microbiol. Immunol.* **2011**, *349*, 245-263.
- (8) Gilmore, T. D.; Wolenski, F. S. NF-kappaB: where did it come from and why? *Immunol. Rev.* **2012**, *246*, 14-35.
- (9) Rayet, B.; Gelinas, C. Aberrant rel/nfkb genes and activity in human cancer. *Oncogene* **1999**, *18*, 6938-6947.
- (10) Gilmore, T. D. Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene* **2006**, *25*, 6680-6684.
- (11) Perkins, N. D.; Gilmore, T. D. Good cop, bad cop: the different faces of NF-kappaB. *Cell Death Differ*. **2006**, *13*, 759-772.
- (12) Gilmore, T. D.; Gerondakis, S. The c-Rel Transcription Factor in Development and Disease. *Genes Cancer.* **2011**, *2*, 695-711.
- (13) Gilmore, T. D.; Kalaitzidis, D.; Liang, M. C.; Starczynowski, D. T. The c-Rel transcription factor and B-cell proliferation: a deal with the devil. *Oncogene* 2004, 23, 2275-2286.
- (14) Gilmore, T. D. Introduction: The Rel/NF-kappaB signal transduction pathway. *Semin. Cancer Biol.* **1997**, *8*, 61-62.

- (15) Gilmore, T. D.; Koedood, M.; Piffat, K. A.; White, D. W. Rel/NF-kappaB/IkappaB proteins and cancer. *Oncogene* **1996**, *13*, 1367-1378.
- (16) Gilmore, T. D. NF-kappa B, KBF1, dorsal, and related matters. *Cell* **1990**, *62*, 841-843.
- (17) Sun, S. C. Non-canonical NF-kappaB signaling pathway. *Cell Res.* 2011, 21, 71-85.
- Beg, A. A.; Ruben, S. M.; Scheinman, R. I.; Haskill, S.; Rosen, C. A.; Baldwin, A. S., Jr I kappa B interacts with the nuclear localization sequences of the subunits of NF-kappa B: a mechanism for cytoplasmic retention. *Genes Dev.* **1992**, *6*, 1899-1913.
- (19) Beg, A. A.; Baldwin, A. S., Jr The I kappa B proteins: multifunctional regulators of Rel/NF-kappa B transcription factors. *Genes Dev.* **1993**, *7*, 2064-2070.
- (20) Zabel, U.; Henkel, T.; Silva, M. S.; Baeuerle, P. A. Nuclear uptake control of NF-kappa B by MAD-3, an I kappa B protein present in the nucleus. *EMBO J.* **1993**, *12*, 201-211.
- (21) Huguet, C.; Crepieux, P.; Laudet, V. Rel/NF-kappa B transcription factors and I kappa B inhibitors: evolution from a unique common ancestor. *Oncogene* **1997**, *15*, 2965-2974.
- (22) Baeuerle, P. A.; Henkel, T. Function and activation of NF-kappa B in the immune system. *Annu. Rev. Immunol.* **1994**, *12*, 141-179.
- (23) Siebenlist, U.; Franzoso, G.; Brown, K. Structure, regulation and function of NF-kappa B. *Annu. Rev. Cell Biol.* **1994**, *10*, 405-455.
- (24) Karin, M. How NF-kappaB is activated: the role of the IkappaB kinase (IKK) complex. *Oncogene* **1999**, *18*, 6867-6874.
- (25) Israel, A. The IKK Complex, a Central Regulator of NF-κB Activation. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000158-a000158.
- (26) Nishikori, M. Classical and Alternative NF-kB Activation Pathways and Their Roles in Lymphoid Malignancies. J. Clin. Exp. Hematop. 2005, 45, 15-24.
- (27) Gamble, C.; McIntosh, K.; Scott, R.; Ho, K. H.; Plevin, R.; Paul, A. Inhibitory kappa B Kinases as targets for pharmacological regulation. *Br. J. Pharmacol.* 2012, *165*, 802-819.
- (28) Karin, M.; Delhase, M. The I kappa B kinase (IKK) and NF-kappa B: key elements of proinflammatory signalling. *Semin. Immunol.* **2000**, *12*, 85-98.
- (29) May, M. J.; Ghosh, S. Rel/NF-kB and IkB proteins: an overview. *Semin. Cancer Biol.* **1997**, *8*, 63-73.

- (30) Zandi, E.; Rothwarf, D. M.; Delhase, M.; Hayakawa, M.; Karin, M. The IkappaB kinase complex (IKK) contains two kinase subunits, IKKalpha and IKKbeta, necessary for IkappaB phosphorylation and NF-kappaB activation. *Cell* **1997**, *91*, 243-252.
- (31) Hacker, H.; Karin, M. Regulation and function of IKK and IKK-related kinases. *Sci. STKE* **2006**, *2006*, re13.
- Mercurio, F.; Zhu, H.; Murray, B. W.; Shevchenko, A.; Bennett, B. L.; Li, J.; Young, D. B.; Barbosa, M.; Mann, M.; Manning, A.; Rao, A. IKK-1 and IKK-2: cytokine-activated IkappaB kinases essential for NF-kappaB activation. *Science* 1997, 278, 860-866.
- (33) Hayden, M. S.; Ghosh, S. Signaling to NF-kappaB. Genes Dev. 2004, 18, 2195-2224.
- (34) Zandi, E.; Chen, Y.; Karin, M. Direct phosphorylation of IkappaB by IKKalpha and IKKbeta: discrimination between free and NF-kappaB-bound substrate. *Science* **1998**, *281*, 1360-1363.
- (35) Delhase, M.; Karin, M. The I kappa B kinase: a master regulator of NF-kappa B, innate immunity, and epidermal differentiation. *Cold Spring Harb. Symp. Quant. Biol.* 1999, 64, 491-503.
- (36) Johnson, L. N.; Noble, M. E.; Owen, D. J. Active and inactive protein kinases: structural basis for regulation. *Cell* **1996**, *85*, 149-158.
- (37) Pandey, M. K.; Sung, B.; Kunnumakkara, A. B.; Sethi, G.; Chaturvedi, M. M.; Aggarwal, B. B. Berberine modifies cysteine 179 of IkappaBalpha kinase, suppresses nuclear factor-kappaB-regulated antiapoptotic gene products, and potentiates apoptosis. *Cancer Res.* 2008, 68, 5370-5379.
- (38) Gupta, S. C.; Prasad, S.; Reuter, S.; Kannappan, R.; Yadav, V. R.; Ravindran, J.; Hema, P. S.; Chaturvedi, M. M.; Nair, M.; Aggarwal, B. B. Modification of cysteine 179 of IkappaBalpha kinase by nimbolide leads to down-regulation of NF-kappaB-regulated cell survival and proliferative proteins and sensitization of tumor cells to chemotherapeutic agents. J. Biol. Chem. 2010, 285, 35406-35417.
- (39) Byun, M. S.; Choi, J.; Jue, D. M. Cysteine-179 of IkappaB kinase beta plays a critical role in enzyme activation by promoting phosphorylation of activation loop serines. *Exp. Mol. Med.* **2006**, *38*, 546-552.
- (40) Harikumar, K. B.; Kunnumakkara, A. B.; Ahn, K. S.; Anand, P.; Krishnan, S.; Guha, S.; Aggarwal, B. B. Modification of the cysteine residues in IkappaBalpha kinase and NFkappaB (p65) by xanthohumol leads to suppression of NF-kappaB-regulated gene products and potentiation of apoptosis in leukemia cells. *Blood* 2009, *113*, 2003-2013.
- (41) Xu, G.; Lo, Y. C.; Li, Q.; Napolitano, G.; Wu, X.; Jiang, X.; Dreano, M.; Karin, M.; Wu, H. Crystal structure of inhibitor of kappaB kinase beta. *Nature* **2011**, 472, 325-330.

- (42) Nolen, B.; Taylor, S.; Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **2004**, *15*, 661-675.
- (43) Boutros, T.; Chevet, E.; Metrakos, P. Mitogen-Activated Protein (MAP) Kinase/MAP Kinase Phosphatase Regulation: Roles in Cell Growth, Death, and Cancer. *Pharmacol. Rev.* 2008, 60, 261-310.
- (44) Pearson, G.; Robinson, F.; Beers Gibson, T.; Xu, B. E.; Karandikar, M.; Berman, K.; Cobb, M. H. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.* **2001**, *22*, 153-183.
- (45) Dhillon, A. S.; Hagan, S.; Rath, O.; Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **2007**, *26*, 3279-3290.
- (46) Ferrell, J. E., Jr; Bhatt, R. R. Mechanistic studies of the dual phosphorylation of mitogen-activated protein kinase. *J. Biol. Chem.* **1997**, *272*, 19008-19016.
- (47) Kalia, M.; Kukol, A. Structure and dynamics of the kinase IKK-beta--A key regulator of the NF-kappa B transcription factor. *J. Struct. Biol.* **2011**, *176*, 133-142.
- (48) Kornev, A. P.; Taylor, S. S.; Ten Eyck, L. F. A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14377-14382.
- (49) Steichen, J. M.; Iyer, G. H.; Li, S.; Saldanha, S. A.; Deal, M. S.; Woods, V. L., Jr; Taylor, S. S. Global consequences of activation loop phosphorylation on protein kinase A. J. Biol. Chem. 2010, 285, 3825-3832.
- (50) Scheeff, E. D.; Bourne, P. E. Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.* **2005**, *1*, e49.
- (51) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "gatekeeper door": exploiting the active kinase conformation. *J. Med. Chem.* **2010**, *53*, 2681-2694.
- (52) Palermo, N. Y.; Natarajan, A. Beyond the frog: the evolution of homology models of human IKKbeta. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 6081-6084.
- (53) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* 2000, 28, 235-242.
- (54) Chemical Computing Group Inc. Molecular Operating Environment (MOE), 2011.10. **2011**, .
- (55) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- (56) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* 2003, 66, 27-85.
- (57) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198-210.
- (58) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. J. Comput. Chem. 2005, 26, 1668-1688.
- (59) Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. Use of amino acid composition to predict ligand-binding sites. *J. Chem. Inf. Model.* **2007**, *47*, 400-406.
- (60) Durand, E.; Derrez, E.; Audoly, G.; Spinelli, S.; Ortiz-Lombardia, M.; Raoult, D.; Cascales, E.; Cambillau, C. Crystal structure of the VgrG1 actin cross-linking domain of the Vibrio cholerae type VI secretion system. *J. Biol. Chem.* **2012**, *287*, 38190-38199.
- (61) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003**, *24*, 1016-1025.
- (62) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- (63) Homeyer, N.; Horn, A. H.; Lanig, H.; Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.* 2006, *12*, 281-289.
- (64) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K.; Yao, X.; Takada, S. CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J. Chem. Theory Comput.* 2011, *7*, 1979-1989.
- (65) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. J. Phys. Chem. B 2008, 112, 9020-9041.
- (66) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.
- (67) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712-725.
- (68) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004, 25, 1605-1612.

- (69) Prall, M. VMD: a graphical tool for the modern chemists. *J. Comput. Chem.* **2001**, *22*, 132-134.
- (70) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314-3321.
- (71) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122-129.
- (72) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824-19839.
- (73) Tsui, V.; Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275-291.
- (74) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified Generalized Born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383-394.
- Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* 2000, 18, 113-135.
- (76) Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774-797.
- (77) Krissinel, E. Crystal contacts as nature's docking solutions. *J. Comput. Chem.* **2010**, *31*, 133-143.
- (78) Mandal, S.; Moudgil, M.; Mandal, S. K. Rational drug design. *Eur. J. Pharmacol.* **2009**, 625, 90-100.
- (79) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71-7007-9-71.
- (80) Borhani, D. W.; Shaw, D. E. The future of molecular dynamics simulations in drug discovery. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 15-26.
- (81) Hao, G. F.; Yang, G. F.; Zhan, C. G. Structure-based methods for predicting target mutation-induced drug resistance and rational drug design to overcome the problem. *Drug Discovery Today* **2012**, *17*, 1121-1126.
- (82) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* 2011, 21, 150-160.

- (83) Mobley, D. L.; Chodera, J. D.; Dill, K. A. The Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J. Chem. Theory Comput.* **2007**, *3*, 1231-1235.
- (84) Tozzini, V. Minimalist models for proteins: a comparative analysis. *Q. Rev. Biophys.* **2010**, *43*, 333-371.
- (85) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annu. Rev. Phys. Chem.* 2011, 62, 41-64.
- (86) Li, F.; Gangal, M.; Juliano, C.; Gorfain, E.; Taylor, S. S.; Johnson, D. A. Evidence for an internal entropy contribution to phosphoryl transfer: a study of domain closure, backbone flexibility, and the catalytic cycle of cAMP-dependent protein kinase. *J. Mol. Biol.* 2002, *315*, 459-469.
- (87) Li, F.; Gangal, M.; Jones, J. M.; Deich, J.; Lovett, K. E.; Taylor, S. S.; Johnson, D. A. Consequences of cAMP and catalytic-subunit binding on the flexibility of the A-kinase regulatory subunit. *Biochemistry* 2000, *39*, 15626-15632.
- (88) Polley, S.; Huang, D.; Hauenstein, A. V.; Fusco, A. J.; Zhong, X.; Vu, D.; Schroefelbauer, B.; Kim, Y.; Hoffmann, A.; Verma, I. M.; Ghosh, G.; Huxford, T. A structural basis for IkB Kinase 2 Activation via oligomerization dependent trans autophosphorylation. *PLoS Biol.* **2013**, *11*, 1-13.
- (89) Liu, S.; Misquitta, Y. R.; Olland, A.; Johnson, M. A.; Kelleher, K. S.; Kriz, R.; Lin, L. L.; Stahl, M.; Mosyak, L. Crystal Structure of a Human IkappaB Kinase beta Asymmetric Dimer. *J. Biol. Chem.* 2013, 288, 22758-22767.

CHAPTER 4 Impact of Intracellular Ionic Strength on Dimer Binding in the NF-kB Inducing Kinase

4.1 Introduction

Understanding the signaling pathways associated with the onset and progression of oncogenesis has been a target of many diverse disciplines. Much emphasis has been placed on elucidating the nuclear factor (NF)- κ B signaling pathway since aberrations within this pathway are associated with many different kinds of immune responses.^{1–3} Sequestered in an inactive state in the cytoplasm by its inhibitory subunits (I κ B α , I κ B β , and I κ B ϵ), the NF- κ B pathway includes a family of five transcription factors, NF- κ B1 (p50 and p105), NF- κ B2 (p52 and p100), RelA (p65), RelB, and c-Rel.^{4–7} These transcription factors are assembled into operational-specific and functional-specific dimers. The dimers include both heteroligomers and homoligomers that, depending upon the conformational and oligomeric architecture, can either induce or repress gene regulation and expression.^{8–12} These transcription factors differ on how and where they bind to DNA and different combinations lead to different specificities of gene expression.⁹ As a central activator of genes involved in inflammation and immune function, the NF- κ B family represents a vital signaling cascade that must be further analyzed to prevent dysregulations in gene expression.

Although there are many ways in which NF- κ B activation can be triggered, the release of the NF- κ B dimers are regulated by two main pathways.¹³ In the classical, or canonical, pathway, the rate-limiting step of the signaling mechanism is the signal-induced degradation of the I κ Bs, which happens upon the activation of the catalytic subunits of the I κ B kinase (IKK) complex. Comprised of two homologous catalytic subunits, IKK α and IKK β , and a regulatory subunit IKK γ , activated enzymes of the IKK complex phosphorylate I κ B, leading to the selective release of the

NF-κB dimers. These NF-κB dimers translocate to the nucleus to regulate gene expression. The signaling cascades that result from the activation of IKKβ regulate the classical pathway by targeting NF-κB1. In contrast, the alternative, or non-canonical, pathway is dependent upon the activation of the NF-κB Inducing Kinase (NIK), an upstream kinase that selectively phosphorylates IKK α , which activates NF-κB2.^{14–17} Although some studies have suggested that NIK serves specific functions in the classical pathway, as the critical component of the non-canonical pathway, NIK serves a promising target for the treatment of autoimmune disorders and cancers. ^{18–22}

The human sequence of NIK, composed of 947 amino acids, is comprised of an N-terminal domain that binds to TRAF3 (30–120), a negative regulatory domain (121–318), a kinase domain (390–660), and a non-catalytic C-terminus (661-947).²³ The negative regulatory domain contains two structural motifs, a basic region (127-147)—similar to that of a leucine zipper—and a proline-rich repeat sequence (250-317). The two regions of the negative regulatory domain have been shown to control the signaling function of NIK as deletions of either motif greatly enhances the NF-KB inducing activity of NIK.²⁴ The C-terminal region is responsible for the signaling function of NIK and is critical for its recognition of other proteins within the pathway, including IKK α and NF- κ B2. As a member of the mitogen-activated protein (MAP) kinase kinase kinase (MAP3K) family, NIK contains the conserved DFG-APE motif (534-566) shared among serine/threonine kinases. Kinases in this family typically are activated via phosphorylation of serines or threonines homologous to Ser549, Thr552, and Thr559 in NIK.

Previous structural and biochemical studies investigating the regulation of NIK have drawn interest into understanding how NIK maintains its structural integrity for its function.²⁵ A crystal structure of truncated NIK, encompassing residues 330-679, was shown to retain catalytic

activity.²³ However, unlike other kinase domains, this catalytically competent conformation is maintained by an N-terminal extension prior to the kinase domain rather than through a phosphorylation event. Mutations that mimic the state of being phosphorylated, termed phosphomimetic mutations, including S549D, T552E/E, and T559D/E, were considered, however, it was determined that the activity of the mutants was similar to that of the wild-type (WT).²⁶ The crystal structure contained two molecules of NIK in an asymmetric arrangement that formed a head-to tail homodimer with a buried solvent-accessible surface area between the interface of 2900 $Å^2$.

Additional studies were carried out to further investigate whether the observed dimer arrangement represented the actual dimer interface or a nonspecific crystallographic artifact. Chromatographic analyses indicated that the protein in solution formed both monomers and dimers. While the full NIK protein is known to be an oligomer, what the dimeric arrangement of the crystal corresponds to remains elusive.²⁵ It has been demonstrated that dimerization, along with other oligomeric arrangements, is significant among many protein kinases as the oligomeric state represents a specific functional state that couples with local conformational changes within a protomer, as small as local sidechain or segment movements to larger domain movements, that dictate how the protein interacts with other biomolecules.^{27–29}

The active site of NIK has been characterized as having structural features that resemble a catalytically active conformation similar to other Ser/Thr kinases in the active state.^{23,26,30–32} Small molecule binding sites in NIK have been identified in crystallographic studies. The nucleotide binding mode has been shown to be consistent with that of protein kinases that undergo phosphorylation in the activation segment, including the cyclic AMP-dependent kinase.^{26,33} Several inhibitors have been identified to bind within the catalytic segment. Despite the diversity

of the compounds in regards to the structure and activity properties, the compounds have a similar scaffold that relates to the purine base of ATP, containing a nitrogen heterocycle that binds near the hinge region between the N and C lobes and interacts with the backbone of Leu471, Leu472, and Glu470. The crystal structure of the truncated NIK oligomer shows that the activation segment of one subunit interacts with the N-terminal extension of the other. In this structure, each subunit is crystallized with ATP- γ -S in the activation segment, however, the conformation of the oligomer interface may hinder the binding of other substrates.

Previous work investigating the conformational-activity and structure-activity relationships of IKKB identified that protein-protein complexes crystallized with surface mutations or phosphomimetic mutations do not always represent the true conformational state and that dynamics are essential.³⁴ Additionally, it was demonstrated that through the use of molecular simulation and by introducing mutations and the binding of small molecules to the crystal structure, allosteric responses could be observed on both a local and a global scale. As the function of a protein is determined by its structure, which is impacted by the surrounding environment, the activity of proteins is extremely sensitive to changes in the environment. Specifically, changes in the intracellular temperature, pressure, pH, and ionic strength can all affect the structure and function of proteins.³⁵⁻⁴⁰ The aforementioned conditions can also influence crystal growth as oligomerizing may be significant for crystallization. Since the activity of a protein is measured by its ability to effectively bind and communicate with other molecules, it is essential to understand the environmental sensitivity of the protein recognition of other molecules involved in abnormal signal transduction mechanisms. Previous studies have observed that cancerous cells, relative to normal cells, have different acidic environments and intracellular ionic compositions ⁴¹⁻⁴⁴,

however, how this directly impacts protein interactions with other molecules within the signaling cascade have not been investigated.

In this work, computer simulation and modeling techniques are employed to investigate the crystallographic arrangement corresponds to a stable dimer and offer additional insight into how the structure-function properties of NIK are dependent upon the intracellular microenvironment and contribute to the its stabilization. By investigating both the inactive and active states of NIK in conjunction with different intracellular ionic strengths, the assessment of the protein-protein interactions provide further resolution of the oligomeric properties of the truncated construct and explicitly highlight the role of the microenvironment in the maintenance of the active conformational states of NIK. The inactive state of NIK, also referred to as the wildtype (WT) system, is compared to an active state of NIK which has a phosphomimetic mutation on Ser549 (S549D). Furthermore, to evaluate the impact of the environment on NIK's structure, 100 mM of sodium chloride was added to both the inactive and active states, represented as WT [Na+] and S549D [Na+] systems, respectively, and compared to the neutralized systems with 0 mM sodium chloride.

4.2 Material and Methods

The crystal structure of NIK, determined with two phosphomimetic mutant (S549D) molecules, was obtained from the Protein Data Bank (PDB ID: 4DN5)²⁶ and prepared using Molecular Operating Environment ⁴⁵ (MOE). Unresolved bonds and atoms were added using MOE and minimized under the Amber99 force field. Mutations were introduced to model both the inactive (wild-type) state and active state, via S549D phosphomimetic mutations. Both models of the inactive and active states were neutralized with 8 Na⁺ ions and solvated in 14 Å of TIP4P-Ew⁴⁶ water beyond the solute in a cuboid box, resulting in an addition of 27560 water molecules. From

these, two additional models were created by solvating with 100 mM of sodium chloride by adding 136 molecules of NaCl (68 of each ion).

The energy of each solvated system was minimized holding a harmonic restraint on the protein of 500, 200, and 20 kcal/mol-A², using 500 steepest descent steps and 500 conjugate gradient steps per harmonic restraint and a non-bond cutoff of 10 Å, followed by 50 ps of NPT MD with a 20 kcal/mol-A² restraint. Subsequently, the energy of each system was further minimized under decreasing restraints on the protein of 10, 5, and 0 kcal/mol using 250 steepest descent steps and 250 conjugate gradient steps. After the energy minimization, the systems were heated gradually to 300K over 50 ps, followed by 1 ns of NPT MD simulation. Four independent simulations were performed for 215 ns at 300K using AMBER 14⁴⁷ in the presence of 0 mM and 100 mM of sodium chloride and under the Amber99SB⁴⁸ force field. The temperature was controlled using Langevin dynamics with a collision frequency of 1 ps⁻¹. The Berendsen barostat was used with isotropic position scaling with a pressure relaxation time of 2 ps to maintain the pressure at 1 bar.

Trajectory analysis was performed using AmberTools and visualized with VMD⁴⁹. Thermodynamic parameters and root mean-squared deviations (RMSDs) from the starting structure were performed using the CPPTRAJ module⁵⁰ of AmberTools. Hydrogen bonding analysis was performed with the CPPTRAJ module of AmberTools, where a hydrogen bond was defined by a cutoff distance of 3.5 Å and a donor-hydrogen-acceptor angle of 180°±60°. Free energy calculations were carried out using MMPBSA.py.⁵¹ The electrostatic solvation energies were calculated for every ps using three GB models: the pairwise GB models of Hawkins and co-workers^{52–54} (HCT) and two modified models developed by Onufriev and coworkers⁵⁵ (OBC1 and OBC2), setting the dielectric constants to 1 and 80 for the molecule and the solvent, respectively.

Since only relative free energy trends were of interest, solute entropy was neglected. The PISA (Protein Interfaces, Surfaces, and Assemblies) software was used on static structures of the initial crystal structure and the final frames of the four independent simulations to calculate properties of the macromolecular interface.⁵⁶

4.3 **Results and Discussion**

To assess the stability of the simulation, thermodynamic parameters are extracted from the trajectories and plotted as a function of time (Table 4.1). Analysis of the thermodynamic parameters shows that the total energy, temperature, and density fluctuated around a common value, reflecting that the simulations are stable under the isobaric-isothermal ensemble and that the properties investigated have physical significance.

Table 4.1 Average thermodynamic parameters from simulations.

	Density (g/mL)	Total Energy (kcal/mol)	Temperature (K)
WT	1.032 ± 0.001	-264465.5 ± 352.6	300.01 ± 0.96
WT [Na+]	1.038 ± 0.001	-276116.7 ± 351.1	300.01 ± 0.96
S549D	1.033 ± 0.001	-267543.1 ± 343.2	300.01 ± 0.97
S549D [Na+]	1.038 ± 0.001	-279172.3 ± 354.4	300.01 ± 0.96

4.3.1 Structural Fluctuation

Local side-chain dynamics are structurally significant as they often couple with larger conformational dynamics and are central for understanding correlations between protein-function. The flexibility of NIK was observed by measuring the RMSD to compare its deviations from both the starting structure and the average structure. As shown in Figure 4.1, the RMSD of WT [Na+] generally deviated less than the WT, but the mutant S549D [Na+] deviated more than the S549D. This indicates that the introduction of the mutation increases the sensitivity of the protein to solvent ions. The fluctuations of the neutral WT and the S549D converged and stabilized early in the

simulation in which the mutant deviated less than the WT; however, the addition of the ion buffer perturbed this relationship. By comparing the neutral systems with the [Na+] systems, it is clear that the presence of the [Na+] buffer stabilizes the WT more than the mutant. To further characterize these motions, the root mean-squared fluctuation (RMSF) was measured to identify whether the deviations were consistent across the full structure. The RMSF plot indicates that the N-terminal segment and part of the kinase domain (Val338 to Ala438) is stabilized throughout the simulation, whereas the activation loop and the C-terminal (Ser581 to Pro675) are more dynamic regions (Figure 4.2). Both systems containing [Na+] resulted in significantly lower deviations, suggesting that [Na+] maintains an accessible area for catalytic activity. The results show that the presence of the ionic environment results in a reduced flexibility of the NIK structure, which may potentially mediate its catalytic activity. As discussed earlier, the C-terminal plays a role in controlling the signaling function of NIK while the N-terminal maintains its catalytically competent conformation; here the results agree.



Figure 4.1 Root mean square deviations (RMSDs) obtained from the 215 ns simulation. (A) The RMSDs were calculated from the starting structure of the inactive (WT) and phosphomimetic mutant (S549D) states in a neutral solution (color coded in blue and green, respectively) and a 100 mM NaCl solution ([Na+]; color coded in red and purple, respectively). For comparison, plots B and C highlight the trends between the inactive and the active states, whereas the impact of the ionic environment is highlighted in plots D and E. The darker colored lines represent the moving average per 100 frames.



Figure 4.2 Root mean square fluctuations (RMSF) of each protomer of NIK. The RMSFs over the 215 ns were calculated per residue for each protomer (332-675) for the modeled inactive states, WT and WT [Na+] (represented in blue and red, respectively), and active states, S549D and S549D [Na+] (represented in green and purple, respectively).

4.3.2 Changes in Solvent Accessibility

Driving forces that promote the functional and dysfunctional association or dissociation of proteins can be measured by physicochemical properties.^{57–59} In particular, the solvent accessible surface area is useful for detecting and analyzing changes in the protein conformation induced as a result of point-mutations, ligand binding as well the binding of other proteins. ⁶⁰ The structure of the NIK protein was crystallized with two molecules related by a 2-fold symmetry, which is a common point group symmetry observed in biological homodimers.^{61,62} Together, the two molecules had a surface area of 29656 Å² in which 2932 Å² is buried between the two units (Table 4.2) However, this was based on a static structure that was crystallized with magnesium, g-ATP, and crystallizing agents.

	Surface Area	Buried Area	ΔG_{int}	ΔG_{diss}	T Δ S _{diss}
Model ^{ab}	(Å ²)	(Å ²)	(kcal/mol)	(kcal/mol)	(kcal/mol)
4DN5 ^a	29656.2	2931.8	-4.9	3.9	13.9
WT	29509.2	3692.0	-5.2	3.5	13.8
WT [Na+]	30740.6	2894.5	-8.9	2.8	14.1
S549D	33479.4	492.8	-1.6	-11.2	13.2
S549D [Na+]	31878.6	2829.6	1.8	-4.5	13.8

Table 4.2 Summary of the dimer interface of NIK.

^a The crystal structure of the NIK dimer.

^b Predictions of the structural and chemical properties correspond to the last trajectory (215 ns). The free energy of formation (ΔG_{int}) and free energy of dissociation (ΔG_{diss}) are reported in kcal/mol.

Using a static image from the last frame of simulation, the changes in the total and buried surface area of the four models are summarized in Table 4.2. Looking at the changes between the crystal structure and WT, the change in the surface area is small (-147 Å²), however the change in the buried surface area increases by 760 Å². This is interesting because the crystal structure corresponds to a phosphomimic state of NIK, as reflected by S549D, which the surface area increases by 3823 Å² and the buried surface area decreases by 2439 Å² (meaning that the unit is dissociating/associating).

For the models that include an ionic strength relative to the crystal structure, the surface area of WT [Na+] and S549D [Na+] increases by 1084 Å² and 2222 Å², respectively, and buried surface area decreases at by 37 Å² and 102 Å², respectively, suggesting that the presence of ions may have an explicit role in stabilizing or destabilizing interactions between the dimer interface. Comparing the changes between systems with and without an ion buffer, the change in the surface area of WT to WT [Na+] increases by 1231 Å² and the buried surface area decreases by 798 Å². In contrast, for the phosphomimic mutant model, the surface area decreases by 1601 Å² (from S549D to S549D [Na+]) and the buried surface area increases by 2337 Å², which means that the dimer interface becomes more exposed.

By analyzing the dynamic changes of a reference protein structure, a mutant or modified state can be examined in the absence of a corresponding experimentally resolved structure to observe how the reference system responds to changes. To characterize the effect of the ionic composition on the dimerization, the accessible surface area (ASA) was measured to determine if either the phosphomimetic mutation or the ionic strength induced conformational changes that expose the dimer to more solvent (Figure 4.3). The neutral WT and S549D have an average ASA of 28359±450 and 29481±390 Å², respectively, and a greater average ASA of 28840±511 and 29912±591 Å², respectively, for the [Na+] systems. After 8 ns, the ASA for both states fluctuate around a common value, in which the ASA of S549D is nearly 3000 $Å^2$ greater than the WT. In contrast, the ASA for the [Na+] systems increases consistently, in which the S549D [Na+] is around 2000 Å² greater than the WT [Na+]. Comparing each state with and without the presence of an ion buffer, the trends in the ASA diverge with separations from 1000-3000 \AA^2 for the WT and 1000-2000 Å² for the S549D model after 6 ns. These trends indicate that the S549D phosphomimetic mutation induces conformational changes that make the NIK dimer more exposed to solvent and that the presence of the ion buffer impacts the structure more than the mutation alone.

4.3.3 Hydrogen-Bonding Analysis/Dimer vs Monomer

Structural-based chromatographic spectroscopy and spectrometry studies were used to determine whether the two asymmetric molecules resolved in the crystal structure of the truncated NIK protein are representative of the homodimer as expressed in the full-length protein or a crystallographic artifact. ²⁶ However, the oligomeric arrangement was unable to be resolved as it was determined that NIK in solution is both monomeric and dimeric. To quantify the gain or loss of interactions between the dimer interface upon mutation, a hydrogen bond analysis was

performed. As illustrated in Table 4.3, upon the activation of NIK, hydrogen bonding interactions between the dimer interface decrease in comparison to the WT. Interestingly, three strong hydrogen bond donor-acceptor pairs, (Phe436-Arg368; Glu434-Ser367; Val435-Arg368), were identified in the neutral WT that were not observed in the other models. Furthermore, the loss of Thre559, NIK's phosphorylation site, as an acceptor of Lys373 in both [Na+] systems shows that the presence of sodium alters the phosphorylation activity of NIK. The results suggest that the presence of the sodium buffer alters hydrogen bonding networks. Regarding the dimerization, the trends indicate that the crystallized structure of NIK may resemble the homodimer.

Y	X-H		WT		S549D
(Unit A)	(Unit B)	WT	[Na+]	S549D	[Na+]
Ser371	Glu560	81%	80%	75%	97%
Gly558	Ser371	77%	66%	67%	72%
Thr561	Lys373	54%	3%	19%	7%
Ser371	Thr559	39%	47%	36%	47%
Glu395	Arg408	38%	8%	30%	14%
Thr561	Pro372	34%	-	7%	3%
Arg432	Phe411	32%	23%	35%	8%
Glu461	Arg408	30%	36%	18%	19%
Trp464	Gly412	29%	9%	14%	18%
Ser410	Arg432	27%	43%	11%	42%
Glu395	Arg405	25%	14%	18%	47%
Asp554	Arg368	23%	-	12%	49%
Gly412	Trp464	19%	14%	24%	3%
Glu396	Arg405	19%	25%	20%	14%
Pro370	Pro557	16%	22%	29%	11%
Pro370	Met563	15%	3%	9%	1%
Asp519	Lys373	15%	-	7%	-
Glu413	Lys430	14%	6%	12%	5%
Hie402	Gln403	11%	6%	0%	9%
Pro370	Val568	11%	-	8%	-
Glu560	Pro370	10%	12%	7%	11%
Gly558	Pro370	10%	2%	18%	2%
Arg408	Glu461	10%	1%	1%	-
Asn466	Arg408	10%	-	-	-
Leu551	Arg368	-	-	-	51%
Gln403	Gln403	2%	3%	2%	40%

Table 4.3 Hydrogen bonding analysis of the NIK dimer interface.^a

Table 4.3 (cont'd)

^a The hydrogen bond acceptor (Y) and donor (X-H) pairs observed from the simulations are denoted by percent occupations.

4.3.4 Aggregation Propensity Sensitive to Ion Buffer

To characterize the aggregation propensity of the two dimer states, the radius of gyration was measured to differentiate the effects caused by the ionic strength versus the mutation (Figure 4.4). Relative to the respective neutral systems, the presence of the ionic buffer in the WT and S549D marginally enhances the rate of the dimer dissociation. For the neutral systems, the radius of gyration plateaus after 8 ns indicating that the dissociation of the dimer is less dependent upon the mutation and that the oligomerization is sensitive to the ionic strength. There are several exposed cavities between the dimer interface in which solvent molecules are occupied for the duration of the simulation (Figure 4.5). It should be noted that the presence of sodium ions within these hydrated cavities was observed.



Figure 4.3 Solvent residing between the dimer interface.



Figure 4.4 Changes in the compactness of the crystal structure.

4.3.5 Comparisons Between the Dimer Binding Energies

To estimate the effects of the ionic strength in terms of the energetic property between the inactive and active state, the relative binding free energies were evaluated (Table 4.4; Figure 4.5). Using the pairwise screening approach (HCT), relative to the WT, the binding free energy of the S549D decreases. The presence of ions enhances the association and dissociation of the WT and S549D, respectively. The electrostatic component of the binding free energy is sensitive to input parameters involved within the calculation—including the internal dielectric constant, the probe radius, and the force field—and can lead to unreliable absolute free energy predictions. For this only relative binding free energetics are considered. In order to evaluate the consistency of the trends between the active and inactive states, additional free energy GB solvation models, including the OBC*n* and GB*n* models, were used to calculate the dimer binding free energies; in all cases the trends held.



Figure 4.5 Binding free energy trends with the HCT and OBC Generalized Born models.

Table 4.4 Binding free energies of the NIK dimer.^a

Model	НСТ	OBC1	OBC2
WT	-126.8	-80.9	-75.2
WT [Na+]	-118.1	-76.1	-71.2
S549D	-126.1	-76.3	-68.5
S549D [Na+]	-134.4	-81.0	-75.0

^a All units are reported in kcal/mol.

4.4 Conclusions

The structure of a protein determines the binding ability and specificity of small molecules and other proteins. Disruptions that cause changes in the native structure of NIK usually serve as the basis for mutations and incorrect gene expressions that lead to diseases and cancers. Changes in the intracellular compositions of healthy cells are a potential factor for disrupting protein structure. The results in this study indicate that an increase in [Na+] tends to further aid in the stabilization of the inactive state of NIK and the destabilization of active state. This implies that increases in sodium concentration in cells can actually disrupt protein structure and lead to mutations and diseases. This claim is also supported by studies that indicate that there is a considerable increase in the sodium in many tumor cells⁴³, which is most likely through sustained depolarization of their cell membranes and high mitotic activity of the cancer cells. These shifts in ionic concentrations within the cells may change the environment in which proteins such as NIK function. These could lead to changes in the function of the protein and the onset of cancer.

This computational study was conducted to determine structure-function properties of the NIK protein. NIK is of interest due to its important role in the NF-kB pathway, a pathway that regulates many inflammatory and autoimmune processes. The results show a homodimeric structure with sensitivity towards ionic presence. Furthermore, the introduction of the phosphomimetic mutation on the 549th residue significantly impacted the behavior and subsequently the function of NIK. The S549D mutation induces conformational changes that make the NIK dimer more exposed to solvent. The increase in solvent-accessibility due to the mutation also serves to make NIK even more sensitive to an ionic presence. The addition of the ions then causes the structural flexibility of NIK to increase. These changes brought on by the ionic sensitivity of NIK then cause structural fluctuations, specifically in the C-terminal and N-terminal regions. These fluctuations in structure then result in changes in kinase domain, allowing for the surrounding residues to acclimate in order to allow NIK to maintain the active conformation. This is further supported in that the presence of the sodium buffer alters hydrogen bonding networks and results in changes in the bonding patterns. Together, these results provide further insight on how NIK maintains an active conformation.

REFERENCES

REFERENCES

- (1) Bonizzi, G.; Karin, M. The Two NF-κB Activation Pathways and Their Role in Innate and Adaptive Immunity. *Trends Immunol.* **2004**, *25* (6), 280–288.
- (2) Courtois, G.; Gilmore, T. D. Mutations in the NF-kappaB Signaling Pathway: Implications for Human Disease. *Oncogene* **2006**, *25* (51), 6831–6843.
- (3) Hoesel, B.; Schmid, J. a. The Complexity of NF-κB Signaling in Inflammation and Cancer. *Mol. Cancer* 2013, *12* (1), 86.
- (4) Rayet, B.; Gelinas, C. Aberrant Rel/nfkb Genes and Activity in Human Cancer. *Oncogene* **1999**, *18* (49), 6938–6947.
- (5) Karin, M. NF-kappaB and Cancer: Mechanisms and Targets. *Mol. Carcinog.* **2006**, *45* (6), 355–361.
- (6) Naugler, W. E.; Karin, M. NF-kappaB and Cancer-Identifying Targets and Mechanisms. *Curr. Opin. Genet. Dev.* **2008**, *18* (1), 19–26.
- Karin, M.; Lin, A. NF-kappaB at the Crossroads of Life and Death. *Nat. Immunol.* 2002, 3 (3), 221–227.
- (8) Saccani, S.; Pantano, S.; Natoli, G. Modulation of NF-κB Activity by Exchange of Dimers. *Mol. Cell* 2003, 11 (6), 1563–1574.
- (9) Wang, V. Y.-F.; Huang, W.; Asagiri, M.; Spann, N.; Hoffmann, A.; Glass, C.; Ghosh, G. The Transcriptional Specificity of NF-κB Dimers Is Coded within the κB DNA Response Elements. *Cell Rep.* **2012**, *2* (4), 824–839.
- (10) Chen, F. E.; Ghosh, G. Regulation of DNA Binding by Rel/NF-kappaB Transcription Factors: Structural Views. *Oncogene* **1999**, *18* (49), 6845–6852.
- (11) Wietek, C.; O'Neill, L. A. J. Diversity and Regulation in the NF-κB System. *Trends Biochem. Sci.* 2007, *32* (7), 311–319.
- (12) Thanos, D.; Maniatis, T.; Il-, P. M. a. NF-KB: A Lesson in Family Values Minireview. *Cell* **1995**, *80*, 529–532.
- (13) Chariot, A. The NF-κB-Independent Functions of IKK Subunits in Immunity and Cancer. *Trends Cell Biol.* 2009, 19 (8), 404–413.
- (14) Xiao, G.; Harhaj, E. W.; Sun, S.-C. NF-κB-Inducing Kinase Regulates the Processing of NF-κB2 p100. *Mol. Cell* 2001, 7 (2), 401–409.
- (15) Sun, S.-C. Non-Canonical NF-κB Signaling Pathway. Cell Res. 2011, 21 (1), 71–85.
- (16) Ling, L.; Cao, Z.; Goeddel, D. V. NF-kappaB-Inducing Kinase Activates IKK-Alpha by Phosphorylation of Ser-176. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (7), 3792–3797.

- (17) Malinin, N. L.; Boldin, M. P.; Kovalenko, A. V; Wallach, D. MAP3K-Related Kinase Involved in NF-kappaB Induction by TNF, CD95 and IL-1. *Nature* **1997**, *385* (6616), 540–544.
- (18) Ramakrishnan, P.; Wang, W.; Wallach, D. Receptor-Specific Signaling for Both the Alternative and the Canonical NF-kB Activation Pathways by NF-kB-Inducing Kinase. *Immunity* 2004, 21 (4), 477–489.
- (19) Smith, C.; Andreakos, E.; Crawley, J. B.; Brennan, F. M.; Feldmann, M.; Foxwell, B. M. J. NF- B-Inducing Kinase Is Dispensable for Activation of NF- B in Inflammatory Settings but Essential for Lymphotoxin Receptor Activation of NF- B in Primary Human Fibroblasts. *J. Immunol.* 2001, *167* (10), 5895–5903.
- (20) Thu, Y. M.; Richmond, A. NF-κB Inducing Kinase: A Key Regulator in the Immune System and in Cancer. *Cytokine Growth Factor Rev.* **2010**, *21* (4), 213–226.
- (21) Keats, J. J.; Fonseca, R.; Chesi, M.; Schop, R.; Baker, A.; Chng, W.-J.; Van Wier, S.; Tiedemann, R.; Shi, C.-X.; Sebag, M.; Braggio, E.; Henry, T.; Zhu, Y.-X.; Fogle, H.; Price-Troska, T.; Ahmann, G.; Mancini, C.; Brents, L. a.; Kumar, S.; Greipp, P.; Dispenzieri, A.; Bryant, B.; Mulligan, G.; Bruhn, L.; Barrett, M.; Valdez, R.; Trent, J.; Stewart, a. K.; Carpten, J.; Bergsagel, P. L. Promiscuous Mutations Activate the Noncanonical NF-κB Pathway in Multiple Myeloma. *Cancer Cell* **2007**, *12* (2), 131–144.
- (22) Yin, L.; Wu, L.; Wesche, H.; Arthur, C. D.; White, J. M.; Goeddel, D. V; Schreiber, R. D. Defective Lymphotoxin-Beta Receptor-Induced NF-kappaB Transcriptional Activity in NIK-Deficient Mice. *Science* 2001, 291 (5511), 2162–2165.
- (23) de Leon-Boenig, G.; Bowman, K. K.; Feng, J. a; Crawford, T.; Everett, C.; Franke, Y.; Oh, A.; Stanley, M.; Staben, S. T.; Starovasnik, M. a; Wallweber, H. J. a; Wu, J.; Wu, L. C.; Johnson, A. R.; Hymowitz, S. G. The Crystal Structure of the Catalytic Domain of the NF-κB Inducing Kinase Reveals a Narrow but Flexible Active Site. *Structure* **2012**, *20* (10), 1704–1714.
- (24) Xiao, G.; Sun, S. C. Negative Regulation of the Nuclear Factor Kappa B-Inducing Kinase by a Cis-Acting Domain. *J. Biol. Chem.* **2000**, 275 (28), 21081–21085.
- (25) Tao, Z.; Ghosh, G. Understanding NIK Regulation from Its Structure. *Structure* **2012**, *20* (10), 1615–1617.
- (26) Liu, J.; Sudom, A.; Min, X.; Cao, Z.; Gao, X.; Ayres, M.; Lee, F.; Cao, P.; Johnstone, S.; Plotnikova, O.; Walker, N.; Chen, G.; Wang, Z. Structure of the Nuclear Factor κB-Inducing Kinase (NIK) Kinase Domain Reveals a Constitutively Active Conformation. *J. Biol. Chem.* **2012**, 287 (33), 27326–27334.
- (27) Byron, O.; Vestergaard, B. Protein–protein Interactions: A Supra-Structural Phenomenon Demanding Trans-Disciplinary Biophysical Approaches. *Curr. Opin. Struct. Biol.* 2015, 35, 76–86.

- (28) Keskin, O.; Nussinov, R. Similar Binding Sites and Different Partners: Implications to Shared Proteins in Cellular Pathways. *Structure* **2007**, *15* (3), 341–354.
- (29) Rothweiler, U.; Åberg, E.; Johnson, K. a.; Hansen, T. E.; Jørgensen, J. B.; Engh, R. a. P38α MAP Kinase Dimers with Swapped Activation Segments and a Novel Catalytic Loop Conformation. J. Mol. Biol. 2011, 411 (2), 474–485.
- (30) Li, K.; McGee, L. R.; Fisher, B.; Sudom, A.; Liu, J.; Rubenstein, S. M.; Anwer, M. K.; Cushing, T. D.; Shin, Y.; Ayres, M.; Lee, F.; Eksterowicz, J.; Faulder, P.; Waszkowycz, B.; Plotnikova, O.; Farrelly, E.; Xiao, S. H.; Chen, G.; Wang, Z. Inhibiting NF-κB-Inducing Kinase (NIK): Discovery, Structure-Based Design, Synthesis, Structure-Activity Relationship, and Co-Crystal Structures. *Bioorganic Med. Chem. Lett.* **2013**, *23* (5), 1238– 1244.
- (31) Castanedo, G. M.; Blaquiere, N.; Beresini, M.; Bravo, B.; Brightbill, H.; Chen, J.; Cui, H.-F.; Eigenbrot, C.; Everett, C.; Feng, J.; Godemann, R.; Gogol, E.; Hymowitz, S.; Johnson, A.; Kayagaki, N.; Kohli, P. B.; Knüppel, K.; Kraemer, J.; Krüger, S.; Loke, P.; McEwan, P.; Montalbetti, C.; Roberts, D. A.; Smith, M.; Steinbacher, S.; Sujatha-Bhaskar, S.; Takahashi, R.; Wang, X.; Wu, L. C.; Zhang, Y.; Staben, S. T. Structure-Based Design of Tricyclic NF-κB Inducing Kinase (NIK) Inhibitors That Have High Selectivity over Phosphoinositide-3-Kinase (PI3K). *J. Med. Chem.* 2017, 60 (2), 627–640.
- (32) Mortier, J.; Masereel, B.; Remouchamps, C.; Ganeff, C.; Piette, J.; Frederick, R. NF-κB Inducing Kinase (NIK) Inhibitors: Identification of New Scaffolds Using Virtual Screening. *Bioorg. Med. Chem. Lett.* **2010**, 20 (15), 4515–4520.
- (33) Johnson, L. N.; Noble, M. E. M.; Owen, D. J. Active and Inactive Protein Kinases: Structural Basis for Regulation. *Cell (Cambridge, Massachusetts)* **1996**, *85* (2), 149–158.
- (34) Jones, M. R.; Liu, C.; Wilson, A. K. Molecular Dynamics Studies of the Protein–Protein Interactions in Inhibitor of κB Kinase-β. J. Chem. Inf. Model. 2014, 54 (2), 562–572.
- (35) Zhang, J. Protein-Protein Interactions in Salt Solutions. In *Protein-Protein Interactions Computational and Experimental Tools*; Cai, W., Ed.; InTech, 2012.
- (36) Date, M. S.; Dominy, B. N. Modeling the Influence of Salt on the Hydrophobic Effect and Protein Fold Stability. *Commun. Comput. Phys.* **2013**, *13* (1), 90–106.
- (37) Möller, J.; Schroer, M. A.; Erlkamp, M.; Grobelny, S.; Paulus, M.; Tiemeyer, S.; Wirkert, F. J.; Tolan, M.; Winter, R. The Effect of Ionic Strength, Temperature, and Pressure on the Interaction Potential of Dense Protein Solutions: From Nonlinear Pressure Response to Protein Crystallization. *Biophys. J.* 2012, *102* (11), 2641–2648.
- (38) Kříž, Z.; Klusák, J.; Krištofiková, Z.; Koča, J. How Ionic Strength Affects the Conformational Behavior of Human and Rat Beta Amyloids – A Computational Study. *PLoS One* 2013, 8 (5), e62914.
- (39) Aykut, A. O.; Atilgan, A. R.; Atilgan, C. Designing Molecular Dynamics Simulations to

Shift Populations of the Conformational States of Calmodulin. *PLoS Comput. Biol.* **2013**, *9* (12), e1003366.

- (40) Kony, D. B.; Hünenberger, P. H.; van Gunsteren, W. F. Molecular Dynamics Simulations of the Native and Partially Folded States of Ubiquitin: Influence of Methanol Cosolvent, pH, and Temperature on the Protein Structure and Dynamics. *Protein Sci.* 2007, *16* (6), 1101–1118.
- (41) Cheung, C. Y.; Ko, B. C. NFAT5 in Cellular Adaptation to Hypertonic Stress -Regulations and Functional Significance. J. Mol. Signal. 2013, 8 (1), 5.
- (42) Fischer, O. M.; Hart, S.; Gschwind, A.; Ullrich, A.; Prenzel, N. Oxidative and Osmotic Stress Signaling in Tumor Cells Is Mediated by ADAM Proteases and Heparin-Binding Epidermal Growth Factor Oxidative and Osmotic Stress Signaling in Tumor Cells Is Mediated by ADAM Proteases and Heparin-Binding Epidermal Growth Factor. *Mol. Cell. Biol.* 2004, 24 (12), 5172–5183.
- (43) Nagy, I. Z.; Lustyik, G.; Nagy, V. Z.; Zarándi, B.; Bertoni-Freddari, C. Intracellular Na+:K+ Ratios in Human Cancer Cells as Revealed by Energy Dispersive X-Ray Microanalysis. J. Cell Biol. 1981, 90 (3), 769–777.
- (44) Uhlik, M. T.; Abell, A. N.; Johnson, N. L.; Sun, W.; Cuevas, B. D.; Lobel-Rice, K. E.; Horne, E. a; Dell'Acqua, M. L.; Johnson, G. L. Rac-MEKK3-MKK3 Scaffolding for p38 MAPK Activation during Hyperosmotic Shock. *Nat. Cell Biol.* 2003, 5 (12), 1104–1110.
- (45) Chemical Computing Group Inc. Molecular Operating Environment (MOE). Scientific Computing & Instrumentation. Chemical Computing Group Inc: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7 2016.
- (46) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. J. Chem. Phys. 2004, 120 (20), 9665–9678.
- (47) Case, D. A.; Darden, T.; Iii, T. E. C.; Simmerling, C.; Brook, S.; Roitberg, A.; Wang, J.; Southwestern, U. T.; Duke, R. E.; Hill, U.; Luo, R.; Irvine, U. C.; Roe, D. R.; Walker, R. C.; Legrand, S.; Swails, J.; Cerutti, D.; Kaus, J.; Betz, R.; Wolf, R. M.; Merz, K. M.; State, M.; Seabra, G.; Janowski, P.; Paesani, F.; Liu, J.; Wu, X.; Steinbrecher, T.; Gohlke, H.; Homeyer, N.; Cai, Q.; Smith, W.; Mathews, D.; Salomon-ferrer, R.; Sagui, C.; State, N. C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Berryman, J.; Kollman, P. A. AMBER 14. University of California, San Francisco.
- (48) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* 2006, 65 (3), 712–725.
- (49) Prall, M. VMD: A Graphical Tool for the Modern Chemists. J. Comput. Chem. 2001, 22 (1), 132–134.

- (50) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput. 2013, 9 (7), 3084– 3095.
- (51) Miller III, B. R.; McGee Jr., T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* 2012, 8 (9), 3314–3321.
- (52) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246* (1,2), 122–129.
- (53) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100* (51), 19824–19839.
- (54) Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2001**, *56* (4), 275–291.
- (55) Onufriev, A.; Bashford, D.; Case, D. a. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins Struct. Funct. Bioinforma.* 2004, 55 (2), 383–394.
- (56) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372* (3), 774–797.
- (57) Pechmann, S.; Levy, E. D.; Tartaglia, G. G.; Vendruscolo, M. Physicochemical Principles That Regulate the Competition between Functional and Dysfunctional Association of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (25), 10159–10164.
- (58) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci.* U. S. A. **1996**, *93* (1), 13–20.
- (59) Marsh, J. a.; Teichmann, S. a. Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding. *Structure* **2011**, *19* (6), 859–867.
- (60) Palma, R.; Curmi, P. M. G. Computational Studies on Mutant Protein Stability: The Correlation between Surface Thermal Expansion and Protein Stability. *Protein Sci.* 2008, 8 (4), 913–920.
- (61) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. A Dissection of Specific and Non-Specific Protein-Protein Interfaces. *J. Mol. Biol.* **2004**, *336* (4), 943–955.
- (62) Goodsell, D. S.; Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105–153.

CHAPTER 5 Selectivity in ROS-Induced Peptide Backbone Bond Cleavage²

5.1 Introduction

Radical-mediated modifications to lipids, proteins, and nucleic acids have been identified to influence the progression of diseases, including degenerative disorders, cancers, diabetes as well as aging.¹⁻¹⁰ Damaged nucleic acids and lipids have specific repair mechanisms; however, damaged proteins are usually marked for degradation. The removal pathways for defective proteins are not always efficient and can lead to accumulations that interfere with cellular functions.¹¹ Irreversible backbone and side chain oxidations can occur as a result of exposure to different types of radical species, including reactive oxygen species (ROS), reactive nitrogen species (RNS), and species generated from the radiolysis of water.^{12,13} Production of ROS in organisms occurs naturally resulting from cell metabolism under normal physiological conditions and can be enriched in alternative environments induced by stress.¹⁴⁻¹⁹

Although many external sources of ROS exist, ROS can be produced in excess by numerous physiological processes.^{2,14} Normally, the prooxidative activity is balanced with antioxidative defense mechanisms.^{14,20,21} Imbalances induced from excessive ROS result in damaged proteins and other biomolecules necessary for maintaining healthy human physiological functions. Two hypotheses suggest why amyloid deposits come into existence after this damage occurs: the amyloid cascade hypothesis and an alternative hypothesis.^{22,23} In the amyloid cascade hypothesis, β -amyloid is viewed as the cause of neurodegeneration, whereas, in the alternative hypothesis, β -amyloid is the body's defense mechanism against imbalanced stress. One example

² This chapter is presented in its entirety from: Stringfellow, H. M.; Jones, M. R.; Green, M. C.; Wilson, A. K.; Francisco, J. S. Selectivity in ROS-Induced Peptide Backbone Bond Cleavage. *J. Phys. Chem. A* **2014**, *118* (48), 11399–11404.

of excess ROS production occurs in dysfunctional or defective complex III of the mitochondrial electron transport chain.²⁴⁻²⁹ In humans, neuronal cells secrete extracellular proteins in misfolded agglomerations (a structure termed β -amyloid) which function as an extracellular shield against more deleterious lipid oxidation capable of damaging vital cellular membranes.^{18-20,30} The initial deposits of β -amyloids incite a positive feedback mechanism that increases β -amyloid production.³¹ Higher concentrations of β -amyloid in the brain could result in the dysfunction of neuronal cells, leading to neurodegeneration and dementia that eventually lead to an Alzheimer's diagnosis.^{3,32-35} A generic β -amyloid structure contains multiple β -pleated sheet motifs that contribute to its structural stability.³⁶⁻³⁹ These β -pleated sheets consist of straight " β -strand" components connected by random coils (" β -turn" structures) that are positioned in either parallel or antiparallel orientations to each other.⁴⁰

All protein side chains are vulnerable towards oxidation; however, each side chain contributes to a different reactivity with ROS. For example, amino acids containing sulfurs are defenseless towards oxidation, in which cysteines and methionines are amended to disulfides and sulfoxide derivatives, respectively.^{2,3,41} Protein backbones can be oxidized directly by ROS as well as by internal free-radical transfers⁴² and are known to undergo fragmentation upon oxidation. Berlett and Stadtman suggested that proteins undergo oxidation upon attack by ROS generated by Fenton catalysis.^{2,4} Superoxide radicals attack the protein deposits at the α -carbon sites to form alkoxyl radicals (Figure 5.1) that further endure fragmentation via two pathways in which complete and terminated oxidation of the protein can occur. In the diamide pathway (Pathway [A]), the carbon-carbon bond of the peptide backbone (i) can undergo homolytic cleavage to produce a diamide (ii) and an isocyanate radical (iii), which then undergoes termination to form an isocyanate (iv). Alternatively in the α -amidation pathway (Pathway [B]), the carbon-nitrogen

bond encompassing the α -carbon site of radicalization can undergo homolytic cleavage to produce an N- α -ketoacyl derivative (v) and an amide radical, which terminates via the addition of peroxyl radical to form an amide (vii). The present study focuses on the selectivity of these two pathways that bifurcate from the formation of the alkoxyl radical at each α -carbon site and also examines the selectivity of each potential radicalization site.



Figure 5.1 Proposed reaction pathways. The diamide (pathway [A]) and α -amidation (pathway [B]) pathways were modeled in this study. The different sites of radicalization C₁, C₂, and C₃ represent C α 1, C α 2 and C α 3, respectively.

Quantum mechanics and molecular mechanics have provided insight about general peptide structures. Most investigations of small peptides focus on the folding and misfolding mechanisms rather than on structural disruptions.⁴³⁻⁴⁸ A challenge for *ab initio* modeling of peptides is to simulate all of the appropriate features of a generic peptide without allowing size limitations to alter any of those intrinsic characteristics. As the molecule size directly affects the computational expense of first principles calculations, it is important to maintain balances among the molecular

modeling options without affecting relevant chemical features of the peptide. *Ab initio* works by Francisco and co-workers show that the α -carbon site is the preferred site of radical attack in a capped tripeptide model.^{49,50} The 42 atom capped tripeptide model is an appropriate molecular size for *ab inito* methods given the computational cost of the methods chosen. Previously, it has been demonstrated that capping the ends of smaller peptide structures is important to avoid introducing additional ionic influences into the energies of neutral peptide systems modeled quantum mechanically.^{46,51,52} In terms of side-chains, peptide models of trialanine structures have been shown to replicate the β -turn features of the β -pleated sheet;⁴⁵ larger side-chain functional groups, however, have the potential to produce steric hindrance that limits radical formation. Although glycine residues have been observed to be less reactive than alanine, they have also been found to undergo intramolecular α -hydrogen abstraction reactions, introducing additional charge onto the peptide structure.⁵³⁻⁵⁵

In this study, pathways [A] and [B] were modeled to determine the effects of ROS attack at N-terminal, internal, and C-terminal carbon sites. Two models representing a β -strand and a β turn (both important motifs of the β -amyloid fibril) were compared for each pathway to elucidate potential site preference for backbone bond cleavage. The results of these calculations suggest a thermodynamic preference for pathway [A]. The impact of method and basis set choice on the energetics of the system were also investigated.

5.2 Methodology

The model system was constructed of 42 atoms to mimic a peptide backbone composed of three α -carbon sites (Figure 5.2). In each model of the parent strand, one α -carbon was radicalized via the substitution of an alkoxyl radical in place of a hydrogen at each site, as per the Berlett and Stadtman mechanism.² The model features N-terminal, C-terminal, and internal α -carbon sites.

Methyl caps on both sides of the peptide minimize possible steric, charge, and α -hydrogen transfer effects. Alanine residues eliminate the possibility of side chain interactions. Both β -strand and β turn structural conformations are modeled to provide insight into possible site preference for backbone cleavage within a secondary structure. In the model system, all equilibrium geometries were optimized using second order Møller-Plesset perturbation correction (MP2)⁵⁶⁻⁵⁹ and density functional theory (DFT) using the B3LYP functional^{60,61} with the $6-31G(d)^{62,63}$ and $6-31+G(d)^{64-1}$ ⁶⁶ basis sets. To assess the effectiveness of DFT in calculating the energetics of this system, the B3LYP functional was utilized similarly with the same basis sets. Coupled cluster with singles, doubles, and perturbed triples $(CCSD(T))^{67-69}$ with the 6-31G(d) basis set were computed from MP2/6-31G(d) stationary points and served as the reference values for this system. Reaction barriers (ΔE_{rxn}) for the N-Ca and C-Ca bond fission per site and conformation were determined from optimizations of the transition states that were carried out with the respective method in conjunction with the 6-31G(d) basis set using the Berny algorithm from the parent strand with an elongated bond length of the cleaved bond. All reaction enthalpies included zero-point and thermal corrections to 298 K.

5.3 **Results and Discussion**

5.3.1 Pathway Favorability

The CCSD(T) results are summarized in Table 5.1. The results of the method and basis set comparisons are summarized in Table 5.2 and Table 5.3. The results generated through the use of both B3LYP and MP2 using 6-31G(d) and 6-31+G(d) basis sets were unable to reproduce the reference values that the CCSD(T) data generated, meaning that they were determined to be inadequate in describing this type of radical-peptide system. Thus the CCSD(T) data will be the

only data referenced in this discussion hereafter. Both pathways are exothermic, as expected from the predictions of Berlett and Stadtman.



Figure 5.2 Modeled backbone conformations. (1) Represents the β -strand or "sheet-like" component of a β -pleated-sheet. (2) The β -turn structure. In both conformations, the first α -carbon has an alkoxyl radical in place of a hydrogen at C α 1.

Table 5.1 Calculated reaction enthalpies for Pathway [A] and Pathway [B] at the CCSD(T) level of theory.

		Pathway [A]	Pa	thway [B]	
Reaction	Site	β-strand	β-turn	β-strand	β-turn
	Ca1	4.8	30.1	38.8	43.9
$\Delta H_{rxn(1)}$	Ca2	23.7	23.6	40.5	33.6
	Ca3	-0.4	31.3	30.1	36.7
	Cal	-30.4	-60.2	-33.9	-33.9
$\Delta H_{rxn(2)}$	Ca2	-48.8	-44.8	-35.2	-34.8
	Ca3	-45.6	-45.6	-18.2	-49.4
	Ca1	-25.6	-30.2	4.9	10.0
ΔH_{total}	Ca2	-25.1	-21.1	5.3	-1.2
	Ca3	-46.0	-14.3	11.8	-12.6
		C-Ca bond fission	N-	Ca bond fission	
٨F		β-strand	β-turn	β-strand	β-turn
ΔLrxn	Ca1	8.8	5.5	22.9	24.1
	Ca2	8.7	6.0	23.6	26.0
	Ca3	7.2	5.7	24.5	27.6

^aAll energies are in kcal/mol.

Overall, Pathway [A] is much more energetically favorable than Pathway [B] by 30.4 to 57.4 kcal/mol for the β -strand conformation and by 1.7 to 40.2 kcal/mol for the β -turn conformation. Pathway [A] is to be the more energetically favorable pathway because of the nature of the structures formed through the process of complete fragmentation. The carbon immediately adjacent to the alkoxyl-radicalized α -carbon (referred to from this point on as "C α ") is doubly-bonded to an O atom, so that the resulting carbonyl functional group withdraws electron density from that C α atom, straining the C α -C bond to the point of cleavage. Reaction A2 and B2, the second steps of each pathway, involve the termination of a radical on the peptide fragment structure involved in the reaction. It is striking that in Pathway [A] a radical is propagated through the production of •OOH, while in Pathway [B] both products are completely terminated.

From this study of a trialanine peptide model it has been determined that there are no absolute trends regarding conformational preference. The β-strand conformation *generally* favors fragmentation via Pathway [A], while the β-turn conformation *generally* favors fragmentation via Pathway [B]. Intramolecular noncovalent interactions are more prevalent in the β-turn structures than in the β-strand structures. In the results of our trialanine peptide study, intramolecular noncovalent interactions stabilize the biomolecule and reduce the favorability of C-Cα bond fission (ΔE_{rsn} in Table 5.1) for the β-turn conformation. These results are verified by a study conducted by Chin et al.⁴⁶ on the local secondary structure conformations of dipeptides and tripeptides, that show that tripeptides are indeed more stable in the β-turn conformation than in the β-strand conformation due to an increase in the intramolecular noncovalent interactions between the backbone and side chains. Other studies investigating reactivity and conformational specificity in peptide fragments also report similar advantages in side-chain stability from intramolecular noncovalent interactions.^{54,55} Additionally, α-carbon hydrogen abstractions from the peptide

backbones in the β -turn conformation do not induce major conformational changes in the gas phase due to intramolecular noncovalent stabilization.⁵³ However, in spite of this similarity in results for the β -turn conformation following C-C α bond fission, our results also identify a striking difference with regard to N-C α bond fission. The intramolecular noncovalent stabilization that caused the β turn conformation to have less energetic gain from undergoing C-C α bond fission does not impact the favorability of the β -turn conformation to undergo N-C α bond fission.

The C α 3 site is the most favorable site of radicalization in Pathway [A] for the β -strand conformation, but the C α 1 site is the most favorable site of radicalization for the β -turn conformation. The greatest difference between the most energetically favorable site and the next closest value is in Pathway [A], where reacting at the C α 3 site in the β -strand structure produces an enthalpy 20.4 kcal/mol lower in energy than the next most favorable site. In Pathway [B], the β -turn structure has the next greatest difference in enthalpy; reacting at the C α 3 site has a 11.4 kcal/mol advantage over the next lowest value. Site selectivity is not consistent from one conformation or reaction to the next; however, it is still an important factor in determining overall favorability of the reactions in question.

These results are corroborated with the calculated reaction barriers for the C-C α and N-C α bond fissions, as shown in Table 5.1. For all three sites of radicalization, the C-C α bond fission has a much lower reaction barrier than that of N-C α bond scission, by an average difference of 15.4 kcal/mol for the β -strand and by 20.2 kcal/mol for the β -turn. The reaction barriers reveal that the C-C α bond fission reaction barrier is consistently lower in the β -turn conformation than in the β -strand conformation by an average of 2.5 kcal/mol. Conversely, the N-C α bond fission reaction barrier is consistently lower in the β -turn conformation by an average of 2.5 kcal/mol. Conversely, the N-C α bond fission reaction barrier is consistently lower in the β -turn conformation by an average of 2.5 kcal/mol. Conversely, the N-C α bond fission reaction barrier is consistently lower in the β -turn conformation by an average of 2.5 kcal/mol. Conversely, the N-C α bond fission reaction barrier is consistently lower in the β -turn conformation by an average of 2.5 kcal/mol.

type indicates that intramolecular noncovalent stabilization contributes to the stability of β -turn reactants relative to products for Pathway [B] (which involves N-C α fission) but not for Pathway [A]. In Pathway [A], another factor must be contributing to the relative stability of the β -strand to increase the reaction barrier of the C-C α bond fission.

5.3.2 Role of Structure on Bond Strength and Site Reactivity

The three radical α -carbon sites do not harbor identical bond strengths. This is consistent with previous findings that not all α -carbon radicals are equally stable.^{16,70} The C-C α bond at site Ca3 within the β -strand conformation is the strongest C-Ca bond to be broken from the structures analyzed. The differences in enthalpy between the β -strand and β -turn models reveal that due to structural differences, the C-C α bond at C α 3 is much stronger in the β -strand conformation than in the β -turn conformation of the peptide. This is because more energy is released upon β -strand C-C α bond fission than through β -turn C-C α bond fission at the same relative position. The average energy of the C-Ca bonds in the model is 18.9 kcal/mol. The stronger C-Ca bond at the Ca2 site of the two conformations is in the β -turn structure. The data further indicates that internal C-C α bond fission releases more energy in the β -turn conformational construct than in the β -strand counterpart. The strongest N-C α bond site of all the sites is located at the C α 3 site. For this type of bond fission, the conformation of the model does not play a significant a role in the strength of the bond. The N-Ca bonds have an average energy of 37.3 kcal/mol. The C-Ca bonds have a greater range of energies throughout the molecule than do the N-C α bonds. The relative location of the backbone C-Ca bond within the peptide structure plays a greater role in determining the energy of that bond than does the relative location of the backbone N-C α bond analog.

5.3.3 Insights into Pathway Preferences

The importance of the second reaction (Reaction A2 or B2) in defining the energetic favorability of the reaction pathway varies. The isocyanate radical termination step is influential in determining the favorability of completely following reaction Pathway [A]. Similarly, in Pathway [B], the reaction enthalpy of the pathway becomes comparably favorable because of the amide radical termination step.

For all sites of ROS-induced bond fission (all three C α sites), Pathway [A] is more energetically favorable than the Pathway [B]. In Pathway [A], the location of the site at which bond fission occurs plays a key role in identifying the strongest bonds. In the β -strand, the C-C α 3 is the strongest bond, stronger than the other C-C α bonds by an average of 14.7 kcal/mol, and stronger than any of the N-C α bonds by an average of 36.9 kcal/mol. In the β -turn, the strengths of all N-C α and C-C α bonds are more similar. For the β -turn structure, the C-C α 2 bond is strongest, stronger than the other C-C α bonds by an average of 7.1 kcal/mol, and stronger than the N-C α bonds by an average of 14.5 kcal/mol. In Pathway [B], the most important step in determining the overall energetic favorability of the pathway is the re-association of the reactive amide radical.

The difference in reaction barriers of peptide backbone bond fission at each site of radicalization plays an important role in overall pathway preference. The differences in the barriers are consistent by site and conformation to determining this preference. The reaction barrier of C-C α bond fission is notably lower than that of N-C α bond fission, by an average difference of 15.4 kcal/mol in the β -strand conformation and by 20.2 kcal/mol in the β -turn conformation. These trends confer a definite energetic preference for Pathway [A] over Pathway [B]. Both the individual reaction steps and the first energy barrier of Pathway [A] are lower in energy than Pathway [B]. The present study neglects the electronic nature of neighboring amino acid side-
chains, which is known to influence backbone cleavage. Within the current approach, the sampled conformations (β -strand and β -turn) are appropriate for describing potential configurational states of the trialanine peptide due to the lack of flexibility and reactivity of the side-chain. The introduction of alternative side-chains would require a more elaborate approach that assesses conformational rearrangements and radical migrations. Assessing how the selectivity in backbone fragmentation changes as a result of neighboring side-chains is suggested as important follow-up.

5.3.4 Evaluation of Methods

Enthalpies computed using MP2 and B3LYP (Table 5.2) were able to point to the presence of selectivity between the Pathway [A] and Pathway [B]. However, these methods were unable to reproduce the selectivity trends generated by the CCSD(T) calculations (summarized in Table 5.1). Although adding diffuse functions had a slight impact on the enthalpy per each individual fragments, the additional basis functions had small significance in the prediction of the reaction enthalpies.

The uniqueness of the CCSD(T) results in describing the trends of ROS-induced peptide backbone fragmentation implies that the self-consistent single-, double-, and triple-electron coupling contributes significantly. Notably, less expensive methods to observe this chemistry are unable to replicate the reactions and thus cannot be used to correctly predict the favored sites or conformations of ROS-induced peptide backbone bond cleavage. The results also indicate that in spite of the attractiveness of using MP2, a more comprehensive method than B3LYP but a lessexpensive method than CCSD(T), to determine the relative site and conformational favorability, MP2 was not observed to provide any additional insight than B3LYP towards elucidating the trends observed with CCSD(T).

		l	3-strand			β-turn			
Site	Method/Basis	$\Delta H_{[A1]}$	$\Delta H_{[A2]}$	ΔH_{tot}	$\Delta H_{[A1]}$	$\Delta H_{[A2]}$	ΔH_{tot}		
	CCSD(T)/6-31G(d)	4.8	-30.4	-25.6	30.3	-60.2	-30.2		
	MP2/6-31G(d)	-7.4	-42.6	-50.1	2.8	-43.1	-40.3		
Ca1	MP2/6-31+G(d)	-9.6	-40.0	-49.6	4.5	-42.6	-38.2		
	B3LYP/6-31G(d)	-6.8	-53.5	-60.3	2.7	-56.8	-54.1		
	B3LYP/6-31+G(d)	-10.0	-53.3	-63.4	-1.4	-57.2	-58.6		
	CCSD(T)/6-31G(d)	23.7	-48.8	-25.1	23.6	-44.8	-21.1		
	MP2/6-31G(d)	-5.8	-43.5	-49.3	-5.3	-41.1	-46.4		
Ca2	MP2/6-31+G(d)	-7.5	-40.7	-48.2	-4.0	-40.6	-44.6		
	B3LYP/6-31G(d)	-5.4	-54.5	-59.9	-6.9	-54.6	-61.5		
	B3LYP/6-31+G(d)	-8.7	-54.3	-63.0	-11.4	-52.6	-64.0		
	CCSD(T)/6-31G(d)	-0.4	-45.6	-46.0	31.3	-45.6	-14.3		
	MP2/6-31G(d)	-7.7	-41.4	-49.1	1.4	-41.4	-40.0		
Ca3	MP2/6-31+G(d)	-8.0	-40.2	-48.2	-2.8	-40.2	-43.0		
	B3LYP/6-31G(d)	-4.1	-54.8	-58.8	0.1	-54.8	-54.6		
	B3LYP/6-31+G(d)	-7.4	-54.0	-61.4	-3.7	-54.0	-57.8		

Table 5.2 Calculated reaction enthalpies for Pathway [A] and Pathway [B].

^aAll energies are in kcal/mol.

Pathway [A]

			β-strand			β-turn			
Site	Method/Basis	$\Delta H_{[B1]}$	$\Delta H_{[B2]}$	ΔH_{tot}	$\Delta H_{[B1]}$	$\Delta H_{[B2]}$	ΔH_{tot}		
	CCSD(T)/6-31G(d)	38.8	-33.9	4.9	43.9	-33.9	10.0		
	MP2/6-31G(d)	13.5	-47.3	-33.8	18.3	-47.3	-29.0		
Ca1	MP2/6-31+G(d)	13.8	-47.3	-33.6	19.2	-47.3	-28.1		
	B3LYP/6-31G(d)	16.2	-25.9	-9.6	13.8	-21.6	-7.7		
	B3LYP/6-31+G(d)	14.2	-26.3	-12.1	10.0	-21.6	-11.6		
	CCSD(T)/6-31G(d)	40.5	-35.2	5.3	33.6	-34.8	-1.2		
	MP2/6-31G(d)	15.3	-48.9	-33.6	26.1	-49.6	-23.4		
Ca2	MP2/6-31+G(d)	16.5	-48.8	-32.4	26.3	-48.6	-22.4		
	B3LYP/6-31G(d)	13.7	-23.5	-9.8	21.7	-26.0	-4.3		
	B3LYP/6-31+G(d)	11.3	-23.3	-12.0	9.0	-25.5	-16.5		
	CCSD(T)/6-31G(d)	30.1	-18.2	11.8	36.7	-49.4	-12.6		
	MP2/6-31G(d)	24.7	-49.2	-24.5	19.4	-46.6	-27.1		
Ca3	MP2/6-31+G(d)	15.9	-49.1	-33.2	16.1	-46.6	-30.5		
	B3LYP/6-31G(d)	23.4	-23.8	-0.4	16.8	-25.4	-8.6		
	B3LYP/6-31+G(d)	20.7	-23.5	-2.7	13.2	-25.1	-12.0		

^aAll energies are in kcal/mol.

		β-st	β-strand		urn
Site	Method/Basis	$\Delta E^{\ddagger}_{[A1]}$	$\Delta \mathrm{E}^{\ddagger}$ [B1]	ΔE^{\ddagger} [A1]	ΔE^{\ddagger} [B1]
Ca1	CCSD(T)/6-31G(d)	8.8	22.9	5.5	24.1
	MP2/6-31G(d)	11.0	26.9	8.2	29.2
	B3LYP/6-31G(d)	3.5	20.9	2.1	21.4
Cα2	CCSD(T)/6-31G(d)	8.7	23.6	6.0	26.0
	MP2/6-31G(d)	10.7	28.0	10.1	35.4
	B3LYP/6-31G(d)	3.9	21.7	1.3	19.5
Cα3	CCSD(T)/6-31G(d)	7.2	24.5	5.7	27.6
	MP2/6-31G(d)	9.0	28.5	8.1	35.1
	B3LYP/6-31G(d)	2.3	20.2	1.3	19.9

Table 5.3 Calculated reaction barriers for reactions [A1] and [B1].

^aAll energies are in kcal/mol.

5.4 Conclusion

This study utilized *ab initio* and density functional methods to investigate site, pathway, and conformational trends associated with the ROS-mediated fragmentation of the trialanine peptide. The results show that the reactions associated with the two pathways of the Berlett and Stadtman mechanisms of ROS-induced peptide backbone bond cleavage are exothermic and provide a quantified estimate of their relative selectivity. Pathway [A] is consistently more energetically favorable than Pathway [B], and site preferences are not determined strictly by their positions as N-terminal, internal, or C-terminal. Energetic analysis reveals that intramolecular noncovalent stabilization contributes to conformational selection for fragmentation following Pathway [A] only. Calculation of reaction barriers confirms the energetic analysis that C-C α bond fission (following Pathway [A]) is much more favorable to occur than N-C α bond fission (following Pathway [B]). The results identify the key backbone bonds (the C-C α 3 bond in the β strand and the C-C α 2 bond in the β -turn) that contribute to the overall stability of the β -pleated sheet motif. Furthermore, because this work identifies the most favorable types of fragmentation within the β -pleated sheet, this research can be utilized to predict how secreted amyloid precursor proteins will fragment in response to ROS bombardment. The ability to predict the structural features involved in these fragmentation mechanisms is significant towards understanding the role of oxidative stress in neurodegenerative disease pathologies.

REFERENCES

REFERENCES

- (1) Nathan, C.; Cunningham-Bussel, A. Beyond Oxidative Stress: An Immunologist's Guide to Reactive Oxygen Species. *Nat. Rev. Immunol.* 2013, *13*, 349-361.
- (2) Berlett, B. S.; Stadtman, E. R. Protein Oxidation in Aging, Disease, and Oxidative Stress. *J. Biol. Chem.* 1997, 272, 20313-20316.
- (3) Stadtman, E. R.; Berlett, B. S. Reactive Oxygen-Mediated Protein Oxidation in Aging and Disease. *Drug Metab. Rev.* 1998, *30*, 225-243.
- (4) Stadtman, E. R. Role of Oxidant Species in Aging. *Curr. Med. Chem.* 2004, *11*, 1105-1112.
- (5) Cai, Z.; Yan, L. J. Protein Oxidative Modifications: Beneficial Roles in Disease and Health. *J. Biochem. Pharmacol.* 2013, *1*, 15-26.
- (6) Cecarini, V.; Gee, J.; Fioretti, E.; Amici, M.; Angeletti, M.; Eleuteri, A. M.; Keller, J. N. Protein Oxidation and Cellular Homeostasis: Emphasis on Metabolism. *Biochim. Biophys. Acta* 2007, *1773*, 93-104.
- (7) Sung, C.; Hsu, Y.; Chen, C.; Lin, Y.; Wu, C. Oxidative Stress and Nucleic Acid Oxidation in Patients with Chronic Kidney Disease. Oxid. Med. Cell. Longev. 2013, 2013, 1-15.
- (8) Quaye, I. K. Oxidative Stress in Human Health and Disease. In *Insight and Control of Infectious Disease in Global Scenario* [Online]; Priti, R., Ed. InTech: Rijeka, Croatia, 2012; Chapter 6, pp 97-120. http://www.intechopen.com/books/insight-and-control-of-infectious-disease-in-global-scenario/oxidative-stress-in-health-and-disease (accessed February 18, 2014).
- (9) Aruoma, O. I. Free Radicals, Oxidative Stress, and Antioxidants in Human Health and Disease. *J. Am. Oil Chem. Soc.* 1998, 75, 199-212.
- (10) Hybertson, B. M.; Gao, B.; Bose, S. K.; McCord, J. M. Oxidative Stress in Health and Disease: The Therapeutic Potential of Nrf2 Activation. *Mol. Aspects Med.* 2011, *32*, 234-246.
- (11) Dunlop, R. A.; Brunk, U. T.; Rodgers, K. J. Oxidized Proteins: Mechanisms of Removal and Consequences of Accumulation. *IUBMB Life* 2009, *61*, 522-527.
- (12) Sayre, L. M.; Moreira, P. I.; Smith, M. A.; Perry, G. Metal Ions and Oxidative Protein Modification in Neurological Disease. *Ann. Ist. Super. Sanita* 2005, *41*, 143-164.
- (13) Cabiscol, E.; Tamarit, J.; Ros, J. Oxidative Stress in Bacteria and Protein Damage by Reactive Oxygen Species. *Int. Microbiol.* 2000, *3*, 3-8.

- (14) Halliwell, B. Reactive Oxygen Species in Living Systems: Source, Biochemistry, and Role in Human Disease. *Am. J. Med.* 1991, *91*, 14S-22S.
- (15) Davies, K. J. Protein Damage and Degradation by Oxygen Radicals. I. General Aspects. J. Biol. Chem. 1987, 262, 9895-9901.
- (16) Davies, M. J.; Dean, R. T. Radical-mediated protein oxidation: from chemistry to medicine; Oxford science publications; Oxford University Press: Oxford; New York, 1997; pp 443.
- (17) Garrison, W. M. Reaction Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins. *Chem. Rev. (Washington, DC, U.S.)* 1987, 87, 381-398.
- (18) Imlay, J. Cellular Defenses Against Superoxide and Hydrogen Peroxide. *Annu. Rev. Biochem.* 2008, 77, 755.
- (19) Davies, M. J. The Oxidative Environment and Protein Damage. *Biochim. Biophys. Acta*. 2005, *1703*, 93-109.
- (20) Elias, R. J.; Kellerby, S. S.; Decker, E. A. Antioxidant Activity of Proteins and Peptides. *Crit. Rev. Food Sci. Nutr.* 2008, *48*, 430-441.
- (21) Gracy, R. W.; Talent, J. M.; Kong, Y.; Conrad, C. C. Reactive Oxygen Species: The Unavoidable Environmental Insult? *Mutat. Res.* 1999, 428, 17-22.
- (22) Jomova, K.; Vondrakova, D.; Lawson, M.; Valko, M. Metals, Oxidative Stress and Neurodegenerative Disorders. *Mol. Cell. Biochem.* 2010, *345*, 91-104.
- (23) Hardy, J.; Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science (Washington, DC, U.S.)* 2002, 297, 353-356.
- (24) Turrens, J. F. Mitochondrial Formation of Reactive Oxygen Species. J. Physiol. (Oxford, U.K.) 2003, 552, 335-344.
- (25) Yamamori, T.; Yasui, H.; Yamazumi, M.; Wada, Y.; Nakamura, Y.; Nakamura, H.; Inanami, O. Ionizing Radiation Induces Mitochondrial Reactive Oxygen Species Production Accompanied by Upregulation of Mitochondrial Electron Transport Chain Function and Mitochondrial Content under Control of the Cell Cycle Checkpoint. *Free Radical Biol. Med.* 2012, *53*, 260-270.
- (26) Chen, Q.; Moghaddas, S.; Hoppel, C. L.; Lesnefsky, E. J. Ischemic Defects in the Electron Transport Chain Increase the Production of Reactive Oxygen Species from Isolated Rat Heart Mitochondria. *Am. J. Physiol.* 2008, *294*, C460-C466.
- (27) Chen, Q.; Vazquez, E. J.; Moghaddas, S.; Hoppel, C. L.; Lesnefsky, E. J. Production of Reactive Oxygen Species by Mitochondria. *J. Biol. Chem.* 2003, 278, 36027-36031.

- (28) Sanz, A.; Stefanatos, R.; McIlroy, G. Production of Reactive Oxygen Species by the Mitochondrial Electron Transport Chain in Drosophila Melanogaster. *J. Bioenerg. Biomembr.* 2010, *42*, 135-142.
- (29) Liu, Y.; Fiskum, G.; Schubert, D. Generation of Reactive Oxygen Species by the Mitochondrial Electron Transport Chain. *J. Neurochem.* 2002, *80*, 780-787.
- (30) Wang, L. L.; Xiong, Y. L. Inhibition of Lipid Oxidation in Cooked Beef Patties by Hydrolyzed Potato Protein Is Related to Its Reducing and Radical Scavenging Ability. J. Agric. Food Chem. 2005, 53, 9186-9192.
- (31) Qin, L.; Liu, Y.; Cooper, C.; Liu, B.; Wilson, B.; Hong, J. Microglia Enhance β-Amyloid Peptide-Induced Toxicity in Cortical and Mesencephalic Neurons By Producing Reactive Oxygen Species. J. Neurochem. 2002, 83, 973-983.
- (32) Davies, M. J.; Fu, S.; Wang, H.; Dean, R. T. Stable Markers of Oxidant Damage to Proteins and Their Application in the Study of Human Disease. *Free Radical Biol. Med.* 1999, 27, 1151-1163.
- (33) Allan Butterfield, D. Amyloid β-Peptide (1-42)-Induced Oxidative Stress and Neurotoxicity: Implications for Neurodegeneration in Alzheimer's Disease Brain. A Review. *Free Radical Res.* 2002, *36*, 1307-1313.
- (34) Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. Ø; Riekel, C. The Structural Biology of Protein Aggregation Diseases: Fundamental Questions and Some Answers. *Acc. Chem. Res.* 2006, *39*, 568-575.
- (35) Conrad, C. C.; Marshall, P. L.; Talent, J. M.; Malakowsky, C. A.; Choi, J.; Gracy, R.
 W. Oxidized Proteins in Alzheimer's Plasma. *Biochem. Biophys. Res. Commun.* 2000, 275, 678-681.
- (36) Qiang, W.; Yau, W.; Luo, Y.; Mattson, M. P.; Tycko, R. Antiparallel β-Sheet Architecture in Iowa-Mutant β-Amyloid Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109, 4443-4448.
- (37) Paravastu, A. K.; Leapman, R. D.; Yau, W.; Tycko, R. Molecular Structural Basis for Polymorphism in Alzheimer's β-Amyloid Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 18349-18354.
- Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. 3D Structure of Alzheimer's Amyloid-β(1-42) Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* 2005, *102*, 17342-17347.
- (39) Coles, M.; Bicknell, W.; Watson, A. A.; Fairlie, D. P.; Craik, D. J. Solution Structure of Amyloid β-Peptide(1-40) in a Water-Micelle Environment. Is the Membrane-Spanning Domain Where We Think It Is? *Biochemistry* 1998, *37*, 11064-11077.

- (40) Soto, C. Plaque Busters: Strategies to Inhibit Amyloid Formation in Alzheimer's Disease. *Mol. Med. Today* 1999, *5*, 343-350.
- (41) Levine, R. L.; Moskovitz, J.; Stadtman, E. R. Oxidation of Methionine in Proteins: Roles in Antioxidant Defense and Cellular Regulation. *IUBMB Life* 2000, *50*, 301-307.
- (42) Faraggi, M.; Tal, Y. The Reaction of the Hydration Electron with Amino Acid, Peptides, and Proteins in Aqueous Solution: II. Formation of Radicals and Electron Transfer Reactions. *Radiat. Res.* 1975, 62, 347.
- (43) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Peptide Folding Simulations: No Solvent Required? *Comput. Phys. Commun.* 1999, *123*, 97-102.
- (44) Lamm, M. S.; Rajagopal, K.; Schneider, J. P.; Pochan, D. J. Laminated Morphology of Nontwisting β-Sheet Fibrils Constructed via Peptide Self-Assembly. *J. Am. Chem. Soc.* 2005, *127*, 16692-16700.
- (45) Tashiro, S.; Kobayashi, M.; Fujita, M. Folding of an Ala-Ala-Ala Tripeptide into a β-Turn via Hydrophobic Encapsulation. J. Am. Chem. Soc. 2006, 128, 9280-9281.
- (46) Chin, W.; Piuzzi, F.; Dimicoli, I.; Mons, M. Probing the Competition Between Secondary Structres and Local Preferences in Gas Phase Isolated Peptide Backbones. *Phys. Chem. Chem. Phys.* 2006, *8*, 1033-1048.
- (47) Azimi, S.; Rauk, A. The Binding of Fe(II)-Heme to the Amyloid Beta Peptide of Alzheimer's Disease: QM/MM Investigations. J. Chem. Theory Comput. 2012, 8, 5150-5158.
- (48) Bakker, J. M.; Plützer, C.; Hünig, I.; Häber, T.; Compagnon, I.; von Helden, G.; Meijer, G.; Kleinermanns, K. Folding Structures of Isolated Peptides as Revealed by Gas-Phase Mid-Infrared Spectroscopy. *ChemPhysChem* 2005, *6*, 120-128.
- (49) Doan, H. Q.; Davis, A. C.; Francisco, J. S. Primary Steps in the Reaction of OH Radicals with Peptide System: Perspective from a Study of Model Amides. J. Phys. Chem. A 2010, 114, 5342-5357.
- (50) Green, M. C.; Federov, D. G.; Kitaura, K.; Francisco, J. S.; Slipchenko, L. V. Open-Shell Pair Interaction Energy Decomposition Analysis (PIEDA): Formulation and Application to the Hydrogen Abstraction in Tripeptides. *J. Chem. Phys.* 2013, *138*, 074111.
- (51) Green, M. C.; Stelzleni, S.; Francisco, J. S. Spectral Marker for C_{α} Damage in Beta Peptides. *J. Phys. Chem. A* 2013, *117*, 550-565.
- (52) Turecek, F.; Syrstad, E. A.; Seymour, J. L.; Chen, X.; Yao, C. Peptide Cation-Radicals. A Computational Study of the Competition Between Peptide N-C_α Bond Cleavage and Loss of the Side Chain in the [GlyPhe-NH2 + 2H]^{+•}. J. Mass Spectrom. 2003, 38, 1093-1104.

- (53) Cheng, W.; Jang, S.; Wu, C.; Lin, R.; Lu, H.; Li, F. Site Specificity of the ^αC-H Bond Dissociation Energy for a Naturally Occurring β-Hairpin Peptide An *Ab Initio* Study. *J. Comput. Chem.* 2009, *30*, 407-414.
- (54) Bhattacharya, A.; Clawson, K. J.; Bernstein, E. R. Influence of Turn (or Fold) and Local Charge in Fragmentation of the Peptide Analogue Molecule CH₃CO-Gly-NH₂ Following Single-Photon VUV (118.22 nm) Ionization. *J. Phys. Chem. A* 2011, *115*, 10679-10688.
- (55) Bhattacharya, A.; Shin, J.; Clawson, K. J.; Bernstein, E. R. Conformation Specific and Charge Directed Reactivity of Radical Cation Intermediates of α-Substituted (Amino, Hydroxy, and Keto) Bioactive Carboxylic Acids. *Phys. Chem. Chem. Phys.* 2010, *12*, 9700-9712.
- (56) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation By Direct Methods. *Chem. Phys. Lett.* 1988, *153*, 503-506.
- (57) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A Direct MP2 Gradient Method. *Chem. Phys. Lett.* 1990, *166*, 275-280.
- (58) Saebø, S.; Almlöf, J. Avoiding the Integral Storage Bottleneck in LCAO Calculations of Electron Correlation. *Chem. Phys. Lett.* 1989, *154*, 83-89.
- (59) Head-Gordon, M.; Head-Gordon, T. Analytic MP2 Frequencies Without Fifth-Order Storage. Theory and Application to Bifurcated Hydrogen Bonds in the Water Hexamer. *Chem. Phys. Lett.* 1994, 220, 122-128.
- (60) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* 1993, 98, 5648.
- (61) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B.* 1988, *37*, 785.
- (62) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta.* 1973, 28, 213-222.
- (63) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements. J. Chem. Phys. 1982, 77, 3654-3665.
- (64) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li-F. J. Comput. Chem. 1983, 4, 294-301.
- (65) Krishnan, R.; Frisch, M. J.; Pople, J. A. Contribution of Triple Substitutions to the Electron Correlation Energy in Fourth Order Perturbation Theory. *J. Chem. Phys.* 1980, 72, 4244.

- (66) Gill, P. M. W.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. The Performance of the Becke-Lee-Yang-Parr (B-LYP) Density Functional Theory with Various Basis Sets. *Chem. Phys. Lett.* 1992, 197, 499-505.
- (67) Cížek, J. On the Use of the Cluster Expansion and the Technique of Diagrams in Calculations of Correlation Effects in Atoms and Molecules. In *Advances in Chemical Physics;* Hariharan, P.C., Ed.; Wiley Interscience: New York, 1969; Vol. 14, p. 35.
- (68) Purvis, G. D., III; Bartlett, R. G. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* 1982, *76*, 1910-1918.
- (69) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* 1989, *157* (6), 479–483.
- (70) Headlam, H. A.; Mortimer, A.; Easton, C. J.; Davies, M. J. β-Scission of C-3 (β-Carbon) Alkoxyl Radicals on Peptides and Proteins: A Novel Pathway Which Results in the Formation of α-Carbon Radicals and the Loss of Amino Acid Side Chains. *Chem. Res. Toxicol.* 2000, *13*, 1087-1095.

CHAPTER 6 Partition Coefficients for the SAMPL5 Challenge using Transfer Free Energies

6.1 Introduction

Theoretical approaches can be a useful partner for predicting physiochemical properties important in experimental design, from drug development, to studies in toxicology and environmental science. Such predictive approaches can provide guidance in high-throughput screening for rational design.^{1,2} An important step in establishing such utility, however, is to ensure that the approaches are suitable and well-tested.

A popular route towards rationale design is the use of mathematical models to determine quantitative structure-activity relationships (QSAR) or quantitative structure-property relationships (QSPR), based on parameters or descriptors that correlate physical and chemical properties with experimental observations. Though useful, it is difficult for these predictive models to determine properties coupled with changes in the electronic environment such as those that arise from solute-solute and solute-solvent interactions.

Measurements of solvation properties of molecules in different solution phases and the distribution equilibria between the phases are of strong interest in drug discovery as these properties are critical for drug profiling. Promising drug candidates can be discarded as a result of inaccurate predictions of such properties; particularly of interest are the partition (P) and distribution (D) coefficients. The partition coefficient (P) between two phases *x* and *y*, respectively, is the ratio of the concentration of solute C in each phase, where the subscript 0 represents the neutral, unionized state of the solute.

$$P = \frac{[C_0]_y}{[C_0]_x} \qquad D = \frac{\sum_i [C_i]_y}{\sum_i [C_i]_x}$$

In contrast, the distribution coefficient (D) is the actual partitioning or distribution of the total analytical concentration between two solvents at a fixed pH, which includes all chemical states in solution (i.e., unionized and ionized states). Although many QSAR and QSPR techniques exist, because these approaches are based on defining statistical relationships via parameters trained to fit known experimental properties, this may not be the most suitable approach for predicting properties for molecules that undergo chemical transformations in different environments, such as compounds with multiple protonation states or compounds that can undergo structural rearrangements.^{3–5} Additionally, many QSAR and QSPR models are parametrized to predict octanol/water partition coefficients as there is an abundance of reliable experimental data for available to parameterize this predictive approach. For this, it is ideal to have a simple approach that relies on less parameterization and is transferable for use with other solvents.

The partition coefficient between two immiscible solvents, such as water and cyclohexane, is expressed as the equilibrium distribution between the concentrations of a solute in each solvent and is related to the change in energy associated with the solute-solvent interactions, which is expressed as the free energy difference, ΔG , of the solute in each solvent,

$$\log P = \log \left(\frac{[\text{solute}]_{\text{cyclohexane}}}{[\text{solute}]_{\text{water}}} \right) = \left(\Delta G_{\text{water}} - \Delta G_{\text{cyclohexane}} \right) \frac{\log_{10} e}{kT}$$

where e is Euler's number, k is Boltzmann's constant, and T is temperature. Predicting the free energy of solvation for a molecule using chemometric techniques can be challenging due to the difficulties in parameterizing the solute-solvent interactions, with the many intermolecular forces that contribute to solvation.

Another route that can be used to predict the partition coefficient is to use quantum mechanical (QM) approaches, accounting for solvation via either an explicit or implicit route. For explicit solvation, individual solvent molecules are included in a calculation. While these

approaches can account for non-covalent solute-solvent interactions, they can become computationally intensive due to the number of solvent molecules that may be needed within the solvation shell as well as the amount of sampling that would be required. For implicit solvation, a representation of the solvent is utilized which neglects explicit contributions of the solute-solvent interactions, resulting in a lower computational cost approach. Implicit solvation approaches, continuum solvent models, are commonly used and include the Conductor-like Polarizable Continuum Model (CPCM)⁶, the COnductor-like Screening MOdel (COSMO)⁷, and the Solvation Model based on Density (SMD)⁸, for predicting the free energies of hydration. Of these solvation models, SMD is a more portable model for the prediction of solvation free energies, as it uses the electron density of the solute in contrast to partial charges as used in CPCM and COSMO. In previous studies predicting the free energies of solvation of small molecules, it has been shown that the increasing quality of a basis set can improve the prediction, including *ab initio* approaches⁹ as well as in hybrid QM and molecular mechanics (MM) approaches such as the QM/MM-non-Boltzmann-Bennett method.^{10,11}

For this investigation, emphasis is on the 13 molecule subset (Batch 0) of the SAMPL5 distribution coefficient molecule set. Using this subset, a variety of hybrid DFT functionals with different basis sets were used to predict the cyclohexane-water partition coefficient. A vertical solvation approach, using gas-phase optimized geometries to predict the free energy in solution, was used for predicting the cyclohexane-water transfer free energy needed for the calculation of the partition coefficient. The approach that was chosen was applied to the full molecule set of 53 compounds, which corresponds to Submission #40 in the SAMPL5 Distribution Coefficient Challenge.

6.2 Methods

The initial structures issued with the SAMPL5 challenge data set were used as the reference state for all calculations. Gas phase geometry optimizations and frequency calculations were performed on the 53 molecules of the SAMPL5 set (shown in Error! Reference source not found.) using B3LYP¹²⁻¹⁴ in conjunction with cc-pVTZ basis sets.¹⁵⁻¹⁸ B3LYP/cc-pVTZ was selected due to its well-established success in the prediction of ground state gas-phase structures for molecules such as those in the SAMPL5 set. Frequencies were examined to ensure that equilibrium stationary points were reached. For second-row species such as sulfur, the recommended form of the correlation consistent basis sets, the augmented tight-d basis sets, ccpV(T+d)Z,¹⁹ was used to avoid the deficiencies noted in the original form of the correlation consistent basis sets for second-row atoms. The correlation consistent basis sets were selected due to their demonstrated behavior with a broad range of functionals^{20–29}, known to converge with respect to increasing basis set size (i.e., cc-pVnZ, where n=D, T, Q) to the Complete Basis Set (CBS), or Kohn-Sham limit, for numerous properties such as thermochemical properties. Though DFT structures are generally known to reach convergence using a triple-zeta basis set, for molecules that include sulfur or transition metal species or molecules of increasing size, energetic properties determined using DFT may not reach convergence unless a basis set of at least quadruple-zeta quality basis sets are used. Thus, basis sets through quadruple-zeta quality were considered in this work.



Figure 6.1 Structure of the 53 compounds investigated in the SAMPL5 challenge. The 13 molecules represent (left) correspond to the Batch 0 subset.

Many prior studies seeking to predict the solvation free energies of small organic molecules often chose hybrid DFT methods, however, an optimal functional or optimal functional class has not yet been agreed upon.^{8,30–40} In order to identify which functional would serve best for the blind prediction, single point calculations were carried out on the optimized geometries using several different hybrid DFT functionals, (B97-1⁴¹, B98⁴², B3PW91^{12,43}, M05⁴⁴, M05-2X⁴⁵, M06⁴⁶, M06-

2X, M06-HF⁴⁷, ω B97⁴⁸, and ω B97X-D⁴⁹) in combination with the cc-pVnZ basis sets (where n=D, T, and Q) for the subset of 13 molecules (noted Batch 0 in **Error! Reference source not found.**). These functionals can be classified as three different types of hybrid functionals: global hybrid generalized gradient approximation (GH-GGA) includes B97-1, B98, and B3PW91; global hybrid meta-GGA (GH-mGGA) includes M05, M05-2X, M06, M06-2X, and M06-HF; and range separated hybrid (RSH)-GGA includes ω B97 and ω B97X-D. These functionals were chosen due to their popularity and utility for organic species, as well as to provide a variety of classes of hybrid functionals. All calculations were performed using the "Ultrafine" integration grid as it is known that HM-GGA functionals are sensitive to the integration grid size⁵⁰ and finer grids can improve numerical accuracy. These single point calculations were carried out in implicit solvent using the SMD solvation model for both water and cyclohexane. The partition coefficients were estimated from the difference in the transfer free energy in water and in cyclohexane.

6.3 Results and Discussion

For the Batch 0 subset of molecules, several hybrid DFT functionals were tested with double, triple, and quadruple- ζ level basis sets. In these calculations, only a single conformation, protonation state, and tautomer were considered thus there is not a statistical uncertainty associated with the calculations. Rather, there is model uncertainty that arises from the assumption of a single geometry, the DFT method, the basis set, and the solvent model. From the results shown in Table 6.1, overall, each approach predicts the partition coefficient within a mean absolute deviation (MAD) in the range of 1.5 - 2.0 logP units in reference to experiment, which corresponds to a 2.0 - 2.7 kcal mol⁻¹ variation in the transfer free energy. While QSPR methods are able to predict

within 0.3 - 1.0 log units from experiment, the physical significance of the predictions are questionable as these models have adroit tactics at modeling noise.⁵¹

SAMPL5	logD		B97-1		 B98			_	B3PW91			
ID	Exp.	DZ ^a	ΤZ	QZ	DZ	ΤZ	QZ		DZ	ΤZ	QZ	
003	1.9 ± 0.1	3.1	2.7	2.6	3.0	2.6	2.5		3.0	2.7	2.6	
015	-2.2 ± 0.3	-3.1	-3.6	-3.6	-3.1	-3.6	-3.7		-3.2	-3.6	-3.7	
017	2.5 ± 0.3	1.3	1.0	1.0	1.2	1.0	1.0		1.0	0.9	0.9	
020	1.6 ± 0.3	0.9	0.3	0.1	0.9	0.2	0.1		0.8	0.2	0.1	
037	-1.5 ± 0.1	-3.9	-4.2	-4.3	-4.0	-4.3	-4.4		-4.0	-4.3	-4.3	
045	-2.1 ± 0.2	-1.1	-1.5	-1.6	-1.1	-1.6	-1.7		-1.2	-1.6	-1.7	
055	-1.5 ± 0.1	-2.6	-3.1	-3.2	-2.7	-3.2	-3.3		-2.7	-3.2	-3.2	
058	0.8 ± 0.1	2.7	2.3	2.2	2.7	2.2	2.2		2.6	2.2	2.2	
059	-1.3 ± 0.3	-0.4	-0.5	-0.5	-0.4	-0.5	-0.5		-0.5	-0.6	-0.6	
061	-1.5 ± 0.1	-0.8	-1.2	-1.3	-0.8	-1.3	-1.4		-0.9	-1.3	-1.4	
068	1.4 ± 0.3	1.5	0.9	0.8	1.5	0.8	0.7		1.4	0.8	0.7	
070	1.6 ± 0.3	5.5	5.1	5.0	5.5	5.1	5.0		5.3	4.9	4.9	
080	-2.2 ± 0.2	1.1	0.5	0.4	1.0	0.4	0.3		1.1	0.5	0.4	
MSD^{b}		0.5	0.1	0.0	0.5	0.0	-0.1		0.4	0.0	-0.1	
MAD ^c		1.5	1.5	1.5	1.5	1.5	1.5		1.5	1.5	1.5	

Table 6.1 Comparison of predicted logP with experimental logD.

SAMPL5	logD		M06 M06-2X			M06-HF			F			
ID	Exp.	DZ	ΤZ	QZ		DZ	ΤZ	QZ		DZ	ΤZ	QZ
003	1.9 ± 0.1	3.0	2.7	2.7		2.8	2.3	2.3		1.8	1.4	1.3
015	-2.2 ± 0.3	-3.1	-3.4	-3.4		-3.5	-4.0	-3.9	-	4.6	-5.3	-5.1
017	2.5 ± 0.3	1.3	1.2	1.2		0.6	0.3	0.5	-	1.0	-1.4	-1.3
020	1.6 ± 0.3	0.9	0.4	0.3		0.5	-0.2	-0.2	-	0.6	-1.5	-1.6
037	-1.5 ± 0.1	-4.0	-4.2	-4.2		-4.4	-4.8	-4.7	-	5.4	-6.2	-6.0
045	-2.1 ± 0.2	-1.0	-1.4	-1.4		-1.4	-1.9	-1.9	-	-2.4	-3.1	-3.0
055	-1.5 ± 0.1	-2.7	-3.1	-3.1		-2.9	-3.4	-3.4	-	3.6	-4.2	-4.1
058	0.8 ± 0.1	2.8	2.4	2.4		2.3	1.8	1.8		1.0	0.5	0.4
059	-1.3 ± 0.3	-0.3	-0.4	-0.4		-0.6	-0.8	-0.7	-	-1.4	-1.5	-1.5
061	-1.5 ± 0.1	-0.8	-1.3	-1.4		-1.0	-1.6	-1.7	-	1.7	-2.4	-2.5
068	1.4 ± 0.3	1.7	1.1	1.1		1.2	0.5	0.5	-	0.3	-0.8	-1.0
070	1.6 ± 0.3	5.6	5.4	5.3		5.0	4.6	4.6		3.3	2.8	2.6
080	-2.2 ± 0.2	1.1	0.5	0.5		0.8	0.1	0.1		0.1	-0.8	-0.7
MSD^{b}		0.5	0.2	0.2		0.1	-0.4	-0.3	-	0.9	-1.5	-1.6
MAD ^c		1.5	1.5	1.5		1.5	1.5	1.5		1.6	1.9	1.9

Table 6.1 (cont'd)

SAMDI 5	logD			D07			ωD	07V I	<u> </u>
	logD		<u>ן</u> דר	<u>DD97</u>	07		010 7	77	07
<u> </u>	$\frac{\text{Exp.}}{10 \pm 0.1}$	1	20	12	$\frac{QL}{24}$	2	0	1 <u>Z</u>	$\frac{QL}{24}$
005	1.9 ± 0.1		2.9	2.3	2.4 4.0	2	.0 5	2.3	2.4
015	-2.2 ± 0.3		-3.3	-3.9	-4.0	-3	.5	-3.9	-3.9
017	2.5 ± 0.5		0.9	0.0	0.0	0	.9	0.7	0.0
020	1.0 ± 0.3		0.7	-0.1	-0.2	0	.0 2	-0.1	-0.2
037	-1.5 ± 0.1		-4.1	-4.5	-4.5	-4	.∠	-4.3	-4.5
045	-2.1 ± 0.2		-1.3	-1.8	-1.9	-1	.4	-1.8	-1.9
055	-1.5 ± 0.1		-2.1	-3.2	-3.3	-2	.8	-3.3	-3.4
058	0.8 ± 0.1		2.5	2.0	2.0	2	.4	2.0	2.0
059	-1.3 ± 0.3		-0.5	-0./	-0./	-0	.6	-0./	-0./
061	-1.5 ± 0.1		-1.0	-1.6	-1./	-1	.0	-1.5	-1.6
068	1.4 ± 0.3		1.3	0.5	0.4	I T	.2	0.6	0.5
070	1.6 ± 0.3		5.2	4.8	4.7	5	.2	4.8	4.7
080	-2.2 ± 0.2		1.0	0.3	0.2	0	.8	0.2	0.2
1 caph			0.0	~ ^	0.0	0	•	• •	0.0
MSD ⁰			0.3	-0.2	-0.3	0	.2	-0.2	-0.3
MAD			1.5	1.5	1.5	1	.5	1.5	1.5
SAMPL5	logD		M05		. <u> </u>	A05-22	X		
ID	Exp.	DZ	ΤZ	QZ	DZ	ΤZ	QZ	7	
003	1.9 ± 0.1	3.0	2.7	2.6	2.8	1.9	2.0)	
015	-2.2 ± 0.3	-3.1	-3.5	-3.7	-3.5	-4.6	-4.4	4	
017	2.5 ± 0.3	1.3	1.2	1.1	0.6	-0.1	0.0)	
020	1.6 ± 0.3	0.9	0.4	0.2	0.5	-0.8	-0.8	3	
037	-1.5 ± 0.1	-3.7	-4.0	-4.1	-4.4	-5.1	-5.0)	
045	-2.1 ± 0.2	-1.0	-1.4	-1.5	-1.4	-2.4	-2.4	4	
055	-1.5 ± 0.1	-2.6	-3.0	-3.1	-2.9	-3.8	-3.8	8	
058	0.8 ± 0.1	2.8	2.4	2.3	2.3	1.3	1.3	3	
059	-1.3 ± 0.3	-0.3	-0.4	-0.4	-0.6	-1.0	-1.()	
061	-1.5 ± 0.1	-0.7	-1.2	-1.3	-1.0	-1.9	-1.9	9	
068	1.4 ± 0.3	1.6	1.0	0.9	1.2	0.0	0.0)	
070	1.6 ± 0.3	5.5	5.3	5.0	5.0	4.1	4.0)	
080	-2.2 ± 0.2	1.2	0.7	0.6	0.8	-0.3	-0.3	3	
MSD ^b		0.6	0.2	0.1	0.1	-0.8	-0.8	3	

MAD^c 1.5 1.5 1.5 1.6 1.6 ^aThe columns labeled DZ, TZ, and QZ correspond to the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets.

^bThe mean signed deviation of the predicted logP from the experimental logD. ^cThe mean absolute deviation of the predicted logP from the experimental logD.



Figure 6.2 Subset of molecules (Batch0) LogP: effect of increasing basis sets. The predicted logP with respect to increasing basis set size (DZ, TZ, QZ) are represented by the red box, green triangle, and purple 'X', respectively.

For the Batch 0 subset of molecules and the functionals considered, there is not a systematic underestimation or overestimation in the prediction of the logP. The predicted logP for the GH-GGA and RSH-GGA functionals are similar, with a MAD of 1.5 logP units. Each of the tested approaches are able to predict the correct sign of logP in water and cyclohexane (hydrophobicity/hydrophilicity) for each molecule, with the exception of molecules 017, 020, 058, 068, and 080. Predicting the correct sign of logP is important as this reflects the preference of the molecule to reside in either the organic or aqueous phase.

The lowest MAD observed in this study is 1.5 logP units, which is given by eight out of the ten functionals tested. As shown in Table 6.1, most of the functionals result in a low mean signed deviation (MSD) compared with the magnitude of the MAD. This indicates that there is not a consistent deviation from the experimental values. Rather, the small magnitude of the MSD is the result of deviations above and below experiment such that they largely cancel for the Batch 0 subset of molecules. The largest MSDs observed are for the M06-HF functional with the magnitude of the MSDs within 50% or more of the MADs. Thus, the M06-HF functional has a larger systematic error resulting in underestimation of the logP value, i.e. negative MSDs in comparison to the other functionals. The calculated values for logP are plotted with respect to the molecule number in Figure 6.2 while the numerical results are shown in Table 6.1. As shown, the quality of the basis set used does not impact the predictions of hydrophobicity or hydrophilicity since the sign of the deviations is mostly unchanged when changing basis. For M06-HF, although it appears that the calculated logP is consistently underestimated with the TZ and QZ basis sets, the functional tends to predict molecules to be more hydrophilic than those predicted with the other functionals. This bias towards hydrophilicity would become a problem for high-throughput drug screening as ideal compounds are neither too hydrophobic nor too hydrophilic. Predictions that are too hydrophilic could result in discarding potential compounds.

The correlation consistent basis sets used for this work were constructed in a systematic fashion to recover correlation energy for *ab initio* methods and overall convergence is commonly demonstrated for properties determined with increasing size of basis sets. As illustrated in Figure 6.2, the increasing quality of the basis sets lowers the logP values. This behavior is clearly convergent, yet convergence to the Kohn-Sham limit does not necessarily result in improved results with respect to experiment.

In examining the Batch 0 subset, trends in the predictions of logP for each molecule are consistent for each functional with only a few exceptions. Molecules 020, 068, and 070 have deviations that differ in sign as a function of basis set and functional. This is simply the result of the predictions being so close to experiment that even a small variation in the predicted logP values can change the sign of the deviation. Increasing the basis set size typically improves the predictions of logP for the HG-GGA functionals and the RSH-GGA functionals, while the magnitude of the signed deviation increases slightly for the GH-mGGA functionals. Overall, the small variation between the logP values obtained with the TZ and QZ basis sets indicates that the TZ basis set is already near the Kohn-Sham limit. A greater percentage of exact exchange from the functionals may help in the prediction of logP, for example, from M05 to M05-2X, a better prediction of logP is obtained for molecules 003, 045, 058, 059, 070, and 080. However, for some molecules too much exact exchange overestimates the solvation in water and underestimates the logP by at least 1.0 to 2.0 logP units. For example, each functional underestimates the logP for molecules 015, 017, 020, 037, and 055, as each overestimates the solvation in water relative to cyclohexane. Additionally for these molecules, increasing the size of the basis set predicts the

molecules to be more hydrophilic. In these cases, the functionals perform best using a DZ basis set. In the case for molecule 068, each functional predicts within experimental error using a DZ basis set, with the exclusion of M06-HF.

Overall, similar predictions of logP are obtained with the GH-GGA functionals and the RSH-GGA functionals. For the GH-mGGA functionals, predictions of logP obtained using M05 and M06 are similar. Using M05-2X and M06-2X yield similar predictions of logP as well. M06-HF stands out the most in contrast to the other functionals as it consistently overestimates the solvation of the molecule in water. From the results, it is evident that increasing the amount of exact exchange, as well the quality of the basis set, overestimates the solvation energy in water relative to the solvation energy in cyclohexane. It is believed that this bias in overestimation in water is due to the parameter fitting for the SMD solvation model. The SMD solvation model, which was reported to achieve an accuracy for predicting the solvation free energies (mean unsigned error of 0.6 to 1.0 kcal/mol for neutral solutes), was more heavily parameterized against molecules solvated in water in contrast to cyclohexane.⁸ This situation can be advantageous for molecules that are slightly hydrophobic. However, this is not optimal since there is a consistent overestimation of the energy from being solvated in water as this results in misleading predictions of the equilibrium distribution of a molecule in different solvents. Although there are cases in which the GH-mGGA functionals perform best, the GGA functionals were constructed with less parameter fitting and may serve as a stronger class of functionals for initial starting guesses when predicting the solvation properties of molecules in which experiment is unavailable.

B3PW91 was chosen as the method for predicting the transfer free energy of the remaining SAMPL5 subsets because of its overall consistent behavior across the period table, and for a number of energetic properties. Some of the molecules within the SAMPL5 set contain sulfur.

Previous studies have shown that for molecules containing sulfur, the BP3W91 functional yields more accurate energetics than the B3LYP functional when using correlation consistent basis sets.^{21,52} The results submitted for the SAMPL5 challenge are shown in Table 6.2. Using B3PW91 and a quadruple- ζ basis level basis set has an MAD of 1.9 and was able to estimate the logP within 2.0 logP units from the experimental logD for about 60% of the molecule set. Although this approach lacks contributions from other protonation states, tautomers, and additional protomers, it provided a good starting estimate of logP. Regarding the outliers that this approach predicted over 2.0 logP units in reference to the experimental logD, it is believed that the prediction can be improved by including the chemical contributions from tautomerization, protonation, as well as with additional conformational sampling as many of the outliers that are greater than 3.0 logP units from the experimental logD are less rigid than other molecules within the dataset. For several of the molecules, the source of the larger deviations are known whereas other relationships between the prediction and the structure will require further investigations. The largest outlier is molecule 083, with a deviation of 9.0 logP units, is a result of modeling of the incorrect tautomer. The tautomeric state issued in the SAMPL5 data set was not the preferred tautomer. The results obtained from using the triple- ζ and the quadruple- ζ level basis sets are very similar. Rather than using a quadruple- ζ level basis set, it is recommended to use a triple- ζ level size and the correlation consistent basis sets that include tight d functions for molecules containing sulfur.²⁹

ID	LogD ^{Exp.}	LogP ^{Calc.}	ID	LogD ^{Exp.}	LogP ^{Calc.}
002	1.4 ± 0.3	-1.0	055	-1.5 ± 0.1	-3.2
003	1.9 ± 0.1	2.6	056	-2.5 ± 0.1	-1.6
004	2.2 ± 0.3	4.1	058	0.8 ± 0.1	2.2
005	-0.9 ± 0.1	1.1	059	-1.3 ± 0.3	-0.6
006	-1.0 ± 0.1	-0.9	060	-3.9 ± 0.2	-1.7
007	1.4 ± 0.3	3.5	061	-1.5 ± 0.1	-1.4
010	-1.7 ± 0.4	-2.4	063	-3.0 ± 0.4	-6.0
011	-3.0 ± 0.1	3.2	065	0.7 ± 0.2	-1.4
013	-1.5 ± 0.4	3.1	067	-1.3 ± 0.3	0.4
015	-2.2 ± 0.3	-3.7	068	1.4 ± 0.3	0.7
017	2.5 ± 0.3	0.9	069	-1.3 ± 0.3	-0.1
019	1.2 ± 0.4	5.9	070	1.6 ± 0.3	4.9
020	1.6 ± 0.3	0.1	071	-0.1 ± 0.5	-0.9
021	1.2 ± 0.3	0.7	072	0.6 ± 0.3	3.3
024	1.0 ± 0.4	1.9	074	-1.9 ± 0.3	-6.3
026	-2.6 ± 0.1	-2.8	075	-2.8 ± 0.3	-1.4
027	-1.9 ± 0.1	2.7	080	-2.2 ± 0.2	0.4
033	1.8 ± 0.2	3.4	081	-2.2 ± 0.3	-5.3
037	-1.5 ± 0.1	-4.3	082	2.5 ± 0.4	6.4
042	-1.1 ± 0.3	0.7	083	-1.9 ± 0.4	-10.9
044	1.0 ± 0.4	0.9	084	0.0 ± 0.2	2.6
045	-2.1 ± 0.2	-1.7	085	-2.2 ± 0.4	0.7
046	0.2 ± 0.3	-1.0	086	0.7 ± 0.2	1.1
047	-0.4 ± 0.3	0.2	088	-1.9 ± 0.3	-1.0
048	0.9 ± 0.4	0.8	090	0.8 ± 0.2	1.0
049	1.3 ± 0.1	1.9	092	-0.4 ± 0.3	-1.1
050	-3.2 ± 0.6	-5.4			
	MSD ^a	0.5			
	MAD^{b}	1.9			

Table 6.2 Overview of the results submitted to the SAMPL5 challenge.

^aThe mean signed deviation of the predicted logP from the experimental logD. ^bThe mean absolute deviation of the predicted logP from the experimental logD.

6.4 Conclusion

In this study, several DFT functionals were used to predict the partition coefficients in cyclohexane and water of molecules in the SAMPL5 molecule set, a diverse set of molecules, representing a variety of protonation states and tautomerization states. Using the SMD implicit solvation model and a vertical solvation approach, the free energy of transfer was predicted,

resulting in a mean absolute deviation of 1.9 logP units from the experimental logD. The results highlight that the performance of density functionals does not consistently overestimate or underestimate the logP for some molecules in the SAMPL5 set. Functionals in the GH-GGA and the RSH-GGA class perform similarly, whereas the performance of GH-mGGA does not provide similar results as GH-GGA and RSH-GGA functionals. The results show that functionals that include a larger percentages of exact exchange tend to predict logP values that overestimate the hydrophilicity. For molecules that are similar to those studied in this work, using a GH-GGA functional, such as B3PW91, in conjunction with cc-pVTZ can be used for predicting transfer free energy of small organic molecules.

Moving forward, alternative strategies should be considered to try to improve the predictions. The acid dissociation constants for many of the ionizable compounds are not known. It is possible that these theoretical predictions could be improved by accounting for the multiple ionization states as this would more accurately estimate the distribution.⁵³ Although assumptions were made regarding the protonation, tautomeric, and conformational states, the results highlight the ability of hybrid DFT approaches and the SMD implicit solvent model for estimating logP. As these approaches are able to predict close to experimental logD, the predicted logP underestimates the energetic contributions related to the structural and environmental heterogeneity in solution that is reflective of experiment. To attempt to further reduce the logP deviations from the experimental logD, the performance of *ab initio* electronic correlation methods could be considered for predicting the logP. While the use of these electronic structure methods would increase the computational cost, these predictions may provide a stronger approach that allows for systematic improvement.

REFERENCES

REFERENCES

- Nieto-Draghi, C.; Fayet, G.; Creton, B.; Rozanska, X.; Rotureau, P.; de Hemptinne, J.-C.; Ungerer, P.; Rousseau, B.; Adamo, C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. *Chem. Rev.* 2015, *115* (24), 13093–13164.
- (2) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* 2012, *112* (5), 2889– 2919.
- (3) Martin, Y. C. Let's Not Forget Tautomers. J. Comput. Aided. Mol. Des. 2009, 23 (10), 693–704.
- (4) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR Modeling: Where Have You Been? Where Are You Going To? J. Med. Chem. 2014, 57 (12), 4977–5010.
- (5) Gramatica, P.; Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. J. Chem. Inf. Model. 2016.
- (6) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995–2001.
- (7) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. J. Chem. Soc., Perkin Trans. 2 1993, No. 5, 799–805.
- (8) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* 2009, *113* (18), 6378– 6396.
- Riojas, A. G.; Wilson, A. K. Solv-ccCA: Implicit Solvation and the Correlation Consistent Composite Approach for the Determination of pKa. J. Chem. Theory Comput. 2014, 10 (4), 1500–1510.
- (10) König, G.; Pickard IV, F. C.; Mei, Y.; Brooks, B. R. Predicting Hydration Free Energies with a Hybrid QM/MM Approach: An Evaluation of Implicit and Explicit Solvation Models in SAMPL4. J. Comput. Aided. Mol. Des. 2014, 28 (3), 245–257.
- (11) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. J. Chem. Theory Comput. 2014, 10 (4), 1406–1419.
- (12) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct

Asymptotic Behavior. Phys. Rev. A 1988, 38 (6), 3098–3100.

- (13) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B.* **1998**, *37*, 785–789.
- (14) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. J. Phys. Chem. 1994, 98 (45), 11623–11627.
- (15) Jr, T. H. D. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. J. Chem. Phys. **1989**, *90*, 1007–1023.
- Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. J. Chem. Phys. 1994, 100 (4), 2975.
- (17) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. J. Chem. Phys. 1993, 98 (2), 1358.
- (18) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. V. Core-Valence Basis Sets for Boron through Neon. J. Chem. Phys. 1995, 103 (11), 4572.
- (19) Dunning, T. H. J.; Peterson, K. A.; Wilson, A. K.; Jr, T. H. D.; Jr., T. H. D. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum through Argon Revisted. J. Chem. Phys. 2001, 114 (21), 9244–9253.
- (20) Wang, N.; Wilson, A. Effects of Basis Set Choice upon the Atomization Energy of the Second-Row Compounds SO2, CCl, and ClO2 for B3LYP and B3PW91. J. Phys. Chem. A 2003, 107, 6720–6724.
- Wang, N. X.; Wilson, A. K. Density Functional Theory and the Correlation Consistent Basis Sets: The Tight D Effect on HSO and HOS. J. Phys. Chem. A 2005, 109 (32), 7187– 7196.
- (22) Prascher, B. P.; Wilson, A. K. The Behaviour of Density Functionals with Respect to Basis Set. V. Recontraction of Correlation Consistent Basis Sets. *Mol. Phys.* 2007, 105 (19–22), 2899–2917.
- (23) Wang, N. X.; Wilson, A. K. The Behavior of Density Functionals with Respect to Basis Set. I. The Correlation Consistent Basis Sets. J. Chem. Phys. **2004**, 121 (16), 7632–7646.
- (24) Laury, M. L.; Boesch, S. E.; Haken, I.; Sinha, P.; Wheeler, R. A.; Wilson, A. K. Harmonic Vibrational Frequencies: Scale Factors for Pure, Hybrid, Hybrid Meta, and Double-Hybrid Functionals in Conjunction with Correlation Consistent Basis Sets. *J. Comput. Chem.* 2011, *32* (11), 2339–2347.

- (25) Prascher, B. P.; Wilson, B. R.; Wilson, A. K. Behavior of Density Functionals with Respect to Basis Set. VI. Truncation of the Correlation Consistent Basis Sets. J. Chem. Phys. 2007, 127 (12), 124110.
- (26) Wang, N. X.; Venkatesh, K.; Wilson, A. K. Behavior of Density Functionals with Respect to Basis Set. 3. Basis Set Superposition Error[†]. J. Phys. Chem. A **2006**, 110 (2), 779–784.
- Wang, N. X.; Wilson, A. K. Behaviour of Density Functionals with Respect to Basis Set: II. Polarization Consistent Basis Sets. *Mol. Phys.* 2005, *103* (2–3), 345–358.
- (28) Jiang, W.; Laury, M. L.; Powell, M.; Wilson, A. K. Comparative Study of Single and Double Hybrid Density Functionals for the Prediction of 3d Transition Metal Thermochemistry. *J. Chem. Theory Comput.* **2012**, *8* (11), 4102–4111.
- (29) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum through Argon Revisited. J. Chem. Phys. 2001, 114 (21), 9244.
- (30) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* 2009, 5 (4), 1016–1026.
- (31) Rayne, S.; Rayne, S.; Forest, K. Accuracy of Computational Solvation Free Energies for Neutral and Ionic Compounds: Dependence on Level of Theory and Solvent Model. *Nat. Preced.* **2010**, 1–22.
- (32) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute–Water Clusters. *J. Chem. Theory Comput.* **2005**, *1* (6), 1133–1152.
- (33) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous Solvation Free Energies of Ions and Ion–Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *J. Phys. Chem. B* **2006**, *110* (32), 16066–16081.
- (34) Takano, Y.; Houk, K. N. Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules. J. Chem. Theory Comput. 2005, 1 (1), 70–77.
- (35) Guthrie, J. P.; Povar, I. A Test of Various Computational Solvation Models on a Set of "difficult" Organic Compounds. *Can. J. Chem.* **2009**, *87* (8), 1154–1162.
- (36) Tekarli, S. M.; Drummond, M. L.; Williams, T. G.; Cundari, T. R.; Wilson, A. K. Performance of Density Functional Theory for 3d Transition Metal-Containing Complexes: Utilization of the Correlation Consistent Basis Sets. *J. Phys. Chem. A* 2009, *113* (30), 8607–8614.
- (37) Riley, K. E.; Op't Holt, B. T.; Merz, K. M. Critical Assessment of the Performance of

Density Functional Methods for Several Atomic and Molecular Properties. J. Chem. Theory Comput. **2007**, *3* (2), 407–433.

- (38) Martell, J. M.; Goddard, J. D.; Eriksson, L. A. Assessment of Basis Set and Functional Dependencies in Density Functional Theory: Studies of Atomization and Reaction Energies. J. Phys. Chem. A 1997, 101 (10), 1927–1934.
- (39) Cohen, A. J.; Mori-Sanchez, P.; Yang, W.; Mori-Sánchez, P. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112* (1), 289–320.
- (40) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals. J. Phys. Chem. A 2007, 111 (42), 10439–10452.
- (41) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and Assessment of New Exchange-Correlation Functionals. *J. Chem. Phys.* **1998**, *109* (15), 6264–6271.
- (42) Schmider, H. L.; Becke, A. D. Optimized Density Functionals from the Extended G2 Test Set. J. Chem. Phys. **1998**, 108 (23), 9624–9631.
- (43) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. J. Chem. Phys. 1993, 98, 5648–5652.
- (44) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Exchange-Correlation Functional with Broad Accuracy for Metallic and Nonmetallic Compounds, Kinetics, and Noncovalent Interactions. J. Chem. Phys. 2005, 123 (16), 161103.
- (45) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. J. Chem. Theory Comput. 2006, 2 (2), 364–382.
- (46) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241.
- (47) Zhao, Y.; Truhlar, D. G. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. J. Phys. Chem. A 2006, 110 (49), 13126–13130.
- (48) Chai, J. Da; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. J. Chem. Phys. **2008**, 128 (8), 1–15.
- (49) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* 2008, 10 (44), 6615.

- (50) Wheeler, S. E.; Houk, K. N. Integration Grid Errors for Meta-Gga-Predicted Reaction Energies: Origin of Grid Errors for the M06 Suite of Functionals. J. Chem. Theory Comput. 2010, 6 (2), 395–404.
- (51) Palmer, D. S.; Mitchell, J. B. O. Is Experimental Data Quality the Limiting Factor in Predicting the Aqueous Solubility of Druglike Molecules? *Mol. Pharm.* **2014**, *11* (8), 2962–2972.
- (52) Denis, P. A.; Ventura, O. N. Density Functional Investigation of Atmospheric Sulfur Chemistry. I. Enthalpy of Formation of HSO and Related Molecules. *Int. J. Quantum Chem.* **2000**, *80* (3), 439–453.
- (53) Kah, M.; Brown, C. D. LogD: Lipophilicity for Ionisable Compounds. *Chemosphere* **2008**, 72 (10), 1401–1408.

CHAPTER 7 CONCLUDING REMARKS

Proteins are dynamic in nature and can have many conformational-dependent functional states, including active, inactive, and intermediate states. These conformational states can represent an array of structural features that distinguish the ability of the protein to bind other molecules. Elucidating the transitions between the conformational states of protein complexes is critical for effective drug discovery, design, and delivery methods, as it would allow direct correlation between conformational-activity and structure-activity relationships. In this dissertation, theoretical investigations were employed to examine structure-function relationships of proteins and peptides. More specifically, investigations were carried out on oncogenic protein-kinases involved in the signaling of the NF-κB pathway.

NF-κB transcription factors are sequestered in the inactive state by IκB proteins and are released upon activation of the canonical or the non-canonical pathway. The canonical pathway is dependent upon the activation of IKK β and the non-canonical pathway is dependent upon the activation of NIK. In Chapter 3, molecular dynamics simulations of IKK β were used to investigate how conformational changes and protein-protein interactions within multimeric assemblies are influenced by changes in the interacting subunits, as dimerization of IKK β is critical for its activation, although it is not required for its reactivity.¹ An approach was designed to characterize the changes in the inactive and active states by contrasting and quantifying the changes in proteinprotein association and dissociation upon post-translational modification.

Thermodynamic properties were monitored for simulation stability. Hydrogen-bonding interactions, salt-bridges, and conformational changes were compared among the monomer, the two different dimers and the tetrameric arrangement of IKK β and suggest that the tetrameric structure mediates a global stability for the enzyme. Additionally, the results highlight that

modeling a single monomer of the protein kinase is insufficient for capturing the associated structure-activity relationships as the oligomeric assembly is significant for function.

In Chapter 4, molecular dynamics simulations and modeling techniques were employed to differentiate the structural dynamics of the NF-kB Inducing Kinase (NIK) in its native and mutant form, and in the absence and presence of salt concentration in efforts to probe whether changes in the ionic environment stabilize or destabilize the NIK dimer. Analyses of structure-activity and conformational-activity relationships indicate that the protein-protein interactions are sensitive to changes in the ionic strength. Ligand binding pockets either compress or expand, affecting both local and distal intermolecular interactions, yielding further insight into how changes within the intracellular microenvironment affect molecular mechanisms and how small molecules may bind.

Ligand-binding sites in rational drug design are often proposed from static conformational states from X-ray crystallography that neglect intrinsic dynamics induced by natural molecular and physiological environments. As the environment of a molecular system directly impacts the structure, it is reasonable that this impacts the activity. Although changes in the ionic concentration are known to impact protein-ligand binding, the specific role and mechanism that assist and hinder ligand binding remain elusive for many protein kinases. Additionally, cancer cells are unique because they are highly expressed with features that aid the cell in maintaining its oncogenic homeostatic conditions, including the expression of heat shock proteins, ion transporters, and other complexes that regulate the intracellular environment (eg. pH and ionic strength). Using the modeling schemes proposed in Chapters 3 and 4, future investigations should include investigating related Serine/Threonine and Tyrosine protein kinases to better understand structure-activity and conformational-activity relationships mediated by changes in the intracellular environment for

these protein kinases and how these changes may induce alternative mechanisms leading to oncogenic activity.

Another aspect of this work includes the assessment of different theoretical methods for the prediction of physical properties of peptides and drug molecules. In Chapter 5, ab initio and density functional methods were used to characterize the selectivity of radical-mediated fragmentation by evaluating site, conformation, and pathway trends of small trialanine peptides resembling a β -strand and a β -turn.² Comparisons of reaction enthalpies show that the diamide pathway is more energetically favorable than the α -amidation pathway and that both pathways are site and conformationally selective. An evaluation of electronic structure methods illustrates that more approximate methods can provide insight on which fragmentation pathway is preferred but are unable to discern site or conformational selectivity and that sophisticated methods that account for electron correlation are needed for reactions of radicals to proteins. All results and conclusions in Chapter 5 were based on gas-phase computations, whereas the target is at understanding larger biomolecules. While these findings offer a fundamental insight into how proteins can respond to reactive oxygen species, producing fragments with predictable terminal functional groups, this study neglects solvent-induced conformational rearrangement and the impact of varying side chains on the reaction pathways. These are of interest for further investigations because they will allow for an enhanced understanding of protein fragmentation as well as the ability to predict the structural features involved in such mechanisms, which is significant towards understanding oxidative stress. As indicated in the results, additional methods should be considered in future work including both quantum mechanics and molecular mechanics methods to account for conformational rearrangements.

In Chapter 6, several density functional methods were employed to estimate the partition coefficients of drug-like molecules from the transfer free energies in cyclohexane and water.³ Three classes of hybrid density functionals were assessed and the results highlight that functionals with less parameterization outperformed those deemed as more sophisticated functions. The results also highlight some of the inadequacies of the modeling schemes. Moving forward with first principle predictions of physiochemical properties, partition coefficients are a measurement of only the neutral forms of the solute, while distribution coefficients include all charged and neutral tautomers of the solute distributing between two immiscible solvents. Future investigations should include efforts at predicting the distribution coefficients by considering reactions of the of the solute ion solution and by further assessing the performance of other *ab initio* and density functional methods as well as solvation models.
REFERENCES

REFERENCES

- (1) Jones, M. R.; Liu, C.; Wilson, A. K. Molecular Dynamics Studies of the Protein–Protein Interactions in Inhibitor of κB Kinase-β. J. Chem. Inf. Model. **2014**, *54* (2), 562–572.
- (2) Stringfellow, H. M.; Jones, M. R.; Green, M. C.; Wilson, A. K.; Francisco, J. S. Selectivity in ROS-Induced Peptide Backbone Bond Cleavage. J. Phys. Chem. A 2014, 118 (48), 11399–11404.
- (3) Jones, M. R.; Brooks, B. R.; Wilson, A. K. Partition Coefficients for the SAMPL5 Challenge Using Transfer Free Energies. *J. Comput. Aided. Mol. Des.* **2016**, *30* (11), 1129–1138.