Computational Modeling of the Structure, Function and Dynamics of Biomolecular Systems

YASHRAJ KULKARNI
Proteins are a structurally diverse and functionally versatile class of biomolecules. They perform a variety of life-sustaining biological processes with utmost efficiency. A profound understanding of protein function requires knowledge of its structure. Experimentally determined protein structures can serve as a starting point for computer simulations in order to study their dynamic behavior at a molecular level. In this thesis, computational methods have been used to understand structure-function relationships in two classes of proteins - intrinsically disordered proteins (IDP) and enzymes.

Misfolding and subsequent aggregation of the amyloid beta (Aβ) peptide, an IDP, is associated with the progression of Alzheimer’s disease. Besides enriching our understanding of structural dynamics, computational studies on a medically relevant IDP such as Aβ can potentially guide therapeutic development. In the present work, binding interactions of the monomeric form of this peptide with biologically relevant molecular species such as divalent metal ions (Zn$^{2+}$, Cu$^{2+}$, Mn$^{2+}$) and amphiphilic surfactants were characterized using long timescale molecular dynamics (MD) simulations. Among the metal ions, while Zn$^{2+}$ and Cu$^{2+}$ maintained coordination to a well-defined binding site in Aβ, Mn$^{2+}$-binding was observed to be comparatively weak and transient. Surfactants with charged headgroups displayed strong binding interaction with Aβ. Complemented by biophysical experiments, these studies provided a multifaceted perspective of Aβ interactions with the partner molecules.

Triosephosphate isomerase (TIM), a highly evolved and catalytically proficient enzyme, was studied using empirical valence bond (EVB) calculations to obtain deeper insights into the catalytic reaction mechanism. Multiple structural features of TIM such as the flexible loop and preorganized active site residues were investigated for their role in enzyme catalysis. The effect of substrate binding was also studied using truncated substrates. Finally, using enhanced sampling methods, dynamic behavior of the catalytically important loop 6 was characterized. The importance of structural stability and flexibility on protein function was illustrated by the work presented in this thesis, thus furthering our scientific understanding of proteins at a molecular level.

Keywords: Molecular Dynamics, Empirical Valence Bond, Enzyme Catalysis, Amyloid Beta, Aβ, Triosephosphate Isomerase, TIM, Computational Biochemistry, Computational Enzymology

Yashraj Kulkarni, Department of Chemistry - BMC, Biochemistry, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

© Yashraj Kulkarni 2020

ISSN 1651-6214
urn:nbn:se:uu:diva-398169 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-398169)
Dedicated to my dear Ajja.
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
Additional Publications


Contribution report

Contributions to the articles that are part of this thesis are listed here.

Paper I: Worked on the computational aspect of the study. Performed part of the simulations and analysis.

Paper II: Worked on the computational aspect of the study. Performed part of the simulations and analysis.

Paper III: Performed force field parameterization of the system, and part of the simulations and analysis.

Paper IV: Performed all of the simulations and analysis.

Paper V: Designed the simulations. Performed part of the simulations and analysis.

Paper VI: Performed the EVB simulations and relevant analysis.
6. Conclusions and Future Perspectives .................................................. 57
7. Sammanfattning på svenska .............................................................. 59
Acknowledgments ......................................................................................... 61
References ..................................................................................................... 65
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aβ</td>
<td>Amyloid-β Peptide</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>BE-METAD</td>
<td>Biased Exchange Metadynamics</td>
</tr>
<tr>
<td>DDM</td>
<td>Dodecyl β-D-maltoside</td>
</tr>
<tr>
<td>DHAP</td>
<td>Dihydroxyacetone Phosphate</td>
</tr>
<tr>
<td>EM</td>
<td>Electron Microscopy</td>
</tr>
<tr>
<td>EVB</td>
<td>Empirical Valence Bond</td>
</tr>
<tr>
<td>FEP</td>
<td>Free Energy Perturbation</td>
</tr>
<tr>
<td>GA</td>
<td>Glycolaldehyde</td>
</tr>
<tr>
<td>GAP</td>
<td>D-glyceraldehyde 3-phosphate</td>
</tr>
<tr>
<td>HREX</td>
<td>Hamiltonian Replica Exchange</td>
</tr>
<tr>
<td>IDP</td>
<td>Intrinsically Disordered Protein</td>
</tr>
<tr>
<td>LFER</td>
<td>Linear Free Energy Relationships</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular Mechanics</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum Mechanics</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>TIM</td>
<td>Triosephosphate Isomerase</td>
</tr>
<tr>
<td>TST</td>
<td>Transition State Theory</td>
</tr>
<tr>
<td>US</td>
<td>Umbrella Sampling</td>
</tr>
</tbody>
</table>
1. Introduction

Biomolecules form the chemical basis of all life processes. Evolution has gradually driven the organization and specialization of these molecules, rendering them self-sustainable in the process. Biomolecules are mainly classified as carbohydrates, lipids, nucleic acids, and proteins, and can exist in simple (e.g., monomeric), or complex (e.g., oligomeric, polymeric) forms. They are also observed to display a trend in which simple building blocks compound to form large complex structures that perform distinct functional roles.¹

Proteins are a structurally diverse and functionally versatile class of biomolecules. As enzymes, they catalyze biochemical reactions; as structural proteins, they render structural integrity to cell membranes by providing support to lipid molecules; as motility proteins, they facilitate cellular or organismal movement; and as membrane proteins, they function as receptors or signal transducers, and also as ion-transport channels.¹ The list is not exhaustive, as proteins are involved in many more vital life processes. The diversity in protein function is achieved mainly by virtue of the broad range of three-dimensional structures formed by polypeptides. Each of the twenty naturally occurring amino acids, which function as monomeric units of polypeptides, have distinct physical and chemical properties that allow proteins to fold into specific structures.² Research efforts addressing the question of protein folding have employed experimental and computational methods³⁻⁵ to study the features of a protein sequence that direct it to fold into its preferred or “correct” structure.

Understanding the significance of structural features of a protein for its efficient functioning is another question that, in this context, is important as well as relevant. Studying protein structures at atomic resolution is essential for understanding structure-function relationships of proteins. X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM) methods are some of the most popular techniques used today to generate three-dimensional structural models of biomolecules, ranging from short peptides to viral particles. These models have provided valuable insights to our understanding of biomolecular structure and function.
X-ray crystallography is a widely-used technique used to determine high-resolution structures of proteins. This method involves the exposure of protein crystals to incident X-ray beams resulting in diffraction patterns. The data is then used to calculate the electron density, and subsequently determine atomic coordinates. Structures of enzymes with or without bound ligands can be obtained using this method. NMR spectroscopy is used to obtain an ensemble of structures from a solution of isotope-labeled peptides. This technique is generally used for determining structures of short polypeptides or fragments of large protein complexes. Cryo-EM on the other hand, which uses protein samples prepared at cryogenic temperatures, is generally used for structure determination of larger protein complexes such as multimeric structures.

Structural models provide a static snapshot of a protein at an atomic resolution. Computer simulations, which model the dynamics of a protein structure using physics-based theories, have been of great help in providing deeper insights into protein function using structural models. These methods, when used in conjunction with experimental methods, provide a holistic view of the system under study, thereby allowing the determination of the role of structural rigidity and flexibility in protein function.

The work presented in this thesis deals with: (A) observing the structural variations in proteins caused by changes in their external environment or inherent characteristics, and (B) studying the effect of these structural and dynamic changes on the desired function of the proteins. Chapter 2 contains a description on the diversity in protein structure and function in two types of proteins – intrinsically disordered proteins (IDPs), and enzymes. While IDPs are known to change their structural forms at different physicochemical conditions, enzymes have a distinct three-dimensional structure that might undergo localized conformational changes as part of its catalytic function. The common factor between both these systems is that structural perturbations in a protein can affect its function. The amyloid beta (Aβ) peptide, an IDP, and triosephosphate isomerase (TIM), a highly catalytically proficient enzyme, have been chosen as model systems for the investigations.

Chapter 3 contains a detailed description and discussion about the various computational methods used in my work. Chapters 4 and 5 provide a detailed summary of my research on the Aβ peptide and TIM respectively, where I have used long timescale molecular dynamics (MD) simulations to study conformational dynamics of the Aβ peptide, and empirical valence bond (EVB)
calculations to model enzyme catalysis in TIM. Chapter 6 summarizes key takeaways from the research showcased in this thesis. It also contains my reflections on the scientific learnings from the studies, and comments on the future prospects of these areas of research.
2. Structure-Function Relationships in Proteins

Proteins are an indispensable chemical component of every biological process. As mentioned before, the structural and functional versatility of proteins is borne out of the chemical diversity of their constituent amino acids. Hence, knowing the structure of a protein is key to determining its function. This paradigm had its foundations laid when the first three-dimensional structure of myoglobin was solved by Kendrew et al. in 1958. Ever since, there has been a tremendous growth in the number of protein structures generated and deposited in the Protein Data Bank (PDB). This wealth of structural data has been instrumental towards enhancing our understanding of functional aspects of proteins such as enzymes or cell surface receptors. Structure-based drug discovery has enabled high-throughput screening of drug candidates, and has resulted in a number of effective drugs targeting fatal diseases such as HIV, tuberculosis, and cancer. Significant advances have been made in the field of enzymology too, where clarity on the mechanistic aspects of enzyme-catalyzed reactions could be achieved as a result of structural analysis of the catalytic active site.

On the other end of the spectrum, recent evidences resulting largely from studies of disordered proteins or disordered regions of folded proteins have challenged this paradigm, attributing functional capability and diversity of proteins to disorderliness in their structures. Peptides involved in signaling pathways are intrinsically disordered, allowing them to function in multiple pathways. Flexible loops in highly-ordered globular proteins such as enzymes are known to play a major role in their catalytic function.

We notice that both order and disorder in a protein structure are crucial for its desired function. This acts as a motivation to perform detailed studies on not just the static structure of a protein, but also on its dynamic behavior. Compu-
tational approaches to simulate structural dynamics and study its effect on protein function thus become instrumental in furthering our understanding and lending clarity to the structure-function paradigm.

In my effort to use computational approaches to better our understanding of protein structural dynamics and its effect on the protein function, I have chosen two contrasting classes of proteins – an intrinsically disordered protein (IDP), and an enzyme – as systems of investigation. In the first part of this chapter, I will discuss about the various salient features associated with IDPs and their implications on living organisms, and give a brief account of their structural properties. The second part of this chapter will feature a detailed overview of enzyme catalysis and its key concepts that are relevant to the research work presented in the thesis.

2.1 Intrinsically Disordered Proteins

Proteins that do not exist in a single, well-defined equilibrium structure but instead as dynamic conformational ensembles are regarded as intrinsically disordered proteins (IDPs). These proteins are said to be disordered as they do not have a definite tertiary structure and attain multiple secondary structural forms, resulting in a dearth of structural order. This disorderliness in an IDP though, is defined by the sequence and composition of its constituent amino acids, and hence the word intrinsic is used to refer to this innate characteristic of the protein.

IDPs can extend their primary amino acid sequence into a random coil, but can also collapse to form a molten globule. The protein sequences of IDPs and structured/ordered proteins have marked differences in their amino acid composition, aromaticity, hydrophobicity, and charge, to name a few. Noticeably, IDPs lack bulky hydrophobic and aromatic residues that are crucial in forming and stabilizing the hydrophobic cores of most folded globular proteins. On the other hand, they are rich in polar amino acids and also structure-breaking amino acids such as glycine and proline. A combination of these two factors grants these proteins their conformational instability, thereby making them functionally versatile.
2.1.1 Biological Function of IDPs
Owing to their structural plasticity, IDPs and intrinsically disordered regions of folded proteins possess the ability to carry out multiple biological functions. More specifically, they are involved in various signaling and regulatory pathways, via interactions with other proteins, nucleic acids, and ligands. Some notable examples of their functions include regulation of transcription and translation, protein phosphorylation, ribosome assembly, and chaperone-assisted protein folding.\textsuperscript{28,32}

IDPs also undergo folding upon binding to a suitable partner, in a process called coupled folding and binding. The entropic cost to fold a disordered protein is paid by the binding enthalpy generated in the process. This phenomenon is utilized by a number of protein–nucleic-acid and protein-protein interactions that are part of transcription and translation. Complexes resulting from such interactions tend to have high specificity and low affinity, enabling targeted association during process initiation and easier dissociation during process termination.\textsuperscript{28,32}

2.1.2 Involvement of IDPs in Diseases
Proper protein function is key to maintenance of life, and protein dysfunction can result in deleterious health conditions. While structural variability confers functional versatility in IDPs, it also holds the potential to trigger protein misfolding, which can further snowball into protein aggregation resulting in the formation of fibrils. Misfolding can be caused by both intrinsic (e.g. point mutations) and extrinsic factors (e.g. interactions with other proteins or small molecules), acting independently or in complex association with one another.\textsuperscript{32}

Protein misfolding and aggregation is associated with the pathogenesis of some of the most prominent neurodegenerative diseases such as Alzheimer’s disease (AD), Parkinson’s disease (PD), Huntington’s disease, and prion diseases.\textsuperscript{38} The mechanism usually involves proteins converting from their soluble, functional states into highly ordered, filamentous protein aggregates. These aggregates, known as amyloid fibrils, tend to accumulate in various organs and tissues.\textsuperscript{32} The defining structural feature of a fibril is a core cross-β-sheet structure formed by β-strands of individual protein units running perpendicular to the fibril’s axis (Figure 1).\textsuperscript{32,39}
The increase in propensity for an IDP to undergo aggregation is attributed to stabilization of a partially folded conformation, which triggers oligomerization and fibrillation by enabling a monomer to make electrostatic, hydrogen bond, and hydrophobic interactions with other monomers. The transformation of a disordered structure into a partially folded conformation, caused by mutations in the protein sequence or changes in the environment, thus acts as a prerequisite for fibrillation. Therefore, in order to elucidate these structural transformations, it is essential to perform studies on the structure and dynamics of IDPs under varying physiological conditions.

2.1.3 Amyloid β Peptide

Alzheimer’s disease (AD) is a neurodegenerative disorder clinically characterized by dementia, cognitive and behavioral impairment, social and occupational dysfunction, and eventual death. The amyloid-β (Aβ) peptide, according to the amyloid hypothesis, is considered to be the main cause of this disease. Misfolding of the Aβ peptide triggers its aggregation, resulting in the formation of senile plaques. The formation of these plaques in the brains of patients affected by AD is considered to be associated with the progression of the disease. As the Aβ peptide is a key component of the plaques and hence an important indicator of AD, it is highly useful to study its structural properties in order to understand its role in AD pathology at a molecular level.
The Aβ peptide is a soluble disordered peptide of length ranging from 39 to 43 amino acid residues. It is produced by proteolytic cleavage of a transmembrane protein called amyloid precursor protein (APP), catalyzed by the enzymes β-secretase and γ-secretase. Faulty metabolism and an imbalance between the production and clearance of this peptide, resulting in aggregation and fibril formation, is associated with AD pathology.

Due to the disordered nature of Aβ peptide, methods such as NMR spectroscopy and MD simulations are preferred for its structural characterization. Models of functionally relevant peptide fragments, as well as the two most common variants – Aβ(1-40) and Aβ(1-42) – have been derived using NMR methods and studied for structural changes induced by variation in their physical and chemical environments. Although Aβ(1-40) is found more abundantly than Aβ(1-42), the latter is known to possess greater propensity for oligomerization and eventual aggregation. For this reason, both these variants are studied extensively for their structural dynamics and aggregation pathways.

Aβ(1-40) is structurally characterized as being composed of a hydrophilic N-terminus region, and two hydrophobic segments forming the central and C-terminal regions (Figure 2). In Aβ(1-42), the C-terminus is extended further by two hydrophobic residues, isoleucine and alanine, which reduces C-terminal flexibility by forming a β-hairpin at residues 38-41. This could explain why Aβ(1-42) has greater tendency to aggregate. However, both the variants, Aβ(1-40) and Aβ(1-42), show multiple discrete structural transitions between α-helical, β-sheet, and random coil forms. It is therefore important to study the structural dynamics of Aβ peptide as a monomer and identify the conformations that qualify as a misfolded state bearing potential to undergo oligomerization.

Figure 2. Primary sequence of the Aβ(1-40) peptide. Charged amino acid residues are individually indicated in bold and colored in red (negatively charged), deep blue (positively charged) and light blue (histidines). The background color provides a general indication of the hydrophilic (light orange) and hydrophobic (light yellow) regions of the peptide.
2.2 Enzymes

Enzymes are biological catalysts capable of significantly accelerating the rates of biologically relevant chemical reactions, by up to $10^{19}$ orders of magnitude, in some cases.\textsuperscript{57,58} In a living organism, they are ubiquitous and influence every important life process, such as glucose metabolism, replication of genetic material, and protein synthesis. The chemistry responsible for sustaining life involves biochemical transformations that cannot occur at physiologically meaningful timescales without the help of enzymes. With a large number of reactions simultaneously taking place in a cell, enzymes, in addition to their catalytic efficiency, exhibit remarkable specificity towards their respective substrates.\textsuperscript{1} This combination of the efficiency and specificity of enzymes, in addition to their biodegradable nature, makes them ideal candidates for use in industrial applications in today’s world where sustainability is of paramount importance.

Barring a few catalytic RNA molecules, the majority of naturally occurring enzymes are comprised of proteins having a distinct three-dimensional structure called a \textit{fold}. This structural fold, defined by the composition and sequence of amino acids in the protein, is very critical for the enzyme to function efficiently. The structure-function paradigm discussed earlier in this chapter stems from the observations made in enzymes that hint at a direct relationship between their structure and biochemical function. However, a similar structural fold can exhibit multiple different protein functionalities (protein ‘moonlighting’), whereas functionally related enzymes might also possess different structures.\textsuperscript{59,60}

As already stated in a broader context, both order and disorder are necessary for efficient protein function. Investigating the role of enzyme structure and dynamics on its function using computational models is useful for understanding the origins of enzyme catalysis from a different perspective. The ability to analyze enzymatic reactions at an atomic level empowers us to perform mechanistic studies and derive information about thermodynamic quantities. The following sections will discuss the basics of enzyme catalysis and describe how quantities determined by biochemical experiments can be viewed in conjunction with those obtained from theoretical simulations.
2.2.1 Enzyme Kinetics

Enzyme catalysis, in simple terms, refers to an increase in the rate of a chemical reaction facilitated by an enzyme. Chemical kinetics deals with the study of the rates of chemical reactions, and attempts to establish a relationship between the rate of a reaction and the concentration of reactants. These relationships can be described by rate equations or rate laws. They carry vital information regarding the reaction mechanisms and corresponding kinetic rate constants, often acting as a bridge between experimental and computational enzymology. The utility of these rate laws are enhanced when they are complemented by other analyses from the physical organic chemistry toolkit, such as linear free energy relationships (LFER) and kinetic isotope effects (KIE). The Michaelis-Menten relationship can be applied to analyze the kinetics of an enzyme-catalyzed reaction.

**Michaelis-Menten Relationship**

A simple enzyme-catalyzed reaction involving the conversion of one substrate molecule to a product molecule, without any intermediate states, can be written as,

\[
E + S \xrightarrow[k_1]{k_{-1}} ES \xrightarrow[k_2]{k_{-2}} E + P
\]

where E, S, ES and P represent the enzyme, substrate, the enzyme-substrate complex, and the product, respectively. \(k_1, k_2\) and \(k_{-1}\) denote the rate constants for the respective reaction steps. While the formation of the ES complex, also known as the Michaelis complex, is reversible and in equilibrium with substrate release from the enzyme, the step leading to product formation is assumed to be irreversible. Thus, the rate of product formation, and change in concentration of the ES complex is given by the following relationships:

\[
\frac{d[P]}{dt} = k_2 [ES]
\]  
(2.1)

\[
\frac{d[ES]}{dt} = k_1 [E][S] - k_{-1} [ES] - k_2 [ES]
\]  
(2.2)

According to the steady state approximation, the ES complex is in fast equilibrium with the free enzyme and substrate, thus implying [ES] to be constant. This results in the following expression:

\[
\frac{d[ES]}{dt} = 0 = k_1 [E][S] - k_{-1} [ES] - k_2 [ES]
\]  
(2.3)
Separating the concentration terms and the rate terms, we get
\[
\frac{k_{-1} + k_2}{k_1} = \frac{[E][S]}{[ES]} \tag{2.4}
\]
The quantity on the left hand side of equation (2.4) can be defined as the Michaelis constant, \(K_M\).
\[
K_M = \frac{k_{-1} + k_2}{k_1} \tag{2.5}
\]
The initial concentration of the enzyme, \([E_0]\), can be written as the sum of concentrations of the free enzyme \([E]\) and enzyme-substrate complex \([ES]\):
\[
[E_0] = [E] + [ES] \tag{2.6}
\]
Substituting the expressions of \(K_M\) and \([ES]\) as obtained from equations (2.5) and (2.6) into equation (2.4), and rearranging the terms results in the following expression for \([ES]\):
\[
[ES] = \frac{[E_0][S]}{K_M + [S]} \tag{2.7}
\]
Substituting equation (2.7) in equation (2.1), we obtain an expression for the rate of the reaction, \(v_0\) in terms of initial concentrations of the enzyme and substrate:
\[
v_0 = \frac{d[P]}{dt} = \frac{k_2[E_0][S]}{K_M + [S]} \tag{2.8}
\]
The maximum velocity of an enzymatic reaction, \(V_{max}\), is given by:
\[
V_{max} = k_2 [E_0] \tag{2.9}
\]
Replacing \(k_2\) in equation (2.8) with equation (2.9), we obtain the final form of the Michaelis-Menten equation,
\[
v_0 = \frac{V_{max}[S]}{K_M + [S]} \tag{2.10}
\]
The Michaelis-Menten relationship, represented in a schematic form in Figure 3, can be used to obtain key kinetic parameters that serve as a guide to evaluate the catalytic efficiency of an enzyme. These parameters are \(k_{cat}\), \(K_M\), and \(k_{cat}/K_M\).

\(k_{cat}\) is a first-order rate constant that indicates the rate of conversion of an enzyme-substrate complex to free enzyme and product. Although for practical purposes it is considered as the rate constant for the chemical step, \(i.e.\ k_2\), it is theoretically the lower limit of the rate constant of the rate-limiting step post ES formation. At saturating levels of the substrate, \(k_2\) becomes equivalent to \(k_{cat}\). This kinetic constant is of great importance in computational enzymology.
as it can be used to obtain the free energy of activation with the help of transition state theory (TST), thus providing a bridge between the kinetic world of experiment and the thermodynamic world of simulation.

\[ v_0 = \frac{V_{\text{max}} [S]}{K_M} \]

\[ V_{\text{max}} = k_{\text{cat}} [E_0] \]

\[ K_M \] is the concentration of the substrate at which the rate of enzyme-catalyzed reaction is half of the maximum rate. Known as the Michaelis constant, it is used as an indicator of the binding affinity of the enzyme for a particular substrate. Although here too, it is important to note that this assumption is valid only when the reaction is two-step and the rate of substrate dissociation far exceeds the rate of product formation \((k_{-1} \gg k_2)\), which is rarely the case for enzyme catalyzed reactions.\(^{61,62}\)

\[ k_{\text{cat}}/K_M \] is usually considered to be the catalytic efficiency of an enzyme towards a certain substrate. As it reflects both the binding and catalytic events, it functions as an apparent second-order rate constant, indicating how substrate binding affects the catalytic rate. In cases where the catalytic rate far exceeds the rate of substrate dissociation, the reaction rate is said to have reached the diffusion limit.

Computational enzymology deals with the calculation of thermodynamic parameters wherein a single enzyme-substrate system is studied in isolation.
This is contrary to experiments, where kinetic parameters obtained for the enzymatic reaction in solution are macroscopic in nature. In order to utilize data obtained from experiments for validating computational models, it is essential to make a connection between chemical kinetics and thermodynamics.

**Linking Kinetics with Thermodynamics**

The Arrhenius equation\(^{63}\) explains the effects of temperature dependence on the rate of enzyme-catalyzed reactions,

\[
k = A e^{-E_a/RT} \tag{2.11}
\]

where \(k\) is the observed rate constant, \(A\) is the reaction specific pre-factor, \(R\) is the universal gas constant, \(T\) is the temperature in Kelvin, and \(E_a\) is the activation energy of the reaction. It implies an increase in the rate of the reaction with an increase in temperature.

![Figure 4. Schematic representation of a free energy profile for the conversion of substrate S to product P. The profile shown in black corresponds to the uncatalyzed reaction in solution (in this case, water, indicated by “w” in the subscripts). The profile shown in green corresponds to the enzyme-catalyzed reaction.](image)

The Eyring-Polanyi equation\(^{64,65}\) is similar in nature, relating the rate of the reaction with its Gibbs free energy,

\[
k = \frac{k_B T}{h} e^{-\Delta G^\ddagger / RT} \tag{2.12}
\]

where \(k\) is the rate of the reaction at temperature \(T\), \(k_B\) is the Boltzmann constant, \(h\) is the Planck constant, and \(\Delta G^\ddagger\) is the Gibbs activation free energy. This relationship allows us to obtain thermodynamic parameters such as the
According to Transition State Theory (TST), a chemical reaction proceeds from the reactant state to the product state via a metastable transition state existing as a hypersurface in phase space. The transition state is understood to be in quasi-equilibrium with the reactant state, meaning that crossing the barrier always leads to the product state. The shortcomings of TST, however, include not considering re-crossing events (formation of reactants even after crossing the barrier) and quantum tunneling. The free energy profiles of an uncatalyzed reaction and its corresponding enzyme-catalyzed reaction are schematically represented in Figure 4.

2.2.2 Enzyme Catalysis
Enzymes are biological catalysts that function to increase the rate of a chemical reaction by lowering the free energy barrier between the reactant and transition states without undergoing any permanent change themselves. An enzyme initially forms a stable complex with the reactant followed by the chemical step that results in product formation. The final step involves product release, and this cycle is repeated again for the next reactant molecule. The catalytic effect of the enzyme on the reaction is on display over the entire process from substrate binding to product release. Elucidating the way in which catalysis occurs is an important area of study to understand the origins of enzyme catalysis.

One of the earliest models proposed to explain enzyme catalysis was Fischer’s “Lock and Key” hypothesis, which was based on shape complementarity between the active site of the enzyme and the substrate. This was later followed by Koshland’s induced fit model, which suggested that the substrate molecule induces conformational changes in the enzyme to accommodate itself in the active site. However, studies showing tight binding between enzyme and transition state analogues indicated that there are other factors influencing enzyme catalysis.

**Transition State Stabilization**
The theory of transition state stabilization was first propounded by Pauling, in which he proposed that the active site of an enzyme binds to the transition state with greater affinity than to the substrate. In this case, the free energy
of the transition state in complex with the enzyme is lower than that of the transition state for an uncatalyzed reaction in solution. This is made possible as a result of a favorable active site environment in which suitable polar and non-polar residues are immaculately arranged in three-dimensional space to facilitate stabilization of the transition state as compared to the reactant and product states. The theory has since garnered support from the concepts of electrostatic stabilization\textsuperscript{74} and active site preorganization,\textsuperscript{75,76} which state that the stabilizing charges in an active site are preorganized for preferential binding to the transition state and do not have to reorient themselves during the reaction. More recently, these theories have been validated by experimental and computational studies.\textsuperscript{81–87}

**Ground State Destabilization**

The active site of an enzyme can also function to destabilize the bound substrate, thus lowering the free energy difference between the reactant state (or ground state) and the transition state. Ground state destabilization implies that the free energy of the transition state is not affected by the enzyme, and catalysis occurs solely by raising the free energy of the ground state in relation to the transition state. Though this theory has been proposed to work in some cases,\textsuperscript{84,85} newer studies have shown that ground state destabilization is not as significant as transition state stabilization.\textsuperscript{86–89}

2.2.3 Structural and Dynamic Effects on Enzyme Function

The effect of protein structure – the order and disorder therein – on its function has already been emphasized in this chapter. It is no different for enzymes, which contain distinct structural scaffolds that help in properly orientating relevant amino acids at the active site. Some disordered regions in enzymes such as loops play a hugely important role in regulating the catalytic activity. The growth of structural and computational biology in recent decades has resulted in large amounts of data that act as evidence to prove that dynamic effects leading to conformational changes in an enzyme are critical for catalysis.

The significance of dynamic effects – correlated motions\textsuperscript{90} and promoting vibrations\textsuperscript{91} to name a few – on enzyme function have been shown using numerous computational studies.\textsuperscript{92–95} My work on modeling the enzyme-catalyzed reaction in triosephosphate isomerase (TIM) has also yielded interesting insights with regard to the effect of active site architecture (Papers III and IV) and loop motion (Paper VI) on catalysis.
2.2.4 Triosephosphate Isomerase

Triosephosphate isomerase (TIM) is an enzyme that reversibly catalyzes the conversion of dihydroxyacetone phosphate (DHAP) to D-glyceraldehyde-3-phosphate (GAP) (Figure 5).\textsuperscript{96–99} It is a glycolytic enzyme which performs an important function of regulating the cellular concentrations of GAP and DHAP, as these molecules link the glycolytic pathway to other metabolic pathways such as the pentosephosphate pathway, gluconeogenesis, and lipid metabolism.\textsuperscript{1}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{reaction_scheme.png}
\caption{Reaction scheme showing the reversible interconversion of DHAP and GAP catalyzed by TIM.}
\end{figure}

Structurally, TIM is a homodimer consisting of two identical subunits or monomers about 250 residues in length (248 residues in yeast TIM\textsuperscript{100}) (Figure 6A). A notable feature of its structure is the TIM barrel fold, which is the most common structural fold observed in nature, found in approximately 10% of all known proteins.\textsuperscript{101–103} It comprises of an eightfold repeating pattern of β-strands and α–helices, also represented as the (βα)\textsubscript{8}-fold. The β-strands and α–helices are linked by flexible loops, referred to as loop 1, loop 2, and so on until loop 8.\textsuperscript{104}

The motion of loops 6 and 7 has significant relevance to the catalytic function of TIM.\textsuperscript{97} Various conformational states of these loops in ligand-bound and “apo” enzymes have been studied with the help of structural models.\textsuperscript{100,105–107} Closure of TIM loop 6 upon substrate binding forms non-covalent stabilizing interactions between the enzyme and the phosphodianion moiety of the substrate.\textsuperscript{108} This is a classic example of ligand-driven conformational change,\textsuperscript{109} a feature also observed in other enzymes such as glycerol-3-phosphate dehydrogenase (GPDH)\textsuperscript{110,111} and orotidine 5’-monophosphate decarboxylase (OMPDC)\textsuperscript{112}. Structural analysis has also helped to elucidate the active site architecture of TIM, with key amino acid residues such as E165, H95, and K12 (residue numbering from yeast TIM\textsuperscript{100}, the same numbering will be used...
at all places in this thesis) surrounding the bound substrate and creating a favorable environment for the chemical reaction to occur (Figure 6B).

Figure 6. (A) Cartoon representation of the dimeric form of yeast TIM (PDB ID: 1NEY).13,100 (B) Close-up view of the active site of TIM showing key amino acid residues (grey color, stick representation) and the bound substrate DHAP (yellow color, ball-and-stick representation). Figure generated using PyMOL.41

TIM catalyzes the isomerization reaction 10^9 times faster than the uncatalyzed reaction in solution.97,114–117 The catalytic proficiency of this enzyme is a result of optimal arrangement of active site residues for transition state stabilization, and dynamic properties of a flexible loop that is involved in reorganization of the active site.97,113 Thus, we observe both structural order and disorder at play in TIM for efficient catalysis of the isomerization reaction. Chapter 5 will provide a more detailed overview of this enzyme system with greater relevance to my investigations on TIM-catalyzed reactions using computational approaches.
3. Computational Approaches

The development of new theories in modern physics and chemistry around the early 20th century heralded a new age in scientific research that prompted the use of theoretical calculations to study natural phenomena. In the early stages of these fields, calculations were performed manually in order to derive equations supporting new theories such as quantum mechanics. However, the advent of powerful computers over the last few decades has catalyzed the maturation of this field.

The study of sub-atomic particles, atoms, and molecules pervades various disciplines of science. All known matter is composed of these very components arranged in various permutations and combinations. This fundamental property of matter makes theoretical chemistry a very powerful tool to study any reasonably sized molecular system using mathematical models at relevant timescales. The field of computational biochemistry has made and continues to make good use of theoretical models in order to study the structural and dynamic properties of biomolecular systems. Furthermore, it also enables the investigation of their function at an atomic level.

This chapter contains a brief overview of the computational approaches that are most commonly used on biomolecular systems. A more detailed description of the underlying principles is included here for those methods that have been utilized in the research work presented in this thesis.

3.1 Quantum Mechanics

Quantum mechanics (QM) is a theory that can be used to provide a physically meaningful description of chemical systems, particularly at the sub-atomic level. One of the fundamentals of QM is the description of a system in the form of a wave function (Ψ). The Schrödinger equation (Equation 3.1) can be solved to obtain information about the current state of a system.

\[ \hat{H} \Psi = E \Psi \]

(3.1)
Here, $\hat{H}$ is the Hamiltonian operator that acts upon the wave function $\Psi$, resulting in $\Psi$ multiplied by a scalar quantity $E$, which is the energy associated with the wave function. It is often difficult to know $\Psi$ for non-trivial systems, making it extremely difficult to solve equation (3.1) analytically. Hence, in order to solve the equation numerically, approximations of certain aspects of the theory are made such that they do not significantly influence the final solution. The most commonly used approximation is the Born-Oppenheimer (BO) approximation, which treats the atomic nuclei as stationary particles relative to the fast-moving electrons, thereby decoupling their motions.

3.1.1 *Ab initio* methods

The Latin term *ab initio* translates to “from first principles”. *Ab initio* methods, thus, make use of the basic principles of quantum mechanics to determine the state of a system. The BO approximation of the Schrödinger equation is usually a starting point for these methods. More approximations are introduced to relatively simplify the process of solving the Schrödinger equation for non-trivial systems, without compromising a lot on the accuracy. The Hartree-Fock (HF) method, for example, is a classic *ab initio* method that introduces an approximation by neglecting the phenomenon of electron correlation and instead treats each electron with an effective average potential. Post-HF methods include Möller-Plesset perturbation theory (MPn), configuration interaction (CI), and coupled cluster theory (CC). These methods treat electron correlation using different kinds of approximations that improve the accuracy of solutions relative to HF, but are computationally more expensive.

3.1.2 Density Functional Theory

Density Functional Theory (DFT) is used to solve for energies of molecular systems by calculating the electron density. It serves as a computationally less intensive option in comparison to *ab initio* methods that make use of the electron wave function. In this theory, the ground state energy of a system is calculated as a sum of three components: (A) the attraction between nuclei and electrons, depending on the electron densities, (B) the kinetic energy of electrons, and (C) the interaction energy between electrons. The challenge in DFT lies in successfully arriving at a functional that accurately describes the electron kinetic energy and electron interaction energy. This is done by carrying out parametrization of the functional to reproduce experimentally determined properties of known compounds, thus making DFT an empirically corrected QM approach.
We now begin to observe that introducing approximations to fundamental theories has many advantages. Firstly, it allows us to develop algorithms and mathematical models that are less intensive to compute. This saves time as well as valuable computational resources. Secondly, the results from these approximated methods can be used for practical purposes without compromising on accuracy, as they do not deviate significantly from the exact method.

The QM-based methods discussed so far are useful for modeling chemical reactivity, as they can describe bond cleavage and formation. However, due to their high computational cost, these approaches are also limited by the number of atoms in a system that can be studied with acceptable computational efficiency. Therefore, if the objective of a simulation is to model the motion or dynamics of a protein or an enzyme system in solvent, we need to opt for a more approximate method.

3.2 Molecular Mechanics

Molecular mechanics (MM) makes use of a classical mechanical framework to describe a molecular system. Approximations are introduced in order to describe interatomic interactions, and Newtonian equations of motion are implemented for modeling dynamic behavior. In this framework, atoms are treated as balls or whole particles, whereas the bonds connecting them are treated as springs. Consequently, subatomic particles such as protons, neutrons and electrons are not explicitly considered. However, information about the mass and charge (partial charge, when considered as part of a molecule) of an atom is included in a key function called a force field. A force field is central to any MM-based method, and consists of a number of individual functions and parameters that describe various bonded and non-bonded interactions that are possible between intra- or intermolecular atoms. The standard form of a force field is given by:

\[
V = \sum_{\text{bonds}} \frac{1}{2} k_b \cdot (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_{\phi} \cdot (\phi - \phi_0)^2 \\
+ \sum_{\text{dihedrals}} k_{\psi} \cdot (1 + \cos(n \cdot \psi - \delta)) + \sum_{\text{improvers}} \frac{1}{2} k_{\theta} \cdot (\theta - \theta_0)^2 \\
+ \sum_{\text{nonbonded}} \frac{q_i q_j}{4 \pi \varepsilon_0 r_{ij}} + A_{ij} \frac{1}{r_{ij}^{12}} + B_{ij} \frac{1}{r_{ij}^{6}}
\]

(3.2)
Here, the terms $k_b, k_\phi, k_\psi$ and $k_\theta$ represent the force constants for the potentials describing the harmonic motion of the bonds, angles, proper dihedrals, and improper torsions, respectively. Coulomb’s law is used to calculate the electrostatic interactions between two atoms. The values $q_i$ and $q_j$ represent the partial atomic charges for the two atoms $i$ and $j$. $A_{ij}$ and $B_{ij}$ are the repulsive and attractive terms, respectively, of the Lennard-Jones potential that is used to describe van der Waals interactions between the atoms $i$ and $j$. Solving for equation (3.2) gives the overall potential energy of the system.

The parametrization of these functional terms is done by fitting to data determined by experiments and/or by performing high-level QM calculations on the molecular fragments of interest. At present, many force fields have been developed that are used to simulate dynamics of biomolecular systems. Some prominent ones include CHARMM36m, AMBER-ff14SB, and OPLS-AA/m, the latter two of these having being extensively used in my investigations presented as part of this thesis. The individual functions and parameters are modified differently in each of these force fields, but they largely draw inspiration from equation (3.2).

Since the MM force field is a relatively simpler function to solve than the ones we encounter in QM-based methods, it can be employed for simulating large biomolecular systems comprising thousands of atoms. In order to study changes in atomic positions and variations in molecular conformations, the potential energy of a system needs to be calculated repeatedly in a deterministic fashion. The outcome of this exercise is molecular dynamics (MD). This involves the integration of Newton’s equations of motion, thereby allowing the system to propagate in time. A trajectory resulting from a MD simulation contains information about the spatial coordinates of all atoms included in the system. This information is vital for analyzing conformational transitions in local or global regions of a biomolecular system, and for identifying energetically stable structural conformations. The major disadvantage of MM in comparison to QM is its inability to describe bond breaking and bond formation events of a chemical reaction, although reactive force fields such as ReaxFF or specialized approaches such as the empirical valence bond (EVB) approach have been developed that permit changes in bonding patterns to be modeled also in an MM framework.
3.2.1 Enhanced Sampling Methods

Exploration of the potential energy surface of a system using conformational sampling remains a daunting task for MM-based methods. A single MD trajectory allows for a deterministic exploration of the various energy minima for a system, propagating from a solitary starting point. This is theoretically possible if the trajectory is allowed to continue indefinitely, but in practice is difficult to obtain a comprehensive scan of the energy landscape as a result of a single long timescale simulation. A usual workaround to this problem is to run multiple simulations with non-identical starting points (initial velocities) that propel the system in different directions in the quest to sample the different energy minima accessible to the system. This manner of improving the sampling of the conformational space is fairly common, as demonstrated in all the papers included in this thesis. However, over time, there has been a push towards the development and implementation of innovative methods into MM in an effort to maximize conformational sampling of the system at minimal computational cost. Some of these methods, such as Hamiltonian Replica Exchange molecular dynamics (HREX-MD) and Biased Exchange Metadynamics (BE-METAD), have been used in Paper VI to extensively sample the conformational diversity of loop 6 in TIM.

Enhanced sampling methods can be broadly classified into two groups based on whether they require a biasing collective variable (CV) or not. In a general sense, a CV is a function of the coordinates of the system which is able to explicitly describe the transition of a system from a known initial state to a known final state. This is particularly advantageous if there is prior knowledge about the initial and final states of the system following, for example, a conformational change. However, the effectiveness of a CV-dependent enhanced sampling method – such as metadynamics – relies heavily on the quality of the chosen CV. On the other hand, methods that are not constrained by CVs – such as replica exchange MD – can explore the potential energy surface of a molecular system without being biased towards sampling a pre-defined conformational change.

**Replica Exchange Molecular Dynamics**

In replica exchange MD (REMD), enhancement of the conformational sampling of the system is achieved by running multiple replicas of simulations at different temperatures in parallel. An exchange of system configurations is done at a certain frequency between pairs of replicas, after which the ex-
changed replicas are simulated at *new* temperatures. This aids in the convergence of various replicas towards a common energy minimum, and also allows for greater sampling of the conformational landscape of the system. It also helps in preventing systems from getting trapped in conformations corresponding to local energy minima.\textsuperscript{138,139}

**Metadynamics**

Metadynamics, a CV-based enhanced sampling approach, makes use of a history-dependent bias potential acting on the CV to accelerate the sampling of rare events in a simulation. This is usually done by flattening the free energy landscape of the CV by populating it with Gaussians, thereby allowing for better sampling of all conformations along the CV space.\textsuperscript{137,138} The simulations are run until the bias potential converges to the negative of the real free energy of the CV. In systems exhibiting complicated motions, the sampling efficiency can be further improved by using a combination of metadynamics and replica exchange methods.\textsuperscript{137,138}

### 3.2.2 Calculation of Free Energies

The free energy of a system is a fundamental concept in physical chemistry that can be used to describe a process in thermodynamic terms. In various computational approaches, calculation or estimation of the free energy of a process (for example, a chemical reaction) is of paramount importance. Two of the most commonly used methods for free energy calculations – free energy perturbation (FEP), and umbrella sampling (US) are discussed for a two-state system in this section.

**Free Energy Perturbation**

The Zwanzig equation\textsuperscript{140} can be used to evaluate the free energy difference between the two states 1 and 2 of a system by calculating the ensemble average of the potential energy difference sampled at state 1.

$$
\Delta G = -RT \ln \left( e^{-\frac{(U_2-U_1)}{RT}} \right)_1
$$

Here, $\Delta G$ is the free energy difference, $R$ is the universal gas constant, $T$ is the temperature in Kelvin, and $U_1$ and $U_2$ are the potential energies corresponding to states 1 and 2, respectively. This relation is however accurate only if the two states overlap significantly with respect to the conformational spaces they represent. To overcome this problem, a number of intermediate states are introduced between states 1 and 2 in order to gradually perturb the system.
A new potential is created as a linear combination of the two potentials, with \( \lambda \) being the coupling parameter whose values range from 0 to 1.

\[
U(\lambda) = (1 - \lambda)U_1 + \lambda U_2
\]  

(3.4)

The free energy of the system going from state 1 to state 2 can then be calculated as a sum over all intermediate steps.

\[
\Delta G_{1 \rightarrow 2} = -RT \sum_{\lambda=0}^{1} \ln \langle e^{-\frac{(U(\lambda + \Delta\lambda) - U(\lambda))}{RT}} \rangle_{\lambda}
\]

(3.5)

This approach is referred to as free energy perturbation (FEP), and is highly instrumental in studying enzymatic chemical reactions, especially when the initial and final states are known beforehand.

**Umbrella Sampling**

While the FEP methodology entails the generation of intermediate states for calculation of the free energy, it also presents a challenge of being able to sufficiently sample rare events associated with or corresponding to these intermediate states. These could include, for example, crossing of the energetic barrier to identify the transition state in a chemical reaction. The umbrella sampling (US) approach allows for ample sampling of the system near the intermediate states. This is done by adding a biasing potential \( W \) to the sampling potential \( U_1 \) for the purpose of improving the sampling, resulting in an effective potential \( U_{total} \).

\[
U_{total} = U_1 + W
\]

(3.6)

Evaluation of the free energy from US calculations is eventually done by removing the energy contribution associated with the biasing potential from the total energy.

**3.3 Multiscale Modeling**

Methods based on the principles of quantum mechanics are capable of describing chemical reactions that involve the breaking and forming of bonds. However, the high precision of these calculations comes at significant computational cost, such that their computational efficiency reduces dramatically with increasing system size. This compels users to truncate the size of the system to include only the reactive species for a fast and accurate description of reaction energetics, thereby losing information regarding the effect of surrounding
atoms on the reacting atoms.\textsuperscript{141,142} Molecular mechanics, on the other hand, oversimplifies the model of an atom by treating it as a whole particle with pre-defined bonding patterns. This rules out the possibility of modeling chemical reactions using conventional MM-based methods (with the exception of specialist approaches such as those mentioned in the introduction to Section 3.2). A hybrid method called QM/MM (quantum mechanics / molecular mechanics) was developed to solve this issue by amalgamating scalability, efficiency, and theoretical rigor.

3.3.1 The QM/MM Approach

In a \textit{multiscale} model,\textsuperscript{143,144} the system being studied is partitioned into two regions – the reactive region, and the surrounding region. The reactive region (also termed as the QM region) is a small group of atoms that are actively involved in the reactive step, and are therefore treated at the more accurate QM level theory. The remainder of the system, \textit{i.e.} the surrounding region (also termed as the classical region), is treated at the more approximate MM level of theory.

The potential energy of the system ($U_{\text{total}}$) is, hence, a sum of the potential energies of the classical ($U_{\text{MM}}$) and QM ($U_{\text{QM}}$) regions, as well as a coupling ($U_{\text{QM/MM}}$) between the two regions.

\begin{equation}
U_{\text{total}} = U_{\text{QM}} + U_{\text{MM}} + U_{\text{QM/MM}}
\end{equation}

(3.7)

The QM/MM approach thus extracts the best of both QM- and MM-based methods, by modeling chemical reactivity as accurately as possible and also by describing the neighborhood of the reactive center.

3.3.2 The Empirical Valence Bond Approach

The empirical valence bond (EVB) approach was developed by Warshel and Weiss in 1980 as a simple and reliable approach to study solvation effects in enzyme-catalyzed chemical reactions.\textsuperscript{134,135} While the QM/MM approach manages to improve system scalability and optimization of computational power for accurate description of enzymatic reactions, calculation of energies for the QM region can still prove to be significantly time-consuming and computationally expensive. EVB presents itself as a semi-empirical approach that, in a QM/MM-like framework, makes use of classical force fields and empirically-fitted parameters to describe the chemistry in an enzymatic system,
thereby allowing for fast and extensive conformational sampling of the reactive region.

The EVB approach defines a chemical reaction as a mixture of independent valence bond states that are coupled by reaction-specific parameters. These diabatic states correspond to the reactant and product states of a reaction, and can also represent any reaction-specific intermediate states. Each of these diabatic states are represented by parabolic functions, which in turn are described using classical force fields and are combined with the help of the EVB mapping parameters $H_{ij}$ and $\alpha$ (for descriptions, see below), to produce an adiabatic free energy surface. This is schematically represented in Figure 7 for a simple two-state reaction.

![Figure 7](image)

*Figure 7.* Schematic representation of the EVB ground state adiabatic free energy surface ($\epsilon_g$, black) for a simple 2-state reaction (equation 3.12). The EVB diabatic parabolas for the two states are represented as $\epsilon_{rs}$ for the reactant state (red) and $\epsilon_{ps}$ for the product state (blue). The reaction coordinate can be geometric in nature, but can also be the energy gap between the two diabatic states. The off-diagonal term, $H_{12}$, describes the coupling between the diabatic states $\epsilon_{rs}$ and $\epsilon_{ps}$. See main text for detailed description of all parameters. Figure adapted from ref. 145 with permission from Elsevier.

As discussed in Section 3.2, molecular mechanics uses a harmonic approximation to describe bond stretch and hence cannot model chemical events. The EVB approach employs the Morse potential to specifically represent the energy of breaking and forming bonds in a reaction. This ensures that calculation of reaction energetics is not performed at a significant cost to computing power, as it would be with standard QM/MM approaches.
Within an EVB framework, the energy of a diabatic state is expressed mathematically as:

\[ H_{ii} = \varepsilon_i = U_{\text{intra}}^i(R, Q) + U_{\text{inter}}^i(R, Q, r, q) + U_{\text{solvent}}^i(r, q) + \alpha_{\text{gas}}^i \] (3.8)

where \( R \) and \( Q \) denote the atomic coordinates and charges of the reacting atoms, and \( r \) and \( q \) denote the atomic coordinates and charges of all the remaining surrounding atoms which include non-reacting protein atoms as well as solvent molecules. The \( U_{\text{intra}} \) term represents the intramolecular interactions of the solute system, \( U_{\text{inter}} \) represents the interaction between atoms of the solute and the surrounding solvent, and \( U_{\text{solvent}} \) is the energy associated with the solvent. The \( \alpha_{\text{gas}}^i \) term denotes the gas-phase energy, which is a constant value when all the fragments (reacting atoms, solute, and solvent) are considered to be at infinite separation. This can be used to adjust the parabolas with respect to each other for a reference reaction, the free energies of which are known from experiments or QM calculations. The off-diagonal term \( H_{ij} \) is the coupling parameter that combines the diabatic states to obtain an adiabatic energy surface, and is represented by a simple exponential function:

\[ H_{ij} = Ae^{-a|\Delta R'|} \] (3.9)

where \( A \) is a constant and \( \Delta R' \) is the distance between atoms of reacting species.

Using the expressions for \( H_{ii} \) (representing the diabatic states) and \( H_{ij} \) from equations (3.8) and (3.9), the Hamiltonian matrix can be constructed as

\[ H = \begin{bmatrix} \varepsilon_1 & H_{ij} \\ H_{ji} & \varepsilon_2 \end{bmatrix} \] (3.10)

where \( \varepsilon_1 \) and \( \varepsilon_2 \) represent the energies of the valence bond states. The adiabatic ground state energy \( E_g \) can be calculated by using this Hamiltonian.

\[ \begin{vmatrix} \varepsilon_1 - E_g & H_{ij} \\ H_{ij} & \varepsilon_2 - E_g \end{vmatrix} = 0 \] (3.11)

Solving this eigenvalue problem for \( E_g \) results in the following solutions which correspond to the ground state (lowest eigenvalue) and transition state energies.\(^{146}\)

\[ E_g = \frac{1}{2} \left[ (\varepsilon_1 + \varepsilon_2) - \sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4 \cdot H_{ij}^2} \right] \] (3.12)

The activation free energy, \( \Delta G^\ddagger \), can be calculated by gradually changing the system from one state to another using a “mapping potential”, \( U_{\lambda} \), along the coordinate of \( \lambda \).

\[ U_{\lambda} = \lambda U_{\varepsilon_1} + (1 - \lambda) U_{\varepsilon_2} \] (3.13)
By incrementally varying the value of $\lambda$, the free energy can be obtained using the FEP/US method as outlined in Section 3.2.2. The free energy difference between each $\lambda$ window is evaluated as

$$\Delta G(\lambda_i) = -RT \cdot \ln \sum_{n}^{i-1} \left( e^{-\frac{U_{n+1} - U_n}{RT}} \right) \frac{\langle e^{-U_n} \rangle}{\langle e^{-U_n} \rangle} \frac{\langle e^{-U_n} \rangle}{\langle e^{-U_n} \rangle}$$

(3.14)

The free energy gap, $U_{\text{gap}}$, is used as a generalized reaction coordinate to construct the free energy profile in EVB. It is the difference in free energies of the two diabatic states for a given set of coordinates $r$.

$$U_{\text{gap}}(r) = U_{\varepsilon_2}(r) - U_{\varepsilon_1}(r)$$

(3.15)

EVB calculations provide us data regarding the free energies of a catalyzed reaction, and comparing free energy profiles of different enzyme variants gives a good account of the effect of mutations on the catalytic step. In addition to this, analysis of energy contributions coming from various parts of the system – the reacting atoms, surrounding protein atoms, and solvent molecules – can provide greater insights into the catalytic mechanism. For example, the linear response approximation (LRA)\textsuperscript{147} can be used to calculate differences in interaction free energies between any two states. These energies can be delineated to get energy contributions of specific amino acids or water molecules to the reacting species.

$$\Delta G_{\text{LRA}} = \frac{1}{2} \cdot (\langle U_2 - U_1 \rangle_2 + \langle U_2 - U_1 \rangle_1)$$

(3.16)

The work showcased in this thesis has used the EVB approach to provide a microscopic rationale to the catalytic mechanism of enzyme-catalyzed reactions in TIM, especially with regard to the analyses involving linear free energy relationships (LFER) which correlate the computed values of activation free energies and thermodynamic free energies. Chapter 5 will expand on the utility of this analysis with respect to the work done in Papers III and IV.
4. Conformational Dynamics of Aβ peptide

The amyloid-β peptide (Aβ) is an intrinsically disordered protein of length ranging from 39 to 43 amino acid residues. The close association of this peptide with the pathology of Alzheimer’s disease (AD), a neurodegenerative disorder, has garnered tremendous interest among various disciplines of the scientific community. The Aβ peptide, owing to lack of a distinct stable structural form, is prone to aggregation into oligomers and extremely stable fibril-like structures. From a medical perspective, it is of utmost value to understand how Aβ aggregation is related to AD progression. It is still unclear whether Aβ misfolding and aggregation is a cause, an effect, or a combination of both for disease development.

Structural biophysicists and biochemists are particularly interested in the process of misfolding and aggregation, the various structural forms that Aβ attains during its conformational transitions, the kinetic and thermodynamic parameters associated with these events, and the effect of external influences. Computational approaches such as MD simulations and other enhanced sampling methods can be useful in not only visualizing the structural changes in the Aβ peptide as monomers or oligomers, but also in studying the influence of external factors that are both physical (temperature) and chemical (small molecules, metal ions, surfactants, lipids) in nature.

In my research investigations on the Aβ peptide, I have used computational approaches to: (A) characterize the interactions between Aβ and divalent metal ions, and (B) study the structural implications of Aβ binding to amphiphilic surfactants. These studies were carried out in collaboration with experimental biophysicists, who used a range of techniques to obtain information about binding affinity, secondary structure, aggregation kinetics, and images of aggregates. Long timescale MD simulations were vital for configurational sampling of Aβ peptide bound to metal ions and surfactant micelles, and provided greater support to the findings by complementing well with experimental results.
4.1 Binding Interactions with Metal Ions (Paper I)

The possibility of metals playing a role in the pathogenesis of AD has been evidenced by numerous clinical and biochemical studies. The fact that a monomeric Aβ peptide bears a net negative charge of around -3 at physiological pH warranted an understanding of its binding interactions with positively charged small molecules such as polyamines, and with metal ions.

The study reported in Paper I characterized binding interactions of the Aβ peptide with divalent metal ions using biophysical experiments as well as computer simulations, with my contribution being to the latter. Specifically, zinc (Zn²⁺), copper (Cu²⁺), and manganese (Mn²⁺) ions were chosen for the study. The binding affinity of the Aβ peptide towards Zn²⁺ and Cu²⁺ ions was known to be in the micromolar to nanomolar range, indicating a probable well-defined binding site in the Aβ peptide that is composed of three histidines (H₆, H₁₃, H₁₄) and the N-terminal aspartate (D₁). Beyond this, the binding of Mn²⁺ ions to Aβ peptide had not been extensively studied in the literature. Manganese is known to exist in multiple oxidation states and plays a key role in various biological processes. Thus, this work was significant as it was aimed at characterizing the binding site for all the three metal ions, and also study the effect of Mn²⁺-binding on Aβ aggregation.

As mentioned in Section 2.1.3, two isoforms of the Aβ peptide are commonly studied – Aβ(1-40) and Aβ(1-42). In this study, Aβ(1-40) was used for various biophysical experiments, whereas a simpler model comprising the hydrophilic N-terminal region Aβ(1-16) was used in MD simulations. This truncated region of the peptide was chosen due to its likelihood to contain a specific binding site for metal ions in the form of polar amino acid residues. Additionally, it also meant a smaller system size for the simulations which ensured longer sampling times with lesser expenditure of computational resources. This strategy was drafted keeping in mind the primary objective of the study, which was to characterize the binding site of metal ions, and not sampling the configurational space of the Aβ(1-40) peptide.

Long timescale MD simulations of Aβ(1-16) in complex with metal ions (Zn²⁺, Cu²⁺, Mn²⁺) in a 1:1 stoichiometry resulted in extensive configurational sampling. Among all the metal ions, binding of Zn²⁺ ion to Aβ(1-16) was the most stable, with H₆, H₁₃, H₁₄ and E₁₁ forming the coordination center (Figure 8A). Cu²⁺ too was fairly stable and had a coordination center similar to
that of Zn$^{2+}$, except for the dissociation of H13 (Figure 8B). Contrary to both Zn$^{2+}$ and Cu$^{2+}$, Mn$^{2+}$ was observed to interact with the residues D1, E3, D7 and E11 with occasional changes in binding modes (Figure 8C). These theoretical observations were found to be in good agreement with experimental observations which showed that while binding of Zn$^{2+}$ and Cu$^{2+}$ to Aβ peptide is stable and at a well-defined coordination center, Mn$^{2+}$-binding is relatively weak and transient in nature. It was further concluded that Mn$^{2+}$-binding might not have an influence on Aβ aggregation.

![Figure 8](image_url)

Figure 8. Aβ(1-16) peptide in complex with (A) Zn$^{2+}$ (orange), (B) Cu$^{2+}$ (green), and (C) Mn$^{2+}$ (purple) ions. Shown here are representative structures of the top-ranked clusters obtained in each case upon performing clustering analysis of MD simulation trajectories. The starting conformation in each case is the divalent metal coordinating by H6, E11, H13 and H14. Distances indicated in the figure correspond to average values over all trajectory frames for each system. Figure adapted and reprinted from ref. 148 with permission from Elsevier.

Binding of Aβ to metal ions can influence initiation or acceleration of peptide aggregation by trapping the monomeric peptide in a partially folded conformation. Although this computational study could not explore the effect of metal ion binding on Aβ aggregation, it was able to confirm the composition and stability of the binding site for Zn$^{2+}$ and Cu$^{2+}$ ions. It also provided first-hand information about the characteristics of Mn$^{2+}$-binding to Aβ peptide.

4.2 Binding Interactions with Surfactants (Paper II)

The Aβ peptide is largely composed of hydrophobic amino acid residues, except for the N-terminal region, which is relatively hydrophilic as a result of a higher concentration of polar and charged amino acid residues (Figure 2). Likewise, lipid molecules that form vesicles and membranes are amphiphilic, meaning that they are made up of a long hydrophobic tail and a polar head-group which could carry a positive, negative or neutral charge. The similarity
of this differential in charge, polarity and hydrophobicity across a monomeric Aβ peptide and a lipid molecule can potentially lead to binding interactions between their respective hydrophilic and hydrophobic regions. Understanding the nature of interactions between the Aβ peptide and lipid membranes is therefore of significant interest.

The goal of this study, as reported in Paper II, was to characterize the binding interactions between the Aβ(1-40) peptide and micelles of amphiphilic surfactants. These surfactants share molecular similarities with lipid molecules, and their micellar forms function as biomembrane mimetics. As was the case in the previous study (Paper I) concerning metal ion interactions, a combination of biophysical experiments and computer simulations was used in this work as well. However, this study used the full length Aβ(1-40) peptide for MD simulations as both the hydrophilic and hydrophobic binding interactions had to be characterized. Surfactants differing in physical and chemical properties such as charge on the headgroup, length of the aliphatic chain, micellar size, and aggregation number were selected. For the computational study, a non-ionic and a negatively charged surfactant in the form of dodecyl β-D-maltoside (DDM) and sodium dodecyl sulfate (SDS) respectively, were chosen to analyze the effect of electrostatics and micellar surface area on Aβ binding.

An NMR study showed that Aβ attains an α-helical structure – particularly in the core and C-terminal regions – when bound to SDS micelles, with the N-terminal region remaining relatively disordered. Hence, two different configurations of the Aβ(1-40) were studied – one in which the peptide was a random coil and bound to the micellar surface, and the other in which the α-helical portion of the peptide was buried into the micelle. MD simulations showed that the random coil peptide remained bound to the micellar surface with its hydrophobic region interacting with shallow hydrophobic grooves on the micelle. This was observed in the cases of both DDM and SDS micelles (Figure 9A and 9B).

The α-helical peptide, on the other hand, displayed peculiar behavior during the simulations, the most noticeable one being the formation of a kink in the α-helix at residues N27 and K28. This feature was consistent across both micelle systems (Figure 9C and 9D). However, the defining difference between the interactions of Aβ(1-40) with DDM and SDS micelles was the binding affinity which was computationally estimated. The DDM-bound peptide
moved towards the micellar surface, with the N-terminal region showing weak interactions with the non-ionic headgroup of DDM. On the other hand, the peptide was able to tightly bind to the SDS micelle, with its hydrophilic and disordered N-terminal region making strong electrostatic contacts with the negatively charged headgroups and the α-helical part binding to deep hydrophobic grooves on the micelle.

*Figure 9.* Top-ranked cluster representations from MD simulation trajectories of (A) a disordered Aβ(1-40) peptide in complex with a DDM micelle, (B) a disordered Aβ(1-40) peptide in complex with an SDS micelle, and (C) an α-helical Aβ(1-40) peptide in complex with a DDM micelle, and (D) an α-helical Aβ(1-40) peptide in complex with an SDS micelle. Cartoon representation of the peptide is shown in purple, with the hydrophobic regions highlighted in yellow. The micelles are shown in transparent surface representation, with carbon atoms shown in white, oxygen atoms shown in red, and sulfur atoms shown in yellow. Reprinted (adapted) with permission from Ref. 168. Copyright 2018 American Chemical Society.

The study thus underlined the importance of electrostatic and hydrophobic interactions between the Aβ peptide and surfactants. Computational methods were able to elucidate the structure and dynamics of these binding partners, and in the process support relevant data obtained from experiments. The scope of these results could be extended also to lipid molecules of cell membranes. Membrane surfaces formed by negatively charged groups could potentially act as strong anchor points for free Aβ peptides and act as surface catalysts for Aβ aggregation. The findings of this study can be of potential use for AD research.
4.3 Perspectives

Scientific studies on Alzheimer’s disease and the Aβ peptide are being carried out globally by a large number of research groups. Each incremental step towards understanding Aβ aggregation and its association with AD pathology can potentially go a long way towards reaching the ultimate objective of this research, which is to develop a cure or preventive therapy for this neurodegenerative disease. The collaborative work showcased in this chapter is an effort to answer the questions pertaining to the nature of Aβ binding to physiologically relevant partner molecules. In these investigations, computational methods in the form of conventional MD simulations have played a key role in putting the pieces of the jigsaw puzzle together with experimental observations, and furthering our understanding of the structure and dynamics of IDPs.
5. Computational Modeling of Enzyme Catalysis

Scientific understanding of the origins of enzyme catalysis is an endeavor that presents challenges of various kinds for different classes of enzymes. While some enzymes make use of external cofactors and metal ions to achieve catalysis, there are other enzymes that are able to efficiently utilize their structural scaffold to catalyze a chemical reaction. Triosephosphate isomerase (TIM), an enzyme belonging to the latter category, has been the subject of numerous mechanistic studies mainly due to its excellent catalytic proficiency. There exists a wealth of biochemical and structural knowledge about this enzyme, making it an ideal model for performing computational investigations to get more insights into its catalytic origins.

Figure 10. Reaction scheme of the mechanism of TIM-catalyzed reversible isomerization of D-glyceraldehyde-3-phosphate (GAP) to dihydroxyacetone phosphate (DHAP) via enediolate intermediates. Reprinted (adapted) with permission from Ref. Copyright 2019 American Chemical Society.

The reaction catalyzed by TIM is the reversible conversion of dihydroxyacetone phosphate (DHAP) to D-glyceraldehyde-3-phosphate (GAP), via enzyme-bound enediolate intermediates (Figure 10). Although the reaction scheme showcases the mechanism as a series of simple proton transfer steps, the actual events occurring in the enzyme are much more complex and involve a proper orchestration of the ensemble of amino acid residues that form the active site. The work described in this chapter is predominantly dedicated to computational modeling of TIM-catalyzed substrate deprotonation using the empirical valence bond (EVB) method.
5.1 Modeling the Hydrophobic Clamp of TIM (Paper III)

The most noteworthy structural feature of TIM that is crucial for ensuring efficient catalysis is the large conformational change of “loop 6” to close over the active site. Loop 6 is important for two reasons. Firstly, it bears a bulky non-polar residue in the form of I170 which desolvates the active site upon closure of loop 6. This, in conjunction with the action of another bulky non-polar residue, L230, located on the other side of the active site, forms a hydrophobic cage around the general base E165. Secondly, the main chain nitrogen atom of G171 makes a hydrogen bond interaction with the phosphodianion oxygen of the substrate, thus anchoring the substrate to the closed loop (Figure 11A).

*Figure 11. Structure of the active site in yeast TIM (PDB ID: 1NEY) in Michaelis complex with DHAP (green). Main chain atoms of the protein are shown in cartoon representation (blue). E165, G171 and S211 are shown in grey, whereas the hydrophobic residues I170 and L230 are shown in yellow. (A) Side view of the active site illustrating hydrogen bond interactions between substrate phosphodianion and main chain nitrogens of G171 and S211. (B) Top view of the active site illustrating the hydrophobic clamp formed by the sidechains of I170 and L230 around E165 and the substrate. Figure generated using PyMOL.*

Deprotonation of the carbon acid of substrates DHAP and GAP by a general base in the form of E165 is the rate-limiting step in TIM-catalyzed isomerization. The ability of E165 to abstract a proton from the substrate is influenced by structural changes around the active site of the enzyme. One of these is the closure of loop 6 that brings the bulky hydrophobic sidechain of I170 in close proximity to E165, desolvates the active site, and increases the basicity of
E165. In the catalytically competent conformation of the active site, I170 and L230 flank the general base and bound substrate on both sides, thus resulting in a hydrophobic clamp (Figure 11B).

The effect of this hydrophobic clamp on the rate-limiting deprotonation step was studied using EVB calculations by calculating the free energy barriers in wild-type TIM, as well as the mutants I170A, L230A, and the double mutant I170A/L230A. The extent of catalytic effect in the wild-type enzyme on substrate deprotonation was also calculated by comparing the $\Delta G^\ddagger$ and $\Delta G^0$ values with those for the uncatalyzed reaction in solution. Kinetic data available from biochemical experiments on these variants of $Tbb$TIM were used to validate the EVB model. The calculations successfully reproduced experimentally determined $\Delta G^\ddagger$ values for deprotonation of both DHAP and GAP in all four enzyme variants (Table 1).

Table 1. Activation ($\Delta G^\ddagger$) and Gibbs free energies ($\Delta G^0$) for TIM-catalyzed deprotonation of DHAP and GAP in wild-type (WT) and mutant variants of TIM. All energies are shown in kcal•mol$^{-1}$.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Catalyst</th>
<th>$\Delta G^\ddagger_{\text{calc}}$</th>
<th>$\Delta G^0_{\text{calc}}$</th>
<th>$\Delta G^\ddagger_{\text{exp}}$</th>
<th>$\Delta G^\ddagger_{\text{calc}}$ - $\Delta G^\ddagger_{\text{exp}}$</th>
<th># of waters$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHAP</td>
<td>CH$_3$CH$_2$CO$_2^-$ in water</td>
<td>25.2 ± 0.2</td>
<td>18.9 ± 0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>WT-TIM</td>
<td>14.5 ± 1.4</td>
<td>5.6 ± 1.8</td>
<td>14.1</td>
<td>0.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>I170A</td>
<td>16.3 ± 1.5</td>
<td>7.6 ± 1.4</td>
<td>15.8</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>L230A</td>
<td>16.7 ± 0.8</td>
<td>8.6 ± 0.8</td>
<td>16.6</td>
<td>0.1</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>I170A/L230A</td>
<td>18.5 ± 1.0</td>
<td>11.0 ± 1.3</td>
<td>17.4</td>
<td>1.1</td>
<td>2.6</td>
</tr>
<tr>
<td>GAP</td>
<td>CH$_3$CH$_2$CO$_2^-$ in water</td>
<td>24.1 ± 0.2</td>
<td>16.1 ± 0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>WT-TIM</td>
<td>12.9 ± 0.8</td>
<td>2.5 ± 0.9</td>
<td>12.9</td>
<td>0.0</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>I170A</td>
<td>16.2 ± 1.7</td>
<td>5.7 ± 1.9</td>
<td>16.0</td>
<td>0.2</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>L230A</td>
<td>14.9 ± 0.8</td>
<td>3.1 ± 1.0</td>
<td>14.2</td>
<td>0.7</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>I170A/L230A</td>
<td>16.5 ± 1.4</td>
<td>5.4 ± 1.8</td>
<td>16.3</td>
<td>0.2</td>
<td>3.0</td>
</tr>
</tbody>
</table>

$^a$ Activation barriers calculated from experimentally determined kinetic parameters$^{175,178,179}$ using Transition State Theory (TST). $^b$ Average number of water molecules within 4 Å of either of the E165 carboxylate oxygen atoms at the transition state.

The alkyl ammonium side chain of K12 (Figure 12) has a dominant contribution towards stabilization of the negative charge developed at the deprotonated carbon of the substrate. However, the electrostatic contribution from this residue and other key active site residues did not vary significantly between the enzyme variants to cause the observed changes in the free energies. Further analysis of data extracted from the EVB simulations revealed a direct correlation between calculated values of $\Delta G^\ddagger$ and the number of water molecules in
the vicinity of the carboxylate group of E165 across all TIM variants (Table 1). This finding reinforced our hypothesis regarding the role of the hydrophobic clamp to increase the basicity of E165. The residues I170 and L230 thus function as an important building block to facilitate optimal stabilizing interactions between the active site residues and the transition state.

5.2 Role of Key Active Site Residues in TIM (Paper IV)

To further our understanding of the influence of other key amino acid residues in the active site, the EVB setup validated in the previous study (Paper III)\textsuperscript{177} was subjected to model catalysis in some more mutants of TIM.\textsuperscript{174} K12, with its positively charged sidechain, is extremely important for electrostatic stabilization of the negative charge developing at the deprotonated carbon of the enediolate intermediate.\textsuperscript{180–182} It is strategically placed at such a location that it also offers stabilization to the negative charge of the phosphodianion moiety of the substrate. K12 is also ion-paired with a second-shell residue E97. This interaction is key to maintaining the correct spatial position of K12. In addition to these two residues, the role of P166 was also investigated (Figure 12). P166, located adjacent to the catalytic general base E165, plays an important role in positioning the side chain of E165 into the active site for performing the proton transfer steps. This involves a combination of conformational changes in loops 6 and 7 that trigger a steric clash between the side chain of P166 and the carbonyl oxygen of G209 on loop 7.

The effects of K12 and E97 on TIM-catalyzed deprotonation of DHAP and GAP were studied using the single mutants K12G, E97A, E97D, E97Q, and the double mutant K12G/E97A. These mutations were chosen based on existing experimental data which could be used to further validate the EVB model. The K12G substitution was found to significantly affect electrostatic stabilization of the transition state, which was reflected in the increases in calculated activation barriers for substrate deprotonation in the K12G mutants as compared to wild-type TIM (Figure 13).

The E97A, E97D, and E97Q substitutions had little to no direct influence on the stabilization of the transition state, and the double substitution of K12G/E97A had a similar effect as K12G on the calculated activation free energy. This confirmed that K12 contributes directly towards the stabilization
of the anionic enediolate intermediate, and E97 is involved in the preorganization of the K12 side chain, but does not influence transition state stabilization.

**Figure 12.** Structure of the active site in yeast TIM (PDB ID: 1NEY)\(^1\)\(^{13,100}\) in Michaelis complex with DHAP. Shown here are sidechains of key active-site residues: K12, E97, E165, P166, I170, and L230. Reprinted (adapted) with permission from Ref. \(^{174}\). Copyright 2019 American Chemical Society.

The P166A substitution provided an exciting opportunity to model the catalytic effect of this critical residue using EVB. Structural data of this mutant in *Tbb*TIM\(^{183}\) showed the E165 side chain to have attained a “swung out” conformation from its catalytically competent “swung in” conformation as observed in *y*TIM\(^{100}\) (Figure 14A). This feature was captured in the EVB calculations as well (Figure 14B), and the effect of perturbed donor-acceptor distances between E165 and the substrate on the deprotonation step was visible in terms of the 2.2 and 2.7 kcal\(\cdot\)mol\(^{-1}\) increases in the activation barriers for DHAP and GAP respectively (Figure 13).

In addition to this, the mutants studied in Paper III\(^{177}\) – I170A, L230A, and I170A/L230A – were also revisited and modeled using EVB. The calculated values of \(\Delta G^\ddagger\) for all enzyme variants examined in this study were in excellent agreement with experimental data (Figure 13). The highlight of this study was the extended Brønsted relationship for TIM-catalyzed substrate deprotonation that showed strong linear correlation between the computed values of activation free energies and thermodynamic free energies for wild-type TIM and nine other mutants (\(\beta = 0.73\) for DHAP, \(\beta = 0.74\) for GAP) (Figure 15).
Figure 13. Activation free energies for TIM-catalyzed deprotonation of (A) DHAP and (B) GAP, determined by EVB calculations (blue) and derived from experimental kinetic parameters (tan). Reprinted (adapted) with permission from Ref. 174. Copyright 2019 American Chemical Society.

Figure 14. Structure overlays of wild-type and P166A TIM illustrating the “swung in” and “swung out” conformations of the E165 0073idechain. (A) Overlay of X-ray structures of ligand-bound wild-type yeast TIM (PDB ID: 1NEY, dark green, E165 “swung in”) and ligand-bound P166A TbbTIM (PDB ID: 2J27, tan, E165 “swung out”). (B) Overlay of top-ranked cluster representative structures from EVB simulations at the Michaelis complex for wild-type (light green, E165 “swung in”) and P166A TIM (pink, E165 “swung out”). (C) Overlay of all structures shown in panels (A) and (B). Reprinted (adapted) with permission from Ref. 174. Copyright 2019 American Chemical Society.
This computational study thus underlined the contribution of non-polar residues (P166, I170, L230) towards the generation of a catalytic cage, and the role of electrostatic preorganization of polar active site side chains (K12) in transition state stabilization.

5.3 Effect of Dianion Binding on Catalysis (Paper V)

While the chemical step in TIM-catalyzed isomerization occurs at one end of the substrate molecule, the phosphodianion group at the other end does not actively participate in the chemical reaction. It is, however, essential for anchoring the substrate to the enzyme by making hydrogen bond interactions with residues on loops 6 and 7. The effect of substrate phosphodianion on catalysis in TIM has been experimentally investigated by using substrate fragments glycolaldehyde (GA) and phosphite dianion (HPi) (Figure 16).\textsuperscript{184,185} However, knowing its direct effect on the chemical step required a computational study in which the substrate deprotonation step could be modeled for GA with and without the presence of the HPi fragment.
Figure 16. Chemical structures of the substrate fragments glycolaldehyde (GA), phosphite dianion (HP\textsubscript{i}), and the full substrate GAP.

The EVB approach was employed to model the energetics for the deprotonation of substrate fragments GA and GA•HP\textsubscript{i} in wild-type TIM. The calculated activation barriers of 15.0 ± 2.4 and 15.5 ± 3.5 kcal\textperiodcentered mol\textsuperscript{-1} for GA and GA•HP\textsubscript{i} respectively showed a small increase of ≤2.6 kcal\textperiodcentered mol\textsuperscript{-1} in comparison to the corresponding value for the full substrate GAP\textsuperscript{113,172}, which was attributed to an entropic effect resulting from increased conformational flexibility of the substrate fragments in the active site of TIM. The increase in calculated Δ\textsubscript{G}\textsuperscript{‡} could not fully account for the intrinsic binding energy of the dianionic phosphoryl group. This study\textsuperscript{186} concluded that the dianion binding energy used to drive a conformational change in the enzyme is fully expressed at the Michaelis complex, thereby corroborating the role of ligand-gated conformational change in enzyme catalysis.

5.4 Dynamics of Loop 6 in TIM (Paper VI)

Closure of loop 6 is important for effective substrate binding and generation of a catalytically competent hydrophobic active site environment. Many earlier studies based on NMR data\textsuperscript{107,187,188} structural models of TIM in open-loop and closed-loop conformations\textsuperscript{97,100,105–107} and computational studies employing simplified models or insufficient simulation timescales\textsuperscript{189} had led to the description of loop 6 dynamics as a simple two-state rigid-body motion\textsuperscript{190}. On the contrary, conformational sampling from EVB calculations in Papers III to V\textsuperscript{174,177,186} revealed that loop 6 undergoes large conformational changes and takes on multiple different conformations in its open form. In order to understand the enigma of loop 6 dynamics and comprehensively characterize its conformational transitions, a detailed computational study was performed using long timescale conventional MD simulations and an assortment of enhanced sampling methods\textsuperscript{191}. 

54
Crystal structures of unliganded and ligand-bound TIM having loop 6 in open and closed states were chosen as starting points for conventional MD simulations for extensive conformational sampling of loop 6. In all cases, the loop opened up substantially within the initial 10-20 ns of simulation time as the process of loop closure was too slow to be captured within reasonable simulation timescales using unbiased MD simulations. Enhanced sampling approaches such as Hamiltonian Replica Exchange MD (HREX-MD) and Biased Exchange Metadynamics (BE-METAD) were employed to speed up the sampling rate and capture the closed state of loop 6. Altogether, it was observed that the conformational landscape for the “open” state of loop 6 was relatively much larger than that for the closed state (Figure 17). It was also observed that loop 6 existed in various open conformations which do not qualify as being catalytically competent. Markov state models (MSMs), constructed using data obtained from conventional MD simulations, revealed swift transitions of loop 6 between various open conformations.

Figure 17. Principal component analysis of conventional MD simulations on (A) apo and (B) DHAP-bound structures of wild-type TIM. The principal components are based on the distances between Cα atoms of loops 6 and 7. Shown here are free energy surfaces indicating the conformational space explored by these loops. PC1 effectively describes the opening and closing motion of loop 6. The open and closed states of the loops derived from the respective crystal structures for apo-TIM (PDB ID: 1YPI,13,106 ▼) and DHAP-TIM (PDB ID: 1NEY,13,100 ▲) are projected here. Reprinted (adapted) with permission from Ref. 191. Copyright 2018 American Chemical Society.

The link between loop dynamics and catalysis in TIM was also explored by performing EVB calculations to model the deprotonation of DHAP at various conformational states of loop 6. A significant falloff in catalysis was observed when loop 6 conformations deviated from the catalytically competent closed
state. These open-loop conformations sampled from BE-METAD had the substrate and key active site sidechains such as I170 displaced from their original positions, resulting in disruption of key enzyme-substrate interactions and flooding of the active site with water molecules. This study showed with conclusive evidence for the first time that the motion of loop 6 in TIM is highly complex and dynamic, and is not a simple two-state rigid body motion as previously assumed in literature. Closure of TIM loop 6 into a catalytically competent conformation was observed to be a rare event at simulation timescales, with its odds only slightly improving in the presence of a substrate or ligand. Simulating the dynamics of flexible enzyme loops is therefore challenging, and demands strategic use of computational approaches and resources.

5.5 Perspectives

Understanding the origins of enzyme catalysis has been the holy grail of biochemical investigations ever since Fischer proposed the “Lock and Key” hypothesis. Improvements in experimental techniques over the past century have guided significant advancements in elucidation of catalytic mechanisms in enzymes. The past few decades have witnessed computational approaches gain traction in the field of enzymology, targeted at characterizing the role of dynamics in enzyme catalysis. The work presented in this chapter is built on the foundations laid by pioneering experimental, structural, and computational studies on TIM as a model enzyme. The enzyme uses preorganized active site residues for transition state stabilization, but also utilizes the motion of a highly flexible loop to prepare the active site for efficient catalysis. Each of these aspects – the hydrophobic clamp generated by loop 6, the preorganization of key active site residues, the role of substrate phosphodianion in catalysis, and the nature of loop motion in TIM – have been examined using computational approaches. This work has complemented experimental data and provided newer insights into a long-studied system, thus underscoring the importance of collaborative research in science of the 21st Century.
6. Conclusions and Future Perspectives

Proteins exist in diverse structural forms and show tremendous functional versatility by performing a variety of life-sustaining biological processes with utmost efficiency. Biochemical and structural studies have provided valuable insights towards determining the link between protein structure and function. More recently, developments in computational infrastructure have catalyzed the implementation of theoretical models for accurate description of molecular systems. Collaboration among the fields of experimental biochemistry, biophysics, structural biology, and computational biochemistry have been instrumental in characterizing structure-function relationship in proteins, as exemplified by the work showcased in this thesis.

Knowledge of the three-dimensional structure of a protein is crucial in determining its biological effect. Computational investigations included in this thesis have been targeted at modeling the dynamics of two protein systems – IDPs and enzymes – that lie on opposite ends of the flexibility spectrum. The Aβ peptide, an IDP associated with the progression of Alzheimer’s disease, was studied for global structural perturbations resulting from its interactions with biologically-relevant partner molecules such as metal ions and amphiphilic surfactants. Complemented by data from biophysical experiments, these studies were able to definitively characterize the binding interactions.

Triosephosphate isomerase (TIM), a catalytically proficient enzyme existing in a well-defined structural fold, was the subject of the second set of studies aimed at uncovering the role of key active site features in enzyme catalysis. Despite being a globally ordered structure, TIM consists of a local disordered region in the form of the highly flexible loop 6. As described in Chapter 5, the disorderliness of this loop is key to generating a catalytically competent active site. Preorganization of active site sidechains is an important function of the structural fold. Thus, it was observed that both structural order and disorder play equally important roles in proper functioning of a protein. These studies provided deeper insights into enzyme catalysis and enriched our understanding of chemical reaction mechanisms in highly evolved enzymes.
Simulations of biomolecular systems function as a computational microscope for the study of their dynamic behavior and the resultant effect on biological function. A limiting factor, however, is the inability of current methods to sample events at biologically relevant timescales. Advancements in computer hardware and algorithms have shown encouraging signs of moving towards the ultimate objective of computational modeling. The growing use of machine learning methods and artificial intelligence augurs well for the field in terms of method development and derivation of biologically relevant trends from large amounts of simulation data. The synergistic use of experimental methods and computational modeling can potentially lead to a profound understanding of the structure-function relationship in biomolecules. It can also pave ways for applied scientific endeavors such as enzyme design, and development of therapeutics targeting IDPs or structured proteins.

Min forskning fokuseras på användandet av beräkningsmetoder till modellering av dynamiken i proteiner för att se effekten den har på dess funktion. Ordning och oordning i en proteinstruktur är avgörande för att bestämma dess funktion. I denna forskning valdes därför två klasser av proteiner tillhörande motsatta ändar av flexibilitetspektrumet att studeras – ”intrinsically disordered proteins” (IDP) och enzymer.

IDP:er är funktionellt viktiga peptider som har möjlighet att genomgå globala strukturella förändringar för att interagera med olika målmolekyler. Denna inre oordning kan också leda till felaktig veckning och efterföljande aggregering av peptiden vilken är förknippad med patogenesen för framträdande neurodegenerativa sjukdomar. Amyloid beta (Aβ) peptiden är en IDP som är associerad med utvecklingen av Alzheimers sjukdom. För att studera de strukturella förändringarna i denna peptid orsakad av dess interaktion med metalljoner (artikel I) och amfifila ytaktiva medel (artikel II) användes datorsimuleringar.

Enzymer är biologiska katalysatorer som påskyndar hastigheterna för biologiskt relevanta kemiska reaktioner med många storleksordningar. Enzymer består av välordnade strukturella veckningar och är selektiva för att katalysera
en specifik kemisk reaktion. Kombinationen av ett enzyms effektivitet och specificitet, vid sidan av dess biologiskt nedbrytbara natur, gör dem till ideala kandidater för tillämpning i industriin inte minst då hållbarhet idag är av stor vikt. Design och anpassning av nya enzymer kan vara möjlig endast om vi har en djup förståelse för ursprunget till enzymkatalys.

Den andra delen av denna avhandling beskriver en samling beräkningsstudier utförda för att modellera och studera katalys i enzymet triosefosfatisomeras (TIM). TIM är ett mycket utvecklat, katalytiskt skickligt enzym som verkar i glykolysen. Beräkningsmodellering ger en spännande möjlighet att ta del av molekylära detaljer i enzymkatalys i TIM. Enzymet använder en kombination av förorganiserade aktiva platsrester och rörelsen av en mycket flexibel loop (loop 6) för att förbereda det aktiva sätet för effektiv katalys. Var och en av dessa aspekter - den hydrofoba klämman som genereras av loop 6 (artikel III), förorganisationen av resterande viktiga aktiva platser (artikel IV), rollen för substratfosfodianion i katalys (artikel V) och arten av loop-rörelse i TIM (artikel VI) - har studerats. Detta arbete illustrerade den funktionella betydelsen av strukturell stabilitet och dynamik i ett enzym, vilket ger nya insikter om katalys i ett modellenzymsystem.

Forskningen som presenteras i denna avhandling exemplifierar ett givande samarbete mellan områdena experimentell biokemi, biofysik, strukturell biologi och beräkningsbiokemi. Förutom att förstå de strukturella funktionerna hos den medicinskt relevanta Aβ peptiden, har beräkningsmodelleringen av kemi i TIM givit nya insikter för att förstå ursprunget till enzymkatalys. Dessa undersökningar har därför bidragit till att karakterisera struktur-funktionsförhållanden i proteiner och därmed förbättra vår vetenskapliga förståelse av proteiner på molekylär nivå.
Acknowledgments

The work presented in this thesis has resulted from four years of my PhD training. During this period, I have interacted professionally and personally with many people who have been of immense help and support to me. I would like to thank them (as many as I can) here.

First of all, I would like to thank my supervisor, Prof. Lynn Kamerlin. Thank you for accepting me as a Masters student in the group and supporting me to carry out a PhD under your supervision. Thank you for your constant support and guidance over the past five years, and letting me attend many interesting conferences worldwide! The experience at Kamerlin lab has been memorable and worth cherishing, and I have learnt about many things both within and outside the scientific domain. A big thank you once again!

I would like to thank my co-supervisor, Prof. Sherry Mowbray, for her support, guidance, and giving me fond memories of the Protein Engineering course, both as a student and as a lab teacher.

I have had the good fortune of working with extremely knowledgeable and collegial collaborators. I thank Prof. Astrid Gräslund for her support as a co-supervisor, and for all the help and guidance with the Aβ project. Thank you for also letting me do experiments in your lab. Sebastian, thank you for teaching me experiments (especially NMR) at DBB. Cecilia and Nicklas, thank you for your help in the lab and interesting discussions. I thank Prof. John Richard for a successful collaboration on the TIM project. Thank you also for your insights on enzyme catalysis in TIM. I have learnt a lot from our scientific and non-scientific conversations. I thank Prof. Dr. Birgit Strodel for the collaboration on both Aβ and TIM projects, and for helpful discussions. I would like to thank all the co-authors on each of the publications listed in this thesis. Thank you Prof. Erik Johansson and Vimal for giving me an opportunity to work on the DNA Polymerase project.
A big thank you to all former and current members of the Kamerlin lab. I owe a lot to my seniors in the group who patiently taught me how to run simulations and helped me in my projects. Fernanda, you were among the first people I interacted with in the group. Thank you for helping me understand what I was doing during my initial phase in the group. Alex, yo, thanks for all the fun and sorry I did not join the beer clubs that often! Beat, you were my first supervisor in the group. Thank you for letting me work in your projects and teaching me how to work in Linux. Also, thanks for testing CADEE on TIM! Paul, thank you for being there whenever I needed help with parametrization or troubleshooting my simulations. David, I had a great time working with you on the PON1 project. Miha, thanks a ton for making it much easier for me to work with EVB. It was a pleasure interacting with you whenever we did. Fabian, thank you for the “pretty-picture” settings for PyMOL. Dennis, thanks for your help with the micelle models. Anna, thanks for your feedback on various projects during the group meetings. Klaudia and Irek, thank you for organizing the board game nights at your place. Malin, thank you for helping me out with stuff whenever I needed. Andrey, unfortunately we could not interact much, but I wish you good luck for your PhD. Qinghua, thank you for all your help during the last three years. You were my first point of contact if I was stuck with any problem with the simulations. It was a pleasure sharing the office with you. I will miss the badminton and table tennis games. I wish you the best for the future. Dušan, thank you for contributing to the TIM projects, and for all the fun times. Anil, thank you so much for sharing my joys and sorrows, for the insightful scientific discussions, and for taking over the GPDH project. I will especially miss our discussions on cricket. Adrian, thanks for lightening my mood with all the awesome fun, and for always being ready to go for a coffee whenever we asked you to join. Good luck with your projects! Marina, Jasmine, Michal, I am sorry we could not interact very often. Thank you for the discussions during the group meetings, fikas, and the cruise trip. I wish all the very best to you. Sebastian, thank you for sharing your knowledge on ASR. Cátia and Rita, it was nice interacting with you during fikas. All the best for your GTPase project. Thank you Bruno, Lucas, and Neli for all the interactions. Good Sir Rory, thank you for all the “sorries”! It was fun having you here over the spring. I wish you good luck with the projects you will work on after joining the group.

I want to thank all the project students who worked alongside me on my various projects. It was a pleasure to supervise all of them. Agata, thank you for staying motivated towards the project and running the Aβ simulations. I am
happy to see you doing well, and wish you the best for your PhD. **Fabian,** thank you for your help with testing force fields on the TIM system. Good luck for your PhD. **Nathalie,** thank you for testing the various TIM mutants. I would like to thank all my colleagues at **ICM** for the helpful discussions and lovely interactions. I thank **Prof. Mikael Widersten** and **Prof. Helena Danielson** for their scientific and non-scientific help and support at **Kemi-BMC.** I would also like to thank my colleagues here whom I interacted with frequently. **Thilak, Eldar, Leandro,** and **Caroline,** thank you for your company in the gym and for all the interesting discussions. **Joana,** thank you for your help with the course lab. **Gina,** thank you so much for your help in the wet lab. I wish you the best for your PhD.

My research work was funded by a fellowship award from the **Sven and Lilly Lawski Foundation,** as well as by the **Swedish Research Council (VR).** I thank both the funding agencies for their generous support. I also thank the **Swedish National Infrastructure for Computing (SNIC)** for the computational resources.

My life in Uppsala would have been incomplete without my friends outside work. I huge thank you to everyone whom I interacted with over the course of the past six years. Please forgive me if I miss mentioning your name. Firstly, I thank my closest friends who were like a family to me here, and made me feel like at home away from home. **Sandesh,** I don’t know if I can thank you enough. You were my first “buddy” when I came to Uppsala and helped me with every little thing. Thank you for letting me share your apartment, for treating me with delicious meals, for teaching me how to cook, and for all the fun. The list is endless, and I am happy to have shared these years with you. All the best with finishing your PhD. **Phani,** thank you for enriching my life with your “thug”ness and always motivating me to publish papers. Please do not give up on the “My Group” idea! Good luck with your PhD. **Punya** and **Smruti,** thank you for all the awesome get-togethers and for giving a homely feeling whenever I visited you. **Malavika,** thank you for maintaining the lunch group even though it reduced to just three people. Hope you finish your PhD soon! Just do it! **Narayan,** thank you for all the questions and cinnamonos. Good luck for the future. **Vivek,** thank you for being an awesome gym partner and for helping me carry things when I moved to the new apartment. That makes me thank **Hemanth** as well, for all the help with everything and always being available. **Oksana,** it was nice interacting with you over lunches and fikas while you were here. Good luck for your PhD. **Sangeetha** (aka bouncer),
thank you for all the fun times, bringing delicious food to 83B, and for accompanying me to the gym. Wish you the best for your PhD as well. Pruthvi, thank you for all our interactions, even though we did not meet that often during our PhD days. All the best for your future endeavors. Anitha, thank you for all the discussions on music and cricket, and for teaching me how to make a pizza from scratch. Loora, thank you for visiting India with Madle. I hope you travel there once again. Cecilia, ett stort tack till dig for helping me with the Swedish sammanfattning! Annika, tack så mycket for your help at such short notice! Ana, Sudarsan and Kristine, thank you for your company whenever you visited our apartment. I also thank my friends from the cricket group for all the awesome games we played together. Those were some of my most favorite days here. Thank you, Srinivas, Sethu, Hari, and Prasad. I thank Prof. Suparna Sanyal and Prof. Biplab Sanyal for making me a part of Uppsala Indian Choir, and for all the lovely events and melodious evenings spent at Sanyal Bari. Thank you to all other members of UIC too.

I thank all my friends from RVCE, who stayed in contact with me for all these years. Raghu, thank you so much for coming to Scandinavia time and again and taking me on some of the most memorable road trips and hikes of my life. Thank you also for hosting me in Houston and for the amazing week with Sambit and Ashlesh. I shall always remember that trip. Srinidhi, thank you for hosting me in Oslo and all the fun times during all the trips. I also thank all my friends in India who gave me so much love every time I made a trip back home. It really meant a lot. Special thanks to Mr. Shanti Agarwal for his support and encouragement. I thank Mr. Sanjay, Ms. Regina, and everyone else at Ovobel for all their help with various things.

I would like to thank all my teachers, from my school days, to the Bachelors, Masters and PhD programs, who have played an important role in my scholarly development. Thank you!

Finally, I would like to thank my family. Aai and Baba, thank you for everything. This would not have been possible without your love and support over all these years. Thank you Mummy and Papa for your encouragement. I thank my extended family as well. In the end, I would like to thank my best friend, who also happens to be my dear wife, Divya, for all the love, for being a strong pillar of support, and for standing by me during the good and bad times. Thank you for inspiring me to do science the right way, and for teaching me to never give up in life. Best wishes for your PhD!
References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1885

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)