Electric Fields: A Metric for Molecular-level Understanding of Protein Mechanisms

Yi Zheng

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Valerie Vaissier Welborn, Chair
T. Daniel Crawford
Nicholas J. Mayhall
James M. Tanko

Apr 24, 2024
Blacksburg, Virginia

Keywords: protein mechanism, electric field, molecular dynamics simulations.

Electric Fields: A Metric for Molecular-level Understanding of Protein Mechanisms

Yi Zheng

(ABSTRACT)

Determining the molecular mechanisms at the origin of protein function remains a challenge due to the complex non-covalent interactions that shape their structure. Since the non-covalent interactions arise from charge fluctuations, electric fields can be used as a tool to quantify the interactions between a target and its environment. The contribution of each component of the system is reflected in the direction and strength of the electric field exerted on the target, which can be calculated from molecular dynamics simulations.

The interactions experienced by ligands in enzymatic active sites determine the catalytic activity of the enzyme. Ligands in synthetic enzymes lack interactions with the protein scaffold, which limit their efficiency. To substitute for the role of non-effective protein scaffold, we introduced a polar DNA fragment to the enzyme vicinity, inducing electrostatic interactions that will facilitate the reaction. We found that the introduction of a DNA fragment enhanced the original interactions between the residues in the active site and the ligand, without creating new interaction hot spots. Using electric fields, we calculated a reduction in activation energy of 2.0 kcal/mol when introducing the DNA fragment, indicating a promising avenue for catalytic improvement.

Inspired by the success in using electric fields to understand enzyme catalysis in the context of electrostatic preorganization theory, we generalized these fundamental concepts to another type of proteins: voltage-gated ion channels. Our results indicate that electric fields also report on channel activity. We find an asymmetry in the number of active residues for

channel function between the four domains and between the two gating motifs of the permeation pathway, with domain I being the major contributor in both cases. The importance of residues for channel activity is not a simple linear correlation of their distance with the functional motif, but a relationship dominated by non-covalent interactions.

Finally, we investigate the effects of loop dynamics on enzyme product inhibition. We modify the chemical nature of the unstructured loops that obstruct the active site of DszB by glycosylating serine and threonine residues. We monitor the corresponding variations in loop dynamics and their effect on the interaction between the enzyme and the product.

Overall, promising results were found using electric fields in the investigation of protein mechanisms that are mainly dominated by non-covalent interactions and provide insight into the role of the individual components in the system.

Electric Fields: A Metric for Molecular-level Understanding of Protein Mechanisms

Yi Zheng

(GENERAL AUDIENCE ABSTRACT)

Although weaker than covalent interactions, non-covalent interactions play a crucial role in molecular biological processes, especially in protein mechanisms. In order to modify the properties of proteins to our advantage, we need a metric with which we can map these interactions onto the protein structure. Different types of non-covalent interactions share one similarity: they originate from the change of electron distribution of interacting atoms, therefore can be captured by analyzing the protein- generated electric fields.

Synthetic enzymes are designed to better adapt to varying environments and catalyze a broader reaction range. However, they are less effective than natural enzymes because the protein scaffold does not contribute to catalysis. Indeed, protein scaffolds in natural enzymes generate an electric field that lowers the reaction activation energy in the active site. Protein scaffolds in synthetic enzyme do not generate such electric fields. To address this issue, we modified the environment of synthetic enzyme KE15, introducing a polar DNA fragment to induce interactions in the active site. This modification strengthen the interactions between protein and ligand, leading to a decrease in the energy required for the reaction.

While enzymes are famous for their generation of electric fields facilitating function, we demonstrated that this phenomenon also exist in voltage-gated ion channels $Na_v1.7$. Residues were found to exert an electric field that can facilitate ion permeation. This is not simply because of their distance to the key regions, but a result of the non-covalent interactions regulating the mechanism, with different regions showing asymmetric importance in the process.

Since the governing non-covalent interactions are relatively weak, proteins are flexible, especially protein loops. In enzyme DszB, this loop flexibility enables a conformational change when the ligand binds the active site. The change in loop conformation traps the product inside the active site, limiting enzymatic turnover. To prevent active site obstruction by these flexible loops, we attached glucose to a few loop residues to modify the hydrophobicity profile near the active site. The introduction of hydrophilic glucoses helps to pull the loops towards the solvent, rather than towards the active site, limiting product inhibition while preserving catalytic activity.

Overall, our results show that electric field can be applied as a general method for protein studies, relating structure to function.

Dedication

To my eight-year-old self, she was curious, brave, and stubborn

Acknowledgments

The time was late autumn of 2019 when I sat beside the window thinking about my future. When the excitement of starting graduate school wears off, I started panicking about not fitting as a graduate student. I haven't joined any research group at the time, so I told myself I should at least try to see what researches are like before I make any decisions. This is the thought that brings me to the where I am today.

First and foremost, I would like to thank my advisor Prof. Welborn. It was the research with her that colored the gloomy first year and the years after. Without her patient and guidance I would not be able to go this far.

My profound gratitude also goes to my family without whose support none of these would have happened. To my parents Yan Jia and Bizhou Zheng, for all the calls and mails of tea cans in the five years I cannot go home, and pictures of delicious home-made food I can only taste through a phone screen. For telling me I can be whoever I want to be. To my brother Zi Zheng, for the questions and talks, the courage you showed me possible, the matching pajamas and plushy, and the Lego we built separately but together. To my grandparents Shuzhen Qu and Chunhua Jia, for all the face-timing from the bus station and breakfast store, and the quiet companions when I am writing. To my late grandma Songzhen Huang, for defending me when I was a little girl.

To my dearest friends Yuting(Bones) Cao and Hejiazi(Lisa) Ding. For all the thoughts, stories, ideas, and jokes (especially the terrible ones) Bones and I shared. For always be there for me early in the morning and late at night despite the time zone separating us, and the never repeated goodnight quotes Bones has. For the trips Lisa and I took together, the lakes we saw, the mountains we climbed, and the incredibly unfortunate few days of

Portland's winter that makes me laugh whenever I look back.

To the Welborn group, and everyone in the office, for tolerating my random questions, and the conversations during coffee break (when coffee machine was still functioning).

Also to a few people whose life have never cross path with mine, but still left profound impact:

To Anton Chekhov and Fyodor Dostoevsky. The literature I grew up reading show me worlds I have not experienced, inspire me to think and question, give me strength, hold my hand through the seemingly darkest time before dawn.

To Bach, Hilary Hahn, and the band Kino. Bach's Chaconne has the power to calm me down no matter the circumstances, and Hahn played my favourite recording. Musics I listened to the most when working are from Kino.

To David E. Newton. For writing a book I picked up at the age of eight, and decided to never stop on the road of science.

Allow me to end with a quote from Chekhov that touch my life:

Lice consume grass, rust consumes iron, and lying the soul.

Contents

List of Figures

List of Tables

List of Abbreviations

C_α The carbon in amino acids connecting the carboxyl and amino group

AMOEBA Atomic Multipole Optimized Energetics for Biomolecular Applications

CNS Central nervous system

DBT dibenzothiophene

DEKA The residues Asp, Glu, Lys, Ala forming the selectivity filter

DszB 2'-Hydroxybiphenyl-2-sulfinate desulfinase

EEID The outer ring structure above the selectivity filter

GC Gating charge

HBP 2-hydroxybiphenyl

HBPS 2'-hydroxybiphenyl- 2-sulfinic acid

IG Intracellular gate

KE15 Kemp Eliminase 15

MD Molecular dynamics

$Na_v1.7$ Human voltage-gated sodium ion channel 1.7

$Na_v s$ Voltage-gated sodium ion channels

PD Pore domain

PNS Peripheral nervous

RMSD Root mean square standard deviation

SF Selectivity filter

VSDs Voltage-sensing domains

VSE Vibrational Stark effect

Chapter 1

Introduction

1.1 Non-covalent Interactions In Proteins

Non-covalent interactions are ubiquitous in biological systems, making the basis for protein folding, enzyme catalysis, molecular recognition and other protein activities.[1–4] Interest in non-covalent interactions came to its blossom in the 20^{th} century, but only in recent decades have we begun to truly uncover their role in chemistry and biology.[1,5,6] Indeed, protein properties and function depend on their structure, which is determined by non-covalent interactions. Therefore, we need a model for the systematic measurement of non-covalent interactions to gain insight into protein mechanisms.

The beginnings of modern chemistry were based on the characterization of covalent interactions. Despite the many successes in this field, covalent bonds alone are not sufficient to paint a complete picture of the chemistry of a real system in which the molecule of interest is interacting with its environment. Compared to covalent interactions, non-covalent interactions are weaker and lead to the formation of molecular clusters, but not to new molecules.[1,7] Although non-covalent interactions are individually weak, there are usually numerous and their collective effect can be significant. Indeed, non-covalent interactions give proteins their structural flexibility and the ability to change their conformation between different states, a process crucial to protein function.

There are different types of non-covalent interactions, which usually include van der Waals, hydrogen bonds, electrostatic and hydrophobic interactions (Figure 1.1).[8,9]

Van der Waals interactions occur when two atoms or molecules are brought close together, resulting in charge fluctuations that create instantaneous dipole moments.[10–12] This interaction is negligible at infinite distance. When atoms are brought closer together, the interaction energy decrease (attractive interaction) until reaching minimum energy, then increase dramatically (repulsive interaction) as the interatomic distance continues to decrease.[13] The attractive component of van der Waals forces is due to dispersion interactions.[14] The temporary dipole raised from the asymmetric distribution of charges are able to perturb the charge distribution of another atom or molecule, inducing a dipole that creates this attractive interaction. The shorter-range, repulsive component of van der Waals forces is due to the repulsion of the electronic clouds (Pauli's principle) on the respective atoms.[15] When atoms approach each other, electrons are prohibited from occupying the same state, resulting in strong repulsive interactions. The combination of these attractive and repulsive forces give rise to the form of van der Waal interactions. This universal interaction is generally considered to be significant only between non-polar groups. Since almost half of the amino acid side chains are non-polar, van der Waals interactions are key to protein structure and function. Further, in allosteric proteins (i.e. proteins that change shape and binding ability upon binding with ligands), van der Waals interactions are thought to be the driving force for their allosteric coupling.[16,17] They also contribute significantly to the stabilization of proteins complexes and small molecules.[18,19]

Hydrogen bonds are interactions between two atoms, one of which is a proton donor and the other a proton acceptor.[20] Hydrogen bonds are usually written as D-H$\cdots$A, where D denotes the proton donor and A is the acceptor. The formation and distance of this donor-acceptor pairing depends on the electronegativity difference between the D-H group, but usually result

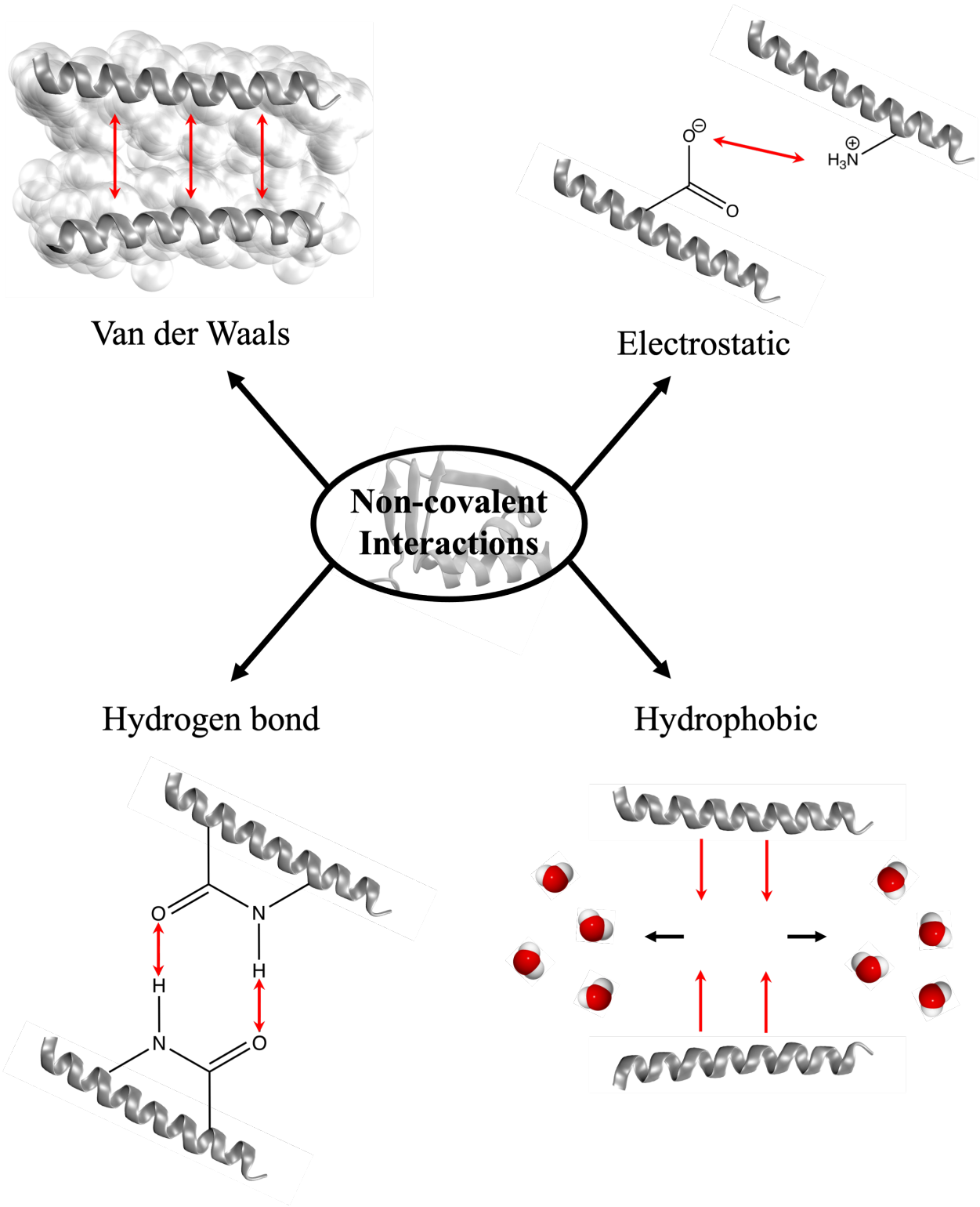

Figure 1.1: Non-covalent interactions is fundamental for protein structural stability and functionality.

in an interaction distance that is smaller than that of van der Waals interactions.[21–23] There are many hydrogen bond donors and acceptors that exist in proteins.[24] The most basic of these are the amino and carbonyl groups that give proteins their secondary structure.[25,26] Hydrogen bonds are not limited to the protein backbone; they also play a role in side-chain interactions for protein stability.[27,28] Finally, proteins are usually present in an aqueous environment, where water plays a big role in stabilizing their structure.[27] This is because water molecules can be proton donors and acceptors, increasing the number of potential hydrogen bonds it can form with amino acids.[29]

Compared to the previous two types of non-covalent interactions, electrostatic interactions are longer-range and can be strong beyond closest neighbors.[30] Electrostatic interactions arise from multipoles that are separated by a distance. Permanent multipoles result from the asymmetric distribution of electron clouds in molecules where the electron density is pulled in the direction of electronegative atoms.[13] Additionally, atoms and molecules can be perturbed by electric fields in the environment, resulting in an induced asymmetric charge distribution and the formation of induced dipoles. A large part of the environment of proteins consists of polar molecules, such as polar residues, water molecules, nucleic acid and phospholipids. Therefore, permanent and induced electrostatic interactions play a significant role in protein structure and function. For example, electrostatic interactions are often the driving force in the protein-ligand recognition processes,[31,32] where the binding infinity correlates with the strength of the electrostatic interactions.[33,34] Studies on membrane proteins have also shown that electrostatic interactions play an important role in their function, facilitating transport of charged ligands across the membrane.[35–37]

Non-polar compounds form aggregates in aqueous environments to avoid contact with water, a phenomenon known as the hydrophobic effect.[38] However, the attractive interaction between non-polar molecules is not the result of a real force, but of the rearrangement of

water molecules.[38–40] As already mentioned, there are almost as many non-polar amino acid side chains as polar ones, hence the importance of hydrophobic interactions for the stability and functionality of proteins in water. Structurally, hydrophobic amino acids have the tendency to aggregate, forming a hydrophobic core in water-soluble proteins.[4] In these type of proteins, the hydrophobic interactions can be the driven force in protein folding.[23,41] Functionally, it facilitates biological processes such as protein-ligand interaction or gating in channel proteins.[42,43]

Because of the remarkably diverse role that proteins play in all living organisms, they are at the center of our efforts to understand biological and chemical processes that are essential for life. Knowledge of the functional properties and mechanisms of proteins has a wide range of applications, from drug development to the catalysis of industrial reactions. The goal of tuning the properties of proteins to our advantage is not realistic without a model to predict their structure and function at the molecular level. However, the complexity of the interactions that regulate the system poses a challenge to the experimental or computational interpretation of protein structure and mechanism through each individual interaction.[5,38] A metric is required to provide detailed measurements of the strength of these interactions. Although they fall into several categories, all non-covalent interactions share a common property: they arise from the fluctuation of multipoles between interacting atoms. These multipoles are modulated by the electric field generated by the protein environment, which can be calculated or measured.[1,38,44] Analyzing the contribution of individual residues to the protein electric fields offers a new and physically-grounded approach to protein modeling.

1.2 Protein structure

Only two amino acids, asparagine and cysteine, had been discovered when Mulder presented the first study on the structure of proteins.[45] Today we know the complete list of the 20 standard amino acids that are the building blocks of proteins.[4,10,40]

The primary structure (Figure 1.2A) of a protein refers to the amino acid sequence of a polypeptide chain. The common structure of amino acids consists of a central carbon (C_α) to which an amino group (NH_2) and a carboxyl group (COOH) are attached. In a polypeptide chain, peptide bonds are formed between the amino and carboxyl groups of the neighboring amino acids, eliminating one water molecule each time. The protein backbone describes these linked C_α, N, C, and O. The side-chain, starting from the C_α, is unique for each amino acid.[46] This structure is shown in Figure 1.3, where the side-chain, denoted as R, connects to the C_α. Based on the side-chain properties, amino acids can be divided into two groups: 11 non-polar and 9 polar amino acids. Under physiological pH (pH=7.4), some side chains are charged: positively for arginine (Arg), lysine (Lys), histidine (His), and negatively for aspartic acid (Asp), glutamic acid (Glu). Amino acids interact with each other to form the protein secondary structure (Figure 1.2B).

The two most common elements of protein secondary structures, the α- helix (Figure 1.3A) and the β-sheet (Figure 1.3B), were first modeled by Linus Pauling and Robert Corey.[25,26] It was assumed that hydrogen bonds and electrostatic interactions determine the formation of the secondary structure and thus the structural stability of the protein.[23,25,26,30] Other patterns can be found in proteins, albeit much less frequently, including the 3_{10}-helix or the π-helix that form more tightly or more loosely packed turns.[46] Between the secondary structure elements are the patternless, flexible loop regions of a protein, which often contribute to the active site.[47,48] Some of the combinations of secondary structure elements in a

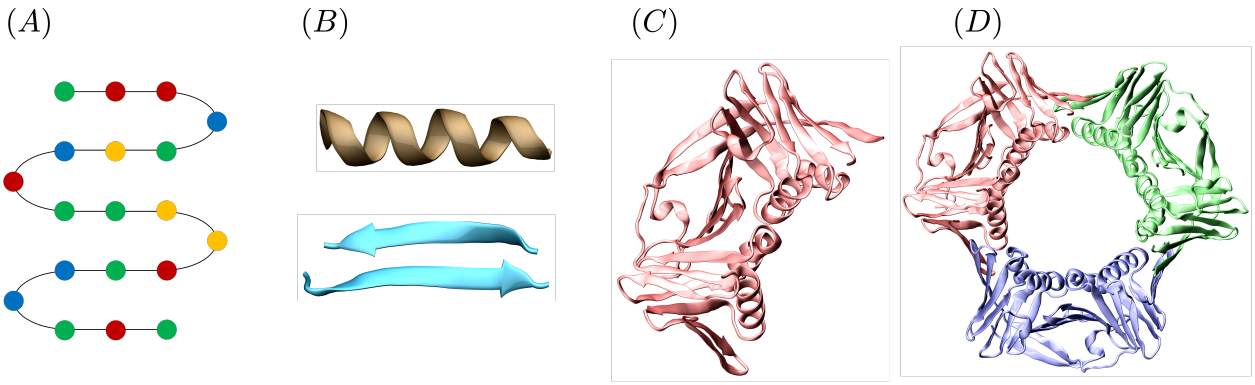

Figure 1.2: Level of protein organization, (A) to (D) represent the primary, secondary, tertiary and quaternary structure respectively. Each circle in (A) represents an amino acid. (B) to (D) are modeled from PDB: 1AXC. The upper coil in (B) is a typical α-helix, the lower structure is a β-sheet. (C) is one of the domains of the protein, and (D) shows the quaternary structure of 1AXC, which consists of 3 domains.

particular spatial arrangement indicate specific protein functions, but predicting the activity of a protein based on its secondary structure is still a challenge.[49,50]

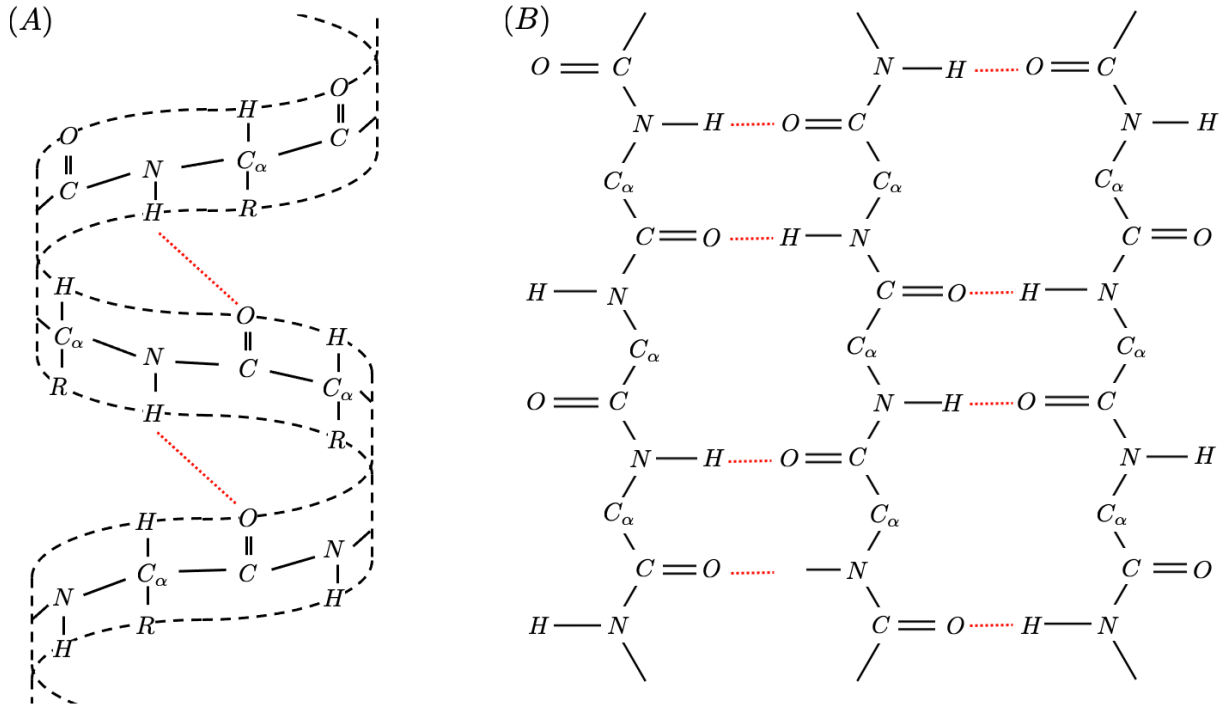

Figure 1.3: Illustration of α-helix (A) and β-sheet (B). The hydrogen bonds are shown as red dotted lines. R represents the side-chain of amino acids.

The three-dimensional packing of protein structural elements is called the tertiary structure (Figure 1.2C). This is the folding of the protein that leads to the basic unit of protein function: domains.[51] Examples of these functional units are the active sites of enzymes or other binding sites for small molecules, which are generally conserved throughout evolution.[51,52] Finally, the quaternary structure (Figure 1.2D) of proteins refers to the spatial arrangement of multiple polypeptide chains, each consisting of primary, secondary and tertiary structures.[53] Each chain, called a subunit, can either function individually or work together in a biological process.[46] Compared to the secondary structure, the formation of the tertiary and quaternary structure depends on the properties of the side chains of the amino acids.[51]

The binding of molecules is often necessary for proteins to fulfill their function. This includes the binding of ligands from small molecules (i.e. ions, metals, oxygen), lipid membranes, nucleic acids or even other proteins.[54] Enzymes are proteins whose function is to catalyze reactions. The reaction rate in the active site of the enzyme is significantly higher than the normal rate. A notable case of protein catalysis is carbonic anhydrase, which hydrates carbon dioxide at a rate of 10^6 molecules per second.[55] This strong catalytic ability makes enzyme desirable for broader applications, to catalyze reactions not found in nature. In natural enzymes, the whole protein (i.e., active site and scaffold) contribute to the reduction in free energy barrier, leading to an acceleration of the reaction. This reduction in energy barrier predominantly comes from a stabilization of the transition state, compared to the reactant state. Our ability to control structure-function relationship in enzymes will enable us to adapt to industrial applications. Another class of proteins functions as a transporter for ions and small molecules. Membrane transport proteins can be divided into three categories based on their permeation mechanism: pumps, carriers and channels.[55] Although all facilitate the transport of ligands across the membrane, pumps require energy input (ATP or light absorption), while carriers and channels enable passive transport. These proteins

are essential for cellular functions and therefore play a key role in our understanding of life.

1.3 Vibrational Stark Effect

Although non-covalent interactions are generally characterized into several types, they are electrostatic in nature as they arise from interactions between charged species. The interactions from each component of the system can be quantified from their contribution to the electric field at that location.[44] Experimentally, we can measure electric fields via vibrational Stark spectroscopy experiments.

The vibrational Stark effect (VSE) describes the perturbation of a probe's vibrational frequency when an external electric field is applied. The energy (U) of a dipole ($\vec{\mu}$) in an external electric field ($\vec{E}_{ext}$) is quantified as:[56]

$$U = \vec{\mu} \cdot \vec{E}_{ext}. \tag{1.1}$$

Note that we use here the chemistry convention where the dipole is oriented from less electronegative atoms (positive) to more electronegative atoms (negative). The physics convention is the opposite, where the dipole points from negative charge to positive charge.[56]

Changes in the vibrational states lead to changes in the dipole moment. Therefore the application of an electric field at different vibrational states leads to an energy difference caused by the change in dipole moment change (Figure 1.4). The orientation of the dipole with respect to the electric field will cause the dipole to be stabilized or destabilized by a various amount. The energy difference between the two states is greater if $\vec{E}_{ext}$ and $\vec{\mu}$ have a different direction and smaller if they are the same. Experimentally, this energy difference (ΔU) can be quantified by measuring the vibrational frequency shift between

different vibrational states ($\Delta\overline{v}$):[57,58]

$$\Delta U = hc\Delta\overline{v} = \Delta\vec{\mu} \cdot \vec{E}_{ext}, \tag{1.2}$$

where h is the Plank's constant and c is the speed of light. The dipole difference can be calculated on the basis of this method from the vibrational frequency shift and the known external electric field acting on it. This can be done by calibrating the vibrational probes in known solvents.[59,60]

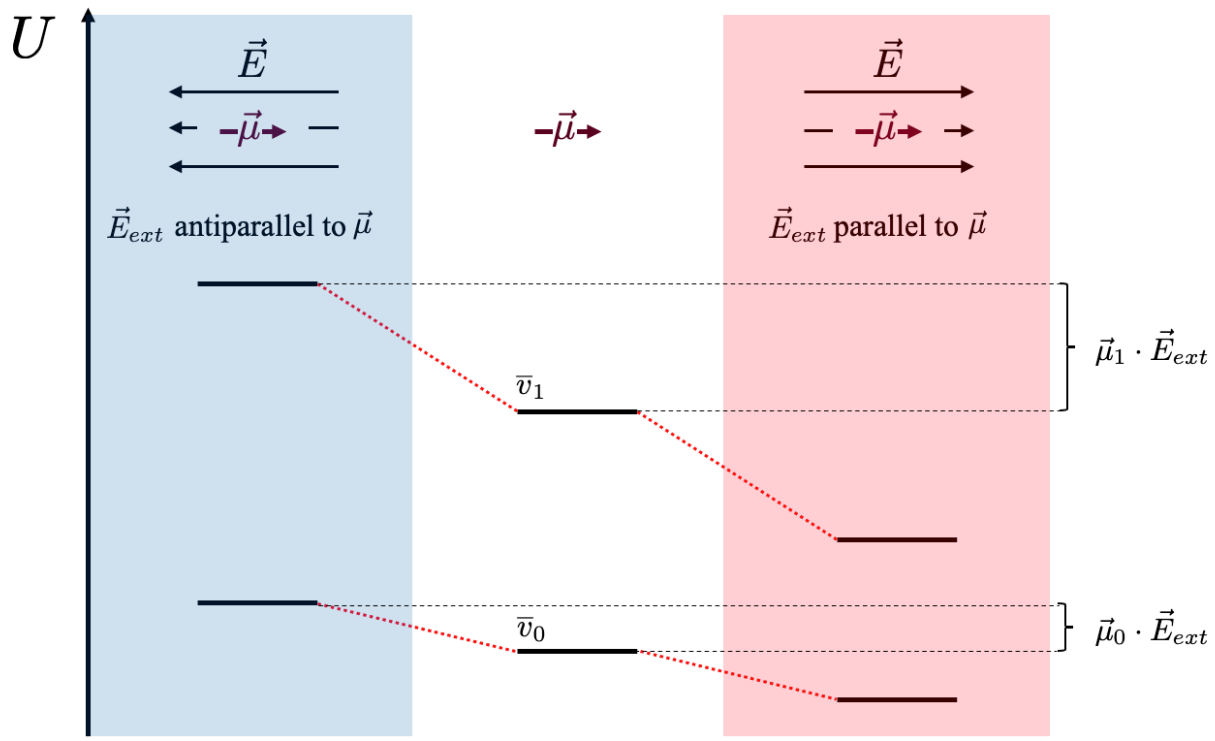

Figure 1.4: The vibrational Stark effect. The dipole moment is different at different vibrational states (denoted $\overline{v}_0$ for ground state and $\overline{v}_0$ for first vibrational state. When applying an external electric field onto a dipole, the shift in vibrational energy is given by $\Delta\vec{\mu} \cdot \vec{E}_{ext}$. If the electric field aligned with the direction of the dipole, the transition energy between states is smaller than when the electric field is in the opposite direction of the dipole.

Knowing the dipole difference for a given probe, VSE can be applied in reverse to measure

an unknown electric field.[58] In this case, the shift in frequency recorded is proportional to the electric field exerted onto the vibrational probe. If we use a protein ligand as a probe, the electric field it experiences when binding to a protein can be quantified.[61] The application of VSE to quantify non-covalent interactions environment is based on the linearity between the vibrational frequency and the field demonstrated by Boxer and co-workers.[44,58,61–63]

VSE has been applied for the study of protein function such as enzyme catalysis.[58] Enzymes catalyze reactions by decreasing the free energy barrier of the reaction compared to the uncatalyzed reaction.[64] This is achieved by generating electric fields that stabilize the reaction transition state, as predicted by Warshel in his electrostatic preorganization theory.[65] This theory has been verified by Boxer and co-workers through the study of ketosteroid isomerase.[63] Using the carbonyl vibration as a Stark probe, a direct link was established between the free energy barrier and the electric field exerted by the protein, with a decrease in the magnitude of the electric field indicating a decrease in the catalytic ability of the enzyme.

Electric fields can not only be measured experimentally, but also calculated computationally, providing a good representation of the charge distribution in the studied system.[66–70]

1.4 Specific Aims

The complexity of non-covalent interactions involved in the regulation of protein mechanism create challenges in understanding protein function. This dissertation will discuss the use of electric fields as a metric to study the structure-function relationship in proteins. Chapter 3 focuses on the modification of synthetic enzyme's catalytic performance. Chapter 4 focuses on relating ion channel structure to ion selectivity and permeation. Chapter 5 focuses on the improvement of natural enzyme slow turnover.

1.4.1 Specific Aim 1: Characterizing the change in enzymatic performance of synthetic enzyme KE15 with/without DNA fragment in its greater environment

Synthetic enzymes are promising tools as industrial reaction catalysts, but still suffer from low catalytic performance compared to natural enzymes. The low enzymatic performance in synthetic enzyme is the result of a inert protein scaffold for the catalytic activity. Since the catalytic activity of enzyme is governed by electrostatic interactions, introduction of a polar fragment such as DNA to the enzyme vicinity should substitute for the ineffective protein scaffold in stabilizing the transition state. This modification of synthetic enzyme's environment was demonstrated on Kemp eliminase enzyme KE15, where the electric fields in the enzyme active site were calculated. The difference between the electric fields in the transition and reactant states showed performance enhancement of KE15 with the introduction of DNA fragment.

1.4.2 Specific Aim 2: Building structure-function relationship for the key regions of $Na_v1.7$ ion channel mechanism

Sodium ion channels play a key role in the cellular communication of excitable cells. They consist of not one, but four domains, which creates challenges to unveil the governing operating principles of these systems. Electric fields have been used to study enzymes, but not yet to study other proteins. We demonstrate that electric field can also be used to rationalize the human voltage gated sodium ion channel $Na_v1.7$.

1.4.3 Specific Aim 3: Probing the product inhibition in enzyme DszB when active site obstructing loops are glycosylated

Although DszB creates a favorable environment to stabilize the transition state, conformational dynamics prevents the product from leaving the active site. This product inhibition needs to be addressed to improve the catalytic activity of the enzyme for industrial applications. We use artificial glycosylation to modify conformational dynamics and decrease active site obstruction. The potential trade-off between the attenuation of product inhibition and the catalytic efficiency of the enzyme, if any, can be calculated with electric fields.

Chapter 2

Methods

2.1 Molecular Dynamics (MD) Simulations

Proteins are crucial macromolecules for cellular activity.[71,72] Although the study of proteins began long before clear structural data was available, the lack of structural information posed a challenge in the field. After the invention of X-ray diffraction in 1913, the long-awaited breakthrough came in 1957, revealing the first protein structure at atomic resolution.[73,74] Advances in computer resources at the same time opened the door to computational simulations for protein research. Macrobiomolecules such as proteins and lipids are known to be flexible. Molecular dynamics (MD) is one of the most common computational methods that can capture the ensemble of conformations of these systems at equilibrium. MD simulations were first applied to protein studies in 1977, followed by application to various subfields such as ligand binding, enzymes or ion channels.[75,76]

In MD, atomic motion is modeled based on Newton's second law:[77,78]

$$\vec{F}_i = -\frac{\partial U}{\partial \vec{r}_i} = m_i \frac{d^2 \vec{r}_i}{dt^2}, \tag{2.1}$$

where $\vec{F}_i$ is the force acting on atom i with mass m_i, U the potential energy of the system, $\vec{r}_i$ the coordinate of atom i, and t the time. U is a function of all atomic coordinates, which couples the set of equations for each atom (N equations for a system with N atoms).

This means that these equations cannot be solved analytically[79–81] and numerical integration schemes, with a discretized time, are instead required.[82] In this case, the atomic coordinates are updated after every time step δt and stored into a trajectory file. While U is a decisive factor for the accuracy of the simulation (see section 2.2), so is δt.[13] If it is too large, the stability of the system is jeopardized; if it is too small, the computational costs increase dramatically. In order to achieve a balance between these two sides, the time step should be of the same order of magnitude as the atomic motion. Since the bond vibration is in the time scale of 10^{-15} seconds, the time step was set to 1 fs.

The initial conformation of the system can be obtained experimentally or computationally. In the case of proteins, the coordinates are often obtained from crystallography or cryogenic electron microscopy (cryo-EM). To start an MD simulation, the atoms (i) are then randomly assigned an initial velocity (v_i) from a Maxwell-Boltzmann distribution:[13]

$$P(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp\left(-\frac{1}{2}\frac{m_i v_i^2}{k_B T}\right), \tag{2.2}$$

where $P(v_i)$ is the probability that atom i has velocity v_i under temperature T given the mass of the atom is m_i, and k_B is the Boltzmann constant.

Note that the system does not reach equilibrium rapidly as the protein and solvent molecules first relax from the initial conformation.[83–85] This phase of the simulation, called equilibration phase, is usually discarded when calculating for equilibrium properties.[81,84] Various observables can be used to determined the duration of the equilibration phase. For example, root mean square standard deviation (RMSD) or potential energy can be used to evaluate convergence to a stationary, equilibrated state.[85–88] The duration of the equilibrium phase also depends on the ensemble in which the coordinates and velocities of the atoms are computed.

The equilibration phase is followed by the production phase, where measurements are carried

out.

Different ensembles are possible for MD simulations. Biological processes normally have a constant pressure and a constant temperature. Therefore, the isothermal-isobaric ensemble (NPT) is most commonly used for the simulation of proteins. This is achieved with the introduction of barostats and thermostats in the system.[81] Barostats, such as Berendsen or Nose-Hoover barostat, maintain a constant average pressure by varying the volume of the system (i.e, scaling of the coordinates).[79,89,90] Thermostats, such as Andersen and Nose-Hoover thermostat, maintain a constant average temperature by modulating atomic velocities.[79,81,90–92]

Figure 2.1 shows the general schematic of an MD simulation.

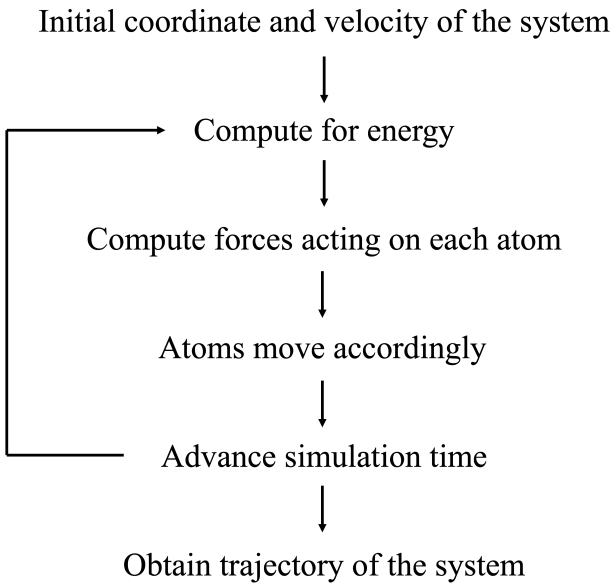

Figure 2.1: Typical sequence of a molecular dynamics simulation process. The initial model of the system provides the atomic coordinates, then initial velocities are assigned to the individual atoms. The atomic coordinates are updated using equation 2.1, and the new set of coordinates is stored in a trajectory archive. The simulation time then moves forward, force acting on atoms are calculated from the energy, and repeats the previous steps.

Computational resources not only influence the time scale of a simulation, but also limit the

number of atoms a system can have. Boundary conditions are applied to bridge the gap between the limited model system size and target condensed phase system. The shape of a simulated cell must allow seamless stacking of its replica in all directions. Various cell shapes are possible for the corresponding system of interest; the cube (rectangular) being the simplest and most commonly used cell shape. Alternatives include the hexagonal prism, the truncated octahedron and the rhombic dodecahedron.[13]

2.2 MD Simulations with AMOEBA Polarizable Force Field

The calculation of the force acting on each atom is the linchpin for the accuracy of an MD simulation. In MD, the force is calculated as the derivative of the potential energy with respect to the atomic coordinates. In classical MD, the potential energy U is an empirical function of the coordinates, which is called the *force field*.

2.2.1 Covalent Interactions

A force field usually consists of several empirical terms, each representing a type of intra- or intermolecular interaction (Figure 2.2). The parameters are determined from quantum mechanical calculations and experimental data. There exists many force fields that model proteins differently depending on the properties they focus on. Despite differences, they all contain the following three terms accounting for covalent interactions: bond stretching, angle bending, and dihedral rotation.[93] A general form of such terms is:

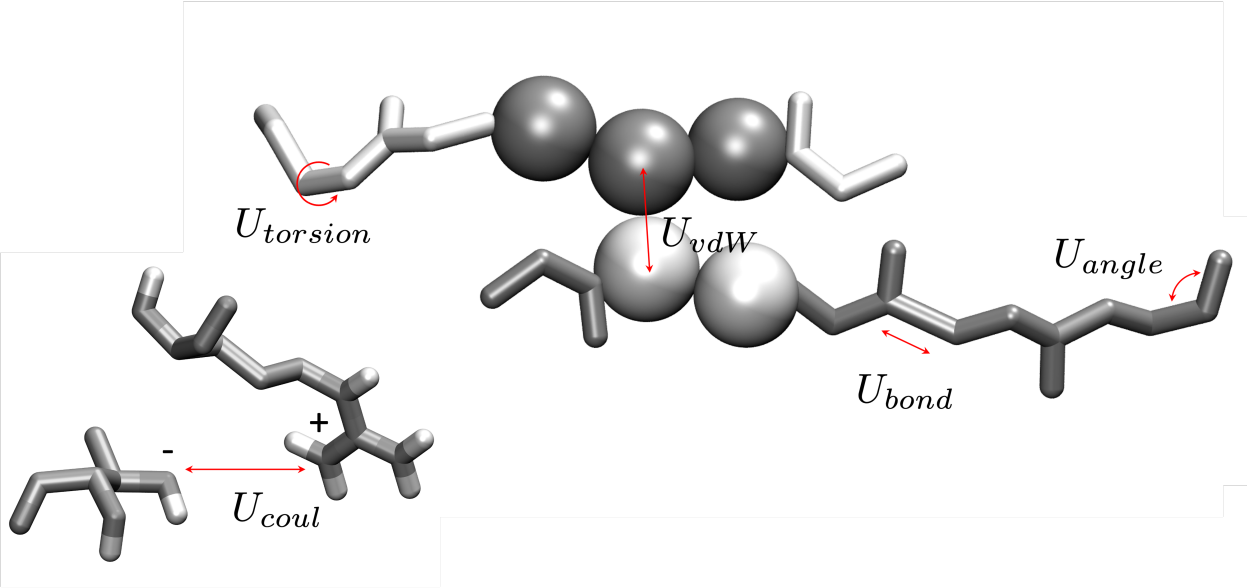

Figure 2.2: A schematic to illustrate the common terms included in a force field: bond stretching (U_{bond}), angle bending (U_{angle}), dihedral rotation ($U_{torsion}$), and non-bonding interactions (U_{vdW}, U_{coul}).

$$U = \sum_{bonds} K_b(r - r_0)^2 + \sum_{angles} K_a(\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\phi - \delta)], \tag{2.3}$$

where the first term describe the energy contribution from bond stretching. K_b is the force constant, r_0 is the equilibrium bond length, and r is the current bond length. Similarly in the second term which describe the energy contribution from angle bending, K_a is the force constant, θ_0 is the equilibrium angle, and θ is the current angle. The third term accounts for the periodic dihedral rotation. V_n represents the height of the potential energy (amplitude), n is the periodicity, ϕ is the torsion angle and δ is the phase.

Additional terms exist depending on the force field. For example, the AMOEBA polarizable force field contains a bond-angle cross term ($b\theta$) and an out-of-plane bending (oop) for intramolecular interactions:[94]

$$U_{b\theta} = K_{b\theta}[(b - b_0) + (b' - b'_0)](\theta - \theta_0) \tag{2.4}$$

$$U_{oop} = K_\chi \chi^2, \tag{2.5}$$

where K is the force constant, χ is the angle between vector jl and plane ijk for a squential bonded center i, j, k, l, and Wilson—Decius—Cross function is applied to restrain the out-of-plane bending.

2.2.2 van der Waals Interactions

Non-bonded interactions are not to be neglected by force fields. The famous Lennard-Jones 12-6 potential is commonly used to describe the van der Waals interactions between the atom pair i and j:

$$U = \sum_{i<j}^{N} 4\epsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}, \tag{2.6}$$

where ϵ_{ij} is the potential well depth, and σ_{ij} is the van der Waals equilibrium distance between ij. However, depending on the character of the system under study, other forms can also be used, such as Lennard-Jones 9-6 and 12-10.[95,96] AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) polarizable force field adopts Halgren buffered 14-7 function for softer repulsive region when accounting for the vdW interactions:[94,97]

$$U = \sum_{i<j}^{N} \epsilon_{ij} \left(\frac{1+\delta}{\rho_{ij}+\delta} \right)^7 \left(\frac{1+\gamma}{\rho_{ij}^7+\gamma} - 2 \right), \tag{2.7}$$

where δ and γ are buffering constants, and ρ_{ij} is given as:

$$\rho_{ij} = \frac{r_{ij}}{r_{ij}^0}, \tag{2.8}$$

where r_{ij}^0 is the minimum energy distance.

2.2.3 Electrostatic Interactions and ELECTRIC

The last general term present in all force fields account for the electrostatic interactions:

$$U = \sum_{i<j}^{N} \frac{q_i q_j}{r_{ij}}, \tag{2.9}$$

where q_i and q_j is the charges on atom i and j.

The uniqueness of AMOEBA polarizable force field, however, lays in its name. A polarizable force field provides a more accurate description of electrostatics and is therefore a more robust force field for protein simulations. Early force fields modeled permanent electrostatics as the interaction of point charges around each atomic center, neglecting the anisotropic distribution of charges. AMOEBA addresses this issue by using atomic multipoles, truncated at the quadrupole moment, to account for local atomic charge distributions:[98–102]

$$M_i = [q_i, \mu_{ix}, \mu_{iy}, \mu_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixy}, ..., Q_{izz}]^T, \tag{2.10}$$

where q_i is the point charge of atom i, μ_i the dipole moment, Q_i the quadrupole moment.

For atom i, the permanent electrostatic potential term due to atom j, U_{ij}^{perm}, is then given by:[103,104]

$$U_{ij}^{perm} = M_i^T T_{ij} M_j, \tag{2.11}$$

where the interaction matrix is (r_{ij} is the distance between atoms i and j):

$$T_{ij} = \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \cdots \\ \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i x_j} & \frac{\partial^2}{\partial x_i y_j} & \frac{\partial^2}{\partial x_i z_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i x_j} & \frac{\partial^2}{\partial y_i y_j} & \frac{\partial^2}{\partial y_i z_j} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i x_j} & \frac{\partial^2}{\partial z_i y_j} & \frac{\partial^2}{\partial z_i z_j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \frac{1}{r_{ij}}. \tag{2.12}$$

Under the influence of an external field, the electron density undergoes a redistribution known as electrostatic polarization. A common approach for modeling this phenomenon is to characterize this charge distortion within the individual atoms as an induced dipole, or to represent the flow of charge between the atoms as a fluctuation charge model.[101] AMOEBA uses the first approach and the induced dipole at site i, $\vec{\mu}_i^{ind}$, can be expressed as:[105]

$$\vec{\mu}_i^{ind} = \alpha_i \vec{E}_i, \tag{2.13}$$

where α is the polarizability. The electric field ($\vec{E}_i$) experienced at atom i by atom j is then given as the sum of the permanent (perm) and induced (ind) electric field:

$$\vec{E}_i = \vec{E}_i^{perm} + \vec{E}_i^{ind}. \tag{2.14}$$

The x component of each term can be expressed in detail respectively as:

$$E_{i,x}^{perm} = \sum_j E_{(i,j),x}^{perm} = \sum_j \left(-T_x q_j + \sum_m T_{xm}\mu_{j,m} - \frac{1}{3}\sum_m \sum_n T_{xmn} Q_{j,mn} \right), \tag{2.15}$$

and

$$E_{i,x}^{ind} = \sum_j E_{(i,j),x}^{ind} = \sum_j \sum_m T_{xm}\mu_{j,m}^{ind}, \tag{2.16}$$

where $m, n \in \{x, y, z\}$.

In AMOEBA, the induced dipole is based on Thole's damping scheme.[106] The charge distribution is "smeared" in each interacting ij pair, with a charge distribution function in the form of:[105]

$$\rho = \frac{3\alpha}{4\pi} \exp\left(-au^3\right), \tag{2.17}$$

where α is the polarizability, a is the smeared charge distribution parameter, and u is the effective distance as a function of atomic polarizability α_i and α_j given by:

$$u = r_{ij}(\alpha_i \alpha_j)^{(1/6)}. \tag{2.18}$$

Overall, AMOEBA polarizable force field can be expressed as follows:

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind}. \tag{2.19}$$

Electric fields can be obtained both experimentally computationally. Our code ELECTRIC

(Electric fields Leveraged from multipole Expansion Calculations in Tinker Rapid Interface Code) calculates electric fields under the framework of the AMOEBA force field.[107] The electric field can then be decomposed into contributions from individual components of the system. Each component is recognized in the ELECTRIC code through an input coordinate file provided by the user.

The electric field exerted from sites k to site i can be projected onto a normalized directional vector defined by atom i and k:

$$E_{proj}^{ik} = \left(\frac{\vec{E}^{i} + \vec{E}^{k}}{2} \right) \cdot \vec{u}^{ik} \tag{2.20}$$

Chapter 3

The Effect of Environmental DNA on the Catalytic Activity of Synthetic Enzyme KE15

3.1 Background

Discussion about the use of enzymes outside their natural environment has never stopped since the discovery of their remarkable catalytic ability. They are highly efficient, selective, non-toxic, biodegradable and can operate at very low concentrations and ambient temperature.[64,108] These properties make them valuable for commercial applications.[109–112] The challenges in the application of enzymes lie in their sensitivity to their environment. Indeed, the pH or temperature used in commercial applications can lead to structural changes in the enzymes that impair their catalytic performance.[113–115] Furthermore, natural enzymes cannot catalyze non-metabolic reactions, which limits their application in industry.[108,116] Ideally, we would like to be able to design synthetic enzymes that could adapt to a wider range of reactions and are more resistant to reaction conditions. However, the catalytic rate of current synthetic enzymes is still far from satisfactory compared to natural enzymes.[117,118]

The Kemp Eliminases (KE) is a family of synthetic enzymes computationally designed by

Baker's group.[119–121] They catalyze the Kemp Elimination reaction, a proton abstraction from carbon (Figure 3.1), which no natural enzymes targets in metabolic pathways. This makes KE a paradigm of synthetic enzymes for both theoretical and experimental studies.

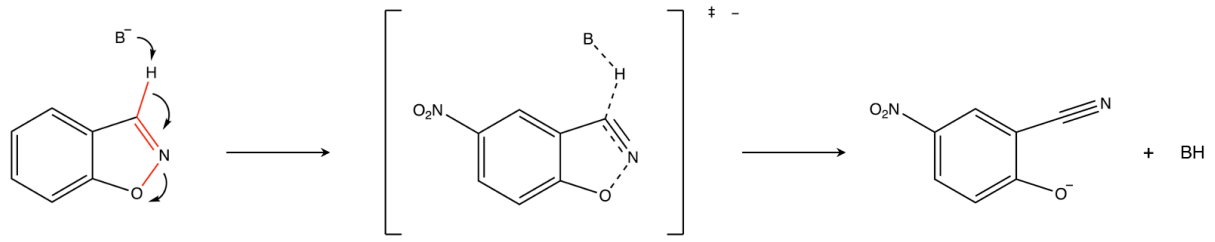

Figure 3.1: The Kemp Elimination reaction mechanism. First step of the reaction is the deprotonation of the 5-benzisoxazole ring by a catalytic base (B^-). The ring opens with the breakage of C-H, N-O and the formation of a C-N triple bond. The three active bonds of the reaction are colored in red.

Numerous studies investigated KEs to better understand their limited efficiency.[119,122–125] They found that while a few protein residues in the active site facilitate the evolution of the reaction, most of the protein scaffold was inert.[120,125] This is because only two or three active site residues are actually designed to stabilize the transition state. The rest of the protein sequence is adapted from existing proteins to accommodate the designed residues.[126–129] In natural enzymes, although the number of active sites residues directly involved in the reaction are limited, the entire protein participates in the enzymatic function. Residues throughout the protein exert an electric field that stabilize the transition state in the active site. This stabilization of the transition state lowers the reaction energy barrier, which accelerates the evolution of the reaction.[130] In contrast, the design strategy employed for KE implies that a majority of the synthetic enzyme is designed for protein stability rather than function (i.e., acceleration of the reaction rate).[131–133] This was confirmed in several studies where KE scaffold residues did not generate electric fields that stabilize the reaction dipole change in the active site.[134–136]

Laboratory directed evolution (LDE) has been used to improve the performance of these enzymes by selecting beneficial mutations, often in the protein scaffold.[120,137–139] Despite its success in some cases, LDE does not guarantee performance enhancement, is expensive and poses a risk to the structural stability of the protein.[140]

Modifying the environment in which the enzyme operates, rather than the enzyme itself, offers several advantages. First, this approach has low risk in destabilizing the enzyme structure. Mutation, even only for single residue, can lead to changes in backbone strain, functional area, or other crucial factors of protein stabilization.[141–143] Second, there is less trade off between protein environment modification and enzyme catalytic activity when using these enzymes in industry. Indeed, the change of environment in commercial applications often lead to low commercial performance.[109]

In this chapter, we introduce a polar molecule in the environment to stabilize the transition state in the active site in place of the protein scaffold (Figure 3.2). DNA, with its polar structure and high programmability, is an ideal candidate for this approach.[144–147]

DNA fragments consist of two parts, the backbone with alternating deoxyribose and negatively charged phosphate groups and the paired nucleotide bases that connect the two strands of DNA. While the polar phosphate-sugar backbone induces electrostatic interactions to stabilize the transition state, the predictable base pairing of DNA has greatly simplified the process of structure design.[148,149] The numerous examples of DNA-protein attachments that occur either naturally or artificially, show that the structural stability of the enzyme is not negatively affected by the presence of DNA fragments in its environment.[150–153]

Here we demonstrate our enzyme performance enhancement approach by placing a 10 base-pair DNA fragment at four randomly selected positions around the KE15 protein scaffold, shown in Figure 3.3. We monitor the effect of DNA on the catalytic activity of the synthetic

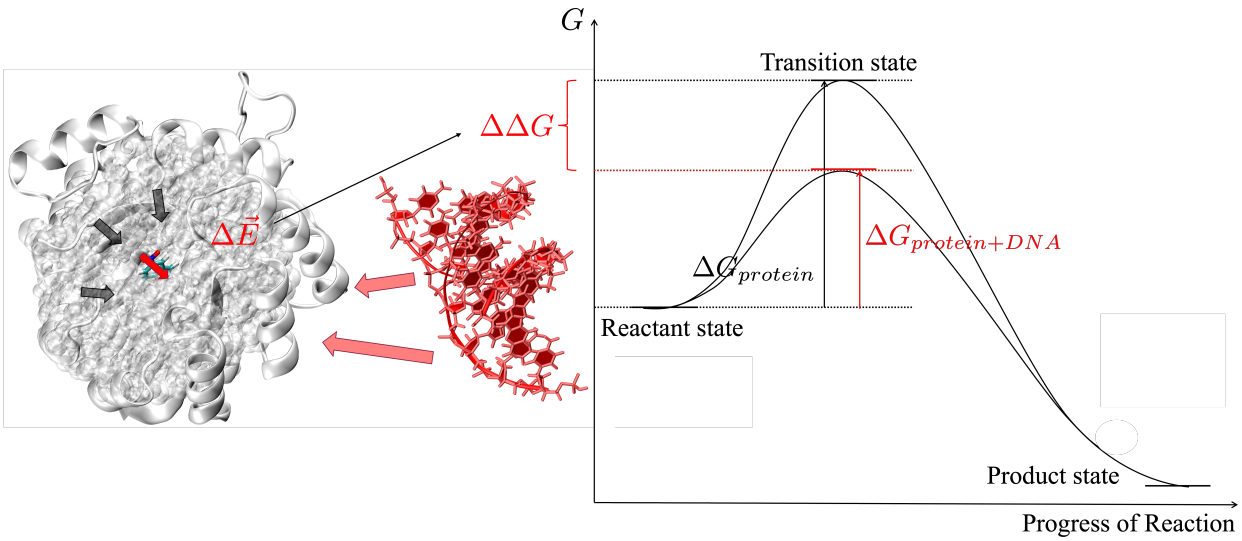

Figure 3.2: DNA-induced enzyme performance enhancement. In the framework of electrostatic preorganization theory, the DNA fragment around the enzyme can induce changes in electric fields from the protein scaffold to the ligand. This external factor substitute for the missing role of protein scaffold, reducing the activation energy.

enzyme KE15, by computing the electric fields generated in each system. More specifically, we calculate the electric fields from individual residues projected onto the three bonds that break and form during the Kemp Elimase reaction (bonds C-H, N-O, and C-N).

3.2 Computational Details

3.2.1 Conformational Ensemble

For each of the four KE15-DNA complex systems, 25 uncorrelated conformations were generated for both reactant state and transition state. We generated the conformational ensemble using the backrub algorithm of the ROSETTA package from 25 × 1000 trials.[154,155] The protein backbone was rotated around the axis defined by C_α, followed by a sampling step to optimize the placement of side chain rotamers using the fixbb algorithm.[156] The

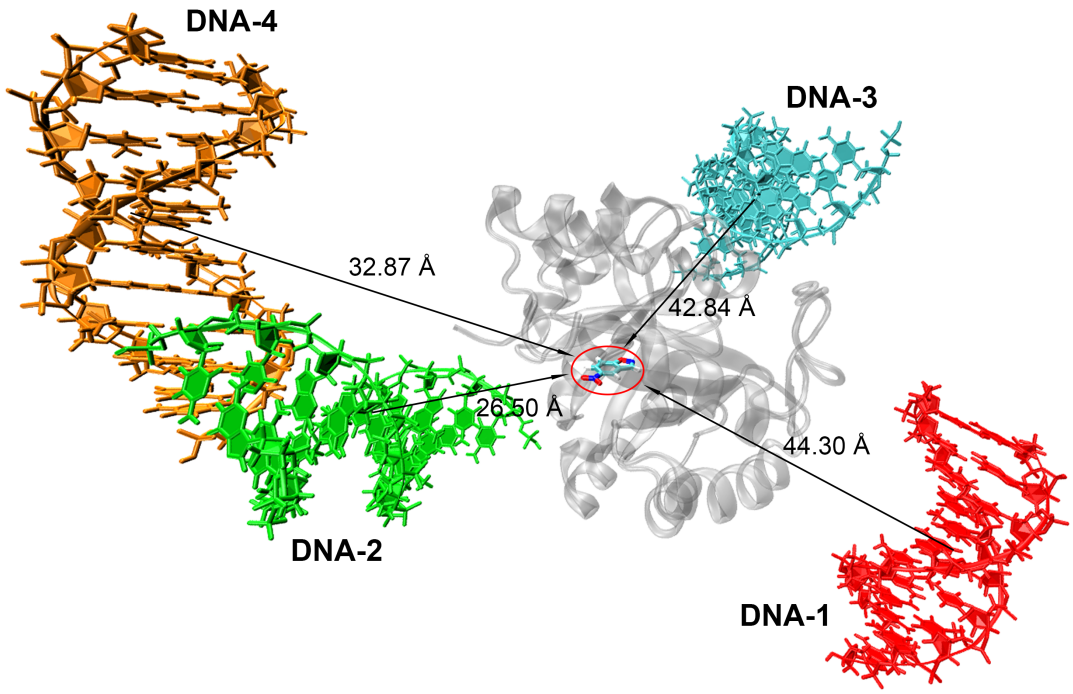

Figure 3.3: The position of DNA fragment (GGTCATGACC-CCAGTACTGG) around KE15. Each KE15-DNA complex system is colored differently for DNA fagement. Distance between the DNA fragment and the ligand is labeled. The ligand is circled in red at the center of KE15 protein scaffold (1THF barral scaffold). Reproduced with permission from Zheng, Y.; Vaissier Welborn, V. Tuning the Catalytic Activity of Synthetic Enzyme KE15 with DNA. *J. Phys. Chem. B.* **2022**, 126, 18, 3407–3413. Copyright 2022 American Chemical Society

conformational ensemble created with this approach explored the side chain conformations that would occur on a nanosecond to microsecond time scale.[157,158] The DNA fragment was placed around the KE15 protein scaffold using PACKMOL.[159] The simulation box size was $80 \times 106 \times 72$ Å^3 for KE15+DNA-1, $86 \times 97 \times 79$ Å^3 for KE15+DNA-2, $103 \times 97 \times 89$ Å^3 for KE15+DNA-3, $86 \times 88 \times 87$ Å^3 for KE15+DNA-4, each with 10 Å of solvent buffer between KE15-DNA complex and the edge of the box.

3.2.2 Molecular Dynamics

We used Tinker 8 for the MD simulations of four KE15-DNA complex systems in both reactant and transition state.[160] The geometry of the ligand structure in the transition state was the same as in the reactant state. The difference between the two states was in the atomic multipoles, so the electrostatics of the two states were different.[125] Solvation of the KE15-DNA complex was performed with GROMACS with the water solvent pre-equilibrated to accelerate the equilibration process.[161,162] The MD simulations were performed under isothermal-isobaric conditions (1 atm, 300K) with the AMOEBA polarizable force field. Each simulation lasted 100 ps with the time step of 1 fs (50 ps of the equilibration phase and 50 ps of the production phase). The time length of equilibration was based on the energy convergence. The time evolution of the RMSD (Figure S1 to Figure S4) and energy (Figure S5 to Figure S8) can be found in Supporting Information. In total, the 25 conformations in each KE15-DNA complex system generated 1.25 ns of production phase. As mentioned above, the conformational ensemble was generated from an algorithm describing side-chain conformations in the time domain up to microseconds. Therefore, the production phase data were sufficient to study the catalytic performance of the enzyme.

3.2.3 Electrostatic Stabilization of the Transition State

The effect of the electric field in the transition state free energy of stabilization was calculated as follow:

$$\Delta G^{\ddagger}_{elec} = \sum_{m} \left(\vec{\mu}^{\ddagger}_{m} \langle \vec{E}^{\ddagger}_{m} \rangle - \vec{\mu}_{m} \langle \vec{E}_{m} \rangle \right), \tag{3.1}$$

where the summation is over the three active bonds C-H, O-N, and C-N triple bonds, denoted by m. $\vec{\mu}$ is the bond dipole and $\vec{E}$ is the electric field, $\ddagger$ denotes the transition states. $\langle \vec{E}_m \rangle$ is the electric field averaged over the conformational ensemble. The bond dipoles are listed in Table 3.1.

Table 3.1: Net bond dipole of the three active bonds of Kemp Elimination reaction using AMOEBA forec field electrostatics.[124]

The norm of $\vec{\mu}$ (Debye)		
Bond	Reactant state	Transition state
C-H	-1.0	1.0
O-N	-1.7	2.3
C-N	2.0	0.4

The electric field projected onto the active bonds were calculated with equation 2.20, while the electric field experienced by each site was a combination of induced and permanent field (Eq.2.14). The calculation for eclectic fields were performed with ELECTRIC.[107]

3.3 Results and Discussion

3.3.1 Residue Contribution to the Transition State Electric Fields

First, we calculated the electric fields in the active site of KE15 exerted from the protein scaffold, the DNA fragment and the solvent (see Figure 3.4). The electric fields were projected onto the three active bonds (C-H, O-N, C-N) in the reactant and transition states.

In this chapter, we projected the electric fields such that a positive signal means an easier redistribution of charge in the transition state. That is, for each of the active bonds, a positive projected electric field indicates that the field facilitates the reaction, promoting the breaking of the C-H, and O-N bonds and the formation of the C-N triple bond.

Previous study analyzed the contribution of each KE15 residue to the electric fields in the active site.[124] We use this study as a reference to understand the effect of introducing DNA on the electric fields. Significant contributors to the electric field projection were similar with or without the DNA's presence, indicating individual residue's role in transition state stabilization remains the same. However, the magnitude of their contribution showed a significant increase. For example, the catalytic base Asp 48 contributed significantly to the transition state electric fields without DNA. With DNA, we see a 296% increase in the magnitude of the contribution of Asp 48 on the fields projected onto the CN bond. An increase of 65% and 21% was also observed for the contribution of Asp 48 on the fields projected onto the CH and ON bonds, respectively. Other key residues, such as Asp 98, Ser 169 and Asp 219, had a similar increase in contribution when in presence of DNA. Changes in the DNA fragment placement varied the magnitude of the transition state electric field projection. For example, the catalytic base Asp 48 made a consistently high contribution. Its contribution to C-H bond breaking was highest in DNA-4 with 27% compared to DNA-1,

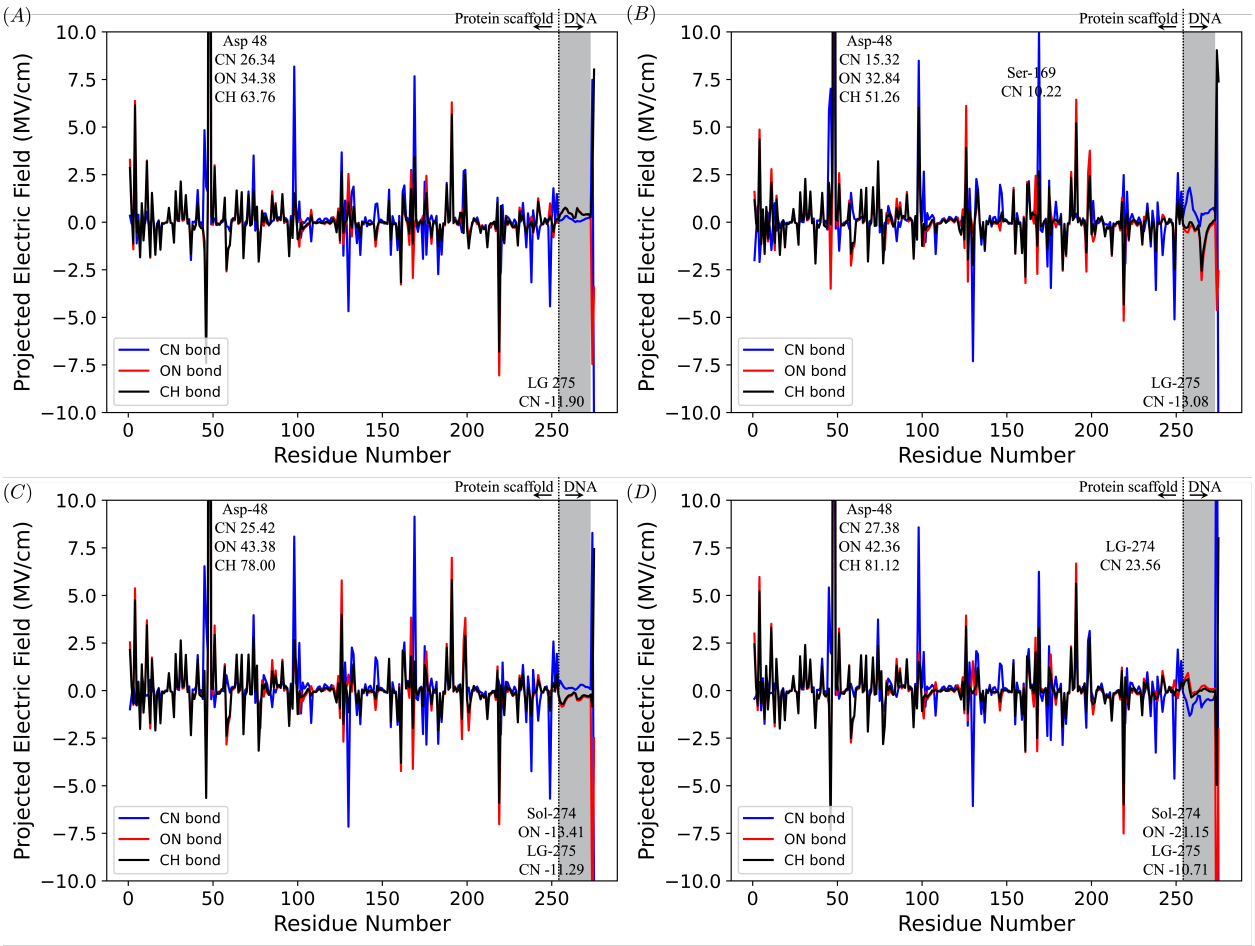

Figure 3.4: Contribution of individual residue to transition state electric fields projected onto the three active bond, (A) to (D) each correspond to a DNA position. Peak value higher than 10 MV/cm were cut off in order to display lower peaks, the magnitude of the cut off peaks were on the side. The grey bar colored the residues from the DNA fragment (resid 254 to 273). Residue 1 to 253 is the protein scaffold. Solvent (water and counterions) is residue 274, ligand is residue 275.

58% compared to DNA-2 and 25% compared to DNA-3.

The negative contribution from Asp 219 indicates that the electron flow is directed against CH and ON bond breaking, while Asp-98 has the opposite effect. This difference suggests that the DNA fragment does not discriminate between residues that favor or disfavor the stabilization of the transition state, but that it enhances the interactions that were originally present. Note also that unlike Ser-169, which is in the active site, Asp-98 and Asp-219 are located in the scaffold (Figure 3.5) and are exposed to water. One possible role of these residues in KE15 is that they can deliver their effect through solvent. When exposed to water, these surface residues may be able to reorient the water molecules outside the active site, potentially creating an environment facilitating the transition state stabilization.

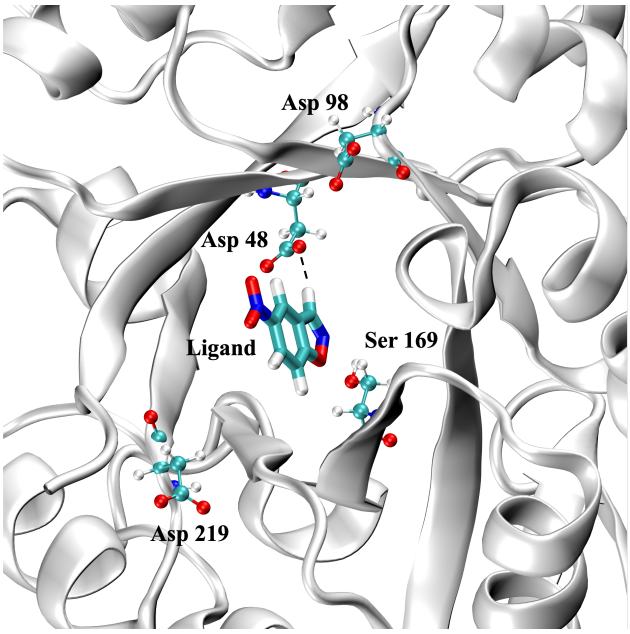

Figure 3.5: Position of residues significant for the transition state projected electric fields. Asp-48 is the active base, with the interaction demonstrate by black dashed line.

The contribution from residues 254 to 273 were no higher than 3 MV/cm, indicating that the DNA fragment in all four KE15-DNA complexes do not make direct contributions to the electric field projected onto the ligand. The solvent, however, had a negative contribution

to the electric field projection in all four complexes. This means the water molecules and counterions together were destabilizing the transition state, favoring the electron flow in the reactant state.

3.3.2 Transition State Stabilization Free Energy

Quantification of the overall performance of the enzyme can be found in Table 3.2, where we analyzed the transition state stabilization free energy. $\Delta G^{\ddagger}_{elec}$ was first calculated individually for each of the three active bonds, and summed into a total $\Delta G^{\ddagger}_{elec}$ as shown by Equation 3.1. Significant reduction of the stabilization free energy came from the active base, DNA, and solvent. The solvent shown the preference of destabilizing the transition state in the analysis of electric field projection, while DNA fragment has no direct contribution. Therefore, part of the DNA fragment's role is to replace the solvent demoting the transition state electron flow. As discussed earlier, the active base experience notable enhancement for electric field magnitude, indicating DNA's ability to tune enzyme performance was partially achieved through strengthening the interaction between the active base and the ligand.

In DNA-1 and DNA-4, DNA fragment decreased the $\Delta G^{\ddagger}_{elec}$ by 2 kcal/mol, where the placement of DNA-2 and DNA-3 gained a 4.2 kcal/mol and 2.3 kcal/mol increase, respectively. Comparison of the four DNA fragment position suggested the difference was not a result of its distance with the ligand, but the fragment orientation with the scaffold. Specifically, DNA-1 and DNA-3 were placed 10 Å further away from the ligand then DNA-2 and DNA-4, suggesting no significant connection between the distance and the decrease of $\Delta G^{\ddagger}_{elec}$. As shown in Figure 3.3, DNA-1 and DNA-4 were located on the side of the scaffold whereas DNA-2 and DNA-3 were facing the active site without scaffold in between. The polar DNA fragment can disrupt the favorable environment protein created for the catalytic function.

Table 3.2: Stabilization free energy of the transition state for each KE15-DNA system. In each active bond, the total was calculated by the sum of base (Asp 48), active site (residue 5, 46, 78, 101, 126, 167, 168, 169, 197, 198, 199, 201, 220), scaffold (residue 1 to 253, excluding base and active site), DNA (residue 254 to 273), and solvent (residue 274). The error was calculated from the mean of time and ensemble averages. The reference system is KE15 without DNA moelcule from ref[124]

$\Delta G^{\ddagger}_{elec}$ (kcal/mol)					
C-H	KE15[124]	KE15+DNA-1	KE15+DNA-2	KE15+DNA-3	KE15+DNA-4
Base	-4.6	-5.1±0.05	-4.9±0.06	-7.1±0.07	-5.9±0.06
Active site	0.1	0.1±0.03	-0.3±0.02	0.2±0.03	0.6±0.03
Scaffold	-1.0	-0.4±0.06	-1.0±0.07	-0.5±0.08	-0.5±0.07
DNA	N/A	-0.8±0.001	1.1±0.004	0.7±0.001	-0.4±0.002
Solvent	0.1	0.1±0.04	-0.3±0.03	0.5±0.03	0.6±0.04
Total	-5.4	-6.1±0.09	-5.4±0.1	-6.2±0.1	-5.6±0.1
O-N	KE15[124]	KE15+DNA-1	KE15+DNA-2	KE15+DNA-3	KE15+DNA-4
Base	-6.0	-5.6±0.04	-5.9±0.05	-7.7±0.05	-6.8±0.05
Active site	1.1	1.2±0.05	0.1±0.05	1.2±0.05	1.4±0.05
Scaffold	-1.6	-0.3±0.08	-1.4±0.07	-0.7±0.08	-0.7±0.07
DNA	N/A	-1.5±0.002	2.7±0.007	1.7±0.003	-1.2±0.005
Solvent	3.2	2.0±0.05	1.2±0.05	2.8±0.06	4.3±0.06
Total	-3.3	-4.2±0.1	-3.3±0.1	-2.7±0.1	-3.0±0.1
C-N	KE15[124]	KE15+DNA-1	KE15+DNA-2	KE15+DNA-3	KE15+DNA-4
Base	2.0	1.1±0.03	1.7±0.04	2.3±0.05	1.4±0.03
Active site	0.0	0.4±0.04	0.7±0.03	0.6±0.04	0.4±0.03
Scaffold	0.8	1.0±0.06	0.9±0.06	1.1±0.07	1.4±0.06
DNA	N/A	0.3±0.002	0.4±0.005	-0.1±0.004	-0.5±0.004
Solvent	0.2	-0.2±0.04	-0.7±0.04	-0.9±0.04	-0.2±0.05
Total	3.0	2.6±0.09	3.0±0.08	3.0±0.1	2.5±0.09
Total	KE15[124]	**KE15+DNA-1**	KE15+DNA-2	KE15+DNA-3	KE15+DNA-4
Base	-8.6	-9.6±0.07	-9.1±0.09	-12.5±0.10	-11.3±0.08
Active site	1.2	1.7±0.08	0.5±0.06	2.0±0.07	2.4±0.06
Scaffold	-1.8	0.3±0.11	-1.5±0.11	-0.1±0.13	0.2±0.11
DNA	N/A	-2.0±0.002	4.2±0.01	2.3±0.004	-2.1±0.006
Solvent	3.5	1.9±0.08	0.2±0.07	2.4±0.08	4.7±0.09
Total	-5.7	**-7.7**±0.17	-5.7±0.17	-5.9±0.20	-6.1±0.18

The KE15-DNA complex system for DNA position 1 to 3 also yield a significant decrease of $\Delta G^{\ddagger}_{elec}$ from solvent. The addition of DNA fragment partially replaced the solvent molecules that were ineffective for enzyme catalysis, reorient them to enhance the stabilization of transition state. This is consistent with previous analysis. Overall, the DNA fragment can not only significantly affect the activation energy, but also improve the performance of protein scaffold and the solvent. Lastly, the active site and the protein scaffold did not have additional effect to enhance KE15's catalytic performance.

When looking at the change of total $\Delta G^{\ddagger}_{elec}$ of the reaction comparing to the reference system (without DNA), DNA fragment either have no effect for enzyme performance such as DNA-2 and DNA-3, or have additional reduction of the free energy requirement for transition state stabilization. This reveals one additional advantage of this method: the catalytic performance does not get worse. KE15-DNA-1 complex system yield 2 kcal/mol difference in $\Delta G^{\ddagger}_{elec}$, where the DNA fragment was further away from the scaffold than DNA-2 or DNA-4. This is converted to a 30-fold increase for the reaction kinetic rate under catalyzation. This indicate the second advantage of this method. No constrain for the DNA position to be within close proximity of the scaffold is required. DNA fragments are flexible, therefore no distance limitation can reduce the design effect significantly. The only factor requires control in this method is the orientation of the DNA fragment with respect to the scaffold.

3.4 Summary

By placing DNA fragments in the vicinity of the scaffold, we demonstrate an alternative approach to improve the performance of the computationally designed KE15 enzyme. The polar DNA fragment leads to a change in the electrostatic environment of the enzyme, which affects the reactivity of the ligand in the active site. The mechanism of this enhancement

has been demonstrated with electric fields.

Electric field projections onto the bonds that break and form during the reaction provide a direct measure of the role of specific residues in stabilizing the transition state. We found that the introduction of a DNA fragment can significantly strengthen the interaction between the catalytic base and the ligand, facilitating the breaking and formation of these bonds. Solvent exposed residues outside of the active site can also impact the transition state stabilization. No new interactions were created, neither between the protein and the ligand nor between the DNA fragment and the ligand. Instead, the surrounding DNA fragment strengthened the original interactions, as can be seen from the overall trend of the contributions of the residues with and without DNA. We found a reduction in total activation energy of up to 2 kcal/mol by using DNA fragments. Other than a significant decrease of the transition state stabilization energy from the active base, the DNA fragment replaced the ineffective water, resulting in the DNA and solvent contributions to the lower transition state stabilization energy.

The placement of DNA fragment plays a crucial role in its effective tuning of KE15 enzyme performance. The fragment facing the active site is much less effective in lowering the activation energy than those facing the protein scaffold. This is because the mechanism of DNA-based KE15 tuning is the enhancement of the original protein-ligand interactions.

Finally, it is worth noting that this difference of 2 kcal/mol was achieved in a preliminary test system without optimizing the position, size or sequence of the DNA. Systematic investigation of the design of the enzyme-DNA complex could lead to further improvements in catalytic performance. Established techniques such as artificial DNA-protein binding and DNA origami provide an experimental approach to create this complex and reduce the challenge of translating such a design in silico to in vitro.

Chapter 4

Electric Fields in Voltage-Gated Sodium Channel Facilitate Function

4.1 Background

Establishing the structure-to-function relationship in proteins is a long-standing aim of molecular biology research.[163,164] In the last two decades, the electrostatic preorganization theory proposed by Warshel, has been widely accepted as the governing principle of enzyme's highly efficient catalytic activity.[65,130] The theory states that enzymes create an electrostatic environment that stabilize the transition state, promoting the evolution of the reaction in the active site. This theory has been validated by measurements and calculations of electric fields in enzymatic active sites.[58,63] Although proteins are diverse and participate in a wide range of biological processes, they share similarities, and their function is generally supported by the entire structure.[30,165] It is possible then, that proteins other than enzymes could also be electrostatically preorganized to facilitate their function. In this chapter, we investigate whether voltage-gated sodium ion channels are preorganized.

Voltage-gated sodium ion channels (Na_vs) are pore-forming membrane proteins. They open, inactivate or close in response to changes in the transmembrane voltage (Figure 4.1).[166] There are two key narrow regions along the Na_vs pore: the selectivity filter (SF) formed by residues Asp, Glu, Lys, Ala (DEKA) and the intracellular gate (IG), that are responsible for

ion selectivity and gating,respectively.[167] When Na_v is open, both the SF and IG are opened and Na^+ can diffuse through the pore. When Na_v is closed, the SF is closed while the IG remains open, preventing the Na^+ from entering the Na_vs. When Na_v is inactivated, the SF is open while the IG is closed, preventing Na^+ from exiting the pore. Note that the Na_vs adopt a similar overall structure when open and inactivated.[167]

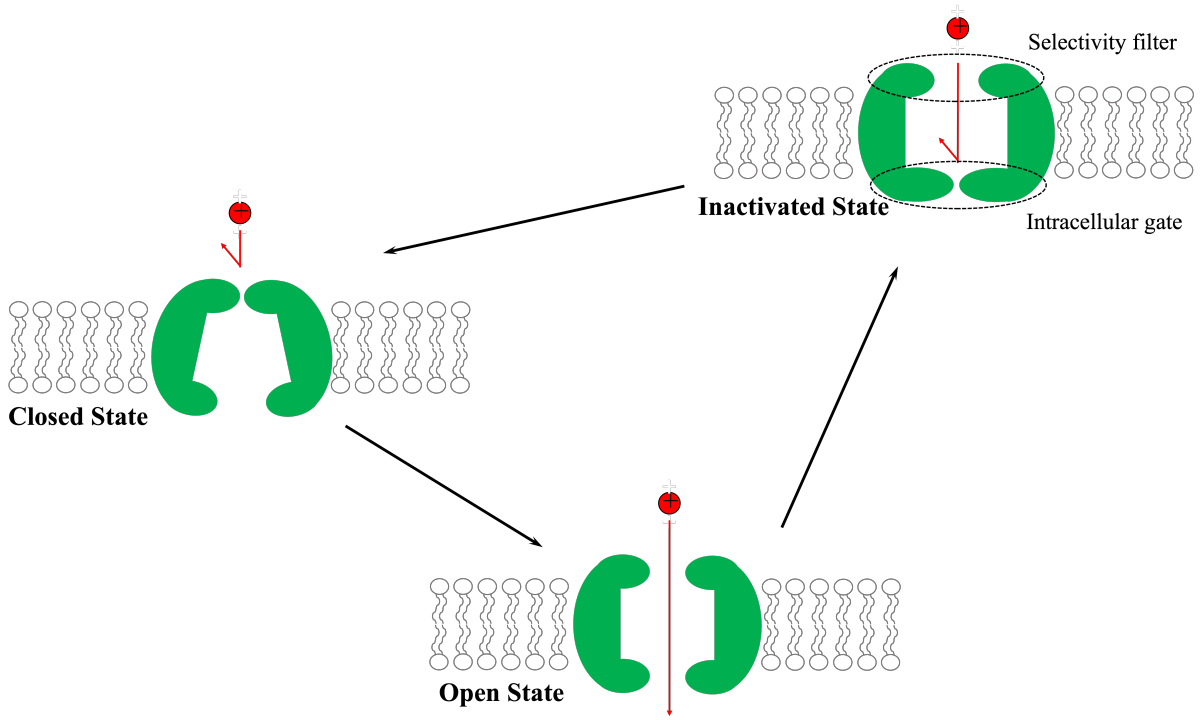

Figure 4.1: Illustration of Na_vs cycle. The Na^+ is represented by the red ball. The open state and inactivated states shares high structural similarity except the "closed" intracelluar gate. Na^+ ions are allowed to pass through the selectivity filter in both state, but cannot exit under inactivate state. Ion permeation completely stop under closed state conformation.

Na_vs are found across species, including mammals, where they are responsible for life-sustaining physiological processes such as brain function, heartbeat or muscle contraction.[168] There are nine Na_vs in mammals ($Na_v1.1$ to $Na_v1.9$) that share structural and functional similarities.[169–172] $Na_v1.1$ to $Na_v1.3$ are expressed in the central nervous system (CNS), $Na_v1.4$ and $Na_v1.5$ in the skeletal and cardiac muscles, $Na_v1.7$ to $Na_v1.9$ in the peripheral nervous

system (PNS) and $Na_v1.6$ in both the CNS and PNS. The main component of Na_vs is a large α subunit that consists of four homologous transmembrane domains (DI to DIV) connected by intercellular loops (Figure 4.2).

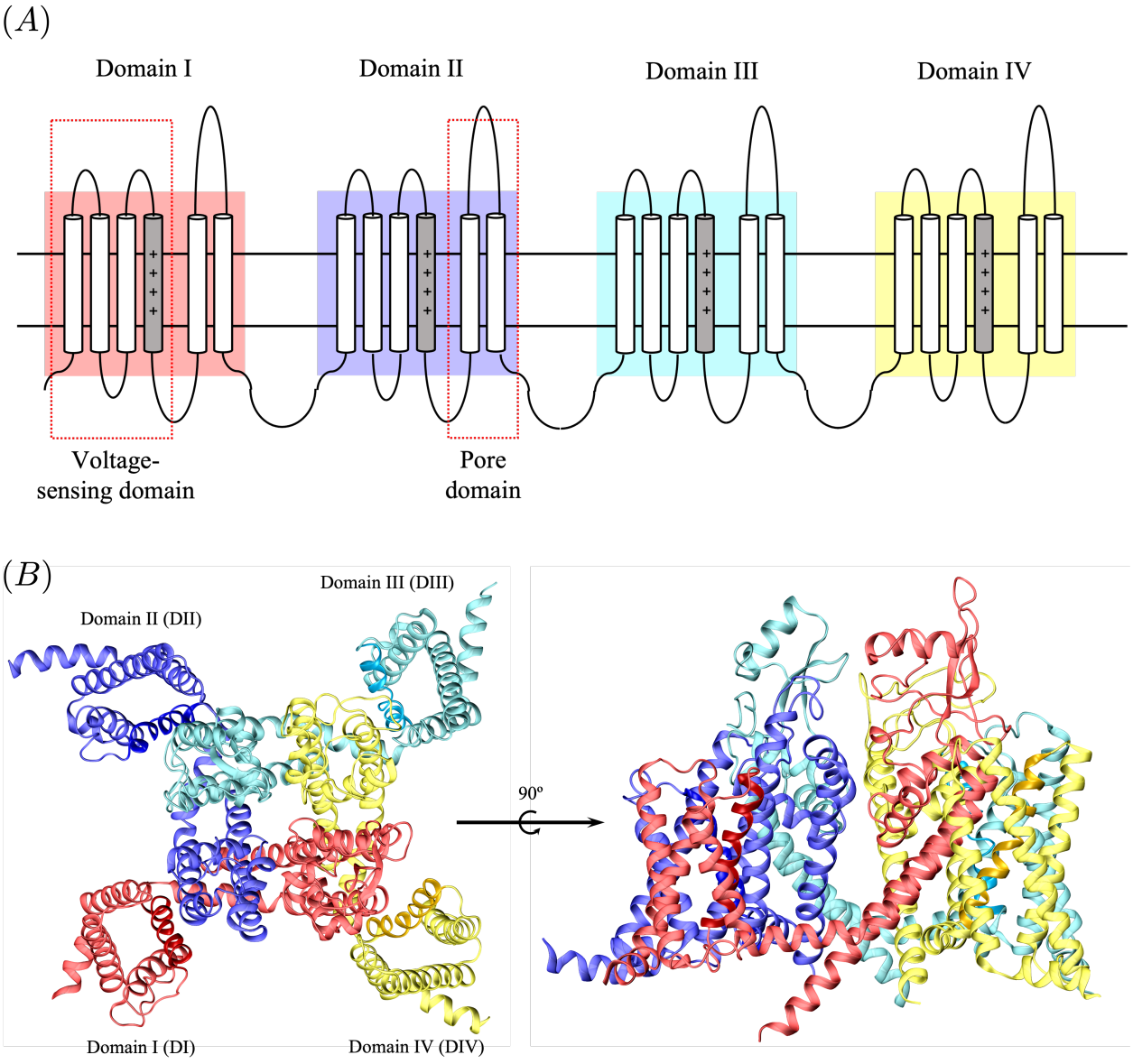

Figure 4.2: Demonstration of Na_vs structure. (A) A sketch demonstrating the structure of Na_vs. Different background indicates the four homologous domains. (B) Top and side view of $Na_v1.7$, each domains are colored correspondingly.

Each of the domains consists of six helices (S1 to S6). Helices S1 to S4 of each domain (DI-

DIV) assemble on the outer region of protein, forming four voltage-sensing domains (VSDs). As the name suggests, the VSDs detect the potential difference across the membrane, triggering a change of state. In particular, S4 slides up the membrane during depolarization, triggering Na_v opening.[173] Every three residues in S4 are positively charged amino acids, called gating charges (GC) residues, that facilitate the motion of S4 and therefore controls the state of Na_vs. Helices S5 and S6 of each domain (DI-DIV) assemble to form one pore domain (PD) at the center of Na_vs, where the SF and IG are located.

Since $Na_v1.7$ is linked to pain signaling in humans, insights into its mechanism at the molecular level would provide valuable information for the treatment of pain syndromes.[174] Although the signature SF and IG motifs are known, they are not specific to $Na_v1.7$, which hinders the design of efficient drugs. Understanding the rest of the protein structure's role in facilitating $Na_v1.7$ function progress slowly, leaving many questions unanswered .[172,175] For example, the Na_vs family lacks symmetry between domains, which raises the question of the properties of individual regions.[176] Comparison between wild type $Na_v1.7$ and its mutants provide valuable insight in studying individual residue's role for the mechanism, but is often time consuming.[172,177]

With the high resolution of the human $Na_v1.7$ recently discovered, atomistic modeling the protein structure is now possible.[178] In this chapter, we calculate electric fields in $Na_v1.7$ to rationalize the specific role of individual residues in gating and Na^+ permeation. Since $Na_v1.7$ is a protein, like enzymes, we anticipate preorganized electric fields, oriented to facilitate the motion of S4 and Na^+ permeation. We use two types of probe to compute the electric fields, in the PD and VSD, respectively. More specifically, we calculate the electric fields (i) projected onto the oxygen of water molecules at the center of SF and IG (Figure 4.3) and (ii) projected along the S4 helix of each domain (Figure 4.4). To validate the electric field results is not simply caused by the distance of residues to the probe, but a true reflection

of their role in $Na_v1.7$ mechanism, we analyzed the electric field projection as a function of probe-residue C_α distance. The result can be found in Supporting Information section 4.

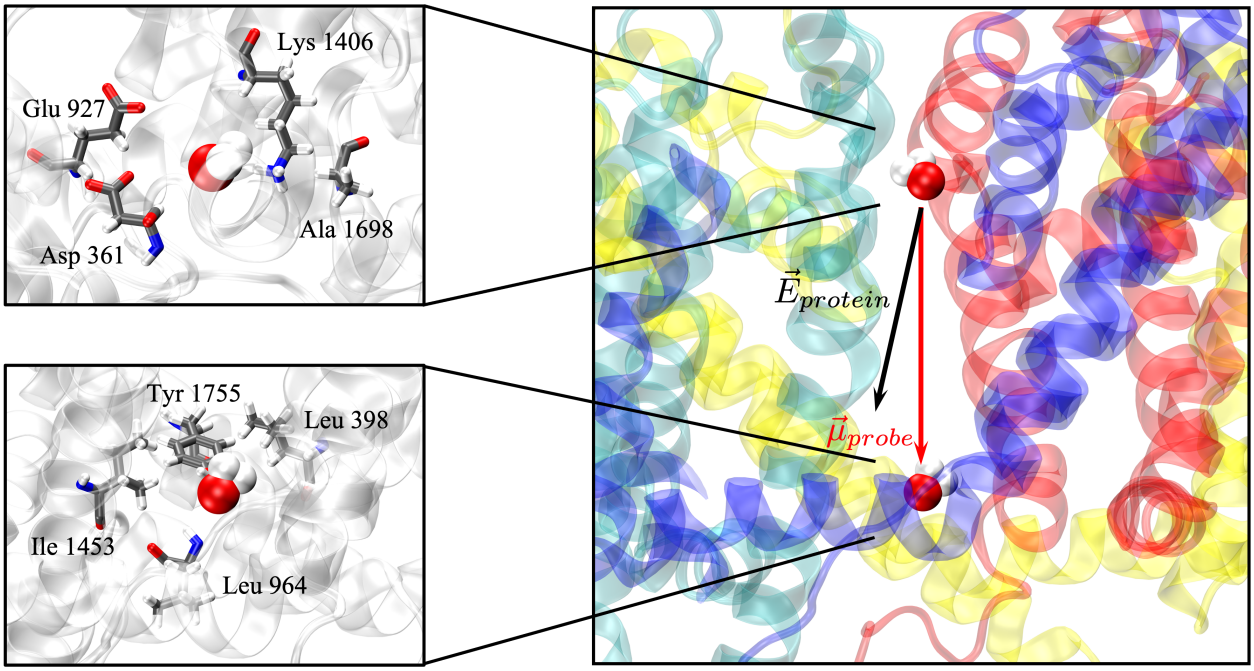

Figure 4.3: Water molecules are located at the center of selectivity filter (SF) and intracellular gate, where oxygen. Electric fields generated by each component of the system ($\vec{E}_{protein}$) are projected onto the unitary vector defining the permeation path ($\vec{u}_{probe}$). Residues forming SF (top left) and IG (bottom left) are annotated.

4.2 Computational Details

4.2.1 Structure preparation

The structure of $Na_v1.7$'s α subunit in complex with two auxiliary β subunits was reported by Shen et al. (PDB ID: 6j8i).[178] The original α subunit structure is the Na_v in inactivated state with a mutation ($Glu^{406}\rightarrow Lys$, or E406K) in residue 406. The N-terminal 113 residues (1—113), DI-DII linker (residue 418—725), DII-DIII linker (residue 973—1174), C-terminal

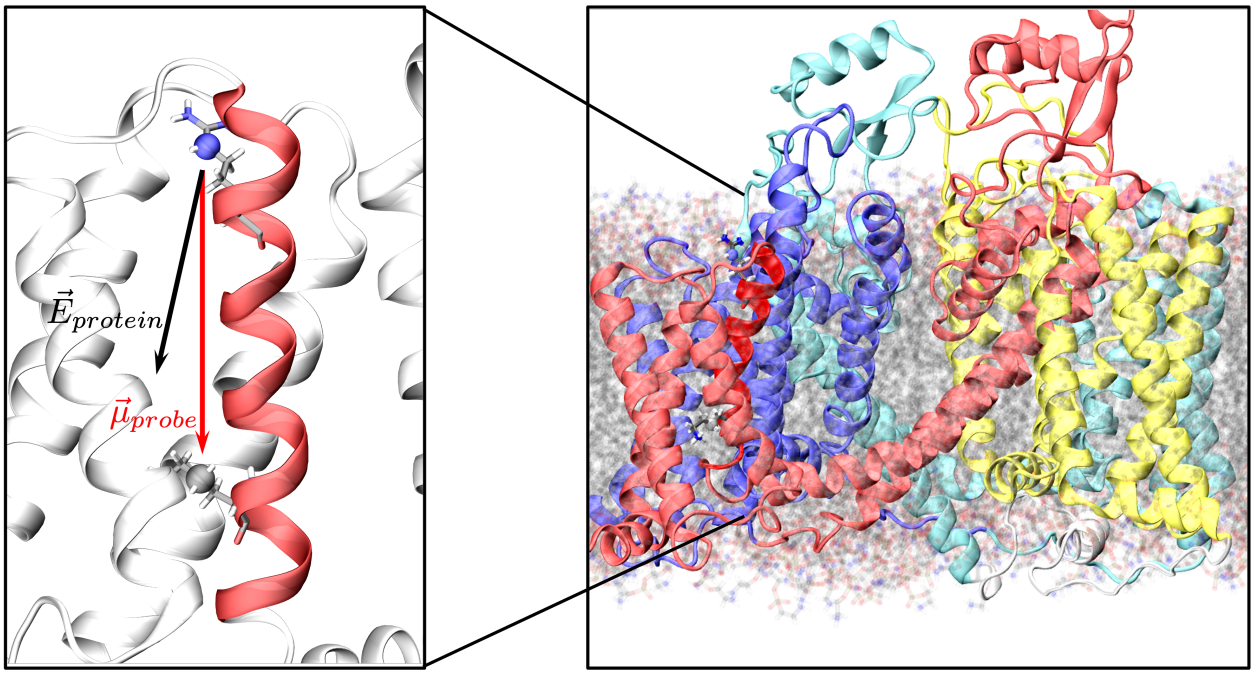

Figure 4.4: Demonstration of the VSD probe of Domain I is shown on the left. Electric fields generated by each component of the system ($\vec{E}_{protein}$) are projected onto the unitary vector defining the direction of voltage-sensing on segment 4 of the VSD ($\vec{u}_{probe}$)

.

residues (1769—1988) and 5 residues on L3-4 loop of VSDII (826—830) are missing in the reported structure of α subunit.

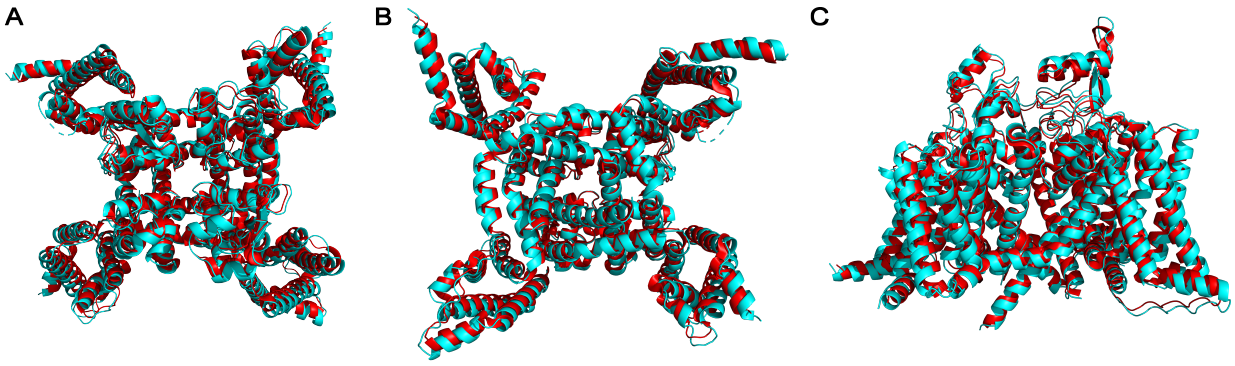

Figure 4.5: Comparison of α subunit structure from the Protein Data Bank (red) and SWISS-MODEL (cyan): A. top view; B. bottom view; and C. side view.[179] The RMSD for atoms in the two structures is 1.614 Å.

The starting structure of our study is the α subunit remodeled with SWISS-MODEL web-server.[180–184] The remodeled structure added residue 826—830 that were missing in the original structure. Although SWISS-MODEL was able to add residue 826—830, in silico prediction methods decrease in folding modeling accuracy for residue numbers higher than 12 to 13.[185] Therefore, no other missing residues were added to our $Na_v1.7$ structure. No significant conformational differences were found between the remodeled structure and the original structure (Figure 4.5).

4.2.2 Conformational Ensemble

The conformation ensemble was generated using Rosetta molecular modeling program.[186] The Rosetta's Backrub algorithm was used first to sample backbone conformations. And then the Fixbb algorithm was used to repack and minimize the side chains around the new backbone conformations. This routine allowed the protein backbone to be rotated around $C\alpha$ axis, then explore the side-chain conformations otherwise occurred on time scale up to

microseconds. The 50 conformers with the lowest energy were generated from this procedure and used for the following simulations. The simulation box dimension was $143 \times 143 \times 133$ $Å^3$.

4.2.3 Molecular Dynamics

The $Na_v1.7$ was embedded in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) membrane bilayer. The 50 conformers were solvated in 0.15M NaCl solution using the Membrane Builder of CHARMM-GUI webserver.[187–193] The system was first relaxed with NVT MD equilibrations (375 ps with 1 fs timestep followed by 1500 ps with 2 fs timestep at 303.15 K) and NPT MD productions (90 ns with 2 fs timestep at 303.15 K and 1 bar) were performed with GROMACS simulation package7 with CHARMM36m forcefield.[194] The last timestep of each simulation was used as input for NPT MD production (300 ps with 1 fs timestep at 300K and 1 atm) using Tinker software package with AMOEBA polarizable force field.[195,196] In total our simulation gave a 15 ns long MD trajtory. The AOMEBA parameters for POPE lipids are reported by Chu et al.[197]

4.2.4 Electric Fields Quantification

Electric fields contributed by each component of the system is calculated as its projection onto a probe (Equation 2.20). The calculation was made possible by the ELECTRIC code.[107] The projected electric fields from individual residues are averaged over the conformation ensemble and 250 ps with 50 ps equilibration. The time length of equilibration was based on the energy convergence. The time evolution of the RMSD (Figure S10) and energy (Figure S9) can be found in Supporting Information.

4.3 Results and Discussion

4.3.1 Pore Domain

We investigated the contribution of each residue to the electric field projected onto the permeation pathway in the pore domain, as shown in Figure 4.6. The PD probes were defined such that the direction points down, towards the inside of the cell membrane. Therefore, a positive projected electric field indicates that the field is oriented from SF to IG. Since Na^+ moves in the direction of the electric field, a positive projection indicates that Na^+ is favored to move from SF to IG as well.

We observe that a specific region of the channel plays a significant role in the projected electric fields (i.e, we do not see a uniform distribution of contributions). 1.22% of the residues contributed to 30.4% of the absolute electric fields of the protein ($\vec{E}^{abs}_{protein} = \sum \vec{E}^{abs}_{res_i}$). Now considering the direction of the projection of the electric field, these 1.22% of residues contributed to -7.54 MV/cm, while the whole protein structure gave -1.15 MV/cm. There are two main contributions to the fields: contribution from the SF area and the IG area. At the SF area, the contributing residues located on top of the probe had the electric field contribution of -15.7 MV/cm. Therefore the negative contributions indicate the residues' role in strongly promoting the ion movement toward these SF residues. Gln 360 and Phe 1405 sit below the SF residues with opposite electric field signs. The +4.2 MV/cm from Gln 360 suggest a role of regulating ion along the permeation path, while the -3.3 MV/cm from Phe 1405 has the opposite effect. The rest of the residues are located in IG area, with the sum of electric field contribution being +9.1 MV/cm. These residues together favors the ion migration towards the IG residues.

The residues with contributions of more than 3 MV/cm to the projected electric fields are

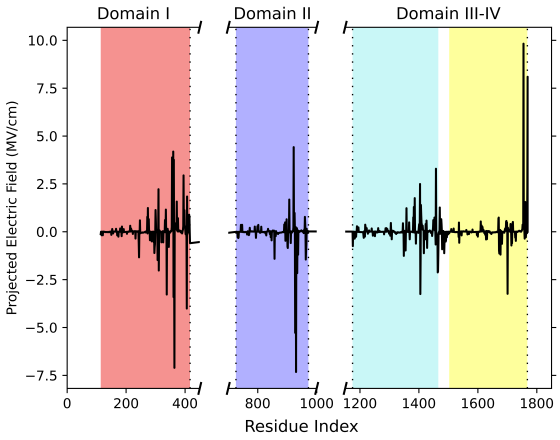

Figure 4.6: Contribution of individual residue to electric fields projected onto the ion permeation path. Each domain is coloured respectively (red, blue, cyan, and yellow). Residue 1 to 1768 correspond to $Na_v1.7$, residue 1769 represents the solvent (water and ions), and residue 1770 represent the POPE membrane.

summarized in Table 4.1. Most of these residues are located on P-loops, the pole loop structure connecting S5 and S6. Previous literature has shown the importance of P-loops for channel selectivity.[198–200] Our calculations suggested the specific residues that bind Na^+: Asp-361 and Glu-927 of the selectivity filter contributed -3.408 and -5.276 MV/cm to the electric field projection, respectively. In addition to these well-studied DEKA residues, the outer carboxylate ring residues were also found with a large contribution to the electric field projection.

The outer carboxylate ring structure exit in all nine Na_vs proteins, these residues located above the selectivity filter were understood to play an important role channel selectivity.[201] However, the ring structure in $Na_v1.7$ is different from the other eight Na_vs in the family. It consist of four residues, two glutamic acids (E) and two aspartic acids (D) from the four domains DI to DIV, which is often referred to as the EEDD ring.[202] In $Na_v1.7$, the aspartic acid from DIII is mutated to isoleucine (I), forming a EEID ring.[202] The three EED residues in $Na_v1.7$ showed a strong attractive interaction with Na^+ in the pore, with Glu-

364 contributing -7.1 MV/cm, Glu-930 contributing -7.3 MV/cm and Asp-1701 contributing -3.2 MV/cm. The conformation of the inactivated and open channels are very similar in the selectivity filter region, suggesting that the main function of the EEID ring in both of the channel state is to direct ion flow toward themselves. However, whether the EEID ring is selective towards Na^+ cannot be determined based on our current results, as it requires further investigation of the electric fields projections under different ion environment.

Opposite effects were observed for Arg 356, Gln 360, Tyr 362, Arg 922, Asp 1458, and Tyr 1755, as they positively contribute to the electric field projection onto the permeation axis. As can be seen in Figure 4.7, Arg 356, Gln 360, Tyr 362 and Arg 922 are all located around the SF and its outer EEID ring, implying that their role is to facilitate ion currents from SF to IG inside the channel pore (i.e., not binding). Asp 1458 and Tyr 1755 are located near the inside of the cell membrane, with Tyr 1755 being one of the residues that form the intracellular gate. Their positive contribution favors ion movement into the cell, suggesting that once the ion has passed the selectivity filter, the protein structure mainly facilitates the uptake of Na^+ ions. 11 of the significant contributions are located near the SF, while only 3 are from the IG region. This suggests that the residues in the SF are a lot more active than their counterpart in the IG.

The $Na_v1.7$ structure we studied is a E406K mutant from the wild type protein, prolonging the inactivated state of $Na_v1.7$. This is one of the disease mutation that lead to primary erythermalgia, a pain syndrome caused by a hyperpolarization ship in for activation.[178] Lys 406 is located in the IG region with Asp 1458 and Tyr 1755. While the contribution to the electric field projection from Asp 1458 and Tyr 1755 were positive, the contribution from Lys 406 was negative. This suggests Lys 406 favors the ion movement from IG to SF, a reverse direction than the regular permeation path, relating to the mutation to extend the inactivation period.

Finally, there is an apparent asymmetry between the domains contributing to the projection of the electric fields as there are more significant residues in DI. Among the residues with positive contribution, those near the SF belong to DI and DII, while those near the exit of the $Na_v1.7$ pore belong to DII and DIV. This means that the asymmetric domains are each involved in different steps of Na^+ uptake and cooperate for a complete permeation mechanism.

Table 4.1: Significant residues identified using projected electric fields of the pore domain probe. Solvent includes water, Na^+, and Cl^-. The error is calculated as the error of the mean over 300 ps and 50 ensemble averages. The standard deviation (σ) describe the electric field fluctuation through time.

PD probe				
Residue	E^{ik}_{proj} (MV/cm)	σ (MV/cm)	Location	Functionality
Asp 338	-3.3 ± 0.017	0.8	DI, Eβ1b—P1	
Arg 356	3.9 ± 0.032	1.5	DI, P—loops	Conserved residue, Inv. amino acid
Gln 360	4.2 ± 0.097	4.6	DI, P—loops	Conserved residue, Invariant amino acid
Asp 361	-3.4 ± 0.109	5.2	DI, P—loops	**DEKA**, Conserved residue, Invariant amino acid
Tyr 362	3.8 ± 0.125	5.9	DI, P—loops	
Glu 364	-7.1 ± 0.118	5.6	DI, P—loops	**EEID**,Conserved residue, Invariant amino acid
Lys 406	-4.0 ± 0.030	1.4	DI, S6	Conserved residue, Invariant amino acid
Arg 922	4.4 ± 0.035	1.7	DII, P—loops	Conserved residue
Glu 927	-5.3 ± 0.214	10.1	DII, P—loops	**DEKA**, Conserved residue, Invariant amino acid
Glu 930	-7.3 ± 0.071	3.4	DII, P—loops	**EEID**, Conserved residue, Invariant amino acid
Phe 1405	-3.3 ± 0.107	5.1	DIII, P—loops	Invariant amino acid
Asp 1458	3.3 ± 0.022	1.0	DIII, S6	Conserved residue, Invariant amino acid
Asp 1701	-3.2 ± 0.019	1.0	DIV, P—loops	**EEID**, Conserved residue, Invariant amino acid
Tyr 1755	9.8 ± 0.266	12.6	DIV, S6	**Intracellular gate**, Conserved residue, Invariant amino acid
solvent	8.1 ± 0.140	6.6		
membrane	6.3 ± 0.001	0.033		

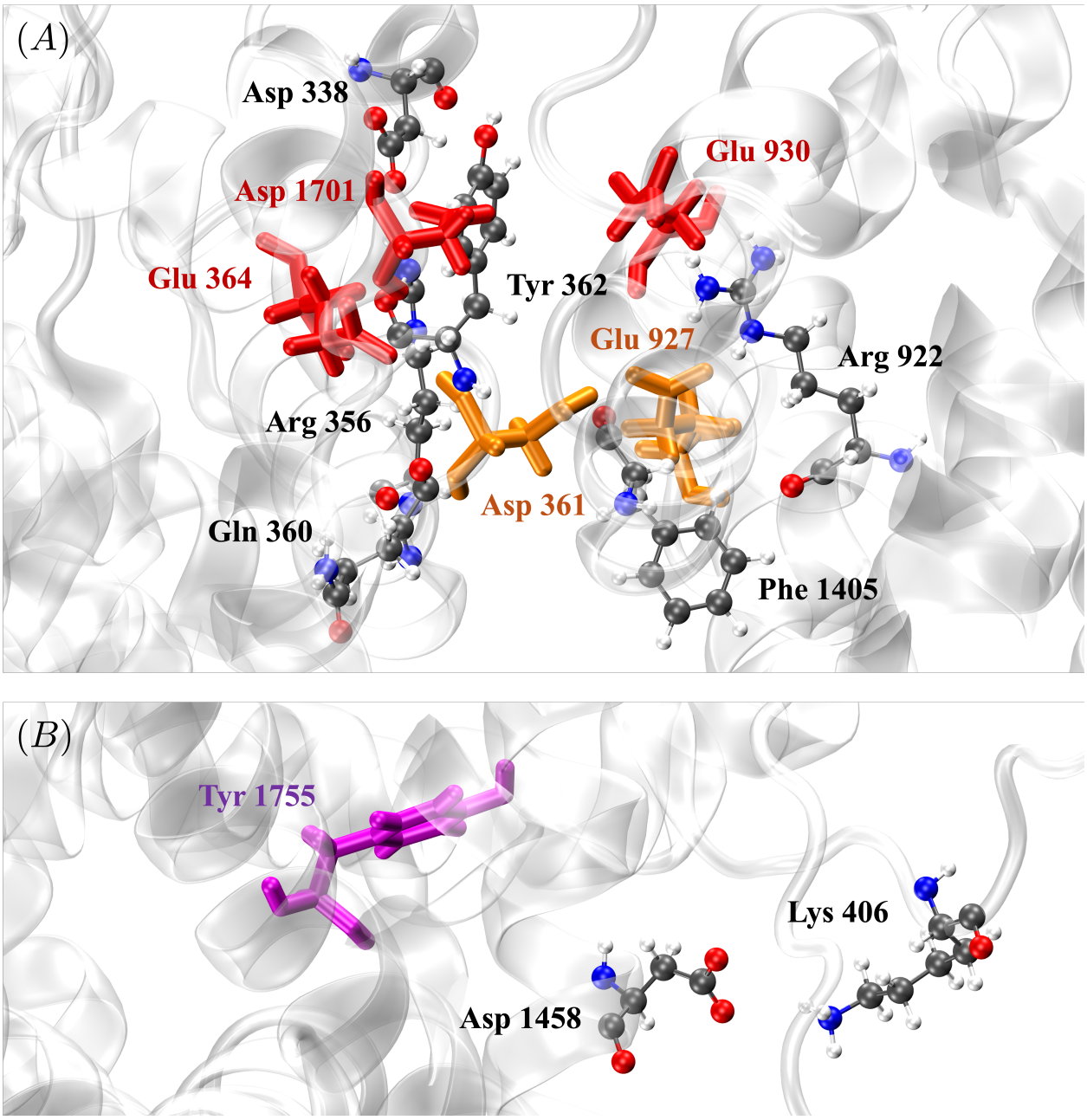

Figure 4.7: Residues with high contribution to the electric field projection. Residues belong to selectivity filter are colored in orange, the outer ring (EEID) in red, intracelluar gate in purple. (A) The selectivity filter region. (B) The intracelluar gate region.

Residues that are invariant across homologous proteins or conserved through evolution usually have structural or functional significance to the protein.[203–205] However, notice that residues Asp 338 and Tyr 362 have been identified as significantly contributing despite not being conserved or invariant, indicating their unique role in $Na_v1.7$. Their positions are separated by the outer ring, with Asp 338 at the top and Tyr 362 at the bottom of the ring structure. Their electric field projections also have different signs, with Asp 338 acting with the outer ring and Tyr 362 opposite. This suggests that when the two residues involved in regulating ion motion, the ions are regulated toward the EEID ring.

The solvent, including the ions and the water, is an important contributor to the projected electric fields. On average, there are 8 water molecules that significantly contribute to the electric field projections. They are mostly located around SF or IG and all within the channel pore. However, some conformers have significantly fewer contributing water molecules than others. The side chain of the IG residues and the residues around them are directed into the pore so that fewer water molecules can occupy the space inside. The membrane also plays an important role in the electric fields generated by the protein when considering the ion permeation pathway. The positive contributions of solvent and membrane indicate that they facilitate the transport of sodium ions from the SF to the IG.

4.3.2 Voltage-sensing Domain

Figure 4.8 shows the contribution of the channel residues to the electric fields projected along the S4 helix in each domain. The probes were chosen such that positive projections points towards the inside of the membrane. Therefore, a negative magnitude of the electric field indicates that the corresponding residue favors the upward movement of S4, towards the outside of the cell. We observe a similar trend for all four probes: the main contributions

come from residues within the same domain. When probing DI, residues from DII can also show contributions to the electric field projections, similarly when probing other domains. The explanation for these seemingly distant intermolecular interactions between residues is the conformation of the $Na_v1.7$. Indeed, $Na_v1.7$ has swapped arrangements of the S1-S4 helices and S5-S6 helices of each domain such that S1-S4 and S5-S6 are connected but spatially separated from each other. The VSD formed by S1-S4 is located on the outside of the channel, while the PD formed by S5-S6 extends to the inside of the channel. As shown in Figure 4.2 B, the PD of one domain is spatially closer to the VSD of the neighboring domains. This indicates the two parts of channel function (i.e. ion permeation and voltage sensing) is not isolated, as VSD is under the influence of PD. The four domains showed different total electric field projections with$\vec{E}^{DI}_{protein}$ = +3.79 MV/cm, $\vec{E}^{DII}_{protein}$ = -34.68 MV/cm, $\vec{E}^{DIII}_{protein}$ = -22.32 MV/cm, $\vec{E}^{DIV}_{protein}$ = -67.39 MV/cm. Domain I plays an overall supporting role for the inward S4 movement, while domain IV is calculated to have a strong preference for the opposite direction. From the literature, it appears that the S4 segment performs an outward movement due to the depolarization process. Combined with our calculations, we can conclude that the response of DI is more significant in inactivation compared to DII to DIV, while DIV is the crucial domain for the open state function $Na_v1.7$ function. This asymmetry in the domain function is consistent with our results in the study of the pore domain.

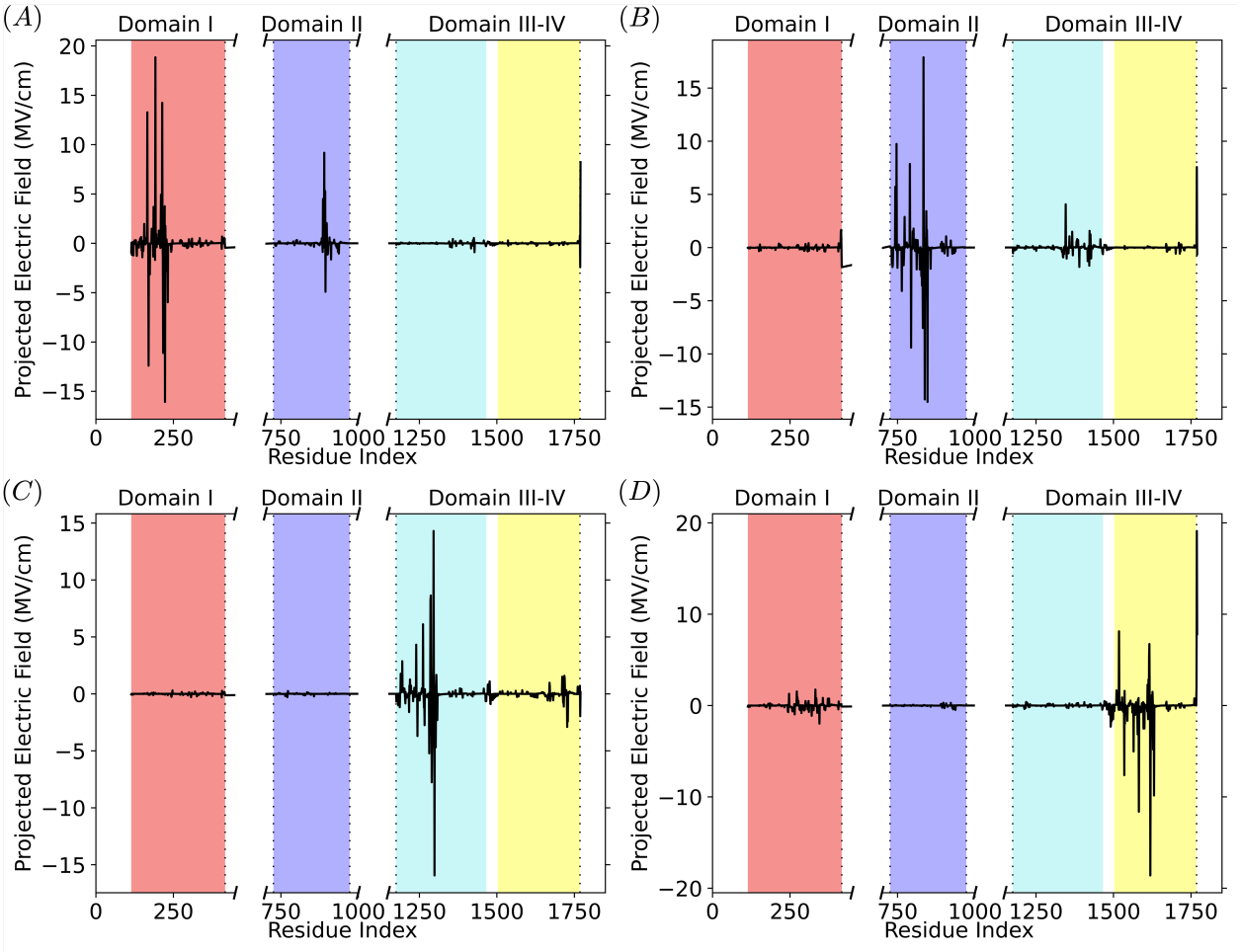

Figure 4.8: Contribution of individual residue to electric fields projected onto S4 helix of 4 different domains, (A) to (D). Residues located at each domain are represented by red, blue, cyan, and yellow background colors. The last 2 residues are solvent (water and counter ions) 1769 and membrane (POPE) 1770.

Table 4.2: Significant residues identified using projected electric fields of the voltage sensing domain probe (DI). Solvent includes water, Na^+, and Cl^-. The error is calculated as the error of the mean over 300 ps and 50 ensemble averages. The standard deviation (σ) describe the electric field fluctuation through time.

VSD probe DI				
Residue	E^{ik}_{proj} (MV/cm)	σ (MV/cm)	Location	Functionality
Glu 166	13.3 ± 0.033	1.6	DI, S2	Conserved residue, May facilitate gating charge transfer
Lys 170	-12.4 ± 0.026	1.2	DI, S2	Conserved residue, Invariant amino acid
Arg 174	-3.2 ± 0.006	0.3	DI, S2	Conserved residue, Invariant amino acid
Asp 186	3.7 ± 0.007	0.3	DI, S2-3—S3	Conserved residue, Invariant amino acid
Asp 192	18.8 ± 0.081	3.8	DI, S3	Conserved residue, May facilitate gating charge transfer
Val 210	4.9 ± 0.026	1.2	DI, S4	Conserved residues
Arg 214	14.2 ± 0.129	6.1	DI, S4	**Gating charge residue**
Arg 217	-11.1 ± 0.041	1.9	DI, S4	**Gating charge residue**
Leu 219	-3.7 ± 0.023	1.1	DI, S4	Conserved residue
Arg 220	-3.7 ± 0.057	2.7	DI, S4	**Gating charge residue**
Leu 222	3.8 ± 0.016	0.8	DI, S4	Conserved residue, Invariant amino acid
Lys 223	-16.1 ± 0.118	5.6	DI, S4	**Gating charge residue**
Lys 232	-6.0 ± 0.014	0.7	DI, S4-5	Conserved residue, Invariant amino acid
Gln 886	4.5 ± 0.063	3.0	DII, S5	Conserved residue, Invariant amino acid
Lys 890	9.2 ± 0.035	1.6	DII, S5	Conserved residue
Lys 893	5.3 ± 0.019	0.9	DII, S5	Conserved residue
Glu 894	-4.9 ± 0.016	0.7	DII, S5	Conserved residue
membrane	8.2 ± 0.002	0.1		

Table 4.3: Significant residues identified using projected electric fields of the voltage sensing domain probe (DII). Solvent includes water, Na^+, and Cl^-. The error is calculated as the error of the mean over 300 ps and 50 ensemble averages. The standard deviation (σ) describe the electric field fluctuation through time.

VSD probe DII				
Residue	E^{ik}_{proj} (MV/cm)	σ (MV/cm)	Location	Functionality
Asp 743	5.7 ± 0.012	0.6	DII, S0—S1	Conserved residue, Invariant amino acid
Asp 747	9.8 ± 0.024	1.1	DII, S1	Conserved residue, May facilitate gating charge transfer
Glu 764	-4.1 ± 0.013	0.6	DII, S1—S2	Conserved residue, May facilitate gating charge transfer
Glu 790	7.9 ± 0.027	1.2	DII, S2	Conserved residue, May facilitate gating charge transfer
Lys 794	-9.4 ± 0.024	1.1	DII, S2	Invariant amino acid
Glu 829	-3.5 ± 0.009	0.4	DII, S3—S4	Conserved residue
Ser 832	-7.5 ± 0.042	2.0	DII, S4	Conserved residue
Leu 834	17.9 ± 0.051	2.4	DII, S4	Conserved residue, invariant amino acid
Arg 838	-14.3 ± 0.070	3.3	DII, S4	**Gating charge residue**
Leu 839	-5.3 ± 0.029	1.4	DII, S4	Conserved residue
Phe 843	3.4 ± 0.014	0.7	DII, S4	Conserved residue, invariant amino acid
Lys 847	-14.5 ± 0.050	2.4	DII, S4	**Gating charge residue**
Lys 1346	4.1 ± 0.005	0.2	DIII, S5—Eβ3a	Conserved residue, invariant amino acid
solvent	7.6 ± 0.009	0.4		

Table 4.4: Significant residues identified using projected electric fields of the voltage sensing domain probe (DIII). Solvent includes water, Na^+, and Cl^-. The error is calculated as the error of the mean over 300 ps and 50 ensemble averages. The standard deviation (σ) describe the electric field fluctuation through time.

VSD probe DIII				
Residue	E^{ik}_{proj} (MV/cm)	σ (MV/cm)	Location	Functionality
Glu 1240	4.3 ± 0.009	0.4	DIII, S2	Conserved residue, May facilitate gating charge transfer
Lys 1244	-3.7 ± 0.007	0.3	DIII, S2	Conserved residue, invariant amino acid
Asp 1262	6.1 ± 0.016	0.7	DIII, S3	Conserved residue, Invariant amino acid
Asp 1282	-5.2 ± 0.018	0.9	DIII, S3—S4	
Ile 1286	8.3 ± 0.040	1.9	DIII, S4	Conserved residue
Lys 1287	9.1 ± 0.074	3.5	DIII, S4	**Gating charge residue**
Arg 1290	-7.9 ± 0.113	5.4	DIII, S4	**Gating charge residue**
Thr 1291	-5.0 ± 0.077	3.7	DIII, S4	Conserved residue, Invariant amino acid
Leu 1292	-3.6 ± 0.036	1.7	DIII, S4	Conserved residue, Invariant amino acid
Arg 1296	14.3 ± 0.152	7.2	DIII, S4	**Gating charge residue**
Arg 1299	-15.9 ± 0.068	3.2	DIII, S4	**Gating charge residue**
Arg 1303	-4.7 ± 0.008	0.4	DIII, S4	
Glu 1727	-2.9 ± 0.008	0.4	DIV, P2—S6	

Table 4.5: Significant residues identified using projected electric fields of the voltage sensing domain probe (DIV). Solvent includes water, Na^+, and Cl^-. The error is calculated as the error of the mean over 300 ps and 50 ensemble averages. The standard deviation (σ) describe the electric field fluctuation through time.

VSD probe DIV				
Residue	E^{ik}_{proj} (MV/cm)	σ (MV/cm)	Location	Functionality
Asp 1518	8.1 ± 0.043	2.0	DIV, S1	May facilitate gating charge transfer
Glu 1535	-7.6 ± 0.012	0.6	DIV, S1	May facilitate gating charge transfer, conserved residue
Lys 1565	-5.0 ± 0.030	1.4	DIV, S2	Conserved residue, Invariant amino acid
Trp 1578	-3.2 ± 0.026	1.3	DIV, S3	Conserved residue, Invariant amino acid
Asp 1582	-11.6 ± 0.051	2.4	DIV, S3	May facilitate gating charge transfer
Ala 1615	5.4 ± 0.019	0.9	DIV, S4	Conserved residue, Invariant amino acid
Arg 1616	6.7 ± 0.057	2.7	DIV, S4	**Gating charge residue**
Arg 1619	-18.6 ± 0.161	7.6	DIV, S4	**Gating charge residue**
Ile 1620	-5.5 ± 0.051	2.4	DIV, S4	**Gating charge residue**
Lys 1628	-4.8 ± 0.014	0.7	DIV, S4—S5	Conserved residue
Arg 1631	-9.8 ± 0.030	1.4	DIV, S4 — S5	Conserved residue, Invariant amino acid
solvent	19.1 ± 0.011	0.5		
membrane	7.8 ± 0.002	0.1		

Tables 4.2-4.5 list the residues with over 3 MV/cm magnitude contributions to the electric fields projections when investigating the VSD. Similar to the PD, more residues belong to DI (17) than the other three domains (13 in DII and DIV, 12 in DIII). The sum of their contributions also suggested the opposite role between DI and DIV, contributed +24.95 MV/cm and -19.00 MV/cm, respectively. While 46% of the significant contributors are located along S4 helices, we also observed contributions from S2 and S3 in each domain, indicating significant interactions between S4 and S2-S3. These residues are mostly close to the bottom of the channel protein. The literature suggests that while S4 move across the membrane in the mechanism of voltage sensing, residues in S2 and S3 form intermolecular interactions with the gating charge residues on S4 to facilitate this mechanism.[171,206,207] Previously we discussed that total electric field projections in DI favors the inactivated state S4 conformations, while DIV favors the open state. Here the electric field contributions from S2 and S3 residues from each domains aligned with the total electric field, as these residues in DI had a overall positive contribution (promoting downward movement of S4), while those in DIV had a overall negative contributions (promoting upward movement of S4). Additionally, mutations of Leu 834 (+17.91 MV/cm) or Gln 886 (+4.50 MV/cm), both key contributors, were found causing paroxysmal pain syndrome primary erythermalgia, a neural disorder caused by hyperactive $Na_v1.7$.[208,209] Our calculation revealed their contribution to the electric field projection is against the upward movement of S4 upon open state. When mutated, these two residues lost their function to suppress S4 conformation transition to open sate, resulting in the hyperactive behavior. While no diseases is found caused by mutation of Arg 1303 currently, its corresponding residue Arg 1319 when mapping onto $hNa_v1.2$ is one of the cause of seizures (benign familial infantile seizures) when mutated.[208] Interestingly, Arg 1303 was also one of the two residues we identified as important electric field projection contributor but not an invariant or conserved residue. As motioned in previous section, this suggests the two residues are not preserved through all nine homologs. Additional cross

comparison between Na_vs homologs on such residues may reveal their functional shift upon mutation.

4.3.3 Side-chain Dihedral Angle Effect on Projected Electric Fields Fluctuation

In Tables 4.1-4.5 we present the fluctuations of the electric fields, computed as the standard deviation. We observe a trend that residues with high standard deviations hold crucial role for $Na_v1.7$. For example, gating charge residues Arg 214 and Arg 1926 have a standard deviation over 6 MV/cm, and Glu 927 (σ = 10 MV/cm) and Tyr 1755 (σ = 12.6 MV/cm) each belong to the gating motif of the permeation path (i.e. SF and IG).

Amino acid rotamers (i.e., side chain conformations) have been extensively studied to characterize protein conformational properties for protein structure prediction or design.[156,210,211] Each rotamer is characterized by five side chain torsions, called χ_1 to χ_5. Here, we analyze the time evolution of these torsions for residues that exhibit large electric field fluctuations. We show in Figure 4.9 the evolution of χ_x (where $x = 1, 2, 3, 4, 5$) and the projected electric fields. In the four examples we present, χ_5 was stable, χ_1 and χ_3 were stable except for Lys 233, but χ_2 and χ_4 exhibit significant change in all cases. More importantly, an increase or decrease in the magnitude of the projected electric field when a rotation of 60° or more happened. For Lys 223, the electric field projection showed a gradual increase between 170 ps to 220 ps. During the same time period, χ_1 χ_3 χ_4 changed over 120°. This relation was also presented in Arg 1296, where a 40 MV/cm difference in electric field projection took place at the same time as the side-chain conformation change. In Arg 214 and Arg 1290, the correlation is not as obvious as in Lys 223 and Arg 1296, probably due to the y-axis scaling. In Figure 4.10 (A) and (B) we show the same configuration of Arg 214 and Arg 1290 with

expanded y axis. Rotation of Arg 214 χ_2 from -180° to -60° and χ_4 from 180° to 90° resulted in the increase of average electric field from 35 MV/cm to 45 MV/cm. In Arg 1290, a 10 MV/cm change in electric field magnitude was also connected to a rotation of the side-chain. When the χ angle recover to it starting position, the electric field magnitude also increased back to the original average.

Similar observations can be made for residues located in the PD (Figure 4.11). In Glu 927, χ_1 and χ_3 both exhibit significant changes at the 210 ps time mark. The electric field projection also drastically changed at that time. In Tyr 1755, however, the side chain conformation change was more gradual, with a smaller electric field fluctuation magnitude comparing to the other example we studied.

Overall, our results indicate that electric fields fluctuations can be used to quantify key rotamer changes. It is important for protein modeling because proteins are known to be flexible, and the changes in structure indicate a change in activity.

4.4 Summary

Electric fields projections provide a map between protein structure and function. In particular, we showed that inactivated $Na_v1.7$ generate preorganized electric fields that support Na^+ binding in the selectivity filter and Na^+ permeation through the pore. Residues around the selectivity filter were found to promote ion migration towards the SF and the outer carboxylate ring. Significantly less residues were found active near the IG, reflecting a inactivated state behavior.

Further, we showed that the motion of S4 in each domain is also supported by the electric field generated by $Na_v1.7$. Specifically, DI residues favored the inactivated state (i.e., suppressing

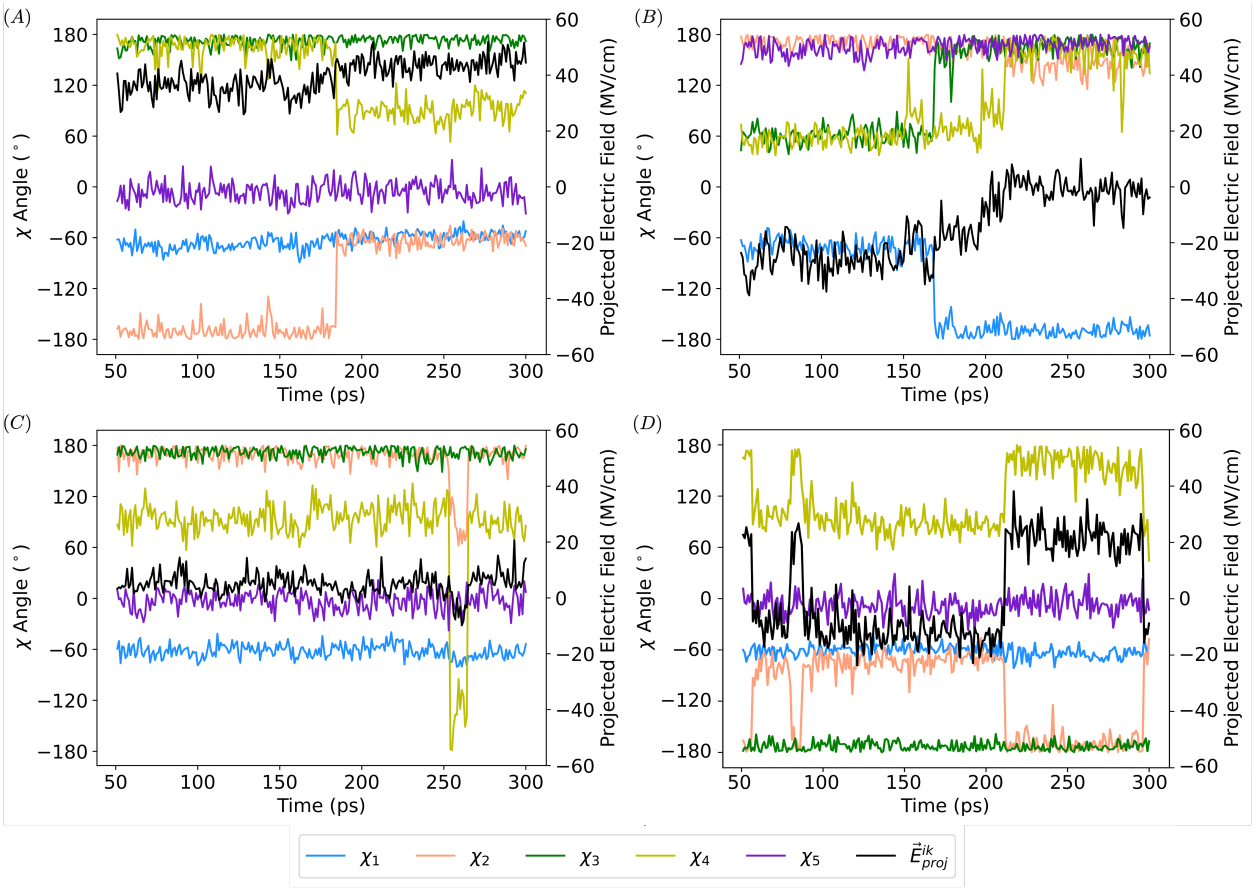

Figure 4.9: The time evolution of the projected electric fields (in black) and the residue χ angles in VSD. χ_1 to χ_5 angles are colored blue, peach, green, yellow, and purple. (A) Arg 214, (B) Lys 223, (C) Arg 1290, (D) Arg 1296. The first 50 ps of the simulation were consider for equilibration, so the production phase started from 51 ps.

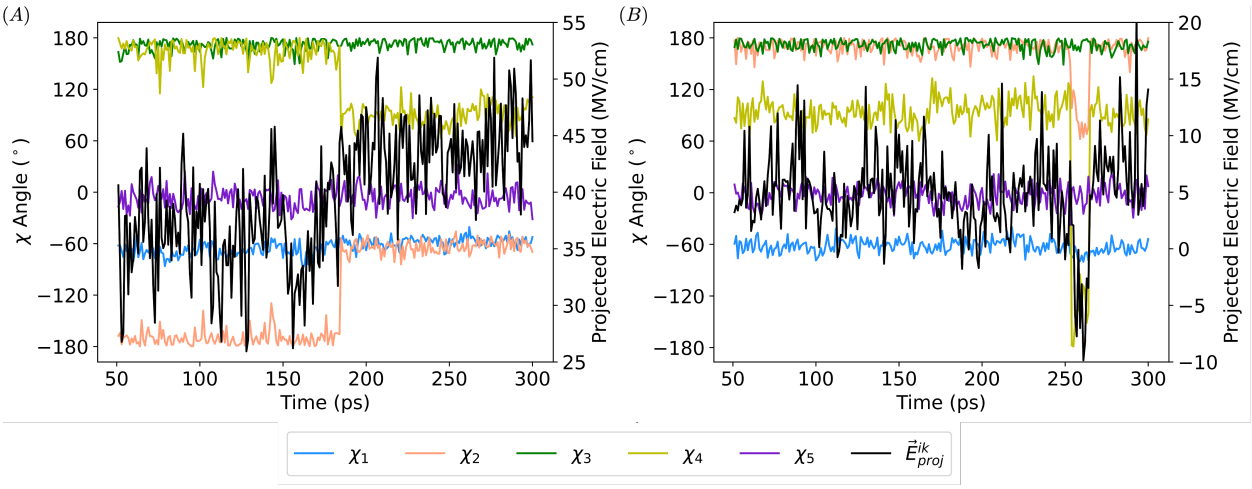

Figure 4.10: The time evolution of the projected electric fields (in black) with exapnded y axis and the residue χ angles in VSD. χ_1 to χ_5 angles are colored blue, peach, green, yellow, and purple. (A) Arg 214, (B) Arg 1290. The first 50 ps of the simulation were consider for equilibration, so the production phase started from 51 ps.

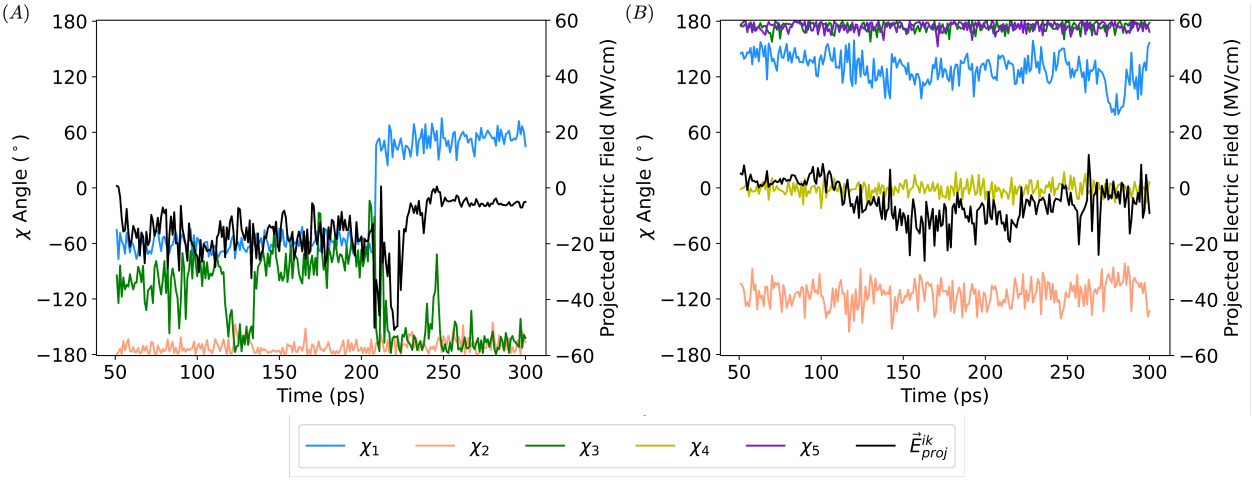

Figure 4.11: The time evolution of the projected electric fields (in black) and the residue χ angles in PD. χ_1 to χ_5 angles are colored blue, peach, green, yellow, and purple. (A) Glu 927, (B) Tyr 1755. The first 50 ps of the simulation were consider for equilibration, so the production phase started from 51 ps.

the upward movement of S4), while DIV residues facilitated the upward movement of S4. Two disease inducing mutations validate our results.

The strength and direction of the electric fields is related to the residue side-chain conformation. The side-chain rotation results in the variation of interactions between residue, a change captured by our calculations. More than 60 $^{\circ}\chi$ rotations were reflected in the electric field signal.

Overall, our approach provides a valuable guide for studying channel mechanisms, because the magnitude and sign of the electric field projection reveals the role of individual residue in the preorganized channel.

Chapter 5

Loop Dynamics and Interactions Upon Glycosylation of Enzyme DszB

5.1 Background

2'-Hydroxybiphenyl-2-sulfinate desulfinase (DszB) is a bacterial enzyme involved in a four-step pathway, the 4S pathway, that converts dibenzothiophene (DBT) to 2-hydroxybiphenyl (HBP). DszB catalyzes the last and limiting step of this pathway, the desulfurization of 2'-hydroxybiphenyl-2-sulfinic acid (HBPS) to HBP (Figure 5.1). The 4S pathway is widely studied for its application to biodesulfurization, a method to remove sulfur from the organosulfur compounds found in fossil fuels.[212–214] Compared to hydrodesulfurization, biodesulfurization can operate under ambient temperature and pressure, significantly reducing the cost of this process.[214] Moreover, polycyclic organic sulfur compounds such as DBT and its derivatives are much harder to be removed compared to inorganic sulfur compounds.[215,216] Enzymes specifically targeting DBT are therefore a promising alternative to hydrodesulfurization.

The main difficulty in using the 4S pathway as an industrial biodesulfurization process is its slow turnover rate, where researchers identified product (HBP) inhibition of DszB as the main cause.[217–220] Product inhibition is beneficial in biological systems because it serves as negative feedback to control the metabolic pathways.[221–223] However, this is not desired for industrial applications.[109,224] Lee and co-workers reported the conformational shifts that occur in DszB

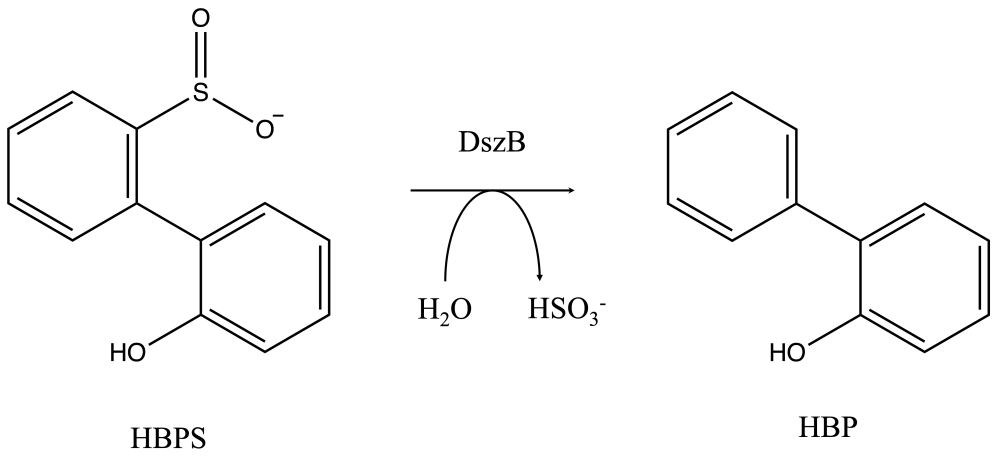

Figure 5.1: Rate limiting step of 4S pathway. The sulfate group cleave from 2'-hydroxybiphenyl-2-sulfinic acid (HBPS), forming 2-hydroxybiphenyl (HBP) product.

upon binding of the ligand.[215,225] Specifically, the 3 loops shown in Figure 5.2 were found to have distinct conformations in the enzyme apo and bound states. These conformational changes cause the trapping of the product in the active site, prohibiting catalyst renewal. The 3 loops exhibit different conformational changes: loop 1 (residues 50 to 60) and loop 3 (residues 180 to 200) adopt an α helix structure in the bound state but lack secondary structure in the apo state. Loop 2 (residues 135 to 150) maintains its α helix structure in both states, but folds inward when the ligand binds. These conformational changes are not substrate specific: when alternative substrates bind to the active site of DszB, identical conformational changes were observed.[215,216,225] This indicates that product inhibition is not specific to the binding of reactant HBPS, but simply a result of the loops conformational changes. Therefore, product inhibition in DszB can be mitigated by regulating the dynamics of these 3 loops. Traditional mitigation methods include mutagenesis on the product binding site. However, this sacrifices the enzyme catalytic performance because product binding site share residues with the substrate binding site.[42,226,227]

Glycosylation describes the protein posttranslational modification process where carbohy-

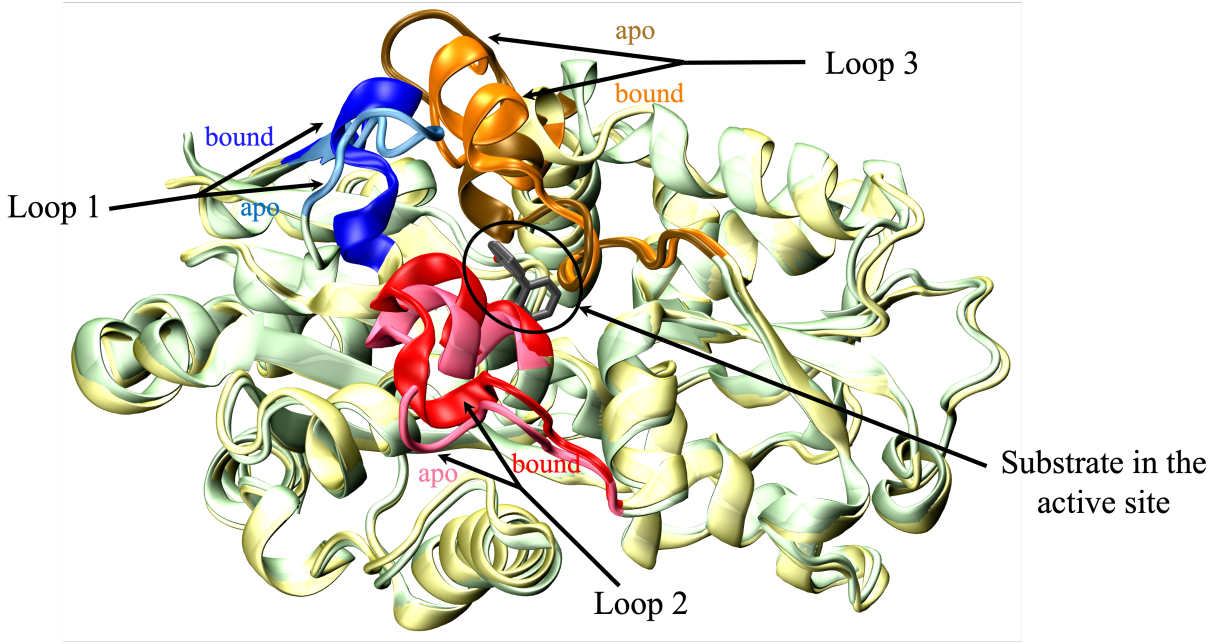

Figure 5.2: Structure overlap of bound (2DE3) and apo states (2DE2) of DszB. The bound state is shown in yellow and the apo state in green. Loop 1 (residues 50 to 60), 2 (residues 135 to 150), and 3 (residues 180 to 200) are the three loops identified as the primary cause of product inhibition, labeled on the figure. The apo state have the active site open, allowing substrate to enter, followed by the conformational change to the bound state. When the reaction is finished, the bound state conformation trapped the product within the active site, resulting in product inhibition of DszB.

drates (glycans) are covalently linked to amino acid side-chains. Protein glycosylation can be categorized by side chain-glycan linkage type. N-linked glycosylation refers to the attachment of glycans to the amide nitrogen of asparagine whereas O-linked glycosylation refers to the attachment of glycans to the hydroxyl oxygen of serine or threonine.[4,228] This naturally occurring post-translational modification plays a crucial role in protein recognition, and provides additional structural stability to the protein folded state through covalent and non-covalent bonds.[229–231] In this chapter, we investigate the role of glycans artificially engineered to the loops of DszB on protein dynamics. Protein glycosylation is one of the natural modifications in increasing protein solubility.[232] Therefore, once glycosylated, solvent exposed structures such as loops on the surface of the protein is subject to change for its movement. Since glycans will be placed outside the active site, they can modulate the loops dynamics without interfering with the chemistry of the reaction (Figure 5.3).

Although not the primary tool for the preliminary analysis of loops dynamics, electric field calculations are crucial for the evaluation of the enzyme catalytic activity, while providing key insight into the role of individual residues. This approach can help us determine any potential trade-off between the mitigated product inhibition and enzyme performance.

5.2 Computational Details

5.2.1 Conformational Ensemble

A total of 6 glucose molecules were covalently attached to the DszB loops with β1-4 linkage, one on each residue. The glycosylation sites on each loop are Ser 53 and Thr 58 for loop 1, Thr 148 for loop 2, Ser 181, Ser 194, and Thr 196 for loop 3. Overall, our six systems were: (I) unglycosylated, wild type DszB protein, denoted DszB, (II) unglycosylated, wild type DszB

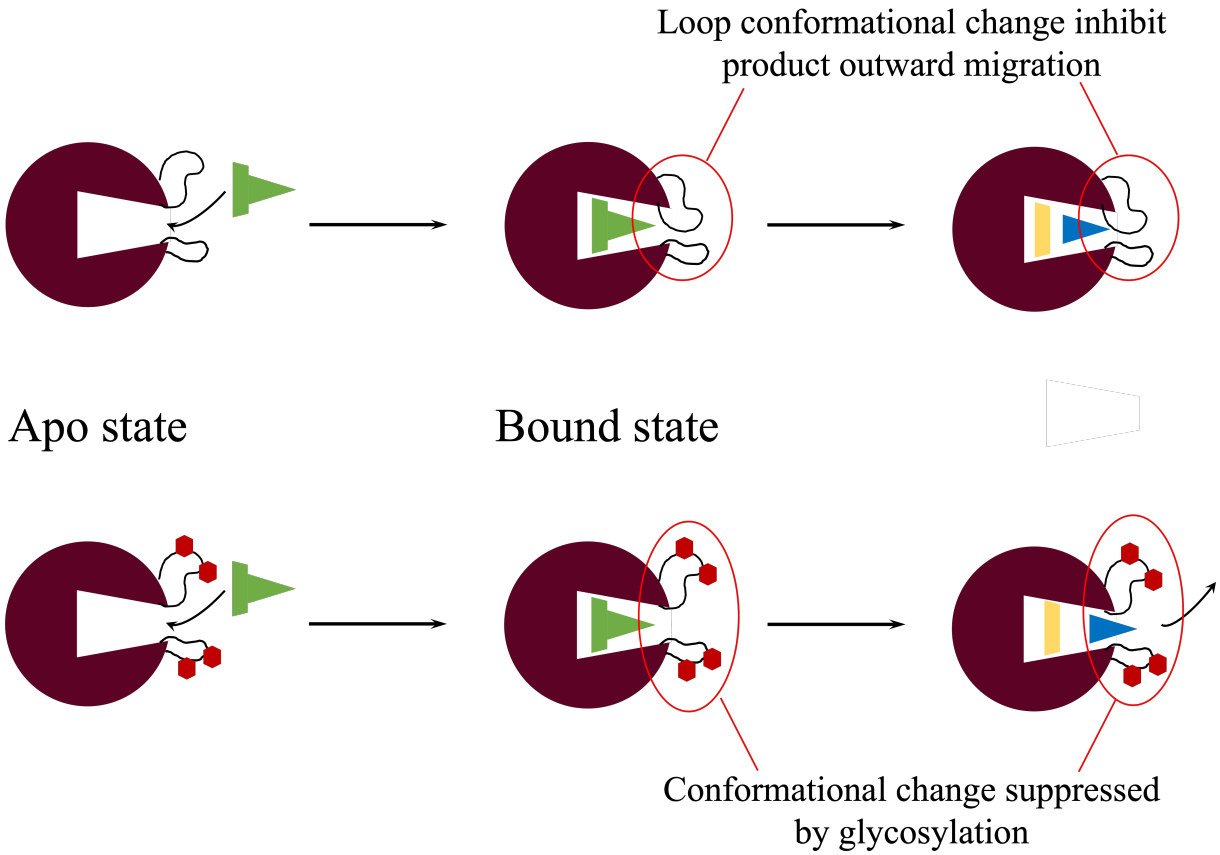

Figure 5.3: Demonstration of glycosylation mitigated product inhibition. Reactant enter the active site at apo state, follow by the change of DszB conformation for reaction. When reaction is finished, the loops in wild type conformation prevent product from exiting the active site. Glycosylation changes loop dynamics, open the pathway for product outward migration.

protein with HBP product in the active site, denoted DszB HBP (III) DszB protein with three glucosylated Ser residues, denoted DszB SERglc (IV) DszB protein with three glucosylated Thr residues, denoted DszB THRglc (V) DszB protein with three glucosylated Ser residues with HBP product in the active site, denoted DszB HBP SERglc and (VI) DszB protein with three glucosylated Ser residues with HBP product in the active site, denoted DszB HBP THRglc. The ROSETTA package was used to generate 25 uncorrelated structures for each of the system, where we select the 5 lowest energy conformations.[154–156] ROSETTA Backrub and Fixbb algorithms samples the protein backbone and side-chain systematically by first rotate residues using C_α as axis, then repack the side-chains for minimized energy under new backbone conformations.[154–156] This step created a comformational ensemble equivalent to large time-scale simulations with shorter simulation time requirement. The simulation box size was $76 \times 96 \times 86$ Å^3.

5.2.2 Molecular Dynamics

We used GROMACS to solvate the 5 structures of each system with pre-equilibrated water and ions (Na^+, Cl^-).[161,162] The ion concentrations were set to 0.16 M mimicking the physiological conditions. Molecular dynamic simulations were performed using Tinker with AMOEBA polarizable force field.[160,233] The MD simulations were performed under isothermal-isobaric conditions (1 atm, 300K) with a time step of 1 fs. The 5 structures of each system were simulated for 80 ns including 20 ns for equilibration. The time length of equilibration was based on the RMSD. The time evolution of the RMSD (Figure S11) can be found in Supporting Information. While the ROSETTA algorithm explore conformations up to microseconds time scale, our parallel simulations made a total of 300 ns long MD trajectories for the 6 systems.

Evaluation of loop dynamics include distance measurement and hydrogen bond analysis. In the calculation of distance, the loops and active sites residues were represented by the C_α of the residues, the product was represented by its geometry center. The hydrogen bond was defined as donor (D) acceptor (A) distance smaller than 3 Å, with the D-H$\cdots$A angle less than 30°. Calculation of hydrogen bond was done with VMD.[234] The number of hydrogen bonds was averaged over the 5 structures in each system. Both distance and the number of hydrogen bonds were averaged over the production phase.

5.2.3 Catalytic Activity Evaluation

In future work, the catalytic activity will be analyzed with electric fields. Individual residue's contribution to the projection in active bond C-S provide their role in reactant-protein interactions. The ELECTRIC code will be used to calculate the electric field projection from individual residues using Eq. 2.20 and Eq.2.14.[107]

5.3 Results and Discussion

5.3.1 The Effect of Glycosylation on Reducing Active Site Obstruction

We studied Ser and Thr O-linked glycosylation on the 3 loops in two separate systems, each with three glucoses attached. There was no attempt for N-linked loop residues glycosylation because there are no Asn residues on the 3 loops.

In Figure 5.4 we presented the root mean square fluctuation (RMSF) of the protein backbone (C_α) in each system, a measurement of atom displacement during the production phase. The

loop 1 (Asp 50 to His 60) of DszB wild type without product showed the highest RMSF, while loop 3 (Ser 180 to Ala 200) was slightly higher than 2 (Gln 135 to Val 150). The higher fluctuation of loop 1 and 3 in this case is likely the result of the lack of secondary structure in the apo state. When the product is present in the active site, there is a decrease in RMSF in loop 1, but loop 2 and 3 remain similar. As loop 1 is the shortest loop, interactions between the product and the loops can fix the movement of loop 1 better than loop 2 or 3. Loop 3 experienced the most changes upon glycosylation, also the longest loop among the three. In addition, there are two Ser glycosylation sites on loop 3 and we observed the most significant changes in loop dynamics on Ser glycosylated loop 3. In comparison, Thr has only one glycosylation site on each loop, and Ser has one on the other two loops.

To better analyze the potential active site obstruction, we computed the time evolution of the distance between loops and active site residues (Figure 5.5 to Figure 5.7). This was measured as the average distance between the Cα of the loop residues to the active site residues (Cys 27, Arg 70, Gly 73 and His 60). We see that the introduction of the product in the active site pulled all three loops closer, confirming hindrance to product exit.[215,216,225] His 60 is located on loop 1, a loop that folds inward to introduce His 60 into the active site upon substrate binding. A 2 to 3 Å decrease of the loop 2 and 3-His 60 distance demonstrates the DszB-HBP system experienced a similar conformational change as described in the literature. The introduction of glycans affected these conformational changes, increasing the loop-active site distances. The loop-active site distance increase was most significant in loop 3, where the Ser and Thr glycosylation each led to 2 and 1 Å addition of distance. In loop 1 and 2, the increase in distance was only moderate with a 0.5 Å addition for both Ser and Thr glycosylation. This difference indicates the changes of loop hydrophobicity is a determining factor for the degrees of active site obstruction. 50% of the loop residues have non-polar side chains, and glycosylation increases the polarity of these residues dramatically, reducing their overall

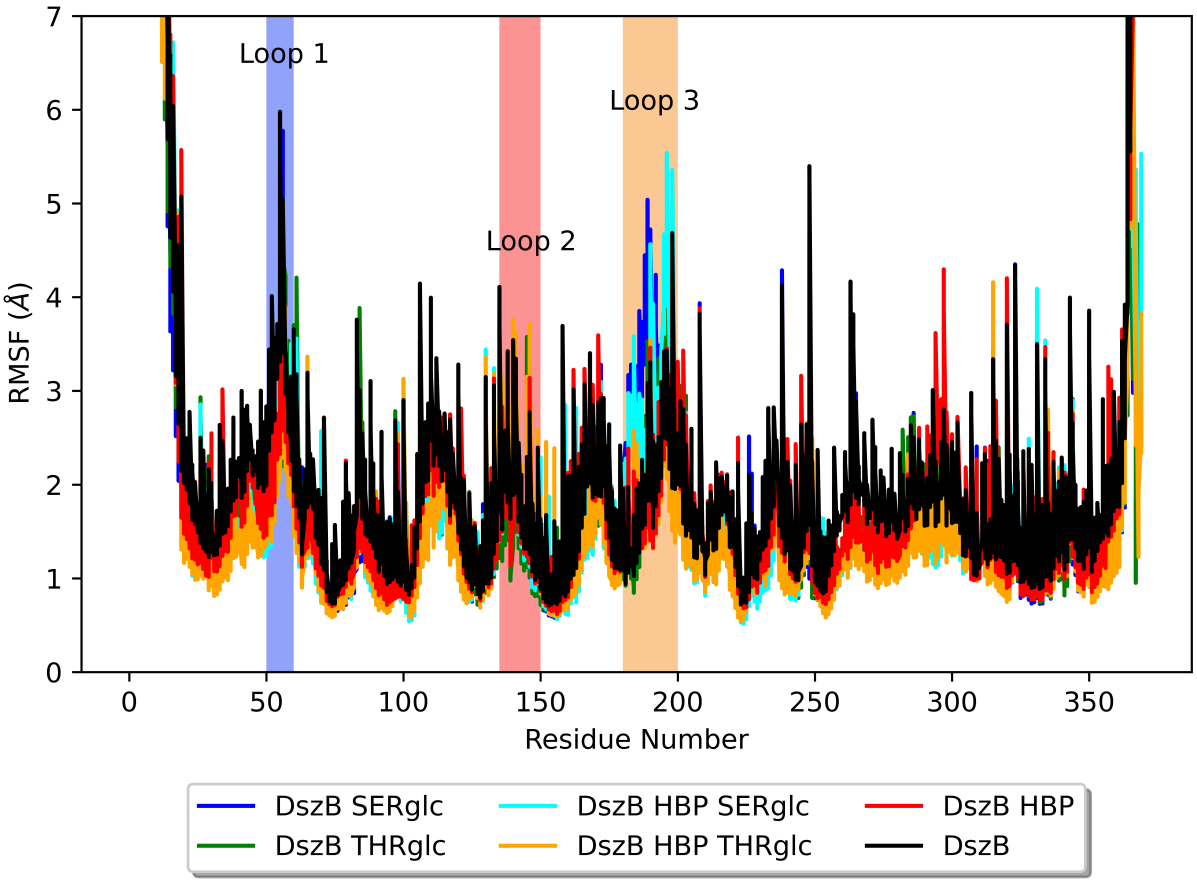

Figure 5.4: Root mean square fluctuation of the enzyme backbone over the MD production phase. Loop 1 (11 residues long), 2 (16 residues long), and 3 (21 residues long) are labeled with corresponding background color blue, red, orange.

hydrophobicity. As a result, the three loops maximize exposure to the solvent and move away from the active site. In loop 2, wild type DszB consists of 56.3% non-polar residues, 9% more than loop 1 and 4% more than loop 3. Therefore, a one residue glycosylation showed the least effect on loop-active site distance distance change. Interestingly, the Ser glycosylation system was found exhibiting a 0.5Å distance increase in loop 2, yet loop 2 had no Ser-glucose attachment. The flexibility of protein structure means the dynamics of one region can affect another. Here the increase in loop 2-active site distance is a result of dynamic correlations.

To better understand the interaction landscape between the product and the loops, we analyzed product-loop distances, as shown in Figure 5.8. Ser glycosylation added 2 to 3 Å in all three loops, implying the potential for weaker hydrophobic interactions between product and loops, therefore an increased likelihood of product exit from the active site. However, Thr glycosylation only showed a significant effect on loop 2, moderately stabilizing loop 3, while loop 1 exhibited no obvious change. One potential explanation of the better stabilization in loop 2 is its α-helix structure in apo state.

The degree of active site obstruction was also evaluated by the change of the amount of hydrogen-bond interactions between product and loop residues (Figure 5.9). Loop 3 is the main loop interacting with the product, while loop 1 has no interactions and loop 2 only interacted with the product in wild type. Glycosylation significantly decreased the number of hydrogen bonds, with Ser glycosylation system (reduced the total number of hydrogen bonds by 97.6%) more effective than Thr glycosylation (reduced the total number of hydrogen bonds by 75.1%). However, the number of hydrogen bonds between product and active site residues increased with glycosylation as shown in Figure 5.10. This suggests that the active site residues can still interact with the closely-related substrate when DszB is glycosylated. Hydrogen bonds were mainly formed between the side chain N of Gly 73 and the hydroxyl

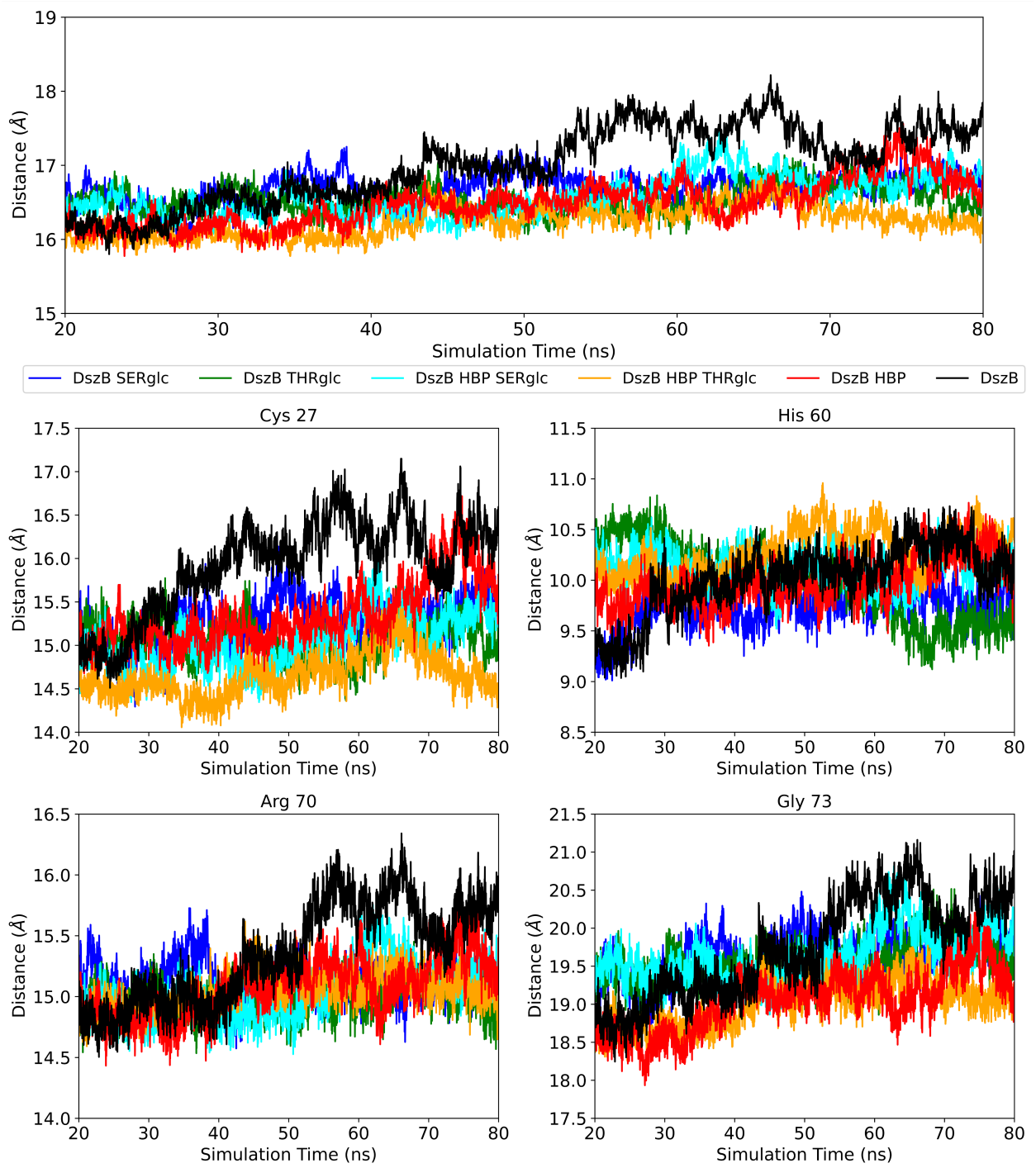

Figure 5.5: Distance between Loop 1 residues and the active site as a function of time. Top: Distance averaged over active site residue Cys 27, Arg 70, and Gly 73, as His 60 is only introduced to the active site upon bounded state. Bottom: Distance by each residues separately.

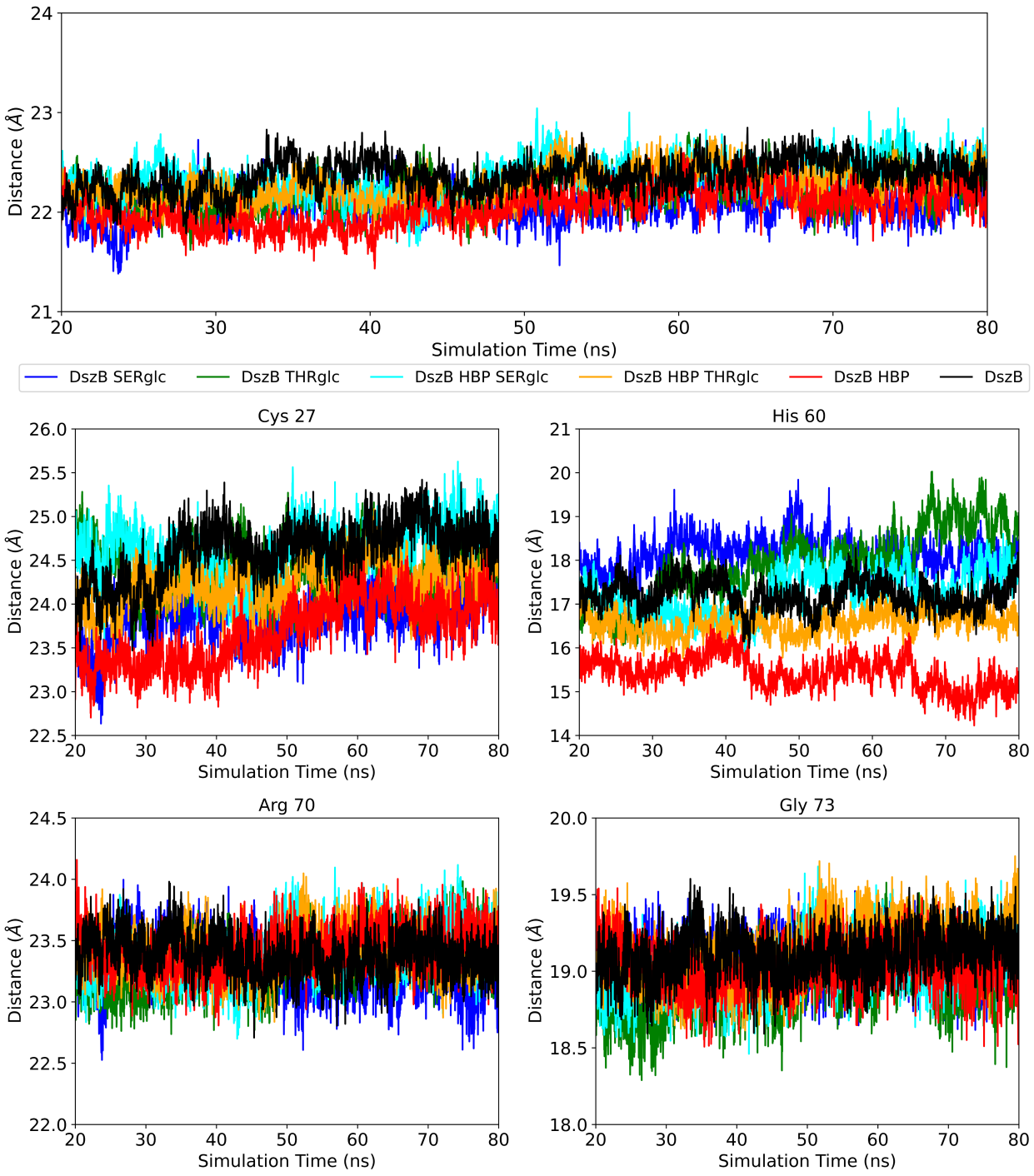

Figure 5.6: Distance between Loop 2 residues and the active site as a function of time. Top: Distance averaged over active site residue Cys 27, Arg 70, and Gly 73, as His 60 is only introduced to the active site upon bounded state. Bottom: Distance by each residues separately.

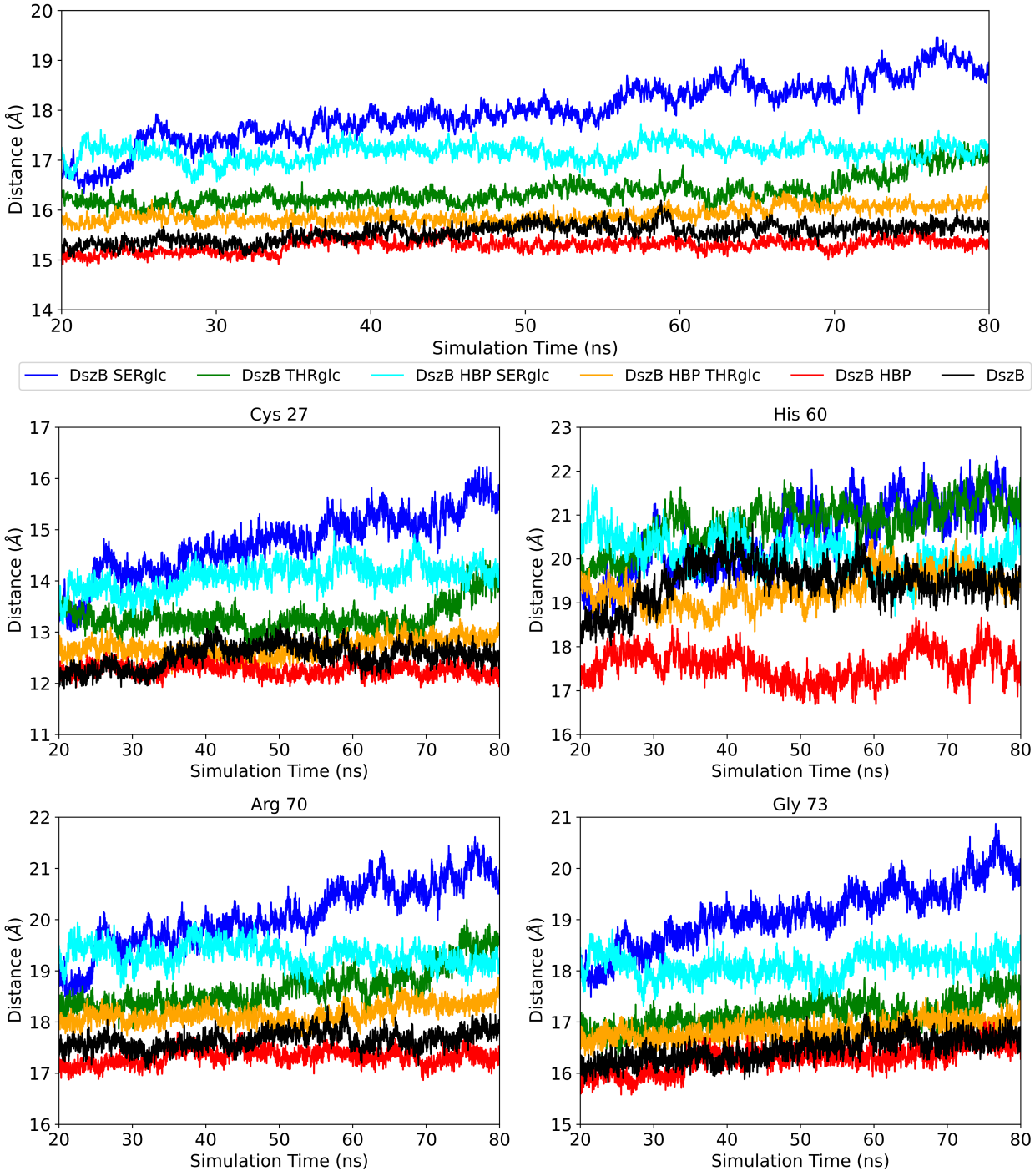

Figure 5.7: Distance between Loop 3 residues and the active site as a function of time. Top: Distance averaged over active site residue Cys 27, Arg 70, and Gly 73, as His 60 is only introduced to the active site upon bounded state. Bottom: Distance by each residues separately.

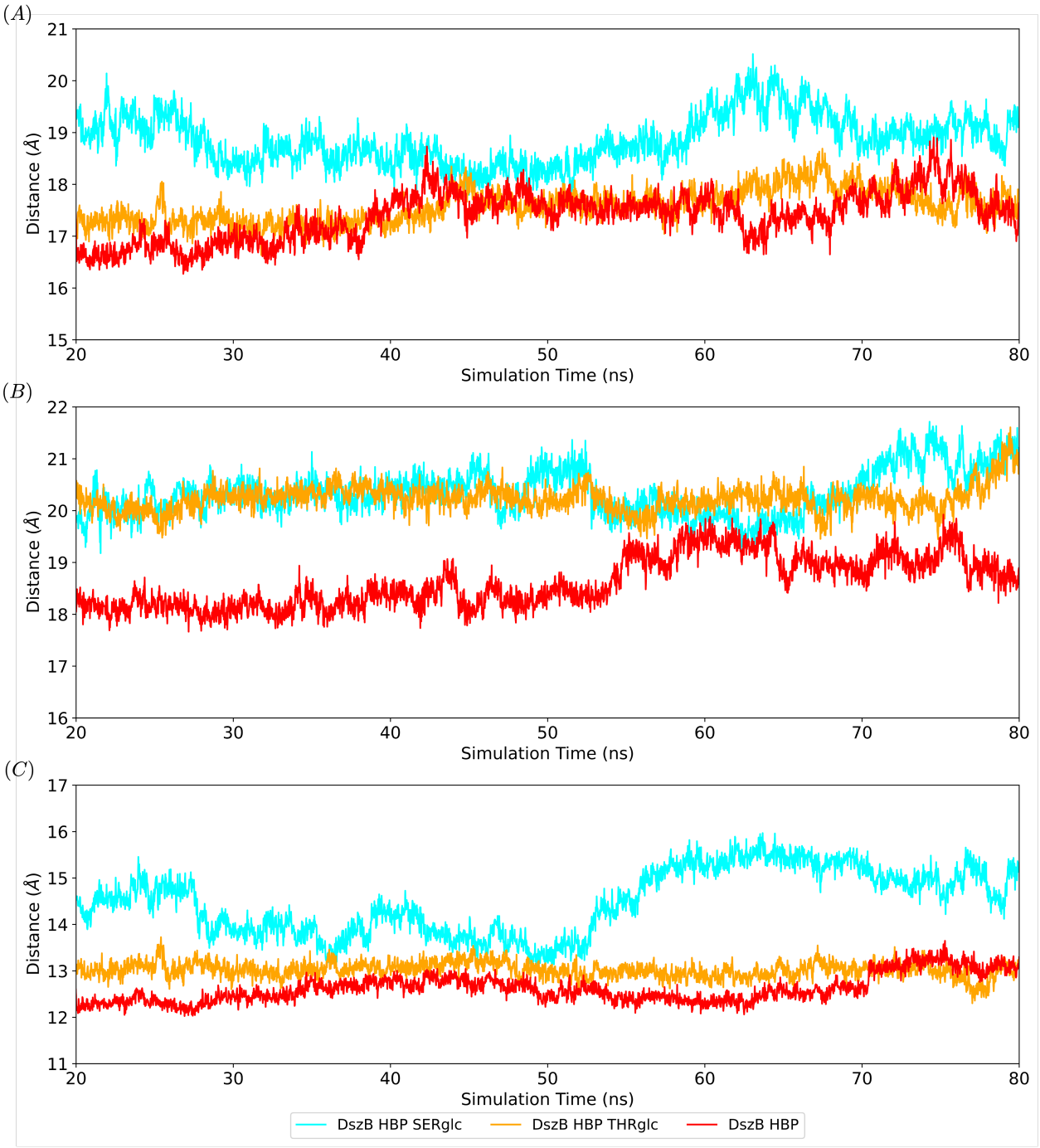

Figure 5.8: Distance between the product and loop residues as a function of time. (A) to (C) represent loop 1 to loop 3.

O of the product. Unlike Cys 27 which is actively involved in the reaction, Gly 73 stabilizes the transition state.[216,219,235] Our results suggests that Gly 73 is also involved in fixing the product position in the active site, indicating an overlap between the substrate and product binding site. While introduction of glucose to the loop can effectively distance the loops from the product, the additional Ser glycosylation site on loop 3 can further weaken the interactions, therefore facilitating the outward migration of the product.

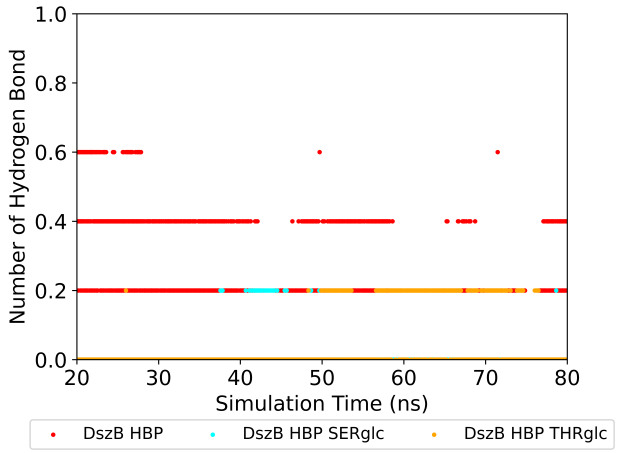

Figure 5.9: The average number of hydrogen bonding interactions between product and loop 3 residues as a function of time.

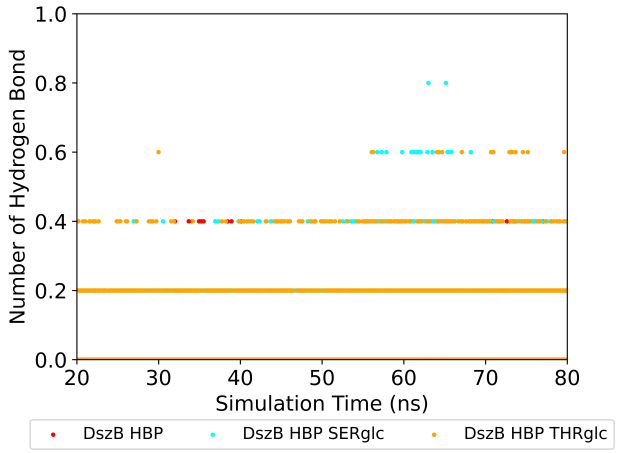

Figure 5.10: The average number of hydrogen bonding interactions between product and active site residues as a function of time.

In short, our preliminary result indicate that glucose artificially attached to the loops of the DszB gave an up to 3 Å increase in the product-loops distance and a 97.6% decrease in the amount of hydrogen bonds between the product and the loops, suggesting the effective reduction of active site obstruction by single glucose glycosylation of loop residues.

5.3.2 Evaluation of DszB Catalytic Activity

Our preliminary results demonstrated that loop residue glycosylation is an effective method for reducing the active site obstruction. However, the success of glycosylation mitigated product inhibition based on another assumption: glycosylation do not sacrifice the enzyme catalytic activity. Analysis of the glycosylated DszB performance requires the calculation of the electric fields in the active site. Evaluation of the individual residue contribution to the electric field projections onto active bond offers valuable insight into residues' role in the catalytic process as demonstrated in Chapter 3. Here, the active bond is the C-S bond being cleaved that produce the HBP product.

Simulation of the reactant state of the enzyme requires parameterization for the product,[125,160,236,237] currently under process.

5.4 Summary

Loops on proteins are usually without secondary structure, therefore displaying high flexibility. In DszB, the dynamics of 3 loops are the main cause of product inhibition, where the product is trapped inside the active site. We demonstrated that glycosylation of the loop residues effectively facilitates the reduction of active site obstruction.

Upon glycosylation, the three loops showed a general trend of extending into the solvent,

moving away from the active site. We observed a 3 Å increase in the loop-product distance, indicating a weakening in the hydrophobic interactions between product and loops 1 and 3. Loop 3 residues, especially Gly 183, is the major contributor to the interactions trapping the product, and we saw a 97.6% decrease in the number of hydrogen bonds upon glycosylation. Although loop 2 has no Ser glycosylation site, this loop still showed a moderate increase in loop-active site distance in the Ser-glycosylation system, a result from the correlation of dynamics between loops. Based on the current result, a priority of the future study is the verification whether this artificial glycosylation affects the catalytic performance of DszB enzyme.

Finally, extending the monosaccharide to glycan chains on the three loops may further stabilize the loop dynamic, improving on the catalytic activity of DszB. If successful, this method can help overcome one of the biggest challenges on desulfurization of fossil fuels for environmental protection.

Chapter 6

Conclusions

Proteins are being intensively studied as a keystone for biological processes. In order to optimize their properties for various applications, we need to understand the molecular mechanisms that govern their function. Protein stability and function are regulated by non-covalent interactions, a class of interactions that arise from multipole fluctuations between atoms. These fluctuations are modulated by the electric fields emanating from the protein and its environment. Therefore, we can analyze the strength and direction of electric fields to map structure to function in proteins.

Here, we studied proteins with two types of function: catalysis and iontransport. In Chapter 3, we introduced a DNA fragment into the environment of the synthetic enzyme KE15 to leverage the electrostatic preorganization principle. We showed that this is an effective approach to optimize the performance of the computational enzyme without destabilizing the protein structure. The role of the individual components of the system (i.e. individual protein residues, DNA fragment, solvent) in stabilizing the transition state was measured using the projection of the electric field onto the active bonds. The comparison with the reference system (without DNA fragment) showed that the introduction of DNA can significantly stabilize the transition state of the ligand. This is achieved by strengthening existing interactions between ligand and scaffold residues inside (e.g. catalytic base) or outside (e.g. solvent-exposed residues) the active site. Although there were no direct interactions between the DNA fragment and the ligand, its presence displaces the ineffective solvent, creating an

overall favorable environment for transition state stabilization.

The reduction in reaction activation energy by 2 kcal/mol with a non-optimized DNA fragment suggests a promising future for this approach. We found that the placement of the DNA fragment facing protein scaffold can effectively increase the catalytic activity of the synthetic enzyme, while its distance has only limited effects. This gives us an advantage in the artificial binding of the DNA fragment. A DNA fragment can be attached to the other end of a peptide chain linked to the N-terminus of the protein scaffold. This experimental approach in constructing enzyme-DNA complex give us control over the position of the DNA fragment. DNA nanotechnology offers the possibility to design DNA structures that can maximize the tuning effect. Based on our results, a systematic study of the structure and placement of the DNA fragments will lead to a deeper understanding of this DNA-based performance enhancement strategy.

Electrostatic preorganization theory is well accepted for enzymes but seldom used for other type of proteins. Chapter 4 focused on establishing the structure-to-function relationship of the voltage-gated ion channel $Na_v1.7$ with electric fields, a step towards the generalization electrostatic preorganization theory to ion channels. We found that the residues around the SF and IG residues generate electric fields that facilitate Na^+ binding and permeation through the pore. For example, the outer ring above the SF promotes Na^+ movement toward the DEKA residues. We also found that the SF region is more active than the IG region and that each domain contributes unequally to the fields. Indeed, we observed a higher number of active residues in DI than in the other three domains, both for ion permeation and voltage sensing. In addition, DI is the only domain that favors the inactivated state conformation for the VSD. Electric field projection in DII to DIV indicates that they promote the open state conformation.

Electric field projection provided an insight into the role of disease mutation residues. Among

the residues that significantly contributed to the electric fields, 3 were found to be related to the hyperpolarization disease mutation. Finally, we found that electric field strength and direction is related to residue side-chain rotamers, where significant changes in torsions ($\chi > 60°$) correlate with significant changes in electric fields.

Electric field calculations in $Na_v1.7$ raised additional questions. First, whether the outer carboxylate ring or other residues of the SF region facilitate binding migration for all types of ions, or they preserve selectivity for Na^+ ions. Second, is the lower number of active residues in the IG region a result of the inactivated state conformation, or the state changes achieved by changing the electric field directions? Third, non-conserved/variant residues were found contributing to the electric field. Cross comparison between Na_vs will provide additional knowledge on their unique functions and could lead to the rational design of therapeutic strategies specific to human $Na_v1.7$.

Protein loops are generally flexibleand without secondary structure. In DszB, the hydrophobic interactions between the product in the active site and these loop residues prevent the product from escaping, resulting in product inhibition of the enzyme. In chapter 5, we present a synthetic loop glycosylation strategy as a solution to reduce such interactions in DszB. We attached glucose to a few loop residues to change the chemical natural of the loops and weaken their interactions with the product. Loop glycosylation created a 3 Å difference in loop-product distance. This decrease in distance reduced up to 97.6% of hydrogen bonds between them. The effect of increasing the number of glucose attached to one loop were demonstrated with loop 3, which experienced the most significant distancing from the active site. While our strategy mitigated the obstruction of product exit pathway, glycosylation also increased the interactions between the active site residues and the product, suggesting a preservation of catalytic activity.

This will be further investigated by modeling the reactant in the active site and analyzing

its interactions with active site residues. This will be done by computing the electric fields along the reactant active bonds between DszB wild type and glycosylated DszB.

Overall, this dissertation focuses on building protein structure-function relationships using electric fields as a metric. Each component in the system reveals its mechanistic contribution through their contribution to the field magnitude and direction. Current works explored its application in enzymes and ion channels, where computational results can serve as a theoretical guide for experimental studies and large scale application. In its broader applications, electric fields can also be used in polymers, as non-covalent interactions also significantly effect polymer properties.

Reference

(1) Müller-Dethlefs, K.; Hobza, P. Noncovalent Interactions: A Challenge for Experiment and Theory. *Chemical Reviews* **1999**, *100*, 143–168.

(2) Lehn, J.-M. Supramolecular Polymer Chemistry—Scope and Perspectives. *Polymer International* **2002**, *51*, 825–839.

(3) Riley, K. E.; Hobza, P. Noncovalent Interactions in Biochemistry. *WIREs Computational Molecular Science* **2011**, *1*, 3–17.

(4) Kuriyan, J.; Konforti, B.; Wemmer, D., *The Molecules of Life*; W.W. Norton Company: 2012.

(5) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *Journal of the American Chemical Society* **2010**, *132*, 6498–6506.

(6) Adhav, V. A.; Saikrishnan, K. The Realm of Unconventional Noncovalent Interactions in Proteins: Their Significance in Structure and Function. *ACS Omega* **2023**, *8*, 22268–22284.

(7) Israelachvili, J. N., *Intermolecular and Surface Forces*; Elsevier: 2011.

(8) Spolar, R. S.; Ha, J. H.; Record, M. T. Hydrophobic Effect in Protein Folding and Other Noncovalent Processes Involving Proteins. *Proceedings of the National Academy of Sciences* **1989**, *86*, 8382–8385.

(9) Persch, E.; Dumele, O.; Diederich, F. Molecular Recognition in Chemical and Biological Systems. *Angewandte Chemie International Edition* **2015**, *54*, 3290–3327.

(10) Pollard, T.; Earnshaw, W.; Lippincott-Schwartz, J.; Johnson, G., *Cell Biology: Third Edition*; Elsevier Inc.: United States, 2016.

(11) Hermann, J.; DiStasio, R. A.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chemical Reviews* **2017**, *117*, 4714–4758.

(12) Israelachvili, J. N. The nature of van der waals forces. *Contemporary Physics* **1974**, *15*, 159–178.

(13) Leach, A. R. In *Molecular Modelling: Principles and Applications*; Pearson: 2001, pp 353–406.

(14) London, F. The general theory of molecular forces. *Transactions of the Faraday Society* **1937**, *33*, 8b–26.

(15) Parsegian, V. A., *Van der Waals forces: a handbook for biologists, chemists, engineers, and physicists*; Cambridge university press: 2005.

(16) Vesper, M. D.; de Groot, B. L. Collective Dynamics Underlying Allosteric Transitions in Hemoglobin. *PLoS Computational Biology* **2013**, *9*, ed. by Livesay, D. R., e1003232.

(17) Shiina, M.; Hamada, K.; Inoue-Bungo, T.; Shimamura, M.; Uchiyama, A.; Baba, S.; Sato, K.; Yamamoto, M.; Ogata, K. A Novel Allosteric Mechanism on Protein–DNA Interactions underlying the Phosphorylation-Dependent Regulation of Ets1 Target Gene Expressions. *Journal of Molecular Biology* **2015**, *427*, 1655–1669.

(18) Ramaswamy, A.; Balasubramanian, S.; Rajagopalan, M. In *Molecular Docking for Computer-Aided Drug Design*; Elsevier: 2021, pp 31–55.

(19) Heller, G. T.; Aprile, F. A.; Vendruscolo, M. Methods of Probing the Interactions Between Small Molecules and Disordered Proteins. *Cellular and Molecular Life Sciences* **2017**, *74*, 3225–3243.

(20) Smith, R.; Inomata, H.; Peters, C. In *Introduction to Supercritical Fluids - A Spreadsheet-based Approach*; Elsevier: 2013, pp 333–480.

(21) Herschlag, D.; Pinney, M. M. Hydrogen Bonds: Simple after All? *Biochemistry* **2018**, *57*, 3338–3352.

(22) Karas, L. J.; Wu, C.-H.; Das, R.; Wu, J. I.-C. Hydrogen Bond Design Principles. *WIREs Computational Molecular Science* **2020**, *10*.

(23) Karshikoff, A., *Non-covalent Interactions In Proteins*; World Scientific Publishing Company: 2006.

(24) Hubbard, R. E.; Kamran Haider, M. Hydrogen Bonds in Proteins: Role and Strength, 2010.

(25) Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proceedings of the National Academy of Sciences* **1951**, *37*, 205–211.

(26) Pauling, L.; Corey, R. B. The Pleated Sheet, A New Layer Configuration of Polypeptide Chains. *Proceedings of the National Academy of Sciences* **1951**, *37*, 251–256.

(27) Pace, C. N. et al. Contribution of Hydrogen Bonds to Protein Stability. *Protein Science* **2014**, *23*, 652–661.

(28) Escobedo, A. et al. Side Chain to Main Chain Hydrogen Bonds Stabilize a Polyglutamine Helix in a Transcription Factor. *Nature Communications* **2019**, *10*.

(29) Buckingham, A.; Del Bene, J.; McDowell, S. The hydrogen bond. *Chemical Physics Letters* **2008**, *463*, 1–10.

(30) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical Reviews* **2018**, *118*, 1691–1741.

(31) Jones, K. A.; Kadonaga, J. T.; Rosenfeld, P. J.; Kelly, T. J.; Tjian, R. A Cellular DNA-Binding Protein that Activates Eukaryotic Transcription and DNA Replication. *Cell* **1987**, *48*, 79–89.

(32) Brand, I. Achievements and Challenges in Resolving Molecular Aspects of DNA-Protein Interactions. *Current Opinion in Electrochemistry* **2023**, *40*, 101333.

(33) Kangas, E.; Tidor, B. Optimizing Electrostatic Affinity in Ligand–Receptor Binding: Theory, Computation, and Ligand Properties. *The Journal of Chemical Physics* **1998**, *109*, 7522–7545.

(34) Nguyen, H.; Lan, P. D.; Nissley, D. A.; O'Brien, E. P.; Li, M. S. Electrostatic Interactions Explain the Higher Binding Affinity of the CR3022 Antibody for SARS-CoV-2 than the 4A8 Antibody. *The Journal of Physical Chemistry B* **2021**, *125*, 7368–7379.

(35) Honig, B. H.; Hubbell, W. L.; Flewelling, R. F. Electrostatic Interactions in Membranes and Proteins. *Annual Review of Biophysics and Biophysical Chemistry* **1986**, *15*, 163–193.

(36) Van den Bogaart, G.; Meyenberg, K.; Risselada, H. J.; Amin, H.; Willig, K. I.; Hubrich, B. E.; Dier, M.; Hell, S. W.; Grubmüller, H.; Diederichsen, U.; Jahn, R. Membrane Protein Sequestering by Ionic Protein–Lipid Interactions. *Nature* **2011**, *479*, 552–555.

(37) Burns, D. B.; Zydney, A. L. Contributions to Electrostatic Interactions on Protein Transport in Membrane Systems. *AIChE Journal* **2001**, *47*, 1101–1114.

(38) Laberge, M. Intrinsic Protein Electric Fields: Bbasic Non-covalent Interactions and Relationship to Protein-Induced Stark Effects. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1998**, *1386*, 305–330.

(39) Israelachvili, J.; Pashley, R. The Hydrophobic Interaction is Long Range, Decaying Exponentially with Distance. *Nature* **1982**, *300*, 341–342.

(40) Privalov, P. L.; Gill, S. J. In *Advances in Protein Chemistry*; Elsevier: 1988, pp 191–234.

(41) Ptitsyn, O.; Rashin, A. A model of myoglobin self-organization. *Biophysical Chemistry* **1975**, *3*, 1–20.

(42) Aryal, P.; Sansom, M. S.; Tucker, S. J. Hydrophobic Gating in Ion Channels. *Journal of Molecular Biology* **2015**, *427*, 121–130.

(43) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein–Ligand Binding. *Proceedings of the National Academy of Sciences* **2007**, *104*, 808–813.

(44) Fried, S. D.; Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Accounts of Chemical Research* **2015**, *48*, 998–1006.

(45) HARTLEY, H. Origin of the Word 'Protein'. *Nature* **1951**, *168*, 244–244.

(46) Branden, C. I.; Tooze, J. In *Introduction to protein structure*; Garland Science: 2012, pp 3–12.

(47) Shi, L.; Javitch, J. A. The Binding Site of Aminergic G Protein–Coupled Receptors: The Transmembrane Segments and Second Extracellular Loop. *Annual Review of Pharmacology and Toxicology* **2002**, *42*, 437–467.

(48) Awasthi, M.; Gulati, S.; Sarkar, D. P.; Tiwari, S.; Kateriya, S.; Ranjan, P.; Verma, S. K. The Sialoside-Binding Pocket of SARS-CoV-2 Spike Glycoprotein Structurally Resembles MERS-CoV. *Viruses* **2020**, *12*, 909.

(49) Bailey, T. L. STREME: Accurate and Versatile Sequence Motif Discovery. *Bioinformatics* **2021**, *37*, ed. by Birol, I., 2834–2840.

(50) Fang, J.; Haasl, R. J.; Dong, Y.; Lushington, G. H. Discover Protein Sequence Signatures from Protein-Protein Interaction Data. *BMC Bioinformatics* **2005**, *6*.

(51) Skipper, L. In *Encyclopedia of Analytical Science*; Elsevier: 2005, pp 344–352.

(52) Fairman, J. W.; Noinaj, N.; Buchanan, S. K. The structural biology of -barrel membrane proteins: a summary of recent reports. *Current Opinion in Structural Biology* **2011**, *21*, 523–531.

(53) Ouellette, R. J.; Rawn, J. D. In *Principles of Organic Chemistry*; Elsevier: 2015, pp 371–396.

(54) Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences* **2016**, *17*, 144.

(55) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L., *Enzymes: Basuc Concepts and Kinetics*; Springer Berlin Heidelberg: 2018, pp 233–266.

(56) Jones, L.; Laverman, L., *Chemical principles*, 7th ed.; W.H. Freeman: New York, NY, 2016.

(57) Hush, N. S.; Reimers, J. R. Vibrational Stark Spectroscopy. 1. Basic Theory and Application to the CO Stretch. *The Journal of Physical Chemistry* **1995**, *99*, 15798–15805.

(58) Fried, S. D.; Boxer, S. G. Electric Fields and Enzyme Catalysis. *Annual Review of Biochemistry* **2017**, *86*, 387–415.

(59) Fried, S. D.; Bagchi, S.; Boxer, S. G. Measuring Electrostatic Fields in Both Hydrogen-Bonding and Non-Hydrogen-Bonding Environments Using Carbonyl Vibrational Probes. *Journal of the American Chemical Society* **2013**, *135*, 11181–11192.

(60) Bagchi, S.; Fried, S. D.; Boxer, S. G. A Solvatochromic Model Calibrates Nitriles' Vibrational Frequencies to Electrostatic Fields. *Journal of the American Chemical Society* **2012**, *134*, 10373–10376.

(61) Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G. Vibrational Stark Spectroscopy in Proteins: A Probe and Calibration for Electrostatic Fields. *The Journal of Physical Chemistry B* **1999**, *103*, 9813–9817.

(62) Bublitz, G. U.; Boxer, S. G. Effective Polarity of Frozen Solvent Glasses in the Vicinity of Dipolar Solutes. *Journal of the American Chemical Society* **1998**, *120*, 3988–3992.

(63) Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme Electric Fields Power Catalysis in the Active Site of Ketosteroid Isomerase. *Science* **2014**, *346*, 1510–1514.

(64) Beilen, J. B.; Li, Z. Enzyme technology: an overview. *Current Opinion in Biotechnology* **2002**, *13*, 338–344.

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

(66) Onsager, L. Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society* **1936**, *58*, 1486–1493.

(67) Marini, A.; Muñoz-Losa, A.; Biancardi, A.; Mennucci, B. What is Solvatochromism? *The Journal of Physical Chemistry B* **2010**, *114*, 17128–17135.

(68) Lee, H.; Lee, G.; Jeon, J.; Cho, M. Vibrational Spectroscopic Determination of Local Solvent Electric Field, Solute–Solvent Electrostatic Interaction Energy, and Their Fluctuation Amplitudes. *The Journal of Physical Chemistry A* **2011**, *116*, 347–357.

(69) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory. *Science* **2006**, *313*, 200–204.

(70) Ringer, A. L.; MacKerell, A. D. Calculation of the Vibrational Stark Effect Using a First-Principles Quantum Mechanical/Molecular Mechanical Approach. *The Journal of Physical Chemistry Letters* **2011**, *2*, 553–556.

(71) Petsko, G. A.; Ringe, D., *Protein structure and function*; New Science Press: 2004.

(72) Finkelstein, A. V.; Ptitsyn, O., *Protein physics: a course of lectures*; Elsevier: 2016.

(73) Bragg, L. W. The Structure of Some Crystals as Indicated by Their Diffraction of X-rays. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1913**, *89*, 248–277.

(74) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662–666.

(75) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585–590.

(76) Sinha, S.; Tam, B.; Wang, S. M. Applications of Molecular Dynamics Simulation in Protein Study. *Membranes* **2022**, *12*, 844.

(77) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biology* **2011**, *9*.

(78) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **1990**, *347*, 631–639.

(79) Tuckerman, M. E.; Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *The Journal of Physical Chemistry B* **1999**, *104*, 159–178.

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

(81) Frenkel, D.; Smit, B., *Understanding molecular simulation*, 2nd ed.; Computational science series; Academic Press: San Diego, CA, 2001.

(82) Binder, K.; Horbach, J.; Kob, W.; Paul, W.; Varnik, F. Molecular dynamics simulations. *Journal of Physics: Condensed Matter* **2004**, *16*, S429–S453.

(83) Zuckerman, D. M., *Statistical physics of biomolecules: an introduction*; CRC press: 2010.

(84) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2019**, *1*.

(85) Stella, L.; Melchionna, S. Equilibration and sampling in molecular dynamics simulations of biomolecules. *The Journal of Chemical Physics* **1998**, *109*, 10115–10117.

(86) Sargsyan, K.; Grauffel, C.; Lim, C. How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2017**, *13*, 1518–1524.

(87) Schreiner, W.; Karch, R.; Knapp, B.; Ilieva, N. Relaxation Estimation of RMSD in Molecular Dynamics Immunosimulations. *Computational and Mathematical Methods in Medicine* **2012**, *2012*, 1–9.

(88) Arnittali, M.; Rissanou, A. N.; Harmandaris, V. Structure Of Biomolecules Through Molecular Dynamics Simulations. *Procedia Computer Science* **2019**, *156*, 69–78.

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

(90) Hoover, W. G.; Ladd, A. J. C.; Moran, B. High-Strain-Rate Plastic Flow Studied via Nonequilibrium Molecular Dynamics. *Physical Review Letters* **1982**, *48*, 1818–1820.

(91) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* **1980**, *72*, 2384–2393.

(92) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **1984**, *52*, 255–268.

(93) González, M. Force Fields and Molecular Dynamics Simulations. *École thématique de la Société Française de la Neutronique* **2011**, *12*, 169–200.

(94) Qi, R.; Wang, Q.; Ren, P. General van der Waals Potential for Common Organic Molecules. *Bioorganic amp; Medicinal Chemistry* **2016**, *24*, 4911–4919.

(95) Hwang, M. J.; Stockfisch, T.; Hagler, A. Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules. *Journal of the American Chemical Society* **1994**, *116*, 2515–2525.

(96) Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy Parameters in Polypeptides. 10. Improved Geometrical Parameters and Nonbonded Interactions for use in the ECEPP/3 Algorithm, with Application to Proline-Containing Peptides. *The Journal of Physical Chemistry* **1992**, *96*, 6472–6484.

(97) Halgren, T. A. The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *Journal of the American Chemical Society* **1992**, *114*, 7827–7843.

(98) Stone, A. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chemical Physics Letters* **1981**, *83*, 233–239.

(99) Williams, D. E. Representation of the Molecular Electrostatic Potential by Atomic Multipole and Bond Dipole Models. *Journal of Computational Chemistry* **1988**, *9*, 745–763.

(100) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chemical Reviews* **2013**, *114*, 779–814.

(101) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371–394.

(102) Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. Biomolecular Electrostatics and Solvation: a Computational Perspective. *Quarterly Reviews of Biophysics* **2012**, *45*, 427–491.

(103) Dykstra, C. E. Intermolecular Electrical Interaction: a Key Ingredient in Hydrogen Bonding. *Accounts of Chemical Research* **1988**, *21*, 355–361.

(104) Applequist, J. A Multipole Interaction Theory of Electric Polarization of Atomic and Molecular Assemblies. *The Journal of Chemical Physics* **1985**, *83*, 809–826.

(105) Ren, P.; Ponder, J. W. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. *Journal of Computational Chemistry* **2002**, *23*, 1497–1506.

(106) Thole, B. Molecular Polarizabilities Aalculated with a Modified Dipole Interaction. *Chemical Physics* **1981**, *59*, 341–350.

(107) Nash, J.; Barnes, T.; Welborn, V. ELECTRIC: Electric fields Leveraged from multipole Expansion Calculations in Tinker Rapid Interface Code. *Journal of Open Source Software* **2020**, *5*, 2576.

(108) Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angewandte Chemie International Edition* **2020**, *60*, 88–119.

(109) Chapman, J.; Ismail, A.; Dinu, C. Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks. *Catalysts* **2018**, *8*, 238.

(110) Bommarius, A. S.; Paye, M. F. Stabilizing Biocatalysts. *Chemical Society Reviews* **2013**, *42*, 6534.

(111) Singh, R.; Kumar, M.; Mittal, A.; Mehta, P. K. MicrobialEnzymes: Industrial Progress in 21st Century. *3 Biotech* **2016**, *6*.

(112) Raveendran, S.; Parameswaran, B.; Ummalyma, S. B.; Abraham, A.; Mathew, A. K.; Madhavan, A.; Rebello, S.; Pandey, A. Applications of Microbial Enzymes in Food Industry. *Food Technology and Biotechnology* **2018**, *56*.

(113) Lambert, W. The effect of pH on the foam fractionation of -glucosidase and cellulase. *Bioresource Technology* **2003**, *87*, 247–253.

(114) Daniel, R. M.; Peterson, M. E.; Danson, M. J.; Price, N. C.; Kelly, S. M.; Monk, C. R.; Weinberg, C. S.; Oudshoorn, M. L.; Lee, C. K. The molecular basis of the effect of temperature on enzyme activity. *Biochemical Journal* **2009**, *425*, 353–360.

(115) Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays in Biochemistry* **2015**, *59*, 1–41.

(116) Erb, T. J.; Jones, P. R.; Bar-Even, A. Synthetic metabolism: metabolic engineering meets enzyme design. *Current Opinion in Chemical Biology* **2017**, *37*, 56–62.

(117) Nanda, V.; Koder, R. L. Designing Artificial Enzymes by Intuition and Computation. *Nat. Chem.* **2009**, *2*, 15–24.

(118) Ward, T. Artificial Enzymes Made to Order: Combination of Computational Design and Directed Evolution. *Angew. Chem. Int. Ed.* **2008**, *47*, 7802–7803.

(119) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190–195.

(120) Khersonsky, O.; Kiss, G.; Rothlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. Bridging the Gaps in Design Methodologies by Evolutionary Optimization of the Stability and Proficiency of Designed Kemp Eliminase KE59. *Proceedings of the National Academy of Sciences* **2012**, *109*, 10358–10363.

(121) Khersonsky, O.; Röthlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss, G.; Houk, K.; Baker, D.; Tawfik, D. S. Optimization of the In-Silico-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution. *Journal of Molecular Biology* **2011**, *407*, 391–412.

(122) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grütter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013**, *503*, 418–421.

(123) Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proceedings of the National Academy of Sciences* **2010**, *107*, 16869–16874.

(124) Vaissier, V.; Sharma, S. C.; Schaettle, K.; Zhang, T.; Head-Gordon, T. Computational Optimization of Electric Fields for Improving Catalysis of a Designed Kemp Eliminase. *ACS Catalysis* **2017**, *8*, 219–227.

(125) Bhowmick, A.; Sharma, S. C.; Head-Gordon, T. The Importance of the Scaffold for de Novo Enzymes: A Case Study with Kemp Eliminase. *Journal of the American Chemical Society* **2017**, *139*, 5793–5800.

(126) Fazelinia, H.; Cirino, P. C.; Maranas, C. D. OptGraft: A Computational Procedure for Transferring a Binding Site onto an Existing Protein Scaffold. *Protein Sci.* **2008**, 180–195.

(127) Malisi, C.; Kohlbacher, O.; Höcker, B. Automated Scaffold Selection for Enzyme Design. *Proteins* **2009**, *77*, 74–83.

(128) Huang, X.; Han, K.; Zhu, Y. Systematic Optimization Model and Algorithm for Binding Sequence Selection in Computational Enzyme Design. *Protein Sci.* **2013**, *22*, 929–941.

(129) Nosrati, G. R.; Houk, K. N. SABER: A cComputational Method for Identifying Active Sites for New Reactions. *Protein Sci.* **2012**, *21*, 697–706.

(130) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chemical Reviews* **2006**, *106*, 3210–3235.

(131) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Iite Remodeling. *Nature Chemical Biology* **2013**, *9*, 494–498.

(132) Tantillo, D. J.; Jiangang, C.; Houk, K. N. Theozymes and Compuzymes: Theoretical Models for Biological Catalysis. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743–750.

(133) Welborn, V. V.; Head-Gordon, T. Computational Design of Synthetic Enzymes. *Chem. Rev.* **2018**, *119*, 6613–6630.

(134) Korendovych, I. V.; DeGrado, W. F. Catalytic Efficiency of Designed Catalytic Proteins. *Current Opinion in Structural Biology* **2014**, *27*, 113–121.

(135) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovnik, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D. Impact of Scaffold Rigidity on the Design and Evolution of an Artificial Diels-Alderase. *Proceedings of the National Academy of Sciences* **2014**, *111*, 8013–8018.

(136) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angewandte Chemie International Edition* **2013**, *52*, 5700–5725.

(137) Brustad, E. M.; Arnold, F. H. Optimizing Non-natural Protein Function with Directed EWvolution. *Current Opinion in Chemical Biology* **2011**, *15*, 201–210.

(138) Dalby, P. A. Strategy and Success for the Directed Evolution of Enzymes. *Current Opinion in Structural Biology* **2011**, *21*, 473–480.

(139) Turner, N. J. Directed Evolution Drives the Next Generation of Biocatalysts. *Nature Chemical Biology* **2009**, *5*, 567–573.

(140) Dunn, M. R.; Otto, C.; Fenton, K. E.; Chaput, J. C. Improving Polymerase Activity with Unnatural Substrates by Sampling Mutations in Homologous Protein Architectures. *ACS Chem. Biol.* **2016**, *11*, 1210–1219.

(141) Steward, R. E.; MacArthur, M. W.; Laskowski, R. A.; Thornton, J. M. Molecular basis of inherited diseases: a structural perspective. *Trends in Genetics* **2003**, *19*, 505–513.

(142) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLoS Computational Biology* **2008**, *4*, ed. by Eisenberg, D., e1000002.

(143) .Newton, M. S.; Arcus, V. L.; Gerth, M. L.; Patrick, W. M. Enzyme evolution: innovation is easy, optimization is complicated. *Current Opinion in Structural Biology* **2018**, *48*, 110–116.

(144) Watson, J. D.; Crick, F. H. C. THE STRUCTURE OF DNA. *Cold Spring Harbor Symposia on Quantitative Biology* **1953**, *18*, 123–131.

(145) Sinden, R. R., *DNA structure and function*; Gulf Professional Publishing: 1994.

(146) Feldkamp, U.; Niemeyer, C. M. Rational Design of DNA Nanoarchitectures. *Angewandte Chemie International Edition* **2006**, *45*, 1856–1876.

(147) Lin, C.; Liu, Y.; Yan, H. Designer DNA Nanoarchitectures. *Biochemistry* **2009**, *48*, 1663–1674.

(148) Dey, S.; Fan, C.; Gothelf, K. V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M. A. D.; Saccà, B.; Simmel, F. C.; Yan, H.; Zhan, P. DNA Origami. *Nature Reviews Methods Primers* **2021**, *1*.

(149) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chemical Reviews* **2017**, *117*, 12584–12640.

(150) Zheng, Z.; Wang, Y. DNA Binding Proteins: Outline of Functional Classification. *BioMolecular Concepts* **2011**, *2*, 293–303.

(151) Dhanasekaran, M.; Negi, S.; Sugiura, Y. Designer Zinc Finger Proteins: Tools for Creating Artificial DNA-Binding Functional Proteins. *Accounts of Chemical Research* **2005**, *39*, 45–52.

(152) Lee, D.-k.; Seol, W.; Kim, J.-S. Custom DNA-Binding Proteins and Artificial Transcription Factors. *Current Topics in Medicinal Chemistry* **2003**, *3*, 645–657.

(153) Trads, J. B.; Tørring, T.; Gothelf, K. V. Site-Selective Conjugation of Native Proteins with DNA. *Accounts of Chemical Research* **2017**, *50*, 1367–1374.

(154) Davis, I. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. The Backrub Motion: How Protein Backbone Shrugs When a Sidechain Dances. *Structure* **2006**, *14*, 265–274.

(155) Smith, C. A.; Kortemme, T. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *Journal of Molecular Biology* **2008**, *380*, 742–756.

(156) Shapovalov, M. V.; Dunbrack, R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*, 844–858.

(157) Smith, C. A.; Ban, D.; Pratihar, S.; Giller, K.; Schwiegk, C.; de Groot, B. L.; Becker, S.; Griesinger, C.; Lee, D. Population Shuffling of Protein Conformations. *Angewandte Chemie International Edition* **2014**, *54*, 207–210.

(158) Keedy, D. A. et al. Mapping the Conformational Landscape of a Dynamic Enzyme by Multitemperature and XFEL Crystallography. *eLife* **2015**, *4*.

(159) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(160) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *Journal of Chemical Theory and Computation* **2018**, *14*, 5273–5289.

(161) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(162) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(163) Orengo, C. A.; Todd, A. E.; Thornton, J. M. From protein structure to function. *Current Opinion in Structural Biology* **1999**, *9*, 374–382.

(164) Kessel, A.; Ben-Tal, N., *Introduction to Proteins*; Chapman and Hall/CRC: 2018.

(165) Sheinerman, F. B.; Honig, B. On the Role of Electrostatic Interactions in the Design of Protein–Protein Interfaces. *Journal of Molecular Biology* **2002**, *318*, 161–177.

(166) Huang, G.; Wu, Q.; Li, Z.; Jin, X.; Huang, X.; Wu, T.; Pan, X.; Yan, N. Unwinding and Spiral Sliding of S4 and Somain Rotation of VSD During the Electromechanical Coupling in $Na_v1.7$. *Proceedings of the National Academy of Sciences* **2022**, *119*.

(167) Barker, B.; Young, G.; Soubrane, C.; Stephens, G.; Stevens, E.; Patel, M. In *Conn's Translational Neuroscience*; Elsevier: 2017, pp 11–43.

(168) Hille, B. Ionic Channels in Excitable Membranes. Current Problems and Biophysical Approaches. *Biophysical Journal* **1978**, *22*, 283–294.

(169) Wang, J.; Ou, S.-W.; Wang, Y.-J. Distribution and Function of Voltage-Gated Sodium Channels in the Nervous System. *Channels* **2017**, *11*, 534–554.

(170) De Lera Ruiz, M.; Kraus, R. L. Voltage-Gated Sodium Channels: Structure, Function, Pharmacology, and Clinical Indications. *Journal of Medicinal Chemistry* **2015**, *58*, 7093–7118.

(171) Yu, F. H.; Catterall, W. A. Overview of the Voltage-gated Sodium Channel Family. *Genome Biology* **2003**, *4*, 207.

(172) Flood, E.; Boiteux, C.; Lev, B.; Vorobyov, I.; Allen, T. W. Atomistic simulations of membrane ion channel conduction, gating, and modulation. *Chem. Rev.* **2019**, *119*, 7737–7832.

(173) Yang, N.; Horn, R. Evidence for Voltage-Dependent S4 Movement in Sodium Channels. *Neuron* **1995**, *15*, 213–218.

(174) Dib-Hajj, S. D.; Yang, Y.; Black, J. A.; Waxman, S. G. The NaV1.7 sodium channel: from molecule to man. *Nature Reviews Neuroscience* **2012**, *14*, 49–62.

(175) Zheng, J.; Trudeau, M. C., *Handbook of Ion Channels*; Crc Press: 2015.

(176) Ahern, C. A.; Payandeh, J.; Bosmans, F.; Chanda, B. The Hitchhiker's Guide to the Voltage-Gated Sodium Channel Galaxy. *Journal of General Physiology* **2015**, *147*, 1–24.

(177) Fischer, T. Z.; Waxman, S. G. Familial pain syndromes from mutations of the Nav1.7 sodium channel. *Annals of the New York Academy of Sciences* **2010**, *1184*, 196–207.

(178) Shen, H.; Liu, D.; Wu, K.; Lei, J.; Yan, N. Structures of Human $Na_v1.7$ Channel in Complex with Auxiliary Subunits and Animal Toxins. *Science* **2019**, *363*, 1303–1308.

(179) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(180) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.

(181) Bienert, S.; Waterhouse, A.; de Beer, T. A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research* **2016**, *45*, D313–D319.

(182) Guex, N.; Peitsch, M. C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS* **2009**, *30*.

(183) Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T. QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics* **2019**, *36*, ed. by Elofsson, A., 1765–1771.

(184) Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* **2017**, *7*.

(185) Totrov, M. In *Homology Modeling*; Humana Press: 2011, pp 207–229.

(186) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W. In *Methods in enzymology*, 2011, pp 545–574.

(187) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.

(188) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.

(189) Lee, J.; Hitzenberger, M.; Rieger, M.; Kern, N. R.; Zacharias, M.; Im, W. CHARMM-GUI Supports the Amber Force Fields. *J. Chem. Phys.* **2020**, *153*, 035103.

(190) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and its Application to Yeast Membranes. *Biophys. J.* **2009**, *97*, 50–58.

(191) Lee, J.; Patel, D. S.; Ståhle, J.; Park, S. J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **2018**, *15*, 775–786.

(192) Park, S.; Choi, Y. K.; Kim, S.; Lee, J.; Im, W. CHARMM-GUI Membrane Builder for Lipid Nanoparticles with Ionizable Cationic Lipids and PEGylated Lipids. *J. Chem. Inf. Model.* **2021**, *61*, 5192–5202.

(193) Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007**, *2*, e880.

(194) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.

(195) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J. P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(196) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.

(197) Chu, H.; Peng, X.; Li, Y.; Zhang, Y.; Min, H.; Li, G. Polarizable Atomic Multipole-Based Force Field for DOPC and POPE Membrane Lipids. *Mol. Phys.* **2018**, *116*, 1037–1050.

(198) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity. *science* **1998**, *280*, 69–77.

(199) Zhorov, B. S.; Tikhonov, D. B. Potassium, Sodium, Calcium and Glutamate-Gated Channels: Pore Architecture and Ligand Action. *Journal of Neurochemistry* **2004**, *88*, 782–799.

(200) Tikhonov, D. B.; Zhorov, B. S. Modeling P-Loops Domain of Sodium Channel: Homology with Potassium Channels and Interaction with Ligands. *Biophysical Journal* **2005**, *88*, 184–197.

(201) Khan, A.; Romantseva, L.; Lam, A.; Lipkind, G.; Fozzard, H. A. Role of Outer Ring Carboxylates of the Rat Skeletal Muscle Sodium Channel Pore in Proton Block. *The Journal of Physiology* **2002**, *543*, 71–84.

(202) Xiong, W.; Farukhi, Y. Z.; Tian, Y.; DiSilvestre, D.; Li, R. A.; Tomaselli, G. F. A conserved ring of charge in mammalian Na+ channels: a molecular regulator of the outer pore conformation during slow inactivation. *The Journal of Physiology* **2006**, *576*, 739–754.

(203) Eck, R. V.; Dayhoff, M. O. Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. *Science* **1966**, *152*, 363–366.

(204) Kolosov, P. Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Research* **2005**, *33*, 6418–6425.

(205) Schulz, G. E.; Schirmer, R. H., *Principles of Protein Structure*; Springer New York: 1979.

(206) Pan, X.; Li, Z.; Zhou, Q.; Shen, H.; Wu, K.; Huang, X.; Chen, J.; Zhang, J.; Zhu, X.; Lei, J.; Xiong, W.; Gong, H.; Xiao, B.; Yan, N. Structure of the Human Voltage-Gated Sodium Channel $Na_v1.4$ in Complex with $\beta 1$. *Science* **2018**, *362*.

(207) Lecar, H.; Larsson, H. P.; Grabe, M. Electrostatic Model of S4 Motion in Voltage-Gated Ion Channels. *Biophysical Journal* **2003**, *85*, 2854–2864.

(208) Huang, W.; Liu, M.; Yan, S. F.; Yan, N. Structure-Based Assessment of Disease-Related Mutations in Human Voltage-Gated Sodium Channels. *Protein &amp Cell* **2017**, *8*, 401–438.

(209) Tang, Z.; Chen, Z.; Tang, B.; Jiang, H. Primary erythromelalgia: a review. *Orphanet Journal of Rare Diseases* **2015**, *10*.

(210) Schrauber, H.; Eisenhaber, F.; Argos, P. Rotamers: To Be or not to Be? *Journal of Molecular Biology* **1993**, *230*, 592–612.

(211) Dunbrack, R. L. Rotamer Libraries in the 21st Century. *Current Opinion in Structural Biology* **2002**, *12*, 431–440.

(212) Mohebali, G.; Ball, A. S. Biodesulfurization of Diesel Fuels – Past, Present and Future Perspectives. *International Biodeterioration amp; Biodegradation* **2016**, *110*, 163–180.

(213) Soleimani, M.; Bassi, A.; Margaritis, A. Biodesulfurization of Refractory Organic Sulfur Compounds in Fossil Fuels. *Biotechnology Advances* **2007**, *25*, 570–596.

(214) Monticello, D. J. Biodesulfurization and the Upgrading of Petroleum Distillates. *Current Opinion in Biotechnology* **2000**, *11*, 540–546.

(215) Lee, W. C.; Ohshiro, T.; Matsubara, T.; Izumi, Y.; Tanokura, M. Crystal Structure and Desulfurization Mechanism of 2 -Hydroxybiphenyl-2-sulfinic Acid Desulfinase. *Journal of Biological Chemistry* **2006**, *281*, 32534–32539.

(216) Yu, Y.; Mills, L. C.; Englert, D. L.; Payne, C. M. Inhibition Mechanisms of Rhodococcus Erythropolis 2 -Hydroxybiphenyl-2-sulfinate Desulfinase (DszB). *The Journal of Physical Chemistry B* **2019**, *123*, 9054–9065.

(217) Caro, A.; Boltes, K.; Letón, P.; García-Calvo, E. Description of By-product Inhibiton Effects on Biodesulfurization of Dibenzothiophene in Biphasic Media. *Biodegradation* **2008**, *19*, 599–611.

(218) Abin-Fuentes, A.; Leung, J. C.; Mohamed, M. E.-S.; Wang, D. I.; Prather, K. L. Rate-Limiting Step Analysis of the Microbial Desulfurization of Dibenzothiophene in a Model Oil System. *Biotechnology and Bioengineering* **2013**, *111*, 876–884.

(219) Gray, K. A.; Mrachko, G. T.; Squires, C. H. Biodesulfurization of Fossil Fuels. *Current Opinion in Microbiology* **2003**, *6*, 229–235.

(220) Ansari, F.; Grigoriev, P.; Libor, S.; Tothill, I. E.; Ramsden, J. J. DBT Degradation Enhancement by Decorating Rhodococcus Erythropolis IGST8 with Magnetic Fe3O4 Nanoparticles. *Biotechnol. Bioeng.* **2009**, *102*, 1505–1512.

(221) Bresnick, E.; Karjala, R. J. End-Product Inhibition of Thymidine Kinase Activity in Normal and Leukemic Human Leukocytes*. *Cancer Research* **1964**, *24*, 841–846.

(222) Mosner, J.; Mummenbrauer, T.; Bauer, C.; Sczakiel, G.; Grosse, F.; Deppert, W. Negative Feedback Regulation of Wild-type p53 Biosynthesis. *The EMBO Journal* **1995**, *14*, 4442–4449.

(223) Alves, R.; Savageau, M. A. Effect of Overall Feedback Inhibition in Unbranched Biosynthetic Pathways. *Biophysical Journal* **2000**, *79*, 2290–2304.

(224) Jemli, S.; Ayadi-Zouari, D.; Hlima, H. B.; Bejar, S. Biocatalysts: Application and Engineering for Industrial Purposes. *Critical Reviews in Biotechnology* **2014**, *36*, 246–258.

(225) Lee, W. C.; Ohshiro, T.; Matsubara, T.; Izumi, Y.; Tanokura, M. Crystallization and Preliminary X-ray Analyses of Desulfurization Enzyme DszB and Its C27S Mutant Complexed with Biphenyl-2-Sulfinic Acid. *Acta Crystallographica Section D Biological Crystallography* **2004**, *60*, 1636–1638.

(226) Han, S.-S.; Kyeong, H.-H.; Choi, J. M.; Sohn, Y.-K.; Lee, J.-H.; Kim, H.-S. Engineering of the Conformational Dynamics of an Enzyme for Relieving the Product Inhibition. *ACS Catalysis* **2016**, *6*, 8440–8445.

(227) Hu, X.; Robin, S.; O'Connell, S.; Walsh, G.; Wall, J. G. Engineering of a fungal -galactosidase to remove product inhibition by galactose. *Applied Microbiology and Biotechnology* **2010**, *87*, 1773–1782.

(228) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nature Reviews Nephrology* **2019**, *15*, 346–366.

(229) Krautter, F.; Iqbal, A. J. Glycans and Glycan-Binding Proteins as Regulators and Potential Targets in Leukocyte Recruitment. *Frontiers in Cell and Developmental Biology* **2021**, *9*.

(230) Van Kooyk, Y.; Rabinovich, G. A. Protein-Glycan Interactions in the Control of Innate and Adaptive Immune Responses. *Nature Immunology* **2008**, *9*, 593–601.

(231) Cheetham, J. C.; Smith, D. M.; Aoki, K. H.; Stevenson, J. L.; Hoeffel, T. J.; Syed, R. S.; Egrie, J.; Harvey., T. S. NMR Structure of Human Erythropoietin and a Comparison with Its Receptor Bound Conformation. *Nature Structural Biology* **1998**, *5*, 861–866.

(232) Shental-Bechor, D.; Levy, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences* **2008**, *105*, 8256–8261.

(233) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114*, 2549–2564.

(234) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(235) Geronimo, I.; Nigam, S. R.; Payne, C. M. Desulfination by 2 -hydroxybiphenyl-2-sulfinate desulfinase proceeds via electrophilic aromatic substitution by the cysteine-27 proton. *Chemical Science* **2017**, *8*, 5078–5086.

(236) Hu, Y.; Houk, K. N.; Kikuchi, K.; Hotta, K.; Hilvert, D. Nonspecific Medium Effects versus Specific Group Positioning in the Antibody and Albumin Catalysis of the Base-Promoted Ring-Opening Reactions of Benzisoxazoles. *Journal of the American Chemical Society* **2004**, *126*, 8197–8205.

(237) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation* **2005**, *1*, 1128–1132.

Electric Fields: A Metric for Molecular-level Understanding of Protein Mechanisms

Yi Zheng

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Valerie Vaissier Welborn, Chair
T. Daniel Crawford
Nicholas J. Mayhall
James M. Tanko

Apr 24, 2024
Blacksburg, Virginia

Keywords: protein mechanism, electric field, molecular dynamics simulations

Contents

List of Figures

0.1 Energy and RMSD of Chapter 3

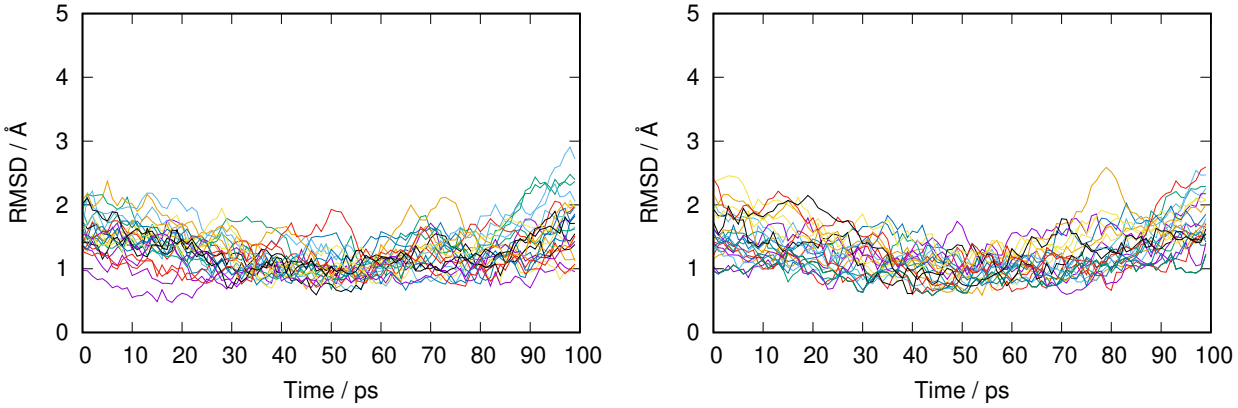

Fig. S1: RMSD of the protein backbone (C_α) and DNA for the KE15 - DNA-1 complex.

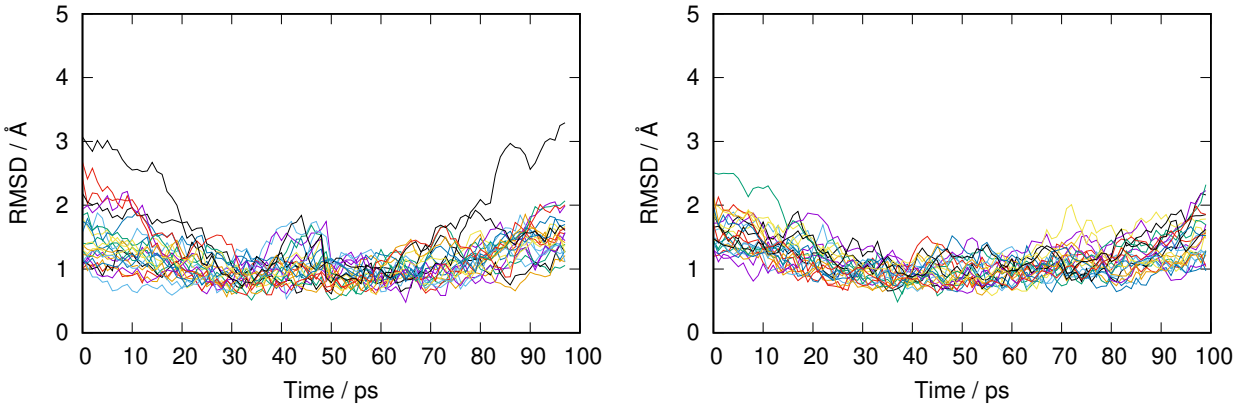

Fig. S2: RMSD of the protein backbone (C_α) and DNA for the KE15 - DNA-2 complex.

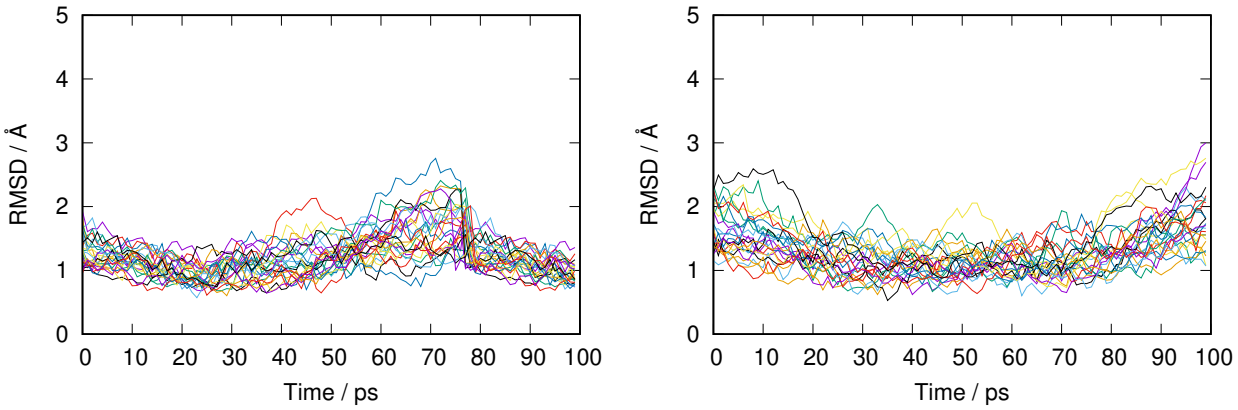

Fig. S3: RMSD of the protein backbone (C_α) and DNA for the KE15 - DNA-3 complex.

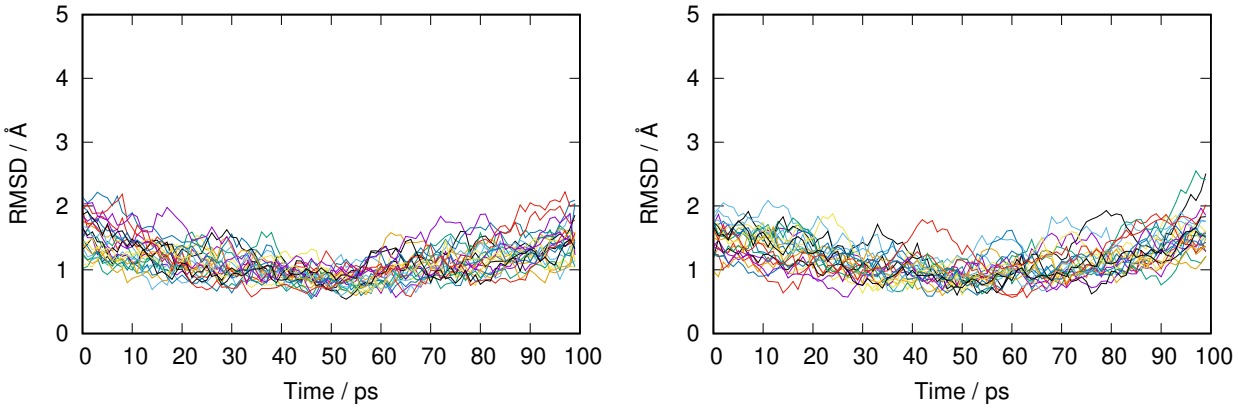

Fig. S4: RMSD of the protein backbone (C_α) and DNA for the KE15 - DNA-4 complex.

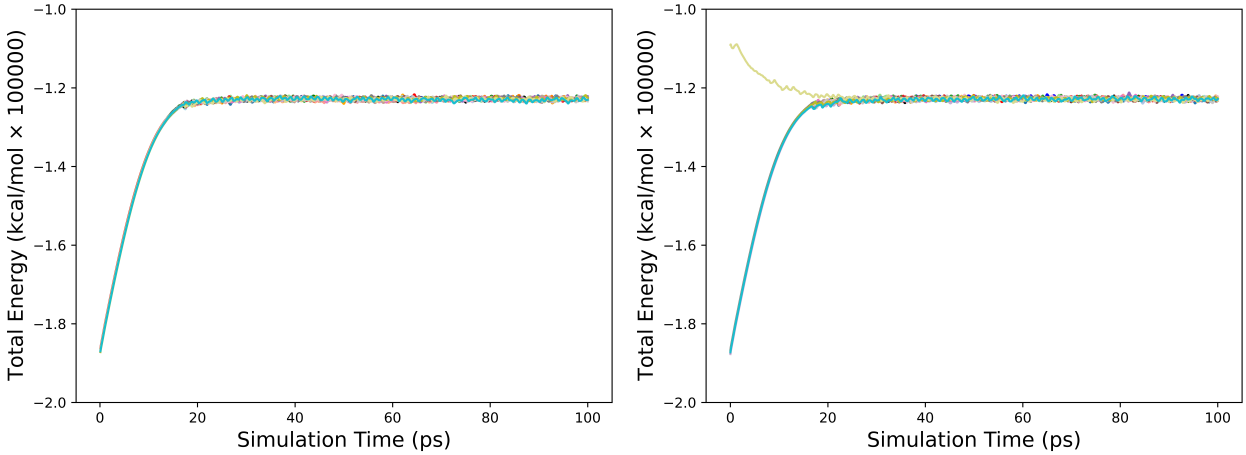

Fig. S5: Convergence MD plots for the KE15-DNA complex with DNA in position 1, reactant and transition state with 25 conformations in each state.

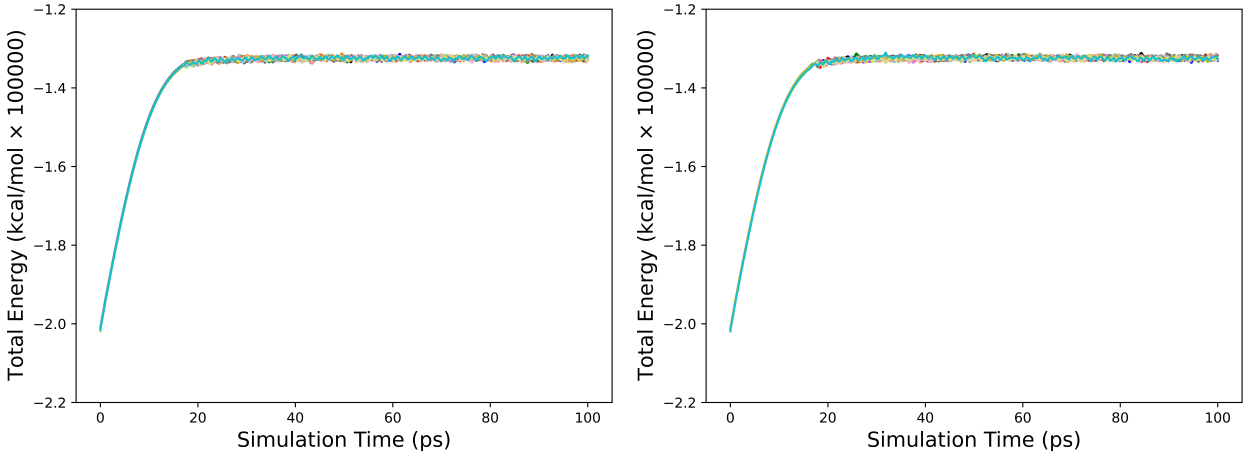

Fig. S6: Convergence MD plots for the KE15-DNA complex with DNA in position 2, reactant and transition state with 25 conformations in each state.

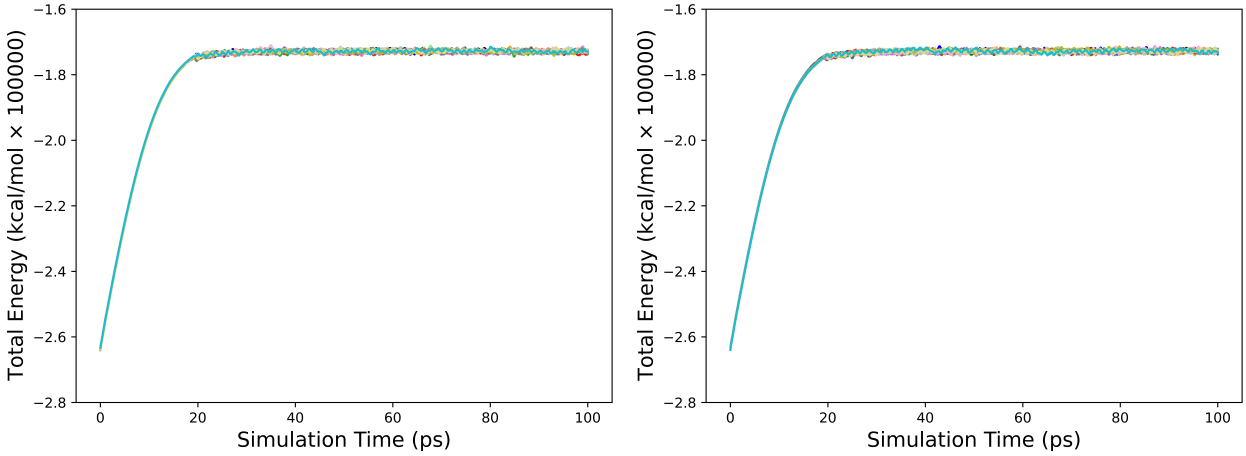

Fig. S7: Convergence MD plots for the KE15-DNA complex with DNA in position 3, reactant and transition state with 25 conformations in each state.

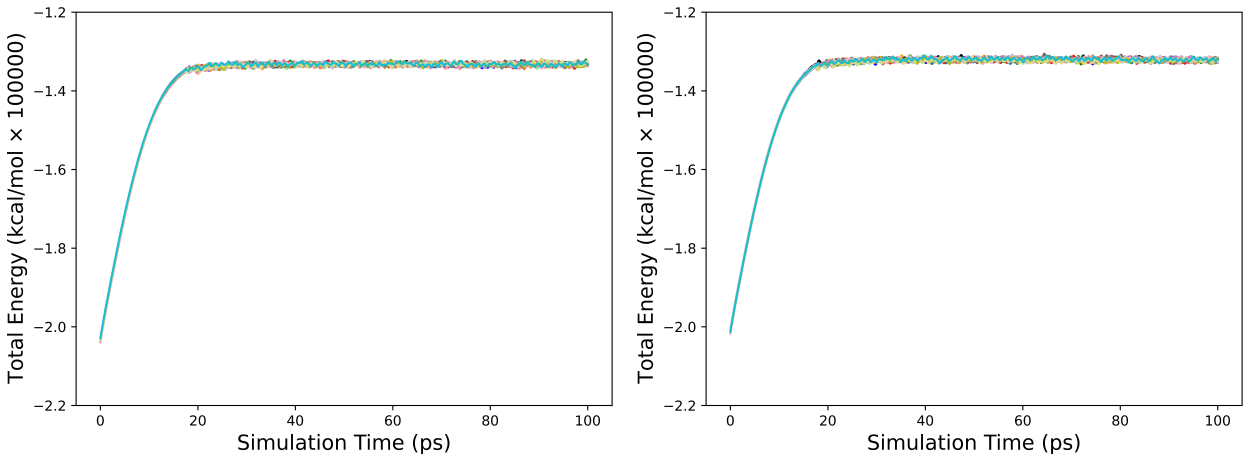

Fig. S8: Convergence MD plots for the KE15-DNA complex with DNA in position 4, reactant and transition state with 25 conformations in each state.

0.2 Energy and RMSD of Chapter 4

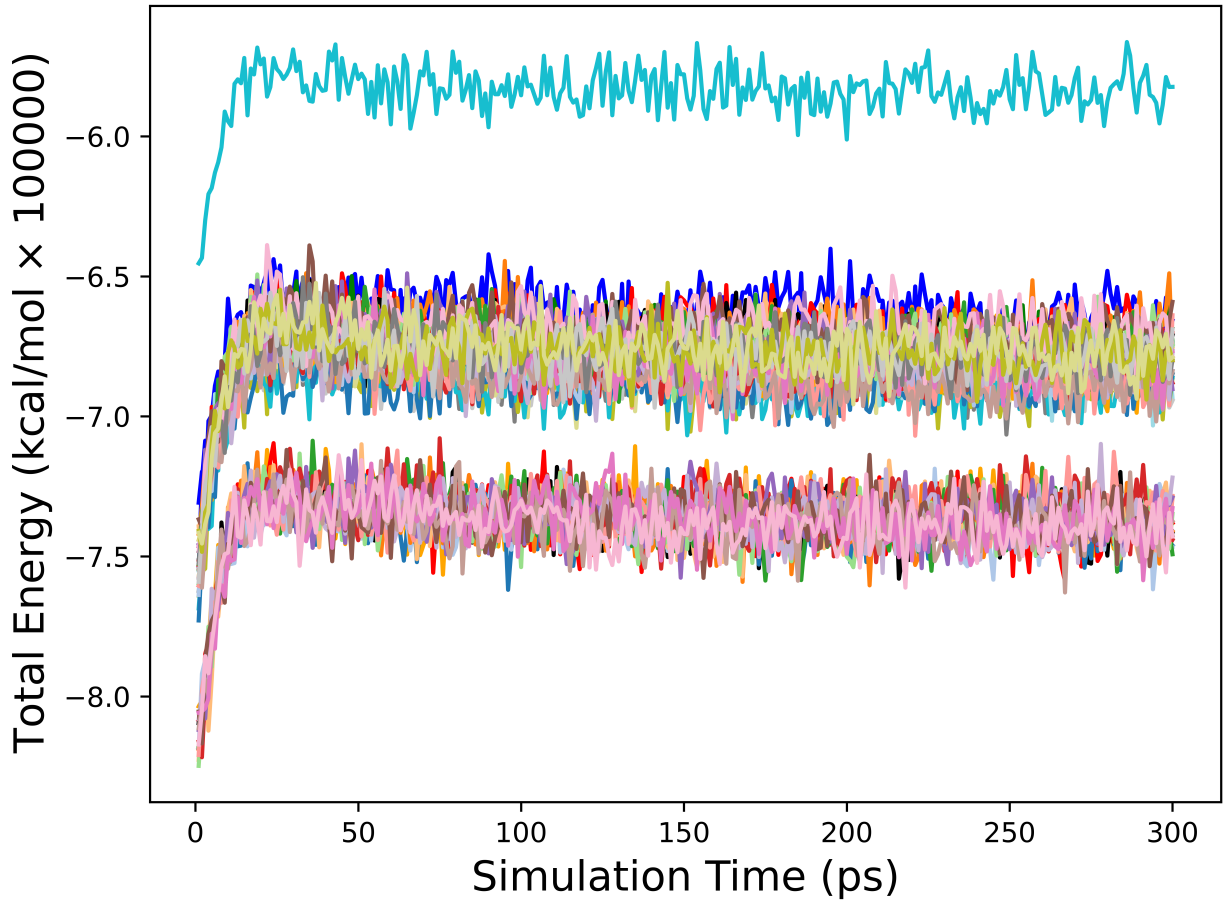

Fig. S9: Convergence MD plots for the 50 conformations of $Na_v1.7$.

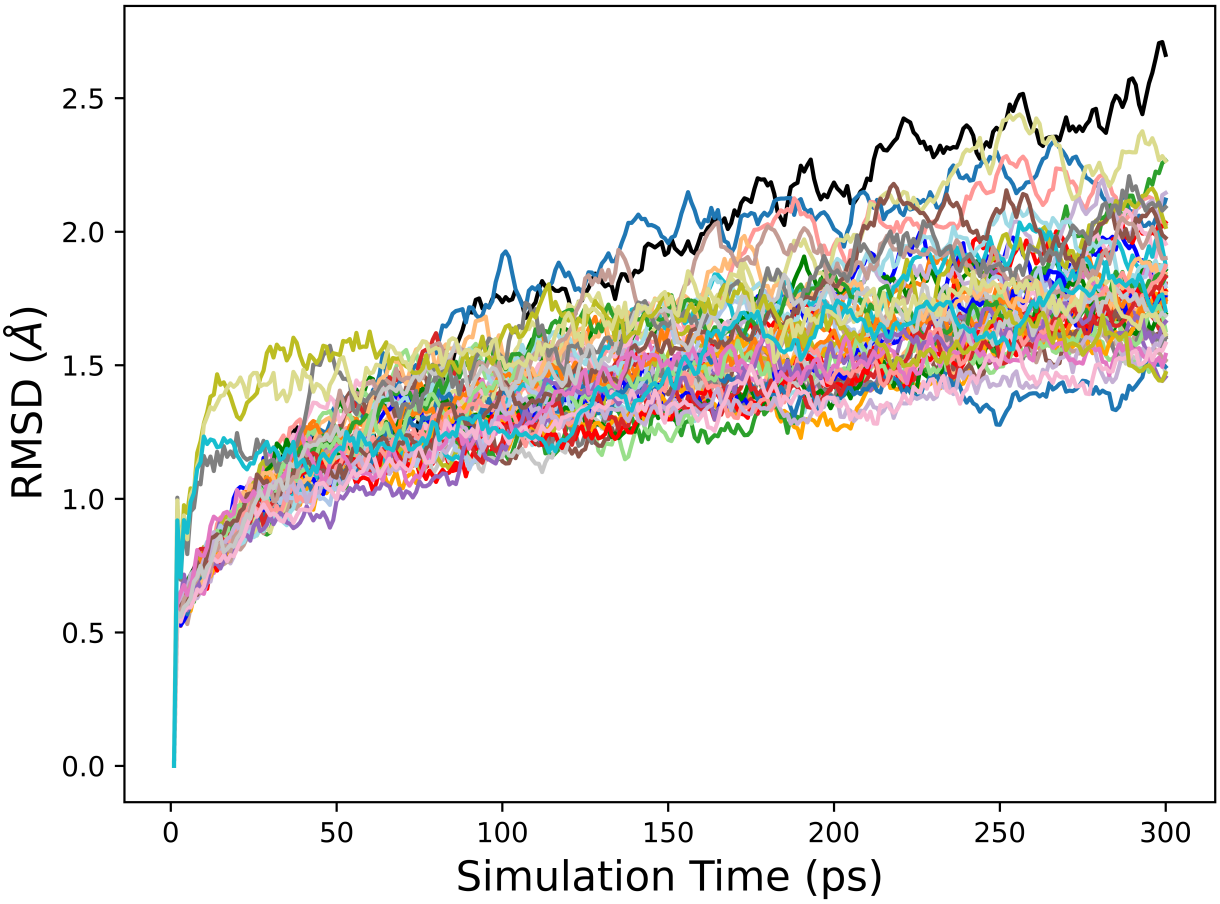

Fig. S10: RMSD of the $Na_v1.7$ backbone (C_α)

0.3 Energy and RMSD of Chapter 5

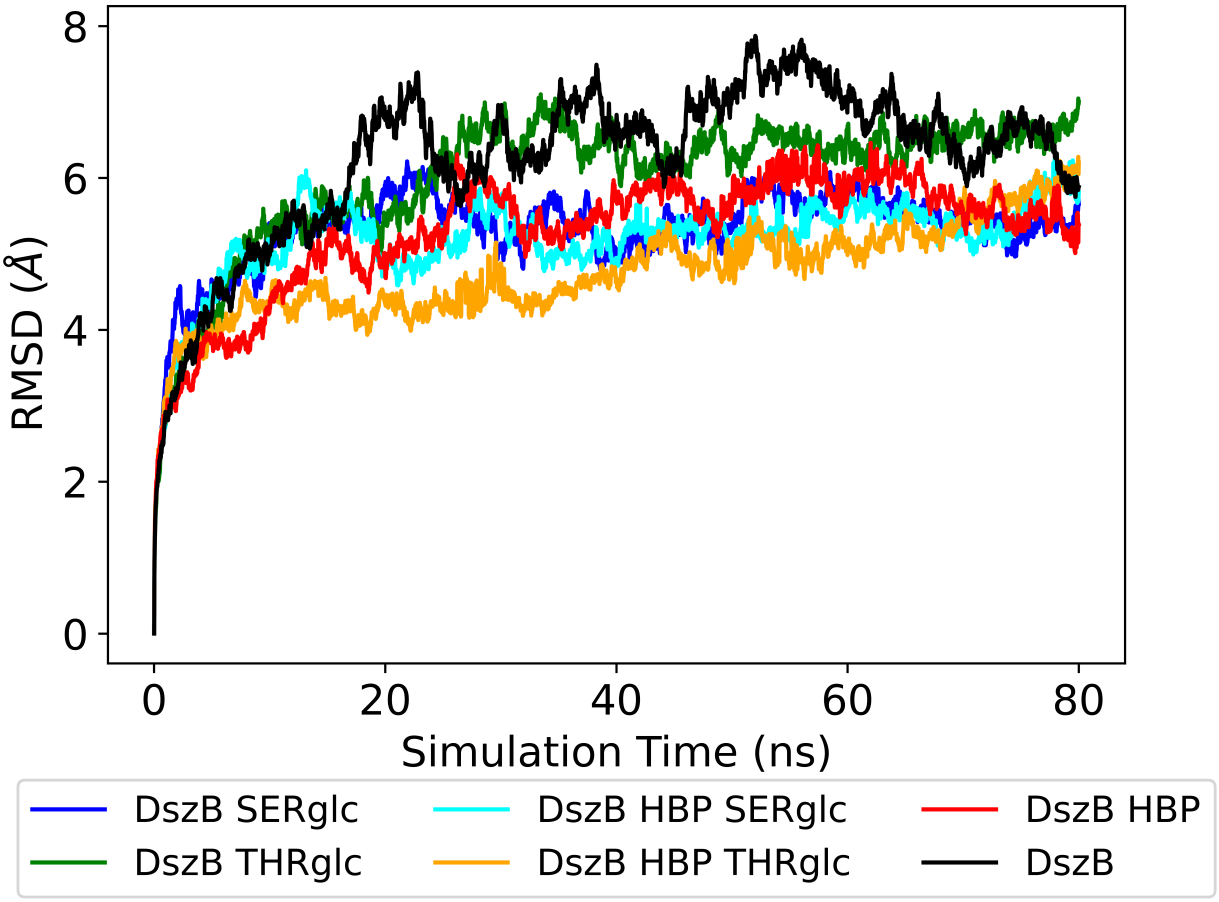

Fig. S11: RMSD of the protein backbone (C_α) for the six DszB system.

0.4 Relation Between Residue Distance and Electric Fields Magnitude

Here we report the relation between the residue-probe distance and their projected electric field contributions (Figures S12 and S13). The residue-probe distance was calculated as the average distance between C_α and probe over the production phase and conformational ensemble. To better analyze the electric fields contribution, we took the absolute value of the electric field projection.

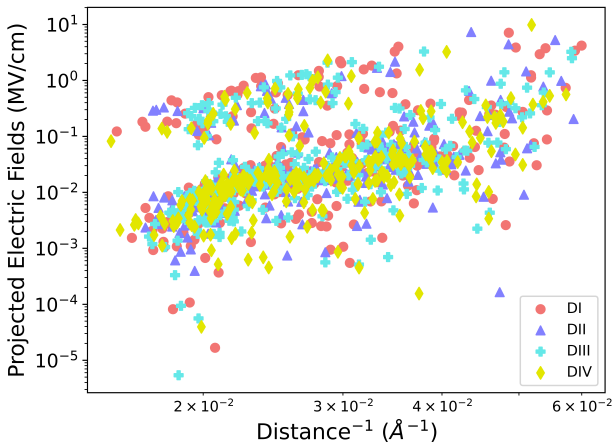

Fig. S12: The absolute value of projected electric field as a function of the distance between the C of each residue and the probe midpoint for the PD. The data points are colored based on the location of residues. DI is represented by red circle, DII by blue triangle, DIII by cyan plus, and DIV by yellow rhombus.

While more residues with significant contributions to the electric field are located on DI, we observe a similar distribution of data point between the 4 domains (Figure S12), indicating the significant contributors are not a result of proximity effect between residues and the probe. In addition, the furthest electric fields contributor to the probe is Asp 338, which has a 12 Å difference in distance from the closest contributor Gln 360. Tyr 1755 contributes the most to the electric field projection, although it is 2.6 Å further away from the probe than

Gln 360. Figure S13 shows the relationship between residue electric field contribution and residue-probe distance when probing each of the 4 VSDs. Similar to PD, residues within the same distance from the probe contribute differently to the electric field. At the same distance, DI experienced a stronger electric field than the other residues, while more contributor for DIII are located closer to the probe. This observation suggests that our calculation cannot be replaced by a distance analysis of residues and channel function motifs.

In addition, there is a clear separation between contributing and non-contributing residues at the same distance. The non-contributing residues all located near 0 MV/cm despite their distance from the probe. If the projection magnitude is greater than 3 MV/cm, we consider them to be significantly contributing, indicating that these residues are actively involved in regulating the channel mechanism. The distribution of these residues on the plot also showed no strong relation between the electric field magnitude and the distance. Thus, the calculation of the electric fields projection is not the reflection of distance, but provides crucial insights into the functional role of the residues.

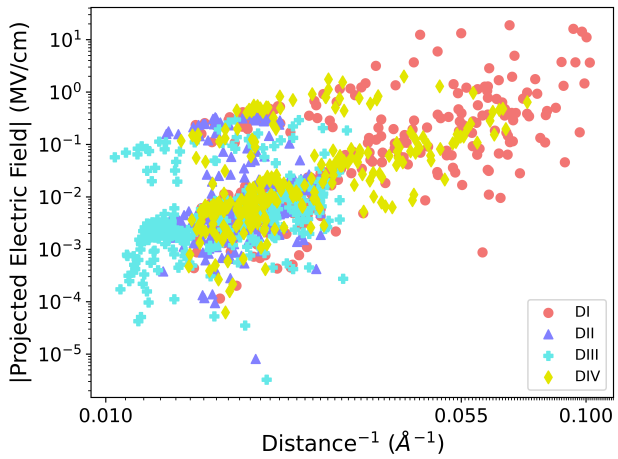

Fig. S13: Overlap of the absolute value of projected electric field as a function of the distance between the C of each residue and the probe midpoint for the 4 VSDs. DI is represented by red circle, DII by blue triangle, DIII by cyan plus, and DIV by yellow rhombus.