Evolutionary Guided Molecular Dynamics Driven Protein Design

by

Ismail Can Kazan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2023 by the
Graduate Supervisory Committee:

Sefika Banu Ozkan, Co-Chair
Giovanna Ghirlanda, Co-Chair
Jeremy Mills
Oliver Beckstein

ARIZONA STATE UNIVERSITY

August 2023

ABSTRACT

Natures hardworking machines, proteins, are dynamic beings. Comprehending the role of dynamics in mediating allosteric effects is paramount to unraveling the intricate mechanisms underlying protein function and devising effective protein design strategies. Thus, the essential objective of this thesis is to elucidate ways to use protein dynamics based tools integrated with evolution and docking techniques to investigate the effect of distal allosteric mutations on protein function and further rationally design proteins. To this end, I first employed molecular dynamics (MD) simulations, Dynamic Flexibility Index (DFI) and Dynamic Coupling Index (DCI) on PICK1 PDZ, Butyrylcholinesterase (BChE), and Dihydrofolate reductase (DHFR) to uncover how these proteins utilize allostery to tune activity. Moreover, a new classification technique (“Controller”/“Controlled”) based on asymmetry in dynamic coupling is developed and applied to DHFR to elucidate the effect of allosteric mutations on enzyme activity. Subsequently, an MD driven dynamics design approach is applied on TEM-1 β-lactamase to tailor its activity against β-lactam antibiotics. New variants were created, and using a novel analytical approach called "dynamic distance analysis" (DDA) the degree of dynamic similarity between these variants were quantified. The experimentally confirmed results of these studies showed that the implementation of MD driven dynamics design holds significant potential for generating variants that can effectively modulate activity and stability.

Finally, I introduced an evolutionary guided molecular dynamics driven protein design approach, integrated co-evolution and dynamic coupling (ICDC), to identify distal residues

that modulate binding site dynamics through allosteric mechanisms. After validating the accuracy of ICDC with a complete mutational data set of β-lactamase, I applied it to Cyanovirin-N (CV-N) to identify allosteric positions and mutations that can modulate binding affinity. To further investigate the impact of mutations on the identified allosteric sites, I subjected putative mutants to binding analysis using Adaptive BP-Dock. Experimental validation of the computational predictions demonstrated the efficacy of integrating MD, DFI, DCI, and evolution to guide protein design. Ultimately, the research presented in this thesis demonstrates the effectiveness of using evolutionary guided molecular dynamics driven design alongside protein dynamics based tools to examine the significance of allosteric interactions and their influence on protein function.

DEDICATION

Dedicated to my parents, my brother, and my late grandmother.

ACKNOWLEDGMENTS

I dedicate this thesis to my loving parents Nejat Kazan, and Dr. Dilek Kazan, my brother Cem Kazan, and my late grandmother Nermin. Without their continues support and love none of this could have been possible. I am forever grateful to have them in my life.

I would like to express my sincere appreciation to Dr. S. Banu Ozkan, my esteemed PhD advisor. Her guidance, support, and unwavering belief in my potential have been invaluable throughout my graduate studies. Her expertise and passion for knowledge have been truly inspiring, motivating me to push beyond my limits. I am immensely thankful for the opportunities she provided, including concrete projects and side projects, which allowed me to develop my skills and deepen my understanding of physics, chemistry, and biology. Moreover, her compassion, understanding, and encouragement have played a pivotal role in my personal and professional growth. I am truly fortunate to have had her as my advisor, and I extend my deepest gratitude for her mentorship and friendship. Her impact on my academic journey will be cherished forever.

I would like to express my heartfelt gratitude and sincere acknowledgment to the exceptional colleagues and coworkers who have played pivotal roles in shaping my journey as a researcher. Throughout my PhD, Dr. Tushar Modi, and Dr. Paul Campitelli served as invaluable friends, offering unwavering guidance, sharing their expertise, and providing crucial scripts and code that significantly aided my work. I am grateful to Dr. Brandon Butler, Dr. Avishek Kumar, for their valuable feedback which greatly improved the quality of my work. I also extend my thanks to the Fulton High-Performance Computing Initiative

for the vital computing resources that enabled me to carry out my computational work. Lastly, I would like to express my gratitude to Bethany Kolbaba Kartchner, Nikhil Ramesh, Jin Lu, and Nick Ose for their friendship over the years and our fruitful collaborations.

I would like to express my appreciation to my esteemed collaborators; and committee members Dr. Giovanna Ghirlanda, Dr. Jeremy Mills, Dr. Oliver Beckstein, Dr. Marcia Levitus, Dr. Mark Hayes, and Dr. Dmitry Matyushov who have played a pivotal role in shaping and enriching my research. Their invaluable guidance, expertise, and collaborative contributions, along with the support and insightful comments have been instrumental in shaping the success and advancement of my research endeavors.

Finally, I would like to express my sincere appreciation to my dear friends Semih Arslanlar, Melike Kabaoglu, Deniz Guclu, Sarp Dag, Emre Olcen, Can Orman, Dr. Can Altinbulakli, and Elcin Tunckol. Their love and encouragement have been an irreplaceable source of strength and motivation. I am truly fortunate to have such incredible individuals in my life, and I am grateful for their presence and support.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Proteins constitute a fundamental component of cellular machinery and hold immense significance in a plethora of biological processes. (Kessel and Ben-Tal, 2018; Lodish et al., 2000). Based on their involvement in specific processes, they are classified as: (a) structural proteins that make up the main structure of connective tissues (Dominguez and Holmes, 2011; Ricard-Blum, 2011), (b) regulatory proteins that are fundamental in regulation of the cell cycle (Bettencourt-Dias et al., 2004; Quon et al., 1998; Tyson et al., 2002), (c) transport proteins that are involved in moving molecules like nutrients and metabolites into and out of cells (André, 1995; Ayrton and Morgan, 2001; Griffith et al., 1992; Jack et al., 2001), (d) immune proteins that play a vital role in protecting the body against foreign invaders (Boulanger, 2009; Kaufmann, 1990; Vierstraete et al., 2004), and (e) enzymes that facilitate a wide range of chemical reactions necessary for the life of the cell. As indicated by these classifications, proteins are important in facilitating multiple biological processes.

In order to fulfill their functional roles, proteins adopt a structure comprising four primary levels. The primary structure of the protein is the 1D (one dimensional) order of amino acids which is referred to as its sequence. The secondary structure is known as repeating localized 2D patterns called alpha helices and beta sheets. The combination of the secondary structure elements makes up the tertiary structure of the protein which describes its 3D fold. Finally, the quaternary structure of a protein is its biological

assembly, and can consist of many tertiary structure subunits as in a complex (Bahar et al., 2017; Kessel and Ben-Tal, 2018). Considering all the levels, the structure of a protein is critical for its function, and its amino acid sequence dictates function. Even a minor change in its 1D amino acid sequence can have profound effects on its 3D fold and activity. Hence, the relationship between the sequence, structure, and function of proteins is of great significance in understanding biological systems. However, understanding this intricate relationship between protein sequence and function poses a formidable obstacle (Kazan et al., 2022; Modi et al., 2021a). Study of proteins, and understanding this relationship can help in prediction of structure and function of new and unknown proteins, providing a foundation for new discoveries in biology and medicine (Hospital et al., 2015).

The study of proteins has a long history, dating back to the early 19th century, when scientists first discovered the presence of proteins. By the 20th century, starting from the first studies related to amino acid sequence and structure determinations in the 1960s, in knowledge of the relationship between sequence, structure and function has been continued to expand (Hospital et al., 2015; Kendrew et al., 1960; Perutz et al., 1960; Stretton, 2002).

1.1 Structural Analysis of Proteins

Historically, investigations pertaining to proteins, their structure, and interactions focused on isolating, purifying, and characterizing individual proteins by using different experimental techniques such as X-ray crystallography, Nuclear magnetic resonance (NMR), Electron microscopy (EM), and Circular dichroism (CD) spectroscopy. Although these methods highlight structural features of proteins, they are expensive and due to technical complexities only limited number of protein targets could be studied at one time.

With the advent of computers and advancements in computational biology, it has become possible to study proteins on a larger scale using computational methods (Geng et al., 2019).

For that reason, beginning from the early 1960s, computational methods have been utilized to tackle various questions about the attributes of proteins. Hydrophobicity, a physicochemical property, is commonly employed to describe the secondary structures of proteins, and this property plays a crucial role in the initial interactions during protein folding. In 1962, the first hydrophobicity scale for amino acids was introduced to aid in prediction of folding/unfolding energetics of the protein by assigning different potentials based on hydrophobic and hydrophilic behavior of amino acid (Cid et al., 1992; Ponnuswamy et al., 1980; Simm et al., 2016; Tanford and Lovrien, 1962; Wilce et al., 1995; Zviling et al., 2005). Peptide bond potential functions give rise to more accurate calculations of protein conformational changes by allowing the energy of a peptide bond rotation to be estimated (Brant and Flory, 2002; Jorgensen and Tirado-Rives, 1988; Némethy and Scheraga, 1965; Zimmerman, 1985). In the late 1960s, Ramachandran plots were introduced. Focused on the conformations of amino acids, Ramachandran plots provide a graphical representation of the allowed and disallowed regions of the torsion angles (Carugo and Djinović-Carugo, 2013; Fowler et al., 2020; Hollingsworth and Karplus, 2010; Kendrew et al., 1958; Ramachandran and Sasisekharan, 1968; Vega et al., 2000).

In the 1970s, disulfide bridge prediction methods, secondary and tertiary structure prediction methods, and protein folding simulations were introduced. These methods allow

for estimation of protein stability, and continue with studies on protein kinetics and folding pathways (Argos et al., 1976; Beale and Buttress, 1969; Cantor and Schimmel, 1980; Chandrasekaran and Balasubramanian, 1969; Chou and Fasman, 1974; Froimowitz and Fasman, 1974; Guzzo, 1965; Holley and Karplus, 1989; Janin et al., 1978; Kao and Karlin, 1986; Kotelchuck and Scheraga, 1969; Levitt and Warshel, 1975; Lewis et al., 1970; McCammon et al., 1977; Némethy and Scheraga, 1977; Nishikawa, 1983; Prothero, 1966; Schiffer and Edmundson, 1967). Beginning in the 1980s and 1990s, rational protein design techniques were introduced. These techniques specialize in engineering proteins tailored for specific functions by modifying the protein structure. Continued by the advancements in protein structure prediction methods, where the structure of the protein is predicted using only the amino acid sequence, these techniques elucidate details on how most proteins fold into their functional forms (Argos et al., 1982; Blake and Johnson, 1984; Blundell et al., 1987; Cohen et al., 1980; Cohen and Kuntz, 1989; Duan and Kollman, 1998; Go, 1983; Jaenicke, 1987; Levitt and Warshel, 1975; Liwo et al., 1999; Sternberg and Thornton, 1978; Zemla et al., 1999).

Although having knowledge of the protein structure is a crucial step in rational design, the lack of sequence-structure correspondence poses a significant challenge, where high-throughput sequencing has generated vast protein sequence datasets without corresponding 3D structures (Marks et al., 2011). Since experimental structures can only be determined for a fraction of proteins, computational methods for protein structure modeling have gained importance, providing models suitable for various applications. One of the

revolutionary computational approaches to study proteins was protein structure modeling, in particular homology modeling.

The term "homology modeling" also known as comparative modeling, refers to the process of modeling the 3D structure of a protein by utilizing structural information from known configurations of similar proteins (Rodriguez et al., 1998). Homology modeling has been a useful tool in predicting the structure of new proteins based on the known structure of similar proteins (Geng et al., 2019; John and Sali, 2003). This method involves utilizing existing structure segments and energy evaluations to build protein structures, which can be accomplished by analyzing established protein structures (for example, Rosetta) (Rohl et al., 2004a). Additionally, deducing co-evolutionary signals between amino acid residues in homologous sequences can aid in the prediction of protein structures from scratch. The accuracy of predicted protein models can vary, but they can potentially be utilized in a variety of ways (Bradley et al., 2005; Marks et al., 2011; Rohl et al., 2004b; Zhang, 2009).

Homology models provide valuable information about the spatial arrangement of important residues in the protein, allowing for the study of binding sites and the design or docking of drugs. Stable and reliable repositories have been developed to provide access to these annotated and evaluated models. However, this approach does have limitations, one of which is that homology modeling assumes that the new protein will have a similar structure to the reference protein. In the recent years deep learning algorithms have surpassed the capabilities of homology modeling by predicting the fold of a protein by using only 1D sequence information (Binder et al., 2022). Although most of the proteins are considered to have a specific fold, they are not static entities. Rather they are considered

to be dynamic, as they constantly undergo conformational changes in response to their environment to perform their biochemical activities. Based on these limitations, in the 2000s, there was a transition towards utilizing computational techniques to investigate not only the structure and function relationship, but also the dynamics of proteins (Geng et al., 2019).

1.2 Dynamic Nature of Proteins and Computational Tools for Exploring Protein Dynamics

While protein structural analysis is valuable for the field, proteins are actually highly dynamic in nature, constantly undergoing a variety of motions and conformational changes. The ability of proteins to undergo conformational changes is essential for their function, as it allows them to carry out various tasks. The conformational changes observed in proteins arise from their interactions with surrounding molecules, and studying these conformations can provide insights into the molecular mechanisms underlying diverse biological processes (Bettati et al., 2011; Dill and Bromberg, 2010; Fuxreiter, 2014). For example, enzymes carry out specific chemical reactions by positioning their active sites in a specific conformation (Daniel et al., 2003; Geng et al., 2019; Kaltenbach and Tokuriki, 2014; Kazan et al., 2023; Kolbaba-Kartchner et al., 2021). Additionally, changes in the dynamic behavior of proteins can be indicative of disease states or other physiological changes in an organism, making the study of protein dynamics an important area of research in biomedicine.

Techniques such as NMR spectroscopy, Fluorescence Resonance Energy Transfer (FRET), Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) are commonly

used experimental techniques to study protein dynamics and provide insight into their biological functions (Schirò et al., 2020). However, computational methodologies often serve as valuable complements to experimental analyses, offering rapid and efficient means to gain deeper insights into protein conformational changes (Andrusier et al., 2008). Hence, in the 2010s with the introduction of novel tools, the availability of increasingly powerful computational resources has accelerated progress in the field of protein dynamics research (Dill and Bromberg, 2010; Dill et al., 2008; Dill and MacCallum, 2012; Eisenberg et al., 2000; García de la Torre et al., 2000; Goddard et al., 2018; Gohlke et al., 2000; Kollman et al., 2000; Kuhlman and Baker, 2000; MacKerell Jr. et al., 2000; Nei and Kumar, 2000; Pandey and Mann, 2000; Schwikowski et al., 2000; Sreerama and Woody, 2000; Wang et al., 2000).

To this end, molecular docking and molecular dynamics (MD) simulations have been utilized to investigate protein behavior at the molecular level to understand their dynamic nature (Bahar et al., 2017; Bolia et al., 2014a; Bolia and Ozkan, 2016; Campitelli et al., 2020; Hansson et al., 2002; Hollingsworth and Dror, 2018; Hollingsworth and Karplus, 2010; Hospital et al., 2015; Kazan et al., 2022; Modi et al., 2021a). The original purpose of these methods was to allow theoretical physicists to study systems consisting of many interacting particles, such as atoms or groups of atoms, using the principles of classical mechanics (Binder, 1995; Durrant and McCammon, 2011; Piana et al., 2014; Rapaport, 2004; Shaw et al., 2008). Expanding on the same principles, the interactions between proteins and peptides/ligands can be investigated by the same tools (Hvidsten et al., 2009). These interactions can be either transient or long lasting, and they are governed by specific

chemical and physical properties of these interacting bodies. The study of protein-protein and protein-ligand interactions is an important field of research, as it provides insights into the molecular mechanisms that regulate various biological processes (Geng et al., 2019; Perez et al., 2016; Vendruscolo and Dobson, 2011).

1.2.1 Molecular Docking

In general, docking is initiated using a pre-existing protein structure and a ligand structure that are both experimentally solved using structural discovery techniques, and comprises two main steps: (i) the rapid creation of an ideal conformation where the protein and ligand are bound together, and (ii) the evaluation of the strength of the interaction between the protein and ligand in the resulting complex (Chodera et al., 2011; Cournia et al., 2017; Mobley and Dill, 2009). The first docking approach was developed in the early 1980s by Kuntz et al. (Kuntz et al., 1982). Since then, different attempts have been conducted to enhance docking algorithms and overcome the difficulties in docking (Brooijmans and Kuntz, 2003; Huang et al., 2006; Jones et al., 1997; Meng et al., 1992, 2011; Morris and Lim-Wilby, 2008; Pagadala et al., 2017; Pinzi and Rastelli, 2019; Taylor et al., 2002).

Most earlier docking methods utilize rigid docking in which the amino acids of receptor of the protein are restricted into rigid bodies, and only the target ligand is allowed to move around the protein's binding site while conducting energy minimization (Gerek and Ozkan, 2010; Totrov and Abagyan, 2008; Zacharias, 2010). Rigid docking presents a significant issue: proteins are not static and undergo a range of conformational changes. The task is arduous, demanding a high degree of accuracy at the expense of computational time, owing

to the intricate nature of the conformational space sampled during a binding event and the complexity of the energy function used to estimate affinities. (Gerek and Ozkan, 2010; Gray et al., 2003; Schneidman-Duhovny et al., 2005).

Therefore, novel docking techniques have been developed to tackle the aforementioned issue raised from rigid docking by incorporating receptor flexibility to a certain degree, which can be broadly classified into two categories: (i) induced fit docking and (ii) ensemble docking (Bolia and Ozkan, 2016; Cummings et al., 2005; Guterres and Im, 2020; Kitchen et al., 2004). These approaches are grounded in biological models that account for the variances between bound and unbound protein conformations (Andrusier et al., 2008; Bienstock, 2012; Lexa and Carlson, 2012; Totrov and Abagyan, 2008; Zacharias, 2010). Induced fit docking approach posits that proteins undergo continuous conformational changes induced by the approaching ligand, and upon attaining a bound conformational state, can maximize its interactions with the ligand molecule to form a complex (Mashiach et al., 2010; Sherman et al., 2006). On the other hand, the ensemble docking techniques leverage several tools, including molecular dynamics, energy minimization, Monte-Carlo minimization, and normal mode analysis to generate conformations prior to modeling binding (Andrusier et al., 2008; Cardozo et al., 1995; Chaudhury and Gray, 2008; Dominguez et al., 2003; Fitzjohn and Bates, 2003; Lexa and Carlson, 2012; Lindahl and Delarue, 2005; Marrink et al., 2007; May and Zacharias, 2008; Meiler and Baker, 2006; Morris et al., 2009; Noid, 2013; Ritchie, 2008; Smith et al., 2005; Wang et al., 2007). Ensemble docking methods differ from explicit protein flexibility modeling by considering protein flexibility before docking using a limited number of discrete protein conformations

(Cavasotto and Abagyan, 2004; Ding and Dokholyan, 2013; Lauck et al., 2010; Österberg et al., 2002). While these methods are popular, they can only model flexibility for a limited number of receptor residues and the time required for docking using these methods increases linearly with the number of structures in the ensemble. They also only permit limited backbone changes and side chain rotation sampling, and therefore cannot sample large-scale backbone conformational changes (Cozzini et al., 2008; Harmalkar and Gray, 2021; Hornak et al., 2006a, 2006b; Liu and Chen, 2016; Maier et al., 2015).

It is important to use effective and intelligent sampling strategies that mimic nature while generating ensembles from any of the aforementioned approaches to consider the dynamic nature of proteins. Because, the success of a docking approach depends on generating a receptor ensemble that encompasses a wide range of binding site conformations observed in nature, while excluding those that predict incorrect poses (Bolia et al., 2014a; Bolia and Ozkan, 2016; Harmalkar and Gray, 2021; Totrov and Abagyan, 2008).

In order to address these difficulties associated with molecular docking, a new flexible docking technique called Adaptive BP-Dock (Adaptive Backbone Perturbation-Dock) was developed by Bolia and Ozkan (2016). This method is based on Perturbation Response Scanning (PRS) (Atilgan et al., 2010; Atilgan and Atilgan, 2009; Bolia and Ozkan, 2016) which calculates the fluctuation responses of residues in a protein using linear response theory (LRT) (Amemiya et al., 2011; Essiz and Coalson, 2009; Ikeguchi et al., 2005; Manson and Coalson, 2012; Yang et al., 2014) and RosettaLigand which account for ligand flexibility and side chain rotamer sampling of the protein receptor. In Adaptive BP-Dock,

the receptor residues of the protein are simultaneously perturbed with a unit force, and a new receptor conformation is obtained using PRS. The ligand orientation is then optimized in this new perturbed receptor conformation through RosettaLigand's flexible ligand protocol (Meiler and Baker, 2006). The approach incorporates full backbone protein flexibility, and full ligand flexibility, while reducing the time and cost associated with traditional experimental methods, making it easier to study protein interactions on a larger scale while simulating the natural course of a binding event (Atilgan et al., 2010; Bolia et al., 2014a; Bolia and Ozkan, 2016; Gerek and Ozkan, 2011, 2010; Kazan et al., 2022).

Previous investigations involving the utilization of Adaptive BP-Dock have effectively demonstrated its rapid testing capabilities on PDZ-peptide, HIV-ligand, and CV-N-glycan systems (Bolia et al., 2014b; Bolia and Ozkan, 2016). Initial implementation of Adaptive BP-Dock exhibited promising outcomes in discerning between binders and non-binders. Nonetheless, enhancements to its prediction accuracy were warranted. A significant challenge persists in its application to protein-ligand systems exhibiting high flexibilities on both receptor and ligand. Moreover, incorporating the conformational changes occurring during the transition from an unbound state to a bound docked state poses a formidable modeling challenge in induced fit approaches. One such example is the lectin-glycan system. This system exhibits intricate complexities stemming from the flexible nature of ligand (glycans) and the promiscuous binding affinities of proteins (lectins) towards ligand (glycans). Therefore, to overcome these limitations, in this thesis, the capability of Adaptive BP-Dock is expanded to utilize the docked poses from end of the simulations to be iteratively fed back as an initial conformation to ensure an induced fit

docking approach and sample conformational changes going from unbound to bound states (Kazan et al., 2022). In addition, a new technique is implemented scaling the displacement of residues due to perturbations to optimize perturbed poses, allowing sampling of a diverse set of conformations. Details of Adaptive BP-Dock is explained in Chapter 2.

The extended version of Adaptive BP-Dock was employed to assess its accuracy in predicting the binding behavior of several protein systems such as WW domain (manuscript under review), PDZ domain of PSD-95 (manuscript in preparation), and CV-N (in chapter 7) (Kazan et al., 2022). The findings revealed a remarkable improvement in the results, indicating that the new enhancements incorporated into Adaptive BP-Dock enabled it to accurately capture the underlying trend in binding (Kazan et al., 2022). Building upon the success achieved with known targets, a blind prediction study was conducted whereby Adaptive BP-Dock was employed to model binding interactions of novel mutants of CV-N towards dimannose. Subsequent experimental validation confirmed the accuracy of the predictions made by Adaptive BP-Dock, thereby demonstrating the success of the extended approach.

1.2.2 Driving Mechanical Insight from Molecular Dynamics (MD) Simulations

While molecular docking provides predicted structures of protein ligand complexes and corresponding empirical binding energies, another commonly used more physical approach is molecular dynamics (MD) simulations to sample protein conformations. MD involves solving classical equations of motion numerically based on the physical force on each particle (Geng et al., 2019; Hospital et al., 2015). However, the integration time steps must be small, typically femtoseconds (10^{-15} s), and understanding biologically relevant events

(e.g., protein folding), ranging to microseconds (10^{-6} s), necessitates a large number of numerical calculations (Dill and Bromberg, 2010; Karplus and McCammon, 2002; Shaw et al., 2008). In the past 40 years, the time scales achieved through atomistic MD simulations have been rapidly increasing, surpassing the growth rate of Moore's law (Karplus and McCammon, 2002; Shalf, 2020; Vendruscolo and Dobson, 2011). As a result, these simulations allow us to observe the changes in the protein structure over time (e.g., dynamics). Moreover, similar to docking, MD simulations can also be used to model protein-ligand and protein-peptide interactions, providing valuable insights into the detailed energetics and mechanics of these interactions at the expense of computation time (Gilson and Zhou, 2007; Guterres and Im, 2020).

Besides investigating the dynamic behavior of proteins, MD simulations are a powerful computational tool for understanding the impact of mutations on protein function (Chen et al., 2019; Jubb et al., 2017; Karplus and Petsko, 1990; Lori et al., 2013; Stefl et al., 2013; Wang et al., 2020, 2011; Xiong et al., 2021). The stability and activity of a protein are heavily dependent on its amino acid sequence, and changes to critical regions of the protein sequence can compromise its function (Campitelli et al., 2020; Modi et al., 2021a). For instance, a mutation can result in misfolding and aggregation, leading to various diseases such as Alzheimer's, Huntington's, cystic fibrosis, and cancer. The effects of a mutation on protein activity or function are influenced by several factors: the residue position where the mutation happens, the sequence background the mutation is added on, and the type of amino acid that is substituted (Kazan et al., 2022; Modi et al., 2021a). Deciphering the intricate connection between mutations and protein function poses a formidable challenge

owing to its multifaceted nature, necessitating a comprehensive investigation of the protein sequence.

The sequence of a protein carries the footprint of the taken evolutionary steps. The variations in the amino acid types would reflect the divergent evolutionary paths taken by various species leading to changes in activity or function. One of the most widely used techniques in investigating evolutionary history is detecting frequency of observing specific amino acid types (conservation) by using multiple sequence alignment (MSA), which involves aligning a large number of related protein sequences in order to detect evolutionary signals that may have been preserved over time. A high conservation suggests that the amino acid type on a location doesn't change over time, and an amino acid change on it would lead to detrimental outcomes. These positions are shown to be commonly located on functionally critical regions of the proteins (e.g., catalytic sites).

Recent studies have demonstrated that mutations occurring in regions distal to binding sites or catalytic regions can still have a significant impact on protein function. These mutations, which are distant from the functionally important sites and are not directly involved in catalysis or binding, but still play important roles in regulating the function of the protein, are named "allosteric mutations" (Kazan et al., 2022; Kolbaba-Kartchner et al., 2021). These distal mutations can serve as reliable indicators of allosteric regulation or the presence of allosteric effects, even in the absence of noticeable structural changes or major alterations in local dynamics near the mutation site.

The term "allostery" was first introduced in 1961 by Monod and Jacob to describe a type of inhibition where the inhibitor is not a structural analog of the substrate (Monod et

al., 1965). In the 1960s, two models: the concerted Monod-Wyman-Changeux (MWC) model, and the sequential Koshland-Némethy-Filmer (KNF) model were developed to explain allosteric effects, and for almost two decades, the concept of allostery was characterized by conformational changes. However, in 1984, Cooper and colleagues proposed a new allosteric model that did not involve conformational changes and introduced the term "dynamic allostery," which emphasized the contribution of entropy to allostery (Cooper and Dryden, 1984; Liu and Nussinov, 2016). Through utilization of dynamic allostery, distal mutations can act as dynamic allosteric regulators of function, affecting the protein's activity through interactions with parts other than functional sites of the protein.

The complexity of the relationship between mutations and protein function is further compounded by the fact that the same mutation can have different effects on function based on the current sequence background of the protein. This variation in the impact of a mutation highlights the limitations of predicting the effects of a mutation based solely on sequence information (Bahar et al., 2017; Kazan et al., 2022; Liu and Nussinov, 2016; Palmer et al., 2015). Accurately identifying key residues and amino acid substitutions that contribute to disease related outcomes, loss of function, or enhanced activity is vital for understanding how proteins function and further lead design efforts.

In this thesis, to overcome these challenges, I utilized protein dynamics analyses tools Dynamic Flexibility Index (DFI) (Kumar et al., 2015b; Larrimore et al., 2017) and Dynamic Coupling Index (DCI) (Campitelli et al., 2020; Larrimore et al., 2017). These tools are explained in Chapter 2 in details. Briefly, DFI provides a means of quantifying

the degree of flexibility of a particular residue, which can be useful in understanding the impact of mutations on protein function. DCI, on the other hand, provides a measure of the degree of dynamic coupling between two residues. Particularly computed dynamic coupling analysis between a position and functionally important residues can help identify distal residues that could affect protein function upon mutations (i.e., allows us to distinguish allosteric residues).

Consequently, in Chapter 3, I applied DFI and DCI on PICK1 PDZ domain to study the change in peptide binding specificity and gain mechanistic insight on the dynamic differences modulating the binding. The PDZ family consists of small modular domains that binds to the C-terminal tail of different proteins to take part in allosteric modulation of various cellular signaling processes (Kennedy, 1995; Ponting, 1997; Stevens et al., 2022a). Given that these domains interact with a diverse range of peptides, it is essential to investigate how different binding partners induce varied allosteric effects on the same PDZ domain. Therefore, PICK1 PDZ domain, which can bind to different ligands, is an ideal model to explore this question and examine the network of interactions that contribute to dynamic allostery. The results indicate the change in binding affinity is reflected by unique dynamical changes at distal regions of PDZ. We showcase these differences focusing on key residues identified in previous experimental studies (Christensen et al., 2019; Kalescky et al., 2015) by pointing out the drastic differences uncovered by dynamic coupling and network analyses.

In Chapter 4, I expanded the DFI and DCI analyses on an enzyme system plant-derived cocaine hydrolase variants Butyrylcholinesterase (pBChE) to investigate its kinetic

behavior. BChE functions in a dimeric form and with DFI and DCI, I uncovered unique dimerization dynamic and intricate change in coupling in-between monomer units indicating the enhancement of activity of mutant BChE is due to changes in its dynamics. Moreover, dynamic changes leading to an enhanced anticholinesterase scavenging ability were investigated. The results indicate that dynamic modeling of the protein can shine light on how new pBChE variant can attain a higher activity.

Moreover, in Chapter 5, DFI and DCI were used to investigate the effect of point mutations on the activity of *Escherichia coli* dihydrofolate reductase (DHFR) from a dynamics point of view. It was demonstrated that mutations distal to functionally important loops, specifically the M20 and FG loops, could have an impact on these loops in DHFR. To investigate this phenomenon, MD simulations were conducted to study the dynamics of these loops in wild-type DHFR using DIF and DCI metrics. It's important to note that the DCI score between two distant, non-interacting residues is not necessarily symmetrical due to the complex conformational dynamics of a protein. To address this, I further extended the DCI metric by introducing a new classification technique termed DCI_{asym}, an asymmetric variant of DCI which quantifies the difference in fluctuation response of residue i when perturbing residue j compared to the response of residue j when perturbing residue i (DCI_{ij} - DCI_{ji}). DCI_{asym} helps determine which of the two residues exerts more influence over the motion between them. Using DCI_{asym}, I assigned “Controller” and “Controlled" labels based on the degree in asymmetric coupling with two functional loops: M20 and FG. If the functional loops have a higher DCI_{asym} score with distal residues, that distal residue is classified as “Controlled” otherwise “Controller”. This “Controller” and

"Controlled" classification discerned residues that are detrimental to function. It also identified residues that are non-conserved and dynamically dominating control over functional loops which are on average showing activity enhancing outcomes. The results further affirms that the "Controller" and "Controlled" classification outperforms other common metrics and can be used as complementary to them.

The findings obtained from the extensive investigation encompassing chapters 3, 4, and 5 have substantiated the computational prowess of the employed tools MD, DFI and DCI, thus solidifying their viability for tackling complex tasks including enzyme design and protein engineering. The subsequent exploration of these challenges in chapters 6 and 7 represents a natural progression building upon the established foundation.

1.2.3 Integration of MD with Co-evolutionary Analysis

In addition to the tools studying protein dynamics, integration of co-evolution between distal locations in proteins can lead to a better understanding of the structure function relationship (Bepler and Berger, 2021; de Juan et al., 2013; Wang et al., 2017; Xu, 2019). This can enhance the identification of allosteric mutations (de Juan et al., 2013; Pollock et al., 2012). The co-evolution of residue pairs is driven by their mutual dependence and the interplay between their impact on function and control of activity (Liu and Nussinov, 2016; Pollock et al., 2012). Therefore, with co-evolution, we can gain insight into the relationship between residue pairs and their interactions (i.e., the contacts they form in the 3-D structure). Although it has been of great interest in finding spatial contacts, co-evolution can also reveal residue pairs that are distant to each other but showing a coherence.

To study co-evolution, various strategies are used to analyze the evolution of protein residues and their interactions over the course of evolutionary trajectories. Some of the common strategies are mutual information (MI) (Ding et al., 2016; S.D. Dunn et al., 2008; Gloor et al., 2005; Shackelford and Karplus, 2007; Simonetti et al., 2013), direct information (DI) (Gouveia-Oliveira and Pedersen, 2007; Hopf et al., 2012; Hopf et al., 2019; Marks et al., 2011; Ovchinnikov et al., 2015), and direct coupling analyses (DCA) (Morcos et al., 2014a, 2011). These methods focus on the statistical dependence between pairs of residues in a protein sequence. This information is used to detect correlations between residues that may be indicative of co-evolution, particularly those in contact in the 3-D structure. Hence, by targeting functional sites in a protein, co-evolution can highlight residues that are evolutionarily coupled to them. I combined co-evolutionary analysis with protein dynamics, and in this thesis I aim to find allosteric mutations that could potentially modulate function distally (de Juan et al., 2013; Pollock et al., 2012).

By incorporating information from both dynamic and coevolutionary features, I hypothesize that the impact of distal mutations on protein function can be predicted. Moreover, residue positions can be altered by amino acid substitutions to enhance or impair protein function. This relationship is explored in chapter 6, where employed MD, DFI an DCI metrics with evolutionary conservation information to tackle an enzyme design challenge with TEM-1 β-lactamase that could potentially enhance the activity towards penam/cephem antibiotics. The active sites of TEM-1 are located in the center of the protein. This creates a unique challenge: i) mutations on the active sites and surrounding residues are previously shown to diminish activity, ii) there are limited number of residues

that are far from the active site. We utilized DFI and DCI applied on MD simulation trajectory to uncover these unique residues in TEM-1. An obstacle in modeling mutations on residues that are rigid is that these residues in TEM-1 are investigated to lead deleterious outcomes upon mutations. Hence, instead of directly mutating these residues we followed another strategy which involves designing residues around these rigid points with Rosetta software. The two sets of variants based on the location they are designed around (rigid or flexible) were subjected to MD simulations and then analyzed by DFI and DCI.

With the development of a novel approach termed dynamic distance analysis (DDA), which highlights the similarities of DFI profiles, I examined the dynamic profiles of dynamics based rigid and flexible designs. With DDA, we demonstrated that alterations by mutations in residues surrounding highly coupled (high DCI with catalytic sites) rigid (low DFI) residues affected enzymatic activity and stability, while targeting flexible (high DFI) uncoupled (low DCI with catalytic sites) residues maintained native-like properties. The results are confirmed by experimental characterization and indicate that evolutionary guided MD driven dynamics designs have a great potential in creating variants which is capable of modulating the activity and stability in a wide range.

As previously discussed, protein design has limitations stemming from the large combinatorics of residues and amino acid types available. From what was learned from applications of evolutionary tools, MD simulations, and post-MD dynamics analyses; I theorize that, an evolutionary guided molecular dynamics driven protein design scheme will narrow down the sample space, and significantly reduce time and resources spent for protein engineering challenges. For this end, following the success of previous enzyme

design challenge, in chapter 7, I developed integrated co-evolution and dynamic coupling (ICDC) (Kazan et al., 2022). ICDC involves the use of dynamic coupling and statistical co-evolution analysis to identify distal residues that modulate binding site dynamics through allosteric mechanisms. To validate ICDC, I analyzed the mutational fitness data of β-lactamase and discovered that rigid positions (low DFI) showing high co-evolution and dynamic coupling (high DCI) with the catalytic sites have significant impacts on function. After the verification with a complete enzyme dataset, I applied ICDC approach to Cyanovirin-N (CV-N), a lectin with specific di-mannose binding. Lectin-glycan systems are extremely challenging to study with high order of randomness stemming from glycans and promiscuity of lectins. Therefore, a computational approach to help design novel variants specifically tailored to glycan targets is necessary. Hence, I employed ICDC to identify allosteric positions and mutations to modulate binding affinity.

The impact of mutations on identified allosteric sites are further explored by subjecting putative mutants to binding analysis using Adaptive BP-Dock. The results of binding prediction with Adaptive BP-Dock revealed a critical residue, I34, which has potential in enhancing, abolishing, and not effecting binding. With these diverse effects, I investigated the mutations further to understand their dynamic characteristics compared to wild type. I34Y mutant show binding enhancing behavior. When the dynamics of the mutant is analyzed with DFI metric, it showed that the mutation rigidifies the binding site residue dynamics relative to wild type. A rigid binding site could potentially maintain interactions with the dimannose easier than a flexible binding pocket. Further investigation of the binding pocket revealed the volume of the pocket of the I34Y mutant is smaller than the

wild type. A smaller binding pocket is related to non mobile behavior seen with DFI and further confirms that the binding is modulated by this distal mutation. The computational predictions are verified by experiments showing the power of implementing MD, DFI, DCI, and evolution together to guide protein design. The results derived from chapter 7 can be used to make predictions about the impact of mutations on protein function, including the effect of mutations on stability, activity, and specificity (Kazan et al., 2022).

Based on all above studies conducted within the scope of this thesis, Chapter 8 concludes the thesis by summarizing the results obtained from using evolutionary guided molecular dynamics driven design with protein dynamics based tools to investigate the role of allosteric interactions in protein function. Overall, the identification of allosteric residues and the study of their impact on protein activity and stability using dynamics based design, and docking are crucial steps in understanding protein function, and the development of new drugs and therapies. The findings presented in this thesis highlight the success of these novel computational tools in characterizing allostery and the vital role it plays in regulating protein dynamics and therefore function.

CHAPTER 2

METHODS FOR INVESTIGATING PROTEIN DYNAMICS AND GUIDE PROTEIN DESIGN

2.1 Modeling Proteins on a Molecular Level

To gain mechanistic insight on intricate behavior of proteins, as mentioned earlier in the preceding section, the dynamic nature of them needs to be investigated. The native structure of the protein (excluding intrinsically disordered proteins) is its completely folded, functional conformation. This conformation is unique and is result of a combination of hydrogen bonding interactions, van der Waals interactions, and disulfide bonds between amino acids. Although native structure is often referred as a single conformation, proteins are dynamic in nature, hence they attain multiple conformations. This plethora of conformations are considered to be a proteins' functional ensemble. Within the conformational ensemble some of the conformations could be similar to each other, and these conformations could be further consolidated and classified as states of the protein. Different proteins attain different states based on the biological process they are serving. Even for a single protein, there could be multiple states and those states could have large conformational differences compared to each other. Investigation of these multitude of states and ensembles require impractical number of experimental resources which can surmount to large magnitude of expenses. Therefore, the use of computational techniques can serve as an initial screening with low cost of setup and high scalability capabilities. These computational tools favor rapid testing and can explore the protein conformational changes on timescales and atomistic details not practical for experimental techniques (Dill

and MacCallum, 2012; Kessel and Ben-Tal, 2018). Computational approaches used in this thesis are explained as follows in detail.

2.1.1 Molecular Dynamics

Molecular Dynamics (MD) simulations, used in this thesis, are one of the most widespread used computational approaches to examine the dynamic nature of proteins (Binder, 1995; Dill and Bromberg, 2010). In general, MD simulation is used to study the motion and behavior of atoms and molecules over time. In the context of proteins, MD simulate the movement of individual atoms in the protein and the surrounding solvent. The basic idea is to numerically solve the equations of motion for each atom, considering interatomic forces and interactions. By integrating these equations over time, the simulations can provide mechanical insights into the behavior and properties of proteins.

By solving Newton's equations of motion for a system of interacting particles, MD tracks the trajectory of every atom through time in these classical systems. Newton's equations of motion consider forces acting on particles. The forces between particles and their corresponding energetic potentials in MD are the results of the interactions within the all-atom system and is often defined by molecular mechanical force fields (Michaud-Agrawal et al., 2011). With these empirical force fields one can describe the total energy (potential) of the system by using the coordinates of atoms and the forces acting on them. Empirical component of force fields come from experiments or computational estimations of complex properties and are integrated as the parameters of the force field. By using the parameter set, the potential energy of the system could be calculated.

The many particle systems are commonly described by bonded and non-bonded interactions. Covalent bonds, electrostatic interactions, van der Waals interactions, Coulomb and Lennard-Jones potentials are some of the components of potential energy functions which sum up to the total potential of the system.

$$U = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\emptyset - \delta_n)]$$

$$+ \sum_{i<j} \sum_{j>i} \left\{ \in_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi \in_o r_{ij}} \right\} \quad (2.1)$$

Where ***U*** is the potential of the system of particles, and the terms in the equation related to bonds, and angles are modeled with harmonic expression. Dihedral potential is estimated using a Fourier term. The final term in the equation is related to van der Waals forces. It is the sum of Lennard-Jones potential, and Coulomb potential. Every term in the force field function could be tailored to a specific protein system or could be generalized. Hence, a diverse number of force fields and a set of parameters describing them have emerged. The history of force fields spans back to 1960s where only several components like Lennard-Jones potentials and dihedral angles of amino acids were used to describe small protein systems. With the advance of computational tools and widespread availability of computer hardware capable of executing complex algorithms, new and more advanced force fields have emerged. Assisted Model Building and Energy Refinement (AMBER) (Salomon-Ferrer et al., 2013a), Chemistry at HARvard Macromolecular Mechanics (CHARMM) (MacKerell Jr. et al., 2000), Optimized Potentials for Liquid Simulations (OPLS) (Bylesjö et al., 2006), and GROningen MOlecular Simulation (GROMOS) (Schmid et al., 2011) are

some of the most popular examples of force field packages encompassing a wide range of protein, protein-ligand, and DNA systems.

In addition to force fields, to generate a simulation condition as close to natural conditions as possible several other factors need to be included. Proteins often naturally exists in solutions and the solution contains various other molecules like water, ions, lipids, other proteins and have intrinsic properties such as temperature and pressure.

Water molecules are the most abundant particles in cells and interacts with every protein system (Kollman et al., 2000). Therefore, an accurate description (model) of water is crucial for effective MD simulations. Water is generally modeled in two different ways: implicit, and explicit. In the implicit approximation water is considered as a field around the protein. Oppositely, in the explicit definition, each and every water molecule around a protein is considered to have an actual 3D structure and interactions (i.e., both with protein and themselves) reproducing some of the characteristics of water. The advancements with these models come with an additional challenge: the complexity of the water model is directly correlated with the computational cost. Hence depending on the system in hand, the complexity of the water model (depending on the number of interaction points in the water molecule model) needs to be selected (Mark and Nilsson, 2001). The three point water model is widely accepted and often provides results correlating highly with its natural form leading to better density and diffusion coefficients. To gather a deeper characteristic on the interactions of water with proteins, four and five point water models could be selected with increasing computational expenses. While water models could be explicit around the proteins, the number of water molecules can immensely impact the

computational time. Hence, a feasible number of water molecules to simulate proteins needs to be evaluated. To achieve this, first a simulation box needs to be defined to restrict the amount of water molecules to a finite amount, meanwhile allowing the simulation to process to take into account the full picture of the surrounding environment.

The simulation box in MD is referred as the large enough container (or water box) which holds both the protein and water molecules. While this boundary condition effectively reduces computational cost, the interaction of water molecules with the box boundary needs to be evaluated. To overcome this issue of molecules interacting with a box wall, periodic boundary condition (PBC) approaches have been developed (Makov and Payne, 1995). PBC allows the simulation box to repeat itself infinitely next to each other. With this approach, when a molecule passes the boundary conditions, it appears on the other side of the box. Although it comes with a clear advantage, the PBC can lead to protein interacting with the clone of itself on the next box. To overcome this, a distance from the protein to the boundary of the box needs to be defined that diminishes the interactions of proteins with themselves. The distance is often selected by considering the electrostatics. The Debye length is a measure of distance where the electric field or the effect of electrostatics of the protein becomes negligible. To this end in the selection of box conditions, a minimum distance (selected as 16Å for the MD simulations in this thesis) from the protein is utilized.

Temperature and pressure are intrinsic properties and defines the thermodynamical properties of the system (Berendsen et al., 1984; Feller et al., 1995). In MD simulations, temperature is commonly referred as thermodynamic temperature and is a measure of the

average kinetic energy of particles. Temperature in MD simulations are regulated by algorithms that add/remove heat from the system called thermostats. Acting as a heat bath thermostat help keep temperature at a desired level. In a familiar manner the pressure is regulated by barostats. In consideration with these properties, proteins can be simulated with MD in nature-like thermodynamical conditions.

In this thesis, for protein systems of DHFR and β-lactamase (Chapter 5 and 6), the MD simulations were conducted with AMBER software package. The initial conformations were solvated with TIP3P three point water models in a simulation box with size calculated by measuring the minimum distance from the proteins. The systems are then neutralized by adding sodium and chlorine ions. For the parametrization of the systems AMBER ff14SB force filed parameter set was used. The initial system is minimized with steepest descent and conjugate gradient algorithms. The system is then heated up to 300K, and simulated in the isothermal, isobaric, constant number of particles ensemble (NPT) under 1 bar pressure. Langevin thermostat, Berendsen barostat is utilized to regulate temperature and pressure, respectively. The production trajectories were run for 2μs.

For Lectin-glycan system (CV-N-dimannose) in Chapter 7, GROMACS (Gromacs version 2018.1) package is used for MD simulation. The solvation box is set using the same technique explained previously. CHARMM36 force field is used for parametrizing the systems. The systems were neutralized with potassium and chlorine ions, followed by steepest descent minimization. Initial system is simulated for 5ns using Berendsen approach and then switched to Nose-Hoover thermostat and Parrinello-Rahman barostat. The production simulations were run for 2 μs at 300K temperature and under 1bar pressure.

2.1.2 Molecular Docking

In this thesis, I utilized molecular docking to model the protein ligand interactions in a coarse grained manner and predict binding affinities. Molecular Docking approach stems from the idea of predicting the binding of a ligand to a protein target (Brooijmans and Kuntz, 2003; Huang et al., 2006). Ligands and proteins can interact in many different conformational patterns. Ligands are often much smaller in size compared to proteins. A ligand which interacts with one protein in a specific way could show a totally different pattern of interaction with another protein. This difference creates an immense challenge on modeling interactions of ligands with proteins as the combinatorics problem is unattainable. Protein Data Bank (PDB) contains structural information on both proteins and ligands. While the number of available solved structures are increasing exponentially every year, considering the number of available conformations of ligand-protein complexes model, acquiring experimental data for a specific ligand-protein system is not probable. Hence, to explore this vast landscape of protein ligand interaction models, molecular docking is crucial with its rapid modeling capabilities.

Molecular docking consists of five vital steps in general. These steps include input preparation, protein and ligand conformation search, docking ligand to conformation, scoring and ranking of docked poses, and finally formal analyses of the docked poses. Input data preparation begins with fetching an X-ray crystallography (Drenth, 2007) or nuclear magnetic resonance spectroscopy (NMR) (Emsley et al., 2013) protein PDB data and continues with either fetching the ligand structure form PDB or creating the ligand synthetically by utilizing computational chemistry approaches (e.g., quantum mechanics

calculations). Conformational exploration step involves searching and generating possible conformations for both protein and ligand individually. A common method is a Monte Carlo scheme which creates samples by introducing random changes (e.g., translational or rotational moves) to the molecules. The third step is docking the ligand to the protein which engages a Monte Carlo sampling method to model the interactions between protein and ligand while translating the ligand in the direction of the protein receptor (Meiler and Baker, 2006). The docked poses generated by the docking step is then scored by selected potential functions. The selection of a potential function has a large impact on the ranking therefore, several different scoring methods have been developed. A knowledge based potential is a function in which each energy term is derived from statistical observations (Alford et al., 2017). Conversely, an empirical scoring function contains terms originated from empirical evidence stemming from experiments (Korb et al., 2009). In this thesis both of these scoring functions were employed to rank protein-ligand docked poses. Finally following the scoring and ranking, a detailed analyses of the top scoring docked poses is necessary. The bound poses can highlight important aspects of the binding modes of the complex.

Each one of the molecular docking steps has their own challenges. Conformational sampling step is often considered to be the bottleneck of docking simulations. As previously discussed in MD simulation section, an all atomistic view of proteins comes with extensive simulation time. To overcome this challenge coarse graining techniques have been employed. I employed Elastic Network models (ENM) and Perturbation Response Scanning (PRS) in this thesis (Atilgan et al., 2001, 2010). ENM coarse grains

the protein by linking atoms (e.g., commonly alpha-carbon atoms) with elastic springs. Despite simplifying the interactions between atoms, it has been shown that ENM captures equilibrium dynamics related motions of the protein (e.g., vibrational frequencies). For a protein with ***N*** number of atoms, the harmonic potential of the whole system, ***U***, could be defined as:

$$U = \sum_{i}^{N}\sum_{j}^{N}\frac{k_{ij}}{2}\left[\left(r_i - r_j\right) - \left(r_i^0 - r_j^0\right)\right]^2 \tag{2.2}$$

$$\left(r_i - r_j\right) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{2.3}$$

Where $r_i - r_j$ is the instantaneous distance between nodes ***i*** and ***j*** in the protein, and $r_i^0 - r_j^0$ is the distance when the system is in equilibrium. The spring constant for the spring connecting the ***i***th and ***j***th node is given as k_{ij}. With the potential equation, the motion of the protein around its equilibrium state can be described by investigating the normal modes. Assuming the protein is inside a potential well and there are outside perturbations to the system, the translational motion of the protein could be ignored. Therefore, by using Taylor's expansion we can expand Eq. 2.2 as:

$$U = \sum U(0) + \frac{1}{2!}\sum\frac{\delta^2 U}{\delta x^2} + \cdots \tag{2.4}$$

Where the first term of the equation is zero and could be neglected from the calculation as the system is in equilibrium, and the terms higher than second order has small contribution to the energy and could be ignored. We can write the revised version of Eq. 2.4 in a matrix form as:

$$U = \Delta\boldsymbol{R}\,\boldsymbol{H}\,\Delta\boldsymbol{R}^T \tag{2.5}$$

In which $\boldsymbol{H}$ is termed as the Hessian matrix containing the second order derivatives of the potential energy. To uncover the normal modes depicting the motion, single value decomposition is applied on Hessian matrix. The decomposition reveals eigenvalues and eigenvectors corresponding to frequencies and the directions of the motion of the protein. The first six eigenvalues and eigenvectors of the decomposition is zero as they are related to translational and rotational motion (translation in x,y,z coordinates plus rotation in x,y,z sums up to six). The eigenvectors with low non zero eigenvalues are related to the motions of the protein in directions that has functional relevance. By excluding the non zero eigenvalues and taking a pseudo inverse of the remainder Hessian matrix we can determine correlations between nodes. The Hessian inverse, $\boldsymbol{H}^{-1}$, which contains covariances is shown to be proportional to a covariance matrix.

Upon the model created by ENM, I apply PRS. PRS estimates the perturbation response of residues due to external forces exerted on each other position in the ENM network. PRS utilizes Linear Response Theory (LRT) (Ikeguchi et al., 2005; Yang et al., 2014), a mathematical framework, to describe the conformational changes resultant from external forces. In the context of molecular docking, PRS technique can be used to estimate the conformational changes of protein due to forces acting on the receptor of the protein by the ligand. With LRT, the forces exerted on the protein by the approaching/interacting ligand could be mimicked. PRS method first measures normal modes of the protein in the equilibrium (e.g., equilibrium positions), and uses these frequencies the protein vibrates at to recalculate new equilibrium positions by using harmonic potential imposed by ENM.

Following from Hessian inverse which contains position covariances, new positions of atoms upon a perturbation could be calculated.

$$\Delta \boldsymbol{R}_{3N\times 1} = \boldsymbol{H}^{-1}_{3N\times 3N}\, \Delta \boldsymbol{F}_{3N\times 1} \tag{2.6}$$

Where $\Delta \boldsymbol{R}$ is the displacement vector, and $\Delta \boldsymbol{F}$ is the external perturbation applied on the critical sites (e.g., active site residues). Updated positions for each residue are then calculated by:

$$\boldsymbol{R}_{3N\times 1} = \boldsymbol{R^0}_{3N\times 1} + \alpha \times \Delta \boldsymbol{R}_{3N\times 1} \tag{2.7}$$

In which the $\boldsymbol{R^0}$ contains the initial coordinates of the residues and $\boldsymbol{R}$ vector represents the new perturbed coordinates of the atoms. The scaling parameter, $\boldsymbol{\alpha}$, is introduced to control the magnitude of the positional changes. Residue positions which the external perturbations are exerted on are selected based on the protein receptor information. The direction of the forces is random, and the magnitude is re-adjusted based on the RMSD change (lower than 2 Angstroms) the applied forces cause to the system (comparing the perturbed pose coordinates with the original coordinates).

While the advantage of PRS is fast calculation of effect of ligand binding, one of the disadvantages is the lack of solvent effects. However, an implementation of PRS combined with other molecular tools including solvent effect could hinder this disadvantage. To achieve this, the novel molecular docking approach termed Adaptive BP-Dock have been developed (Bolia and Ozkan, 2016; Kazan et al., 2022).

Adaptive BP-Dock utilizes binding conformational sampling method PRS by using backbone perturbations (BP) to create perturbed poses of proteins interacting with ligand. This approach excels at adding conformational diversity to classical docking approaches

and induced fit sampling. Adaptive BP-Dock method begins with either a bound or an unbound conformation of a protein. When a bound conformation is used, docking approaches can easily sample the ligand conformation matching the binding pattern. Conversely, if an unbound protein conformation is used as a starting conformation, the protein itself needs to move through conformational landscape to find its lowest energy bound conformation that could be docked with the ligand. This requires sampling the binding induced conformational changes of protein and, PRS, which is included in Adaptive BP-Dock, can help determine these changes. Hence, in Adaptive BP-Dock, the protein structure is initially perturbed by external forces on protein receptor residues echoing the interactions with the ligand. The acquired perturbed pose is then subjected to energy minimization with AMBER software. This step re-models the side chain conformations that were neglected in the backbone perturbation step. Following minimization is the docking step. Docking step in Adaptive BP-Dock is provided by RosettaLigand approach (Meiler and Baker, 2006). RosettaLigand, part of Rosetta software, is a flexible ligand, rigid backbone docking scheme. This Monte Carlo method can sample the conformational changes of the ligand interacting with the receptor of the protein by holding the backbone of the protein rigid and only allowing only side chains to be flexible (i.e., have different torsional angles). RosettaLigand has been shown to be effective in docking small ligands and peptides to proteins while enabling full flexibility on ligand side. After the docking step the docked poses sampled are reranked by empirical scoring function X-score (Wang et al., 2002). X-score energy units (XEUs) calculated for protein-ligand complexes has been shown to provide higher correlation with experimental

binding energies. While the strategy provides a docked pose, the power of Adaptive BP-Dock comes from its iterative sampling. During sampling of conformations, the complex explores a vast energy landscape, and some of the binding modes discovered by the docking algorithm could be a local minima in the landscape rather than a global one. Adaptive BP-Dock enables the complex to escape from these local minima by applying external forces (Brownian kicks) to the system. The predictive power of Adaptive BP-Dock has been tested with its application on lectin-glycan system in this thesis (Chapter 7). The flow chart of the Adaptive BP dock approach, using CV-N as an example, is shown in Figure 2.1.

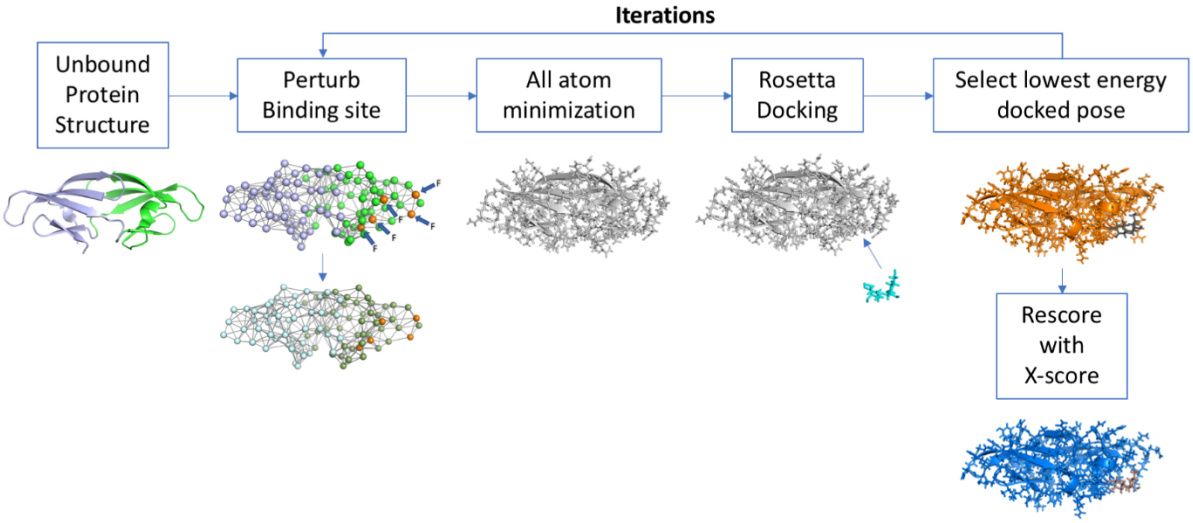


Figure 2.1: Flow chart of Adaptive BP-Dock, flexible docking approach (Kazan et al., 2022).

2.2 Evolutionary guidance in protein design

It is well established in nature that organisms are formed by repeated cycles of mutation, genetic drift, and natural selection, and only the ones that adapt to these changes have been able to survive over time (Hughes, 2005, 1997; Kessel and Ben-Tal, 2018;

Kreitman and Akashi, 1995; Ohno, 2013; Soskine and Tawfik, 2010; Waters and Vierling, 2020). Favorable genetic codes that provide survival are passed onto future generations and new species are evolved, while unfavorable codes are their hosts progressively being lost due to detrimental changes (Jäckel et al., 2008). During evolution, biochemical activities of proteins such as substrate binding and catalytic activities, and other features such as folding, and stability are changed due to mutations (Bordin et al., 2021; Jayaraman et al., 2022). These changes can result in the creation of new proteins with different functions, or the modification of existing proteins to improve their function or adapt to new environments due to insertion, deletion and substitution in amino acid types in their sequence (Bordin et al., 2021). Through studying the evolutionary history of proteins, we can gain insight into the relationship between residue pairs and their interactions (i.e., the contacts they form in the 3-D structure) (Bepler and Berger, 2021; de Juan et al., 2013; Wang et al., 2017; Xu, 2019). These interactions are associated with co-evolution, which is the coordinated changes in amino acid types that are observed in a residue pair. This interplay can result in the evolution of novel protein functions and is a crucial mechanism in the development of new biological systems and adaptation (Kazan et al., 2022; Larrimore et al., 2017). Co-evolved positions are dissected form the evolutionary history which is contained in the pool of similar sequences gathered from investigating protein family homology. The sequences in the pool are consolidated into a multiple sequence alignment (MSA) and the co-evolutionary information is revealed from the location pairs (Bordin et al., 2021; Campbell et al., 2016; de Juan et al., 2013; Rivoire et al., 2016; Salinas and Ranganathan, 2018; Torgeson et al., 2022; Yang et al., 2016).

Co-evolutionary data has emerged as a valuable tool for analyzing the three-dimensional (3-D) structural contacts of proteins (Jumper et al., 2021; Marks et al., 2012, 2011; Wang et al., 2017). By leveraging abundant sequence information, it allows us to calculate primary contacts that closely mimic realistic structural contacts, enabling the creation of accurate contact maps. These contact maps play a pivotal role in protein folding studies where only sequence information is utilized to predict 3D fold (Morcos et al., 2014b; Wang et al., 2016). Such insights provide valuable information about the spatial arrangement of residues and contribute to a deeper understanding of protein structure and function.

In this study, diverse statistical approaches were employed, including the utilization of MISTIC, EVcouplings, and RaptorX webservers. RaptorX server uses a deep neural network. The network leverages both structural and sequence information including multiple ortholog protein families with similar function and phylogeny (Xu, 2019). This strategy has demonstrated exceptional accuracy in predicting contacts compared to alternative methods. The EVcouplings approach is a Mutual Information (MI) based technique that considers both co-evolution and conservation to calculate novel Direct Information (DI) (Hopf et al., 2019). While MI approach effectively captures true contacts, it can inadvertently include both direct and indirect contacts due to its global nature. To address this limitation, the MISTIC web server introduces a correction term into the MI function, compensating for limited statistics in an MSA with a restricted number of sequences (Simonetti et al., 2013). This approach proves particularly advantageous when dealing with rare homologs and MSAs containing multiple gaps in their alignments. By

integrating these various methods, this study aims to provide highly accurate predictions of residue couplings. The analysis of co-evolutionary data, in conjunction with other computational tools, contributes to a deeper understanding of protein dynamics and interactions. These findings enhance our knowledge of protein structure-function relationships and pave the way for further advancements in the field of protein research.

2.3 Dynamic Flexibility Index (DFI) and Dynamic Coupling Index (DCI)

The flexibility of a protein could be investigated by Dynamic Flexibility Index (DFI) to gather position based scores related to its motion (Gerek and Ozkan, 2011; Larrimore et al., 2017). The concept continues from equation (2.6) in which $\Delta \boldsymbol{R}$ is the displacement vector, and Hessian inverse, $\boldsymbol{H}^{-1}$, which contains covariance information. Here we replace $\Delta \boldsymbol{F}$ as the random external Brownian kick applied on the system. The Hessian inverse was previously calculated by using ENM, but the covariance information could also be gathered from MD simulations. Using a position (x, y, x in Euclidean space) based covariance on MD trajectory, $\boldsymbol{H}^{-1}$ can be replaced by **G**, a coordinate covariance matrix.

$$\Delta \boldsymbol{R}_{3N\times 1} = \boldsymbol{G}_{3N\times 3N}\, \Delta \boldsymbol{F}_{3N\times 1} \tag{2.8}$$

When the process is repeated on every residue position to calculate residue response instead of solely on functionally critical residues, we can rewrite equation (2.8) as

$$\left|\Delta \boldsymbol{R}^{j}\right|_{i} = \sqrt{(\langle \boldsymbol{G}\, \Delta \boldsymbol{F}^{j} \rangle)^{2}} \tag{2.9}$$

Where a perturbation is applied on residue $\boldsymbol{j}$ and the magnitude of the average fluctuation response of another residue $\boldsymbol{i}$ is recorded. For every position the DFI score is calculated as:

$$DFI_{i} = \frac{\sum_{j=1}^{N}\left|\Delta \boldsymbol{R}^{j}\right|_{i}}{\sum_{i=1}^{N}\sum_{j=1}^{N}|\Delta \boldsymbol{R}^{j}|_{i}} \tag{2.10}$$

In which ***i*** and ***j*** are residue positions in the protein and ***N*** is the total number of residues. DFI metric measure the flexibility/rigidity of a position, or in other words the resilience of a position to a perturbation. If a position is displaced higher compared to all the other residues in the protein due to a perturbation, DFI score would be high (i.e., residue is flexible). Conversely, low DFI score indicates that the residue location shows low mobility (i.e., residue is rigid). Rigid sites are communication hubs in a protein with many interactions with their surrounding atoms. Conversely, flexible sites are very mobile. Perturbations on rigid locations (i.e., temperature change, or a mutation) would have more impact on the protein's dynamics compared to flexible locations. Hence, they are prone to perturbations.

Understanding and identifying residues that has an impact on protein function is vital. To be able to predict the dynamic allosteric regulation for a given position, Dynamic Coupling Index (DCI) is developed (Larrimore et al., 2017). DCI uses the same fundamental theories as DFI. It measures the dynamic coupling between functionally critical sites and distal residues. This strength of coupling could be understood by investigating the residue response due to a perturbation on functionally critical sites. The fluctuation response of a residue ***i*** is measured by applying perturbations on functionally critical sites one by one, ***j***, and calculating the average displacement.

$$DCI_i = \frac{N}{N_{Functional}} \frac{\sum_{j=1}^{N_{Functional}} \left|\Delta \boldsymbol{R}^j\right|_i}{\sum_{j=1}^{N} |\Delta \boldsymbol{R}^j|_i} \tag{2.11}$$

Where ***N*** is the positions in the protein, and $N_{Functional}$ is the functionally important residues, a subset of ***N***. With the use of DCI, one can highlight the residues which have the

highest dynamic allosteric coupling with active sites, and following, can target these highly coupled sites to further design activity enhancing mutants.

I utilized DFI and DCI approaches in Chapters 3 through 7. DCI highlights the allosteric dynamic coupling between active sties and distal residues. Interestingly the coupling could be asymmetric, meaning DCI value when we perturb an active site, ***i***, and check the response of a distal site, ***j***, could be different than when ***j*** is perturbed and the response of residue ***i*** is examined. We can understand this asymmetry with a novel metric, DCI_{asym}, which is calculated as:

$$DCI_{asym} = DCI_i - DCI_j \tag{2.12}$$

I utilized DCI_{asym} on Chapter 5 to investigate the relation between functionally critical sites and how they dominate control over the rest of the protein and defined novel "Controller" and "Controlled" classifications to identify possible activity enhancing positions.

CHAPTER 3

INVESTIGATING THE ALLOSTERIC RESPONSE OF THE PICK1 PDZ DOMAIN TO DIFFERENT LIGANDS WITH ALL-ATOM SIMULATIONS

This chapter is adapted from: "Stevens, A.O., Kazan, I.C., Ozkan, S.B., He, Y. (2022) Investigating the allosteric response of the PICK1 PDZ domain to different ligands with all-atom simulations Protein Science.31:e4474."

Amy O. Stevens conducted molecular dynamics simulations and network analyses. I. Can Kazan was responsible for the dynamic flexibility index and dynamic coupling index calculations.

To fully comprehend the impact of dynamic allostery in the complex regulation of protein function, one must consider a broader range of proteins. The binding of an allosteric ligand at a specific site on the protein can propagate conformational changes or alter the protein's dynamic behavior, leading to effects at distal sites. Hence, based on the protein's structure, the characteristics of the ligand, and the interactions between them; different binding partners on the same protein can induce different outcomes. In order to unravel the significance of dynamic allostery in regulating protein function, in this chapter, we investigated two protein-ligand systems: PICK1 PDZ-DAT and PICK1 PDZ-GluR2 by utilizing MD simulations. I applied DFI and DCI analysis and discovered that i) different ligands induce different dynamic changes upon binding, ii) both ligands show dynamic

allostery with distal αA helix, and iii) residue I35 is a crucial canonical binding residue for the dynamic allostery.

3.1 Abstract

The PDZ family is comprised of small modular domains that play critical roles in the allosteric modulation of many cellular signaling processes by binding to the C-terminal tail of different proteins. As dominant modular proteins that interact with a diverse set of peptides, it is of particular interest to explore how different binding partners induce different allosteric effects on the same PDZ domain. Because the PICK1 PDZ domain can bind different types of ligands, it is an ideal test case to answer this question and explore the network of interactions that give rise to dynamic allostery. Here, we use all-atom molecular dynamics simulations to explore dynamic allostery in the PICK1 PDZ domain by modeling two PICK1 PDZ systems: PICK1 PDZ-DAT and PICK1 PDZ- GluR2. Our results suggest that ligand binding to the PICK1 PDZ domain induces dynamic allostery at the αA helix that is similar to what has been observed in other PDZ domains. We found that the PICK1 PDZ-ligand distance is directly correlated with both dynamic changes of the αA helix and the distance between the αA helix and βB strand. Furthermore, our work identifies a hydrophobic core between DAT/GluR2 and I35 as a key interaction in inducing such dynamic allostery. Finally, the unique interaction patterns between different binding partners and the PICK1 PDZ domain can induce unique dynamic changes to the PICK1 PDZ domain. We suspect that unique allosteric coupling patterns with different ligands may play a critical role in how PICK1 performs its biological functions in various signaling networks.

3.2 Introduction

PDZ (PSD-95/Dlg1/ZO-1) domains are highly abundant protein–protein interaction domains involved in regulating signaling pathways (Kennedy, 1995; Kim and Sheng, 2004; Morais Cabral et al., 1996; Ponting, 1997; van Ham and Hendriks, 2003; Ye and Zhang, 2013). They play a critical role in many biological processes, such as managing cell polarity, regulating tissue growth and development, trafficking of membrane protein receptors and ion channels, and regulating cellular pathways (Brakeman et al., 1997; Harris and Lim, 2001; Romero et al., 2011). So far, 268 PDZ domains have been identified in 151 unique human proteins (Luck et al., 2012). Despite the broad function and relatively low sequence identity within PDZ domains, the secondary structure is highly conserved. The canonical PDZ domains contain six β-strands and two α-helices and have a single binding site in the hydrophobic groove between the αB helix and the βB strand (Doyle et al., 1996), as shown in Figure 3.1a. PDZ domains most commonly interact with the final three to five C-terminal residues of target proteins via the carboxylate binding loop that is defined by the conserved χ-φ-Gly-φ motif, where χ is any residue and φ is any hydrophobic residue (Pedersen et al., 2014). Various groups have revealed how these highly conserved protein–protein interactions propagate allosteric effects through the PDZ domain (Chen et al., 2007; De Los Rios et al., 2005; Dhulesia et al., 2008; Fuentes et al., 2004; Gianni et al., 2006; Grembecka et al., 2006; Kumawat and Chakrabarty, 2017; Lockless and Ranganathan, 1999; Lu et al., 2016; Miño-Galaz, 2015; Morra et al., 2014; Niu et al., 2007; Tochio et al., 2000; von Ossowski et al., 2006; Walma et al., 2002).

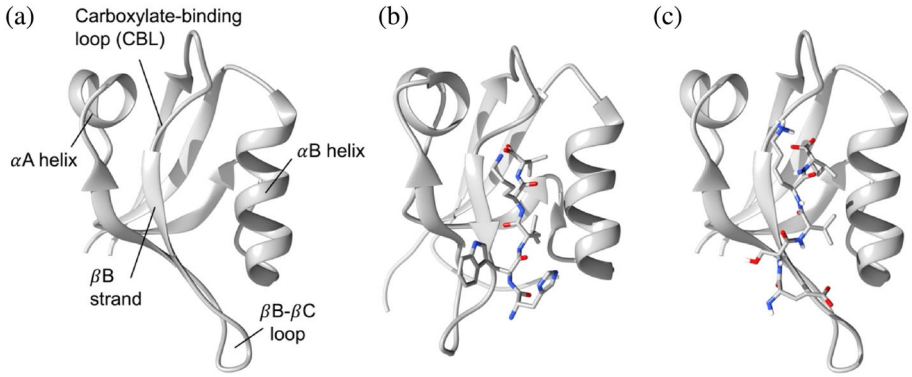


Figure 3.1: The PICK1 PDZ domain. (a) PICK1 PDZ domain with labeled secondary structures (PDB ID: 2PKU, ligand removed). (b) PICK1 PDZ-DAT complex (PDB ID: 2LUI). DAT ligand is the final five C-terminal residues of DAT (HWLKV). (c) PICK1 PDZ-GluR2 complex (PDB ID: 2PKU). GluR2 ligand is the final five C-terminal residues of AMPAR GluR2 (ESVKI). Notably, (b) and (c) are the starting structures of the all-atom MD simulations.

The PDZ domain is considered to be a model system to study allostery within small modular domains. Allostery in the PDZ family was initially brought to the table when Lockless and Ranganathan (Lockless and Ranganathan, 1999) proposed a method to statistically predict allosteric residue networks using multiple sequence alignment. This method is based on networks of energetically coupled residues that are responsible for the propagation of allostery throughout the PDZ domain. This original work sparked a wide interest in studying allostery within the PDZ family.

Many efforts have followed Lockless and Ranganathan's footsteps by applying various computational techniques, including direct coupling analysis (Gianni et al., 2011; Hultqvist et al., 2013), deep coupling scan (Olson et al., 2014), anisotropic thermal diffusion (Ho and Agard, 2010; Ota and Agard, 2005), rigid-residue scan (Kalescky et al., 2016), and interaction correlation via molecular dynamics simulations (Kong and Karplus, 2009; Lu

et al., 2016; Miño-Galaz, 2015), to reveal allosteric networks within the PDZ family. Furthermore, experimental groups have expanded our understanding of allostery in the PDZ family with applications of nuclear magnetic resonance (NMR) (Fuentes et al., 2006, 2004; Petit et al., 2009) and mutational analyses (Chi et al., 2008; Gianni et al., 2011; Hultqvist et al., 2013). Despite the abundance of domains in the PDZ family, these efforts have primary focused on a few well-studied PDZ domains, including Par-6 PDZ (Peterson et al., 2004; Thayer et al., 2017; Whitney et al., 2011) PSD-95 PDZ3 (Bozovic et al., 2020, 2020; Chi et al., 2008; Gerek and Ozkan, 2011; Gianni et al., 2006; Guclu et al., 2021; Kumawat and Chakrabarty, 2020, 2017; Morra et al., 2014; Petit et al., 2009), PTP-1 E PDZ2 (Cilia et al., 2012; Fuentes et al., 2006, 2004; Gerek and Ozkan, 2011; Lu et al., 2016; Morra et al., 2014), PTP-BL PDZ (Gianni et al., 2006; van den Berk et al., 2007). To the best of our knowledge, little attention has yet been given to explore allostery of the PDZ domain in Protein Interacting with C kinase-1 (PICK1).

PICK1 is a scaffolding protein involved in regulating the trafficking of various membrane proteins via endocytosis (Dev et al., 1999; Lu and Ziff, 2005; Rocca et al., 2008). PICK1 is an especially unique PDZ protein as it is the only protein in the human proteome that is comprised of both a PDZ domain and a BAR (Bin/amphiphysin/Rvs) domain (Hanley, 2008; Karlsen et al., 2015; Madsen et al., 2012). The PICK1 PDZ domain forms protein–protein interactions with a variety of integral membrane proteins, including the dopamine transporter (DAT) (Bjerggaard et al., 2004) and the GluR2 subunit of the AMPA receptor (Dev et al., 1999). Widely accepted hypotheses suspect that such PDZ-protein interactions lead to a propagation of signals through PICK1 that alters its

interdomain dynamics (Lu and Ziff, 2005; Rocca et al., 2008). This global transduction of signal through PICK1 could be explained by allostery at the PICK1 PDZ domain. The presence of allostery at the PICK1 PDZ domain would have major implications in our understanding of the biological function of PICK1.

The purpose of this study is to use all-atom molecular dynamics (MD) simulations to reveal how the atomic-level interaction pattern affects the interaction mechanisms and dynamics between the PICK1 PDZ domain and two representative ligands. These ligands include the final five C-terminal residues of two natural ligands: DAT and AMPAR GluR2. The two systems of interest are shown in Figure 3.1b,c. Here, we see that both ligands induce dynamic allostery at the αA helix of the PICK1 PDZ domain. Furthermore, our results suggest that different ligands may trigger different dynamic changes to the PICK1 PDZ domain. Lastly, our work identifies that the hydrophobic core that is formed between the ligands and residue I35 may be key to inducing such dynamic allostery.

3.3 Materials and Methods

3.3.1 Molecular Dynamics

We studied two PICK1 PDZ systems: PICK1 PDZ-DAT complex and PICK1 PDZ-GluR2 complex. The DAT ligand refers to the final five C-terminal residues (HWLKV) of the dopamine transporter (DAT), and the GluR2 ligand refers to the final five C-terminal residues (ESVKI) of the carboxyl tail peptide of the AMPA receptor GluR2 subunit. Experimentally determined crystal structures of the complex systems were used to generate the starting structure for all all-atom molecular dynamics simulations. (PDB ID: 2LUI (Erlendsson et al., 2014) and 2PKU (Pan et al., 2007), respectively). The PDB file of the

PICK1 PDZ-DAT complex (PDB ID: 2LUI) was manually edited by trimming terminal residues to ensure an identical sequence to the PICK1 PDZ- GluR2 system. Each starting structure is shown in Figure 3.1b,c. Each system was prepared using CHARMM-GUI (Jo et al., 2008; Lee et al., 2016). The most recently developed CHARMM36m (Huang et al., 2017) force field with explicit solvent (TIP3P) was used in each simulation with the Groningen Machine for Chemical Simulations (GROMACS) package (Abraham et al., 2015; Berendsen et al., 1995; Szilárd et al., 2015), version 2020.4. Counter ions (Na^+ or Cl) were added to neutralize the systems at 293 K. Steepest-descent minimization and 1-ns MD equilibrium simulations were carried out to generate equilibrated starting structures for the MD simulations. All bonds with hydrogen atoms were converted to constraints with the algorithm LINear Constraint Solver (LINCS). A Nose–Hoover temperature thermostat was used in each simulation. The time step was set as 2 fs, and snapshots were taken every 100 ps. Each system was built in a 90Å 90Å 90Å cubic water box. Each system (PICK1 PDZ-DAT and PICK1 PDZ-GluR2) had four replicates at 7 μs per trajectory, a total of 28 μs (4 x 7 μs) per system.

3.3.2 Defining the Bound State

The PICK1 PDZ-DAT and PICK1 PDZ-GluR2 complex systems had various dissociation events over the four trajectories (Figure 3.2). It is important to define a boundary that separates the bound states from the unbound states. Because the PICK1 PDZ-ligand complexes were very dynamic, we considered the distance distributions (Figures 3.3 and 3.4) of four key binding residue pairs that have been previously identified (Jo et al., 2008; Pan et al., 2007) between the PICK1 PDZ domain and the ligands. For the

PICK1 PDZ-DAT and PICK1 PDZ-GluR2 complexes, residue pairs I37-L2 and I37-V2, respectively, display the clearest distinction on average between the bound state and unbound states. With these state-defining residue pairs, frames were classified bound or unbound.

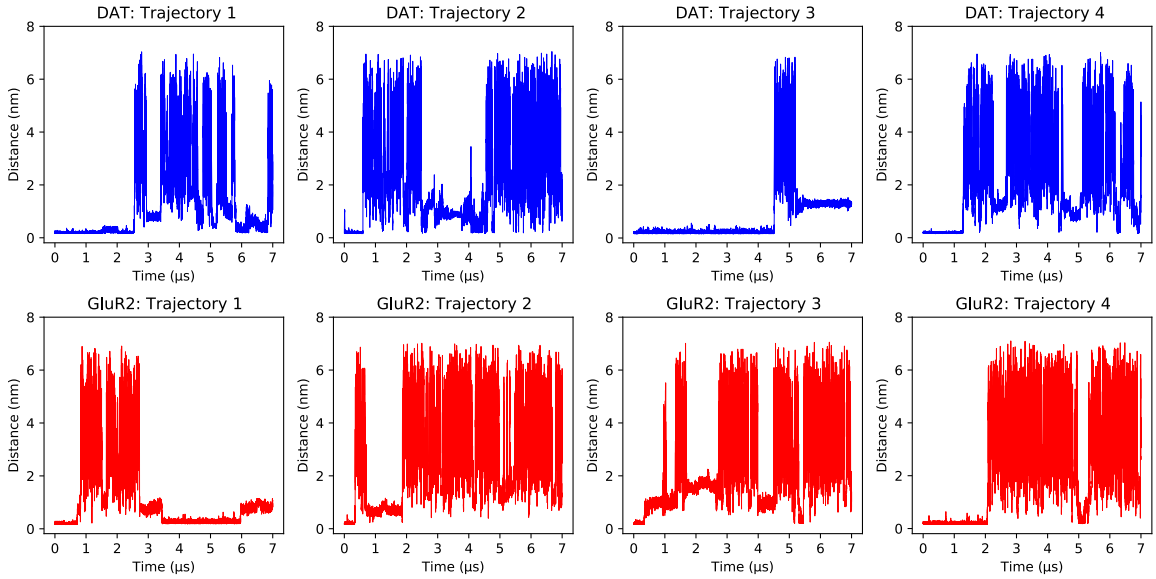


Figure 3.2: Distance between Ile37 of the PDZ domain and the P_{-2} position of the ligand during each trajectory. Spikes in the distance represent ligand dissociation.

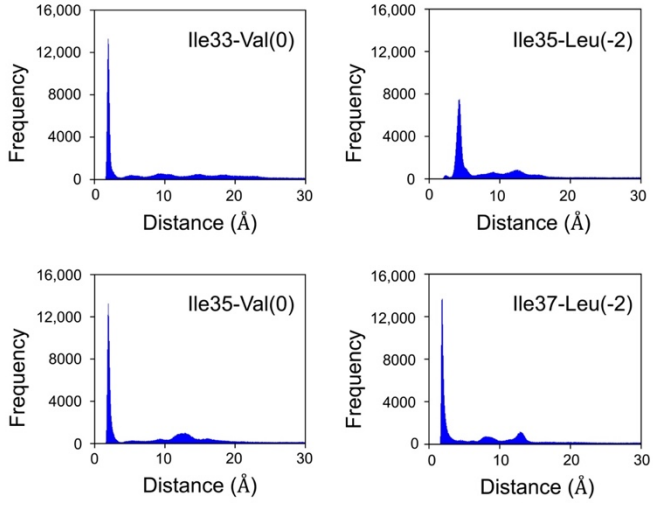


Figure 3.3: Distance distributions for the PDZ-DAT complex system. Distance is calculated using Ile37 of the PDZ domain and Leu_{-2} of the ligand.

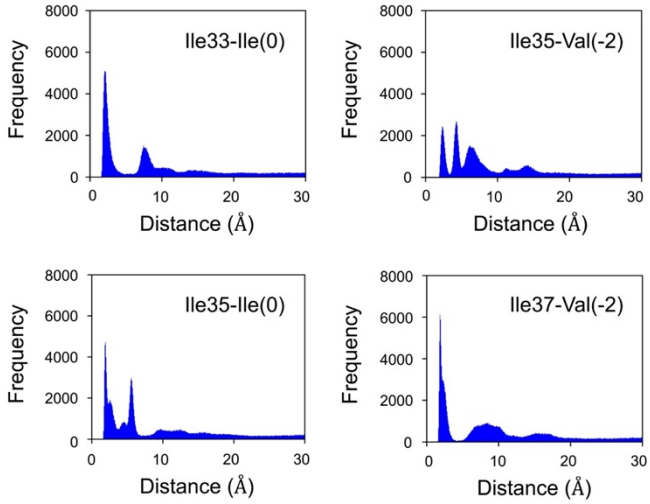


Figure 3.4: Distance distribution for the PDZ-GluR2 complex system. Distance is calculated using Ile37 of the PDZ domain and Val_{-2} of the ligand.

A bound state is defined as a distance less than 5.0 Å between any two atoms in I37 and L2 for the PICK1 PDZ-DAT complex, and a distance less than 5.0 Å between any two atoms on I37-V2 for the PICK1 PDZ-GluR2 complex. To test the accuracy of the defined cutoff, cluster analysis was performed over the bound state trajectories to reveal the most probable positions of DAT and GluR2 about the PIKC1 PDZ domain. In this way, we obtained the five most probable clusters of each ligand.

Figure 3.5 shows the PICK1 PDZ domain in gray while the most probable positions of the DAT (Figure 3.5a) and GluR2 (Figure 3.5b) are shown by unique colors. Our results confirm that the ligands reside in the PICK1 PDZ binding pocket in the defined bound state trajectories.

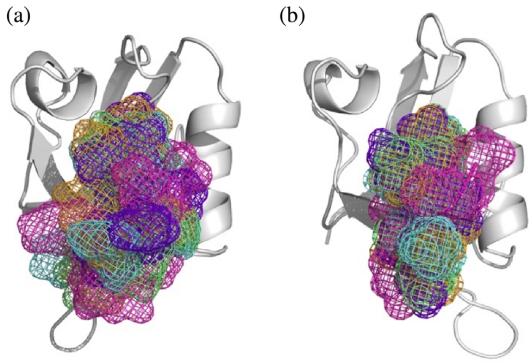


Figure 3.5: Cluster analysis reveals the most probable states of the (a) DAT and (b) GluR2 about the PICK1 PDZ domain after dividing the trajectories into bound states. The PICK1 PDZ domain is shown in gray and each cluster of the ligands is shown in a unique color. (a) Cluster 1 (orange) represents 62.7% of the frames, Cluster 2 (purple) represents 20.4% of the frames, Cluster 3 (pink) represents 8.1% of the frames, and Cluster 4 (green) represents 7.8% of the frames. Cluster 5 was excluded because if represents less than 1% of the frames. (b) Cluster 1 (orange) represents 37.1% of the frames, Cluster 2 (purple) represents 22.0% of the frames, Cluster 3 (pink) represents 20.8% of the frames, Cluster 4 (green) represents 10.8% of the frames, and Cluster 5 (blue) represents 9.3% of the frames.

3.3.3 Dynamic Flexibility Index (DFI)

The DFI metric estimates the resilience of residues within a given protein system. Being a residue specific metric, DFI calculates relative flexibility scores (Larrimore et al., 2017). By incorporating Linear Response Theory (LRT) and Perturbation Response Scanning (PRS) (Atilgan and Atilgan, 2009), DFI calculates the response of a residue due to a perturbation on another residue normalized by the average response of all residues in the protein (Bozovic et al., 2020). Position specific dynamics profiles are calculated by utilizing residue covariances.

$$[\Delta \mathbf{R}]_{3N\times 1} = [\boldsymbol{H}]^{-1}_{3N\times 3N}[\mathbf{F}]_{3N\times 1} \tag{3.1}$$

$$DFI_i = \frac{\sum_{j=1}^{N}|\Delta R^j|_i}{\sum_{i=1}^{N}\sum_{j=1}^{N}|\Delta R^j|_i} \tag{3.2}$$

The Hessian matrix, $\boldsymbol{H}$, contains the second derivative of potentials. Residue covariances are calculated by taking the inverse of the Hessian matrix, $\boldsymbol{H^{-1}}$. The Elastic Network Model (ENM) is commonly used to produce the Hessian matrix. However, to include explicit solvent and better estimate residue interactions, residue covariances can be gathered from an MD simulation production trajectory. In this study, we utilized the MD simulations to calculate residue covariances. $\Delta\boldsymbol{R}$ is a response vector calculated by multiplying the covariance matrix with the force vector, $\boldsymbol{F}$ and contains the residue responses. The collection of DFI values calculated from this approach is further refined with a percentile ranking to normalize the scores. A residue with a DFI score less than 0.2 is considered a rigid location, while a position with a DFI score higher than 0.8 is considered a flexible residue. Rigid residues have been found to be important in protein stability and function (Modi and Ozkan, 2018).

3.3.4 Dynamic Coupling Index (DCI)

Utilizing the same elemental principles as described above, the DCI metric captures the dynamic allosteric coupling of pair of residues in a protein. DCI calculates the response of a residue due to a Brownian force applied to another residue in the same system normalized by the average response of the same residue due to perturbations on the rest of the proteins. The magnitude of the response represents the strength of the dynamic allosteric coupling of a site to another residue being perturbed.

$$DCI_i = \frac{\sum_j^{N_{Functional}} |\Delta R^j|_i / N_{Functional}}{\sum_{j=1}^{N} |\Delta R^j|_i / N} \tag{3.3}$$

A DCI score applied on binding site residues can reveal other residues in the protein that are highly coupled, meaning a binding event or the dynamics of the residue upon

binding will be highly affected. Notably, the DCI score is not an indicator of binding dynamics but rather how the binding dynamics are coupled to the rest of the protein. DCI metric can uncover long range allosteric communications related to the binding event (Atilgan and Atilgan, 2009; Campitelli et al., 2020; Modi et al., 2021b). Residues with a high DCI score indicate strong coupling with binding site and a position with a low DCI score is considered weakly coupled to the binding site.

3.3.5 Network analysis

Network analysis calculates the correlated movements between residues within a protein or protein complex by constructing residue-based and community-based weighted network graphs according to a trajectory. During the calculations, each residue is represented by a node in a network and the links between nodes are the cross-correlation values between these nodes. By using the algorithm developed by McCammon and Harvey (McCammon and Harvey, 1988), the displacement of the Ca atoms are used to assess the magnitude of all pairwise cross-correlation coefficients. If the correlation value is 1, the fluctuations of two Ca atoms are completely correlated. If the correlation value is 1, the fluctuations of two Ca atoms are completely anticorrelated (same period and opposite phase). Lastly, if the correlation value is 0, the fluctuations of two Ca atoms are not correlated. The analysis uses the calculated cross-correlation coefficients to return a community partition with the highest overall modularity value based on Girvan-Newman style clustering (Newman and Girvan, 2004). All the above analysis was carried out using the bio3d package (Grant et al., 2021, 2006; Skjærven et al., 2014).

3.3.6 Local frustration evaluations

To quantify the degree of local frustration associated with the binding of different ligands to the PICK1 PDZ domain, the Frustratometer server (http://frustratometer.qb.fcen.uba.ar/) (Ferreiro et al., 2007; Parra et al., 2016; Rausch et al., 2021) was used to evaluate the two PDZ-ligand complexes investigated here. Default parameters were used when carrying out the assessments of local frustration, for example, a 5 Å radius cutoff value was applied. The PDB structures used in local frustration analysis contained only the PDZ domain, and the ligands have been removed.

3.4 Results

Each trajectory experienced ligand dissociation events (Figure 3.2). These dissociation events present a unique opportunity to explore the switching of dynamic states at the αA helix in real-time. First, we reveal the unique and specific ligand-protein interactions related to the dissociation events by performing hydrogen bond analysis across the two complex systems. Hydrogen bond analysis reveals canonical Class II PDZ-ligand interactions with the carboxylate-binding loop in each system (Figure 3.6). These results are in good agreement with previous experimental work (Christensen et al., 2019).

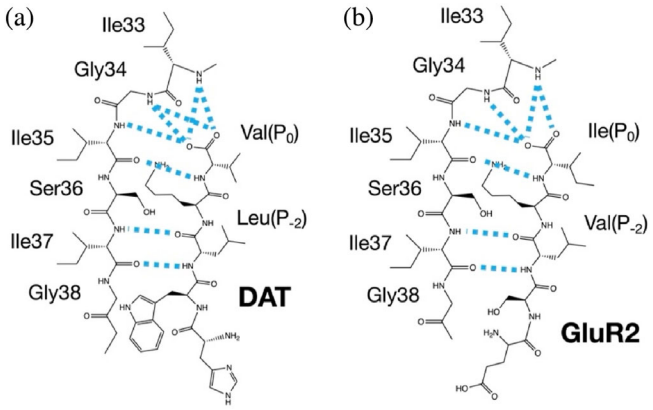


Figure 3.6: Hydrogen bonding network at the binding pocket of the (a) PICK1 PDZ-DAT complex and (b) PICK1 PDZ-GluR2 complex. PDZ-DAT and PDZ-GluR2 display a similar pattern of hydrogen bonding.

Additionally, we performed a statistical analysis to rank the probability of each hydrogen bond forming in the binding pocket (Figure 3.7 and 3.8). The PICK1 PDZ- DAT system has three hydrogen bonds that occur in at least 90% of the bound frames, including I37(N)-L_2(O), L_2(N)-I37(O) and V_0(N)-I35(O) (Figure 3.7). The PICK1 PDZ-GluR2 system has three hydrogen bonds that occur in at least 70% of frames with the ligand bound, including I_0(N)-I35(O), I37(N)-V_2(O) and V_2(N)-I37 (O) (Figure 3.8). These three most probable pairs in each system are in agreement with each other. In both systems, the most probable hydrogen bonds occur between (1) I37 and the residue at position P_2 of the ligand and (2) I35 and the residue at position P_0 of the ligand. These interactions are much more prevalent than interactions between G34-P_0 and I33-P_0.

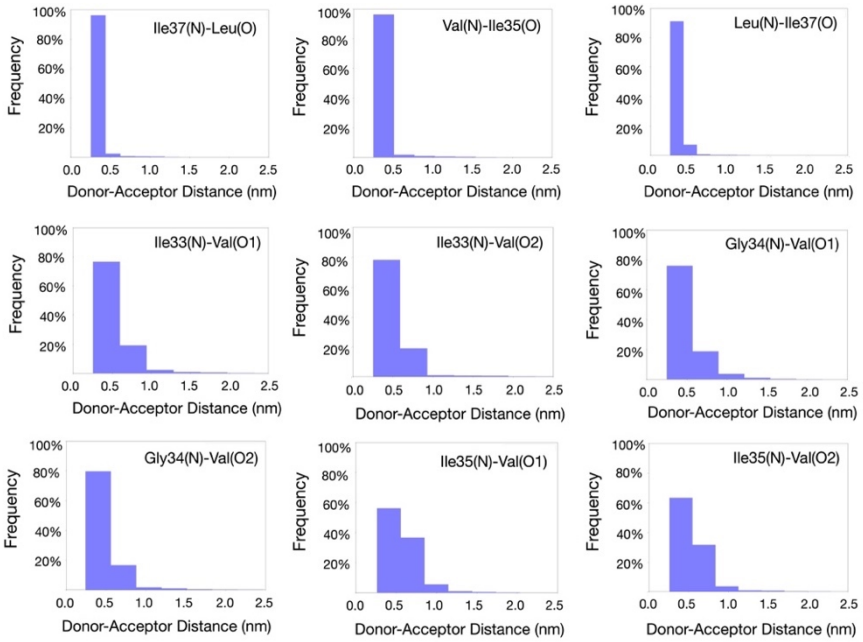


Figure 3.7: Probability of each hydrogen bonding pair within the PICK1 PDZ-DAT complex.

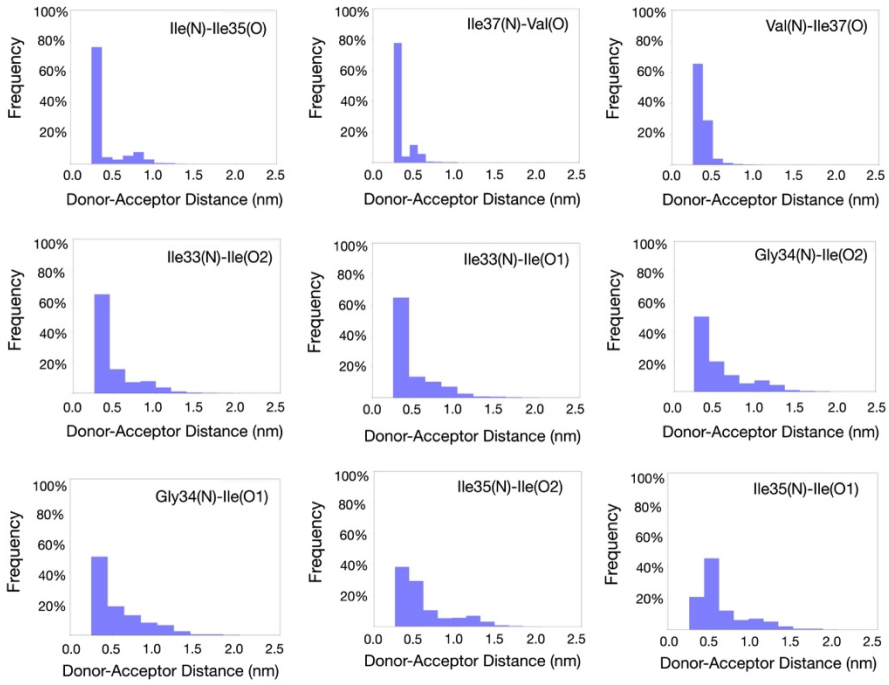


Figure 3.8: Probability of each hydrogen bonding pair within the PICK1 PDZ-GluR2 complex.

While the above analysis reveals the most probable hydrogen bonds within each complex, it is unclear if these interactions are simply essential to the stability of complex formation or, ultimately, if they effect the overall dynamics and subsequent dynamic allostery of the system. To connect the changes in protein-ligand hydrogen bonding interactions (particularly, as related to ligand dissociation) to protein dynamics, we explored the correlation between ligand dissociation and the dynamics of PICK1 PDZ domain by calculating the coupling of various residue-residue distance pairs over the first 3 ms of each trajectory. Five pairs were considered in the coupling calculation: I33-P_0, G34-P_1, I35-P_2, S36-P_3 and I37-P_4. The five PICK1 PDZ residues were chosen because they comprise the βB strand which has been identified as a key player in ligand binding by previous work (Doyle et al., 1996).

These pairs were selected to represent the overall interactions between PICK1 PDZ domain and ligand. Figure 3.9 lists the 20 residue-residue pairs for each system that are most strongly correlated with the distance changes between the five selected pairs. Having the relative highest rank in both systems, we consider the distance between I33 (βB strand) and A58 (αA helix) as directly dependent on the atomic-level interactions between the PICK1 PDZ domain and ligand. Interestingly, the I33-A58 distance can also be used to describe the overall distance between the βB strand and the αA helix. We explore the correlation between the PDZ-ligand interactions and the distance between the βB strand and the αA helix below:

PDZ-DAT

Defining Pair	Correlated Pair	Ranking
I33-P_0	I33-G62*	1.53
I33-P_0	I33-D61	1.52
G34-P_{-1}	I33-A58*	1.31
G34-P_{-1}	I33-T56	1.20
G34-P_{-1}	V23-T56	1.17
I33-P_0	I33-A59	1.10
I33-P_0	I33-A58*	1.08
G34-P_{-1}	I33-A59	1.07
I33-P_0	I33-T63*	1.06
I33-P_0	I33-T56	1.04
G34-P_{-1}	I33-T63	1.03
G34-P_{-1}	I33-L60	1.01
I35-P_{-2}	I33-A59	1.01
S36-P_{-3}	I33-A59	1.01
I33-P_0	L25-G62*	1.00
G34-P_{-1}	I33-G62	0.99
S36-P_{-3}	I33-A58*	0.99
I33-P_0	V23-F53	0.98
G34-P_{-1}	I33-D61	0.98
I35-P_{-2}	I33-A58	0.98

PDZ-GluR2

Defining Pair	Correlated Pair	Ranking
I33-P_0	L25-G62	2.51
I33-P_0	I33-A58*	2.50
I33-P_0	V23-T56	2.46
I33-P_0	V23-P57	2.45
I33-P_0	I33-G62*	2.44
I37-P_{-4}	L25-G62	2.44
I33-P_0	Q26-G62	2.43
I33-P_0	I33-T63*	2.43
I33-P_0	V23-N55	2.43
S36-P_{-3}	I33-A58*	2.43
S36-P_{-3}	L25-G62	2.43
I33-P_0	K27-G62	2.40
I33-P_0	L25-T63	2.40
G34-P_{-1}	I33-A58*	2.40
S36-P_{-3}	V23-P57	2.39
S36-P_{-3}	V23-T57	2.39
G34-P_{-1}	L25-G62	2.38
S36-P_{-3}	I33-A58	2.38
I37-P_{-4}	V23-P57	2.38
I33-P_0	T24-T56	2.37

Figure 3.9: Tabulated ranking of correlated distance pairs. Pairs that occur in both systems are marked (*). Ile33-Ala58 was selected as the representative pair.

Figure 3.10 describes representative dissociation events for the PICK1 PDZ-DAT (Figure 3.10a,b) and PICK1 PDZ- GluR2 systems (Figure 3.10c,d). First, we will consider PICK1 PDZ-DAT system, where the dissociation of the DAT is weakly correlated with the dynamics of the aA helix (Figure 3.10a,b). The distance between I37 of the PICK1 PDZ domain and L_2 of DAT was used to trace the dissociation as defined in the Methods section. At $2.5 ms, the distance between I37 and L_2 spikes as the ligand dissociates from the binding pocket (Figure 3.10a, black). This dissociation is confirmed by hydrogen bond and surface area analysis.

As DAT dissociates, the number of hydrogen bonds and the surface area between the PICK1 PDZ domain and DAT drops to zero (Figure 3.10a, blue and purple, respectively).

The surface area between the PICK1 PDZ domain and DAT was calculated using solvent-accessible surface area. While the dissociation event does not clearly correlate with the RMSD of the αA helix (Figure 3.10a, red), it does result in a distinct increase in distance between αA helix and the βB strand (Figure 3.10a, green).

Next, we will consider the representative dissociation event for the PICK1 PDZ-GluR2 system (Figure 3.10c,d). As shown in Figure 3.10c, the dissociation of the GluR2 is directly correlated with the dynamics of the αA helix. The dissociation of GluR2 at 2.0 μs is confirmed by a sharp distance increase between I37 of the PICK1 PDZ domain and V_2 of GluR2 (Figure 3.10c, black), a loss of hydrogen bonds between the PICK1 PDZ domain and GluR2 (Figure 3.10c, blue), and a loss of surface area contact between the PICK1 PDZ domain and GluR2 (Figure 3.10c, purple). Interestingly, the disruption of PICK1 PDZ-GluR2 interactions is correlated with dynamic changes at the αA helix. Figure 3.10c (red) shows that the RMSD of the αA helix increases with the dissociation of GluR2. Moreover, our analysis reveals a correlation between PICK1 PDZ-GluR2 interactions and the distance between the βB strand and the αA helix (Figure 3.10c, green). This distance separation may play a role in the destabilization of the αA helix.

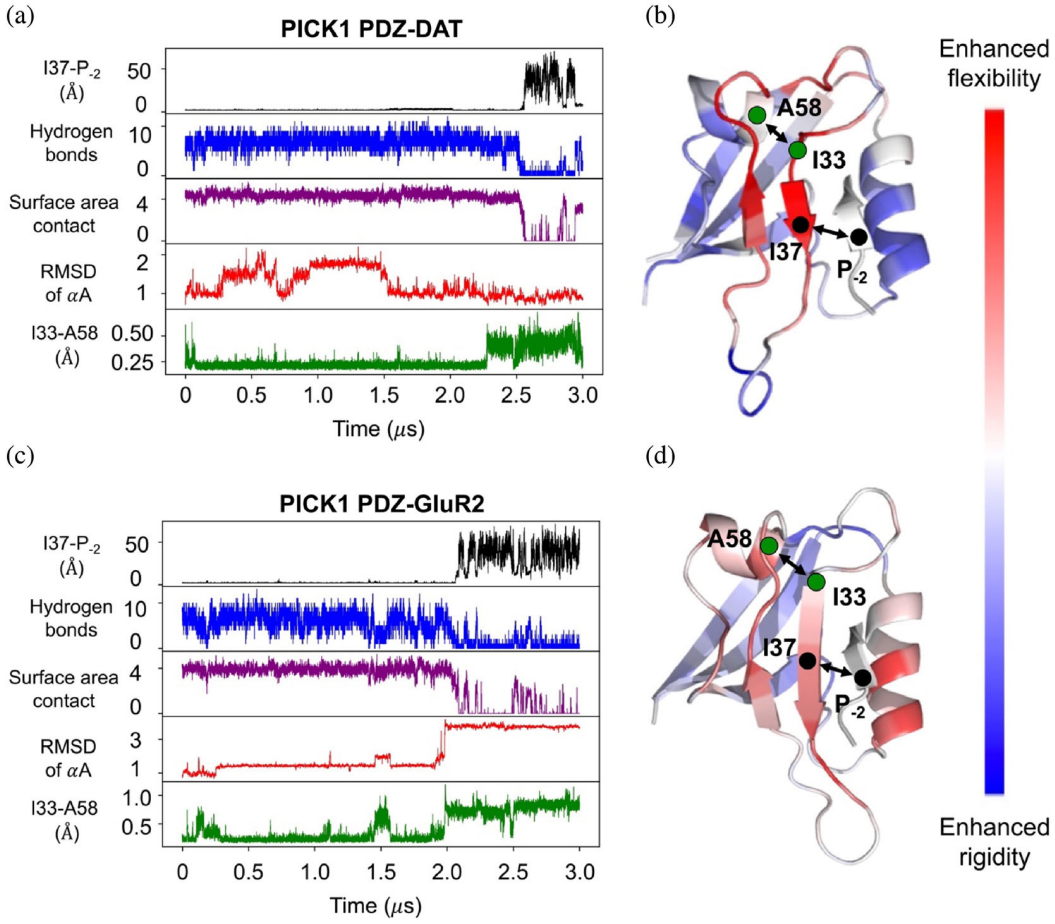


Figure 3.10: Correlation between ligand dissociation and the dynamics of the PICK1 PDZ domain. (a) Representative PDZ-DAT trajectory. At $2.5 μs, the distance between I37 of PICK1 PDZ and L_2 of DAT increases (black, A), the number of hydrogen bonds between the PICK1 PDZ domain and DAT decreases (blue), the surface area contact between the PICK1 PDZ domain and DAT decreases (purple), the RMSD of the αA helix does not appear to correlate with ligand dissociation (red), and the residue-residue distance between I33 of the βB stand and A58 of the αA helix increases (green). (b) ΔDFI between the bound and unbound states of the PICK1 PDZ-DAT trajectory 1. ΔDFI of PDZ-DAT indicates little change in the flexibility of the αA helix upon ligand dissociation. (c) Representative PDZ-GluR2 trajectory. At $2 μs, the distance between I37 of PICK1 PDZ and V_2 of GluR2 increases (black), the number of hydrogen bonds between the PICK1 PDZ domain and GluR2 decreases (blue), the surface area contact between the PICK1 PDZ domain and GluR2 decreases (purple), the RMSD of the αA helix increases (red), and the residue-residue distance between I33 of the βB-stand and A58 of the αA helix increases (green).(d) ΔDFI between the bound and unbound states of the PICK1 PDZ-GluR2 trajectory 4. ΔDFI shows an enhanced flexibility of the αA helix upon GluR2 dissociation.

Finally, we calculated the change in the dynamics flexibility index (ΔDFI) across the bound and unbound states of each system (Figure 3.10b,d). ΔDFI reveals significant changes in dynamics of the PICK1 PDZ domain due to the dissociation of ligands. The important ligand binding regions, including the αB helix and βB strand, show enhanced flexibility upon ligand dissociation. When the interactions are disrupted, the key binding residues gain more conformational freedom, and the flexibility enhances. Thus, enhanced flexibility at the binding site is a direct indicator of a dissociation. More interestingly, ΔDFI also reveals unique changes to the αA helix upon dissociation of each unique ligand. As represented by the RMSD of the αA helix (Figure 3.10a, red), the dissociation of DAT does not enhance the flexibility of the αA helix (Figure 3.10b). Instead, the majority of the αA helix has little change in terms of flexibility while A59 shows enhanced rigidity (Figure 3.10b). Oppositely, there are significant changes in dynamics of the αA helix due to the dissociation of GluR2 (Figure 3.10d). Echoing the RMSD of the αA helix (Figure 3.10c, red) and the distance between I33 and A58 (Figure 3.10c, green), DFI analysis shows enhanced flexibility at the αA helix upon ligand dissociation (Figure 3.10d). As the I33-A58 distance increases, the interactions between the αA helix and the carboxylate-binding loop become weaker to allow more fluctuations. Advancing to a dynamically more flexible regime, the αA helix is observed be to allosterically being altered by the dissociation event.

To further explore the correlation between ligand binding and the dynamics at the αA helix, we performed protein network analysis. Protein network analysis can reveal the coupling of major movements by creating protein structure networks based on the primary motions of each residue. The analysis reveals the residues within the PICK1 PDZ domain

that are most strongly coupled to the ligands' motion. The motions of DAT (Figure 3.11a) and GluR2 (Figure 3.11b) are both coupled to the motion of the distal αA helix and the βB-βC loop of the PICK1 PDZ domain. Interestingly, the motions of DAT are more strongly coupled to the βB and βC strands than are the motions of GluR2.

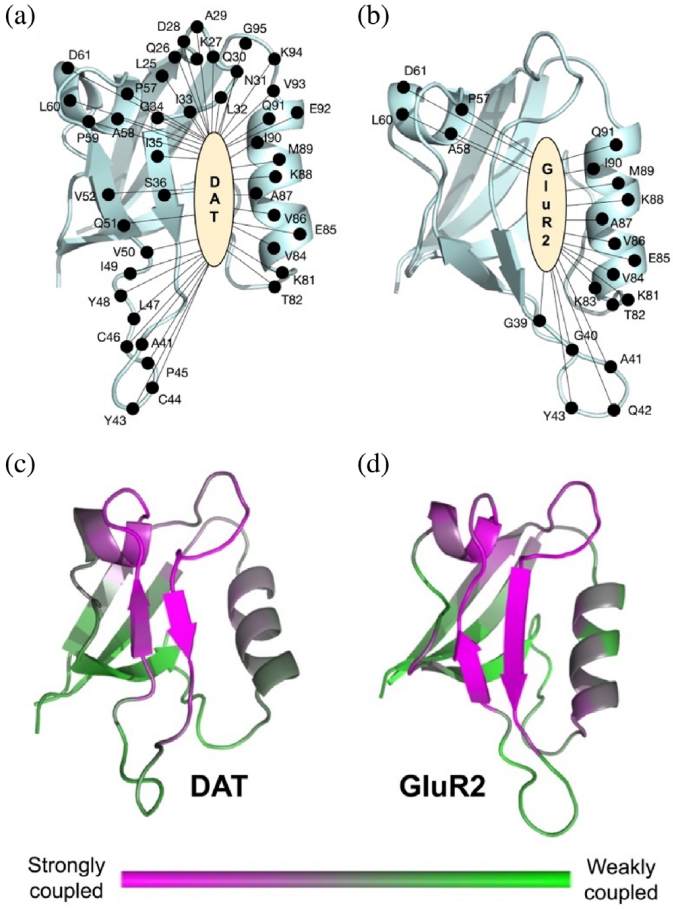


Figure 3.11: Allosteric dynamic coupling within the PICK1 PDZ-ligand systems. (a) Protein structure network analysis of the PICK1 PDZ-DAT system. (b) Protein structure network analysis of the PICK1 PDZ-GluR2 system. The motions of DAT and GluR2 are both coupled with the distal αA helix. (c) DCI analysis of the PICK1 PDZ-DAT system. (d) DCI analysis of the PICK1 PDZ-GluR2 system. In both systems, the binding residues of the PICK1 PDZ domain57 are coupled with the αA helix.

Dynamic coupling index (DCI) was applied to each system to explore the coupling of dynamics between binding site residues and the global protein. The DCI metric has previously been shown to capture allosteric coupling of distal site to critically important residues in a protein. Upon a binding event, the binding site residues experience exerted forces from the ligand so that the dynamics of the system may be affected. Notably, the force exerted by the ligand not only affects the dynamics of the binding site residues but may also affect the dynamics of the global protein due to allosteric communication. The DCI metric measures the coupling strength of a residue to a binding site. A highly coupled residue will experience the repercussions of binding more than weakly coupled residues. As shown in Figure 3.11c,d, DCI analysis on the PICK1 PDZ-DAT and PICK1 PDZ-GluR2 systems reveals a coupling trend that echoes results from network analysis at the αA helix. Both DAT and GluR2 binding residues observes strong coupling to the αA helix.

Time-resolved force distribution analysis (TRFDA) (Costescu and Gräter, 2013) was performed to reveal the punctual stress on each PICK1 PDZ residue as a result of ligand binding as in previous work (Li et al., 2022). TRFDA was performed over each trajectory, and the per trajectory results were summed over each complex system. The summed results are shown in Figure 3.12. The 10 PICK1 PDZ residues that experienced the greatest punctual stress for each system are listed in Figure 3.13.

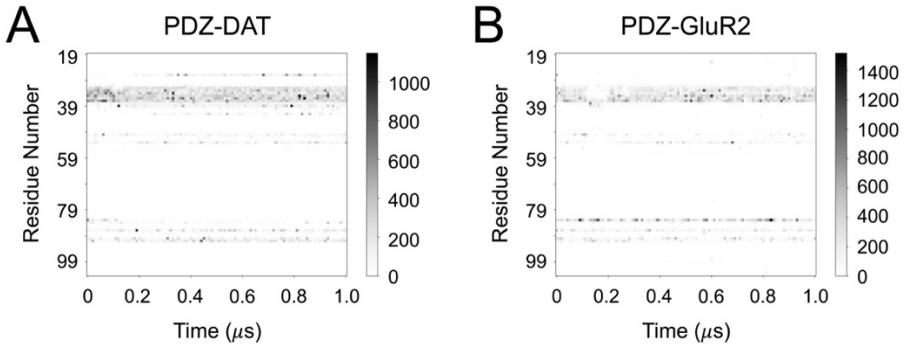


Figure 3.12: Summed TRFDA for each complex system: (A) PICK1 PDZ-DAT and (B) PICK1 PDZ-GluR2.

PICK1 PDZ-DAT		PICK1 PDZ-GluR2	
Ile37	βB strand	Ile37	βB strand
Ser36	βB strand	Lys83	αB helix
Ile35	βB strand	Ser36	βB strand
Ile33	βB strand	Ile33	βB strand
Gly34	βB strand	Ile35	βB strand
Leu32	βB strand	Leu32	βB strand
Lys27	βA-βB loop	Asn31	βB strand
Ile90	αB helix	Gln91	αB helix
Gln91	αB helix	Ala87	αB helix
Ala87	αB helix	Ile90	αB helix

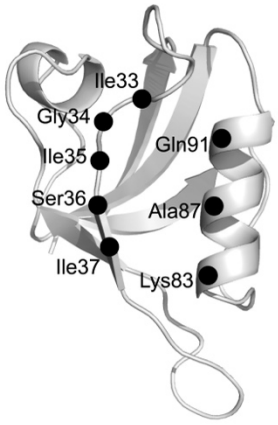


Figure 3.13: Time-resolved force distribution analysis (TRFDA) reveals the top ten PDZ residues with the greatest punctual stress in each complex system. DAT and GluR2 induce the greatest punctual stress on the βB strand.

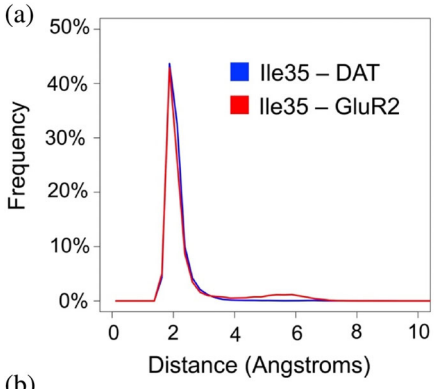
Both DAT and GluR2 induce the greatest punctual stress on the βB strand and αB helix, regions that directly interact with the ligands. In the PICK1 PDZ- DAT system, all six residues that experience the greatest punctual stress comprise the βB strand. Oppositely, GluR2 induces significant punctual stress on K83 of the αB helix. These results point to the different interaction patterns induced by different ligands binding.

Our analysis reveals that DAT and GluR2 can induce unique stresses on the PICK1 PDZ domain, but the specific residues and mechanisms through which dynamic allostery is propagated in the PICK1 PDZ domain remains in question. A recent review of allostery in the PDZ family (Stevens and He, 2022a) notes that A46 (αA helix) of PTP-BL PDZ2 and A347 (αA helix) of PSD-95 PDZ3 have been consistently identified as allosteric residues in a wide array of computational and experimental efforts (Cilia et al., 2012; Dhulesia et al., 2008; Du et al., 2010; Fuentes et al., 2006; Gerek and Ozkan, 2011; Kalescky et al., 2015; Kong and Karplus, 2009; Lee and Zheng, 2010; Li et al., 2022; McLaughlin et al., 2012; Ota and Agard, 2005; Walma et al., 2002). Furthermore, in a recent work exploring the interactions and dynamics between the PICK1 PDZ domain and the small molecule inhibitor BIO124, we propose that a structural alignment of PICK1 PDZ, PTP-BL PDZ2, and PSD-95 PDZ3 suggests that this allosteric alanine residue on the αA helix is evolutionarily conserved across all three PDZ domains (Stevens et al., 2022b). This structural alignment also suggests that the interactions between BIO124 and I35 of the PICK1 PDZ domain may have a role in the propagation of signal to A58 of the αA helix (Stevens et al., 2022a). Notably, A58 forms a van der Waals surface with I35, which is directly involved in ligand binding. Here, our results support the importance of A58 as an allosteric residue in the PICK1 PDZ domain. Distance analysis reveals that I33-A58 distance is coupled with ligand binding, protein network analysis identifies A58 in the network of residues dynamically coupled to the ligand, and DCI analysis indicates A58 is strongly coupled to binding site residues. We suspect that interactions between natural

ligands and I35 of the PICK1 PDZ domain may also have a role in the propagation of signal to the αA helix.

We explore the role of I35 in propagating allosteric signal to the αA helix of the PICK1 PDZ domain. Distance distribution and TRFDA are used to identify the degree of interactions between the ligands and I35. As shown in Figure 3.14a, distance distribution analysis was performed between the ligand and I35 for each system.

(a)



(b)

Punctual Stress (*kJ/mol/nm*)

	Ligand → Ile35
DAT	295.03
GluR2	312.28

Figure 3.14: The role of I35 in propagating allosteric signal. (a) Distance distribution between I35 of the PICK1 PDZ domain and the ligands. (b) Punctual stress on I35 of the PICK1 PDZ domain induced by the ligands

Here, the distance is defined as the shortest distance between any two atoms in the ligand and I35. DAT (blue) and GluR2 (red) both form the close contact ($2 Å) with I35. In addition to exploring the distance distribution between ligands and I35, we also calculated the punctual stress on I35 induced by the ligand by using TRFDA. As shown in Figure 3.13, I35 is one of the top five residues that experiences the greatest punctual stress in each system. Figure 3.14b lists the punctual stress on I35 induced by DAT and GluR2.

GluR2 induces a slightly greater punctual stress on I35 than DAT does. As demonstrated by Figure 3.10, GluR2 is more strongly coupled to the αA helix than DAT is. This stronger coupling between GluR2 and the αA helix may be a result of the strong punctual stress at I35. Together, distance distribution analysis and TRFDA point to the importance of interactions between the ligand and I35 in inducing dynamic allostery at the αA helix of the PICK1 PDZ domain.

As discussed in previous work, the dynamic allostery can be closely related to the local conformational changes resulting from local frustrations. To explore the local frustration regions in PICK1 PDZ domains, the Frustratometer server was used.

It can be seen from Figure 3.15a that the αA helix is indeed a local high frustration region. Moreover, there are other local frustration regions, for example, αB and βB-βC loop, which contain highly frustrated interactions. Interestingly, both of these two regions were identified in our network analysis (Figure 3.11), showing their correlations with the ligands. The tight green lines at the center highlight that the major structural “core” is conserved. The frustration projection on each residue is shown in Figure 3.15b. The ligands are part of the core and, at the same time, trigger frustration on the protein surface.

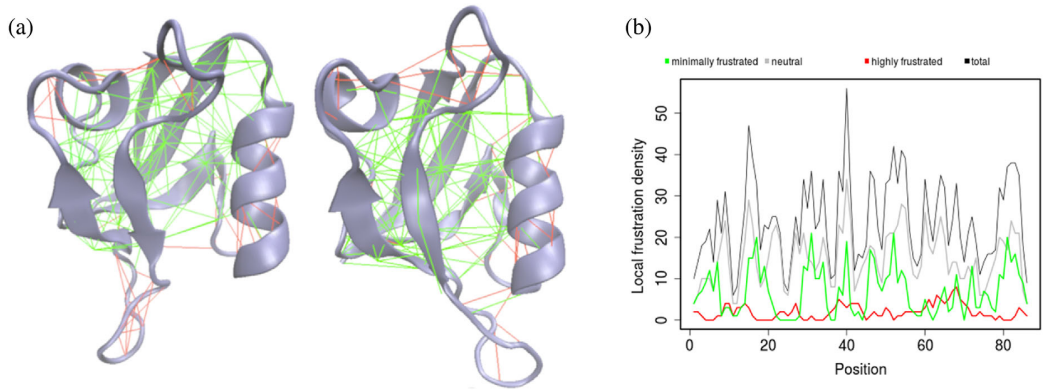


Figure 3.15: Local frustration in allosteric PICK1-PDZ domains. (a) The frustratograms for the individual conformations, with the minimally frustrated interactions in green lines, and the highly frustrated interactions in red lines. Left: PDB ID 2LUI, right: PDB ID 2PKU (b) Quantification of the local frustration projected on each residue of the PICK1-PDZ domain with minimally frustrated interactions (green) or highly frustrated interactions (red)

3.5. Discussion

The purpose of this work is to investigate the dynamic allostery in the PICK1 PDZ domain that can be induced by unique binding partners. We found that (1) the PICK1 PDZ domain exhibits dynamic allostery at the αA helix, (2) the unique interaction patterns between different binding partners and the PICK1 PDZ may induce unique dynamic changes to the PICK1 PDZ domain, and (3) the hydrophobic core that is formed between the ligands and I35 may be key to inducing dynamic allostery at the αA helix.

Our results demonstrate that natural ligands DAT and GluR2 can induce dynamic allostery at the αA helix of the PICK1 PDZ domain. Protein structure network, DCI, TRFDA, and local frustration analysis show that both DAT and GluR2 are dynamically correlated with the αA helix. This dynamic correlation distant from the binding pocket points to the ability of DAT and GluR2 to induce dynamic allostery across the PICK1 PDZ domain. These results are in agreement with previous work which has identified the αA helix as an allosteric region within other PDZ domains, including Par-6 PDZ, PTP-1 E

PDZ2, PTP-BL PDZ1, and AF-6 PDZ (Dev et al., 1999; Gianni et al., 2006; Lu et al., 2016; Miño-Galaz, 2015; Morra et al., 2014; van den Berk et al., 2007; Whitney et al., 2011). Furthermore, dissociation events captured during our simulations presented a unique opportunity to explore dynamic changes to the PICK1 PDZ domain in real time. GluR2 dissociation is directly coupled with increased fluctuations at the αA helix and increased distance between the αA helix and the βB strand. The distant shift of the αA helix and the βB strand agrees with secondary structure shifts seen in previously studied PDZ domains (Kumawat and Chakrabarty, 2017; Miño-Galaz, 2015). Notably, the dissociation of the PICK1 PDZ-DAT complex was not so clearly correlated to dynamic changes at the αA helix. These results suggest that different binding partners may induce different dynamic changes to the PICK1 PDZ domain.

Previous work on the PTP-BL PDZ2 domain (Fuentes et al., 2006; Gianni et al., 2006) and the PSD-95 PDZ3 domain (Lockless and Ranganathan, 1999) has pointed to the importance of structural equivalents of I35 in propagating allosteric signal to the αA helix. Our work suggests that I35 may also be a key residue in propagating signals in the PICK1 PDZ domain. Our results demonstrate that both DAT and GluR2 are dynamically coupled with the αA helix. Distance distribution analysis and TRFDA reveal that DAT and GluR2 form the close contact with and induce the strong punctual stress on I35. These results suggest that interactions between the ligand and I35 are key to inducing dynamic allostery at the αA helix in the PICK1 PDZ domain. The release of the AlphaFold 2 provides a high-resolution solution (Binder et al., 2022; Stevens and He, 2022b) to compare PDZ domains across multiple species and different proteins.

Our results identify dynamic allostery within the PICK1 PDZ domain. By comparing the responses of the PICK1 PDZ domain to the binding of different ligands, we see that the binding of different types of ligands may induce different dynamic changes to PICK1 PDZ domain. Our previous work on the PICK1 protein identified the αA helix of the PDZ domain as a key participant in inter- domain PDZ-BAR and PDZ-linker interactions (Kim and Sheng, 2004). We suspect that the ligand-induced dynamic changes at the αA helix may affect interdomain interactions and ultimately explain the long hypothesized conformational change of PICK1 upon ligand binding (Karlsen et al., 2015; Rocca et al., 2008). An atomic-level resolution of the mechanism behind the PICK1 interdomain dynamics may greatly affect how we understand the PICK1 protein.

3.6 Acknowledgement

This research was funded by the National Science Foundation Graduate Research Fellowship Program (Grant No. DGE-1939267), the National Science Foundation (Grant No. 2137558), the National Science Foundation (Grant No. 1901709), the Leverhulme Trust (RPG- 2017-222), and the Gordon and Betty Moore Foundation (Award: 1715591). This work was also supported by the Substance Use Disorders Grand Challenge Pilot Research Award, the Research Allocations Committee (RAC) Award, and the University of New Mexico Office of the Vice President for Research WeR1 Faculty Success Program. We also acknowledge the Centre of Informatics - Tricity Academic Supercomputer & network (CI TASK) in Gdansk, Poland, for the availability of high-performance computing resources.

CHAPTER 4

PLANT-EXPRESSED COCAINE HYDROLASE VARIANTS OF BUTYRYLCHOLINESTERASE EXHIBIT ALTERED ALLOSTERIC EFECTS OF CHOLINESTERASE ACTIVITY AND INCREASED INHIBITOR SENSITIVITY

This chapter is adapted from "Larrimore, K.E., Kazan, I.C., Kannan, L., Kendle, R.P., Jamal, T., Barcus, M., Bolia, A.,Brimijoin, S., Zhan, C-G., Ozkan, S.B., Mor, T.S. (2017) Plant-Expressed Cocaine Hydrolase Variants Of Butyrylcholinesterase Exhibit Altered Allosteric Efects Of Cholinesterase Activity And Increased Inhibitor Sensitivity. Scientific Reports 7(1), DOI:10.1038/s41598-017-10571-z"

I Can Kazan performed the analysis for Figs 5 and 6 and S2 presented in this work and wrote the corresponding text. Tsafrir S. Mor and Katherine E. Larrimore designed experiments and wrote the main manuscript text. Katherine E. Larrimore performed experiments for Figs 1, 2, 3 and 4. Latha Kannan, R. Player Kendle, Tameem Jamal, and Matthew Barcus assisted in protein preparation. Y.G., Stephen Brimijoin and Chang-Guo Zhan designed the cocaine hydrolase mutants of BChE, $pBChE_{V2-5}$. Tsafrir S. Mor conceived the project.

Chapter 3 presented how different binding partners induce varied allosteric effects on the same PDZ domain and this mechanism is revealed by applying DCI and DFI to evaluate the complex network of interactions that contribute to the dynamic allostery. In this chapter, I focused on examining allosteric mutations that significantly contribute to the

catalytic activity of an enzyme, cocaine hydrolase variants of Butyrylcholinesterase (BChE), which has an ability to hydrolyze cocaine and used for anti-cocaine treatment. Highly efficient cocaine-metabolizing variants of BChE were designed experimentally by our collaborators by introducing mutations. However, the effect of these mutations on the enzymes' sensitivity to anticholinesterases and its specificity to choline ester substrates are unknown. Thus, I developed a coarse-grained ENM models that incorporate specific interactions based on amino-acid types (mutations) to sample conformational dynamics and used DCI and DFI analysis to uncover the crucial role of these mutations in tuning sensitivity of the variants of BChE to choline ester substrates.

4.1 Abstract

Butyrylcholinesterase (BChE) is an enzyme with broad substrate and ligand specificities and may function as a generalized bioscavenger by binding and/or hydrolyzing various xenobiotic agents and toxicants, many of which target the central and peripheral nervous systems. Variants of BChE were rationally designed to increase the enzyme's ability to hydrolyze the psychoactive enantiomer of cocaine. These variants were cloned, and then expressed using the magnICON transient expression system in plants and their enzymatic properties were investigated. In particular, we explored the effects that these site-directed mutations have over the enzyme kinetics with various substrates of BChE. We further compared the affinity of various anticholinesterases including organophosphorous nerve agents and pesticides toward these BChE variants relative to the wild type enzyme. In addition to serving as a therapy for cocaine addiction-related diseases, enhanced

bioscavenging against other harmful agents could add to the practicality and versatility of the plant-derived recombinant enzyme as a multivalent therapeutic.

4.2 Introduction

The human serum enzyme butyrylcholinesterase (BChE) is a promiscuous enzyme capable of binding and/or hydrolyzing a diverse array of compounds including many natural and man-made toxicants of the central and peripheral nervous system, unlike the highly-selective, homologous enzyme, acetylcholinesterase (AChE) (Lockridge, 2015). BChE is capable of counteracting the toxicity of various anticholinesterases by binding to them before they reach their targets in the nervous system. BChE is capable of detoxifying organophosphorous (OP) nerve agents like paraoxon, as well as acetylcholine receptor antagonists, and psychoactive plant alkaloids such as cocaine (Chen et al., 2015; Decker, 2005; Khan et al., 2005; Loizzo et al., 2008). Exogenously-supplied BChE can augment the bioscavenging capacity of the endogenous enzyme and provide broad protection by sequestering the anticholinesterase agents (Doctor and Saxena, 2005; Geyer et al., 2010a, 2010b; Saxena et al., 2015). Moreover, recombinantly-produced BChE variants with improved binding affinities and catalytic prowess can be created to improve on the parameters of the wild type (WT) enzyme.

In addition to improving BChE's binding affinity toward anticholinesterase agents, the hydrolytic activity of human BChE (hBChE) against cocaine has also been a target for improvement. The catalytic activity of WT hBChE against cocaine is measurable, albeit slow, and provides one of the major detoxification pathways for the drug, generating non-psychoactive metabolites (Carmona et al., 2000; Inaba et al., 1978). Mutants of BChE have

been rationally-designed, creating highly efficient recombinant cocaine hydrolases aimed toward an enzyme-based therapy to treat drug overdose and addiction (Pan et al., 2005; Sun et al., 2002a; Xie et al., 1999; Xue et al., 2013, 2011; Zheng et al., 2014, 2008; Zheng and Zhan, 2011). When designing BChE-based cocaine hydrolase mutants, care was taken to ensure that their ability to hydrolyze the crucially important substrate, acetylcholine (ACh), was *not* significantly enhanced.

A low-cost, sustainable, source of recombinant BChE must be readily available to produce clinically useful quantities of BChE mutants. Rapid and high level transient expression of foreign proteins in plants is needed to efficiently screen copious numbers of mutant variants, while maintaining the ability to ramp up production greatly when mutants of particular interest have been established. Mammalian expression systems have been used to produce cocaine hydrolase variants of BChE (Chen et al., 2016), but such platforms can be difficult and expensive to scale up (Connors and Hoffman, 2013). Plant-based recombinant protein production systems, in particular transient expression systems that make use of viral vectors (Fig. 4.1a), have advantages including reduced production costs, similar or cheaper downstream costs, as well as easy scalability (Chen et al., 2015; Mor, 2015; Topp et al., 2016).

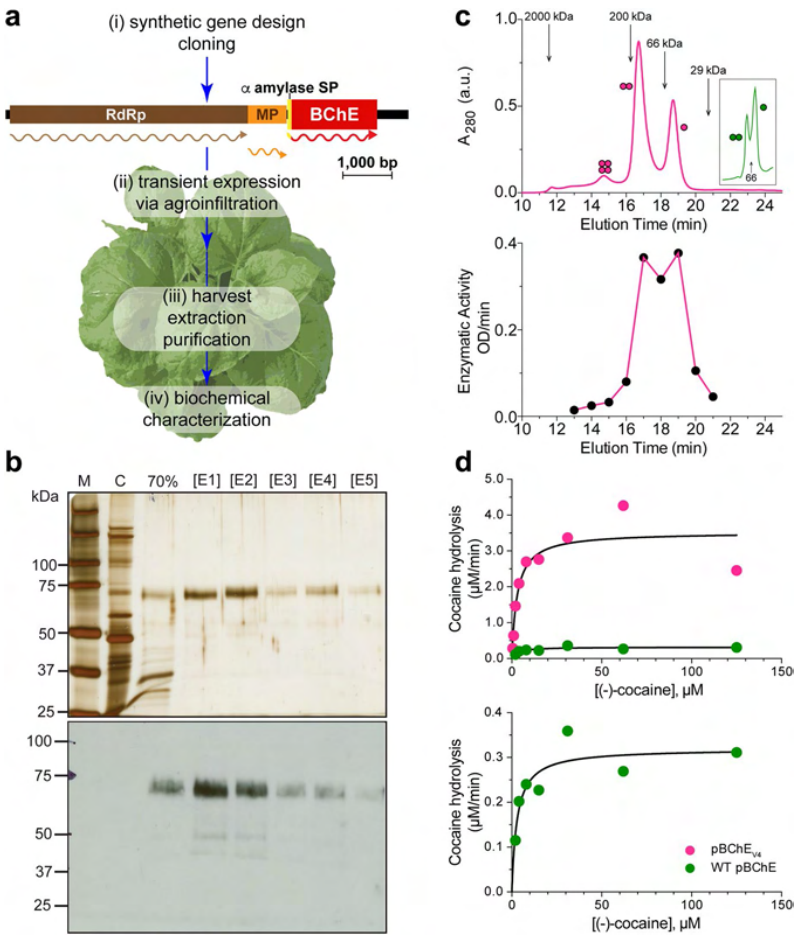


Figure 4.1: Plant production and biochemical characterization of a cocaine hydrolase variant of BChE. (**a**) Plant- based strategy for the production of BChE. (i) Plant-expression optimized synthetic genes encoding human BChE and variants thereof were cloned into the TMV-based MagnICON vector system, which recombines *in vivo* to yield a cell-to-cell-spreading replicon. (ii) WT *Nicotiana benthamiana* plants were infiltrated with agrobacteria harboring the MagnICON vectors (iii) and on peak accumulation day of the transiently- expressed recombinant enzymes, leaf material was harvested, homogenized and the enzymes were purified. Transient expression replicon: RpRd, RNA-dependent RNA polymerase; MP, movement protein gene; α,barley alpha-amylase signal peptide. Wavy lines represent the translation products of the replicon genes.(**b**) Purification of pBChEV4. Leaf extract from pBChEV4 –expressing plants was clarified by 70% $(NH_4)_2SO_4$ precipitation then subject to ConA purification and eluted with stepwise increasing concentrations of methyl- α-D-mannopyranoside ([E1]-[E5]). Samples from these purification steps, protein size markers (M) and an un-infiltrated WT *N. benthamiana* extract control (C) were subject to SDS-PAGE followed by silver-staining (top) or BChE-specific immunoblotting (bottom). Lanes in respective gels were loaded based on equal enzymatic activity. (**c**) Oligomerization of pBChEV4. Purified preparation of pBChEV4

was analyzed by SEC- HPLC; fractions were monitored for total protein content (top) and pooled fractions (0.5 mL every 1 min) for enzymatic activity (bottom). Inset: fractionation pattern of WT pBChE. Molecular mass standards are indicated with arrows. (**d**) Enzymatic hydrolysis of (−)-cocaine by WT pBChE and pBChEV4. Purified samples of WT pBChE (green, 1.21×10^{-1} μM, upper and lower panel) and pBChEV4 (pink, 6.06×10^{-4} μM, upper panel). Curves represent nonlinear regression fitted to the Michaelis-Menten model (Equation 1). Fitting the data to the Radić model (substrate inhibition, Equation 2) does not result in a significantly better fit (based on the extra sum-of-squares F test; $p > 0.12$ and $p > 0.78$ for the mutant and WT enzymes, respectively).

Our lab has previously shown that the tobacco relative *Nicotiana benthamiana* can serve as a source for clinically-relevant quantities of cocaine hydrolase variants of BChE (Geyer et al., 2008; Larrimore et al., 2013). These highly efficient cocaine-metabolizing variants of BChE were designed with the goal of increasing catalytic efficiency of cocaine hydrolysis toward an anti-cocaine treatment. But how the newly introduced mutations affect the enzymes' sensitivity to anticholinesterases and its kinetics with choline ester substrates remains unknown.

Here we report the complex kinetic behavior of the plant-derived cocaine hydrolase variants of BChE (pBChE) and their enhanced anticholinesterase scavenging ability. Using Dynamic Coupling Index (*DCI*) analysis we have evidence that the mutations allosterically affect the catalytic triad not only within a single subunit, but also propagate to neighboring subunits of the BChE oligomer.

4.3 Methods

4.3.1 Dynamic Flexibility Index (DFI) Analysis

Dynamic flexibility index (DFI) metric (Gerek et al., 2013) is based on the Perturbation Response Scanning method (PRS) that couples covariance matrix of residue displacement

with linear response theory (LRT) (Atilgan et al., 2010; Atilgan and Atilgan, 2009; Kumar et al., 2015b).

PRS was originally based on the Elastic Network Model (ENM). In ENM, protein is viewed as an elastic network, in which each amino acid is represented by their C-alpha position, and a harmonic interaction is assigned to pairs of amino-acids within a specified cutoff distance (Atilgan et al., 2010). Simply put, in the WT BChE, two residues that are interacting with each other are represented by a harmonic interaction with the same spring constant (Fig. 4.2). In contrast, a mutation at a given position is considered to destabilize the interactions of the mutational site. Thus, this destabilization is introduced in the ENM as a decrease in spring constant providing a loss in interaction strength with mutated positions (lower right in Fig. 4.2), the mutation positions S199, A227, G287, W328, and G332 are shown as red spheres).

In PRS, we apply a random Brownian kick as a perturbation to a single residue in the chain one at a time, sequentially. This perturbation mimics the external forces exerted on the protein through interactions with another protein, another biological macromolecule or small molecule ligand, *in silico*. The perturbation cascades through the residue interaction network and may introduce conformational changes in the protein. Then, we compute the response fluctuation profile of all other residues to the perturbation as linear response using Equation (4.1) where F is a unit random force on selected residues, H^{-1} is the inverse of the Hessian matrix and ΔR is the positional displacements of the N residues of the protein in three dimensions (Atilgan et al., 2001, 2010; Atilgan and Atilgan, 2009; Gerek and Ozkan, 2011; Gerek et al., 2013).

The response fluctuation profile is used to calculate the *DFI* scores using Equation (4.2), where $\left[\Delta R^j\right]_i$ is the response fluctuation amplitude of position i, upon perturbing position j. Thus, the *DFI* of position i is the total fluctuation response of position i upon perturbing all positions in the chain one at a time. The *DFI* value for each position is normalized to the overall intrinsic flexibility of the protein chain. High and low *DFI* scores could be interpreted as dynamically flexible sites and rigid (hinge) sites, respectively (Kumar et al., 2015a). The *DFI* scores can be converted into percentile ranking scores, namely % *DFI*.

$$[\Delta \mathrm{R}]_{3Nx1} = [H]^{-1}_{3Nx3N}[\mathrm{F}]_{3Nx1} \tag{4.1}$$

$$DFI_i = \frac{\sum_{j=1}^{N}\left[\Delta R^j\right]_i}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\Delta R^j\right]_i} \tag{4.2}$$

4.3.2 Dynamic Coupling Index (*DCI*) Analysis

The position-specific metric *DCI* uses PRS methodology to identify the residues that are allosterically linked to functionally critical positions through residue fluctuation dynamics for a given protein. The index *DCI* computes whether this position exhibits a higher fluctuation response to a perturbation that occurred at functionally critical sites (e.g. binding sites or catalytic site) compared to the perturbations at the other sites of the chain. It is measured as the ratio of average fluctuation responses of a given residue *j* upon perturbations of functionally critical sites to the average response of residue *j* upon perturbations placed on all other residues using:

$$DCI_i = \frac{\sum_{j=1}^{N_{Functional}}\left|\Delta \boldsymbol{R}^j\right|_i / N_{Functional}}{\sum_{j=1}^{N}\left|\Delta \boldsymbol{R}^j\right|_i \ / \ N} \tag{4.3}$$

In Equation (4.3), $\left|\Delta R^{j}\right|_{i}$ is the response fluctuation profile of residue j upon perturbation of residue i. The numerator is the average mean square fluctuation response obtained over the perturbation of the functionally critical residues $N_{functional}$; the denominator is the average mean square fluctuation response over all residues (Gerek and Ozkan, 2011; Kumar et al., 2015b).

These *DCI* profiles can also be converted into rank profiles, which are labeled as % *DCI* profiles. The positions that have higher % *DCI* values are functionally important residues that are not linked to the functional residues by direct covalent bonds or non-covalent interactions (e.g. hydrogen bonding and van der Waals interactions), but are allosterically communicating over longer distances (i.e. allosteric dynamic coupling) via residues that form extensive interaction networks.

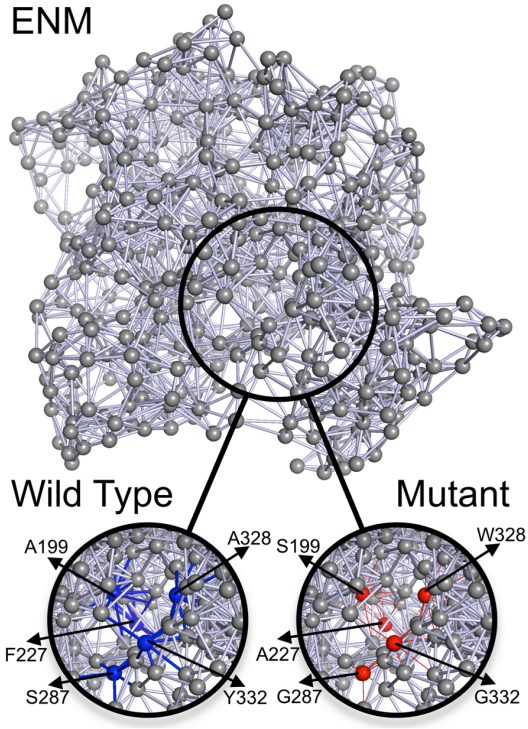


Figure 4.2: Modeling of wild type and mutant by using Elastic Network Model (ENM). WT human BChE and BChEV4 are modeled by ENM. The spheres indicate the locations of alpha carbons of each amino acid and the sticks are representing the harmonic oscillators (i.e., springs) between them. The thickness of the sticks represents the magnitude of the spring constant. For WT hBChE the mutation positions are shown as blue spheres (A199, F227, S287, A328, and Y332) and for the pentavalent mutant BChEV4 the mutation positions are shown as red spheres (S199, A227, G287, W328, and G332). The spring constant for each connection is assumed to be same for the WT (i.e., the thickness of the blue sticks is same as grey sticks). The mutation at a given position are considered to destabilize the interactions of the mutational site. This is incorporated in the model as a decrease in spring constant (low thickness values are shown as red sticks indicating a loss in interaction strength with mutated positions).

4.4 Results and Discussion

4.4.1 Plant Production of a Recombinant Cocaine-Hydrolyzing Human BChE Variant

Several research groups have been working on rational re-design of BChE into a cocaine hydrolase (Pan et al., 2005; Sun et al., 2001; Xie et al., 1999; Zheng et al., 2014; Zlebnik et al., 2014). The group led by Zhan used hybrid quantum mechanical/molecular

mechanical (QM/MM) method-based predictions followed by validation through *in vitro* and *in vivo* experiments. This process provided evidence for a correlation between the measured catalytic efficiency of cocaine hydrolysis and the sum of the enzyme-substrate hydrogen-bonding distances within the first transition state. In successive papers Zhan et al. (2014) reported the further design of BChE variants with ever increasing catalytic efficiency (Gao et al., 2006; Gao and Zhan, 2006; Liu et al., 2013; Pan et al., 2005; Xue et al., 2013, 2011; Yang et al., 2010; Zheng et al., 2010, 2008).

We previously reported on several of these variants (see Methods in Appendix A for a list of variants and their specifically-modified residues as well as Table A.2) using the deconstructed tobacco mosaic virus (TMV)-based expression system in plants (Fig. 4.1a) (Larrimore et al., 2013). This virus-assisted transient expression system exploits plant viral vectors deconstructed for the rapid, industrial-scale expression of foreign proteins (Geyer et al., 2010a, 2010b, 2008). In developing this technology, we focused on a variant, pBChE$_{V4}$ (A199S/F227A/S287G/A328W/Y332G) reported to hydrolyze cocaine close to the upper limit set by substrate diffusion rates. Recently, another BChE variant with a 6th mutation, P285A, was reported with further 2-fold better catalytic efficiency potentially bringing it to the diffusion-limited maximal theoretical ceiling (Zheng et al., 2014).

The pBChE$_{V4}$ was purified as previously reported for its WT counterpart (Geyer et al., 2010a). SDS-PAGE analysis of pBChE$_{V4}$ revealed that it resolved with an apparent molecular mass of ~65–70 kDa. This is similar to previously described plant-derived BChE variants and slightly smaller than the ~85 kDa human BChE monomer, likely due to differences in glycosylation (Fig. 4.1b) (Geyer et al., 2010a; Schneider et al., 2014a). When

highly purified pBChE$_{V4}$ was subjected to SEC-HPLC, about two thirds was dimeric (Fig. 4.1c). Most of the remainder were monomers, but low amounts of tetramers were also detected (Fig. 4.1c). Similar preparations of the WT enzyme, obtained through transient expression using the MagnICON system, showed inverse proportions of monomers and dimers (Fig. 4.1c inset). Interestingly, stable expression of WT enzyme in transgenic plants results in a substantial tetramer fraction (Geyer et al., 2010a, 2010b).

The plant-derived pBChE$_{V4}$ was examined closely for its ability to hydrolyze (−)-cocaine (Fig. 4.1d) and was found to have >2000-fold improved catalytic efficiency against that substrate (k_{cat}/K_M = 1.9 × 10^9 M^{-1} min^{-1}) compared with the WT plant-derived enzyme (k_{cat}/K_M = 9.0 × 10^5 M^{-1} min^{-1}). The higher efficiency is mostly due to a large increase in k_{cat} of pBChE$_{V4}$ as compared to WT pBChE (5805 min^{-1} *vs* 2.6 min^{-1}, respectively) with nearly identical affinity to the substrate (K_M = 3.0 μM *vs* K_M = 2.9 μM, respectively). The catalytic efficiency of the plant-derived variant and its improvement over WT BChE are in agreement with reports of this same variant derived from other sources such as human embryonic kidney-293F cells (Zhan et al., 2014).

4.4.2 Cocaine Hydrolase Variants of BChE Exhibit Altered Allosteric Effects

The specific residues changed on the road to BChE-based cocaine hydrolases included those at the bottom of the catalytic gorge near the π-cation binding site (A328) and in the peripheral anionic site (Y332) (Sun et al., 2002a; Xie et al., 1999). Catalytic activity against (−)-cocaine was further improved through additional mutations to the oxyanion hole (A199) (Gao and Zhan, 2006), entrance to the gorge (S287) (Pan et al., 2005) and non-

active site residues participating in H-bonding (F227) (Zheng et al., 2010). Together, these changes result in increased catalytic efficiency against (−)-cocaine and potentially affect the enzyme's interactions with other substrates and ligands. Indeed, preliminary results with crude preparations revealed such effects (presented in The XIth International Meeting on Cholinesterases, Kazan, Russia, June 4–9, 2012 "Plant-produced butyrylcholinesterase variants as versatile bioscavengers"). It was, therefore, of interest to determine if there were other such allosteric effects on the function of several plant-derived BChE variants.

Table 4.1: Catalytic activity of WT BChE and cocaine hydrolase variants against butyrylthiocholine and acetylthiocholine.

Substrate		WT hBChE	WT pBChE	pBChE$_{V2}$	pBChE$_{V3}$	pBChE$_{V4}$	pBChE$_{V5}$
BTC	**Kinetic behavior**	Substrate activation	Substrate activation	Modified Hill	Modified Hill	Modified Hill	Modified Hill
	k_{cat}/K_M ($M^{-1}min^{-1}$)	1.6×10^9	1.8×10^9	1.8×10^7	4.9×10^7	3.4×10^7	3.2×10^8
	k_{cat} (min)	26241.2	25992.7	664.6	463.0	2806.0	27440.1
	K_M (μM)	16.8 ± 2.9	14.6 ± 1.4	37.1 ± 21.8	9.5 ± 3.2	83.0 ± 21.5	86.2 ± 27.7
	K_{ss} (mM)	2.2 ± 0.4	2.3 ± 0.3	0.4 ± 0.3	0.3 ± 0.1	0.4 ± 0.4	0.4 ± 0.2
	b[a]	3.1 ± 0.2	2.5 ± 0.1	0.6 ± 0.2	0.3 ± 0.1	0.3 ± 0.2	0.7 ± 0.1
	n[b]	n.a.	n.a.	1.1 ± 0.3	1.1 ± 0.2	2.8 ± 0.6	1.0 ± 0.2
	x[b]	n.a.	n.a.	2.3 ± 2.0	2.3 ± 0.3	1.1 ± 0.6	1.5 ± 0.3
	R^2	0.98	0.99	0.83	0.95	0.95	0.95
ATC	**Kinetic behavior**	Substrate activation	Substrate activation	Substrate activation	Modified Hill	Michaelis-Menten	Modified Hill
	k_{cat}/K_M ($M^{-1}min^{-1}$)	2.5×10^8	9.6×10^7	3.7×10^7	9.1×10^6	8.3×10^6	2.1×10^8
	k_{cat} (min)	9185.0	8490.1	3756.3	243.2	1093.3	16154.1
	K_M (μM)	36.7 ± 3.8	89.2 ± 7.8	101 ± 13	26.8 ± 4.8	132 ± 13	77.0 ± 4.5
	K_{ss} (mM)	2.3 ± 0.4	2.7 ± 0.5	2.4 ± 1.0	1.2 ± 0.4	n.a.	7.6 ± 3.3
	b	2.2 ± 0.1	1.9 ± 0.1	1.6 ± 0.1	0.7 ± 0.1	n.a.	1.6 ± 0.3
	n	n.a.	n.a.	n.a.	0.9 ± 0.1	n.a.	1.4 ± 0.1
	x	n.a.	n.a.	n.a.	1.4 ± 0.5	n.a.	2.3 ± 1.1
	R^2	0.99	1.0	0.99	0.98	0.94	0.99

[a] When b > 1, enzyme is exhibiting substrate activation; when b < 1, enzyme is exhibiting substrate inhibition; if b = 1, the enzyme is following Michaelis-Menten kinetics.

[b] n and x represent the Hill coefficients. See Scheme 2 of Supplementary Fig A.1 online and Equation A.3. Positive cooperativity is observed when either n > 1 or x > 1. Negative cooperativity is observed when n < 1 or x < 1.

To rule out artifacts from our novel expression system, we first compared WT human plasma-derived (hBChE) to pBChE. The Michaelis-Menten constant (K_M) of WT hBChE and WT pBChE was determined with the substrate, butyrylthiocholine. Nonlinear regression analysis showed values of 16.8 ± 2.9 μM and 14.6 ± 1.4 μM respectively, similar

to previous reports (Yang et al., 2010) and essentially identical to each other (Figs 4.3 and 4.4, Table 4.1).

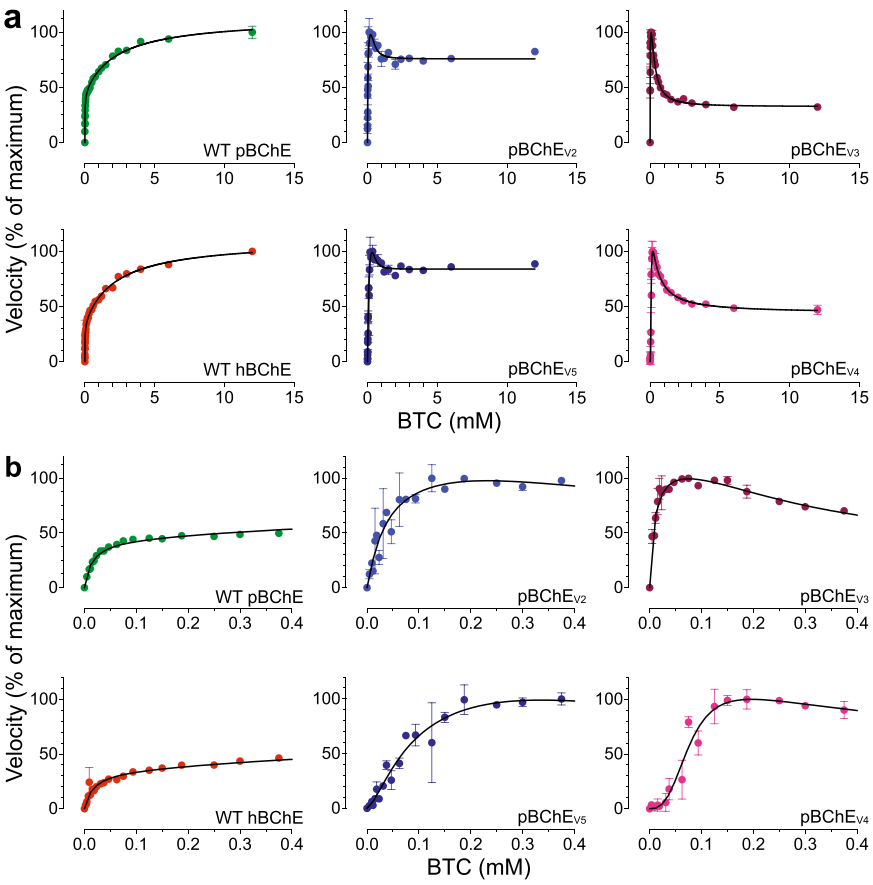


Figure 4.3: BTC hydrolysis by WT hBChE, WT pBChE, and pBChE$_{V2\text{-}5}$. (**a**) Reaction rates are plotted against substrate concentration (mean ± SEM). Plots in (**b**) zoom in on the low range of substrate concentrations. The 100% values and the goodness of fit values are as follows: WT hBChE, 100% = 1.57 ± 0.04 nmol/min, Equation (A.2), R^2 = 0.98; WT pBChE, 100% = 1.21 ± 0.07 nmol/min, Equation (A.2), R^2 = 0.99; pBChE$_{V2}$, 100% = 0.79 ± 0.10 nmol/min, Equation (A.3), R^2 = 0.83; pBChE$_{V3}$, 100% = 0.96 ± 0.01 nmol/min, Equation (A.3), R^2 = 0.95; pBChE$_{V4}$, 100% = 8.9 ± 0.6 nmol/min, Equation (A.3), R^2 = 0.95; pBChE$_{V5}$, 100% = 8.1 ± 0.3 nmol/min, Equation (A.3), R^2 = 0.95.

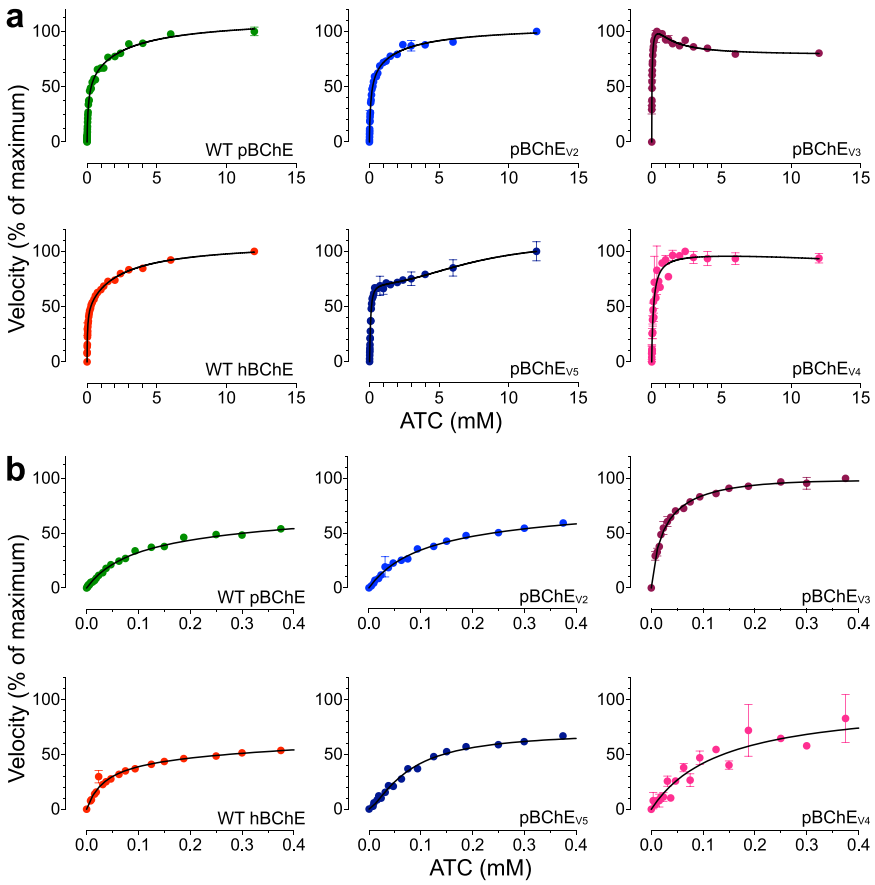


Figure 4.4: ATC hydrolysis by WT hBChE, WT pBChE, and pBChE$_{V2\text{-}5}$. (**a**) Reaction rates are plotted against substrate concentration (mean ± SEM). Plots in (**b**) zoom in on the low range of substrate concentrations. The 100% values and the goodness of fit values are as follows: WT hBChE, 100% = 0.97 ± 0.04 nmol/min, Equation (A.2), R^2 = 0.99; WT pBChE, 100% = 3.0 ± 0.1 nmol/min, Equation (A.2), R^2 = 1.00; pBChE$_{V2}$, 100% = 4.7 ± 0.1 nmol/min, Equation (A.2), R^2 = 0.99; pBChE$_{V3}$, 100% = 1.39 ± 0.00 nmol/min, Equation (A.3), R^2 = 0.98; pBChE$_{V4}$, 100% = 2.6 ± 0.1 nmol/min, Equation (A.1), R^2 = 0.94; pBChE$_{V5}$, 100% = 12.0 ± 0.7 nmol/min, Equation (A.3), R^2 = 0.99.

WT hBChE and WT pBChE also exhibited similar turnover numbers (*kcat* = 2.6 × 104 min−1 and *k*cat = 2.6 × 104 min−1, respectively) and catalytic efficiencies with the substrate analog butyrylthiocholine (BTC, *kcat*/KM = 1.6 × 109 M−1min−1 and *kcat*/KM = 1.8 × 109 M−1 min−1, respectively; Fig. 4.3, Table 4.1).

Catalytic efficiency of pBChE$_{V2}$ (F227A/S287G/A328W/Y332A), pBChE$_{V3}$ (A199S/S287G/A328W/Y332G), and pBChE$_{V4}$ (A199S/F227A/S287G/A328W/Y332G) toward BTC was reduced 100-, 37- and 5-fold, respectively, mostly due to a large reduction in the turnover number but also to small changes in the K_M. An even larger drop (~1000-fold) was observed in the catalytic efficiencies of pBChE$_{V3}$ and pBChE$_{V4}$ toward the substrate analog acetylthiocholine (ATC), but not in the case of pBChE$_{V2}$, which dropped only 3-fold. Of note, the catalytic efficiency of pBChE$_{V5}$ (F227A/S287G/A328W/Y332G) toward both substrates remained equal to the WT enzyme (Table 4.1).

Human AChE and BChE exhibit characteristic allosteric effects due to low-affinity substrate binding at the "peripheral site" (P-site) positioned near the entrance to the catalytic gorge (Barak et al., 1995; Masson et al., 2001). Despite the close homology between the two cholinesterases, their substrates exert opposite allosteric effects. AChE is inhibited by ACh concentrations above 5 mM, but BChE is *stimulated* by similar concentrations of ACh and BCh and their thioester analogues (ATC and BTC, Figs 4.2 and 4.3) (Boeck et al., 2002; Chen et al., 2012; Evron et al., 2007; Masson et al., 2001; Radic et al., 1993; Shafferman et al., 1992; Yang et al., 2010). This remains true regardless of the source of the enzyme, as both WT hBChE and WT pBChE exhibited typical substrate activation against BTC and ATC (Figs 4.3 and 4.4, Table 4.1) (Geyer et al., 2010a). A simple modification of the Michaelis-Menten model (Equation A.1) results in an adequate steady-state description of the phenomena of substrate activation and inhibition in WT cholinesterases (Equation A.2, Scheme 1 in Fig. A.2) (Radic et al., 1993).

In striking contrast, hydrolysis of BTC by pBChE$_{V3}$ and pBChE$_{V4}$ revealed partial substrate inhibition (Fig. 4.3) reminiscent of the kinetics of human AChE with ACh as was previously reported for native and plant-derived human enzyme (Evron et al., 2007; Shafferman et al., 1992). But this inhibition (~40%) was much weaker than that exhibited by AChE (>90%), and their respective peak activities were reached at BTC concentrations of approximately 70 μM and 230 μM respectively (Fig. 4.3).

Even more complex enzymatic behavior was exhibited by pBChE$_{V2}$ and pBChE$_{V5}$, which differ from each other only by, respectively, an alanine or a glycine residue at position 332 (Fig. 4.3). BTC at concentrations higher than approximately 125 to 375 μM had more limited inhibitory effect on these variants (about 20%, Fig. 4.3) as compared to pBChE$_{V3}$ and pBChE$_{V4}$. Interestingly, at still higher substrate concentrations (>2 mM) very slight but highly reproducible substrate activation re-appeared (Fig. 4.3).

Hydrolysis of the smaller substrate ATC also revealed differences between the four variants. In all tested variants, ATC has much weaker inhibitory effect on its hydrolysis. In fact, pBChE$_{V2}$ and pBChE$_{V5}$, which were somewhat inhibited by high concentrations of BTC, were clearly activated by high ATC concentrations, as was the case of WT hBChE or WT pBChE (Fig. 4.4, Table 4.1). Still, pBChE$_{V3}$ was inhibited at high concentrations of ATC (but not to the extent that BTC provoked), while pBChE$_{V4}$ was not allosterically affected by the smaller substrate, exhibiting a hyperbolic Michaelis-Menten kinetic profile (Fig. 4.4).

X Chen et al. (2015) found similar results regarding BTC for one of their cocaine hydrolyzing variants. Their mutant, hCocH, is a mammalian cell-derived equivalent of pBChEV4. They suggested that the change from substrate activation to substrate inhibition was due to destabilization of the rate-limiting step's transition state when a second substrate molecule binds in the peripheral site. This is plausible but remains speculative at this point.

The complex kinetic behavior of certain variants is reflected by the relatively poor fit between experimental data and the standard model for BChE and AChE's allosteric effects (Equation A.2, Scheme 1 in Fig. A.2). A closer look at hydrolysis rates at low BTC concentrations (Figs 4.3b and 4.4b) showed a sigmoidal pattern as BTC concentrations rise. Sigmoidal behavior is characteristic for homo-oligomeric enzymes that exhibit cooperative binding of substrate molecules. Both BChE and AChE are oligomeric, tetramers and dimers being most common *in vivo*. But the common view is that oligomerization status does not affect the enzymatic properties of either enzyme (Blong et al., 1997; Saxena et al., 2003; Velan et al., 1991). BChE purified from transgenic plants is about 50% tetrameric (Geyer et al., 2010a, 2010b), while TMV-assisted transient-expression in plants yields a mixture of monomers and dimers with few tetramers (Fig. 4.1c) (Schneider et al., 2014b).

As will be explored further, it is possible that mutations introduced into BChE to improve its activity toward cocaine also affected subunit interactions, which in turn made the enzyme behave cooperatively. Nonetheless, even monomeric enzymes with multiple substrate binding-sites, like all cholinesterases, can also exhibit cooperative (or anticooperative) binding. We found that including Hill coefficients describing

cooperativity (or anticooperativity) into the standard analysis of uncompetitive inhibition, provides an adequate model to describe the behavior of the BChE variants against BTC and ATC (Scheme 2 in Fig. A.2). While the molecular mechanism is not yet established, the suggested model yields an estimate for factors that are assumed to be negligible for the WT enzymes (Scheme 2 in Fig. A.2). Specifically, the model anticipates the possibility that binding of one substrate molecule at either the peripheral or the active site may alter the binding of a second molecule in the other site. The Hill coefficients (Table 4.1) demonstrate weak positive cooperativity, which would be particularly important at low substrate concentrations. At higher substrate concentrations, effects on k_{cat} are more prominent and result in the observed substrate inhibition (against BTC) and activation (against ATC). A non-equilibrium analysis of the interactions between the peripheral site and the active site, similar to the one offered by Rosenberry (Rosenberry, 2010), should provide further insight into the mechanism involved here.

4.4.3 Inhibition Analysis

The mutations rendering enzyme variants with the ability to efficiently hydrolyze (−)-cocaine had profound allosteric effects on cholinesterase activity and our data suggest similar effects of those mutations on sensitivities to various anticholinesterases. To test this possibility, we studied representatives of several important cholinesterase inhibitor classes including two OPs (paraoxon and Iso-OMPA), a carbamate (neostigmine) and an AChE-specific bisquaternary inhibitor (BW284c51). To this end, BTC hydrolysis was analyzed following a 30-minute incubation with the inhibitors.

Compared to the WT enzyme (either plasma- or plant-derived), pBChE$_{V2}$ and pBChE$_{V5}$ had dramatically increased sensitivity, reflected in decreased IC_{50} values. This was true for all tested anticholinesterase agents except for the AChE-specific inhibitor BW284c51 (Fig. 4.5, Table 4.2). In fact, each of the variants were 40–50 fold more sensitive to both OPs paraoxon and Iso-OMPA than the WT enzyme ($p < 0.0001$). In respect to neostig-mine, the variants were also more sensitive than WT BChE but with smaller differences (10–20 fold). Similarly, increased sensitivities were observed in an earlier plant-derived cocaine-hydrolyzing variant pBChE$_{V1}$ (A328W/ Y332A) previously described by Geyer et al. (2008).

Higher-than WT sensitivities to paraoxon were also seen with pBChE$_{V3}$ and pBChE$_{V4}$ but the increase was not as dramatic as in the other two variants (3–7 fold, Fig. 4.5, Table 4.2). The inhibition rate constants (k_i) for inhibition by paraoxon of WT pBChE and pBChE$_{V4}$ were 3.14×10^6 and 1.7×10^7 M^{-1} min^{-1} respectively. On the other hand, sensitivities to the other OP, iso-OMPA, and to neostigmine were near WT levels.

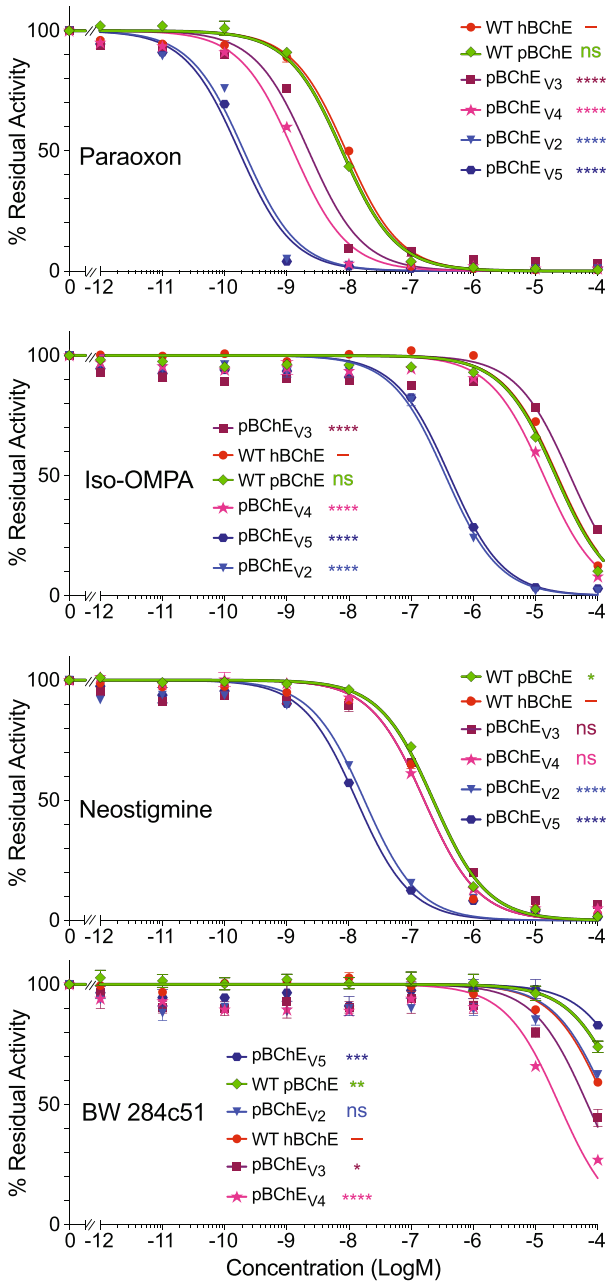


Figure 4.5: Inhibition profiles of WT hBChE, WT pBChE and pBChE$_{V2\text{-}5}$. Residual BTC hydrolytic activity (mean ± SEM) with the indicated concentrations of paraoxon and Iso-OMPA (OP inhibitors), neostigmine bromide (a carbamate inhibitor) and BW (an AChE-specific bis-quaternary inhibitor). The legends in each panel list the traces in order of decreasing IC50. Plots of variants are compared to the human plasma-derived enzyme. ns, no statistical difference; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

Table 4.2: Inhibition of BTC hydrolysis activity. Log IC_{50} values ± SEM of various anticholinesterase inhibitors versus WT hBChE, WT pBChE, and pBChE$_{V2\text{-}5}$. Fold increase in sensitivity relative to WT pBChE is shown in parentheses. Concentration of BChE variants was ~5 mU. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. ****$p < 0.0001$.

	Paraoxon		**Iso-OMPA**		**Neostigmine**		**BW284c51**	
WT hBChE	−8.04 ± 0.04	(0.9)	−4.65 ± 0.03	(0.8)**	−6.79 ± 0.02	(1)****	−3.86 ± 0.03	(2)****
WT pBChE	−8.11 ± 0.03	(1)	−4.77 ± 0.03	(1)	−6.64 ± 0.01	(1)	−3.55 ± 0.05	(1)
pBChE$_{V2}$	−9.69 ± 0.06	(38)****	−6.45 ± 0.05	(48)****	−7.75 ± 0.04	(13)****	−3.82 ± 0.11	(2)*
pBChE$_{V3}$	−8.65 ± 0.07	(3)****	−4.45 ± 0.10	(0.5)***	−6.69 ± 0.05	(1)	−4.16 ± 0.10	(4)****
pBChE$_{V4}$	−8.89 ± 0.05	(6)****	−4.87 ± 0.05	(1)	−6.81 ± 0.03	(1)****	−4.62 ± 0.09	(12)****
pBChE$_{V5}$	−9.77 ± 0.05	(46)****	−6.39 ± 0.06	(42)****	−7.88 ± 0.03	(17)****	−3.31 ± 0.11	(1)*

Particularly interesting was the unexpected and small but statistically significant increase in sensitivity of pBChE$_{V3}$ and pBCHE$_{V4}$ toward the AChE-specific bisquaternary inhibitor BW284c51 as compared to pBChE (4- and 12-fold greater than WT, respectively). Thus, these variants have another AChE-like attribute besides substrate inhibition. The inhibitor spans the length of the catalytic gorge, from the P-site at the surface of the protein to the active site at the bottom of the gorge (Felder et al., 2002). The only mutation common to both pBChE$_{V3}$ and pBCHE$_{V4}$ and absent from the other two variants is A199S. This alanine residue is conserved in both AChE and BChE and is adjacent to the catalytic triad's serine residue (S198). The equivalent position in AChE was not identified in previously published research as relevant for the interaction with BW284C51. The hydroxyl of the serine residue may contribute a new H-bond to interact with the central carbonyl of BW284c51. However, without further data we cannot rule out contributions to the increased sensitivity through direct and/or allosteric effects of other residue changes in addition to the A199S mutation.

The substantially enhanced sensitivity of the cocaine hydrolases, especially pBChE$_{V2}$ and pBChE$_{V5}$ to the OP poisons, has important potential implications for detoxifying these harmful substances. One approach to detoxification is to supply WT BChE to scavenge nerve agents. Another approach is to enhance BChE's binding affinity to anticholinesterase agents and create a more effective bioscavenger. Excitingly, though perhaps not surprising, mutations to BChE intended to enhance cocaine hydrolysis have altered the binding affinity of the cocaine hydrolases toward anti-cholinesterase inhibitors. The enhanced scavenging for OP nerve agents by the cocaine hydrolyzing variants suggests further development for dual use of the biologics.

4.4.4 Dynamic Coupling Index (DCI) Analysis Predicts Allosteric Coupling Between the Pentavalent Mutations of pBChE$_{V4}$ and its Active Site

While all four variants exhibit novel enzymatic properties, the superior efficiency of cocaine hydrolysis of pBChE$_{V4}$ compared to other variants (Fig. 4.1d, unpublished data and Zheng et al. (Zheng et al., 2014, 2008)), prompted further investigation of this variant. Specifically, we reasoned that the altered substrate preference, hydrolysis kinetics, and inhibitor sensitivity suggest that the mutated positions in pBChE$_{V4}$ (A199S/F227A/S287G/A328W/Y332G), all but one being quite distal to the active site, may be allosterically linked to the catalytic locus. To test this hypothesis we used a recently developed metric, the "Dynamic Coupling Index" (*DCI*) (Kumar et al., 2015b) that identifies residues exhibiting significant fluctuation upon perturbation of functionally important loci including the active catalytic site and other substrate binding sites in the protein (Gerek and Ozkan, 2011).

Using *DCI* analysis, we identified positions that dynamically couple to residues of the catalytic triad, i.e. S198, E325 and H438. According to this analysis, positions exhibiting high *DCI* values present residues that are dynamically linked to the active site despite being far away from the catalytic residues.

In Fig. 4.6a, the %*DCI* values for human BChE upon perturbation of the three catalytic residues are color-coded within a spectrum of red-white-blue (from highest to lowest respectively). It appears that the five mutated positions of pBChE$_{V4}$ variant (A199S, F227A, S287G, A328W and Y332G) are highly coupled to the catalytic triad. Conversely, a reciprocal analysis of perturbing the five mutated positions and measuring %*DCI* values for other residues show that the catalytic triad's residues are highly coupled to these five mutated positions (Fig. 4.6b). This reaffirms our hypothesis concerning dynamic interplay between these mutated positions and catalytic residues. Moreover, the strong dynamic coupling between mutational sites and the catalytic site suggests that mutations alter the conformational dynamics of the enzyme, leading to changes in enzymatic function.

The changes in catalytic properties of BChE variants can be partially attributed to the direct allosteric effect of peripheral amino-acid substitution on the catalytic triad suggested by *DCI* analysis (Fig. 4.6). Our results also raise the possibility that such mutations affect the interactions between enzyme subunits – specifically they may lead to increased enzymatic cooperativity (Figs 4.3 and 4.4).

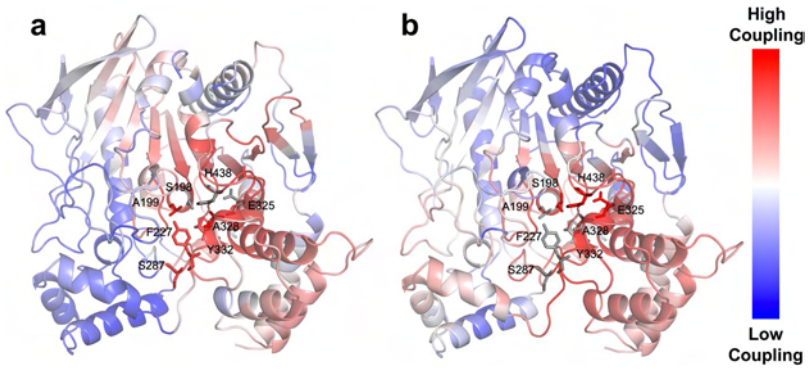


Figure 4.6: %*DCI* profile of WT hBChE. The %*DCI* profiles for hBChE are color-coded in a cartoon diagram from a spectrum of red-white-blue (red -highest, blue -lowest coupling to perturbation locations). (**a**) Upon perturbation of catalytic residues (S198, E325, and H438 shown as grey sticks) the five mutation positions (A199, F227, S287, A328, and Y332 shown as red sticks) shows high coupling (high % *DCI* values). (**b**) Upon perturbation of five mutation positions (A199, F227, S287, A328, and Y332 shown as grey sticks) the catalytic residues (S198, G325, and H438 shown as red sticks) shows high coupling (high % *DCI* values).

4.4.5 Dynamic Flexibility Index (DFI) Analysis Predicts Global Flexibility Changes Upon Introduction of Mutations

To further substantiate our hypothesis and provide mechanistic insights on how these five mutations lead to changes in enzymatic behavior, we explored the conformational dynamics of the WT and the pBChE$_{V4}$ variant using a dynamic flexibility index (*DFI)*. *DFI* computes the fluctuation response of a given position to the perturbations that occur at different parts of the protein using linear response theory, capturing the multi-dimensional effects when the protein structure is displaced out of equilibrium for example when interacting with small molecules or other cellular constituents. *DFI* allows us to identify and map flexible and rigid positions in the structure (Butler et al., 2015; Gerek et

al., 2013). *DFI* can be considered a measure of the local conformational entropy of a given position within the set of interactions governed by the 3D fold of the protein due to its ability to probe the conformational space of a protein at the residue level. For example, we recently used *DFI* to provide mechanistic insights about emergence of new functions during the evolution of several protein families (Kim et al., 2015; Zou et al., 2015) and to explain the molecular basis of single-nucleotide polymorphisms associated with genetic-diseases (Butler et al., 2015; Kumar et al., 2015b, 2015a).

We measured the *DFI* values of residues for WT hBChE and pBChE$_{V4}$, and the % *DFI* profiles shows us the flexibility of the proteins in ranking order (Fig. 4.7a). Examining the flexibility of the monomer-monomer contact (binding) interfaces (Fig. 4.7b), it appears that the dimerization surface of pBChE$_{V4}$ is less flexible in comparison with the WT counterpart. Rigidified monomer-monomer interface is often associated with increased affinity (Alvarez-Garcia and Barril, 2014; Li et al., 2015). The association constant for dimerization depends on the entropic cost at the binding interface: dimerization causes the binding interface to be more rigid and is therefore causing a decrease in entropy (negative entropy change associated with dimerization i.e. $\Delta S_{dimerization} < 0$). Because the entropy level associated with the WT contact surface is higher than in the mutant (i.e., the former is more flexible than the latter). Hence the entropic cost of dimerization is higher in WT than in the mutant (i.e. $\Delta S_{dimerization}$ of WT is more negative than that of the mutant). These results support our observation that preparations of pBChE$_{V4}$ have higher proportion of dimers as compared to pBChE, which is mostly monomeric.

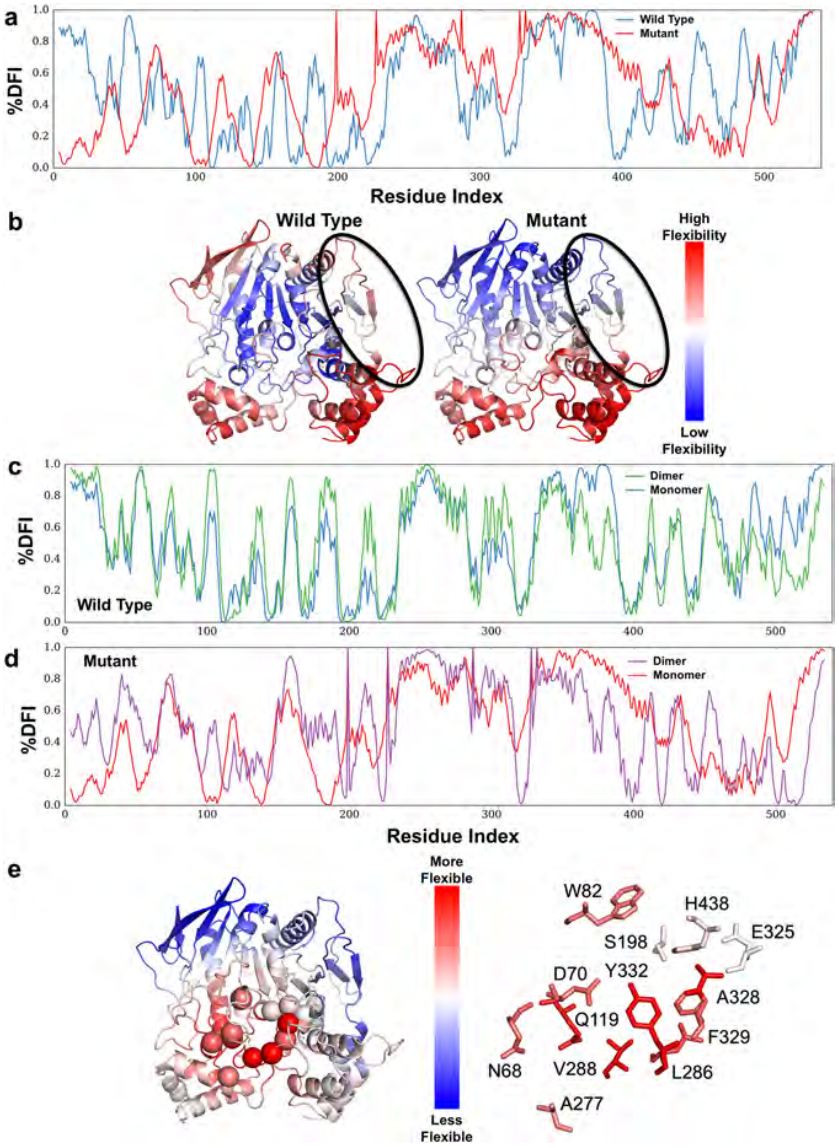


Figure 4.7: %*DFI* profile of WT hBChE and pentavalent mutant. (**a**) The %*DFI* profiles of WT BChE (blue) and BChE$_{V4}$ (x-axis – residue numbers, y-axis – %*DFI* values at each position). (**b**) Color-coded structure diagrams depicting the %*DFI* values at each position. The circled regions are part of the monomer-monomer contact region (V377, D378, T457, K458, A459, I462, Y500, R509, M511, T512, K513, R514, L515). (**c**) The % *DFI* profiles of monomeric (blue) and dimeric (green) WT BChE. (**d**) The %*DFI* profiles of monomeric (red) and dimeric (purple) BChE$_{V4}$. (**e**) Color-coded structure diagrams depicting the values of %*DFI* differences between the dimeric forms of WT BChE and BChE$_{V4}$ at each position. The red-white-blue code reveals loci with increased flexibility (shades of red), decreased flexibility (shades of blue) or no change (white) in the mutant dimer vs. the WT dimer.

The oligomerization of WT BChE is usually not regarded as affecting the enzymatic properties of the enzyme. However, the sigmoidal nature of the enzyme kinetics observed here in the mutant variants (Figs 4.3 and 4.4) suggests a degree of cooperativity. If this is the case, we should see that dimerization induces new conformational dynamics in the mutant but less so in the WT. To test this possibility, we explored how dimerization may affect the dynamics of each monomeric subunit in WT and the mutant using *DFI* analysis. The five residue substitutions are introduced into the Elastic Network Model (ENM) model (see Methods) at the core of the *DFI* analysis as changes in the spring constants of the harmonic oscillators interconnecting the alpha-carbons of the adjoining amino-acids (Fig. 4.2). In other words, since the mutations introduced local destabilization around the mutational sites, we modeled this effect as decreased spring constants for the interactions of the mutational positions (i.e. weakened harmonic interactions of the mutational sites).

With this approach, we can predict global changes in flexibility upon introduction of mutations. The local disruption due to the mutations not only introduce enhanced flexibilities at the mutational sites, but can create a global flexibility change in all positions (i.e. change in *DFI* profile) due to network of interactions. In fact, it appears that the change in the flexibility of one region is compensated by the changes in flexibility of other regions. As could be expected based on the well-documented lack of cooperativity in BChE upon its oligomerization, the *DFI* profile of the WT hBCHE subunit in monomeric and dimeric form are quite similar (Fig. 4.7c). On the other hand, in the case of the BChEV4 mutant, the *DFI* profile of subunit in the dimer form differs notably from that in monomeric form (Fig. 4.7d). This change suggests that dimerization induces new conformational dynamics.

Interestingly, when we map the localized differences in the % *DFI* values between mutant and WT, we observed that the mutations lead to enhanced flexibility near the gorge site in the dimer, but less so in the monomer (Fig. 4.7e). The peripheral anionic site (D70, N68, Q119, A227), cation- π domain (W82, A328), acyl pocket (L289 V288), and phenothiazine ring site (Y332, F329) exhibited increased flexibility upon mutations, while the rigid profile of the catalytic triad (S198, E235, H438) did not change.

The *DFI* analysis suggests that compared to the WT, BChE$_{V4}$ should have a better propensity to dimerize and that within the mutant dimer there is an increase in flexibility near the gorge (Fig. 4.7e). We propose that changes in flexibility might facilitate propagation of conformational changes from one subunit to the other. Thus, at low substrate concentrations, binding of a substrate molecule on one of the subunits might positively affect substrate binding and/or turnover at the catalytic gorge of the other subunit, explaining the sigmoidal kinetic observed at low substrate concentrations (Figs 4.3b and 4.4b). At higher substrate concentrations, allosteric effects within each subunit may lead to inhibition countering the cooperative enhancement and explaining the observed partial substrate inhibition (Figs 4.3a and 4.4a). While speculative, this suggested mechanistic explanation raises several predictions that will be tested by further experimentation and simulation including substrate docking.

4.5 Conclusions

The biochemical characterization of the plant-derived cocaine hydrolases reported here offers not only a better understanding of a novel anti-cocaine treatment, but also possible protection from potent pesticides and other anticholinesterase agents. The outcomes

demonstrate the practicality and versatility of plant-derived recombinant enzymes as potential multivalent biologics. As new mutations are being found to establish even more efficient cocaine hydrolases, the results reported here point toward the importance of testing these enzymes for their altered kinetic behavior toward their substrates and their potential as OP bioscavengers.

4.6 Acknowledgement

We would like to thank Dr. Yang Gao (Mayo Clinic Rochester) for valuable advice especially concerning the setting up of the cocaine hydrolysis assay. Work was supported in part by the National Institute for Drug Abuse Grant DP1 DA031340 awarded to the Mayo Clinic and subcontracted to ASU.

CHAPTER 5

ALLOSTERIC REGULATORY CONTROL IN DIHYDROFOLATE REDUCTASE IS REVEALED BY DYNAMIC ASYMMETRY

This chapter is adapted from: "Kazan, I. Can, Jeremy H. Mills, and S. Banu Ozkan. Allosteric Regulatory Control in Dihydrofolate Reductase is Revealed by Dynamic Asymmetry. Protein Science: e4700., https://doi.org/10.1002/pro.4700"

In chapters 3 and 4, I successfully applied DFI and DCI metrics to investigate the dynamics of various proteins, and how changes in dynamics could relate to protein function. Following, here, in chapter 5, to evaluate the effect of mutations in allosteric residues on enzyme activity, Dihydrofolate reductase (DHFR) from *Escherichia coli* is used as the model enzyme system. DHFR is a vital protein in various biological processes in the cell such as DNA synthesis, and cell growth. To this end, I performed MD simulation to study the dynamics of functional loops in wild-type DHFR and used DFI and DCI analysis to evaluate how mutations on positions distal to functional looops modulate function. Besides two computational metrics DFI and, DCI, I used an asymmetric version of DCI (DCI_{asym}) to uncover and developed a new classification per positions as "Controlled" or "Controller" of functional loops. The classification not only allowed us to predict functionally beneficial or detrimental substitutions, but also shed light on the underlying mechanisms by identifying specific evolutionarily non-conserved residues that

exert control over the dynamics of functional loops, providing valuable guidance aimed at enhancing enzymatic activity in search for therapeutic applications.

5.1 Abstract

We investigated the relationship between mutations and dynamics in *Escherichia coli* dihydrofolate reductase (DHFR) using computational methods. Our study focused on the M20 and FG loops, which are known to be functionally important and affected by mutations distal to the loops. We used Molecular Dynamics simulations and developed position-specific metrics, including the Dynamic Flexibility Index (DFI) and Dynamic Coupling Index (DCI), to analyze the dynamics of wild-type DHFR and compared our results with existing deep mutational scanning data. Our analysis showed a statistically significant association between DFI and mutational tolerance of the DHFR positions, indicating that DFI can predict functionally beneficial or detrimental substitutions. We also applied an asymmetric version of our DCI metric (DCI_{asym}) to DHFR and found that certain distal residues control the dynamics of the M20 and FG loops, whereas others are controlled by them. Residues that are suggested to control the M20 and FG loops by our DCI_{asym} metric are evolutionarily non-conserved; mutations at these sites can enhance enzyme activity. On the other hand, residues controlled by the loops are mostly deleterious to function when mutated and are also evolutionary conserved. Our results suggest that dynamics-based metrics can identify residues that explain the relationship between mutation and protein function or can be targeted to rationally engineer enzymes with enhanced activity.

5.2 Introduction

The human-microbial antibiotic arms race has prompted extensive research efforts aimed at both developing new drugs and gaining a complete understanding of druggable enzymes (Aminov, 2010; Davies and Davies, 2010; Laxminarayan et al., 2013; Martínez, 2008; Weinreich et al., 2006). One such enzyme is dihydrofolate reductase (DHFR), which has been investigated for its fundamental role in 5,6,7,8-tetrahydrofolate (THF) synthesis (Luk et al., 2013; McCormick et al., 2021; Reynolds et al., 2011; Schnell et al., 2004; Thompson et al., 2020). Due to an abundance of biophysical data (Schnell et al., 2004), DHFR from *Escherichia coli* represents an excellent model system for studying the relationship between protein dynamics and function. The catalytic activity of *E. coli* DHFR has been extensively studied. One major achievement in these studies was the crystallization of DHFR in conformations that represent intermediate steps of the enzymatic reaction pathway (Boehr et al., 2006; Sawaya and Kraut, 1997). These experiments revealed that multiple loops in DHFR are implicated in its function. For example, the M20 loop (residues 9-24) controls access to the active site and the FG loop (residues 116-132) stabilizes the M20 loop through hydrogen bonding interactions (Boehr et al., 2006; Cammarata et al., 2015; Sawaya and Kraut, 1997). Mutations on both of these loops have been reported to severely limit the activity of DHFR (Benkovic et al., 1988; Thompson et al., 2020), whereas positions distal to these sites can be altered to enhance activity (Agarwal et al., 2002; Benkovic et al., 1988; Bhabha et al., 2011; Rod et al., 2003; Thompson et al., 2020; Wang et al., 2006; Wong et al., 2005).

The dynamics of *E.coli* DHFR have also been thoroughly studied in an effort to gain insight into the impact of point mutations on its activity (Bhabha et al., 2013; Epstein et al., 1995; Gekko et al., 2000; Rodrigues et al., 2016; Tamer et al., 2019). These studies revealed that mutations in DHFR often modulate the enzyme's activity indirectly and at a distance (Gekko et al., 2000; Tamer et al., 2019). Namely, in a series of computational and experimental studies, mutations distal to the active site of DHFR were shown to alter hydrogen bonding interactions and the rotamers of residues close to the active site through a network of interacting residues (Bhabha et al., 2013; Campitelli and Ozkan, 2020; Epstein et al., 1995; Gekko et al., 2000; Mauldin et al., 2009; Mauldin and Lee, 2010; Rodrigues et al., 2016). The critical role of dynamics in DHFR function has been studied previously (Bhabha et al., 2013; Campitelli and Ozkan, 2020; Epstein et al., 1995; Gekko et al., 2000; Mauldin et al., 2009; Mauldin and Lee, 2010; Rodrigues et al., 2016; Tamer et al., 2019) but a general relationship between dynamics of each position and their contribution to the activity has yet to be elucidated.

We hypothesize that the position-specific dynamic features of DHFR can shed light on the diverse impact of mutations on its activity. Therefore, we thoroughly examined the dynamics of DHFR utilizing three computational metrics: the Dynamic Flexibility Index (DFI) (Kumar et al., 2015b; Larrimore et al., 2017; Modi et al., 2021b), the Dynamic Coupling Index (DCI) (Campitelli et al., 2021; Campitelli and Ozkan, 2020; Larrimore et al., 2017), and an asymmetric version of DCI, which we call DCI_{asym} (Campitelli et al., 2021; Campitelli and Ozkan, 2020; Ose et al., 2022). The DFI metric measures the normalized magnitude of response of a residue to perturbations applied on all other amino

acids; a high DFI value indicates high flexibility, conversely, a low DFI score suggests that a residue is highly rigid. Our DCI metric reports on the dynamic coupling between residues (Butler et al., 2018, 2015; Campitelli et al., 2021; Campitelli and Ozkan, 2020; Kazan et al., 2022; Kolbaba-Kartchner et al., 2021; Larrimore et al., 2017; Modi et al., 2021b; Modi and Ozkan, 2018). A high DCI value indicates high dynamic coupling between residues *i* and *j*, while a low DCI score implies weak coupling between these residues. Due to the complex conformational dynamics of a protein, the DCI score between two distal, non-interacting residues is not necessarily symmetric. We therefore developed a new metric called DCI asymmetry (DCI_{asym}), which reports the difference in fluctuation response of residue *i* when *j* is perturbed versus the response of residue *j* when *i* is perturbed ($DCI_{ij} - DCI_{ji}$). DCI_{asym} can therefore be used to assess which of a pair of residues dominates the control of motion between them.

In this study, we applied DFI, DCI, and DCI_{asym} to molecular dynamics (MD) simulations of DHFR. Because the M20 and FG loops of DHFR are highly important for its function (McCormick et al., 2021; Reynolds et al., 2011; Sawaya and Kraut, 1997; Schnell et al., 2004; Thompson et al., 2020), we used our DCI and DCI_{asym} metrics to assess whether the distal regions of DHFR dynamically modulate these loops. We then compared these analyses to a published deep mutational scanning dataset (Thompson et al., 2020), which allowed us to link the activity of DHFR to its dynamics. Our dynamics metrics provided a link between the previously reported mutational data and collective motions of the enzyme. In particular, DCI_{asym} allowed us to classify a given residue position as "controlled" (i.e., dynamically controlled by the loops) if its fluctuation response to a

perturbation on M20 & FG loops is considerably lower than the response of M20 & FG loops when that residue is perturbed. If the opposite is true, we classified that residue as "controller" (i.e., the residue is dynamically controlling the loop). When we analyzed the mutational outcome of the controlled and controller positions using previous deep sequencing data, we observed that "controller" positions act as allosteric hot spots (i.e., mutations at these positions modulate DHFR activity), whereas mutations on "controlled" positions are usually deleterious. Thus, dynamics based approaches (particularly the "controller" and "controlled" classification) could be used to better understand the relationship between protein dynamics and function in other proteins.

5.3 Computational Methods Used to Determine the Relationship Between Mutations and Dynamics In *Escherichia Coli* Dihydrofolate Reductase (DHFR)

5.3.1 Molecular Dynamics Simulations

We used the AMBER molecular dynamics software (Salomon-Ferrer et al., 2013b) to study the dynamics of *E. coli* DHFR. The protein system is parametrized with ff14SB force field (Maier et al., 2015) and solvated with TIP3P explicit water model using minimum 16Å distance from the protein to define the box size. The solvated protein is neutralized by sodium and chlorine ions and the energy is minimized with a steepest descent algorithm by ten thousand steps. The production trajectory was simulated with Isothermal, isobaric, constant number of particles ensemble (NPT) at 300K and 1 bar pressure. Langevin thermostat was utilized to maintain the kinetic temperature of 300K, and the pressure is regulated by the Berendsen barostat. Additionally, SHAKE algorithm was used constrain the hydrogens. The simulation is run for 2μs until convergence is achieved. We considered

the simulation converged when the root mean square deviation (RMSD) between the highest sampled conformation in consecutive time windows (i.e., the last 300ns windows and the 300ns window sequentially before it) is lower than 1Å. Similar to the procedure described by Sawle and Ghosh (2016), we used window sizes ranging from 100ns up to 1μs to determine convergence.

5.3.2 Dynamic Flexibility Index (DFI)

Our DFI metric calculates the relative flexibility/rigidity of individual residues in an protein (Campitelli et al., 2020; Kumar et al., 2015b; Larrimore et al., 2017; Modi et al., 2021b; Stevens et al., 2022b). The DFI algorithm, which is developed using Linear Response Theory (LRT) and Perturbation Response Scanning (PRS), calculates the average response of a residue as a result of a perturbation on every other residue in a protein (Gerek and Ozkan, 2011). Taking advantage of the residue covariances, DFI provides position specific flexibility profiles.

$$[\Delta \mathrm{R}]_{3Nx1} = [H]^{-1}_{3Nx3N}[\mathrm{F}]_{3Nx1} \tag{5.1}$$

A Hessian matrix, $\boldsymbol{H}$, is compiled from the second derivatives of potentials. The inverse of the Hessian matrix, $\boldsymbol{H^{-1}}$, contains residue covariances. The covariance matrix can be generated from a protein structure by utilizing an Elastic Network Model (ENM) or gathered from a MD simulation of the protein, which implicitly accounts for amino-acid side chain interactions and solvent interactions. We used the latter to calculate the dynamic metrics in this study. The residue response vector, $\Delta\boldsymbol{R}$, is the resultant vector containing the magnitude of responses from multiplying the covariance matrix by the force vector, $\boldsymbol{F}$.

The dynamic flexibility index (DFI) for position i, which computes the normalized fluctuation response of a position upon perturbation on the chain is calculated as

$$DFI_i = \frac{\sum_{j=1}^{N} \lceil \Delta R^j \rceil_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} \lceil \Delta R^j \rceil_i} \quad (5.2)$$

where $\left| \Delta R^j \right|_i = \sqrt{\langle (\Delta R)^2 \rangle}$ is the magnitude of fluctuation response at position i due to a perturbation at position j.

The DFI score yields position specific information about the conformational dynamics of a protein system. Positions displaying low DFI scores are highly rigid. These sites often make more than an average number of interactions with their neighbors, which suggests that they represent crucial dynamic hubs in a protein. Conversely, positions with high DFI scores are often highly mobile regions of a protein. These sites do not contribute to the collective motion of a protein as substantially as the rigid regions.

5.3.3 Dynamic Coupling Index (DCI) and DCI asymmetry (DCI_{asym})

The DCI metric stems from the same fundamental analysis method that is used to carry out a DFI calculation (Campitelli et al., 2021; Larrimore et al., 2017). DCI measures the allosteric coupling between residue pairs. To carry out DCI analysis of DHFR, a random unit force was applied to residues contained in M20 & FG loops and was allowed to propagate through the protein until it reached a residue distal from the initial perturbation location. After probing all active site residues, we calculate a "magnitude of response" to other residues in the protein, which ultimately represents the strength of coupling between each active site residue and all other residues in the protein. A calculated DCI of position *i* suggests its response to a perturbation on position *j* and is calculated as follows:

$$\mathrm{DCI}_{ij} = \frac{|\Delta R^j|_i}{\sum_{j=1}^{N} |\Delta R^j|_i / N} \tag{5.3}$$

where, $\left|\Delta \mathrm{R}^{\mathrm{j}}\right|_{\mathrm{i}} = \sqrt{\langle (\Delta \mathrm{R})^2 \rangle}$ is the magnitude of the fluctuation response at position i due to perturbations at position j normalized over the average response of position i when any position in the protein is perturbed by a random Brownian force. Thus, $\mathrm{DCI}_{ij} > 1$ indicates that position i is more sensitive to perturbations occurring on position j. Alternatively, a position with a DCI_{ij} value lower than 1 is regarded as weakly coupled to the site j. Moreover, the dominance in dynamic control can be determined by calculating the asymmetry between residue locations i and j. DCI_{ij} is defined as the response of residue i when residue j is perturbed and DCI_{ji} represents the response of residue j when residue i is perturbed. DCI asymmetry (DCI_{asym}) (Campitelli et al., 2021; Campitelli and Ozkan, 2020; Ose et al., 2022) of location i is calculated as follows:

$$DCI_{asym} = DCI_{ij} - DCI_{ji} \tag{5.4}$$

Given this definition, DCI_{asym} can take both positive and negative value. Accordingly, we consider residues with DCI_{asym} values around zero (between negative -0.05 to 0.05) to be dynamically coupled with the M20 and FG loops in a symmetric fashion. The residues with DCI_{asym} values higher than 0.05 are considered as "controlled" (e.g., M20 loop controlled) and the ones with DCI_{asym} values lower than -0.05 as "controller" (e.g., M20 loop controller).

5.3.4 Solvent Accessible Surface Area (SASA)

The SASA calculation is employed by using Naccess (Hubbard and Thornton, 1993). Naccess algorithm first creates a sphere with the radius of a water molecule and then rolls the sphere on the surface of the protein. The accessible surface area is calculated per residues by measuring the fraction of residue that is accessible to the solvent.

5.3.5 Network Features

Network Analysis of Protein Structures (NAPS) webserver is utilized to calculate the network features betweenness, closeness, and eigenvector centrality (Chakrabarty and Parekh, 2016). Betweenness measures how often an amino acid lies on the shortest path between two other amino acids in the protein. High-betweenness nodes have been previously shown as important residues for protein structure and function (del Sol and O'Meara, 2005). These residues are crucial in proteins, as the shortest paths between nodes (i.e., distal sites and active sites) pass through these nodes. Closeness metric shows how easily an amino acid can be reached by other amino acids in the protein. Eigenvector centrality measures how well an amino acid is connected to other important amino acids in the protein. Amino acids that are more easily reached by others and well connected to other important amino acids are important for maintaining the overall stability and function of the protein (Chakrabarty and Parekh, 2014; del Sol et al., 2006; van den Bedem et al., 2013).

5.3.6 Number of Contacts

To determine the average number of contacts, we analyzed the MD simulation trajectory by counting the Cα contacts within 10Å for each residue that appeared in over 80% of the frames in the trajectory sampled every 1ns.

5.4 Results and Discussion

5.4.1 Distinguishing Tolerant vs Non-Tolerant Mutations and Understanding Mutational Outcomes Using Dynamic Flexibility Analyses

To gain insight into the impact of mutations on DHFR activity, we first investigated the enzyme using DFI (Figure 5.1). In our DFI analysis, we use Brownian force perturbation to capture each position's response to random perturbations exerted on the protein chain. When a mutation creates a disturbance in the equilibrium dynamics of the protein, the local network of interactions surrounding the mutational site is often altered. Thus, the DFI value of a position can give a first order approximation of the impact that a mutation at that site might have on the enzyme's activity. We previously demonstrated that a high correlation exists between DFI values and modulation of protein function by disease-related mutations (Campitelli et al., 2021; Campitelli and Ozkan, 2020; Modi et al., 2021b; Modi and Ozkan, 2018). Rigid locations identified by our DFI metric are often linked to disease related outcomes when mutated (Butler et al., 2018, 2015; Gerek et al., 2013; Kumar et al., 2015b; Ose et al., 2022); alternatively, flexible locations are less prone to these types of disadvantageous mutations.

Our efforts to study the relationship between dynamics and function in DHFR began with molecular dynamics (MD) simulations of the enzyme using a model of apo DHFR

(PDB ID: 1rx2) from the Protein Databank. We chose to focus our simulations on the apo protein because previous studies using NMR suggested that the apo enzyme also samples bound state dynamics (Beach et al., 2005; Boehr et al., 2006). Thus, we believed that use of an apo structure would provide insight into the dynamics of DHFR in conformations present in both the apo and bound forms. We then analyzed these MD simulations using our DFI metric, which revealed that previously known functionally important M20 and FG loops display dynamics profile different than each other (Figure 5.1). In our previous studies (Gerek and Ozkan, 2011; Kolbaba-Kartchner et al., 2021; Modi et al., 2021b), we observed that residues that directly interact with ligands are often more rigid, and maintenance of this rigidity is important for enzyme function. We observe a similar trend with residue M20 and its neighbors (N18 and A19) in the M20 loop (Figure 5.1C), which directly interact with DHFR's substrate. These positions are more rigid in our analysis than the remainder of residues in the loop. Moreover, residues D122 and T123 in the FG loop, and V13 and G15 on the M20 loop (Figure 5.1C) are also found to be less flexible relative to other residues within these loops. These residues have previously been shown to stabilize the neighboring loops (Luk et al., 2013; Reynolds et al., 2011; Sawaya and Kraut, 1997; Thompson et al., 2020). In agreement with our previous studies (Butler et al., 2018, 2015; Gerek et al., 2013; Kumar et al., 2015b; Ose et al., 2022), substitutions at M20 and FG loop positions with low DFI scores are experimentally shown to drastically diminish (if not abolish) the activity of DHFR (Thompson et al., 2020). Overall, the M20 loop shows a lower average DFI score ($\langle \%DFI \rangle = 0.38$) compared to the FG loop ($\langle \%DFI \rangle = 0.77$) implying modulation of DHFR activity by these loops are different (Figure 5.1B).

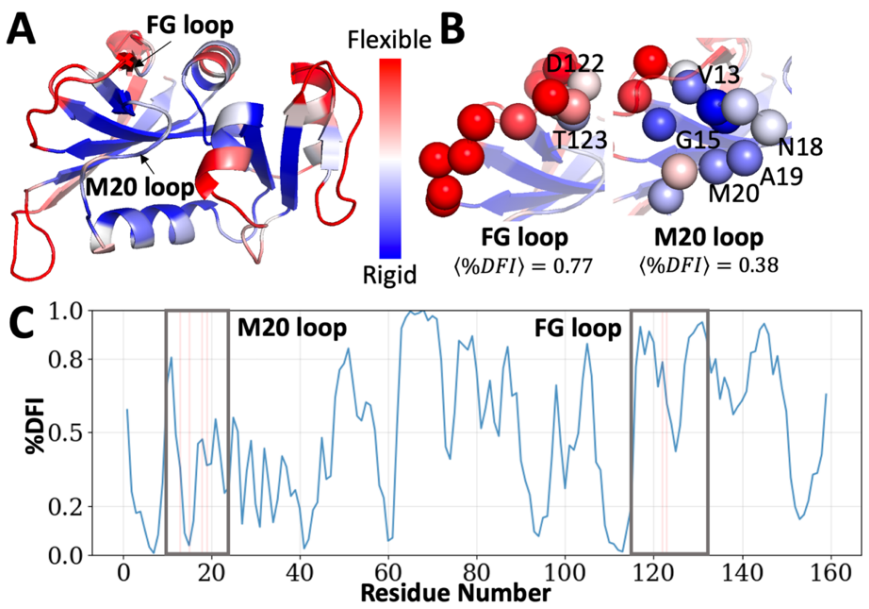


Figure 5.1: A) DFI profile of DHFR projected on the crystal structure with PDB ID: 1rx2. Regions of high flexibility are colored red; regions of medium flexibility are colored white and highly rigid regions are colored blue. B) Functionally critical residues in the M20 (residues 9-24) and FG (residues 116-132) loops are shown as spheres colored by their DFI scores. C) The DFI profile of apo DHFR. D122 and T123 in the FG loop; and V13, G15, N18, and A19 on the M20 loop are highlighted with red colored vertical lines.

To further understand the implication of the conformational dynamics of residues related to function in DHFR, we sought to relate data obtained using our DFI metric to previously reported experimental data (Thompson et al., 2020). Namely, we used the per residue functional classification defined by Thompson et al (2020), in which positions with advantageous mutations (named "Beneficial"), positions with WT-like behavior (named "Tolerant"), positions that possess both advantageous and disadvantageous mutations (named "Mixed"), residues with mostly disadvantageous mutations (named "Restricted"), and locations that exhibited almost no activity when mutated away from the wild-type amino acid (named "Intolerant") are described. When we analyzed positions belonging to

the aforementioned categories from the perspective of their DFI values, we observed the following trends both in the absence (-Lon) and the presence (+Lon) of Lon protease (Figure 5.2A): We found that "Tolerant" and "Beneficial" mutations were more commonly found in residues with high %DFI scores, suggesting that more flexible residues are better able to accommodate mutations without negatively impacting protein function. In contrast, "Intolerant" and "Restricted" mutations were more commonly found in residues with lower %DFI scores, suggesting that more rigid residues are less able to tolerate mutations without negatively impacting protein function. The difference between "Tolerant" and "Intolerant" distributions are statistically significant (p=0.0002). As well as the difference in distributions of "Beneficial" and "Restricted" with a p value of 2e-05. The "Mixed" residues do not show a particular trend towards either rigid or flexible.

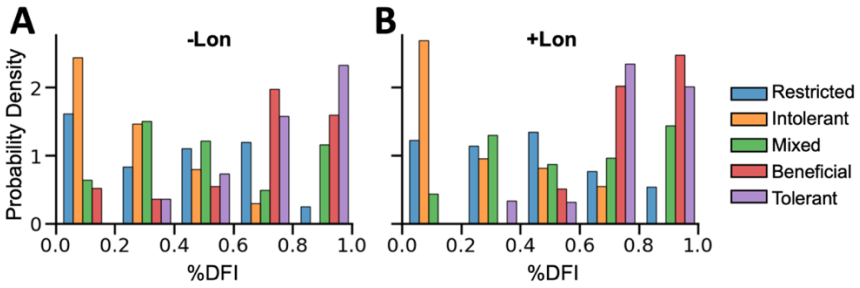


Figure 5.2: DFI score distributions for the five previously defined functional classes with and without the in the presence of Lon protease in Thompson et al. A) In the absence of Lon protease the “Intolerant” labeled residues almost always display very low DFI values, followed by the residues labeled as “Restricted” showing an overall rigid behavior (i.e., %DFI<0.6). Conversely, "Beneficial” and “Tolerant” residues are more commonly found in high DFI regions of the protein (i.e., % DFI≥0.6). The differences in these distributions are statistically significant, with p-values 0.0002 and 2e-05 respectively, calculated by Fisher's exact test using 0.6 %DFI as the threshold value. Residues labeled “Mixed” are distributed across different DFI ranges. B) In the presence of Lon protease, DFI scores of the residues distributed among functional classes are similar to those when the Lon protease is absent.

Moreover, to understand the relationship between DFI and protease sensitivity of residues, we investigated the distribution of DFI values for residues that are tolerant to Lon (i.e., a residue with "Beneficial" label in the absence of Lon, is still "Beneficial" in its presence), and those that are susceptible (i.e., a residue with "Beneficial" label in the absence of Lon, is "Restricted" in its presence) (Figure 5.3). The box plots shows that residues that are susceptible to the presence of Lon protease overall have a slightly higher DFI value compared to residues that are tolerant suggesting that enhanced flexibility, low rigidity might play a role with stability, as our earlier studies showed that rigid sites contribute to overall folding stability of a protein (Butler et al., 2015; Modi et al., 2021a). In summary, these results support the strength of the DFI metric in assessing the functional outcome of mutations on the protein, regardless of whether they proximal to or distal from functionally important loops.

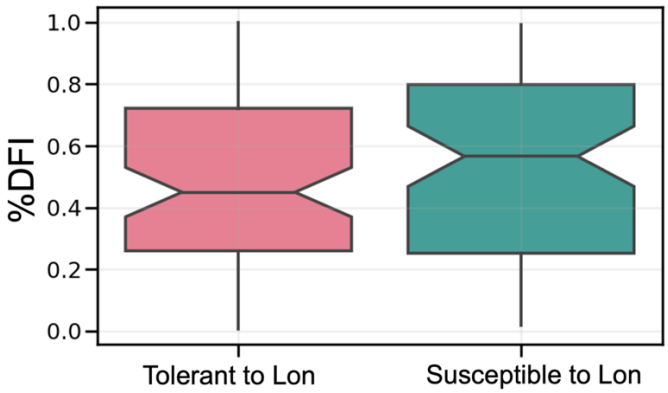


Figure 5.3: Box plot of DFI values for two sets of residues related to their protease sensitivity. Residues that are tolerant to Lon protease have slightly lower DFI scores compared to the one that are susceptible ($p < 0.23$). This shows that the susceptible residues have a higher degree of flexibility than the tolerant residues. This observation is interesting because it may suggest that the degree of flexibility of a residue play a role in its susceptibility to protease activity and its overall stability.

5.4.2 Asymmetry in Dynamic Coupling Reveals Allosteric Mutational Sites

Assessment of DHFR with our DFI metric also raised other important questions: First, how do residues distal to functionally important loops impact the overall enzymatic activity and second, what role do dynamic networks play in the control of DHFR function? Mutations found far from (but dynamically coupled to) functionally important loops, which we term "allosteric mutations", have previously been shown to substantially affect enzyme activity (Benkovic et al., 1988; Campitelli et al., 2021; Gekko et al., 2000; Thompson et al., 2020). The manner in which such this long-range dynamic communication propagates through proteins can be highlighted with our dynamic coupling metric DCI (Campitelli et al., 2020; Modi et al., 2021; Ose et al., 2022). A high DCI value indicates high dynamic coupling between residues *i* and *j*, suggests that strong communication between these residues exists. A low DCI score implies weak coupling between residues and suggests the absence of strong communication between them (Campitelli et al., 2021; Campitelli and Ozkan, 2020; Kolbaba-Kartchner et al., 2021; Larrimore et al., 2017; Modi and Ozkan, 2018).

We applied DCI analysis to the M20 and FG loops to explore how these loops affect protein activity. Dynamic coupling analyses reveal that the M20 and FG loops, despite being close to each other, exhibit different long-distance interactions with the rest of the protein (Figure 5.4A). We also discovered that each loop is dynamically coupled to specific regions within DHFR. For example, helix B, which spans residues 24 to 35, is more coupled to the M20 Loop (Figure 5.4A), while the FG loop is highly coupled to β sheets C and D and the helical "E region" (residues 57 to 85, Figure 5.4C).

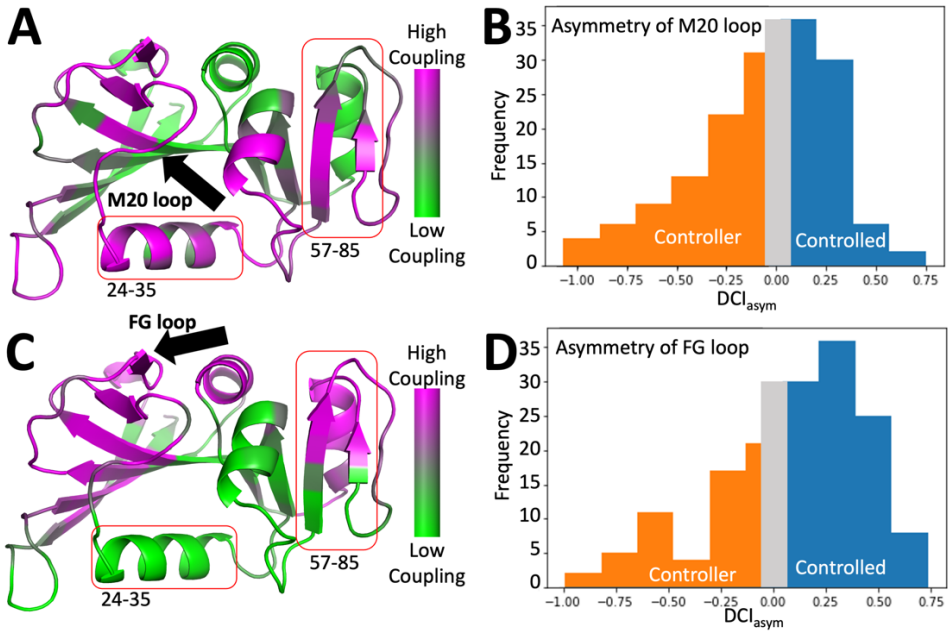


Figure 5.4: An analysis of DHFR using DCI_{asym}. Panels A and C show DCI profiles measuring the dynamic coupling of the M20 (A), and FG (C) loops projected onto the DHFR structure; high coupling is shown in purple, low coupling is shown in green. Panels B and D show the distribution of DCI_{asym} values of all residues calculated by targeting the M20 (B) and FG (D) loops. Positive DCI_{asym} values indicate that the residues within the loop control interactions with other residues while negative values represent residues that control the dynamics of the loop.

The complex network of the protein immediately suggested a disparity in dynamic coupling between positions that could be understood by an asymmetry in communication (Figure 5.4). Since each residue directly contacts a distinct set of neighboring residues, each position in a protein has a unique coupling network. Moreover, the dynamic coupling for position *i* with respect to *j* is not necessarily symmetric to the dynamic coupling of *j* to *i*. Thus, changes at position *i* may have larger effect on the flexibility of position *j* and vice versa. To capture this asymmetry, we created a novel metric DCI_{asym} (Campitelli et al., 2021). If the magnitude of difference between dynamic coupling scores of positions *i* to *j*

(DCI_{ji}) vs the coupling of j to i (DCI_{ij}) is significant, an asymmetry in communication between the two residues will exist. This asymmetry in communication can be informative on why certain amino acid substitutions at particular positions are more deleterious or beneficial to activity, vice versa.

To explore how dynamic coupling between the M20 loop and the B helix or the FG loop and the E helix and the C and D sheets affected enzyme function, dynamic coupling was considered in both directions using our DCI_{asym} metric. We first applied DCI_{asym} to the M20 loop. We consider DCI_{asym} values between -0.05 and 0.05 to suggest that both positions are symmetric in their coupling; in other words, neither position has dominance. Alternatively, a residue would be defined as an "M20 controlled position" when its average DCI_{asym} score (calculated by taking the average of all M20 loop positions; i.e., $\langle DCI_{asym} \rangle$) is positive and larger than 0.05, and as an "M20 controller position" when $\langle DCI_{asym} \rangle$ is negative and lower than -0.05 (Figure 5.4B). The same analysis is repeated on the FG loop (Figure 5.4D). The "controlled"/"controller" categorization of residues is then compared with average selection coefficient values from the work of Thompson et al (2020) (Figure 5.5). Selection coefficient values represent the impact of a mutation to DHFR activity relative to wild type. A mutation with a selection coefficient value around zero (±0.2) is considered as neutral. Values higher than 0.2 are beneficial to function, and conversely values lower than -0.2 are deleterious.

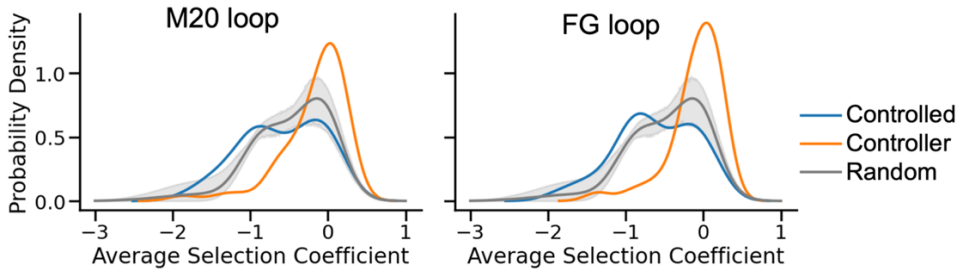


Figure 5.5: Analyses of the "controller" and "controlled" classified average selection coefficient value distributions (+Lon) for the M20 and FG loops. For both M20 and FG loops the average selection coefficient value distributions are different for "controller" and "controlled" labels. The residues with "controller" labels are commonly distributed near either neutral/enhanced (near zero, or positive) region while "controlled" residues display a distribution among negative values (deleterious) (M20 loop: p= 0.005, and FG loop: p= 0.008, Student's t-test). The gray distribution (line as the mean and shade as the variance) is generated by randomly selecting a different subset of residues (excluding "controller" / "controlled" residues) five times. Comparison of the randomly selected positions' average selection coefficient distributions with those of "controller" and "controlled", distributions of both M20 loop and FG loop shows that randomly selecting residues fails to capture the selection coefficient distribution of the "controller" residues (average p values over 5 random samples are 0.028, and 0.001, respectively) and "controlled" residues (p< 0.043 and <0.0425, respectively).

When investigated, the distribution of average selection coefficient values of "controlled" and "controller" residues displayed a different pattern (Figures 5.5 and 5.6A). When residues that control the M20 loop are considered, the peak of the distribution is observed to be above zero, which indicates that mutations at these positions have, on average, a positive impact on the activity. Conversely, sites "controlled" by residues in the M20 loop display a broad distribution with high density around very negative values. This suggests that mutations at these residues have a deleterious effect on protein activity. A similar trend is observed when the FG loop is targeted with DCI and DCI_{asym} analyses (Figure 5.5) (Figure 5.6B). Furthermore, comparison of the average selection coefficient distributions of "controlled" and "controller" residue positions with that of randomly

selected positions reveals the statistical significance of the distribution of our classifications in distinguishing the impact of mutations on activity (Figure 5.5).

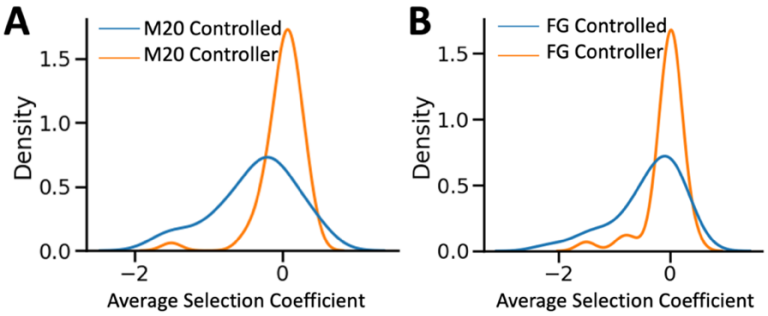


Figure 5.6: The asymmetry labeled average selection coefficient value (in the absence of Lon) distributions for the M20 and FG loops. A) Distribution for M20 loop plot shows "controller" and "controlled" labeled distributions are different (p= 0.005, Student's t-test). B) FG loop value distribution show "controller" and "controlled" labeled distributions are different. (p= 0.008, Student's t-test).

5.4.3 Beneficial Mutations are Enriched at Controller Sites

To further assess the impact of amino acid substitutions on residues with variety of control over the M20 and FG loops, we combined the "controlled" and "controller" designations from each loop. Namely, in this analysis, a residue was defined as a "controller" if it exerts control over both the M20 and FG loops simultaneously and is considered a "controlled" residue otherwise. The average selection coefficient value distributions of "controlled" and "controller" residues differ from each other when viewed in this way (Figure 5.7A). Controller residues generally present more activity-enhancing amino acid substitutions compared to residues in the "controlled" category. Interestingly, the peak of the distribution of "controlled" residue mutations is below the neutral range (near -1.0). This indicates that mutations to "controlled" positions more often yield deleterious outcomes with respect to function. On the other hand, mutation of "controller" positions could gradually modulate function both positively and negatively and could

therefore act like rheostatic switches (Campitelli et al., 2021). This ultimately suggests that the M20 and FG loops themselves are highly conserved due to functional constraints. However, those residues that control the loops can affect the overall enzyme function by distally altering functionally important residues.

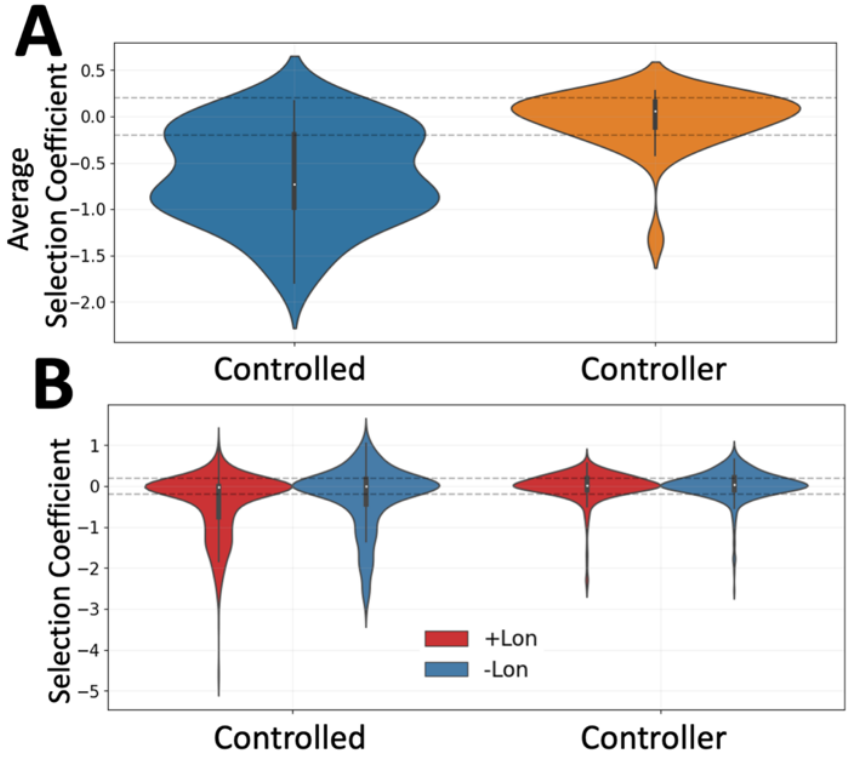


Figure 5.7: Experimentally measured selection coefficient values of "controller" and "controlled" residues of the M20 and FG loops. A) A violin plot of average selection coefficient values of the residues controlling both loops suggests that these residues have a peak on positive values compared to those residues that are controlled by the loops (p=3e-07). This suggests that mutations to residues that are controller of the M20, and FG loops can potentially enhance the activity of DHFR, while mutations to residues controlled by these loops are mostly deleterious. B) A violin plot generated using the selection coefficient of all amino acid substitutions per position. The distribution of selection coefficient values for "controller" residues falls primarily in the neutral to positive range. Alternatively, a broader distribution is observed for residues controlled by both loops; mutations at these residues often have a drastic negative impact on activity.

To remove any bias that arose from averaging, we also obtained the distributions using selection coefficient values for every mutation (as opposed to average values for all mutational outcomes per position). When all selection coefficient values are considered, the differences in asymmetry between “controlled” and “controller” residues is more pronounced (Figure 5.7B). Additionally, when other functionally important sites, GH loop (spanning through residues 142 to 149) and Adenosine Binding Domain (residues 63, 64, and 65) are investigated, the results are similar to those found with M20 and FG loops (Figure 5.8). This striking difference in the distribution of functional outcomes of mutations on “controlled” versus “controller” residues illustrate the importance of dynamic allosteric control (McCormick et al., 2021; van den Bedem et al., 2013). Previously, we explored asymmetry in dynamic coupling by analyzing 591 pathogenic missense variants in 144 human enzymes (Ose et al., 2022). We showed that many mutations, far from the active site, exhibit deleterious behavior (sometimes leading to pathogenicity) due to their high coupling with the active site. Furthermore, we also observed that these mutations are coupled to the active site, but the coupling strength (DCI score) of the mutation sites back to active site is not as high, showcasing an indifference in coupling strength (asymmetry). The "controller" and "controlled" classification developed in this present work highlights the importance of dynamic coupling to active sites, in agreement with previous study. In addition, it highlights the degree to which asymmetry in this coupling can modulate function in a positive or negative direction.

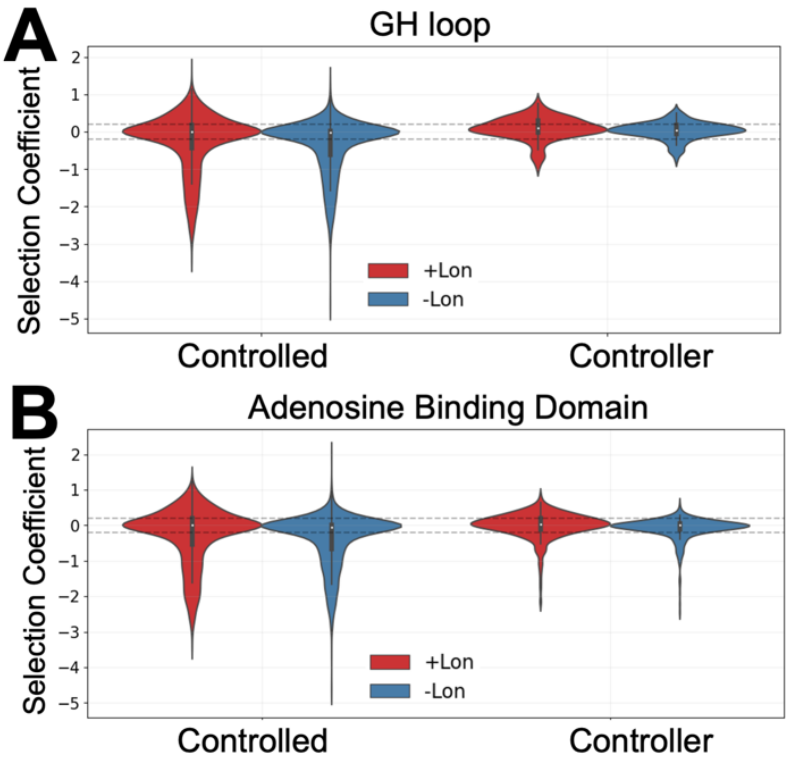


Figure 5.8: Violin plots of experimentally measured selection coefficient values of "controller" and "controlled" residues of the A) GH loop and B) Adenosine Binding Domain. The distribution of selection coefficient values for "controller" residues follows an overall neutral trend. On the other hand, "controlled" residues show a diverse distribution spreading to negative (deleterious) ranges. The difference observed in the distributions of "controlled" and "controller" are statistically significant, with p values 0.003 and 3e-07, for GH loop and Adenosine Binding Domain, respectively.

5.4.4 Leveraging Asymmetry in Dynamic Coupling for Fine-Tuning Function: A Comparative Analysis of Other Metrics and Functional Outcomes

To evaluate the effectiveness of the DCI_{asym} metric in identifying residue positions with diverse activities upon mutations, we compared it with several other metrics that are commonly used to identify functionally critical sites such as solvent-accessible surface area (SASA) (Butler et al., 2015; Cao et al., 2019; Chan and Dill, 1990; Wei et al., 2013), average # of contacts as well as network metrics including betweenness, closeness, and eigenvector centrality values (Figure 5.9) (Chakrabarty and Parekh, 2016; del Sol et al.,

2006; del Sol and O'Meara, 2005) (See methods). After computing these metrics for each residue, we grouped positions sharing similar values using histograms and analyzed the average and variance of the experimental average selection coefficients of the positions residing in each bin/group. The average of SASA values in each bin correlates with average experimental values (R=0.88), indicating that highly accessible residues are more likely to have neutral outcomes upon mutation. The average number (#) of contacts shows a correlation of -0.84 with experimental fitness, but the deviation at medium ranges suggests that, on average, most of these residues are deleterious to function when mutated. The betweenness scores show a relatively low correlation with the experimental values (R=-0.65). Despite its strong correlation with fitness (R=-0.85), the observed average negative fitness values with all eigenvector centrality ranges suggest that underlying factors beyond altered functional outcomes upon substitution are not fully captured by this metric. The closeness measure identifies residue positions with high experimental fitness scores (>0.2) but fails largely to identify residues with deleterious behavior. In contrast, when we analyzed dynamic-based metrics (e.g., DCI_{asym}) of the M20 and FG loops in the same manner, the results show that the DCI_{asym} metric is the most effective in capturing the trend of changing fitness for both the M20 loop and FG loop, with high correlations of -0.98 and -0.97, respectively. This indicates that as a position becomes more controlled, mutations at that site are more deleterious; alternatively, "controller" residues yield more neutral or beneficial outcomes when mutated. These findings suggest that the "controlled" and "controller" classification based on the DCI_{asym} metric can not only provide high accuracy

in identifying and characterizing residues, put can also help identify “controller” sites that may subtly tune the enzyme’s function when mutated.

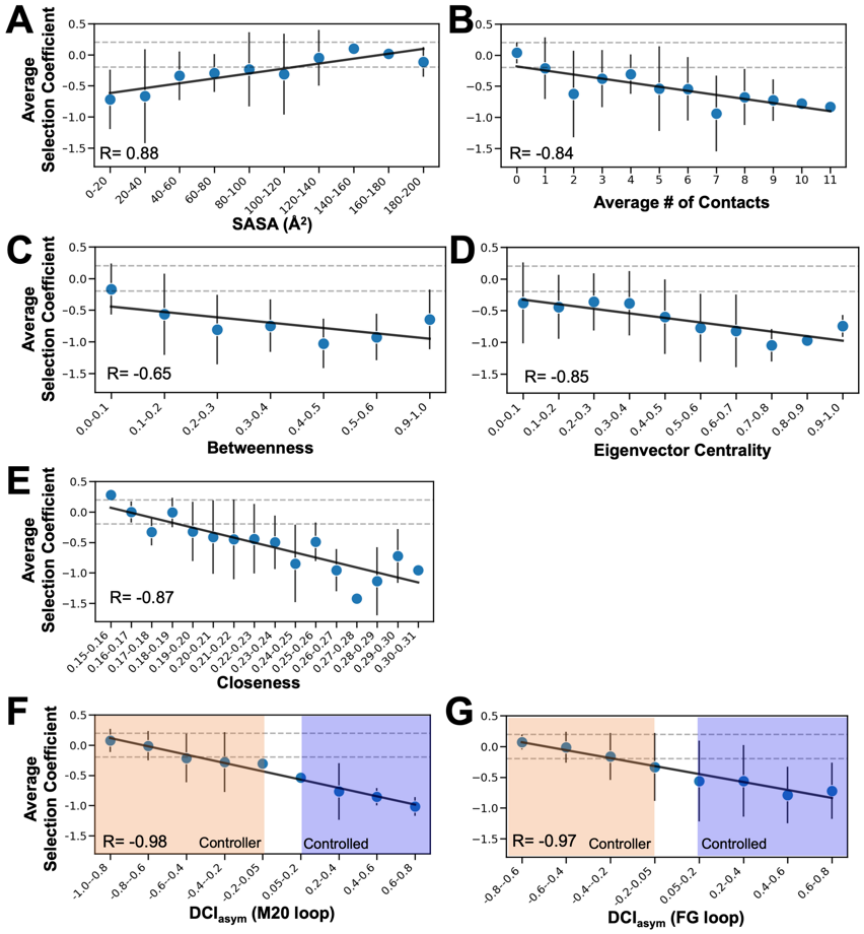


Figure 5.9: Correlation plots of binned structural and dynamic features with average selection coefficients A) Structural features SASA and B) average number (#) of contacts are binned and compared with experimental data. Both SASA and average # of contacts values of residues correlate well with the experimental data. Structural metrics betweenness, closeness, and eigenvector centrality are compared with average selection coefficient. C) Betweenness metric binned every 0.1 range shows that ranges from zero to 0.2 and 0.9 to 1.0 have higher fitness values relative to others (R=-0.65). D) Eigenvector centrality metric is binned every 0.1 range. The eigenvector centrality metric overall shows a good correlation, but all the bins have an experimental value lower than the neutral range (R=-0.85). E) Closeness metric binned every 0.01 range shows that residues with values from 0.15 to 0.19 shows great promise in enhancing the activity (R=-0.87). F) M20 loop and G) FG loop DCI_{asym} value binned every 0.2 window shows that DCI_{asym} values lower than zero yield higher activity compared to those positions in the positive ranged bins. This correlation fits well with the definitions of “controlled” and “controller”.

5.4.5 Examining the Interplay of Asymmetry in Dynamic Coupling and Evolutionary Conservation

To gain insight into how our "controller"/"controlled" categorization relates to a position's conservation, we utilized the ConSurf database (Ben Chorin et al., 2020; Goldenberg et al., 2009) to evaluate conservation of each site. Namely, the distribution of conservation of each residue in DHFR was considered with respect to its asymmetry categorization (Figure 5.10). Previous studies have shown that positions of M20 and FG loops are structurally and evolutionarily conserved (Bhabha et al., 2013; Thorpe and Brooks III, 2004; Weinreich et al., 2006). When the conservation of residues controlled by the M20 and FG loops was investigated, the "controlled" residues were found to be highly conserved, which agrees with our analysis. This suggests that mutations of "controlled" residues often yield deleterious outcomes and are therefore filtered out by natural selection. Alternatively, "controller" sites are found to be non-conserved, indicating that they can accommodate a diverse number of mutations. This behavior is observed in the deep mutational scanning data (Thompson et al., 2020) where mutations on "controller" residues enable enhancement or modular changes in the activity of DHFR. Conservation analyses showed that "controlled" residues are highly conserved, while "controller" sites are non-conserved, allowing for a diverse number of mutations and enabling enhancement or modular changes in the activity of DHFR. Indeed, these results agree with the deep sequencing data where the mutations on controlled sites are usually deleterious; therefore, they are also eliminated during evolution.

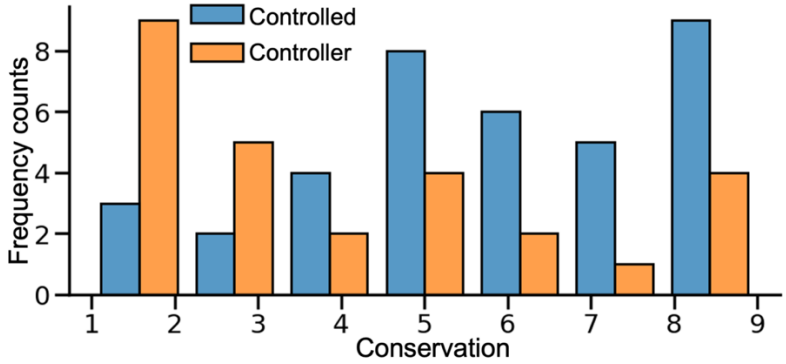


Figure 5.10: Conservation distribution of DHFR positions designated either controlled by or controllers of the M20 and FG loops. Conservation values are obtained using ConSurf database (Ben Chorin et al., 2020; Goldenberg et al., 2009). Residues that are "controller" attain lower values (non-conserved) compared to "controlled" residues which are more distributed on higher (conserved) values. The Student's t-test showed that the difference in distribution was statistically significant (p= 0.003).

To gain a deeper insight into why a distinction between "controlled" and "controller" residue conservation exists, we examined the flexibility of these positions; our previous studies highlighted a strong correlation between conservation and flexibility (Butler et al., 2018; Campitelli et al., 2021; Kazan et al., 2022; Modi et al., 2021a; Ose et al., 2022). The analyses demonstrate that "controlled" residues are often highly rigid with an average %DFI score of 0.2 (Figure 5.11). Mutations occurring at these rigid sites typically have detrimental effects on function. In contrast, the "controller" residues exhibit a higher average %DFI value of 0.8 (Figure 5.11), indicating high flexibility. This flexibility enables these positions to tolerate a broader range of amino acid changes. Consequently, selectively targeting "controller" residues holds promise for random mutagenesis or rational design approaches aiming to finely adjust the activity of DHFR. We believe our dynamics metrics DFI and DCI_{asym} could uncover these positions in other proteins as well.

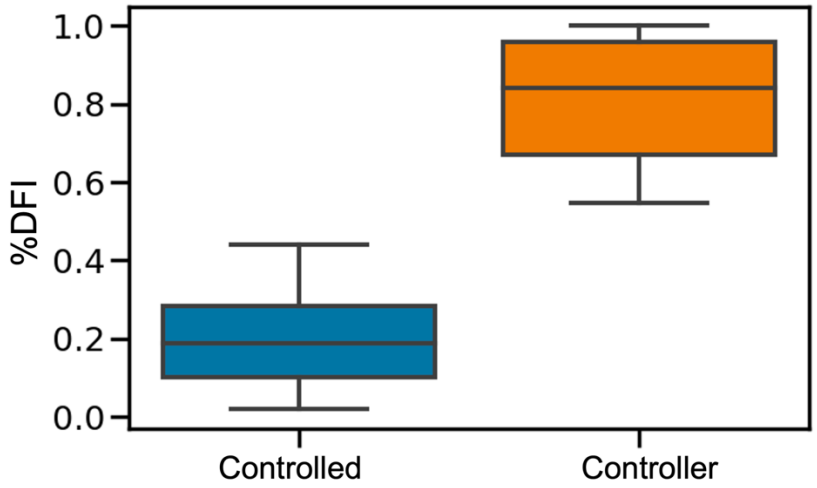


Figure 5.11: A box plot showing %DFI distribution of "controlled" and "controller" residues. These distributions show that "controller" sites attain high %DFI values on average; conversely "controlled" positions are generally found to be rigid.

5.5 Conclusion

In this study, we show that dynamic based metrics can be utilized to better understand the functional outcomes of mutations in DHFR. DFI scores displayed great promise in differentiating positions that might lead to beneficial or deleterious functional changes. The DCI metric revealed that the long-distance dynamic coupling between the M20 loop and other residues in DHFR differs significantly differs from that of FG loop. These diverse allosteric features are further investigated with our novel DCI_{asym} metric. The observed differences between residues that are controlled by or control two important loops in DHFR highlight how mutation of "controller" residues can fine tune enzyme activity through dynamic allostery. In addition, the evaluation of evolutionary conservation of "controlled" versus "controller" positions indicated that the "controller" sites are more amenable to mutations. On the other hand, "controlled" sites are more conserved since mutations to these sites often results in loss of function. Although our study was carried out using

DHFR, the conclusions drawn in this work display great promise for using dynamics metrics to gain a better understanding of how residues distal from functional portions of proteins can potentially modulate protein activity without compromising the fold.

5.6 Acknowledgement

This research was supported by the Gordon and Betty Moore Foundations and National Science Foundation (1901709).

CHAPTER 6

THE ROLE OF RIGID RESIDUES IN MODULATING TEM-1 β-LACTAMASE FUNCTION AND THERMOSTABILITY

This chapter is adapted from: "Kolbaba-Kartchner, B.; Kazan, I.C.; Mills, J.H.; Ozkan, S.B. (2021) The Role of Rigid Residues in Modulating TEM-1 β-Lactamase Function and Thermostability. Int. J. Mol. Sci. 22, 2895."

I Can Kazan shared first-authorship with Bethany Kolbaba-Kartchner. I Can Kazan conducted all computational work related to DFI, DCI, and MD simulations presented here while Bethany Kolbaba-Kartchner performed Rosetta design method and executed experimental characterization.

Previous chapters were focused on elucidating the intricate interplay between protein-ligand interactions, mutations, and the influential factors that drive protein dynamics, ultimately resulting in alterations in protein activity or function. Building upon this foundation, here, I explore a different enzyme system, β-lactamase, leveraging the acquired knowledge to further tackle the challenge of protein design. Bacteria produces β-lactamase to provide a defense mechanism against β-lactam antibiotics. Therefore, it is essential to investigate this enzyme to design more effective antibiotics for the treatment of infections. With this goal in mind, TEM-1 β-lactamase is used as a model system to design novel variants that are similar to a promiscuous and more stable ancestor of TEM-1 β-lactamase.

I applied DFI and DCI to initially uncover residues that could either potentially impact the activity upon mutations (rigid residues) or maintain neutrality (flexible residues). We modeled mutations around these rigid and flexible residues to create new designs using RosettaDesign. I also performed MD simulations of the Rosetta Models. To evaluate the functional outcomes, I developed a novel approach "dynamic distance analysis" (DDA) which quantifies the dynamic similarities (DFI profile similarity form MD) of proteins to each other and highlights the variants with dynamics profiles closest to that of target protein. The results, which were experimentally validated by our collaborators, showed that the integration of MD driven dynamics design holds great promise in creating variants capable of effectively modulating the activity and stability of enzymes across a wide range.

6.1 Abstract

The relationship between protein motions (i.e., dynamics) and enzymatic function has begun to be explored in β-lactamases as a way to advance our understanding of these proteins. In a recent study, we analyzed the dynamic profiles of TEM-1 (a ubiquitous class A β-lactamase) and several ancestrally reconstructed homologues. A chief finding of this work was that rigid residues that were allosterically coupled to the active site appeared to have profound effects on enzyme function, even when separated from the active site by many angstroms. Here, our aim was to further explore the implications of protein dynamics on β-lactamase function by altering the dynamic profile of TEM-1 using computational protein design methods. The Rosetta software suite was used to mutate amino acids surrounding either rigid residues that are highly coupled to the active site or to flexible

residues with no apparent communication with the active site. Experimental characterization of ten designed proteins indicated that alteration of residues surrounding rigid, highly coupled residues, substantially affected both enzymatic activity and stability; in contrast, native-like activities and stabilities were maintained when flexible, uncoupled residues, were targeted. Our results provide additional insight into the structure-function relationship present in the TEM family of β-lactamases. Furthermore, the integration of computational protein design methods with analyses of protein dynamics represents a general approach that could be used to extend our understanding of the relationship between dynamics and function in other enzyme classes.

6.2 Introduction

Since the 1940s, β-lactam antibiotics, which target a key enzyme in bacterial cell wall biosynthesis, have been the antimicrobial weapon of choice in the war against bacterial infection (Coulson, 1985). The widespread use of β-lactams is likely a consequence of the fact that they are inexpensive to produce and have historically been effective in treating most infections. However, as the use of this class of antibiotics became more widespread, so too did the prevalence of β-lactamase enzymes, which hydrolyze the β-lactam ring and render the antibiotic nonfunctional (Coulson, 1985). Additionally, as new β-lactam antibiotics enter into clinical use, the remarkable adaptivity of β-lactamases complicates efforts to develop novel antibiotics that are resistant to degradation by this class of enzyme (Bush, 2018). The TEM family of β-lactamases has been thoroughly studied to gain insight into the manner in which resistance is achieved (Brandt et al., 2017; Brown et al., 2020; Cortina et al., 2018; Cortina and Kasson, 2018; Gobeil et al., 2019). Despite these efforts,

we currently possess an incomplete understanding of the relationship between sequence and function in this enzyme class. A major challenge is that several mutations have been identified that have a significant influence on function, but which are highly distal from the enzyme active site (Singh and Dominy, 2012). In addition, even single point mutations (e.g., the well-characterized, M182T substitution), which have minimal effects on enzymatic function can drastically affect the protein's thermostability (Orencia et al., 2001; X. Wang et al., 2002). Our inability to rationalize the manner in which these thoroughly studied mutations alter enzyme function is suggestive of an incomplete understanding of the sequence-function relationships present in β-lactamases. This in turn limits our ability to develop novel classes of antibiotics that are not substrates for these enzymes (Fair and Tor, 2014).

As explained in Chapter 1, a possible explanation as to how mutations distal to the active site can still exert influence at a great distance is that they serve to reshape the inherent dynamics of the enzyme (Campitelli et al., 2020; Doucet et al., 2007; Gerek et al., 2009; Gerek and Ozkan, 2011; Kim et al., 2015; Larrimore et al., 2017; Modi et al., 2018; Modi and Ozkan, 2018; Zou et al., 2015). In a recent study, we explored this hypothesis in the TEM-1 β-lactamase using two in silico, dynamics-based metrics: the dynamic flexibility index (dfi) (Gerek and Ozkan, 2011; Kumar et al., 2015a), which measures the mobility of each residue, and the dynamic coupling index (dci) (Campitelli et al., 2018; Larrimore et al., 2017), which assesses the coupling between distant residues (Modi and Ozkan, 2018). Using these two metrics, we characterized TEM-1 and a set of ancestrally reconstructed TEM-1 variants that possess vastly distinct physical properties (i.e.,

thermostabilities) and functions (i.e., substrate specificity) despite having almost identical conformations (Risso et al., 2013; Salverda et al., 2010; Stiffler et al., 2015; Zou et al., 2015). A major finding of our previous study was that TEM-1 and its ancestral homologues possessed distinct dynamic profiles and that these differences in dynamics appeared to have profound effects on enzyme function. Namely, rigid residues that are distal from, but highly coupled to, residues in the active site appeared to have substantial effects on protein function (Campitelli et al., 2020, 2018; Li et al., 2015; Modi and Ozkan, 2018). One intriguing hypothesis that might explain these data is that rigid residues can serve as "hubs" of dynamic communication. This notion has also been validated in the context of disease-causing mutations in other proteins, in which mutations to rigid residues that are far from the active site are functionally deleterious (Campitelli et al., 2020; Gerek et al., 2013; Kumar et al., 2015b; Modi and Ozkan, 2018).

More recently, we used both dfi and dci to analyze members of the TEM family that either arose in the clinic or were generated via directed evolution (Modi and Ozkan, 2018). In this study, we observed that mutations known to confer resistance to non-native substrates (1) often occur at particularly rigid residues as judged by our dfi metric and (2) appear to allosterically modify the flexibility of catalytic residues within the active site as suggested by our dci metric (Modi and Ozkan, 2018). Collectively, these studies support the hypothesis that rigid residues are of particular importance to the overall dynamics of proteins and may have a substantial impact on protein function if they are allosterically coupled to the active site. If our hypothesis is correct, mutations that alter the identity of allosteric rigid residues (or those in their vicinity) could have substantial effects on enzyme

activity; however, the ability to thoroughly explore this hypothesis is challenging. Although extensive datasets comprised of clinically derived TEM family variants additional variants generated via directed evolution (Stiffler et al., 2015) exist, the serendipitous identification of proteins with multiple mutations in the vicinity of known rigid residues would be unlikely. One potential solution is to use computational protein design methods to specifically target mutations to regions of interest. A major benefit of this approach is the ability to "pre-screen" each combination of mutations in silico to exclude variants in which protein folding is not predicted to be energetically favorable.

In this work, computational protein design methods were used to alter the environments surrounding two residues that were identified as being rigid and highly coupled to the active site despite being separated from it by a great distance. Dynamic profiles of each designed protein (hereafter referred to as a "design") were then generated and compared to that of an ancestrally reconstructed variant of TEM-1 (the "Gram-negative common ancestor" or GNCA), which possesses increased thermostability, but reduced activity against ampicillin relative to wild type TEM-1 (Risso et al., 2013). Principal component analysis (PCA) was used to identify designs with dynamic profiles that were predicted to be more similar to GNCA than extant TEM-1, and five designs were characterized in the laboratory. All designs exhibited reduced activity against ampicillin relative to TEM-1, but an increase in thermostability was also observed. Reduced activity against ampicillin and increased thermostability relative to TEM-1 are both features of GNCA. Alternatively, when identical design protocols were applied to flexible residues that were not coupled to the active site, native-like catalytic abilities and thermostabilities were maintained. Finally,

in an effort to further link dynamics to enzyme function, we developed a novel analytical approach termed the “dynamic distance analysis” (dda) that was applied retrospectively to our experimentally characterized proteins. The dda analysis appeared to capture functional differences between our designed proteins and could be a useful tool for dynamic profile analysis in future studies. Collectively, our results serve to further highlight the importance of allosteric rigid residues in regulating the dynamics of the TEM-1 β-lactamase.

6.3 Computational Protein Design Methods Used for The Implications of Protein Dynamics on B-Lactamase Function

6.3.1 Molecular Dynamics (MD)

The AMBER software package was utilized for simulating all β-lactamases in this study. Each system was parameterized with the ff14SB force field and the explicit water model TIP3P (Maier et al., 2015; Salomon-Ferrer et al., 2013b). The solvation box was assigned as 16 Å. The system was neutralized by sodium and chloride ions and minimized for 11,000 steps using the steepest descent algorithm. Isothermal, isobaric, and constant number of particles ensemble production trajectories were performed at 300K and 1 bar pressure. For each production, a 1 μs simulation was conducted. The residue covariances were calculated using a 50 ns length window shifted by 10 ns (example: 1–50 ns, 10–60 ns, etc.) over the course of the trajectories.

6.3.2 Dynamic Flexibility Index (dfi)

The dfi metric (Gerek and Ozkan, 2011; Kumar et al., 2015a; Modi and Ozkan, 2018) calculates the relative flexibility/rigidity of a residue in a protein by incorporating the residue covariances. The protein can be modeled with the Elastic Network Model (ENM)

in which harmonic springs connect Cαs (Atilgan et al., 2010). Taking the second derivatives of the potential forms a Hessian matrix, *H* Equation (6.1). The inverse of the Hessian matrix is proportional to the covariance matrix. The models based on ENM cannot capture changes in the dynamics of the designed variants based on Cα positions alone. Therefore, we substituted the inverse of the Hessian with the covariance matrices from MD trajectories to capture the effect of mutations on the protein conformations. The covariance matrix, *G*, contains the residue covariances, obtained by the MD trajectories Equations (6.2) and (6.3) (Campitelli et al., 2020; Kumar et al., 2015b; Larrimore et al., 2017; Modi and Ozkan, 2018; Gerek et al., 2013).

$$[\Delta \mathrm{R}]_{3Nx1} = [H]^{-1}_{3Nx3N}[\mathrm{F}]_{3Nx1} \tag{6.1}$$

$$[\Delta \mathrm{R}]_{3Nx1} = [G]_{3Nx3N}[\mathrm{F}]_{3Nx1} \tag{6.2}$$

$$df i_i = \frac{\sum_{j=1}^{N} [\Delta R^j]_i}{\sum_{i=1}^{N}\sum_{j=1}^{N} [\Delta R^j]_i} \tag{6.3}$$

The residue response vector ($\Delta\mathbf{R}$) is the resultant vector containing the fluctuation responses from multiplying the covariance matrix with the force vector, **F**. $|\Delta \mathbf{R}^j|_i$ denotes the magnitude of the residue response fluctuation vector of position i, when j is exposed to a random force vector.

6.3.3 Dynamic Coupling Index (dci)

The dynamic coupling index (dci) (Campitelli et al., 2020; Larrimore et al., 2017; Modi and Ozkan, 2018) measures the degree of dynamic coupling between two residues. Namely, it captures the strength of displacement of a residue *i* upon perturbation of a

distinct residue *j*, relative to the average fluctuation response of position *i* when all of the positions within a structure are perturbed. Generally, this metric is used to establish the communication between a functionally important residue and other residues within the protein that are many angstroms away. The dynamic coupling index of a given residue i is calculated using the equation below Equation (6.4):

$$dci_i = \frac{\sum_{j=1}^{N_{Functional}} \left|\Delta \boldsymbol{R}^j\right|_i / N_{Functional}}{\sum_{j=1}^{N} |\Delta \boldsymbol{R}^j|_i \ / \ N} \tag{6.4}$$

where $\left|\Delta \mathbf{R}^j\right|_i$ corresponds to the magnitude of the residue response vector ($\Delta\mathbf{R}$) for residue *i* when residue *j* is perturbed. The *dci* score thus provides information on the allosteric behavior of a location associated with active site dynamics. A high *dci* value implies strong coupling between active sites, inversely, a low scoring position is regarded as weakly coupled to the active site (Campitelli et al., 2020; Larrimore et al., 2017; Modi and Ozkan, 2018).

6.3.4 Dynamic Distance Calculation

Principal Component Analysis (PCA) was used to compare and cluster the flexibility profiles of the designed TEM-1 variants with respect to TEM-1 and GNCA. However, because the output of a PCA is dependent on the input data, the calculated distances between any designed protein and TEM-1 or GNCA can change with the inclusion of new or distinct data points (e.g., a different set of designed proteins). To account for this, we employed an iterative, random sampling approach to capture the relative distance of a designed protein from TEM-1 and from GNCA (Figure B.1).

For every designed TEM-1 variant, a dataset containing the target design, TEM-1, GNCA and an additional seven randomly chosen designs was constructed and used to generate a PCA. Namely, the dfi profiles of these ten proteins were merged into a matrix, X, of dimension Equation (6.5):

$$(m \text{ x } n) \tag{6.5}$$

Here, *m* is the total number of datasets that are clustered together, which each have *n* number of attributes (*n* = total number of residues). Singular value decomposition of X was then carried out as follows Equation (6.6):

$$[X]_{mxn} = [U]_{mxm}[\Sigma]_{mxn}[V]_{nxn.} \tag{6.6}$$

Here, *U* and *V* are unitary matrices with orthonormal columns and are called left singular vectors and right singular vectors, respectively, and Σ is a diagonal matrix with diagonal elements known as singular values of *X*.

The singular values of *X*, by convention, were arranged in a decreasing order of their magnitude, $\sigma = \{\sigma_i\}$ representing the variances in the corresponding left and right singular vectors. The set of the highest singular values (representing the largest variance in the orthonormal singular vectors) can be interpreted to show the characteristics in the data *X* and the right singular vectors create orthonormal basis which spans the vector space representing the data. The left singular vectors contain weights indicating the significance of each attribute in the dataset as Equation (6.7):

$$w_i = \sum_{k=1}^{r} \sigma_k |u_{ik}| \tag{6.7}$$

Using these features of the decomposed singular vectors, we created another matrix, X^* using only the highest three singular values which mimics the basic characteristics of the original dataset. It can be represented as Equation (6.8):

$$[X^*]_{m\times r} = [V^*]_{m\times r}[\Sigma^*]_{r\times r} \tag{6.8}$$

Here, Σ* contains only the largest 3 singular values and V^* contains the corresponding right singular vectors. The data were then clustered hierarchically based on the pairwise distance between different proteins in the reconstructed dfi data with reduced dimensions. The distance between designed protein, j_1, and TEM-1, j_2, was computed in the reduced dimension using three principal components Equation (6.9):

$$d_{12} = \sqrt{\sum_{i=1}^{3}\left(X^{*}{}_i^{j1} - X^{*}{}_i^{j2}\right)^2} \tag{6.9}$$

We also calculated the distance between each designed TEM-1 variant and GNCA to measure the similarity in their flexibility profiles. The random selection of dataset was repeated a thousand times to create a diverse distance distribution and we called this distance profile analysis *dynamic distance analysis* (dda). The distributions were fit to a Gaussian mixture model with a Dirichlet prior to estimate the density and the mean of the dynamic distances (Bishop, 2006). The distributions and the mean distances were utilized for selecting the designed proteins that cluster close to GNCA and far from TEM-1 (Figure B.1).

6.3.5 Rosetta Design Protocol

A high-resolution (1.8 Å) structure of TEM-1 (PDB ID: 1btl) was processed to remove waters, non-proteinogenic molecules and a second copy of the protein in the asymmetric

unit. The resulting structure was subjected to an energy minimization using the Rosetta relax protocol; detailed descriptions of all computational protocols used in this study can be found in the Appendix B.

The relaxed 1btl structure was used as an input to the DesignAround protocol within Rosetta using the ref15 score function. This algorithm first identifies spheres with user-defined radii around a defined residue. Residues within these "design spheres" were subjected to in silico mutagenesis, conformational sampling and backbone minimization.

6.4 Results and Discussion

6.4.1 Computational Analysis Using dfi and dci

Our efforts to better understand the relationship between protein dynamics and function began by identifying a TEM-1 variant that could serve as a basis of comparison to the wild type protein. Recently, the putative sequences of ancestral TEM-1 were predicted using Bayesian bioinformatics (Risso et al., 2013). Three ancestral TEM family homologues (the Gram-negative and Gram-positive common ancestor, PNCA; the Gram-negative common ancestor, GNCA, and enterobacteria common ancestor, ENCA) were observed to possess distinct physical and biochemical properties when characterized in the laboratory (Risso et al., 2013). This is likely a consequence of the fact that these proteins are thought to have existed at different times in the evolutionary history of this enzyme (Risso et al., 2013). We chose to focus our efforts on the ancestral homologue GNCA because its properties differ more substantially from TEM-1 than the other variants. Despite sharing > 50% identical residues (Figure 6.1A), nearly identical folds (1.3 Å root-mean-square deviation

(RMSD) over all Cαs, (Figure 6.1B), and conserved catalytic residues (Figure 6.1C), GNCA unfolds at a temperature that is ~35 °C higher than wild type TEM-1.

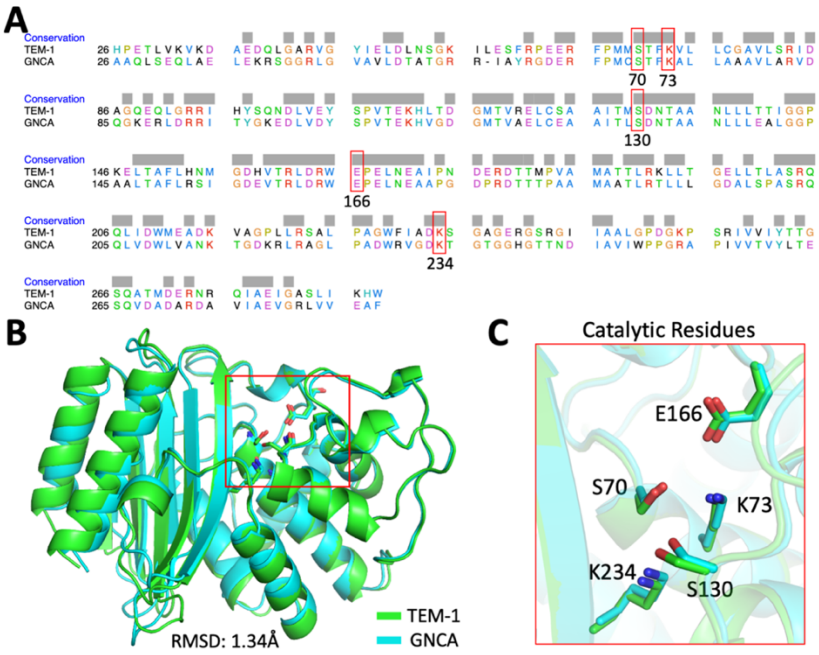


Figure 6.1: Differences in sequence and structure between TEM-1 and its ancestral variant GNCA. A) Sequence alignment (Ambler numbering) of TEM-1 and GNCA shows a 54% sequence identity; conserved active site residues are highlighted in red boxes. B) The crystal structures of TEM-1 (PDB ID: 1btl), green and GNCA (PDB ID: 4b88), cyan are superimposed and the catalytic residues are shown as sticks within a red box. The low root-mean-square-deviation (RMSD) indicates a high conservation of structure. C) Active site residues in TEM-1 and GNCA are shown in green and blue sticks, respectively.

Furthermore, GNCA appears to be a "substrate generalist" in that it possesses measurable (but reduced) activity against penam antibiotics (e.g., penicillin) relative to TEM-1, while simultaneously possessing a far greater ability to degrade cepham antibiotics (e.g., cefotaxime) (Risso et al., 2013).

It is difficult to rationalize the substantial differences in function and stabilities that are observed in GNCA and TEM-1 in light of the high sequence identity and structural similarities that exist for these proteins. Previous studies in our laboratory (Modi and

Ozkan, 2018; Zou et al., 2015) suggested that the inherent dynamics of both TEM-1 and GNCA might play a role in regulating their functions. To further explore this, we analyzed the dynamic profiles of both proteins using two metrics developed in our group: The Dynamic Flexibility Index (dfi) and the Dynamic Coupling Index (dci). The dfi method (Butler et al., 2015; Gerek et al., 2013; Kumar et al., 2015b) is based on Linear Response Theory and Perturbation Response Scanning (Atilgan et al., 2010) and calculates the resilience of a given residue to random force perturbations applied to other residues in the protein. A given amino acids dfi value is therefore related to the relative conformational entropy (i.e., flexibility) of that residue with respect to the rest of the protein. A high dfi value indicates high flexibility; conversely, a low dfi value indicates rigidity. The dci metric (Larrimore et al., 2017; Modi and Ozkan, 2018) is derived from the same theoretical origin as dfi and is used to quantify the degree to which two residues are dynamically coupled in terms of correlated motions. A high dci value between a pair of residues that do not interact directly indicates allosteric coupling and suggests that a perturbation to one residue will be transmitted to the other even over long distances. A low dci score implies a weak coupling between a residue pair, and no strong communication channel between them is expected.

When we applied the dfi and dci analyses to extant TEM-1 and a set of reconstructed ancestral homologues including GNCA (Modi and Ozkan, 2018; Zou et al., 2015), our analyses indicated that rigid residues (i.e., those with low dfi scores) that are highly coupled to the active site can contribute substantially to protein function. In this study, we hoped to

further explore the importance of rigid residues to protein function by altering the identity of amino acids in their vicinity.

We selected two residues in TEM-1 (V44 and V262) as targets for our study. Not only do both residues have low dfi scores (%dfi value < 0.2) (Figure 6.2A), but they are highly coupled to the active site (%dci > 0.7) (Figure 6.2B). These two residues were of particular interest to us because they are over 10 Å away from the active site and are located on adjacent β-strands with side chains facing opposite domains. We also identified three distal, flexible residues in TEM-1 (K55, P226, and K256) with high dfi scores (%dfi > 0.8) (Figure 6.2A) and low coupling to active site residues as evaluated by the dci metric (%dci < 0.4) (Figure 6.2B) and over 10 Å away from the active site to serve as controls. Alteration of the protein environments surrounding allosteric rigid residues would be expected to substantially modify protein function if our hypothesis is correct. Alternatively, modification of amino acids surrounding flexible residues with low dynamic coupling to the active site would be expected to result in proteins with native-like functions. All of the allosteric rigid and uncoupled flexible residues we targeted for design are over 10 Å from the nearest catalytic residue, which suggests that mutations in their vicinities should only have an indirect effect on the active site unless other factors (e.g., dynamic coupling) are at play.

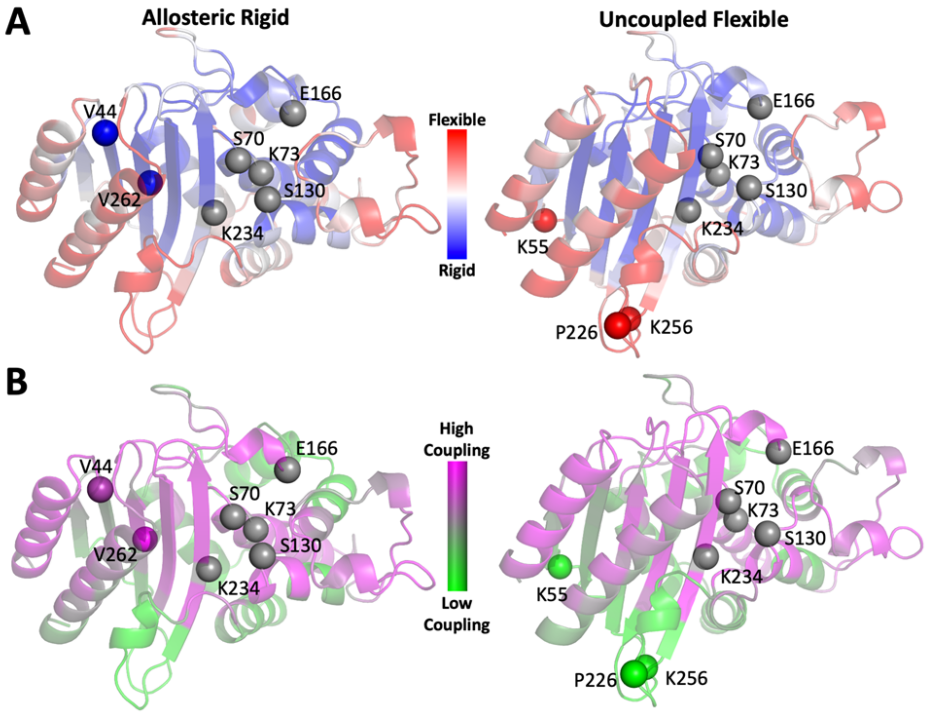


Figure 6.2: The *dfi* (panel A) and *dci* (panel B) values of each residue in TEM-1 are calculated and mapped onto the structure of TEM-1, which is shown as color coded cartoons. Catalytic residues are shown as grey spheres. Rigid and flexible residues used in this study are shown as spheres that are colored by either their *dfi* (panel A) or *dci* (panel B) score. Allosteric rigid residues, V44 and V262, have low *dfi* scores and high allosteric dynamic coupling with the active site residues. Residues K55, P226, and K256 are both highly flexible and exhibit low allosteric dynamic coupling to the active site.

6.4.2 Computational Design of TEM-1 Variants

In order to alter the amino acid compositions surrounding both the rigid and flexible residue positions, we used the Rosetta computational protein design suite (Leaver-Fay et al., 2011). The Rosetta software employs a Monte Carlo sampling protocol to randomize the identity and conformation (rotamer) of a randomly chosen residue; the fitness of the mutated protein is then assessed using the Rosetta energy function (Alford et al., 2017). In

the course of a single design trajectory, the Monte Carlo sampling algorithm is applied iteratively to a set of user-defined residues.

We sought to develop a computational protocol within Rosetta that would substantially alter the chemical properties of the native amino acids without negatively affecting the protein's ability to fold. To do this, the RosettaDesign algorithm (Kuhlman et al., 2003) was used to randomly mutate residues within "design spheres" that had radii from 8–12 Å surrounding each of the target residues (Figure 6.3A). Slight alterations to the conformation of the peptide backbone were allowed only for residues that fell within the design sphere. A second shell was also defined that extended 4 Å beyond the inner design sphere. Residues in this shell were precluded from mutating but were energetically minimized in the context of adjacent, mutated residues. Independent design trajectories were carried out for all rigid and flexible residues. The two rigid (V44 and V262) and three flexible (K55, P226 and K256) residues that served as targets for our studies were also prohibited from mutating during the design calculations (Figure 6.3B). Finally, catalytic residues (S70, K73, S130, E166, K234) were also maintained as their native identities and conformations during the design process. The designed proteins contained between two and eleven mutations with an average of seven mutations per protein. Ultimately, 64 unique designed proteins were generated using this approach.

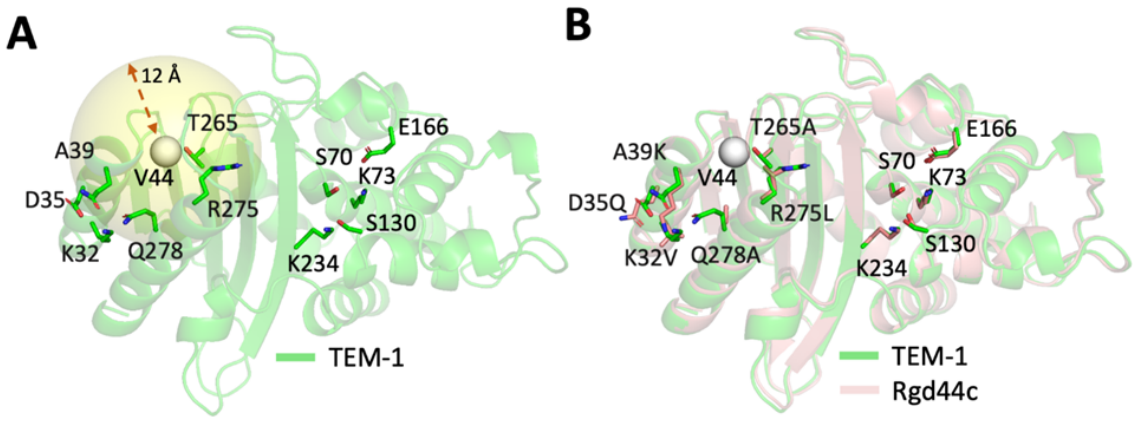


Figure 6.3: Our general computational protein design strategy is shown schematically using the designed protein Rgd44c as an example. (A) Residues within an 8-12 Å sphere surrounding a given residue (V44 in this example) are considered to be candidates for mutation. (B) A combination of mutations surrounding the target residue are generated using the RosettaDesign algorithm and scored using the Rosetta energy function. An overlay of the Rgd44c design model with TEM-1 (panel B) indicates that this design protocol creates a diversity of mutations within the design sphere while leaving active site residues untouched. The target rigid residue (V44) is shown as a white sphere in both panels. Both catalytic and designed residues are shown as sticks.

6.4.3 Selection of The Designed Proteins Using Flexibility Profiles

To assess how the computationally designed mutations affected TEM-1 dynamics, we subjected all designed proteins to a 1 μs molecular dynamics (MD) simulation followed by analysis using the dfi metric (Figure 6.4A). In order to rapidly compare the dfi profiles of our designed proteins to those of TEM-1 and GNCA, we used a 2D principal component analysis (PCA). The PCAs both simplified our data and allowed for the facile visualization of relationships between the calculated dynamic profiles of the designed proteins (Figure 6.4B). PCAs generated from our rigid designs showed a diverse distribution in both the first and second principal components (Figure 6.4C). On the PCA, several designed proteins were positioned relatively closer to GNCA in both components. We chose a subset

of five such designs in which the allosteric rigid residues had been targeted (henceforth referred to as "rigid designs") for experimental characterization (Figure 6.4C).

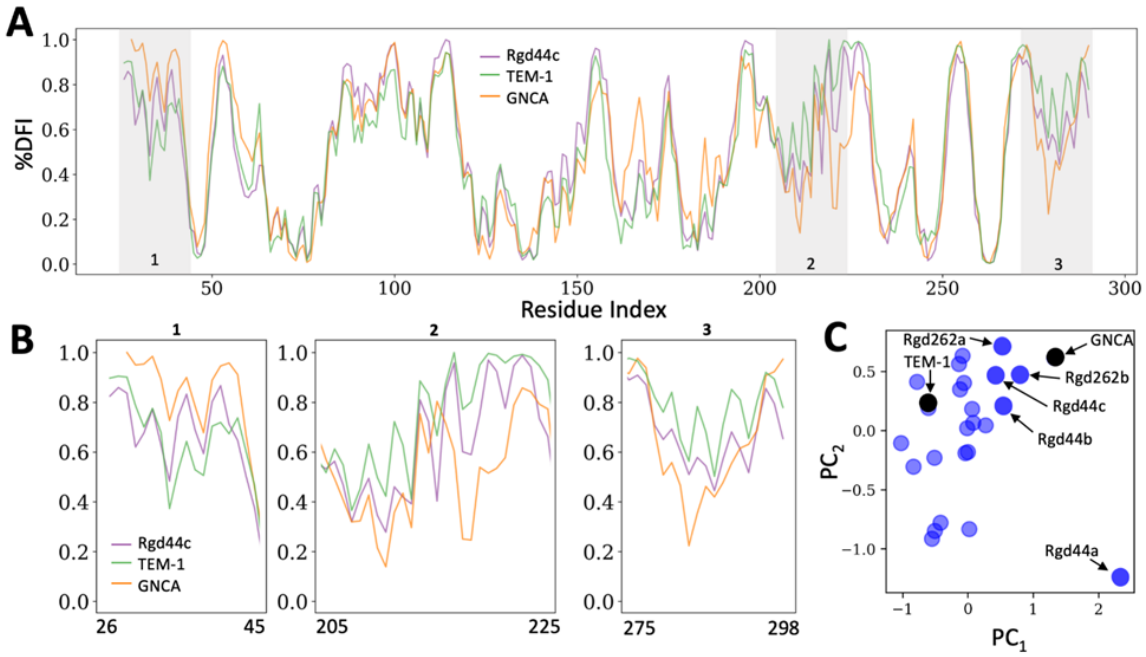


Figure 6.4: Dynamic analyses of TEM-1, GNCA, and the rigid designs. A) Depiction of the *dfi* profile of TEM-1 (green), GNCA (orange) and variant Rgd44c (purple). Rdg44c is chosen as an example for illustrative purposes. B) Portions of the full *dfi* profile of each protein (panel A) are expanded to highlight dynamic differences between the three proteins. A shift towards a GNCA-like *dfi* profile is an indication of a change in dynamical characteristics of a protein. C) Principal Component Analysis (PCA) is applied to the rigid designs. First and second principal components are plotted on the x- and y-axes, respectively. Designs chosen for experimental characterization are highlighted using darker colors and labeled with the design name.

Four of the five rigid designs (Rdg44b, Rdg44c, Rdg262a, and Rdg262b, where the number in each name corresponds to the rigid residue that was targeted in the design calculations) clustered slightly away from TEM-1 and towards GNCA on both axes of the PCA; alternatively, Rdg44a, clustered near GNCA on the first principal axis but appeared as an outlier on the second axis. We hoped that experimental characterization of Rdg44a might help elucidate the parameters captured in each of the two principal components. It

should be mentioned that only four among the five rigid designs that were chosen for characterization had Rosetta scores that were lower (lower Rosetta scores imply lower energies) than TEM-1. The Rosetta score of Rdg262a was higher than TEM-1, but we selected this design for experimental characterization due to the fact that it clustered near GNCA in both axes of the PCA.

To analyze the designed proteins in which flexible, uncoupled residues were targeted (henceforth referred to as "flexible designs"), we generated a PCA in which all flexible design candidates were compared to TEM-1, GNCA and all the rigid designs including those that were not selected for characterization (Figure B.2). Although a wide distribution of flexible designs was observed in this PCA, many of them clustered near TEM-1; a smaller subset clustered near the rigid designs we previously selected for characterization. In an effort to avoid biases that might have arisen if we chose only flexible designs that clustered with TEM-1 for analysis, we opted to experimentally characterize four flexible designs (Flx226a, Flx226b, Flx226c and Flx55) that clustered near the rigid designs chosen for experimental characterization and only one (Flx256) that clustered near TEM-1 (Figure B.2). Although clustering in similar locations in the PCA would suggest that the two proteins should have similar properties, it is difficult to infer what feature is represented on each axis of the PCA. We hoped that the diverse selection of proteins chosen for characterization would therefore provide information regarding whether rigid residues serve as hubs of dynamic control and also whether or not the PCA is a useful metric for discriminating between proteins with different activity and thermostabilities.

6.4.4 Experimental Analysis of The Designed Proteins

As GNCA and TEM-1 differ substantially with respect to thermostability (90.3 °C and 56.4 °C, respectively) and activity against penam β-lactam antibiotics (GNCA is ~2 orders of magnitude less efficient at degrading ampicillin than TEM-1), we chose to focus our analyses of the designed proteins on these characteristics. To do this, genes encoding each of the selected rigid and flexible designs were first cloned into the pET29b expression plasmid. Sequenced confirmed plasmids were transformed into a BL21 Star (DE3) *Escherichia coli* expression strain in preparation for further analyses.

We assessed the resistance of our designed proteins to penam β-lactams by establishing the minimal inhibitory concentration of ampicillin (MIC_{amp}) for each of our designed proteins using the protocol of Wiegand et al. (2008). Briefly, BL21 Star (DE3) cells harboring a pET29b plasmid that contained a gene encoding one of our variants were grown in a liquid medium containing a range of ampicillin concentrations and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), which induced overexpression of our TEM-1 variants. The ability of cells to grow at each ampicillin concentration was determined by measuring the optical density at 600 nm ($O.D._{600}$); the lowest antibiotic concentration that inhibited growth was recorded. All rigid designs exhibited either minimal or no activity against ampicillin (Table 6.1). The two rigid designs that showed the highest activity against ampicillin, Rdg44c and Rdg262b, had MIC_{amp} values of 26 μg/mL, which is two orders of magnitude less efficient than wild type TEM-1 (MIC_{amp} = 1500 μg/mL), but is only half that of GNCA (MIC_{amp} = 43 μg/mL). Alternatively, the

MIC_{amp} values of all the flexible designs were in the range of 375-1500 μg/mL (Table 6.1) which is on par with wild type TEM-1.

Table 6.1: Minimal Inhibitory Concentrations (MIC_{amp}) and melting temperatures of the TEM-1 variants. MIC values are determined in Luria broth. Melting points were determined by circular dichroism. NM indicates that a T_m was not established for this protein due to aggregation during purification.

Variant	MIC_{amp} (mg/ml)	T_m (°C)
GNCA	43	90.3
TEM-1	1500	56.4
Rdg44a	< 2	NM
Rdg44b	< 2	63.1
Rdg44c	26	66.4
Rdg262a	< 2	NM
Rdg262b	26	56.4
Flx226a	1500	57.4
Flx226b	375	53.2
Flx226c	1500	55.6
Flx256	750	58.1
Flx55	750	58.5

Two possible explanations for the lack of activity against ampicillin observed in our rigid designs are: (1) that only poor protein expression was achieved or (2) that they did not fold into native-like structures; neither of these possibilities are directly examined in MIC assays. We therefore expressed and purified each of the designed proteins and assessed their abilities to adopt native-like structures using circular dichroism (CD) spectroscopy. All designed proteins were observed to express soluble (Figure B.3). However, two of the rigid designs, Rdg44a and Rdg262a, had a high propensity to aggregate during the purification process, which precluded further characterization. In

contrast, no aggregation of any of the flexible designed proteins was observed throughout the purification process. We subjected all purified proteins to both wavelength scans and thermal melts using CD, which allowed determination of the melting temperature (T_m) of each protein (Figure B.4). The T_ms of all flexible designs fell into a range (53.2 °C to 58.5 °C) that was within ~3 °C of the T_m of TEM-1 56.4 °C (Table 6.1). Alternatively, the T_ms of the rigid designs varied greatly. Although the least stable of the allosteric rigid designs (Rdg262b) exhibited a T_m that was on par with TEM-1, two others exhibited marked increases in stability. Namely, Rdg44b and Rdg44c were measured to have T_ms of 63.1 °C and 66.4 °C, respectively, which correspond to increases of ~6 °C and 10 °C relative to TEM-1.

The residues targeted for design in this study exhibit a broad distribution of distances from the active site. For example, the two rigid residues (V44 and V262) are closer to the active site than any flexible residues that were targeted for design with distances of 10.1 Å and 17.3 Å, respectively, while the distance of the flexible residues from a catalytic residue ranged from 17.5 Å–22.1 Å. We therefore sought to assess whether or not a correlation existed with respect to the distance from a targeted residue to the active site and altered enzymatic function. To do this, we calculated the distances between the Cαs of all residues mutated during the design process and the Cα of the nearest catalytic residue for all experimentally characterized proteins (Table B.1) using the PyMOL software (The PyMOL Molecular Graphics System, Version 4.3; Schrödinger, LLC: New York, NY, USA).

The two designed proteins that had the shortest distances between a mutated residue and one of the catalytic residues both targeted residue 262 (Rdg262a and b). Rdg262a carries a mutation at position 233, which is directly adjacent in sequence space to catalytic residue 234. Rdg262b contains the next shortest distance between a mutation and an active site residue at 5.8 Å. Rdg262a showed no activity against ampicillin; it is possible that the observed lack of activity is due to the protein's instability and/or propensity to aggregate as observed during purification. Alternatively, Rdg262b possessed an identical T_m to TEM-1 but showed minimal activity against ampicillin despite containing a mutation that is only ~6 Å away from a catalytic residue. On the other end of the spectrum, the nearest mutations to any catalytic residue in two of the flexible designs, Flx226a and c, are 18.5 and 17.5 Å away, respectively. Both of these TEM-1 variants showed near native activity against ampicillin, which is consistent with the fact that mutations that are both distant from and uncoupled to the active site should have little effect on activity.

In the remaining designs, the distribution of distances between the nearest catalytic residue and a designed mutation are much more similar irrespective of whether rigid or flexible residues were targeted. For example, Rdg44a and Flx226b both have mutations that are 12.1 Å from a catalytic residue and Rdg44c and Flx55 have mutations that are 9.7 Å and 9.8 Å away from the catalytic residues, respectively. As these pairs of proteins contain one rigid and one flexible design and also exhibit similar distances between the nearest mutation and any catalytic residue, they appear to provide a direct test of the implications of targeting mutations to flexible vs. rigid residues. Interestingly, Rdg44a was highly unstable and aggregation prone despite only having mutations over 10 Å away from

the catalytic residues. In contrast, Rdg44c had activity against ampicillin that was three orders of magnitude less than the wild type protein, but also showed a 10 °C increase in T_m relative to TEM-1. Alternatively, both flexible designs (Flx226b and Flx55) maintained substantial activity against ampicillin and exhibited T_ms that were within 3 °C of wild type TEM-1 (Table 6.1). These data further support the notion that rigid, highly coupled residues play a large role in determining both the activity and physical properties of TEM-1. Furthermore, the fact that the rigid designs that adopted a native-like fold showed a substantial decrease in activity supports the notion that our dci metric can provide meaningful information regarding residues that may be able to affect protein function via allosteric dynamic coupling to the active site.

6.4.5 Dynamics Analysis of The Designed Proteins

Experimental characterization of our designed proteins demonstrated that the MIC_{amp} values of the rigid designs were significantly reduced relative to both TEM-1 and the flexible designs irrespective of the distances between the nearest mutations and the catalytic residues. This suggests that changes in the local network of interactions surrounding rigid residues that exhibit long-range dynamic coupling with the active site may allosterically alter the flexibility of active site residues. In order to further analyze this possibility using our computational metrics, we calculated the flexibility of the active site residues in both sets of designed proteins using the dfi metric. The dfi values of each catalytic residue in our experimentally characterized proteins were subtracted from those of TEM-1 to generate a Δdfi profile (Figure B.5A). A clear difference between the Δdfi values of the catalytic residues of the rigid and flexible designs was observed (Figure 6.5).

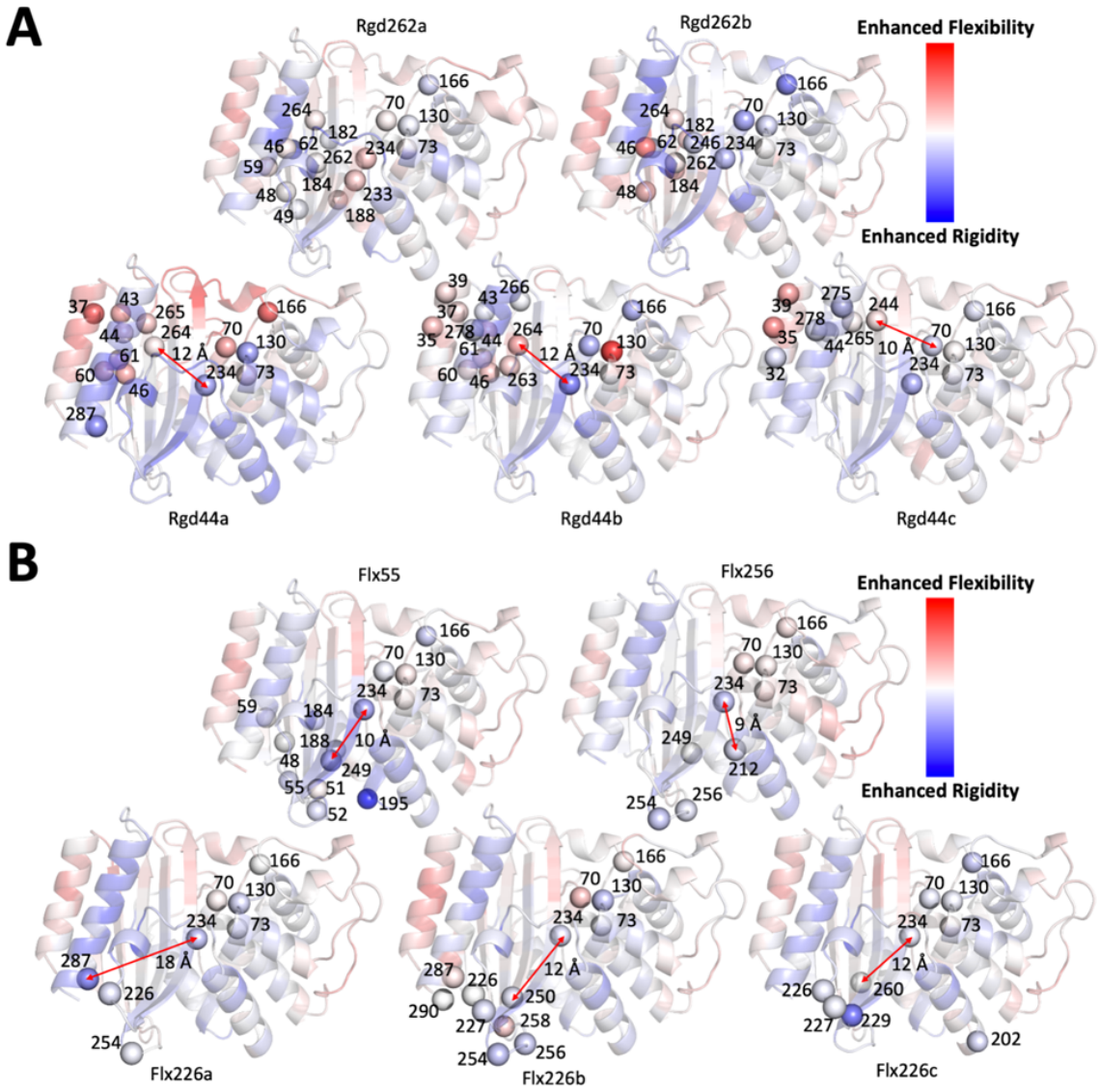


Figure 6.5: The change in the dynamics profiles of experimentally characterized rigid (A) and flexible (B) designs (Δ*dfi* values) are mapped onto the TEM-1 structure. Point mutations around the residues targeted for design and catalytic residues in TEM-1 are shown as spheres and labeled with their residue indices. The distance between the mutations closest to the catalytic residues are marked with red arrows and labeled with the corresponding distance in angstroms. The minimum distance in most designs is larger than 10 Å (Rgd262a and b and Flx256 are exceptions), which suggests that the changes in dynamics of catalytic residues is due to distal allosteric communication with the active site in many instances.

Namely, the catalytic residues in the rigid designs underwent a greater change in relative flexibility (both increases and decreases) compared to the flexible designs. Alternatively, the relative flexibilities of the catalytic residues in the flexible designs exhibited a narrower distribution centered at zero (Figure B.5B). These data support the notion that the rigid residues we chose are highly coupled to the active site (as suggested by our original dci analysis) and also that targeting the local interaction of allosteric rigid residues can indeed alter the flexibilities of residues, even if they are separated by substantial distances.

Our experimental results and the detailed dfi profiling of the experimentally characterized designs brought to light the fact that our initial PCA analysis did not appear to adequately discriminate between the activities of the designed proteins. Although designs in which rigid, coupled residues were targeted often possessed vastly different properties than those in which flexible, uncoupled residues were targeted, many of these designs clustered in similar areas of the PCA (Figure B.2).

Therefore, we sought to develop a new metric that might have a greater discriminatory ability than the PCA alone. We developed an iterative method that we have termed the Dynamic Distance Analysis (dda) in which the "dynamic distance" of a designed protein to either TEM-1 or GNCA is computed relative to those of randomly selected groups of designed proteins. As the distance between any two proteins in a PCA (based on their three principal eigenvectors, see chapter 6.3.1) depends on the component proteins used to generate that PCA, randomly selected sets of designed proteins should yield a much better

picture of the true relationship between a given designed protein and a target protein (TEM-1 and GNCA).

To generate the dda profiles of our designed proteins, we used a bootstrapping approach in which we first generated multiple PCAs using small, randomly chosen subsets of designed proteins and then iteratively measured the distances between the dfi profiles of each designed protein and both GNCA and TEM-1 (Figure B.1). When we clustered the dda profiles of the rigid and flexible designs using a new PCA; a clear separation between the two emerges (Figure 6.6), which correlates well with their biophysical characterization. For example, flexible designs Flx55 and Flx256 cluster together in our dda analysis and also possess similar MIC_{amp} values (750 μg/mL). Similarly, Flx226a and Flx226c, whose MIC_{amp} values are the same as TEM-1 (1500 μg/mL), also appear in very similar regions of the dda PCA. The two rigid designs, Rgd44a and Rgd262a, which exhibited aggregation during purification, are both found as outliers in the dda clustering. Notably, Rgd44c and Rgd262b, which exhibit higher thermostabilities and similar MIC_{amp} values to TEM-1, are also clustered in the same vicinity.

In an effort to assess whether or not the trends observed in the dda analyses of experimentally characterized proteins were universal, we applied dda to all the designed proteins, even those not chosen for characterization. Interestingly, the dynamic distances of the rigid designs are biased away from TEM-1 relative to their flexible design counterparts (Figure B.6A); conversely, the flexible designs form a narrower distribution that is closer to TEM-1. This suggests that flexible residues that are not coupled to the

active site do not likely contribute to the collective motion of the protein as substantially as do rigid residues.

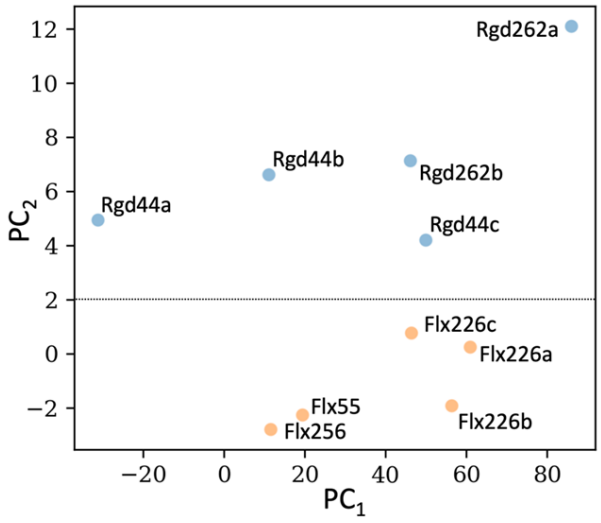


Figure 6.6: Dynamic distances are clustered for all characterized allosteric rigid (blue) and uncoupled flexible (orange) designs. The rigid designs and the flexible designs cluster separately. Designed proteins with similar MIC_{amp} values, (Flx55 and Flx256), (Flx226c and Flx226a), (Rdg262b and Rdg44c) cluster in the same vicinity.

When the distances of our designed proteins to GNCA are considered, the uncoupled flexible designs display a sharp, narrow distribution that is distant from GNCA (Figure B.6B). Alternatively, the distribution of the rigid designs is broad and contains proteins with dynamic profiles that are more like that of GNCA. These data suggest that the re-design of the environment surrounding rigid residues appears to alter the dynamics of TEM-1 more substantially than when the environment surrounding uncoupled flexible residues is targeted.

6.5 Conclusions

The goal of this work was to better understand the relationship between structure and function in the TEM family of β-lactamases. Building on previous evolutionary studies on the β-lactamase enzyme TEM-1 (Zou et al., 2015), we explored the hypothesis that rigid residues can serve to both establish the global dynamic profile of the enzyme and exert substantial influence over physical properties (e.g., substrate specificities) so long as long-range coupling exists between the rigid residues and the active site. To explore this, we used the Rosetta computational protein design software to re-design the local network of interactions surrounding residues that fit the aforementioned criteria. Our designed proteins were analyzed using computational metrics that assessed both the global dynamic profile and the allosteric coupling of each residue to the active site. Based on these metrics, a subset of our designed proteins was selected for experimental characterization.

Ten designed TEM-1 variants were characterized with respect to the minimal inhibitory concentration of ampicillin as well as thermostability. These data suggested that targeting mutations to environments surrounding rigid residues that were highly coupled to the active site often resulted in a substantial shift in protein stability and function; alternatively, targeting flexible, uncoupled residues resulted in protein variants with more native-like activities and thermostabilities. Namely, when mutations were targeted to the vicinity of two rigid residues that do not directly interact with the active site, but which are highly coupled to it, a substantial reduction in TEM-1′s ability to degrade its native substrate was observed in all cases even though native-like folds were maintained in many cases. Alternatively, thermostabilities and activities against TEM-1′s native substrate were

maintained in a set of designed proteins in which residues that were neither rigid nor predicted to be coupled to the active site were targeted for mutagenesis. These results are consistent with our computational analyses of the designed proteins' dynamics. Namely, it appears that altering the local interactions surrounding rigid residues that are highly coupled to the active site can allosterically alter the flexibility profiles of active site residues at a distance, which can in turn alter the biophysical properties of the enzyme. In an effort to identify an analytical method that was more informative as to the activities that designed proteins might possess, we developed a novel metric that measures the "dynamic distance" between two proteins. Many of our designed proteins with similar functional properties were observed to cluster together when analyzed by this algorithm. These results not only further support the potential importance of mutations in the vicinity of rigid residues, but also support the fact that coupling between distal residues and the active site can have profound effects on enzyme activities.

The relationship between protein dynamics and function is highly complex and studying it represents an exceedingly difficult challenge (Knies et al., 2017; Ma et al., 2011; Orencia et al., 2001; Salverda et al., 2010; Singh and Dominy, 2012; Zhang et al., 2020). Our approach represents a new method for exploring this subject in a highly directed manner. We hope that additional application of these methods to distinct residues in TEM-1 will ultimately provide a more complete understanding of the complex dynamic landscape present in this class of proteins. This could not only facilitate a rapid prediction of the biochemical properties of new clinical isolates but could also pave the way for the development of new antibiotics that specifically target new protein conformations

accessible only through alterations of the global dynamic profile. Finally, the methods reported here could also find use in understanding the dynamic profiles of other enzyme classes, which could have profound implications from the perspective of understanding and treating diseases.

6.6 Acknowledgement

This research was funded by The National Science Foundation, grant number 1901709. We thank Jose Sanchez-Ruiz (Universidad de Granada) for the generous gift of the GNCA expression plasmid and Ron Mills for helpful review of the manuscript.

CHAPTER 7

DESIGN OF NOVEL CYANOVIRIN-N VARIANTS BY MODULATION OF BINDING DYNAMICS THROUGH DISTAL MUTATIONS

This chapter is adapted from: "Kazan, I. C.; Sharma, P.; Rahman, M. I.; Bobkov, A.; Fromme, R.; Ghirlanda, G.;Ozkan, S. B. Design of novel cyanovirin- N variants by modulation of binding dynamics through distal mutations. eLife 2022;11:e67474."

I Can Kazan shared first-authorship with Prerna Sharma and Mohammad Imtiazur Rahman. I Can Kazan conducted all computational work presented here, while Prerna Sharma and Mohammad Imtiazur Rahman performed experimental works.

Drawing upon the cumulative findings elucidated in preceding chapters, in the final chapter of this thesis, I introduce an evolutionary guided molecular dynamics driven protein design approach to identify distal residues that modulate binding site dynamics through allosteric mechanisms. To achieve this, I developed integrated co-evolution and dynamic coupling (ICDC) approach to identify distal residues, find amino acid substitutions, and assess the effect of mutations in modulation of function. Integrating the key concepts and findings from preceding chapters, ICDC combines dynamic information calculated by DFI and DCI with coevolutionary coupling information to identify residues that could have diverse effect on binding upon mutations. To validate the effectiveness of ICDC, I analyzed preexisting mutational fitness data of β-lactamase and discovered that rigid positions (low DFI) with high co-evolution and dynamic coupling (high DCI) to the

catalytic sites exert significant influences on function. After confirming the approach with a comprehensive enzyme dataset, I applied ICDC to Cyanovirin-N (CV-N), a lectin with specific dimannose binding; for identification, mutation and assessment of allosteric positions that can modulate binding affinity. Once positions and possible amino acid substitutions were identified with ICDC, the novel variants are modeled with MD simulations. Then, the change in dimannose binding affinity of the variants relative to wild type is modeled by Adaptive BP-Dock as explained in detail in chapter 1 and 2. The predictions were validated by experiments conducted. The findings derived from this meticulous investigation and subsequent engineering of CV-N details the power of utilizing dynamic metrics combined with MD simulations and co-evolution.

7.1 Abstract

We develop integrated co-evolution and dynamic coupling (ICDC) approach to identify, mutate, and assess distal sites to modulate function. We validate the approach first by analyzing the existing mutational fitness data of TEM-1 β-lactamase and show that allosteric positions co-evolved and dynamically coupled with the active site significantly modulate function. We further apply ICDC approach to identify positions and their mutations that can modulate binding affinity in a lectin, Cyanovirin-N (CV-N), that selectively binds to dimannose, and predict binding energies of its variants through Adaptive BP-Dock. Computational and experimental analyses reveal that binding enhancing mutants identified by ICDC impact the dynamics of the binding pocket and show that rigidification of the binding residues compensates for the entropic cost of

binding. This work suggests a mechanism by which distal mutations modulate function through dynamic allostery and provides a blueprint to identify candidates for mutagenesis in order to optimize protein function.

7.2. Introduction

The evolutionary history of a protein comprises the ensemble of mutations acquired during the course of its evolutionary trajectory across different species, and contains valuable information on which residue positions contribute the most to a given protein's 3D-fold and function based on their conservation (Campbell et al., 2016; Rivoire et al., 2016; Yang et al., 2016). Furthermore, the subset of positions that are co-evolved (i.e., correlated mutational sites) provide clues on specific, native-state interactions. Pairwise residue contacts inferred from co-evolved positions within a protein family can be used as distance restraints to accurately model 3D structures (de Juan et al., 2013; Hopf et al., 2019; Kamisetty et al., 2013; Kim et al., 2014; Tripathi et al., 2015). Recent revolutionary successes in accurate predictions of 3D protein structures combine these methods with machine learning strategies, that is, deep learning (Jumper et al., 2021; Wang et al., 2016; Xu, 2019). Co-evolved positions also embed information on protein function, for example, revealing how factors such as binding affinity and specificity are modulated across evolutionary history and species (Rivoire et al., 2016; Salinas and Ranganathan, 2018; Torgeson et al., 2022). However, accessing, interpreting, and applying this information in a predictive manner is very challenging; mutations observed in the evolutionary history are often distal from the functional sites, implying that protein dynamics are responsible for their effects on function and that these sites act as distal allosteric regulators of function

(Campitelli et al., 2020; Modi et al., 2021a; Romero and Arnold, 2009; Salinas and Ranganathan, 2018; Tokuriki et al., 2012; Wei et al., 2016).

Molecular dynamics (MD) simulations can capture protein dynamics and reveal the impact of distal mutations on function (Bowman and Geissler, 2012; Campbell et al., 2016; Campitelli et al., 2020; Kolbaba-Kartchner et al., 2021; Modi et al., 2021a; Yang et al., 2016). However, the computational cost of MD simulations of sufficient length can be prohibitively high; further, it's often far from straightforward to forge a clear connection to function. To bridge this gap, we developed a framework to quickly evaluate MD trajectories and identify the sensitivity of a given position to mutation based on its intrinsic flexibility, which we assess using our dynamic flexibility index (DFI) metric, and on its dynamic coupling with functionally critical positions assessed by dynamic coupling index (DCI) (Campitelli et al., 2018; Gerek and Ozkan, 2011; Kumar et al., 2015b; Larrimore et al., 2017). DFI measures the resilience of a position by computing the total fluctuation response and thus captures the flexibility/rigidity of a given position. Applying DFI to several systems, we showed that rigid positions such as hinge sites contribute the most to equilibrium dynamics, and that mutations at hinge sites significantly impact function regardless of the distance from active sites (Kim et al., 2015; Kolbaba-Kartchner et al., 2021; Modi et al., 2021a, 2021b, 2018; Zou et al., 2021, 2015). DCI measures the dynamic coupling between residue pairs and thus identifies positions most strongly coupled to active/binding sites; these positions point to possible allosteric regulation sites important for modulating function in adaptation and evolution (Butler et al., 2015; Campitelli et al.,

2021; Kuriyan and Eisenberg, 2007; Lu and Liang, 2009; Modi et al., 2021a; Modi and Ozkan, 2018; Ose et al., 2020; Risso et al., 2018; Wodak et al., 2019).

Here, we present integrated co-evolution and dynamic coupling (ICDC) approach to identify distal allosteric sites, and to assess and predict the effects of mutations on these sites on function. We propose a system to classify residue positions in a binary fashion based on co-evolution (co-evolved, 1 or not, 0) and dynamic coupling by DFI and DCI (dynamically coupled 1, or not, 0) with the functionally critical sites. This classification captures the complementarity of dynamics-based and sequence-based methods. We hypothesize that positions belonging to category (**1,1**), that is, positions both co-evolved and dynamically coupled with the functional sites, will have the largest effect on function.

We validate our hypothesis first by analyzing the existing mutational fitness data for TEM-1 β-lactamase, available for every position of the protein (Stiffler et al., 2015). In agreement with our hypothesis, we find that mutations on category (**1,1**) positions significantly modulate the function. A large fraction of mutations enhancing enzymatic activity correspond to category (**1,1**) irrespective of distance from the active site. Second, we apply our ICDC approach to blindly predict and experimentally validate mutations that allosterically modulate dimannose binding in a natural lectin, cyanovirin-N (CV-N). CV-N binds dimannose with nanomolar affinity and remarkable specificity (Barrientos et al., 2003; Botos and Wlodawer, 2005, 2003; Mori and Boyd, 2001; O'Keefe et al., 2003). It is part of the CV-N family, found in a wide range of organisms including cyanobacterium, ascomycetous fungi, and fern (Koharudin et al., 2008; Koharudin and Gronenborn, 2013; Patsalo et al., 2011; Percudani et al., 2005; Qi et al., 2009). While the 3D folds is

remarkably conserved in all experimentally characterized members, the affinity and specificity for different glycans and, in particular, to dimannose varies significantly (Koharudin et al., 2009, 2008; Matei et al., 2016; Woodrum et al., 2013). To design CV-N variants with improved binding affinities for dimannose based on distal allosteric coupling, we binned each position in one of the four categories based on computed DFI, DCI, and co-evolution rates. We explored mutations at these sites based on frequency in the sequence alignment. After obtaining the mutant models through MD simulations, we assessed the impact of each naturally observed mutation on binding affinity by docking dimannose to the mutant models via Adaptive BP-Dock (Bolia et al., 2014a, 2014b; Bolia and Ozkan, 2016). We chose position I34, which belongs to category (**1,1**) and is 16 Å away from the binding pocket, for experimental validation. We found that mutations I34K/L/Y had a diverse effect on glycan binding, either improving by two-fold or abolishing completely. Through experimental (explained in Appendix C) and MD studies we show that the observed improvement in binding affinity is due to changes in the dynamics of residues in the binding pocket; mutation I34Y leads to rigidification of binding sites, thus compensating the entropic cost of binding (Breiten et al., 2013; Chodera and Mobley, 2013; Cornish-Bowden, 2002; Fox et al., 2018). Mutations at an additional position (A71T/S) from category (**1,1**) showed evidence of the same allosteric mechanism governing the modulation of binding dynamics. Overall, this study provides not only a new approach to identify distal sites whose mutations modulate binding affinity, but also sheds light into mechanistic insights on how distal mutations modulate binding affinity through dynamics allostery.

7.3 Methods Used for Modulation of Binding Dynamics Of CV-N

7.3.1 Adaptive BP-Dock

Adaptive backbone perturbation docking, Adaptive BP-Dock in short, allows us to model the interaction between CV-N and glycans *in silico* (Bolia and Ozkan, 2016). Adaptive BP-dock combines the complex simulation of backbone flexibility of a protein into Rosetta's ligand docking application (Davis and Baker, 2009). The common restriction in docking is the implementation of flexibility of receptor and ligand (Davis et al., 2009; Davis and Baker, 2009; DeLuca et al., 2015; Meiler and Baker, 2006). Rosetta included the flexibility of ligand in their monte-carlo sampling approach but lacking full receptor flexibility. This high order challenge is overcome by utilizing Perturbation Response Scanning (PRS) to compute backbone changes during docking (Atilgan and Atilgan, 2009; Bolia et al., 2014; Ikeguchi et al., 2005). This procedure also allows the modeling of transition from an unbound state to a bound state (Bolia and Ozkan, 2016). The computational cost of sampling is reduced by using a coarse-grained approach employing Elastic Network Model (ENM) leading to an efficient way of computing backbone perturbations, mimicking the ligand interacting with receptor (Atilgan et al., 2001, 2010; Atilgan and Atilgan, 2009).

We employed Adaptive BP-Dock in modeling glycan CV-N interactions starting from an unbound conformation of CV-N. The perturbed pose of the protein is calculated using PRS. The structure is then minimized, and the side chains are added at this step. The glycan is docked to the minimized structure using RosettaLigand algorithm. Rosetta samples bound conformations using a knowledge based potential function and calculates

bound pose energies. The lowest energy docked pose is selected and feed back to perturbation step, and the same procedure is followed iteratively until a convergence is reached. At the end of each iteration the lowest energy docked pose is taken and binding score is calculated using an empirical scoring function X-score. X-score energy units (XEUs) has shown to provide higher correlations with experimental results (R. Wang et al., 2002). Adaptive BP-Docks iterative algorithm ensures the sampling does not get trapped in a local minimum and reaches a global minimum. The challenge of unbound/bound modeling is solved using the iterative approach as the conformations are led towards a bound pose with the help of PRS.

7.3.2 Molecular Dynamics (MD)

Gromacs simulations are conducted for P51G-m4 CV-N and all the variants in unbound form, and further for P51G-m4 CV-N, I34 variants I34K, I34L, I34Y, and A71 variants A71S, A7T in bound form. (Abraham et al., 2015; Spoel et al., 2005). For each simulation the all-atom system is parametrized with CHARMM36 force field and explicit water model TIP3P. The solvation box is set to be minimum 16Å from the edge of the protein. The system is neutralized by potassium ions to sustain electroneutrality and minimized with steepest descent for 10000 steps. A short-restrained equilibrium is conducted in the constant number of particles, pressure, and temperature ensemble (NPT) for 5 ns using the Berendsen method at 300K temperature and 1 bar pressure. NPT production trajectories were performed with Nose-Hoover and Parrinello-Rahman temperature and pressure coupling methods for 2μs at 300K and 1 bar. For all cases

periodic boundary conditions and particle-mesh Ewald (PME) with interaction cutoff of 12Å is employed with Gromacs version 2018.1.

7.3.3. Dynamic Flexibility Index (DFI)

DFI is a position specific metric that can measures the resilience of a given position to the force perturbations in a protein. It calculates the fluctuation response of a residue relative to the gross fluctuation response of the protein (Kumar et al., 2015b; Larrimore et al., 2017). DFI calculates residue response due to a perturbation by utilizing covariance matrices.

$$[\Delta \mathrm{R}]_{3Nx1} = [H]^{-1}_{3Nx3N}[\mathrm{F}]_{3Nx1} \tag{7.1}$$

$$DFI_i = \frac{\sum_{j=1}^{N}[\Delta R^j]_i}{\sum_{i=1}^{N}\sum_{j=1}^{N}[\Delta R^j]_i} \tag{7.2}$$

Residue response, $\Delta \boldsymbol{R}$, is calculated using Linear Response Theory (LRT) by applying force, $\boldsymbol{F}$, in multiple directions to mimic isotropic fluctuations. Hessian matrix, $\boldsymbol{H}$, contains second derivatives of potentials. The inverse of Hessian matrix, $\boldsymbol{H^{-1}}$, contains residue covariances, and interpreted as a covariance matrix. The covariance matrices can be gathered from MD simulations, and also by using Elastic Network Model (ENM) of a protein. In this study, MD covariance matrices have been utilized to incorporate residue interactions accurately.

Residues with low DFI score (below 0.2) are considered as hinge points. These points are communication hubs in this 3-D interaction network. Due to high coordination number, the residues exhibiting low DFI values are crucial as information gateways. While they do not exhibit high residue fluctuation to the perturbations, they quickly transfer the

perturbation information to other parts, thus they are in control of collective motion of the protein. A change in low DFI positions (i.e., a mutation) will lead to a transformation in the communication grid and majority of disease-associated (i.e. function altering mutations) are often observed as hinges (Butler et al., 2015; Gerek et al., 2013; Kumar et al., 2015a). The substitution on these site usually alters catalytic activity or binding interaction (i.e., glycans) by modulating equilibrium dynamics (Campitelli et al., 2020).

7.3.4 Dynamic Coupling Index (DCI)

Dynamic Coupling Index (DCI) exploits the same framework of DFI (Campitelli et al., 2020; Larrimore et al., 2017). DCI utilizes the residue response fluctuation upon random force perturbation at a specific residue position to investigate residues that exhibit long-range coupling to each other. In DCI approach, a unit force is applied on functional residues (i.e., binding site residues) one by one and responses of all other residues are calculated.

$$DCI_i = \frac{\sum_{j=1}^{N_{Functional}} \left|\Delta \boldsymbol{R}^j\right|_i / N_{Functional}}{\sum_{j=1}^{N} |\Delta \boldsymbol{R}^j|_i \ / \ N} \tag{7.3}$$

With DCI scheme the residues with high response (high DCI score) indicates high long range dynamic coupling. Residues with high DCI values with binding sites play a critical role in intercommunication of a protein with the binding residues. These coupled residues are of utmost importance in how forces propagate through amino acid chain network on a binding event. Some of the coupled residues are far from the binding site but still encompass modulation capabilities over binding pocket.

7.3.5 Informing Dynamics from Co-evolution

Co-evolutionary data paves the way to assessing 3-D structural contacts by utilizing available sequence information (Hopf et al., 2018; Marks et al., 2012; Morcos et al., 2014a). Sequence information is more abundant compared to resolved protein structures. Exploiting the sequence information, primary contacts comparable to realistic structural contacts can be calculated and a contact matrix is formed. The accuracy of these contact maps is proved to be valuable in protein folding studies (Kryshtafovych et al., 2019; Morcos et al., 2011; Wang et al., 2016). Evolutionary coupling (EC) analysis is used to collect information on how much two residues in a protein sequence is in close proximity in 3-D structure. EC scores could be calculated by many different statistical approaches. In this study EC information is gathered by using RaptorX, EVcouplings, and MISTIC webservers (Hopf et al., 2019; Simonetti et al., 2013; Wang et al., 2017). While the limitation of these methods emerges from sequence homolog availability of a protein in multiple sequence alignment (MSA), RaptorX uses a deep neural network leveraging joint family approach, combining multiple ortholog protein families sharing similar function and phylogeny, to infer possible contacts. This method is proven to produce high accuracy in contact prediction compared to others (Wang et al., 2017). However, for a given MSA containing enough homolog sequences other methods are also strong in predicting spatial contacts. EVcouplings approach uses Direct Information (DI) to calculate co-evolutionary couplings. DI metric is a modified mutual information (MI) score considering consistency between pairwise probabilities and single amino acid frequencies (de Juan et al., 2013; Morcos et al., 2011). Nonetheless, MI, a global approach compared to local DI metric, is

accurate in capturing true contacts, while entangling indirect contacts from direct contacts. MISTIC web server has taken advantage of MI to calculate co-evolutionary couplings (Dunn et al., 2008; Gouveia-Oliveira and Pedersen, 2007; Simonetti et al., 2013). In their MI method they introduced a correction term to MI to surpass the low statistics gathered with an MSA containing limited number of sequences. This approach is very useful in cases where certain homologs are rare and MSA of these homologs have multiple gaps in their alignments. All of these methods are employed in this study to achieve high accuracy predictions in finding residue couplings.

7.4 Results and Discussion

7.4.1. Combining Long-Range Dynamic Coupling Analysis with Co-Evolution Allows to Identify Distal Sites That Contribute to Functional Activity.

With our ICDC approach, we aim to explore the role of dynamics versus evolutionary coupling (EC) as well as the role of rigidity versus flexibility in allosterically modulating active/binding site dynamics. To this extent, we created four unique categories that classify residue positions based on residue DFI score, DCI score, and co-evolutionary score: category (**1,1**) is dynamically and co-evolutionarily coupled rigid sites (exhibiting %DFI values 0.2 or lower, showing 0.7 or higher %DCI with the binding site, and showing 0.6 or higher co-evolution scores with the binding site); category (**1,0**) is dynamically coupled but co-evolutionarily not coupled sites; category (**0,1**) is dynamically not coupled but co-evolutionarily coupled sites; category (**0,0**) is dynamically not coupled, and co-evolutionarily not coupled flexible sites (exhibiting %DFI values 0.7 or higher) (Tables C.1 and C.2); importantly, this classification is based on two independent statistical

approaches thus compensate the noise of individual approaches. Based on our evolutionary analysis (Campitelli et al., 2020; Modi et al., 2021b; Modi and Ozkan, 2018), we hypothesize that category (**1,1**) would impact protein activity or binding affinity the most. To test our hypothesis, we first analyzed the deep mutational scanning data available for the TEM-1 β-lactamase, correlating changes in ampicillin degradation activity (e.g., MIC values) with mutations to all possible amino acids at each position (Stiffler et al., 2015). The experimental results showed that amino acid substitutions at the catalytic site residues of TEM-1 negatively impacted activity. Mutations at other positions also affected activity; while most mutations were deleterious, surprisingly, others resulted in increased activity. The impact of mutations on dynamics and function of TEM-1 have been heavily explored but the distal mutational effects are still poorly understood (Kolbaba-Kartchner et al., 2021; Modi et al., 2021a; Modi and Ozkan, 2018; Salverda et al., 2010; Schneider et al., 2021; Stiffler et al., 2015; Thomas et al., 2010; Zimmerman et al., 2017; Zou et al., 2015). We applied our approach by obtaining DFI, DCI, and co-evolution scores for every position of TEM-1 and binning residue positions into each ICDC category (Table C.1 and Table C.3). We constructed fitness distributions for each category using the experimentally measured single mutant relative fitness values for all mutations per position provided in the dataset (Figure 7.1).

We found that category (**1,1**) positions show the highest impact, both significantly enhancing and reducing ampicillin degradation by TEM-1 (Figure 7.1A&C). In addition, category (**0,0**) residue mutations (i.e., the exact opposite of category (**1,1**)) lie within the neutral-like activity range defined by Stiffler et al (2015), suggesting that mutations on

positions that neither co-evolve nor dynamically couple to active site do not affect the function significantly. Category (**1,0**) residues enhance activity more than those in the neutral category (**0,0**). Mutations in category (**0,1**) positions also modulate function in both positive and negative direction, albeit not as strongly as those in category (**1,1**). However, mutations that negatively impact activity are conspicuously under-represented in the multiple sequence alignment (MSA) of native sequences (Figure 7.1B), particularly in category (**1,1**).

This finding implies nature mostly allows mutations that don't compromise fold and function: Negative selection (i.e., elimination of amino acid types that are detrimental to the folding) is a major force in shaping the mutational landscape (Jana et al., 2014; Modi et al., 2021a; Morcos, 2020; Morcos et al., 2014a, 2013). Thus, the use of conservation information from MSA is a useful tool in eliminating deleterious amino acid substitutions in protein design.

Our ICDC selection criteria effectively identifies residue positions and their amino acid substitutions that could fine-tune function without leading to a functional loss; and category (**1,1**) residues have the largest impact on function irrespective of their distance from active site (Figure 7.1C).

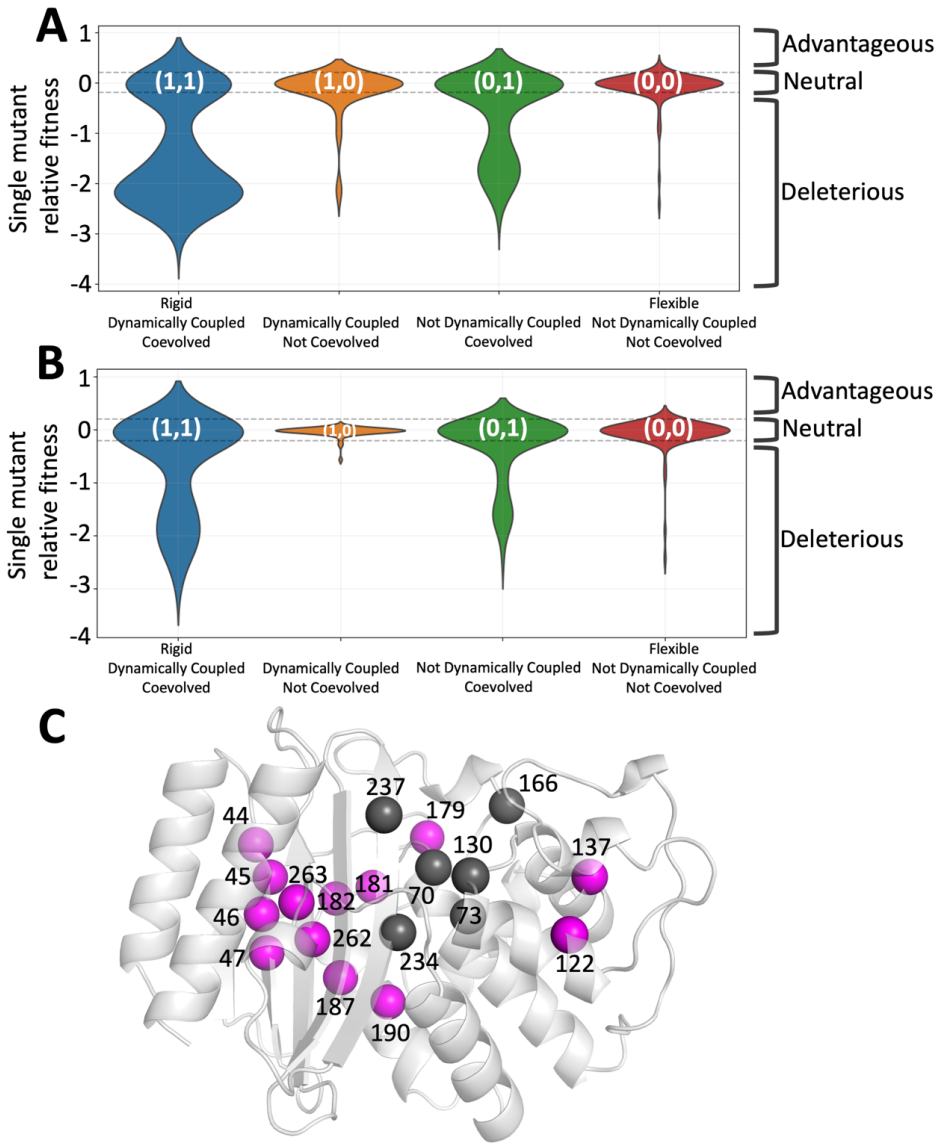


Figure 7.1: ICDC categories based on the dynamics and co-evolutionary analyses applied on TEM-1 β-lactamase. A) The distributions in the form of violin plots are obtained for each ICDC category using all available experimental mutational data (Stiffler et al., 2015) B) Violin plots showing the fitness values for amino acid substitutions observed in the natural sequences. C) The category **(1,1)** positions are mapped on 3-D structure. The catalytic site residues are shown in dark grey whereas category **(1,1)** positions are shown in magenta color. The function altering category **(1,1)** positions are widely distributed over the 3-D structure.

7.4.2 Application of ICDC Approach to Modulate CV-N Binding Affinity Through Distal Mutations

CV-N is a small (11 kDa) natural lectin isolated from cyanobacterium *Nostoc ellipsosporum* which comprises two quasi-symmetric domains, A (residues 1–38/90–101) and B (residues 39–89 respectively), that are connected to each other by a short helical linker. Despite almost having identical structures, the domains show relatively low sequence homology (28% sequence identity and 52% similarity). Functionally, they both bind dimannose, yet the affinity is quite different, with domain B having tighter binding affinity (K_d = 15.3 μM), and domain A showing weak affinity (K_d = 400 μM) (Balzarini, 2007; Bolmstedt et al., 2001; Li et al., 2015).

To simplify our analyses, we used a designed CV-N variant, P51G-m4, that contains a single high-affinity dimannose binding site (domain B), folds exclusively as a monomer in physiological conditions, and is more stable to thermal denaturation than wild type (Fromme et al., 2008, 2007). The binding pocket of domain B of CV-N has been subjected to intense scrutiny to glean information on the origin of its binding specificity for dimannose (Bewley, 2001; Bolia et al., 2014a; Botos and Wlodawer, 2003; Li et al., 2015; Vorontsov and Miyashita, 2009). Previous mutational studies on the binding pocket residues have shown their importance in modulating interaction with dimannose (Barrientos et al., 2006; Bolia et al., 2014a; Chang and Bewley, 2002; Matei et al., 2016). All known substitutions of the binding residues led to decreased binding affinity for dimannose on domain B (Bolia et al., 2014a; Fujimoto and Green, 2012; Kelley et al., 2002; Matei et al., 2016; Ramadugu et al., 2014). Evolutionary analyses shows that the

majority of the binding site residues are conserved in CV-N glycan interactions, suggesting that affinity is already optimized at the binding site (Koharudin et al., 2008; Percudani et al., 2005). We hypothesized that amino acid substitutions at distal positions could enhance the dimannose affinity of CV-N by rigidification of the binding site and applied our ICDC approach to CV-N to identify positions in each category (Table C.2).

We generated models of CV-N variants in each ICDC category by mutating these positions to amino acid types observed in the MSA of CV-N family members, choosing the subset of sequences that have binding sites with identical or similar amino acid composition to P51G-m4 CV-N. As discussed above, this approach allows us to identify amino acid substitutions with the least impact on fold. All the substitutions identified (104 variants in total) were modeled using the crystal structure of P51G-m4 CV-N (Fromme et al., 2008) and subjected to MD simulations (Abraham et al., 2015; Spoel et al., 2005). The best conformation sampled for each variant obtained from equilibrated production trajectories was used as a model for dimannose docking analysis. We evaluated the variants using Adaptive BP-Dock, a computational docking tool that incorporates both ligand and receptor flexibility to accurately sample binding-induced conformations and ranks them using X-scores binding energy units (XEUs). The details of Adaptive BP-Dock are explained in Chapter 2.1.2. In previous work on CV-N this method yielded good correlations with experimentally measured binding affinities (K_d), and established –6.0 XEU as a good threshold to differentiate variants that bind dimannose from 'non-binders' (Bolia et al., 2014b; Li et al., 2015; Woodrum et al., 2013). Here, we applied Adaptive BP-Dock initially on wild-type CV-N and its variants, P51G-m4 and mutDB (a mutant in

which binding by domain B has been obliterated) and the results recapitulate the success of previous studies (Table C.4). This result shows that Adaptive BP-Dock can correctly assess the dimannose binding of CV-N and its variants, thus, we applied it on new P51G-m4 CV-N variants to predict the impact of mutations on dimannose binding. Figure 7.2 shows the distribution of changes in predicted binding energy scores relative to the P51G-m4 energy scores for mutations belonging to each binary category: a positive change in binding score represents an unfavorable effect on binding, and, conversely, a negative change in the score indicates an enhancement in binding.

The substitutions on positions in category (**1,1**) (Figure 7.2) yield a wide range of change in binding energy scores: the tail of the distribution on the positive side reaches nearly a binding score change of 2.0 XEUs and on the negative site values below –0.5 XEUs. Strikingly, the positions in category (**1,1**) yield the most binding enhancing energy scores compared to all other categories, mirroring TEM-1 results. Additionally, the substitutions applied in category (**1,0**) also result in more favorable binding energy scores for dimannose. Mutations in both category (**1,1**) and (**1,0**) present favorable binding energy scores. However, the number of mutations predicted to be enhancing binding in category (**1,1**) is more than those in category (**1,0**) (26% of category (**1,1**) compared to 14% of category (**1,0**)). Interestingly, the mutations in category (**1,0**) that disrupt the binding energy scores is not as strong as category (**1,1**), but similar to category (**0,1**) and (**0,0**). The observed mostly neutral behavior with category (**0,0**) agrees with the same trend obtained with TEM-1 analyses.

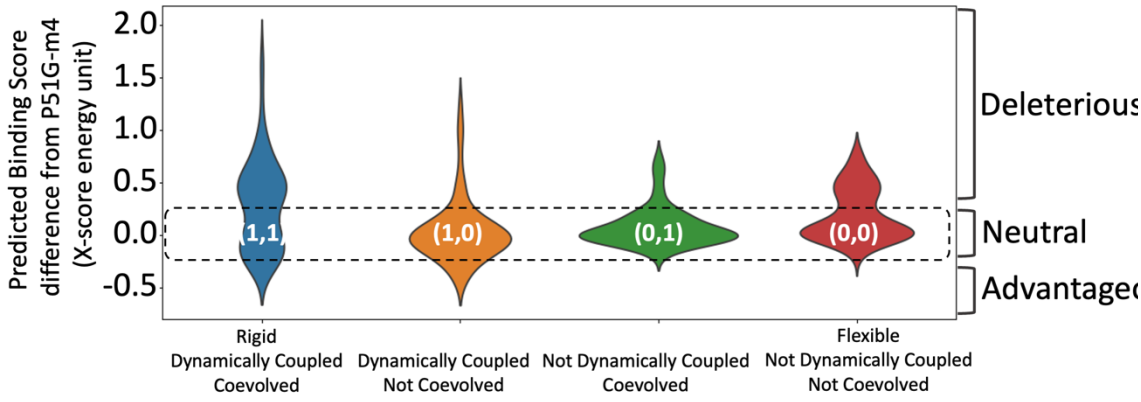


Figure 7.2: Predicted binding energies for each ICDC category. Mutations in category **(1,1)** positions comprise the highest number of binding energy enhancing mutations as well as deleterious mutations. Mutations in category **(0,0)** positions are mostly near neutral (Category **(1,1)** & **(0,0)** *P* value < 0.3).

Overall, the distribution of computational binding scores of dimannose binding to CV-N in each category aligns with the distribution of experimentally characterized TEM-1 fitness results of the same category. However, there are some discrepancies, for example, there are beneficial mutations in category (**0,1**) in TEM-1, but we don't observe the same trend in CV-N. This is due to the initial challenge faced in constructing the MSA of CV-N homologous proteins. There is limited sequence information, and most of the proteins in the CV-N family exhibits binding specificity to a different glycan (Fujimoto and Green, 2012; Koharudin et al., 2009). In contrast, β-lactamase family proteins exhibit highest activity toward penicillin, and they have been subjected to strong natural selection leading to conservation in both fold and function (Salverda et al., 2010; Zou et al., 2021). Hence, the less noise in evolutionary analysis in case of β-lactamase family of proteins allows us to correctly filter deleterious type of substitutions based on the MSA. Regardless, however, in both cases, as hypothesized, substitutions on category (**1,1**) residues impact the function most.

To further investigate the mechanism of functional modulation of category (**1,1**) mutations, we chose the position with highest binding enhancing docking scores, I34, from category (**1,1**). I34 exhibits %DFI values lower than 0.2 (Figure 7.3A), is at least 16 Å away from binding residues (distal), dynamically coupled (Figure 7.3B) and co-evolved with the binding pocket (Tables C.3 and C.5). Moreover, docking scores of I34 variants suggest that the mutations (explained in Appendix C) can modulate binding in a wide range: I34Y variant leads to an increase in binding affinity (beneficial), I34K decreases the binding affinity (deleterious), and I34L yields no change (neutral) (Table 7.1).

To verify the predictions of I34 variants, we first assessed the folding and thermal stability of these mutants by circular dichroism (CD) spectroscopy (Explained in Appendix C). Far-UV CD spectroscopy showed that all mutants are well folded and adopt a fold similar to the parent protein, characterized by spectra with a single negative band centered at 216 nm. We determined the stability of the mutants by CD monitored thermal denaturation; the thermal denaturation curves were analyzed to obtain apparent melting temperature (T_m) values (Explained in Appendix C). We found that the conservative mutation I34L is as stable as P51G-m4, with apparent T_m of 57.8°C and 58°C, respectively. In contrast, I34Y and I34K were less thermostable than P51G-m4 as shown by apparent T_m values of 54.7°C and 47°C, respectively. Not surprisingly, substituting a hydrophobic residue with a basic aliphatic amino acid (lysine) has a large destabilizing effect, while aromatic and polar tyrosine is better tolerated. The trend of thermostability is P51G-m4~I34 L> I34 Y> I34 K (Figure C.1).

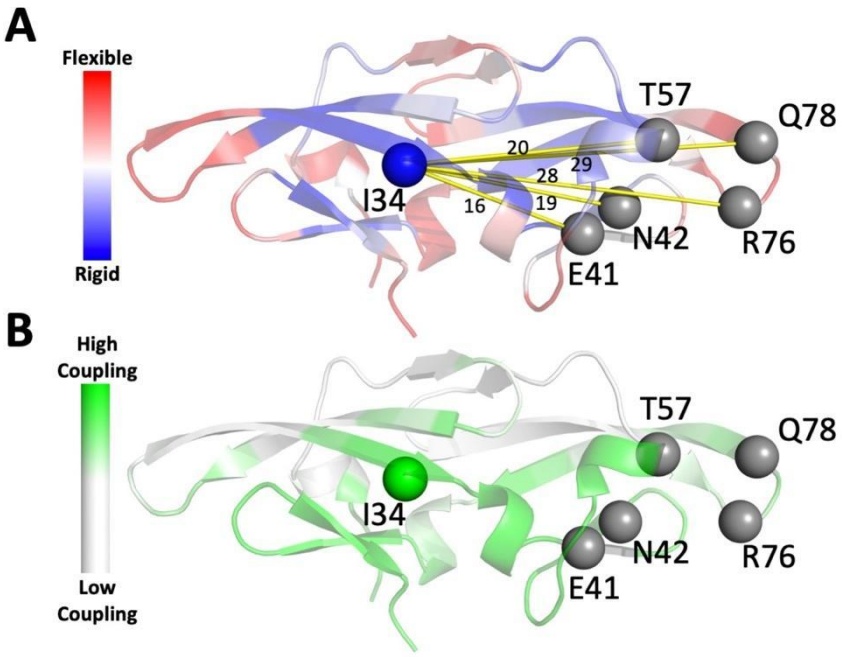


Figure 7.3: DFI and DCI analyses on CV-N. (**A**) Dynamic flexibility index (DFI) profile mapped onto cyanovirin-N (CV-N) structure: red corresponds to high DFI (very flexibile sites), and blue to low DFI values (rigid sites). Position I34 (low DFI score) is highlighted. (**B**) Dynamic coupling index (DCI) profile projected on CV-N structure with green corresponding to sites exhibiting high coupling with binding site residues.

Chemical denaturation experiments (Explained in Appendix C) were used to extract thermodynamic values, after ensuring complete equilibration at each concentration of guanidinium hydrochloride by incubating the samples for 72 hr (Patsalo et al., 2011). The ΔG_{H20} values and C_m values of P51G-m4, I34L, I34Y, and I34K are found as 3.0, 2.94, 2.91, and 2.38 kcal/mol and of 1.45, 1.39, 1.13, and 0.68 M respectively (Table 7.1). The results align with the thermal denaturation results: P51G-m4 is the most stable to denaturant, followed by I34L, I34Y, and I34K (Figure C.2).

Table 7.1: Predicted binding affinities of domain B, experimental ITC data, and chemical denaturation experiments for P51G-m4, and its I34 variants.

Protein	Predicted Binding Score (X-score energy unit)	ITC dimannose K_d (μM)	ITC dimannose ΔH (kcal/mol)	ITC dimannose TΔS (kcal/mol) (T=298K)	ITC dimannose ΔG (kcal/mol)	ΔG_{H2O} (kcal/ mol)	C_m (M)
P51G-m4	-6.62	117 ± 3	-12.3 ± 0.3	-7.00 ± 0.3	-5.30 ± 0.3	3.01 ± 0.047	1.46 ± 0.019
P51G-m4-I34K	-5.85	No-binding	No-binding	No-binding	No-binding	2.40 ± 0.124	0.68 ± 0.015
P51G-m4-I34L	-6.19	148 ± 2	-9.60 ± 0.1	-4.40 ± 0.1	-5.20 ± 0.1	2.95 ± 0.077	1.39 ± 0.009
P51G-m4-I34Y	-6.75	64 ± 5	-4.35 ± 0.1	1.32 ± 0.2	-5.67 ± 0.2	2.91 ± 0.157	1.13 ± 0.017

Next, we evaluated the impact of the mutations on the dimannose binding affinity by isothermal titration calorimetry (ITC) (Explained in Appendix C) (Figure C.3); data were analyzed to extract K_d values listed in Table 7.1. We found that I34Y binds dimannose with tightest affinity (K_d: 64 μM) of all the mutants tested, a two-fold improvement over P51G-m4 (K_d: 117 μM). Binding by I34L is slightly weaker with a K_d of 148 μM. No binding was observed for I34K in these conditions. Thermodynamic values extracted from ITC experiments (Table 7.1), suggesting that entropy changes play an important role in the observed changes in binding affinity: surprisingly, entropy is positive for I34Y, indicating an increase in disorder upon binding.

To glean more information on the mode of binding by I34Y, we determined the X-ray structure (Explained in Appendix C) of the unbound and dimannose-bound form and compared it with the template protein P51G-m4. The fold is highly conserved (Figure 7.4) as shown by main chain RMSD of 0.16 and 0.20 Å with bound and unbound I34Y, respectively, and tyrosine is well tolerated at position I34.

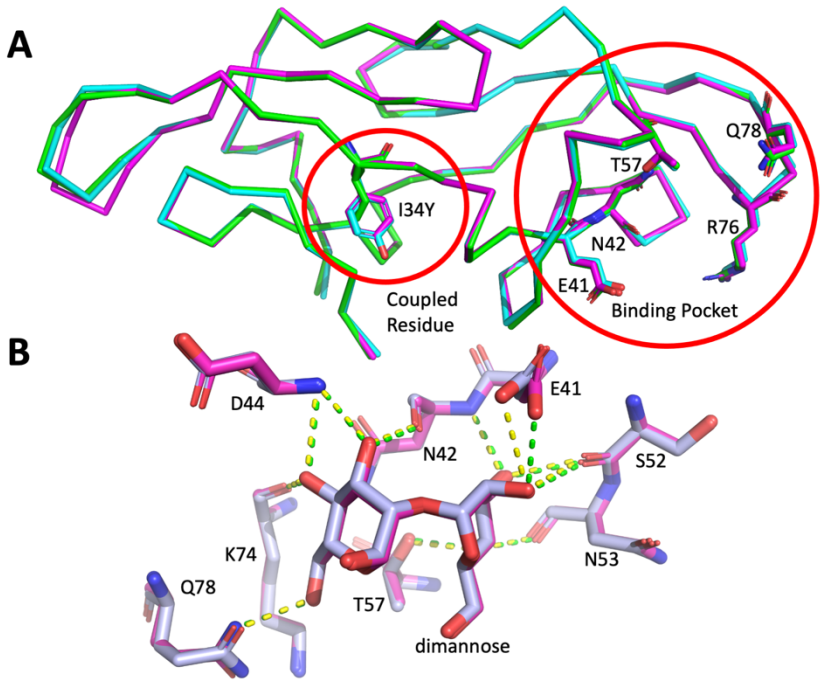


Figure 7.4: The comparison of the crystal structures of P51G-m4 and I34Y. A) The crystal structures of I34Y (bound in magenta and unbound in cyan) and its template protein P51G-m4 (green) are superimposed. B) Overlay of bound structures of I34Y (magenta) and P51G-m4 (grey) (RMSD 0.15 Å); dashed lines depict polar interactions with dimannose.

The binding pocket region is also structurally conserved compared to P51G-m4. Analysis of the polar contacts between dimannose and P51G-m4 and I34Y (Figure 7.4B) shows an identical number of hydrogen bonds (11) with the ligand, indicating a conserved binding pose. We compared the docked pose of I34Y acquired from Adaptive BP-Dock with the bound X-ray structure. The ligand shows an RMSD value of 0.75 Å (Figure C.4). These observations suggest that the increase in binding affinity of I34Y toward dimannose might be mediated by equilibrium dynamics, which are not captured by the crystal structure. This hypothesis is supported by the changes in entropy compensation measured experimentally (ITC) in dimannose binding by P51G-m4 (negative TΔS) and I34Y (positive TΔS).

7.4.3 Molecular Mechanism Governing the Binding Dynamics In I34 Variants

It is interesting to observe that a distal site can modulate binding affinity to a wide range based on amino acid substitutions. This finding has also been observed for allosterically regulated enzymes such as LacI, for which different amino acid substitutions on non-conserved sites lead to gradual changes in function, acting like a rheostatic switch to modulate function through conformational dynamics (Campitelli et al., 2021, 2020; Meinhardt et al., 2013; Miller et al., 2017; Swint-Kruse et al., 1998). To gather atomic level detail on how the substitutions on I34 dynamically modulate the binding affinity, we employed MD simulations in both bound and unbound forms (see Chapter 2 for details of the simulations). The unbound trajectories were analyzed for acquiring binding pocket hydrogen bond distances and pocket volume. Later, to learn about the ligand-induced conformational dynamic changes, the bound trajectories were utilized to estimate computational binding free energies (Deng and Roux, 2009; Okazaki et al., 2006).

Previous computational work in our lab had linked binding affinity in the CV-N family to the accessibility of the binding pocket: A hydrogen bond between the amide hydrogen of N42 and carbonyl oxygen of N53 forms a closed pocket, hindering glycan accessibility, whereas the loss of this hydrogen bond leads to an open pocket (Li et al., 2015). Using the formation of this hydrogen bond in the trajectories of unbound WT and I34Y as metric for assessing open and closed conformations, we found that I34Y variant samples the open binding pocket more often than P51G-m4 (Figure C.5).

Another compelling evidence differentiating I34 variants from P51G-m4 is the change in their binding pocket volumes estimated by POVME pocket volume calculation tool

(Wagner et al., 2017). The calculated pocket volumes for I34Y, I34K, and P51G-m4 were converted into frequencies to obtain probability distributions (Figure 7.5A), revealing that I34Y variant samples a more compact pocket volume compared to P51G-m4. If the pocket is too small or too large, dimannose cannot maximize its interaction with the protein, and a compact conformation enables dimannose to easily make the necessary hydrogen bond interactions with the protein. This optimum pocket volume sampled by I34Y may also explain the different binding energetics observed by ITC, in which a positive entropy change upon binding compensates for the loss in enthalpy compared to P51G-m4 (Table 7.1) (Breiten et al., 2013; Cornish-Bowden, 2002). Pocket volume analysis reveals a larger value for I34K compared to P51G-m4, suggesting that this mutant cannot accommodate the necessary interactions with the dimannose resulting in loss of binding. We applied the same pocket volume calculation to the X-ray structures of P51G-m4 and I34Y variant, and we found volumes of 141 and 114 $Å^3$ for P51G-m4 and I34Y, respectively, in the unbound forms (Figure 7.5B). These volumes correlate well with the mean volumes from MD trajectories, suggesting that the variants modulate the conformational dynamics of binding pocket.

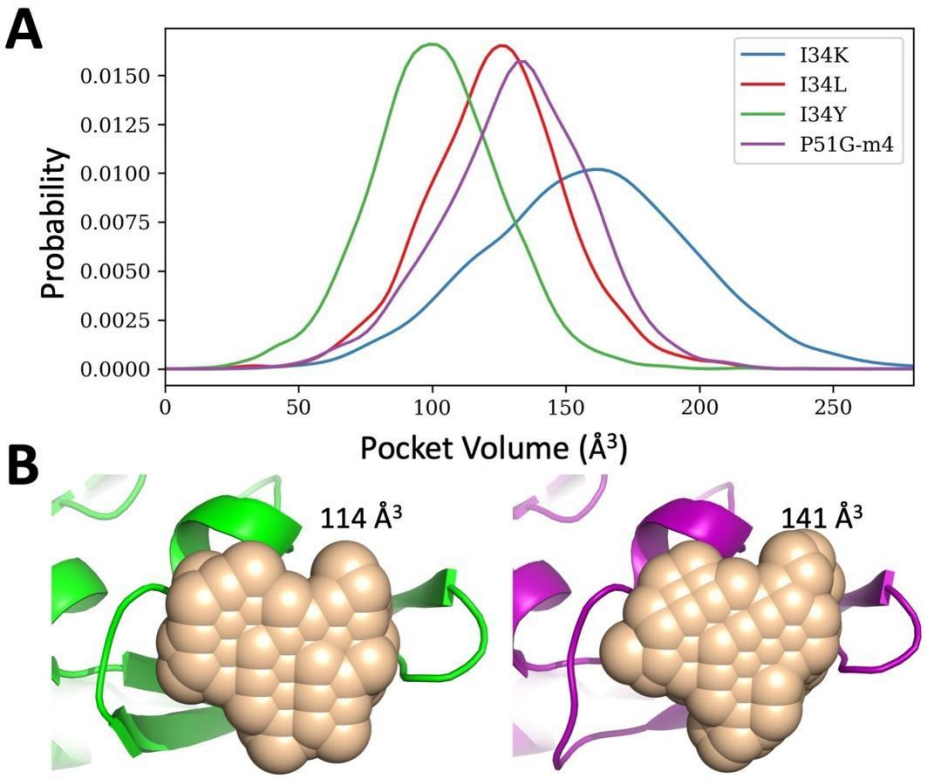


Figure 7.5: Binding pocket volume estimations for P51G-m4 and its variants. A) Probability distribution of the pocket volume analyses obtained from MD simulation trajectories. I34Y populates a conformation with an optimum volume more than others. P51G-m4 and I34L variant sample similar pocket volumes, but I34K variant has a larger pocket volume compared to others. B) Pocket volume comparison of the domain B of solved structures for P51G-m4 (purple) and I34Y variant (green).

Overall, the conformational dynamics analysis of the unbound conformations indicates a shift of the native ensemble toward a smaller pocket volume upon I34Y mutation. This could explain the decrease in the entropic cost of binding observed in ITC results. We also analyzed the binding energetics by carrying out dimannose docking with 2000 different conformations sampled from the binding pocket volume distributions. We found that the small volume restrict accessibility to the side-chain conformations of binding residue R76 in the I34Y variant, yielding different hydrogen bond patterns with the dimannose (Figure C.6) and suggesting a loss in enthalpic contribution.

The bound simulation trajectories were subjected to the MM-PBSA approach to estimate computational binding free energies and related enthalpic and entropic contributions (He et al., 2020; Rastelli et al., 2010). The results are tabulated on Table 7.2. The computed binding free energies capture the trend of experimental binding affinities (R=0.87). The I34Y variant displays a more favorable binding with dimannose compared to wild type. Interestingly, both experimental and computational results show I34Y compensating the enthalpic loss with entropic gain. While I34L variant enthalpic loss is greater than I34Y in computational approach, the overall binding free energy mirrors the ITC results. Additionally, loss of binding of I34K variant overlaps with the ITC data.

Table 7.2: Binding free energies, enthalpy and entropy values for wild type CV-N and its variants calculated with MM-PBSA approach applied on dimannose bound MD simulations.

Protein	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
P51G-m4	-31.13	-18.34	-12.80
I34K	-0.03	-51.07	51.04
I34L	-27.54	-17.97	-9.57
I34Y	-29.76	-15.72	-14.05

* The ΔG scores displayed in this table correlates with experimental binding scores with an R value of 0.87.

7.4.4 Substitutions of I34 Modulates the Conformational Ensemble Leading to Change in Dimannose Binding Affinity

Proteins adapt to a new environment by modulating the native state ensemble through mutations of different positions while keeping the 3D structure conserved (Campitelli et al., 2020; Kuriyan and Eisenberg, 2007; Li et al., 2015; Liu and Nussinov, 2017; Modi and Ozkan, 2018; Risso et al., 2018; Tripathi et al., 2015; Woodrum et al., 2013). As we also observed a similar pattern of conservation of structure yet change in function in our

designed CV-N I34 variants, we further analyzed the flexibility profiles of I34 variants. The DFI profiles clustered using principal component analyses match the 2D map of melting temperature and K_d as reaction coordinates, suggesting a correlation between changes in dynamics and changes in function (Figure 7.6). The 2D map shows I34L, P51G-m4, and I34Y under the same cluster, with I34L and P51G-m4 close, while I34K is markedly different (Figure 7.6A). The dendrogram constructed based on the DFI profiles captures this clustering (Figure 7.6B) with P51G-m4 and I34L variant under the same branch, suggesting their dynamics are very similar; I34Y is under the same main cluster albeit in a different branch. I34K is under a separate branch, indicating different dynamics. This is in agreement with our previous studies, where substitutions on DARC spots modulate binding dynamics reflected in their flexibility profiles to adapt to a new environment (Campitelli et al., 2021; Kumar et al., 2015b; Modi et al., 2021a).

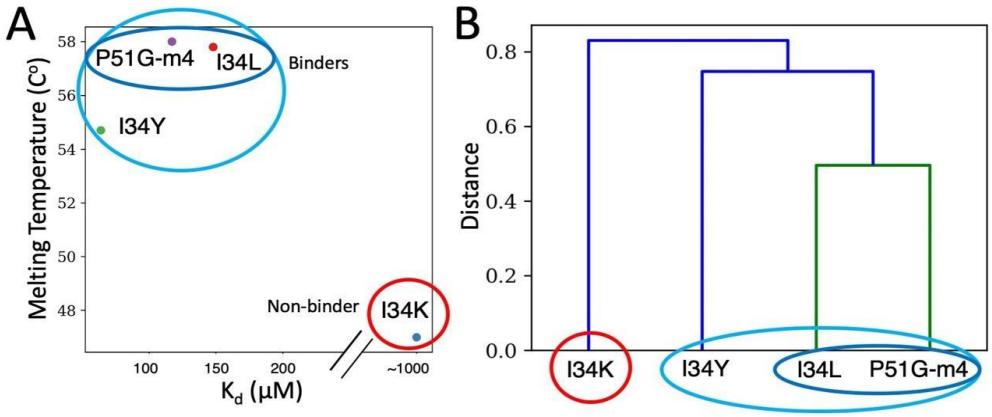


Figure 7.6: Clustering of CV-N variants using DFI profiles and biophysical properties. A) 2D map of K_d and Melting temperature of P51G-m4 and its variants B) PCA clustering on the first two principal components of the DFI profiles as a dendrogram.

We further gleaned a molecular view of the role of flexibility in binding by comparing changes in DFI profiles of the binding site residues with P51G-m4 for each mutant, in the

unbound and bound form (Figure 7.7A and B). We found that flexibility at position T57 is highly dependent on the amino acid at position I34: flexibility increases in I34K, suggesting a higher entropic penalty for binding interactions; It is unchanged in I34L, which has similar binding affinity. In contrast, T57 becomes much more rigid in I34Y mutant. This indicates the rigidification leading to a decrease in the entropic cost can contribute to the binding affinity enhancement of this mutant which is also in agreement with the ITC results.

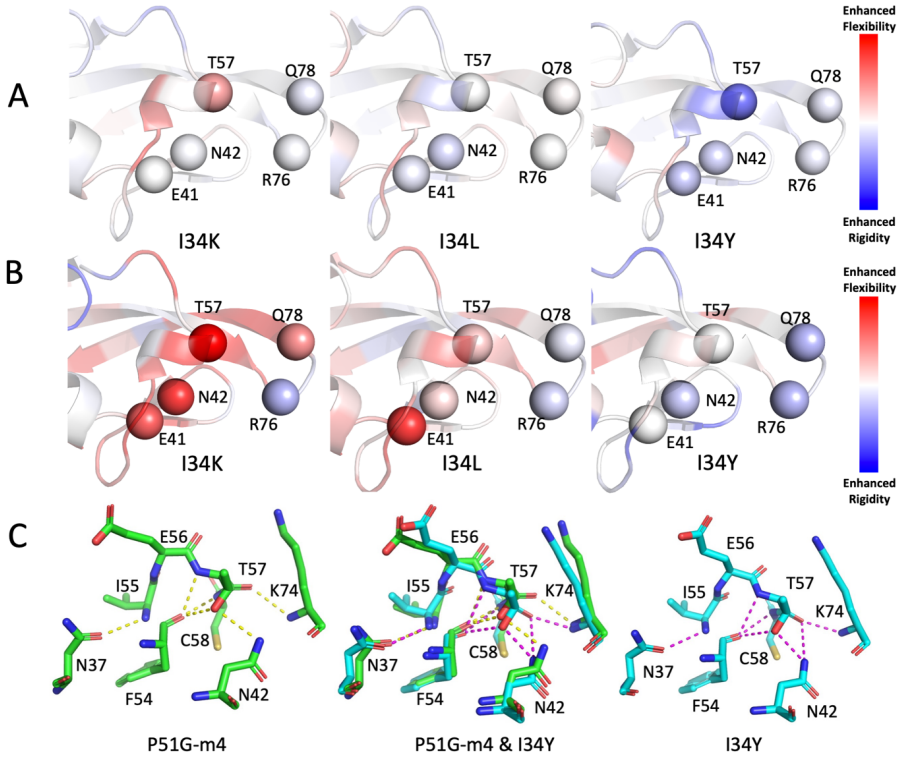


Figure 7.7: Changes in flexibility of the binding site residues upon mutations in bound and unbound forms. A) Change in flexibility of I34K, I34L, and I34Y relative to P51G-m4 in unbound form are shown. Residues E41, N42, and T57 rigidifies on I34Y compared to P51G-m4. B) Change in flexibility of I34K, I34L, and I34Y relative to P51G-m4 in bound form are projected on structure. C) Hydrogen bonding interactions of residues I55, E56, T57, and C58 are shown for P51G-m4 and I34Y variant.

Comparison of the flexibility profiles of the bound form with those of the unbound form reveals that residue I34 in WT drastically gets rigidified upon binding, whereas I34Y variant does not. The decreased flexibility of T57 in the unbound form of I34Y accommodates the interactions with dimannose, contributing to the entropic compensation. In addition to the binding site residues of domain B, the flexibility of the rest of the residues also contributes to the total change in binding free energies. Therefore, we analyzed the correlation between (i) the sum of total change in flexibility of the binding site residues, (ii) the binding site residues and the residues exhibiting highly coupling with the binding pocket, with the experimentally measured binding affinity change. We observe a strong correlation between change in flexibility and change in affinity, as expected I34Y exhibiting tighter binding also gets more rigidified upon binding compared to P51G-m4. Moreover, inclusion of the highly coupled residues in addition to the domain B binding sites in computing the total sum of DFI scores yields a higher correlation with the experimental binding affinity change (Figure C.7A). On the other hand, the correlation between the flexibility change of the randomly selected residues and experimental binding affinities yields poor correlation coefficient (Figure C.7B). These results strongly support the role of dynamic allostery in modulating binding affinity.

The rigidification of T57 in I34Y variant is compelling evidence that the distal mutation is allosterically controlling the binding site dynamics. We further computed the network of interactions that connects the residue position 34–57 and investigated whether distinct pathways emerge after I34Y mutation. We analyzed the hydrogen bond networks, particularly computed the possible network of hydrogen bonds creating pathways from 34

to 57 using the sampled snapshots from the MD trajectories (Figure C.8). This analysis presents a unique pathway from 34 to 57 by first forming a new hydrogen bond between the side chain oxygen of the Tyrosine 34 and the nitrogen of the Tyrosine 100 in I34Y variant. Furthermore, a second pathway is also found which is sampled much more frequently in I34Y variant strengthening the communication between positions 34 and 57.

Thus, both pathways may contribute to the rigidification of T57. We also analyzed the conformations from MD clustered with highest percentage based on alpha carbon RMSD for I34Y and P51G-m4 and compared the hydrogen bond interactions of T57 and its neighboring residues. The closest neighbors of T57; positions I55, E56, and C58 conserved their hydrogen bond interactions with their surrounding residues between P51G-m4 and I34Y. On the other hand, T57 makes an additional hydrogen bond interaction in I34Y compared to P51G-m4 (Figure 7.7C), suggesting that enhancement in hydrogen bond networking of T57 in I34Y leads to rigidification of this position in equilibrium dynamics.

To gain more insight on distal dynamic modulation of binding pocket particularly the decrease of binding site flexibility through distal coupling, we computationally and experimentally characterized another residue, A71, belonging to category (**1,1**) and its mutations: T, S. The docking scores and DFI profiles of A71 variants show high similarity to position I34 ones. The variant A71T is predicted as binding enhancing by our docking scheme displaying a similar binding score as I34Y (A71T predicted binding score: –6.81 XEU), whereas variant A71S is predicted as analogous to I34L variant (A71S predicted binding score: –6.20 XEU). This position is next to residue E72, which is within the

hydrogen bond pathway (Pathway 2) (Figure C.8) identified previously connecting I34 and binding residue T57. Furthermore, the computed binding free energies by MM-PBSA is found to be correlating with position I34 results. The A71T variant shows a binding free energy near I34Y (A71T ΔG: –13.70 kcal/mol with ΔH: –29.33 kcal/mol and TΔS: –15.63 kcal/mol), and A71S close to I34L (A71S ΔG: –9.98 kcal/mol with ΔH: –29.00 kcal/mol and TΔS: –19.02 kcal/mol). All computational analyses suggested that A71 can modulate binding affinity through distal dynamic coupling similar to I34, hence we experimentally characterized these two variants.

The experimental binding affinity by ITC correlates with in silico predictions. When the change in total DFI score upon binding is compared to change in free energy of binding from ITC experiments (Figure C.7), A71T (ΔG: –5.70 kcal/mol with ΔH: –6.00 kcal/mol and TΔS: –0.30) features both a change in total DFI and ΔG closer to I34Y, and A71S (A71S ΔG: –5.10 kcal/mol with ΔH: –9.10 kcal/mol and TΔS: –4.00) shows a score identical to I34L. The entropy of A71T shows a similar change as I34Y experimentally (A71T TΔS: –0.30) indicating that the same compensation mechanism is utilized by another category (**1,1**) residue. A71S is closer to I34L (A71S TΔS: –4.00). Similar to I34Y, the melting temperature of A71T is lower than P51G-m4 (Figure C.1). Results of A71 variants further establish the potential of ICDC and category (**1,1**) residues in diversely tuning the binding affinity of domain B of CV-N through playing enthalpy-entropy compensation of binding process.

Our new ICDC approach suggests that it is possible to identify and incorporate distal mutations into protein design bringing together evolutionary inferences with long-range dynamic communications within the 3D network of interactions.

7.5 Acknowledgements

SBO acknowledges support from the Gordon and Betty Moore Foundations and National Science Foundation (Award: 1715591 and 1901709). This work was supported in part by NIH award 1R21CA207832-01.

CHAPTER 8

FINAL REMARKS

Proteins are highly dynamic in nature, and constantly undergoing variety of motions and conformational changes. One of the reasons for the alteration of protein's dynamic behavior or propagation of conformational changes is an allosteric mutation. Understanding dynamic allostery is important to elucidate the complex regulation of protein function and designing targeted therapeutics. By targeting allosteric sites and exploiting the dynamic properties of proteins, it may be possible to develop drugs that modulate protein function with higher specificity and fewer side effects. In the research conducted for this study, I first employed two protein dynamics analysis tools, namely Dynamic Flexibility Index (DFI) and Dynamic Coupling Index (DCI), which are extensively described in Chapter 2. DFI allows for the quantification of the flexibility level of individual residues, aiding in the comprehension of how mutations on rigid/flexible locations may influence protein function. On the other hand, DCI measures the extent of dynamic coupling of residues distant to functionally significant residues, enabling the identification of crucial residues that may impact protein function when subjected to mutation through dynamic allosteric regulation.

The investigation conducted in Chapter 3 employed DCI and DFI to probe the relationship between the dynamics and the binding of the PICK1 PDZ domain in its interactions with two ligands, DAT and GluR2. Given the substantial complexities associated with PDZ ligand interactions, gaining a comprehensive understanding of the

recognition mechanisms between PDZ and ligands is of utmost importance for therapeutic applications. By comparing the dynamic responses of the PICK1 PDZ domain to binding different ligands, results show that the binding of various ligands can lead to distinct dynamic alterations in the PICK1 PDZ domain. The observations reveal that both ligands elicit dynamic allostery in the αA helix of the PICK1 PDZ domain. Notably, the study identifies the hydrophobic core formed between the ligands and residue I35 as a critical factor in triggering this dynamic allostery.

Understanding the atomic-level mechanism behind the interdomain dynamics of PICK1 PDZ significantly enhances our comprehension of the relationship between dynamic allostery and protein's functionality. In chapter 4, I extended the application of DFI and DCI in understanding enzyme-substrate interactions using Butyrylcholinesterase (BChE) as the model system. BChE showcases its exceptional capacity to neutralize dangerous substances, including paraoxon (an organophosphorous nerve agent), acetylcholine receptor antagonists, and psychoactive plant alkaloids such as cocaine. This characteristic has led to the utilization of BChE in the management of drug overdose and addiction. To facilitate the large-scale production of BChE mutants for clinical applications, the establishment of an easily accessible, economically viable, and environmentally sustainable source of recombinant BChE is of paramount importance. Although mammalian expression systems have been utilized for production of BChE variants targeting cocaine hydrolysis, scaling up such platforms can be challenging and costly. Therefore, plant based recombinant protein production systems are preferred because of reduced production cost, lower down-stream expenses, and the ability of easy

scale-up production. Previously, a close relative of tobacco, *Nicotiana benthamiana*, was utilized to produce large quantities of BChE variants that hold have high catalytic activity to use as treatments. These BChE variants contain multiple mutations. These mutations could be investigated to get a better understanding of how they modulate activity with choline ester substrates and anticholinesterases. Understanding the catalytic mechanism of pBChE is crucial for therapeutic purposes. Therefore, in order to reveal how mutations affect catalytic activity of BChE, I utilized DCI and DFI. The results showed that the mutations impact the activity by modulating the catalytic site dynamics. In addition, BChE functions as a dimer, and the dimerization is shown to connect two catalytic sites and mutations with dynamic allostery, creating a long range modulating effect.

Furthermore, in Chapter 5, I extensively analyzed the complex relationship between mutations and dynamics in the enzyme dihydrofolate reductase (DHFR) using DFI and DCI. I also expanded upon the DCI metric by introducing a new classification technique called DCI_{asym}. This classification, “Controller” / “Controlled”, compared with deep mutational scanning data, not only enabled me to estimate whether point mutations would have beneficial or detrimental effects on function, but it also provided insights into the underlying mechanisms by identifying specific residues that are not conserved throughout evolution but still exert control over the dynamics of functionally critical M20 and FG loops. The results obtained from the comprehensive investigation spanning chapters 3, 4, and 5 have solidified the computational effectiveness of DFI and DCI, establishing their suitability for addressing complex tasks such as enzyme design and protein engineering.

In Chapter 6, my research focused on understanding the significance of dynamics in the utilization of specificity in β-lactamase, particularly the modern TEM-1 variant. I aimed to identify specific residues and mutations, guided by dynamic metrics DFI and DCI, that could potentially enhance the activity of TEM-1 towards penam/cephem antibiotics. The active sites of TEM-1 are located at the core of the protein, presenting a unique challenge. Firstly, mutations on the active sites and surrounding residues have been shown to reduce activity. Secondly, the number of residues distant from the active site that could potentially alter activity is limited. To address these challenges, I employed DFI and DCI to identify these distal residues which are categorized as rigid or flexible, and dynamically coupled and not-coupled. Rigid residues are known to be critical for protein stability and activity and therefore could be target for mutagenesis studies. However, one hurdle in modeling mutations on rigid residues is that such mutations are known to have detrimental effects in TEM-1. Therefore, instead of directly mutating these rigid residues, we adopted an alternative strategy, which involved designing residues around these regions using Rosetta software. We generated two sets of variants based on their design location (rigid or flexible) and coupling with active site (coupled or not-coupled), and subjected them to extensive MD simulations. These simulations of variants were subsequently analyzed using DFI and DCI.

The DFI and DCI analyses applied on MD simulations require using a single value decomposition technique to reduce the number of degrees of freedom of the multifaceted nature of time series dynamics from MD. To achieve this, I used Principal component analysis (PCA) to dissect the dynamic profiles of variants and wild type emerging from the

MD simulations. The results indicate that one such approach has a limitation, which is the number of data points in the input data has a large influence on the PCA components. To tackle this problem, I created a new approach termed “dynamic distance analysis” (DDA). DDA overcomes the requirement of having large number of data points in the input dataset by iteratively calculating PCA components by using subsets of the data. We used DDA analysis to identify the variants based on TEM-1 template that achieve dynamics behavior similar to GNCA. DDA revealed five promising designs. The designs were validated through experimental characterization and demonstrated that MD driven designs have substantial potential in creating variants capable of modulating activity and stability across a wide range.

In chapter 7, I delved into modeling binding characteristics of a small lectin that specifically binds dimannose in order to engineer stable and robust lectins with desired glycan specificity. As a model system, we focused on Cyanovirin-N (CV-N), a lectin consisting of 101 amino acids, which exhibits micromolar binding affinity towards mannose-rich glycans. Specifically, CV-N's highest binding affinity is observed towards di-mannose, making it an ideal target for this study. However, traditional protein design approaches face challenges due to the vast number of potential combinations of residues and amino acid types. Based on insights gained from the application of evolutionary tools, MD simulations, and post-MD dynamics analyses with DFI and DCI, I propose integrated coevolution and dynamic coupling (ICDC) approach for identifying, mutating, and evaluating distal sites to modulate function. Once positions and potential amino acid substitutions are identified using ICDC, the variants are modeled and subjected to MD

simulations. The change in binding affinity relative to the wild type is then modeled using a docking tool developed in house: Adaptive BP-Dock.

The results of binding prediction using Adaptive BP-Dock highlighted a critical residue, I34, which exhibited the potential to enhance, abolish, or have no effect on binding. To gain further insights, I explored the dynamic characteristics of the mutations compared to the wild type. The I34Y mutant displayed enhanced binding behavior. When analyzing the dynamics of the mutant using the DFI metric, it became evident that the I34Y mutation rigidified the dynamics of the binding site residue compared to the wild type. A rigid binding site has the potential to maintain interactions with dimannose more effectively than a flexible binding pocket. Further investigation of the binding pocket revealed that the I34Y mutant had a smaller pocket volume than the wild type, which aligns with the non-mobile behavior observed with DFI. This further confirms that the binding is influenced by this distal mutation. These computational findings were validated through experimental biophysical characterizations. Moreover, the experiments reveal that the entropic cost of binding has changed in this mutant relating back to its change in rigidity, implying that the binding affinity is modulated not by enthalpy but by the change in conformational dynamics. The findings are subjected to computational change in binding free energy calculations with Poisson-Boltzmann Surface Area (MM-PBSA). This estimation showed a similar result to experiments and docking scores, further capturing the enthalpy/entropy compensation behavior seen with experiments. These results, obtained through the investigation and engineering of CV-N, illustrate the power of combining dynamic metrics with MD simulations and co-evolution techniques.

The synergistic integration of computational techniques and the extensive body of work described herein establishes a fertile ground for further development and expansion. Beginning from an investigation of dynamics and its relation to function, an extension to predictive design and computational predictions of activity with docking sets the groundwork for wider applications on many different proteins and enzymes.

This thesis emphasizes that the combined utilization of multiple techniques in relation to both structural, dynamics, and evolutionary side of protein research could complement each other. In future endeavors, the complementarity could be expanded by the introduction of new approaches such as Deep Neural Networks to carefully combine the metrics maintaining the intricate balance in relationship with protein sequence, structure, dynamics, and function. Moreover, in the context of docking methodologies, using an ENM with harmonic spring scaling stemming from amino acid specificity for covariance calculation offers the opportunity to incorporate a broader spectrum of dynamics based conformational sampling. This approach enables the inclusion of a more extensive range of molecular motions and enhances the accuracy of docking predictions. The ongoing development of computational tools, distinguished by their remarkable efficacy in accurately predicting protein binding events and identifying functionally relevant residues, has profound implications for medical research.

REFERENCES

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**:19–25. doi:10.1016/j.softx.2015.06.001

Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S. 2002. Network of coupled promoting motions in enzyme catalysis. *Proceedings of the National Academy of Sciences* **99**:2794–2799. doi:10.1073/pnas.052005999

Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P, Dunbrack RLJr, Das R, Baker D, Kuhlman B, Kortemme T, Gray JJ. 2017. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* **13**:3031–3048. doi:10.1021/acs.jctc.7b00125

Alvarez-Garcia D, Barril X. 2014. Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J Chem Theory Comput* **10**:2608–2614. doi:10.1021/ct500182z

Amemiya T, Koike R, Fuchigami S, Ikeguchi M, Kidera A. 2011. Classification and Annotation of the Relationship between Protein Structural Change and Ligand Binding. *Journal of Molecular Biology* **408**:568–584. doi:10.1016/j.jmb.2011.02.058

Aminov RI. 2010. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Front Microbiol* **1**:134. doi:10.3389/fmicb.2010.00134

André B. 1995. An overview of membrane transport proteins in Saccharomyces cerevisiae. *Yeast* **11**:1575–1611. doi:10.1002/yea.320111605

Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. 2008. Principles of Flexible Protein-Protein Docking. *Proteins* **73**:271–289. doi:10.1002/prot.22170

Argos P, Rao JKM, Hargrave PA. 1982. Structural Prediction of Membrane-Bound Proteins. *European Journal of Biochemistry* **128**:565–575. doi:10.1111/j.1432-1033.1982.tb07002.x

Argos P, Schwarz James, Schwarz John. 1976. An assessment of protein secondary structure prediction methods based on amino acid sequence. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **439**:261–273. doi:10.1016/0005-2795(76)90062-3

Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal* **80**:505–515. doi:10.1016/S0006-3495(01)76033-X

Atilgan C, Atilgan AR. 2009. Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein. *PLOS Computational Biology* **5**:e1000544. doi:10.1371/journal.pcbi.1000544

Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. 2010. Manipulation of Conformational Change in Proteins by Single-Residue Perturbations. *Biophysical Journal* **99**:933–943. doi:10.1016/j.bpj.2010.05.020

Ayrton A, Morgan P. 2001. Role of transport proteins in drug absorption, distribution and excretion. *Xenobiotica* **31**:469–497. doi:10.1080/00498250110060969

Bahar I, Jernigan RL, Dill KA. 2017. Protein Actions: Principles and Modeling. Garland Science.

Balzarini J. 2007. Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nature Reviews Microbiology* **5**:583–597. doi:10.1038/nrmicro1707

Barak D, Ordentlich A, Bromberg A, Kronman C, Marcus D, Lazar A, Ariel N, Velan B, Shafferman A. 1995. Allosteric Modulation of Acetylcholinesterase Activity by Peripheral Ligands Involves a Conformational Transition of the Anionic Subsite. *Biochemistry* **34**:15444–15452. doi:10.1021/bi00047a008

Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM. 2006. Dissecting carbohydrate–Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Engineering, Design and Selection* **19**:525–535. doi:10.1093/protein/gzl040

Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR. 2003. Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Research* **58**:47–56. doi:10.1016/S0166-3542(02)00183-3

Beach H, Cole R, Gill ML, Loria JP. 2005. Conservation of μs−ms Enzyme Motions in the Apo- and Substrate-Mimicked State. *J Am Chem Soc* **127**:9167–9176. doi:10.1021/ja0514949

Beale D, Buttress N. 1969. Studies on a human 19-S immunoglobulin M The arrangement of inter-chain disulphide bridges and carbohydrate sites. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **181**:250–267. doi:10.1016/0005-2795(69)90248-7

Ben Chorin A, Masrati G, Kessel A, Narunsky A, Sprinzak J, Lahav S, Ashkenazy H, Ben-Tal N. 2020. ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci* **29**:258–267. doi:10.1002/pro.3779

Benkovic SJ, Fierke CA, Naylor AM. 1988. Insights into enzyme function from studies on mutants of dihydrofolate reductase. *Science* **239**:1105–1110.

Bepler T, Berger B. 2021. Learning the protein language: Evolution, structure, and function. *Cell Systems* **12**:654-669.e3. doi:10.1016/j.cels.2021.05.017

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**:3684–3690. doi:10.1063/1.448118

Berendsen HJC, van der Spoel D, van Drunen R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**:43–56. doi:10.1016/0010-4655(95)00042-E

Bettati S, Luque FJ, Viappiani C. 2011. Protein dynamics: experimental and computational approaches. *Biochim Biophys Acta* **1814**:913–915. doi:10.1016/j.bbapap.2011.05.003

Bettencourt-Dias M, Giet R, Sinka R, Mazumdar A, Lock WG, Balloux F, Zafiropoulos PJ, Yamaguchi S, Winter S, Carthew RW, Cooper M, Jones D, Frenz L, Glover DM. 2004. Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432**:980–987. doi:10.1038/nature03160

Bewley CA. 2001. Solution Structure of a Cyanovirin-N:Manα1-2Manα Complex: Structural Basis for High-Affinity Carbohydrate-Mediated Binding to gp120. *Structure* **9**:931–940. doi:10.1016/S0969-2126(01)00653-0

Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, Dyson HJ, Godzik A, Wilson IA, Wright PE. 2013. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol* **20**:1243–1249. doi:10.1038/nsmb.2676

Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE. 2011. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**:234–238.

Bienstock RJ. 2012. Computational drug design targeting protein-protein interactions. *Curr Pharm Des* **18**:1240–1254. doi:10.2174/138161212799436449

Binder JL, Berendzen J, Stevens AO, He Y, Wang J, Dokholyan NV, Oprea TI. 2022. AlphaFold illuminates half of the dark human proteins. *Curr Opin Struct Biol* **74**:102372. doi:10.1016/j.sbi.2022.102372

Binder K. 1995. Monte Carlo and Molecular Dynamics Simulations in Polymer Science. Oxford University Press.

Bishop CM. 2006. Pattern Recognition and Machine Learning. Springer.

Bjerggaard C, Fog JU, Hastrup H, Madsen K, Loland CJ, Javitch JA, Gether U. 2004. Surface targeting of the dopamine transporter involves discrete epitopes in the distal C terminus but does not require canonical PDZ domain interactions. *J Neurosci* **24**:7024–7036. doi:10.1523/JNEUROSCI.1863-04.2004

Blake CCF, Johnson LN. 1984. Protein structure. *Trends in Biochemical Sciences* **9**:147–151. doi:10.1016/0968-0004(84)90123-3

Blong MR, Bedows E, Lockridge O. 1997. Tetramerization domain of human butyrylcholinesterase is at the C-terminus. *Biochemical Journal* **327**:747–757. doi:10.1042/bj3270747

Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM. 1987. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**:347–352. doi:10.1038/326347a0

Boeck AT, Schopfer LM, Lockridge O. 2002. DNA sequence of butyrylcholinesterase from the rat: expression of the protein and characterization of the properties of rat butyrylcholinesterase. *Biochemical Pharmacology* **63**:2101–2110. doi:10.1016/S0006-2952(02)01029-8

Boehr DD, McElheny D, Dyson HJ, Wright PE. 2006. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **313**:1638–1642. doi:10.1126/science.1130258

Bolia A, Gerek ZN, Ozkan SB. 2014a. BP-Dock: A Flexible Docking Scheme for Exploring Protein–Ligand Interactions Based on Unbound Structures. *J Chem Inf Model* **54**:913–925. doi:10.1021/ci4004927

Bolia A, Ozkan SB. 2016. Adaptive BP-Dock: An Induced Fit Docking Approach for Full Receptor Flexibility. *J Chem Inf Model* **56**:734–746. doi:10.1021/acs.jcim.5b00587

Bolia A, Woodrum BW, Cereda A, Ruben MA, Wang X, Ozkan SB, Ghirlanda G. 2014b. A Flexible Docking Scheme Efficiently Captures the Energetics of Glycan-Cyanovirin Binding. *Biophys J* **106**:1142–1151. doi:10.1016/j.bpj.2014.01.040

Bolmstedt AJ, O'Keefe BR, Shenoy SR, McMahon JB, Boyd MR. 2001. Cyanovirin-N Defines a New Class of Antiviral Agent Targeting N-Linked, High-Mannose Glycans in an Oligosaccharide-Specific Manner. *Mol Pharmacol* **59**:949–954. doi:10.1124/mol.59.5.949

Bordin N, Sillitoe I, Lees JG, Orengo C. 2021. Tracing Evolution Through Protein Structures: Nature Captured in a Few Thousand Folds. *Frontiers in Molecular Biosciences* **8**.

Botos I, Wlodawer A. 2005. Proteins that bind high-mannose sugars of the HIV envelope. *Progress in Biophysics and Molecular Biology*, Structure-guided design of AIDs Antivirals **88**:233–282. doi:10.1016/j.pbiomolbio.2004.05.001

Botos I, Wlodawer A. 2003. Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *CMLS, Cell Mol Life Sci* **60**:277–287. doi:10.1007/s000180300023

Boulanger LM. 2009. Immune Proteins in Brain Development and Synaptic Plasticity. *Neuron* **64**:93–109. doi:10.1016/j.neuron.2009.09.001

Bowman GR, Geissler PL. 2012. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of Sciences* **109**:11681–11686.

Bozovic O, Zanobini C, Gulzar A, Jankovic B, Buhrke D, Post M, Wolf S, Stock G, Hamm P. 2020. Real-time observation of ligand-induced allosteric transitions in a PDZ domain. *Proc Natl Acad Sci U S A* **117**:26031–26039. doi:10.1073/pnas.2012999117

Bradley P, Misura KMS, Baker D. 2005. Toward high-resolution de novo structure prediction for small proteins. *Science* **309**:1868–1871. doi:10.1126/science.1113801

Brakeman PR, Lanahan AA, O'Brien R, Roche K, Barnes CA, Huganir RL, Worley PF. 1997. Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**:284–288. doi:10.1038/386284a0

Brandt C, Braun SD, Stein C, Slickers P, Ehricht R, Pletz MW, Makarewicz O. 2017. In silico serine β-lactamases analysis reveals a huge potential resistome in environmental and pathogenic species. *Sci Rep* **7**:43232. doi:10.1038/srep43232

Brant DA, Flory PJ. 2002. The Configuration of Random Polypeptide Chains. II. Theory. *ACS Publications*. doi:10.1021/ja01091a003

Breiten B, Lockett MR, Sherman W, Fujita S, Al-Sayah M, Lange H, Bowers CM, Heroux A, Krilov G, Whitesides GM. 2013. Water Networks Contribute to Enthalpy/Entropy Compensation in Protein–Ligand Binding. *J Am Chem Soc* **135**:15579–15584. doi:10.1021/ja4075776

Brimijoin S, Shen ML, Sun H. 2002. Radiometric solvent-partitioning assay for screening cocaine hydrolases and measuring cocaine levels in milligram tissue samples. *Analytical Biochemistry* **309**:200–205. doi:10.1016/S0003-2697(02)00238-5

Brooijmans N, Kuntz ID. 2003. Molecular Recognition and Docking Algorithms. *Annual Review of Biophysics and Biomolecular Structure* **32**:335–373. doi:10.1146/annurev.biophys.32.110601.142532

Brown CA, Hu L, Sun Z, Patel MP, Singh S, Porter JR, Sankaran B, Prasad BVV, Bowman GR, Palzkill T. 2020. Antagonism between substitutions in β-lactamase explains a path not taken in the evolution of bacterial drug resistance. *Journal of Biological Chemistry* **295**:7376–7390. doi:10.1074/jbc.RA119.012489

Bush K. 2018. Past and Present Perspectives on β-Lactamases. *Antimicrobial Agents and Chemotherapy* **62**:10.1128/aac.01076-18. doi:10.1128/aac.01076-18

Butler BM, Gerek ZN, Kumar S, Ozkan SB. 2015. Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association. *Proteins* **83**:428–435. doi:10.1002/prot.24748

Butler BM, Kazan IC, Kumar A, Ozkan SB. 2018. Coevolving residues inform protein dynamics profiles and disease susceptibility of nSNVs. *PLoS computational biology* **14**:e1006626.

Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics* **20**:341–351. doi:10.1002/cem.1006

Cammarata MB, Thyer R, Rosenberg J, Ellington A, Brodbelt JS. 2015. Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry. *J Am Chem Soc* **137**:9128–9135. doi:10.1021/jacs.5b04628

Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, Tokuriki N, Jackson CJ. 2016. The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol* **12**:944–950. doi:10.1038/nchembio.2175

Campitelli P, Guo J, Zhou H-X, Ozkan SB. 2018. Hinge-Shift Mechanism Modulates Allosteric Regulations in Human Pin1. *J Phys Chem B* **122**:5623–5629. doi:10.1021/acs.jpcb.7b11971

Campitelli P, Modi T, Kumar S, Ozkan SB. 2020. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annual Review of Biophysics* **49**:267–288. doi:10.1146/annurev-biophys-052118-115517

Campitelli P, Ozkan SB. 2020. Allostery and Epistasis: Emergent Properties of Anisotropic Networks. *Entropy* **22**:667. doi:10.3390/e22060667

Campitelli P, Swint-Kruse L, Ozkan SB. 2021. Substitutions at Nonconserved Rheostat Positions Modulate Function by Rewiring Long-Range, Dynamic Interactions. *Mol Biol Evol* **38**:201–214. doi:10.1093/molbev/msaa202

Cantor CR, Schimmel PR. 1980. Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function. Macmillan.

Cao H, Wang J, He L, Qi Y, Zhang JZ. 2019. DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. *J Chem Inf Model* **59**:1508–1514. doi:10.1021/acs.jcim.8b00697

Cardozo T, Totrov M, Abagyan R. 1995. Homology modeling by the ICM method. *Proteins: Structure, Function, and Bioinformatics* **23**:403–414. doi:10.1002/prot.340230314

Carmona GN, Jufer RA, Goldberg SR, Gorelick DA, Greig NH, Yu QS, Cone EJ, Schindler CW. 2000. Butyrylcholinesterase accelerates cocaine metabolism: in vitro and in vivo effects in nonhuman primates and humans. *Drug Metab Dispos* **28**:367–371.

Carugo O, Djinović-Carugo K. 2013. Half a century of Ramachandran plots. *Acta Cryst D* **69**:1333–1341. doi:10.1107/S090744491301158X

Cavasotto CN, Abagyan RA. 2004. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *Journal of Molecular Biology* **337**:209–225. doi:10.1016/j.jmb.2004.01.003

Chakrabarty B, Parekh N. 2016. NAPS: Network Analysis of Protein Structures. *Nucleic Acids Research* **44**:W375–W382. doi:10.1093/nar/gkw383

Chakrabarty B, Parekh N. 2014. PRIGSA: Protein repeat identification by graph spectral analysis. *J Bioinform Comput Biol* **12**:1442009. doi:10.1142/S0219720014420098

Chan HS, Dill KA. 1990. Origins of structure in globular proteins. *Proc Natl Acad Sci U S A* **87**:6388–6392.

Chandrasekaran R, Balasubramanian R. 1969. Stereochemical studies of cyclic peptides: VI. Energy calculations of the cyclic disulphide cysteinylcysteine. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **188**:1–9. doi:10.1016/0005-2795(69)90039-7

Chang LC, Bewley CA. 2002. Potent Inhibition of HIV-1 Fusion by Cyanovirin-N Requires Only a Single High Affinity Carbohydrate Binding Site: Characterization of Low Affinity Carbohydrate Binding Site Knockout Mutants. *Journal of Molecular Biology* **318**:1–8. doi:10.1016/S0022-2836(02)00045-1

Chaudhury S, Gray JJ. 2008. Conformer Selection and Induced Fit in Flexible Backbone Protein–Protein Docking Using Computational and NMR Ensembles. *Journal of Molecular Biology* **381**:1068–1087. doi:10.1016/j.jmb.2008.05.042

Chen J, Wang X, Pang L, Zhang JZH, Zhu T. 2019. Effect of mutations on binding of ligands to guanine riboswitch probed by free energy perturbation and molecular dynamics simulations. *Nucleic Acids Research* **47**:6618–6631. doi:10.1093/nar/gkz499

Chen Q, Niu X, Xu Y, Wu J, Shi Y. 2007. Solution structure and backbone dynamics of the AF-6 PDZ domain/Bcr peptide complex. *Protein Sci* **16**:1053–1062. doi:10.1110/ps.062440607

Chen VP, Gao Y, Geng L, Parks RJ, Pang Y-P, Brimijoin S. 2015. Plasma butyrylcholinesterase regulates ghrelin to control aggression. *Proc Natl Acad Sci U S A* **112**:2251–2256. doi:10.1073/pnas.1421536112

Chen X, Fang L, Liu J, Zhan C-G. 2012. Reaction pathway and free energy profiles for butyrylcholinesterase-catalyzed hydrolysis of acetylthiocholine. *Biochemistry* **51**:1297–1305. doi:10.1021/bi201786s

Chen X, Huang X, Geng L, Xue L, Hou S, Zheng X, Brimijoin S, Zheng F, Zhan C-G. 2015. Kinetic characterization of a cocaine hydrolase engineered from mouse butyrylcholinesterase. *Biochem J* **466**:243–251. doi:10.1042/BJ20141266

Chen X, Xue L, Hou S, Jin Z, Zhang T, Zheng F, Zhan C-G. 2016. Long-acting cocaine hydrolase for addiction therapy. *Proc Natl Acad Sci U S A* **113**:422–427. doi:10.1073/pnas.1517713113

Chi CN, Elfström L, Shi Y, Snäll T, Engström A, Jemth P. 2008. Reassessing a sparse energetic network within a single protein domain. *Proc Natl Acad Sci U S A* **105**:4679–4684. doi:10.1073/pnas.0711732105

Chodera JD, Mobley DL. 2013. Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu Rev Biophys* **42**:121–142. doi:10.1146/annurev-biophys-083012-130318

Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. 2011. Alchemical free energy methods for drug discovery: progress and challenges. *Current Opinion in Structural Biology* **21**:150–160. doi:10.1016/j.sbi.2011.01.011

Chou PY, Fasman GD. 1974. Prediction of protein conformation. *Biochemistry* **13**:222–245. doi:10.1021/bi00699a002

Christensen NR, Čalyševa J, Fernandes EFA, Lüchow S, Clemmensen LS, Haugaard-Kedström LM, Strømgaard K. 2019. PDZ Domains as Drug Targets. *Adv Ther (Weinh)* **2**:1800143. doi:10.1002/adtp.201800143

Cid H, Bunster M, Canales M, Gazitúa F. 1992. Hydrophobicity and structural classes in proteins. *Protein Engineering, Design and Selection* **5**:373–375. doi:10.1093/protein/5.5.373

Cilia E, Vuister GW, Lenaerts T. 2012. Accurate prediction of the dynamical changes within the second PDZ domain of PTP1e. *PLoS Comput Biol* **8**:e1002794. doi:10.1371/journal.pcbi.1002794

Cohen FE, Kuntz ID. 1989. Tertiary Structure Prediction In: Fasman GD, editor. Prediction of Protein Structure and the Principles of Protein Conformation. Boston, MA: Springer US. pp. 647–705. doi:10.1007/978-1-4613-1571-1_17

Cohen FE, Sternberg MJE, Taylor WR. 1980. Analysis and prediction of protein β-sheet structures by a combinatorial approach. *Nature* **285**:378–382. doi:10.1038/285378a0

Connors NJ, Hoffman RS. 2013. Experimental treatments for cocaine toxicity: a difficult transition to the bedside. *J Pharmacol Exp Ther* **347**:251–257. doi:10.1124/jpet.113.206383

Cooper A, Dryden DT. 1984. Allostery without conformational change. A plausible model. *Eur Biophys J* **11**:103–109. doi:10.1007/BF00276625

Cornish-Bowden A. 2002. Enthalpy—entropy compensation: a phantom phenomenon. *Journal of Biosciences* **27**:121–126.

Cortina GA, Hays JM, Kasson PM. 2018. Conformational Intermediate That Controls KPC-2 Catalysis and Beta-Lactam Drug Resistance. *ACS Catal* **8**:2741–2747. doi:10.1021/acscatal.7b03832

Cortina GA, Kasson PM. 2018. Predicting allostery and microbial drug resistance with molecular simulations. *Current Opinion in Structural Biology*, Cryo electron microscopy: the impact of the cryo-EM revolution in biology • Biophysical and computational methods - Part A **52**:80–86. doi:10.1016/j.sbi.2018.09.001

Costescu BI, Gräter F. 2013. Time-resolved force distribution analysis. *BMC Biophys* **6**:5. doi:10.1186/2046-1682-6-5

Coulson A. 1985. Beta-lactamases: molecular studies. *Biotechnol Genet Eng Rev* **3**:219–253. doi:10.1080/02648725.1985.10647814

Cournia Z, Allen B, Sherman W. 2017. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J Chem Inf Model* **57**:2911–2937. doi:10.1021/acs.jcim.7b00564

Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA. 2008. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J Med Chem* **51**:6237–6255. doi:10.1021/jm800562d

Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. 2005. Comparison of Automated Docking Programs as Virtual Screening Tools. *J Med Chem* **48**:962–976. doi:10.1021/jm049798d

Daniel RM, Dunn RV, Finney JL, Smith JC. 2003. The Role of Dynamics in Enzyme Activity. *Annual Review of Biophysics and Biomolecular Structure* **32**:69–92. doi:10.1146/annurev.biophys.32.110601.142445

Davies J, Davies D. 2010. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews* **74**:417–433. doi:10.1128/MMBR.00016-10

Davis IW, Baker D. 2009. RosettaLigand Docking with Full Ligand and Receptor Flexibility. *Journal of Molecular Biology* **385**:381–392. doi:10.1016/j.jmb.2008.11.010

Davis IW, Raha K, Head MS, Baker D. 2009. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Science* **18**:1998–2002. doi:10.1002/pro.192

de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. *Nat Rev Genet* **14**:249–261. doi:10.1038/nrg3414

De Los Rios P, Cecconi F, Pretre A, Dietler G, Michielin O, Piazza F, Juanico B. 2005. Functional dynamics of PDZ binding domains: a normal-mode analysis. *Biophys J* **89**:14–21. doi:10.1529/biophysj.104.055004

Decker M. 2005. Novel inhibitors of acetyl- and butyrylcholinesterase derived from the alkaloids dehydroevodiamine and rutaecarpine. *European Journal of Medicinal Chemistry* **40**:305–313. doi:10.1016/j.ejmech.2004.12.003

del Sol A, Fujihashi H, Amoros D, Nussinov R. 2006. Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. *Protein Science* **15**:2120–2128. doi:10.1110/ps.062249106

del Sol A, O'Meara P. 2005. Small-world network approach to identify key residues in protein–protein interaction. *Proteins: Structure, Function, and Bioinformatics* **58**:672–682. doi:10.1002/prot.20348

DeLuca S, Khar K, Meiler J. 2015. Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS One* **10**. doi:10.1371/journal.pone.0132508

Deng Y, Roux B. 2009. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J Phys Chem B* **113**:2234–2246. doi:10.1021/jp807701h

Dev KK, Nishimune A, Henley JM, Nakanishi S. 1999. The protein kinase C alpha binding protein PICK1 interacts with short but not long form alternative splice variants of AMPA receptor subunits. *Neuropharmacology* **38**:635–644. doi:10.1016/s0028-3908(98)00230-5

Dhulesia A, Gsponer J, Vendruscolo M. 2008. Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *J Am Chem Soc* **130**:8931–8939. doi:10.1021/ja0752080

Dill K, Bromberg S. 2010. Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience. Garland Science.

Dill KA, MacCallum JL. 2012. The Protein-Folding Problem, 50 Years On. *Science* **338**:1042–1046. doi:10.1126/science.1219021

Dill KA, Ozkan SB, Shell MS, Weikl TR. 2008. The Protein Folding Problem. *Annual Review of Biophysics* **37**:289–316. doi:10.1146/annurev.biophys.37.092707.153558

Ding F, Dokholyan NV. 2013. Incorporating Backbone Flexibility in MedusaDock Improves Ligand-Binding Pose Prediction in the CSAR2011 Docking Benchmark. *J Chem Inf Model* **53**:1871–1879. doi:10.1021/ci300478y

Ding Y, Tang J, Guo F. 2016. Predicting protein-protein interactions via multivariate mutual information of protein sequences. *BMC Bioinformatics* **17**:398. doi:10.1186/s12859-016-1253-9

Doctor BP, Saxena A. 2005. Bioscavengers for the protection of humans against organophosphate toxicity. *Chemico-Biological Interactions*, Proceedings of the VIII International Meeting on Cholinesterases **157–158**:167–171. doi:10.1016/j.cbi.2005.10.024

Dominguez C, Boelens R, Bonvin AMJJ. 2003. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J Am Chem Soc* **125**:1731–1737. doi:10.1021/ja026939x

Dominguez R, Holmes KC. 2011. Actin structure and function. *Annu Rev Biophys* **40**:169–186. doi:10.1146/annurev-biophys-042910-155359

Doucet N, Savard P-Y, Pelletier JN, Gagné SM. 2007. NMR investigation of Tyr105 mutants in TEM-1 beta-lactamase: dynamics are correlated with function. *J Biol Chem* **282**:21448–21459. doi:10.1074/jbc.M609777200

Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. 1996. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85**:1067–1076. doi:10.1016/s0092-8674(00)81307-0

Drenth J. 2007. Principles of Protein X-Ray Crystallography. Springer Science & Business Media.

Du Q-S, Wang C-H, Liao S-M, Huang R-B. 2010. Correlation analysis for protein evolutionary family based on amino acid position mutations and application in PDZ domain. *PLoS One* **5**:e13207. doi:10.1371/journal.pone.0013207

Duan Y, Kollman PA. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**:740–744. doi:10.1126/science.282.5389.740

Dunn S.D., Wahl LM, Gloor GB. 2008. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* **24**:333–340. doi:10.1093/bioinformatics/btm604

Durrant JD, McCammon JA. 2011. Molecular dynamics simulations and drug discovery. *BMC Biology* **9**:71. doi:10.1186/1741-7007-9-71

Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. 2000. Protein function in the post-genomic era. *Nature* **405**:823–826. doi:10.1038/35015694

Emsley JW, Feeney J, Sutcliffe LH. 2013. High Resolution Nuclear Magnetic Resonance Spectroscopy: Volume 2. Elsevier.

Epstein DM, Benkovic SJ, Wright PE. 1995. Dynamics of the Dihydrofolate Reductase-Folate Complex: Catalytic Sites and Regions Known To Undergo Conformational Change Exhibit Diverse Dynamical Features. *Biochemistry* **34**:11037–11048. doi:10.1021/bi00035a009

Erlendsson S, Rathje M, Heidarsson PO, Poulsen FM, Madsen KL, Teilum K, Gether U. 2014. Protein interacting with C-kinase 1 (PICK1) binding promiscuity relies on unconventional PSD-95/discs-large/ZO-1 homology (PDZ) binding modes for nonclass II PDZ ligands. *J Biol Chem* **289**:25327–25340. doi:10.1074/jbc.M114.548743

Essiz SG, Coalson RD. 2009. Dynamic Linear Response Theory for Conformational Relaxation of Proteins. *J Phys Chem B* **113**:10859–10869. doi:10.1021/jp900745u

Evron T, Geyer BC, Cherni I, Muralidharan M, Kilbourne J, Fletcher SP, Soreq H, Mor TS. 2007. Plant-derived human acetylcholinesterase-R provides protection from lethal organophosphate poisoning and its chronic aftermath. *The FASEB Journal* **21**:2961–2969. doi:10.1096/fj.07-8112com

Fair RJ, Tor Y. 2014. Antibiotics and bacterial resistance in the 21st century. *Perspect Medicin Chem* **6**:25–64. doi:10.4137/PMC.S14459

Felder CE, Harel M, Silman I, Sussman JL. 2002. Structure of a complex of the potent and specific inhibitor BW284C51 with Torpedo californica acetylcholinesterase. *Acta Cryst D* **58**:1765–1771. doi:10.1107/S0907444902011642

Feller SE, Zhang Y, Pastor RW, Brooks BR. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **103**:4613–4621. doi:10.1063/1.470648

Ferreiro DU, Hegler JA, Komives EA, Wolynes PG. 2007. Localizing frustration in native proteins and protein assemblies. *Proceedings of the National Academy of Sciences* **104**:19819–19824. doi:10.1073/pnas.0709915104

Fitzjohn PW, Bates PA. 2003. Guided docking: First step to locate potential binding sites. *Proteins: Structure, Function, and Bioinformatics* **52**:28–32. doi:10.1002/prot.10380

Fowler NJ, Sljoka A, Williamson MP. 2020. A method for validating the accuracy of NMR protein structures. *Nat Commun* **11**:6321. doi:10.1038/s41467-020-20177-1

Fox JM, Zhao M, Fink MJ, Kang K, Whitesides GM. 2018. The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition. *Annual Review of Biophysics* **47**:223–250. doi:10.1146/annurev-biophys-070816-033743

Froimowitz M, Fasman GD. 1974. Prediction of the Secondary Structure of Proteins Using the Helix-Coil Transition Theory. *Macromolecules* **7**:583–589. doi:10.1021/ma60041a009

Fromme R, Katiliene Z, Fromme P, Ghirlanda G. 2008. Conformational gating of dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å resolution. *Protein Science* **17**:939–944. doi:10.1110/ps.083472808

Fromme R, Katiliene Z, Giomarelli B, Bogani F, Mc Mahon J, Mori T, Fromme P, Ghirlanda G. 2007. A Monovalent Mutant of Cyanovirin-N Provides Insight into the Role of Multiple Interactions with gp120 for Antiviral Activity,. *Biochemistry* **46**:9199–9207. doi:10.1021/bi700666m

Fuentes EJ, Der CJ, Lee AL. 2004. Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J Mol Biol* **335**:1105–1115. doi:10.1016/j.jmb.2003.11.010

Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. 2006. Evaluation of Energetic and Dynamic Coupling Networks in a PDZ Domain Protein. *Journal of Molecular Biology* **364**:337–351. doi:10.1016/j.jmb.2006.08.076

Fujimoto YK, Green DF. 2012. Carbohydrate Recognition by the Antiviral Lectin Cyanovirin-N. *J Am Chem Soc* **134**:19639–19651. doi:10.1021/ja305755b

Fuxreiter M. 2014. Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods. CRC Press.

Gao D, Cho H, Yang W, Pan Y, Yang G, Tai H-H, Zhan C-G. 2006. Computational design of a human butyrylcholinesterase mutant for accelerating cocaine hydrolysis based on the transition-state simulation. *Angew Chem Int Ed Engl* **45**:653–657. doi:10.1002/anie.200503025

Gao D, Zhan C-G. 2006. Modeling evolution of hydrogen bonding and stabilization of transition states in the process of cocaine hydrolysis catalyzed by human butyrylcholinesterase. *Proteins* **62**:99–110. doi:10.1002/prot.20713

García de la Torre J, Huertas ML, Carrasco B. 2000. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. *Biophysical Journal* **78**:719–730. doi:10.1016/S0006-3495(00)76630-6

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server In: Walker JM, editor. The Proteomics Protocols Handbook, Springer Protocols Handbooks. Totowa, NJ: Humana Press. pp. 571–607. doi:10.1385/1-59259-890-0:571

Gekko K, Kamiyama T, Ohmae E, Katayanagi K. 2000. Single Amino Acid Substitutions in Flexible Loops Can Induce Large Compressibility Changes in Dihydrofolate Reductase1. *The Journal of Biochemistry* **128**:21–27. doi:10.1093/oxfordjournals.jbchem.a022726

Geng H, Chen F, Ye J, Jiang F. 2019. Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Comput Struct Biotechnol J* **17**:1162–1170. doi:10.1016/j.csbj.2019.07.010

Gerek ZN, Keskin O, Ozkan SB. 2009. Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior. *Proteins* **77**:796–811. doi:10.1002/prot.22492

Gerek ZN, Kumar S, Ozkan SB. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evolutionary Applications* **6**:423–433. doi:10.1111/eva.12052

Gerek ZN, Ozkan SB. 2011. Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning. *PLOS Computational Biology* **7**:e1002154. doi:10.1371/journal.pcbi.1002154

Gerek ZN, Ozkan SB. 2010. A flexible docking scheme to explore the binding selectivity of PDZ domains. *Protein Science* **19**:914–928. doi:10.1002/pro.366

Geyer BC, Fletcher SP, Griffin TA, Lopker MJ, Soreq H, Mor TS. 2007. Translational control of recombinant human acetylcholinesterase accumulation in plants. *BMC Biotechnology* **7**:27. doi:10.1186/1472-6750-7-27

Geyer BC, Kannan L, Cherni I, Woods RR, Soreq H, Mor TS. 2010a. Transgenic plants as a source for the bioscavenging enzyme, human butyrylcholinesterase. *Plant Biotechnol J* **8**:873–886. doi:10.1111/j.1467-7652.2010.00515.x

Geyer BC, Kannan L, Garnaud P-E, Broomfield CA, Cadieux CL, Cherni I, Hodgins SM, Kasten SA, Kelley K, Kilbourne J, Oliver ZP, Otto TC, Puffenberger I, Reeves TE, Robbins N, Woods RR, Soreq H, Lenz DE, Cerasoli DM, Mor TS. 2010b. Plant-derived human butyrylcholinesterase, but not an organophosphorous-compound hydrolyzing variant thereof, protects rodents against nerve agents. *Proc Natl Acad Sci U S A* **107**:20251–20256. doi:10.1073/pnas.1009021107

Geyer BC, Muralidharan M, Cherni I, Doran J, Fletcher SP, Evron T, Soreq H, Mor TS. 2005. Purification of transgenic plant-derived recombinant human acetylcholinesterase-R. *Chemico-Biological Interactions*, Proceedings of the VIII International Meeting on Cholinesterases **157–158**:331–334. doi:10.1016/j.cbi.2005.10.097

Geyer BC, Woods RR, Mor TS. 2008. Increased organophosphate scavenging in a butyrylcholinesterase mutant. *Chemico-Biological Interactions*, Proceedings of the IX International Meeting on Cholinesterases **175**:376–379. doi:10.1016/j.cbi.2008.04.012

Gianni S, Haq SR, Montemiglio LC, Jürgens MC, Engström Å, Chi CN, Brunori M, Jemth P. 2011. Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity. *J Biol Chem* **286**:27167–27175. doi:10.1074/jbc.M111.239541

Gianni S, Walma T, Arcovito A, Calosci N, Bellelli A, Engström A, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW. 2006. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure* **14**:1801–1809. doi:10.1016/j.str.2006.10.010

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**:343–345. doi:10.1038/nmeth.1318

Gilson MK, Zhou H-X. 2007. Calculation of Protein-Ligand Binding Affinities. *Annual Review of Biophysics and Biomolecular Structure* **36**:21–42. doi:10.1146/annurev.biophys.36.040306.132550

Gloor GB, Martin LC, Wahl LM, Dunn SD. 2005. Mutual Information in Protein Multiple Sequence Alignments Reveals Two Classes of Coevolving Positions. *Biochemistry* **44**:7156–7165. doi:10.1021/bi050293e

Go N. 1983. Theoretical Studies of Protein Folding. *Annual Review of Biophysics and Bioengineering* **12**:183–210. doi:10.1146/annurev.bb.12.060183.001151

Gobeil SMC, Ebert MCCJC, Park J, Gagné D, Doucet N, Berghuis AM, Pleiss J, Pelletier JN. 2019. The Structural Dynamics of Engineered β-Lactamases Vary Broadly on Three Timescales yet Sustain Native Function. *Sci Rep* **9**:6656. doi:10.1038/s41598-019-42866-8

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**:14–25. doi:10.1002/pro.3235

Gohlke H, Hendlich M, Klebe G. 2000. Knowledge-based scoring function to predict protein-ligand interactions11Edited by R. Huber. *Journal of Molecular Biology* **295**:337–356. doi:10.1006/jmbi.1999.3371

Goldenberg O, Erez E, Nimrod G, Ben-Tal N. 2009. The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res* **37**:D323–D327. doi:10.1093/nar/gkn822

Gouveia-Oliveira R, Pedersen AG. 2007. Finding coevolving amino acid residues using row and column weighting of mutual information and multi-dimensional amino acid representation. *Algorithms for Molecular Biology* **2**:12. doi:10.1186/1748-7188-2-12

Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**:2695–2696. doi:10.1093/bioinformatics/btl461

Grant BJ, Skjaerven L, Yao X-Q. 2021. The Bio3D packages for structural bioinformatics. *Protein Sci* **30**:20–30. doi:10.1002/pro.3923

Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. 2003. Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. *Journal of Molecular Biology* **331**:281–299. doi:10.1016/S0022-2836(03)00670-3

Grembecka J, Cierpicki T, Devedjiev Y, Derewenda U, Kang BS, Bushweller JH, Derewenda ZS. 2006. The binding of the PDZ tandem of syntenin to target proteins. *Biochemistry* **45**:3674–3683. doi:10.1021/bi052225y

Griffith JK, Baker ME, Rouch DA, Page MGP, Skurray RA, Paulsen IT, Chater KF, Baldwin SA, Henderson PJF. 1992. Membrane transport proteins: implications of sequence comparisons. *Current Opinion in Cell Biology* **4**:684–695. doi:10.1016/0955-0674(92)90090-Y

Guclu TF, Kocatug N, Atilgan AR, Atilgan C. 2021. N-Terminus of the Third PDZ Domain of PSD-95 Orchestrates Allosteric Communication for Selective Ligand Binding. *J Chem Inf Model* **61**:347–357. doi:10.1021/acs.jcim.0c01079

Guterres H, Im W. 2020. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J Chem Inf Model* **60**:2189–2198. doi:10.1021/acs.jcim.0c00057

Guzzo AV. 1965. The Influence of Amino Acid Sequence on Protein Structure. *Biophys J* **5**:809–822.

Hanley JG. 2008. PICK1: a multi-talented modulator of AMPA receptor trafficking. *Pharmacol Ther* **118**:152–160. doi:10.1016/j.pharmthera.2008.02.002

Hansson T, Oostenbrink C, van Gunsteren W. 2002. Molecular dynamics simulations. *Curr Opin Struct Biol* **12**:190–196. doi:10.1016/s0959-440x(02)00308-1

Harmalkar A, Gray JJ. 2021. Advances to tackle backbone flexibility in protein docking. *Current Opinion in Structural Biology*, Theory and Simulation/Computational Methods Macromolecular Assemblies **67**:178–186. doi:10.1016/j.sbi.2020.11.011

Harris BZ, Lim WA. 2001. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* **114**:3219–3231. doi:10.1242/jcs.114.18.3219

He X, Liu S, Lee T-S, Ji B, Man VH, York DM, Wang J. 2020. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **5**:4611–4619. doi:10.1021/acsomega.9b04233

Ho BK, Agard DA. 2010. Conserved tertiary couplings stabilize elements in the PDZ fold, leading to characteristic patterns of domain conformational flexibility. *Protein Sci* **19**:398–411. doi:10.1002/pro.318

Holley LH, Karplus M. 1989. Protein secondary structure prediction with a neural network. *Proceedings of the National Academy of Sciences* **86**:152–156. doi:10.1073/pnas.86.1.152

Hollingsworth SA, Dror RO. 2018. Molecular Dynamics Simulation for All. *Neuron* **99**:1129–1143. doi:10.1016/j.neuron.2018.08.011

Hollingsworth SA, Karplus PA. 2010. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins **1**:271–283. doi:10.1515/bmc.2010.022

Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. 2012. Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell* **149**:1607–1621. doi:10.1016/j.cell.2012.04.012

Hopf TA., Green AG, Schubert B, Mersmann S, Schärfe CPI, Ingraham JB, Toth-Petroczy A, Brock K, Riesselman AJ, Palmedo P, Kang C, Sheridan R, Draizen EJ, Dallago C, Sander C, Marks DS. 2019. The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **35**:1582–1584. doi:10.1093/bioinformatics/bty862

Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, Bonvin AMJJ, Marks DS. 2018. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**. doi:10.7554/eLife.03430

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006a. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**:712–725. doi:10.1002/prot.21123

Hornak V, Okur A, Rizzo RC, Simmerling C. 2006b. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **103**:915–920. doi:10.1073/pnas.0508452103

Hospital A, Goñi JR, Orozco M, Gelpí JL. 2015. Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* **8**:37–47. doi:10.2147/AABC.S70333

Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**:71–73. doi:10.1038/nmeth.4067

Huang N, Shoichet BK, Irwin JJ. 2006. Benchmarking Sets for Molecular Docking. *J Med Chem* **49**:6789–6801. doi:10.1021/jm0608356

Hubbard SJ, Thornton JM. 1993. Naccess, Computer Program; Department of Biochemistry and Molecular Biology, University College London,.

Hughes AL. 2005. Gene duplication and the origin of novel proteins. *Proceedings of the National Academy of Sciences* **102**:8791–8792. doi:10.1073/pnas.0503922102

Hughes AL. 1997. The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London Series B: Biological Sciences* **256**:119–124. doi:10.1098/rspb.1994.0058

Hultqvist G, Haq SR, Punekar AS, Chi CN, Engström Å, Bach A, Strømgaard K, Selmer M, Gianni S, Jemth P. 2013. Energetic pathway sampling in a protein interaction domain. *Structure* **21**:1193–1202. doi:10.1016/j.str.2013.05.010

Hvidsten TR, Lægreid A, Kryshtafovych A, Andersson G, Fidelis K, Komorowski J. 2009. A Comprehensive Analysis of the Structure-Function Relationship in Proteins Based on Local Structure Similarity. *PLOS ONE* **4**:e6266. doi:10.1371/journal.pone.0006266

Ikeguchi M, Ueno J, Sato M, Kidera A. 2005. Protein Structural Change Upon Ligand Binding: Linear Response Theory. *Phys Rev Lett* **94**:078102. doi:10.1103/PhysRevLett.94.078102

Inaba T, Stewart DJ, Kalow W. 1978. Metabolism of cocaine in man. *Clinical Pharmacology & Therapeutics* **23**:547–552. doi:10.1002/cpt1978235547

Jack DL, Yang NM, H. Saier Jr M. 2001. The drug/metabolite transporter superfamily. *European Journal of Biochemistry* **268**:3620–3639. doi:10.1046/j.1432-1327.2001.02265.x

Jäckel C, Kast P, Hilvert D. 2008. Protein Design by Directed Evolution. *Annual Review of Biophysics* **37**:153–173. doi:10.1146/annurev.biophys.37.032807.125832

Jaenicke R. 1987. Folding and association of proteins. *Progress in Biophysics and Molecular Biology* **49**:117–237. doi:10.1016/0079-6107(87)90011-3

Jana B, Morcos F, Onuchic JN. 2014. From structure to function: the convergence of structure based models and co-evolutionary information. *Phys Chem Chem Phys* **16**:6496–6507. doi:10.1039/C3CP55275F

Janin J, Wodak S, Levitt M, Maigret B. 1978. Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology* **125**:357–386. doi:10.1016/0022-2836(78)90408-4

Jayaraman V, Toledo-Patiño S, Noda-García L, Laurino P. 2022. Mechanisms of protein evolution. *Protein Science* **31**:e4362. doi:10.1002/pro.4362

Jo S, Kim T, Iyer VG, Im W. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**:1859–1865. doi:10.1002/jcc.20945

John B, Sali A. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res* **31**:3982–3992. doi:10.1093/nar/gkg460

Jones G, Willett P, Glen RC, Leach AR, Taylor R. 1997. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *Journal of Molecular Biology* **267**:727–748. doi:10.1006/jmbi.1996.0897

Jorgensen WL, Tirado-Rives J. 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *ACS Publications*. doi:10.1021/ja00214a001

Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montaño B, Blundell TL, Ascher DB. 2017. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Progress in Biophysics and Molecular Biology*, Exploring mechanisms in biology: simulations and experiments come together **128**:3–13. doi:10.1016/j.pbiomolbio.2016.10.002

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

Kalescky R, Liu J, Tao P. 2015. Identifying Key Residues for Protein Allostery through Rigid Residue Scan. *J Phys Chem A* **119**:1689–1700. doi:10.1021/jp5083455

Kalescky R, Zhou H, Liu J, Tao P. 2016. Rigid Residue Scan Simulations Systematically Reveal Residue Entropic Roles in Protein Allostery. *PLoS Comput Biol* **12**:e1004893. doi:10.1371/journal.pcbi.1004893

Kaltenbach M, Tokuriki N. 2014. Dynamics and constraints of enzyme evolution. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **322**:468–487. doi:10.1002/jez.b.22562

Kamisetty H, Ovchinnikov S, Baker D. 2013. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *PNAS* **110**:15674–15679. doi:10.1073/pnas.1314045110

Kao PN, Karlin A. 1986. Acetylcholine receptor binding site contains a disulfide cross-link between adjacent half-cystinyl residues. *Journal of Biological Chemistry* **261**:8085–8088. doi:10.1016/S0021-9258(19)83877-2

Karlsen ML, Thorsen TS, Johner N, Ammendrup-Johnsen I, Erlendsson S, Tian X, Simonsen JB, Høiberg-Nielsen R, Christensen NM, Khelashvili G, Streicher W, Teilum K, Vestergaard B, Weinstein H, Gether U, Arleth L, Madsen KL. 2015. Structure of Dimeric and Tetrameric Complexes of the BAR Domain Protein PICK1 Determined by Small-Angle X-Ray Scattering. *Structure* **23**:1258–1270. doi:10.1016/j.str.2015.04.020

Karplus M, McCammon JA. 2002. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol* **9**:646–652. doi:10.1038/nsb0902-646

Karplus M, Petsko GA. 1990. Molecular dynamics simulations in biology. *Nature* **347**:631–639. doi:10.1038/347631a0

Kaufmann SHE. 1990. Heat shock proteins and the immune response. *Immunology Today* **11**:129–136. doi:10.1016/0167-5699(90)90050-J

Kazan IC, Mills JH, Ozkan SB. 2023. Allosteric Regulatory Control in Dihydrofolate Reductase is Revealed by Dynamic Asymmetry. *Protein Sci* e4700. doi:10.1002/pro.4700

Kazan IC, Sharma P, Rahman MI, Bobkov A, Fromme R, Ghirlanda G, Ozkan SB. 2022. Design of novel cyanovirin-N variants by modulation of binding dynamics through distal mutations. *eLife* **11**:e67474. doi:10.7554/eLife.67474

Kelley BS, Chang LC, Bewley CA. 2002. Engineering an Obligate Domain-Swapped Dimer of Cyanovirin-N with Enhanced Anti-HIV Activity. *J Am Chem Soc* **124**:3210–3211. doi:10.1021/ja025537m

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**:662–666. doi:10.1038/181662a0

Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, Shore VC. 1960. Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature* **185**:422–427. doi:10.1038/185422a0

Kennedy MB. 1995. Origin of PDZ (DHR, GLGF) domains. *Trends Biochem Sci* **20**:350. doi:10.1016/s0968-0004(00)89074-x

Kessel A, Ben-Tal N. 2018. Introduction to Proteins: Structure, Function, and Motion, Second Edition. CRC Press.

Khan SB, Azhar-ul-Haq, Perveen S, Afza N, Malik A, Nawaz SA, Shah MR, Choudhary MI. 2005. Butyrylcholinesterase inhibitory guaianolides fromAmberboa ramosa. *Arch Pharm Res* **28**:172–176. doi:10.1007/BF02977710

Kim DE, DiMaio F, Wang RY-R, Song Y, Baker D. 2014. One contact for every twelve residues allows robust and accurate topology-level protein structure modeling. *Proteins: Structure, Function, and Bioinformatics* **82**:208–218. doi:10.1002/prot.24374

Kim E, Sheng M. 2004. PDZ domain proteins of synapses. *Nat Rev Neurosci* **5**:771–781. doi:10.1038/nrn1517

Kim H, Zou T, Modi C, Dörner K, Grunkemeyer TJ, Chen L, Fromme R, Matz MV, Ozkan SB, Wachter RM. 2015. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. *Structure* **23**:34–43. doi:10.1016/j.str.2014.11.011

Kitchen DB, Decornez H, Furr JR, Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **3**:935–949. doi:10.1038/nrd1549

Knies JL, Cai F, Weinreich DM. 2017. Enzyme Efficiency but Not Thermostability Drives Cefotaxime Resistance Evolution in TEM-1 β-Lactamase. *Mol Biol Evol* **34**:1040–1054. doi:10.1093/molbev/msx053

Koharudin LMI, Furey W, Gronenborn AM. 2009. A designed chimeric cyanovirin-N homolog lectin: Structure and molecular basis of sucrose binding. *Proteins: Structure, Function, and Bioinformatics* **77**:904–915. doi:10.1002/prot.22514

Koharudin LMI, Gronenborn AM. 2013. Sweet entanglements—protein: Glycan interactions in two HIV-inactivating lectin families. *Biopolymers* **99**:196–202. doi:10.1002/bip.22106

Koharudin LMI, Viscomi AR, Jee J-G, Ottonello S, Gronenborn AM. 2008. The Evolutionarily Conserved Family of Cyanovirin-N Homologs: Structures and Carbohydrate Specificity. *Structure* **16**:570–584. doi:10.1016/j.str.2008.01.015

Kolbaba-Kartchner B, Kazan IC, Mills JH, Ozkan SB. 2021. The Role of Rigid Residues in Modulating TEM-1 β-Lactamase Function and Thermostability. *International Journal of Molecular Sciences* **22**:2895. doi:10.3390/ijms22062895

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. 2000. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc Chem Res* **33**:889–897. doi:10.1021/ar000033j

Kong Y, Karplus M. 2009. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins* **74**:145–154. doi:10.1002/prot.22139

Korb O, Stützle T, Exner TE. 2009. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *J Chem Inf Model* **49**:84–96. doi:10.1021/ci800298z

Kotelchuck D, Scheraga HA. 1969. THE INFLUENCE OF SHORT-RANGE INTERACTIONS ON PROTEIN CONFORMATION, II. A MODEL FOR PREDICTING THE α-HELICAL REGIONS OF PROTEINS*. *Proc Natl Acad Sci U S A* **62**:14–21.

Kreitman M, Akashi H. 1995. Molecular Evidence for Natural Selection. *Annual Review of Ecology and Systematics* **26**:403–422. doi:10.1146/annurev.es.26.110195.002155

Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. 2019. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function, and Bioinformatics* **87**:1011–1020. doi:10.1002/prot.25823

Kuhlman B, Baker D. 2000. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences* **97**:10383–10388. doi:10.1073/pnas.97.19.10383

Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**:1364–1368. doi:10.1126/science.1089427

Kumar A, Butler BM, Kumar S, Ozkan SB. 2015a. Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. *Curr Opin Struct Biol* **35**:135–142. doi:10.1016/j.sbi.2015.11.002

Kumar A, Glembo TJ, Ozkan SB. 2015b. The Role of Conformational Dynamics and Allostery in the Disease Development of Human Ferritin. *Biophysical Journal* **109**:1273–1281. doi:10.1016/j.bpj.2015.06.060

Kumawat A, Chakrabarty S. 2020. Protonation-Induced Dynamic Allostery in PDZ Domain: Evidence of Perturbation-Independent Universal Response Network. *J Phys Chem Lett* **11**:9026–9031. doi:10.1021/acs.jpclett.0c02885

Kumawat A, Chakrabarty S. 2017. Hidden electrostatic basis of dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A* **114**:E5825–E5834. doi:10.1073/pnas.1705311114

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. 1982. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* **161**:269–288. doi:10.1016/0022-2836(82)90153-X

Kuriyan J, Eisenberg D. 2007. The origin of protein interactions and allostery in colocalization. *Nature* **450**:983–990. doi:10.1038/nature06524

Larrimore KE, Barcus M, Kannan L, Gao Y, Zhan C-G, Brimijoin S, Mor T. 2013. Plants as a source of butyrylcholinesterase variants designed for enhanced cocaine hydrolase activity. *Chem Biol Interact* **203**:217–220. doi:10.1016/j.cbi.2012.09.004

Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, Bolia A, Brimijoin S, Zhan C-G, Ozkan SB. 2017. Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Scientific reports* **7**:10419.

Lauck F, Smith CA, Friedland GF, Humphris EL, Kortemme T. 2010. RosettaBackrub—a web server for flexible backbone protein structure modeling and design. *Nucleic Acids Research* **38**:W569–W575. doi:10.1093/nar/gkq369

Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HF, Sumpradit N, Vlieghe E, Hara GL, Gould IM, Goossens H. 2013. Antibiotic resistance—the need for global solutions. *The Lancet infectious diseases* **13**:1057–1098.

Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban Y-EA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. 2011. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487**:545–574. doi:10.1016/B978-0-12-381270-4.00019-6

Lee H-J, Zheng JJ. 2010. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal* **8**:8. doi:10.1186/1478-811X-8-8

Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL, MacKerell AD, Klauda JB, Im W. 2016. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12**:405–413. doi:10.1021/acs.jctc.5b00935

Levitt M, Warshel A. 1975. Computer simulation of protein folding. *Nature* **253**:694–698. doi:10.1038/253694a0

Lewis PN, G[unk]o N, G[unk]o M, Kotelchuck D, Scheraga HA. 1970. Helix Probability Profiles of Denatured Proteins and Their Correlation with Native Structures*. *Proc Natl Acad Sci U S A* **65**:810–815.

Lexa KW, Carlson HA. 2012. Protein flexibility in docking and surface mapping. *Q Rev Biophys* **45**:301–343. doi:10.1017/S0033583512000066

Li T, Motta S, Stevens AO, Song S, Hendrix E, Pandini A, He Y. 2022. Recognizing the Binding Pattern and Dissociation Pathways of the p300 Taz2-p53 TAD2 Complex. *JACS Au* **2**:1935–1945. doi:10.1021/jacsau.2c00358

Li Z, Bolia A, Maxwell JD, Bobkov AA, Ghirlanda G, Ozkan SB, Margulis CJ. 2015. A Rigid Hinge Region Is Necessary for High-Affinity Binding of Dimannose to Cyanovirin and Associated Constructs. *Biochemistry* **54**:6951–6960. doi:10.1021/acs.biochem.5b00635

LiCata VJ, Allewell NM. 1997. Is substrate inhibition a consequence of allostery in aspartate transcarbamylase? *Biophysical Chemistry*, 10 Years of the Gibbs Conference on Biothermodynamics **64**:225–234. doi:10.1016/S0301-4622(96)02204-1

Lindahl E, Delarue M. 2005. Refinement of docked protein–ligand and protein–DNA structures using low frequency normal mode amplitude optimization. *Nucleic Acids Research* **33**:4496–4506. doi:10.1093/nar/gki730

Liu H, Chen Q. 2016. Computational protein design for given backbone: recent progresses in general method-related aspects. *Current Opinion in Structural Biology*, Engineering and design • Membranes **39**:89–95. doi:10.1016/j.sbi.2016.06.013

Liu J, Nussinov R. 2017. Energetic redistribution in allostery to execute protein function. *Proc Natl Acad Sci U S A* **114**:7480–7482. doi:10.1073/pnas.1709071114

Liu J, Nussinov R. 2016. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput Biol* **12**:e1004966. doi:10.1371/journal.pcbi.1004966

Liu X, Shurong H, Wenchao Y, Lei F, Fang Z, Chang-Guo Z. 2013. Catalytic activities of a cocaine hydrolase engineered from human butyrylcholinesterase against (+)- and (-)-cocaine. *Chemico-biological interactions* **203**. doi:10.1016/j.cbi.2012.08.003

Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. 1999. Protein structure prediction by global optimization of a potential energy function. *Proc Natl Acad Sci U S A* **96**:5482–5485. doi:10.1073/pnas.96.10.5482

Lockless SW, Ranganathan R. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**:295–299. doi:10.1126/science.286.5438.295

Lockridge O. 2015. Review of human butyrylcholinesterase structure, function, genetic variants, history of use in the clinic, and potential therapeutic uses. *Pharmacol Ther* **148**:34–46. doi:10.1016/j.pharmthera.2014.11.011

Lodish HF, Berk A, Zipursky SL, Baltimore D, Darnell JE, Matsudaira P. 2000. Molecular Cell Biology. W.H. Freeman.

Loizzo MR, Tundis R, Menichini Federica, Menichini Francesco. 2008. Natural Products and their Derivatives as Cholinesterase Inhibitors in the Treatment of Neurodegenerative Disorders: An Update. *Current Medicinal Chemistry* **15**:1209–1228.

Lori C, Lantella A, Pasquo A, Alexander LT, Knapp S, Chiaraluce R, Consalvi V. 2013. Effect of Single Amino Acid Substitution Observed in Cancer on Pim-1 Kinase Thermodynamic Stability and Structure. *PLOS ONE* **8**:e64824. doi:10.1371/journal.pone.0064824

Lu C, Knecht V, Stock G. 2016. Long-Range Conformational Response of a PDZ Domain to Ligand Binding and Release: A Molecular Dynamics Study. *J Chem Theory Comput* **12**:870–878. doi:10.1021/acs.jctc.5b01009

Lu H-M, Liang J. 2009. Perturbation-based Markovian transmission model for probing allosteric dynamics of large macromolecular assembling: a study of GroEL-GroES. *PLoS Comput Biol* **5**:e1000526. doi:10.1371/journal.pcbi.1000526

Lu W, Ziff EB. 2005. PICK1 interacts with ABP/GRIP to regulate AMPA receptor trafficking. *Neuron* **47**:407–421. doi:10.1016/j.neuron.2005.07.006

Luck K, Charbonnier S, Travé G. 2012. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett* **586**:2648–2661. doi:10.1016/j.febslet.2012.03.056

Luk LYP, Javier Ruiz-Pernía J, Dawson WM, Roca M, Loveridge EJ, Glowacki DR, Harvey JN, Mulholland AJ, Tuñón I, Moliner V, Allemann RK. 2013. Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. *Proceedings of the National Academy of Sciences* **110**:16344–16349. doi:10.1073/pnas.1312437110

Ma B, Tsai C-J, Haliloğlu T, Nussinov R. 2011. Dynamic allostery: linkers are not merely flexible. *Structure* **19**:907–917. doi:10.1016/j.str.2011.06.002

MacKerell Jr. AD, Banavali N, Foloppe N. 2000. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**:257–265. doi:10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W

Madsen KL, Thorsen TS, Rahbek-Clemmensen T, Eriksen J, Gether U. 2012. Protein Interacting with C Kinase 1 (PICK1) Reduces Reinsertion Rates of Interaction Partners Sorted to Rab11-dependent Slow Recycling Pathway *. *Journal of Biological Chemistry* **287**:12293–12308. doi:10.1074/jbc.M111.294702

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**:3696–3713. doi:10.1021/acs.jctc.5b00255

Makov G, Payne MC. 1995. Periodic boundary conditions in ab initio calculations. *Phys Rev B* **51**:4014–4022. doi:10.1103/PhysRevB.51.4014

Manson AC, Coalson RD. 2012. Response of Rotation–Translation Blocked Proteins Using Langevin Dynamics on a Locally Harmonic Landscape. *J Phys Chem B* **116**:12142–12158. doi:10.1021/jp306030b

Mark P, Nilsson L. 2001. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J Phys Chem A* **105**:9954–9960. doi:10.1021/jp003020w

Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. 2011. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLOS ONE* **6**:e28766. doi:10.1371/journal.pone.0028766

Marks DS, Hopf TA, Sander C. 2012. Protein structure prediction from sequence variation. *Nat Biotechnol* **30**:1072–1080. doi:10.1038/nbt.2419

Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. 2007. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J Phys Chem B* **111**:7812–7824. doi:10.1021/jp071097f

Martínez JL. 2008. Antibiotics and antibiotic resistance genes in natural environments. *Science* **321**:365–367.

Mashiach E, Nussinov R, Wolfson HJ. 2010. FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins: Structure, Function, and Bioinformatics* **78**:1503–1519. doi:10.1002/prot.22668

Masson P, Xie W, Froment M-T, Lockridge O. 2001. Effects of mutations of active site residues and amino acids interacting with the Ω loop on substrate activation of butyrylcholinesterase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1544**:166–176. doi:10.1016/S0167-4838(00)00217-X

Matei E, Basu R, Furey W, Shi J, Calnan C, Aiken C, Gronenborn AM. 2016. Structure and Glycan Binding of a New Cyanovirin-N Homolog. *J Biol Chem* **291**:18967–18976. doi:10.1074/jbc.M116.740415

Mauldin RV, Carroll MJ, Lee AL. 2009. Dynamic Dysfunction in Dihydrofolate Reductase Results from Antifolate Drug Binding: Modulation of Dynamics within a Structural State. *Structure* **17**:386–394. doi:10.1016/j.str.2009.01.005

Mauldin RV, Lee AL. 2010. Nuclear Magnetic Resonance Study of the Role of M42 in the Solution Dynamics of Escherichia coli Dihydrofolate Reductase. *Biochemistry* **49**:1606–1615. doi:10.1021/bi901798g

May A, Zacharias M. 2008. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking. *Proteins: Structure, Function, and Bioinformatics* **70**:794–809. doi:10.1002/prot.21579

McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. *Nature* **267**:585–590. doi:10.1038/267585a0

McCammon JA, Harvey SC. 1988. Dynamics of Proteins and Nucleic Acids. Cambridge University Press.

McCormick JW, Russo MA, Thompson S, Blevins A, Reynolds KA. 2021. Structurally distributed surface sites tune allosteric regulation. *elife* **10**:e68346.

McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. 2012. The spatial architecture of protein function and adaptation. *Nature* **491**:138–142. doi:10.1038/nature11500

Meiler J, Baker D. 2006. ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics* **65**:538–548. doi:10.1002/prot.21086

Meinhardt S, Jr MWM, Parente DJ, Swint-Kruse L. 2013. Rheostats and Toggle Switches for Modulating Protein Function. *PLOS ONE* **8**:e83502. doi:10.1371/journal.pone.0083502

Meng EC, Shoichet BK, Kuntz ID. 1992. Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry* **13**:505–524. doi:10.1002/jcc.540130412

Meng X-Y, Zhang H-X, Mezei M, Cui M. 2011. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer - Aided Drug Design* **7**:146–157. doi:10.2174/157340911795677602

Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. 2011. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**:2319–2327. doi:10.1002/jcc.21787

Miller M, Bromberg Y, Swint-Kruse L. 2017. Computational predictors fail to identify amino acid substitution effects at rheostat positions. *Scientific Reports* **7**:41329. doi:10.1038/srep41329

Miño-Galaz GA. 2015. Allosteric communication pathways and thermal rectification in PDZ-2 protein: a computational study. *J Phys Chem B* **119**:6179–6189. doi:10.1021/acs.jpcb.5b02228

Mionetto N, Morel N, Massoulié J, Schmid RD. 1997. Biochemical determination of insecticides via cholinesterases. 1. Acetylcholinesterase from rat brain: functional expression using a baculovirus system, and biochemical characterization. *Biotechnology Techniques* **11**:805–812. doi:10.1023/A:1018425224892

Mobley DL, Dill KA. 2009. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get." *Structure* **17**:489–498. doi:10.1016/j.str.2009.02.010

Modi T, Campitelli P, Kazan IC, Ozkan SB. 2021a. Protein folding stability and binding interactions through the lens of evolution: a dynamical perspective. *Current Opinion in Structural Biology* **66**:207–215. doi:10.1016/j.sbi.2020.11.007

Modi T, Huihui J, Ghosh K, Ozkan SB. 2018. Ancient thioredoxins evolved to modern-day stability–function requirement by altering native state ensemble. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**:20170184. doi:10.1098/rstb.2017.0184

Modi T, Ozkan SB. 2018. Mutations Utilize Dynamic Allostery to Confer Resistance in TEM-1 β-lactamase. *Int J Mol Sci* **19**. doi:10.3390/ijms19123808

Modi T, Risso VA, Martinez-Rodriguez S, Gavira JA, Mebrat MD, Van Horn WD, Sanchez-Ruiz JM, Banu Ozkan S. 2021b. Hinge-shift mechanism as a protein design principle for the evolution of β-lactamases from substrate promiscuity to specificity. *Nat Commun* **12**:1852. doi:10.1038/s41467-021-22089-0

Monod J, Wyman J, Changeux J-P. 1965. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* **12**:88–118. doi:10.1016/S0022-2836(65)80285-6

Mor TS. 2015. Molecular pharming's foot in the FDA's door: Protalix's trailblazing story. *Biotechnol Lett* **37**:2147–2150. doi:10.1007/s10529-015-1908-z

Mor TS, Sternfeld M, Soreq H, Arntzen CJ, Mason HS. 2001. Expression of recombinant human acetylcholinesterase in transgenic tomato plants. *Biotechnology and Bioengineering* **75**:259–266. doi:10.1002/bit.10012

Morais Cabral JH, Petosa C, Sutcliffe MJ, Raza S, Byron O, Poy F, Marfatia SM, Chishti AH, Liddington RC. 1996. Crystal structure of a PDZ domain. *Nature* **382**:649–652. doi:10.1038/382649a0

Morcos F. 2020. Protein conformations à la carte, a step further in de novo protein design. *PNAS* **117**:8674–8676. doi:10.1073/pnas.2004188117

Morcos F, Hwa T, Onuchic JN, Weigt M. 2014a. Direct Coupling Analysis for Protein Contact Prediction In: Kihara D, editor. Protein Structure Prediction, Methods in Molecular Biology. New York, NY: Springer. pp. 55–70. doi:10.1007/978-1-4939-0366-5_5

Morcos F, Jana B, Hwa T, Onuchic JN. 2013. Coevolutionary signals across protein lineages help capture multiple protein conformations. *PNAS* **110**:20533–20538. doi:10.1073/pnas.1315625110

Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences* **108**:E1293–E1301. doi:10.1073/pnas.1111471108

Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG. 2014b. Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc Natl Acad Sci U S A* **111**:12408–12413. doi:10.1073/pnas.1413575111

Mori T, Boyd MR. 2001. Cyanovirin-N, a Potent Human Immunodeficiency Virus-Inactivating Protein, Blocks both CD4-Dependent and CD4-Independent Binding of Soluble gp120 (sgp120) to Target Cells, Inhibits sCD4-Induced Binding of sgp120 to Cell-Associated CXCR4, and Dissociates Bound sgp120 from Target Cells. *Antimicrobial Agents and Chemotherapy* **45**:664–672. doi:10.1128/AAC.45.3.664-672.2001

Morra G, Genoni A, Colombo G. 2014. Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains. *J Chem Theory Comput* **10**:5677–5689. doi:10.1021/ct500326g

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**:2785–2791. doi:10.1002/jcc.21256

Morris GM, Lim-Wilby M. 2008. Molecular Docking In: Kukol A, editor. Molecular Modeling of Proteins, Methods Molecular Biology™. Totowa, NJ: Humana Press. pp. 365–382. doi:10.1007/978-1-59745-177-2_19

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press.

Némethy G, Scheraga HA. 1977. Protein folding. *Quarterly Reviews of Biophysics* **10**:239–352. doi:10.1017/S0033583500002936

Némethy G, Scheraga HA. 1965. Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. *Biopolymers* **3**:155–184. doi:10.1002/bip.360030205

Newman MEJ, Girvan M. 2004. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **69**:026113. doi:10.1103/PhysRevE.69.026113

Nishikawa K. 1983. Assessment of secondary-structure prediction of proteins comparison of computerized chou-fasman method with others. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **748**:285–299. doi:10.1016/0167-4838(83)90306-0

Niu X, Chen Q, Zhang J, Shen W, Shi Y, Wu J. 2007. Interesting structural and dynamical behaviors exhibited by the AF-6 PDZ domain upon Bcr peptide binding. *Biochemistry* **46**:15042–15053. doi:10.1021/bi701303p

Noid WG. 2013. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* **139**:090901. doi:10.1063/1.4818908

Ohno S. 2013. Evolution by Gene Duplication. Springer Science & Business Media.

Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG. 2006. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **103**:11844–11849.

O'Keefe BR, Smee DF, Turpin JA, Saucedo CJ, Gustafson KR, Mori T, Blakeslee D, Buckheit R, Boyd MR. 2003. Potent Anti-Influenza Activity of Cyanovirin-N and Interactions with Viral Hemagglutinin. *Antimicrobial Agents and Chemotherapy* **47**:2518–2525. doi:10.1128/AAC.47.8.2518-2525.2003

Olson CA, Wu NC, Sun R. 2014. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr Biol* **24**:2643–2651. doi:10.1016/j.cub.2014.09.072

Orencia MC, Yoon JS, Ness JE, Stemmer WP, Stevens RC. 2001. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat Struct Biol* **8**:238–242. doi:10.1038/84981

Ose N, Butler BM, Kumar A, Ozkan SB, Kumar S. 2020. Dynamic Allosteric Residue Coupling Reveals Disease Mechanism for Gaucher Disease and NSNVS Across the Proteome. *Biophysical Journal* **118**:53a. doi:10.1016/j.bpj.2019.11.472

Ose NJ, Butler BM, Kumar A, Kazan IC, Sanderford M, Kumar S, Ozkan SB. 2022. Dynamic coupling of residues within proteins as a mechanistic foundation of many enigmatic pathogenic missense variants. *PLoS Comput Biol* **18**:e1010006. doi:10.1371/journal.pcbi.1010006

Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. 2002. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Structure, Function, and Bioinformatics* **46**:34–40. doi:10.1002/prot.10028

Ota N, Agard DA. 2005. Intramolecular Signaling Pathways Revealed by Modeling Anisotropic Thermal Diffusion. *Journal of Molecular Biology* **351**:345–354. doi:10.1016/j.jmb.2005.05.043

Ovchinnikov S, Kinch L, Park H, Liao Y, Pei J, Kim DE, Kamisetty H, Grishin NV, Baker D. 2015. Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife* **4**:e09248. doi:10.7554/eLife.09248

Pagadala NS, Syed K, Tuszynski J. 2017. Software for molecular docking: a review. *Biophys Rev* **9**:91–102. doi:10.1007/s12551-016-0247-1

Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, Kishony R. 2015. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat Commun* **6**:7385. doi:10.1038/ncomms8385

Pan L, Wu H, Shen C, Shi Y, Jin W, Xia J, Zhang M. 2007. Clustering and synaptic targeting of PICK1 requires direct interaction between the PDZ domain and lipid membranes. *EMBO J* **26**:4576–4587. doi:10.1038/sj.emboj.7601860

Pan Y, Gao D, Yang W, Cho H, Yang G, Tai H-H, Zhan C-G. 2005. Computational redesign of human butyrylcholinesterase for anticocaine medication. *Proc Natl Acad Sci U S A* **102**:16656–16661. doi:10.1073/pnas.0507332102

Pancook, J.D., Pecht, G., Ader, M., Mosko, M., Lockridge, O. and Watkins, J.D., 2003, March. Application of directed evolution technology to optimize the cocaine hydrolase activity of human butyrylcholinesterase. In *Faseb Journal* (Vol. 17, No. 4, pp. A565-A565). 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998 USA: FEDERATION AMER SOC EXP BIOL.

Pandey A, Mann M. 2000. Proteomics to study genes and genomes. *Nature* **405**:837–846. doi:10.1038/35015709

Parra RG, Schafer NP, Radusky LG, Tsai M-Y, Guzovsky AB, Wolynes PG, Ferreiro DU. 2016. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res* **44**:W356-360. doi:10.1093/nar/gkw304

Patsalo V, Raleigh DP, Green DF. 2011. Rational and Computational Design of Stabilized Variants of Cyanovirin-N That Retain Affinity and Specificity for Glycan Ligands. *Biochemistry* **50**:10698–10712. doi:10.1021/bi201411c

Pedersen SW, Pedersen SB, Anker L, Hultqvist G, Kristensen AS, Jemth P, Strømgaard K. 2014. Probing backbone hydrogen bonding in PDZ/ligand interactions by protein amide-to-ester mutations. *Nat Commun* **5**:3215. doi:10.1038/ncomms4215

Percudani R, Montanini B, Ottonello S. 2005. The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins: Structure, Function, and Bioinformatics* **60**:670–678. doi:10.1002/prot.20543

Perez A, Morrone JA, Simmerling C, Dill KA. 2016. Advances in free-energy-based simulations of protein folding and ligand binding. *Curr Opin Struct Biol* **36**:25–31. doi:10.1016/j.sbi.2015.12.002

Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC. 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **185**:416–422. doi:10.1038/185416a0

Peterson FC, Penkert RR, Volkman BF, Prehoda KE. 2004. Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. *Mol Cell* **13**:665–676. doi:10.1016/s1097-2765(04)00086-3

Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. 2009. Hidden dynamic allostery in a PDZ domain. *Proceedings of the National Academy of Sciences* **106**:18249–18254. doi:10.1073/pnas.0904492106

Piana S, Klepeis JL, Shaw DE. 2014. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Current Opinion in Structural Biology*, Folding and binding / Nucleic acids and their protein complexes **24**:98–105. doi:10.1016/j.sbi.2013.12.006

Pinzi L, Rastelli G. 2019. Molecular Docking: Shifting Paradigms in Drug Discovery. *International Journal of Molecular Sciences* **20**:4331. doi:10.3390/ijms20184331

Pollock DD, Thiltgen G, Goldstein RA. 2012. Amino acid coevolution induces an evolutionary Stokes shift. *Proceedings of the National Academy of Sciences* **109**:E1352–E1359. doi:10.1073/pnas.1120084109

Ponnuswamy PK, Prabhakaran M, Manavalan P. 1980. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **623**:301–316. doi:10.1016/0005-2795(80)90258-5

Ponting CP. 1997. Evidence for PDZ domains in bacteria, yeast, and plants. *Protein Sci* **6**:464–468.

Prothero JW. 1966. Correlation between the distribution of amino acids and alpha helices. *Biophys J* **6**:367–370.

Qi X, Yang Y, Su Y, Wang T. 2009. Molecular Cloning and Sequence Analysis of Cyanovirin-N Homology Gene in Ceratopteris thalictroides. *amfj* **99**:78–92. doi:10.1640/0002-8444-99.2.78

Quon KC, Yang B, Domian IJ, Shapiro L, Marczynski GT. 1998. Negative control of bacterial DNA replication by a cell cycle regulatory protein that binds at the chromosome origin. *Proceedings of the National Academy of Sciences* **95**:120–125. doi:10.1073/pnas.95.1.120

Radic Z, Pickering NA, Vellom DC, Camp S, Taylor P. 1993. Three distinct domains in the cholinesterase molecule confer selectivity for acetyl- and butyrylcholinesterase inhibitors. *Biochemistry* **32**:12074–12084. doi:10.1021/bi00096a018

Ramachandran GN, Sasisekharan V. 1968. Conformation of Polypeptides and Proteins**The literature survey for this review was completed in September 1967, with the journals which were then available in Madras and the preprinta which the authors had received.††By the authors' request, the publishers have left certain matters of usage and spelling in the form in which they wrote them. In: Anfinsen CB, Anson ML, Edsall JT, Richards FM, editors. Advances in Protein Chemistry. Academic Press. pp. 283–437. doi:10.1016/S0065-3233(08)60402-7

Ramadugu SK, Li Z, Kashyap HK, Margulis CJ. 2014. The Role of Glu41 in the Binding of Dimannose to P51G-m4-CVN. *Biochemistry* **53**:1477–1484. doi:10.1021/bi4014159

Rapaport DC. 2004. The Art of Molecular Dynamics Simulation. Cambridge University Press.

Rastelli G, Rio AD, Degliesposti G, Sgobba M. 2010. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry* **31**:797–810. doi:10.1002/jcc.21372

Rausch AO, Freiberger MI, Leonetti CO, Luna DM, Radusky LG, Wolynes PG, Ferreiro DU, Parra RG. 2021. FrustratometeR: an R-package to compute local frustration in protein structures, point mutants and MD simulations. *Bioinformatics* **37**:3038–3040. doi:10.1093/bioinformatics/btab176

Reynolds KA, McLaughlin RN, Ranganathan R. 2011. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**:1564–1575.

Ricard-Blum S. 2011. The Collagen Family. *Cold Spring Harb Perspect Biol* **3**:a004978. doi:10.1101/cshperspect.a004978

Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J Am Chem Soc* **135**:2899–2902. doi:10.1021/ja311630a

Risso VA, Sanchez-Ruiz JM, Ozkan SB. 2018. Biotechnological and protein-engineering implications of ancestral protein resurrection. *Current Opinion in Structural Biology*, Engineering and design: New applications • Membranes **51**:106–115. doi:10.1016/j.sbi.2018.02.007

Ritchie DW. 2008. Recent Progress and Future Directions in Protein-Protein Docking. *Current Protein and Peptide Science* **9**:1–15. doi:10.2174/138920308783565741

Rivoire O, Reynolds KA, Ranganathan R. 2016. Evolution-Based Functional Decomposition of Proteins. *PLOS Computational Biology* **12**:e1004817. doi:10.1371/journal.pcbi.1004817

Rocca DL, Martin S, Jenkins EL, Hanley JG. 2008. Inhibition of Arp2/3-mediated actin polymerization by PICK1 regulates neuronal morphology and AMPA receptor endocytosis. *Nat Cell Biol* **10**:259–271. doi:10.1038/ncb1688

Rod TH, Radkiewicz JL, Brooks CL. 2003. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proceedings of the National Academy of Sciences* **100**:6980–6985. doi:10.1073/pnas.1230801100

Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. 2016. Biophysical principles predict fitness landscapes of drug resistance. *Proceedings of the National Academy of Sciences* **113**:E1470–E1478. doi:10.1073/pnas.1601441113

Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G. 1998. Homology modeling, model and software evaluation: three related resources. *Bioinformatics* **14**:523–528. doi:10.1093/bioinformatics/14.6.523

Rohl CA, Strauss CEM, Chivian D, Baker D. 2004a. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins: Structure, Function, and Bioinformatics* **55**:656–677. doi:10.1002/prot.10629

Rohl CA, Strauss CEM, Misura KMS, Baker D. 2004b. Protein structure prediction using Rosetta. *Methods Enzymol* **383**:66–93. doi:10.1016/S0076-6879(04)83004-0

Romero G, von Zastrow M, Friedman PA. 2011. Role of PDZ proteins in regulating trafficking, signaling, and function of GPCRs: means, motif, and opportunity. *Adv Pharmacol* **62**:279–314. doi:10.1016/B978-0-12-385952-5.00003-8

Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10**:866–876. doi:10.1038/nrm2805

Rosenberry TL. 2010. Strategies to resolve the catalytic mechanism of acetylcholinesterase. *J Mol Neurosci* **40**:32–39. doi:10.1007/s12031-009-9250-3

Salinas VH, Ranganathan R. 2018. Coevolution-based inference of amino acid interactions underlying protein function. *eLife* **7**:e34300. doi:10.7554/eLife.34300

Salomon-Ferrer R, Case DA, Walker RC. 2013a. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **3**:198–210. doi:10.1002/wcms.1121

Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC. 2013b. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *Journal of chemical theory and computation* **9**:3878–3888.

Salverda ML, De Visser JAG, Barlow M. 2010. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. *FEMS microbiology reviews* **34**:1015–1036.

Sawaya MR, Kraut J. 1997. Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**:586–603.

Sawle L, Ghosh K. 2016. Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma. *J Chem Theory Comput* **12**:861–869. doi:10.1021/acs.jctc.5b00999

Saxena A, Hastings NB, Sun W, Dabisch PA, Hulet SW, Jakubowski EM, Mioduszewski RJ, Doctor BP. 2015. Prophylaxis with human serum butyrylcholinesterase protects Göttingen minipigs exposed to a lethal high-dose of sarin vapor. *Chem Biol Interact* **238**:161–169. doi:10.1016/j.cbi.2015.07.001

Saxena A, Hur RS, Luo C, Doctor BP. 2003. Natural monomeric form of fetal bovine serum acetylcholinesterase lacks the C-terminal tetramerization domain. *Biochemistry* **42**:15292–15299. doi:10.1021/bi030150x

Schiffer M, Edmundson AB. 1967. Use of Helical Wheels to Represent the Structures of Proteins and to Identify Segments with Helical Potential. *Biophys J* **7**:121–135.

Schirò A, Carlon A, Parigi G, Murshudov G, Calderone V, Ravera E, Luchinat C. 2020. On the complementarity of X-ray and NMR data. *J Struct Biol X* **4**:100019. doi:10.1016/j.yjsbx.2020.100019

Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **40**:843–856. doi:10.1007/s00249-011-0700-9

Schneider JD, Castilho A, Neumann L, Altmann F, Loos A, Kannan L, Mor TS, Steinkellner H. 2014a. Expression of human butyrylcholinesterase with an engineered glycosylation profile resembling the plasma-derived orthologue. *Biotechnol J* **9**:501–510. doi:10.1002/biot.201300229

Schneider JD, Marillonnet S, Castilho A, Gruber C, Werner S, Mach L, Klimyuk V, Mor TS, Steinkellner H. 2014b. Oligomerization status influences subcellular deposition and glycosylation of recombinant butyrylcholinesterase in Nicotiana benthamiana. *Plant Biotechnol J* **12**:832–839. doi:10.1111/pbi.12184

Schneider SH, Kozuch J, Boxer SG. 2021. The Interplay of Electrostatics and Chemical Positioning in the Evolution of Antibiotic Resistance in TEM β-Lactamases. *ACS Central Science* **7**:1996–2008.

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research* **33**:W363–W367. doi:10.1093/nar/gki481

Schnell JR, Dyson HJ, Wright PE. 2004. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annual review of biophysics and biomolecular structure* **33**:119–140.

Schwikowski B, Uetz P, Fields S. 2000. A network of protein–protein interactions in yeast. *Nat Biotechnol* **18**:1257–1261. doi:10.1038/82360

Shackelford G, Karplus K. 2007. Contact prediction using mutual information and neural nets. *Proteins: Structure, Function, and Bioinformatics* **69**:159–164. doi:10.1002/prot.21791

Shafferman A, Velan B, Ordentlich A, Kronman C, Grosfeld H, Leitner M, Flashner Y, Cohen S, Barak D, Ariel N. 1992. Substrate inhibition of acetylcholinesterase: residues affecting signal transduction from the surface to the catalytic center. *The EMBO Journal* **11**:3561–3568. doi:10.1002/j.1460-2075.1992.tb05439.x

Shalf J. 2020. The future of computing beyond Moore's Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **378**:20190061. doi:10.1098/rsta.2019.0061

Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossváry I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC. 2008. Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM* **51**:91–97. doi:10.1145/1364782.1364802

Sherman W, Beard HS, Farid R. 2006. Use of an Induced Fit Receptor Structure in Virtual Screening. *Chemical Biology & Drug Design* **67**:83–84. doi:10.1111/j.1747-0285.2005.00327.x

Simm S, Einloft J, Mirus O, Schleiff E. 2016. 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biological Research* **49**:31. doi:10.1186/s40659-016-0092-5

Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C. 2013. MISTIC: mutual information server to infer coevolution. *Nucleic Acids Research* **41**:W8–W14. doi:10.1093/nar/gkt427

Singh MK, Dominy BN. 2012. The evolution of cefotaximase activity in the TEM β-lactamase. *J Mol Biol* **415**:205–220. doi:10.1016/j.jmb.2011.10.041

Skjærven L, Yao X-Q, Scarabelli G, Grant BJ. 2014. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**:399. doi:10.1186/s12859-014-0399-6

Smith GR, Sternberg MJE, Bates PA. 2005. The Relationship between the Flexibility of Proteins and their Conformational States on Forming Protein–Protein Complexes with an Application to Protein–Protein Docking. *Journal of Molecular Biology* **347**:1077–1101. doi:10.1016/j.jmb.2005.01.058

Soskine M, Tawfik DS. 2010. Mutational effects and the evolution of new protein functions. *Nat Rev Genet* **11**:572–582. doi:10.1038/nrg2808

Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**:1701–1718. doi:10.1002/jcc.20291

Sreerama N, Woody RW. 2000. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Analytical Biochemistry* **287**:252–260. doi:10.1006/abio.2000.4880

Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E. 2013. Molecular Mechanisms of Disease-Causing Missense Mutations. *Journal of Molecular Biology*, Understanding Molecular Effects of Naturally Occurring Genetic Differences **425**:3919–3936. doi:10.1016/j.jmb.2013.07.014

Sternberg MJ, Thornton JM. 1978. Prediction of protein structure from amino acid sequence. *Nature* **271**:15–20. doi:10.1038/271015a0

Stevens AO, He Y. 2022a. Allosterism in the PDZ Family. *Int J Mol Sci* **23**:1454. doi:10.3390/ijms23031454

Stevens AO, He Y. 2022b. Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules* **12**:985. doi:10.3390/biom12070985

Stevens AO, Kazan IC, Ozkan B, He Y. 2022a. Investigating the allosteric response of the PICK1 PDZ domain to different ligands with all-atom simulations. *Protein Science* **31**:e4474. doi:10.1002/pro.4474

Stevens AO, Luo S, He Y. 2022b. Three Binding Conformations of BIO124 in the Pocket of the PICK1 PDZ Domain. *Cells* **11**:2451. doi:10.3390/cells11152451

Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase. *Cell* **160**:882–892. doi:10.1016/j.cell.2015.01.035

Stretton AOW. 2002. The first sequence. Fred Sanger and insulin. *Genetics* **162**:527–532.

Sun H, El Yazal J, Lockridge O, Schopfer LM, Brimijoin S, Pang YP. 2001. Predicted Michaelis-Menten complexes of cocaine-butyrylcholinesterase. Engineering effective butyrylcholinesterase mutants for cocaine detoxication. *J Biol Chem* **276**:9330–9336. doi:10.1074/jbc.M006676200

Sun H, Pang Y-P, Lockridge O, Brimijoin S. 2002a. Re-engineering Butyrylcholinesterase as a Cocaine Hydrolase. *Mol Pharmacol* **62**:220–224. doi:10.1124/mol.62.2.220

Sun H, Shen ML, Pang Y-P, Lockridge O, Brimijoin S. 2002b. Cocaine Metabolism Accelerated by a Re-Engineered Human Butyrylcholinesterase. *J Pharmacol Exp Ther* **302**:710–716. doi:10.1124/jpet.302.2.710

Swint-Kruse L, Matthews KS, Smith PE, Pettitt BM. 1998. Comparison of simulated and experimentally determined dynamics for a variant of the LacI DNA-binding domain, NLac-P. *Biophysical journal* **74**:413–421.

Szilárd P, Abraham MJ, Kutzner C, Hess B, Lindahl E. 2015. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. pp. 3–27. doi:10.1007/978-3-319-15976-8_1

Tamer YT, Gaszek IK, Abdizadeh H, Batur TA, Reynolds KA, Atilgan AR, Atilgan C, Toprak E. 2019. High-Order Epistasis in Catalytic Power of Dihydrofolate Reductase Gives Rise to a Rugged Fitness Landscape in the Presence of Trimethoprim Selection. *Molecular Biology and Evolution* **36**:1533–1550. doi:10.1093/molbev/msz086

Tanford Charles, Lovrien Rex. 1962. Dissociation of Catalase into Subunits. *J Am Chem Soc* **84**:1892–1896. doi:10.1021/ja00869a025

Taylor RD, Jewsbury PJ, Essex JW. 2002. A review of protein-small molecule docking methods. *J Comput Aided Mol Des* **16**:151–166. doi:10.1023/A:1020155510718

Thayer KM, Lakhani B, Beveridge DL. 2017. Molecular Dynamics-Markov State Model of Protein Ligand Binding and Allostery in CRIB-PDZ: Conformational Selection and Induced Fit. *J Phys Chem B* **121**:5509–5514. doi:10.1021/acs.jpcb.7b02083

Thomas VL, McReynolds AC, Shoichet BK. 2010. Structural bases for stability–function tradeoffs in antibiotic resistance. *Journal of molecular biology* **396**:47–59.

Thompson S, Zhang Y, Ingle C, Reynolds KA, Kortemme T. 2020. Altered expression of a quality control protease in E. coli reshapes the in vivo mutational landscape of a model enzyme. *Elife* **9**:e53476.

Thorpe IF, Brooks III CL. 2004. The coupling of structural fluctuations to hydride transfer in dihydrofolate reductase. *Proteins: Structure, Function, and Bioinformatics* **57**:444–457. doi:10.1002/prot.20219

Tochio H, Hung F, Li M, Bredt DS, Zhang M. 2000. Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-9511Edited by P. E. Wright. *Journal of Molecular Biology* **295**:225–237. doi:10.1006/jmbi.1999.3350

Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications* **3**:1257. doi:10.1038/ncomms2246

Topp E, Irwin R, McAllister T, Lessard M, Joensuu JJ, Kolotilin I, Conrad U, Stöger E, Mor T, Warzecha H, Hall JC, McLean MD, Cox E, Devriendt B, Potter A, Depicker A, Virdi V, Holbrook L, Doshi K, Dussault M, Friendship R, Yarosh O, Yoo HS, MacDonald J, Menassa R. 2016. The case for plant-made veterinary immunotherapeutics. *Biotechnol Adv* **34**:597–604. doi:10.1016/j.biotechadv.2016.02.007

Torgeson KR, Clarkson MW, Granata D, Lindorff-Larsen K, Page R, Peti W. 2022. Conserved conformational dynamics determine enzyme activity. *Science Advances* **8**:eabo5546. doi:10.1126/sciadv.abo5546

Totrov M, Abagyan R. 2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology*, Theory and simulation / Macromolecular assemblages **18**:178–184. doi:10.1016/j.sbi.2008.01.004

Tripathi S, Waxham MN, Cheung MS, Liu Y. 2015. Lessons in Protein Design from Combined Evolution and Conformational Dynamics. *Scientific Reports* **5**:14259. doi:10.1038/srep14259

Tyson JJ, Csikasz-Nagy A, Novak B. 2002. The dynamics of cell cycle regulation. *BioEssays* **24**:1095–1109. doi:10.1002/bies.10191

van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS. 2013. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat Methods* **10**:896–902. doi:10.1038/nmeth.2592

van den Berk LCJ, Landi E, Walma T, Vuister GW, Dente L, Hendriks WJAJ. 2007. An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry* **46**:13629–13637. doi:10.1021/bi700954e

van Ham M, Hendriks W. 2003. PDZ domains-glue and guide. *Mol Biol Rep* **30**:69–82. doi:10.1023/a:1023941703493

Vega MC, Martínez JC, Serrano L. 2000. Thermodynamic and structural characterization of Asn and Ala residues in the disallowed II′ region of the Ramachandran plot. *Protein Science* **9**:2322–2328. doi:10.1110/ps.9.12.2322

Velan B, Grosfeld H, Kronman C, Leitner M, Gozes Y, Lazar A, Flashner Y, Marcus D, Cohen S, Shafferman A. 1991. The effect of elimination of intersubunit disulfide bonds on the activity, assembly, and secretion of recombinant human acetylcholinesterase. Expression of acetylcholinesterase Cys-580----Ala mutant. *J Biol Chem* **266**:23977–23984.

Vendruscolo M, Dobson CM. 2011. Protein Dynamics: Moore's Law in Molecular Biology. *Current Biology* **21**:R68–R70. doi:10.1016/j.cub.2010.11.062

Vierstraete E, Verleyen P, Baggerman G, D'Hertog W, Van den Bergh G, Arckens L, De Loof A, Schoofs L. 2004. A proteomic approach for the analysis of instantly released wound and immune proteins in Drosophila melanogaster hemolymph. *Proceedings of the National Academy of Sciences* **101**:470–475. doi:10.1073/pnas.0304567101

von Ossowski I, Oksanen E, von Ossowski L, Cai C, Sundberg M, Goldman A, Keinänen K. 2006. Crystal structure of the second PDZ domain of SAP97 in complex with a GluR-A C-terminal peptide. *FEBS J* **273**:5219–5229. doi:10.1111/j.1742-4658.2006.05521.x

Vorontsov II, Miyashita O. 2009. Solution and Crystal Molecular Dynamics Simulation Study of m4-Cyanovirin-N Mutants Complexed with Di-Mannose. *Biophysical Journal* **97**:2532–2540. doi:10.1016/j.bpj.2009.08.011

Wagner JR, Sørensen J, Hensley N, Wong C, Zhu C, Perison T, Amaro RE. 2017. POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J Chem Theory Comput* **13**:4584–4592. doi:10.1021/acs.jctc.7b00500

Walma T, Spronk CAEM, Tessari M, Aelen J, Schepens J, Hendriks W, Vuister GW. 2002. Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. *J Mol Biol* **316**:1101–1110. doi:10.1006/jmbi.2002.5402

Wang C, Bradley P, Baker D. 2007. Protein–Protein Docking with Backbone Flexibility. *Journal of Molecular Biology* **373**:503–519. doi:10.1016/j.jmb.2007.07.050

Wang DD, Ou-Yang L, Xie H, Zhu M, Yan H. 2020. Predicting the impacts of mutations on protein-ligand binding affinity based on molecular dynamics simulations and machine learning methods. *Computational and Structural Biotechnology Journal* **18**:439–454. doi:10.1016/j.csbj.2020.02.007

Wang J, Ma C, Fiorin G, Carnevale V, Wang T, Hu F, Lamb RA, Pinto LH, Hong M, Klein ML, DeGrado WF. 2011. Molecular Dynamics Simulation Directed Rational Design of Inhibitors Targeting Drug-Resistant Mutants of Influenza A Virus M2. *J Am Chem Soc* **133**:12834–12841. doi:10.1021/ja204969m

Wang J-L, Liu D, Zhang Z-J, Shan S, Han X, Srinivasula SM, Croce CM, Alnemri ES, Huang Z. 2000. Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proceedings of the National Academy of Sciences* **97**:7124–7129. doi:10.1073/pnas.97.13.7124

Wang L, Tharp S, Selzer T, Benkovic SJ, Kohen A. 2006. Effects of a distal mutation on active site chemistry. *Biochemistry* **45**:1383–1392.

Wang R, Lai L, Wang S. 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **16**:11–26. doi:10.1023/a:1016357811882

Wang S, Li W, Zhang R, Liu S, Xu J. 2016. CoinFold: a web server for protein contact prediction and contact-assisted protein folding. *Nucleic Acids Res* **44**:W361-366. doi:10.1093/nar/gkw307

Wang S, Sun S, Li Z, Zhang R, Xu J. 2017. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLOS Computational Biology* **13**:e1005324. doi:10.1371/journal.pcbi.1005324

Wang X, Minasov G, Shoichet BK. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* **320**:85–95. doi:10.1016/S0022-2836(02)00400-X

Waters ER, Vierling E. 2020. Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist* **227**:24–37. doi:10.1111/nph.16536

Wei G, Xi W, Nussinov R, Ma B. 2016. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem Rev* **116**:6516–6551. doi:10.1021/acs.chemrev.5b00562

Wei Q, Xu Q, Dunbrack RL. 2013. Prediction of phenotypes of missense mutations in human proteins from biological assemblies. *Proteins* **81**:199–213. doi:10.1002/prot.24176

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *science* **312**:111–114.

Whitney DS, Peterson FC, Volkman BF. 2011. A conformational switch in the CRIB-PDZ module of Par-6. *Structure* **19**:1711–1722. doi:10.1016/j.str.2011.07.018

Wiegand I, Hilpert K, Hancock REW. 2008. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3**:163–175. doi:10.1038/nprot.2007.521

Wilce MCJ, Aguilar M-Isabel, Hearn MTW. 1995. Physicochemical Basis of Amino Acid Hydrophobicity Scales: Evaluation of Four New Scales of Amino Acid Hydrophobicity Coefficients Derived from RP-HPLC of Peptides. *Anal Chem* **67**:1210–1219. doi:10.1021/ac00103a012

Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I, Karanicolas J, Stock G, Hamm P, Stote RH, Eberhardt J, Chebaro Y, Dejaegere A, Cecchini M, Changeux J-P, Bolhuis PG, Vreede J, Faccioli P, Orioli S, Ravasio R, Yan L, Brito C, Wyart M, Gkeka P, Rivalta I, Palermo G, McCammon JA, Panecka-Hofman J, Wade RC, Di Pizio A, Niv MY, Nussinov R, Tsai C-J, Jang H, Padhorny D, Kozakov D, McLeish T. 2019. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **27**:566–578. doi:10.1016/j.str.2019.01.003

Wong KF, Selzer T, Benkovic SJ, Hammes-Schiffer S. 2005. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proceedings of the National Academy of Sciences* **102**:6807–6812. doi:10.1073/pnas.0408343102

Woodrum BW, Maxwell JD, Bolia A, Ozkan SB, Ghirlanda G. 2013. The antiviral lectin cyanovirin-N: probing multivalency and glycan recognition through experimental and computational approaches. *Biochem Soc Trans* **41**:1170–1176. doi:10.1042/BST20130154

Xie W, Altamirano CV, Bartels CF, Speirs RJ, Cashman JR, Lockridge O. 1999. An Improved Cocaine Hydrolase: The A328Y Mutant of Human Butyrylcholinesterase is 4-fold More Efficient. *Mol Pharmacol* **55**:83–91. doi:10.1124/mol.55.1.83

Xiong W, Liu B, Shen Y, Jing K, Savage TR. 2021. Protein engineering design from directed evolution to de novo synthesis. *Biochemical Engineering Journal* **174**:108096. doi:10.1016/j.bej.2021.108096

Xu J. 2019. Distance-based protein folding powered by deep learning. *Proceedings of the National Academy of Sciences* **116**:16856–16865. doi:10.1073/pnas.1821309116

Xue L, Hou S, Tong M, Fang L, Chen X, Jin Z, Tai H-H, Zheng F, Zhan C-G. 2013. Preparation and in vivo characterization of a cocaine hydrolase engineered from human butyrylcholinesterase for metabolizing cocaine. *Biochem J* **453**:447–454. doi:10.1042/BJ20130549

Xue L, Ko M-C, Tong M, Yang W, Hou S, Fang L, Liu J, Zheng F, Woods JH, Tai H-H, Zhan C-G. 2011. Design, preparation, and characterization of high-activity mutants of human butyrylcholinesterase specific for detoxification of cocaine. *Mol Pharmacol* **79**:290–297. doi:10.1124/mol.110.068494

Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N. 2016. Conformational Tinkering Drives Evolution of a Promiscuous Activity through Indirect Mutational Effects. *Biochemistry* **55**:4583–4593. doi:10.1021/acs.biochem.6b00561

Yang L-W, Kitao A, Huang B-C, Gō N. 2014. Ligand-Induced Protein Responses and Mechanical Signal Propagation Described by Linear Response Theories. *Biophysical Journal* **107**:1415–1425. doi:10.1016/j.bpj.2014.07.049

Yang W, Xue L, Fang L, Chen X, Zhan C-G. 2010. Characterization of a high-activity mutant of human butyrylcholinesterase against (-)-cocaine. *Chem Biol Interact* **187**:148–152. doi:10.1016/j.cbi.2010.01.004

Ye F, Zhang M. 2013. Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem J* **455**:1–14. doi:10.1042/BJ20130783

Zacharias M. 2010. Accounting for conformational changes during protein–protein docking. *Current Opinion in Structural Biology*, Theory and simulation / Macromolecular assemblages **20**:180–186. doi:10.1016/j.sbi.2010.02.001

Zemla A, Venclovas C, Moult J, Fidelis K. 1999. Processing and analysis of CASP3 protein structure predictions. *Proteins* **Suppl 3**:22–29. doi:10.1002/(sici)1097-0134(1999)37:3+<22::aid-prot5>3.3.co;2-n

Zhan M, Hou S, Zhan C-G, Zheng F. 2014. Kinetic characterization of high-activity mutants of human butyrylcholinesterase for the cocaine metabolite norcocaine. *Biochem J* **457**:197–206. doi:10.1042/BJ20131100

Zhang Y. 2009. Protein structure prediction: when is it useful? *Curr Opin Struct Biol* **19**:145–155. doi:10.1016/j.sbi.2009.02.005

Zhang Y, Doruker P, Kaynak B, Zhang S, Krieger J, Li H, Bahar I. 2020. Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr Opin Struct Biol* **62**:14–21. doi:10.1016/j.sbi.2019.11.002

Zheng F, Xue L, Hou S, Liu J, Zhan M, Yang W, Zhan C-G. 2014. A highly efficient cocaine-detoxifying enzyme obtained by computational design. *Nat Commun* **5**:3457. doi:10.1038/ncomms4457

Zheng F, Yang W, Ko M-C, Liu J, Cho H, Gao D, Tong M, Tai H-H, Woods JH, Zhan C-G. 2008. Most efficient cocaine hydrolase designed by virtual screening of transition states. *J Am Chem Soc* **130**:12148–12155. doi:10.1021/ja803646t

Zheng F, Yang W, Xue L, Hou S, Liu J, Zhan C-G. 2010. Design of high-activity mutants of human butyrylcholinesterase against (-)-cocaine: structural and energetic factors affecting the catalytic efficiency. *Biochemistry* **49**:9113–9119. doi:10.1021/bi1011628

Zheng F, Zhan C-G. 2011. Enzyme-therapy approaches for the treatment of drug overdose and addiction. *Future Med Chem* **3**:9–13. doi:10.4155/fmc.10.275

Zimmerman MI, Hart KM, Sibbald CA, Frederick TE, Jimah JR, Knoverek CR, Tolia NH, Bowman GR. 2017. Prediction of new stabilizing mutations based on mechanistic insights from Markov state models. *ACS central science* **3**:1311–1321.

Zimmerman SS. 1985. Chapter 4 - Theoretical Methods in the Analysis of Peptide Conformation In: Hruby VJ, editor. Conformation in Biology and Drug Design. Academic Press. pp. 165–212. doi:10.1016/B978-0-12-304207-1.50010-2

Zlebnik NE, Brimijoin S, Gao Y, Saykao AT, Parks RJ, Carroll ME. 2014. Long-term reduction of cocaine self-administration in rats treated with adenoviral vector-delivered cocaine hydrolase: evidence for enzymatic activity. *Neuropsychopharmacology* **39**:1538–1546. doi:10.1038/npp.2014.3

Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol Biol Evol* **32**:132–143. doi:10.1093/molbev/msu281

Zou T, Woodrum BW, Halloran N, Campitelli P, Bobkov AA, Ghirlanda G, Ozkan SB. 2021. Local Interactions That Contribute Minimal Frustration Determine Foldability. *J Phys Chem B* **125**:2617–2626. doi:10.1021/acs.jpcb.1c00364

Zviling M, Leonov H, Arkin IT. 2005. Genetic algorithm-based optimization of hydrophobicity tables. *Bioinformatics* **21**:2651–2656. doi:10.1093/bioinformatics/bti405

APPENDIX A

EXPERIMENTAL METHODS AND SUPPLEMENT DATA FOR PLANT-EXPRESSED COCAINE HYDROLASE VARIANTS OF BUTYRYLCHOLINESTERASE

A.1 DNA Constructs

Previously, we reported that the full-length human WT BChE gene (UniProt accession number P06276) was optimized for expression in *N. benthamiana* plants (pBChE) (Geyer et al., 2010a, 2010b). A synthetic gene encoding the A328W/Y332A mutant of pBChE (Geyer et al., 2008; Sun et al., 2002a, 2002b) (named here $pBChE_{V1}$) was used as the template for successive rounds of site-directed mutagenesis using the QuickChange method (Stratagene; mutagenic primers are listed in Supplementary Table A.1) yielding the following mutants: Variant 2 ($pBChE_{V2}$): F227A/S287G/A328W/Y332A59, Variant 3 ($pBChE_{V3}$): A199S/S287G/A328W/Y332G15, Variant4 ($pBChE_{V4}$): A199S/F227A/S287G/A328W/Y332G18, and Variant 5 ($pBChE_{V5}$): F227A/S287G/A328W/Y332G (Brimijoin and co-workers, unpublished). The mutated genes (also listed in Supplementary Table A.2) were verified by DNA sequencing. A C-terminal hexahistidine tag was added to the plant-expression optimized variants of BChE. The resulting constructs were then cloned into a deconstructed tobacco mosaic virus (TMV)-based plant expression vector (MagnICON, kind gift of Nomad Inc.) to be used in *Agrobacterium tumefaciens*-mediated transient expression in *N. benthamiana*.

Table A.1: Oligonucleotides used for site-directed mutagenesis. Forward (F) and reverse (R) primers are shown with mutation sites indicated in lowercase

Name	Primer Sequence	Mutation
oTM607	(F) $^{5'}$CTTTGGAGAGTCTtctGGAGCTGCTTCTG$^{3'}$	A199S
oTM608	(R) $^{5'}$CAGAAGCAGCTCCagaAGACTCTCCAAAG$^{3'}$	A199S
oTM609	(F) $^{5'}$CCAATCTGGTTCCgctAATGCTCCTTGG$^{3'}$	F227A
oTM610	(R) $^{5'}$CCAAGGAGCATTagcGGAACCAGATTGG3'	F227A
oTM611	(F) $^{5'}$GGAACTCCTTTGggaGTGAACTTTGGTC$^{3'}$	S287G
oTM612	(R) $^{5'}$GACCAAAGTTCACtccCAAAGGAGTTCC$^{3'}$	S287G
oTM613	(F) $^{5'}$GGATGAGGGTACAtggTTCCTTGTGggtGGAGCGCCTGG$^{3'}$	A328W/Y332G
oTM614	(R) $^{5'}$CCAGGCGCTCCaccCACAAGGAAccaTGTACCCTCATCC$^{3'}$	A328W/Y332G
oTM655	(F) $^{5'}$GGTTCCTTGTGgctGGAGCGCCTGG$^{3'}$	Y332A
oTM656	(R) $^{5'}$CCAGGCGCTCCagcCACAAGGAACC$^{3'}$	Y332A

Table A.2: Cocaine hydrolase variants of butyrylcholinesterase used in this study.

Name	Amino acid mutations	References
$pBChE_{V2}$	F227A/S287G/A328W/Y332A	Pancook, *et al.* 2003
$pBChE_{V3}$	A199S/S287G/A328W/Y332G	Pan, *et al.* 2005
$pBChE_{V4}$	A199S/F227A/S287G/A328W/Y332G	Xue, *et al.* 2013
$pBChE_{V5}$	F227A/S287G/A328W/Y332G	Brimijoin and co-workers, unpublished

A.2 Transient Recombinant Protein Production in Plants

An outline of the expression strategy is shown in Fig. 4.1a. All expression vectors were electroporated into *A. tumefaciens* strain GV3101 electro-competent cells. Transformed strains were screened via antibiotic selection as well as colony screen PCR and only positive colonies were used for downstream studies. Bacteria cultures were grown at 30 °C until mid-logarithmic phase, pelleted by centrifugation at 4,500 × *g* for 20 min at room temperature and then resuspended in infiltration buffer (10 mM 2-(N-morpholino) ethanesulfonic acid (MES), 10 mM magnesium sulfate heptahydrate, pH 5.5). Plants were infected either by needle-less syringe injection or by whole-plant vacuum infiltration.

Leaves infiltrated with each variant were harvested at the respective day of peak expression as determined in previous reports (Larrimore et al., 2013).

A.3 Enzymatic Assays

To evaluate cocaine hydrolysis, a sensitive radiometric assay was used as previously described (Brimijoin et al., 2002). Briefly, [^{3}H](−)-cocaine labeled on the benzene ring (50 Ci/mmol), purchased from PerkinElmer Life Sciences (Boston, MA), was used as a substrate with varying concentrations of (−)-cocaine. In the presence of enzyme this reaction proceeded at room temperature (25°C) until stopped by the addition of 0.02 M HCl. Any neutralized, liberated, labeled benzoic acid was then extracted with a toluene-based fluor and measured by scintillation counting. On the other hand, the substrate would fractionate into the aqueous phase and would not generate scintillation. Enzyme concentrations in the reaction mix were 800 ng/100 µL (1.21×10^{-1} µM) for WT pBChE and 4 ng/100 µL (6.06×10^{-4} µM) for pBChE$_{V4}$.

Choline ester hydrolysis activity was evaluated by a modified Ellman assay (Geyer et al., 2010a, 2007, 2005). Activity was measured using either butyrylthiocholine iodide (BTC, Sigma) or acetylthiocholine iodide (ATC, Sigma) at 30 °C in a Spectramax 190 spectrophotometer (Molecular Devices). Total soluble protein levels were determined by the Bradford protein assay (Bio-Rad Protein Assay Reagent, Bio-Rad) (Mor et al., 2001). The assay was conducted in 96-well plate format over varying concentrations of BTC or ATC in final well volume of 200 µL. To account for product formed by substrate self-hydrolysis, initial velocity of non-enzymatic hydrolysis was subtracted from initial velocity

of the matched enzyme-catalyzed reactions and reaction rates were then plotted as a function of substrate concentration.

Data were plotted using GraphPad Prism software, which was also used to fit the data by non-linear regression. The following models were fitted. For Michaelis-Menten kinetics we used Equation (A.1). For substrate inhibition/activation, we used Equation (A.2), following the model in Scheme 1 (Fig. A.1), as was suggested by Radić and coworkers (Radic et al., 1993).

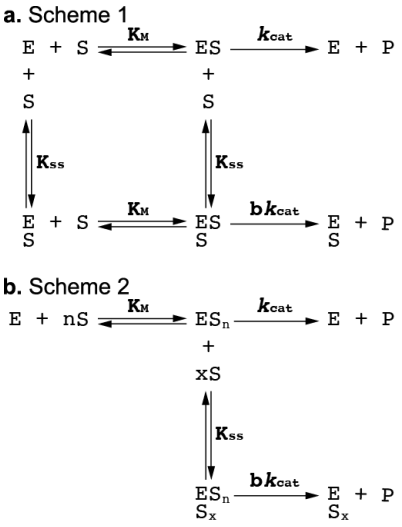


Figure A.1: Schematic diagrams describing the kinetics of cholinesterase-catalyzed hydrolysis of substrates**. (a)** Scheme 1 describes the reaction of cholinesterase (E)-catalyzed hydrolysis of substrates (S). K_{SS} is the dissociation constant of the peripheral site. The hydrolysis capacity (bK_{cat}) reflects the allosteric effect of substrate binding at the peripheral binding site. (**b**) Scheme 2 describes the reaction in terms of uncompetitive substrate inhibition/activation and cooperative substrate binding with characteristic Hill coefficients n and x that describe cooperativity or anticooperativity.

The Radić model (Radic et al., 1993) ascribes the allosteric effect of substrate binding at the peripheral binding site that causes a change in the catalytic rate bk_{cat}. When $b > 1$

we encounter substrate activation (WT BChE, see below), when $b < 1$ we encounter substrate inhibition and when $b = 1$ we have a Michaelian enzyme (Equation A.1). K_{ss} is the dissociation constant of the peripheral site.

Velocity *vs* substrate concentration data of some of the BChE variants described here fitted well to a model initially suggested by LiCata and Allewell (1997) for aspartate transcarbamylase (Scheme 2 in Fig. A.1). This model describes the reaction in terms of uncompetitive substrate inhibition/activation and cooperative substrate binding with characteristic Hill coefficients. The equation describing this model is Equation (A.3).

The Hill coefficients n and x need not be integers. Values greater than one describe cooperativity, while values of less than one describes anti-cooperativity. The parameters *b* and K_{ss} function in the same way as in Equation (A.2).

$$v = \left(\frac{V_{max}[S]}{K_M + [S]} \right) \tag{A.1}$$

$$v = \left(\frac{1 + b[S]/K_{ss}}{1 + [S]/K_{ss}} \right) \left(\frac{V_{max}}{1 + K_M/[S]} \right) \tag{A.2}$$

$$v = \frac{V_{max}(1 + b[S]^x/K_{ss}^x)}{1 + (K_M^n/[S]^n) + ([S]^x/K_{ss}^x)} \tag{A.3}$$

A.4 Inhibition

Inhibition studies were conducted with the OPs paraoxon (diethyl (4-nitrophenyl) phosphate) and iso-OMPA (N- [bis(propan-2-ylamino)phosphoryloxy-(propan-2-ylamino)phosphoryl]propan-2-amine), the carbamate neostigmine bromide ([3-(dimethylcarbamoyloxy)phenyl]-trimethylazanium;bromide), or the reversible

bisquaternary inhibitor BW284c51 (BW, [4-[5-[4-[dimethyl(prop-2-enyl)azaniumyl]phenyl]-3-oxopentyl]phenyl]-dimethyl-prop-2-enylazanium;dibromide). The four inhibitors were purchased from Sigma (St Louis, MO).

Preparations of BChE and variants thereof were incubated in 96-well plate format with indicated concentrations of the inhibitors for 30 min at room temperature followed by activity measurements based on modified Ellman assay using 1 mM BTC as the substrate. IC_{50} values were determined by non-linear regression (GraphPad Prism) fit according to Equation (A.4).

The inhibition rate constant (k_i) of $pBChE_{V4}$ treated with paraoxon was determined as previously described (Mionetto et al., 1997). Inhibition curves were statistically analyzed by the extra sum-of-squares F test (GraphPad Prism) together and were found to be significantly different from each other. Following up with individual comparisons to WT pBChE revealed statistical significance in all except the following: paraoxon inhibition of WT pBChE *vs* WT hBChE, Iso-OMPA inhibition of WT pBChE *vs* $pBChE_{V4}$, and neostigmine inhibition of WT pBChE *vs* $pBChE_{V3}$.

$$\text{residual BChE activity} = \frac{100}{1 + 10^{(log[I] - logIC_{50})}} \tag{A.4}$$

A.5 Purification

All extraction and purification procedures were carried out at 4 °C. Large-scale protein preparations were extracted from plant leaf tissue by blending in the presence of 50 mM sodium phosphate, 150 mM sodium metabisulfite, 1 mM EDTA, pH 8.0. Extract was filtered through double-layer miracloth and centrifuged at 22,000 × *g* for 30 min followed

by pH adjustment to pH 5.0 and further clarification by ammonium sulfate precipitation. The pellet was resuspended in cold 1X phosphate buffered saline (PBS) and dialyzed overnight against 1X PBS, pH 7.4 to remove salts and sodium metabisulfite. The clarified protein preparation was then subjected to sequential affinity chromatography steps with Concanavalin-A-Sepharose followed by procainamide affinity chromatography as previously described (Geyer et al., 2010a).

A.6 SDS-PAGE and Western Blot

Plant-derived protein preparations were resolved by SDS-PAGE on 8% polyacrylamide gels followed by staining with Pierce Silver Stain Kit (Thermoscientific). In parallel, protein was transferred to nitrocellulose membrane and decorated with rabbit polyclonal anti-hBChE antibodies (kindly provided by Dr. Oksana Lockridge) and anti-rabbit IgG-Horse Radish Peroxidase secondary antibodies (Santa Cruz Biotechnology) followed by chemiluminescence analysis using western blotting luminol reagent (Santa Cruz Biotechnology).

A.7 Size Exclusion HPLC

SEC-HPLC fractionation of purified preparations of pBChE$_{V4}$ was carried out as previously described using Alliance HPLC (Waters) with a Shodex KW-803 column (8 × 300 mm, Kawasaki) (Geyer et al., 2010a). All samples were run in filtered, degassed mobile phase buffer (20 mM Na_2HPO_4/NaH_2PO_4, pH 8.0, 200 mM NaCl, 0.04% NaN_3) at a flow rate of 0.5 mL/min. Molecular mass standards used were blue dextran (2000 kDa) and the proteins β-amylase (200 kDa), bovine serum albumin (66 kDa) and carbonic

anhydrase (29 kDa). Fractions were collected and analyzed for cholinesterase activity by the modified Ellman assay.

APPENDIX B

EXPERIMENTAL METHODS, AND SUPPLEMENTS USED FOR THE ROLE OF RIGID RESIDUES IN MODULATING TEM-1 β-LACTAMASE FUNCTION AND THERMOSTABILITY

B.1 Protein Expression and Purification

A pET24b plasmid encoding the gene for GNCA was a generous gift from Professor Jose Sanchez-Ruiz (Universidad de Granada). Genes encoding rigid design variants were codon-optimized for expression in *E. coli* cells. The native TEM-1 N-terminal periplasmic localization signal peptide (MSIQHFRVALIPFFAAFCLPVFA) was appended to the beginning of each gene; to facilitate purification, a C-terminal 6xHis affinity tag was added to the end of each gene. Genes encoding each rigid design were synthesized by IDT (Coralville, IA). The gene for wildtype TEM-1 was amplified from a pET21b vector using PCR. Genes encoding the rigid designs and TEM-1 were subcloned into the pET29b vector using the Gibson Assembly (Gibson et al., 2009) at a site that placed them under the control of the T7lac promoter. Genes encoding the uncoupled flexible residue variants were synthesized and cloned into pET29b vectors by GenScript (Piscataway, NJ, USA).

The sequences of all plasmids containing TEM-1, GNCA, rigid or flexible designs were confirmed by Sanger sequencing and were transformed via electroporation into BL21 Star (DE3) *E. coli* cells. Cells containing plasmids encoding GNCA were grown in lysogeny broth (LB) at 37 °C with shaking at 250 rpm until an $O.D._{600}$ of ~0.8 was reached. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM to induce expression; cells were grown for 3 h post induction. Cells containing plasmids encoding TEM-1 were grown in LB media at 20 °C with shaking at 220 rpm until an $O.D._{600}$ of ~0.8 was reached. Induction was again carried out with 1 mM IPTG and was allowed to proceed for 8–12 h. Cells containing plasmids encoding the rigid and flexible design variants were grown in 2xYT media to confluence overnight and pelleted by

centrifugation. After resuspension in fresh 2xYT media, protein expression was induced with 1 mM IPTG and cells were grown for an additional 20 h at 20 °C with shaking at 220 rpm.

After expression, the cells were pelleted via centrifugation at 4100× *g* for 15 min and the media was discarded. The cells were resuspended in TBS (50 mM Tris pH 8.0, 500 mM NaCl) and were again centrifuged at 4100×*g* for 15 min; the supernatant was discarded. The pellet was incubated at room temperature for 15 min with SET buffer (20% sucrose, 1 mM ethylenediaminetetraacetic acid (EDTA), 30 mM Tris pH 8.0, 1 μM phenylmethylsulfonyl fluoride (PMSF), 1 mg/mL lysozyme). After centrifugation at 4100× *g* for 15 min, the supernatant was decanted and saved. The cells were then shocked to release the periplasmic contents with ice cold 100 mM $MgCl_2$ at a 1:15 ratio of cell pellet weight to solution volume. Cells were vigorously agitated on ice for 15–30 min then centrifuged with the saved soluble fraction from the first stage at 4 °C for 60 min at 12,000×*g*.

The supernatant was then loaded onto a 5 mL nitrilotriacetic acid agarose (Ni-NTA) (Millipore Sigma, Burlington, MA, USA) column, washed with 5 column volumes of a low imidazole buffer (25 mM Tris pH 8.0, 150 mM NaCl, 15 mM imidazole), and eluted with a high imidazole buffer (25 mM Tris pH 8.0, 150 mM NaCl, 500 mM imidazole). All proteins were then subjected to a second purification step using anion exchange chromatography: Proteins were concentrated to a volume of 0.5–1 mL, diluted into the loading buffer (50 mM Tris, pH 9.0, 50 mM NaCl) and loaded directly onto the 5 mL Hi Trap Q Fast Flow column (Millipore Sigma, Burlington, MA, USA). The column was

washed with 5 column volumes of the loading buffer and eluted with 50 mM Tris, pH 9.0 250 mM NaCl. Protein purity was verified by SDS-PAGE (Figure B.3).

B.2 Circular Dichroism Characterization of Protein Folding and Stability

Far-ultraviolet circular dichroism (CD) measurements were performed in triplicate on a Jasco J-815 spectrophotometer (Jasco, Inc, Easton, MD, USA) equipped with a Peltier temperature controller. Wavelength scans were measured from 300 to 180 nm at room temperature with 1 nm steps using a 1 nm bandwidth, 5 nm/min scan rate; reported data represent an average of three independent scans. Thermal melts were monitored by the absorption signal at 222 nm with a temperature slope of 5 °C/min. For wavelength scans and thermal melts, the purified protein was in a TBS buffer (10mM Tris 50 mM NaCl, pH 7.0) in a cuvette with a 1 mm pathlength. Protein concentrations were calculated in triplicate using the absorbance at 280 nm and absorption coefficients as calculated by the ProtParam tool in the Expasy software suite (Gasteiger et al., 2005). Protein concentrations ranged between 0.18–0.25 mg/mL for all scans. Thermal melt curves were fitted using nonlinear regression least squares fit with the Hill equation in the GraphPad Prism version 9.0.0 for Windows, GraphPad software, San Diego, California, USA.

B.3 MIC Assays

Minimal inhibitory concentrations of ampicillin (MIC_{amp}) were performed in triplicate on 96-well plates (Wiegand et al., 2008). For each designed protein, TEM-1 and GNCA, five colonies were picked from a fresh agar plate and used to inoculate a 5 mL culture of LB, which was grown to confluence overnight at 37 °C. Overnight cultures were diluted in LB with 1 mM IPTG to a final working concentration of 5×10^5 cfu/mL. Three stock

solutions of ampicillin were independently prepared at 6000 μg/mL in LB with 1 mM IPTG and each solution was subsequently diluted in steps of 0.5 through the addition of LB with 1 mM IPTG to yield a final range of concentrations of 6–3000 μg/mL. The ampicillin concentrations for GNCA and the rigid designs were prepared at 400 μg/mL in LB with 1 mM IPTG and each solution was diluted in steps of 0.6 for a final concentration range of 2–200 μg/mL. The 96-well plates were covered with a fitted lid and incubated at 37 °C for 20 h. All optical density measurements were carried out at 600 nm using a SpectraMax M5 (Molecular Devices, LLC, San Jose, CA, USA); the absorbance of the buffer was subtracted from each measurement. To establish the lowest concentration of antibiotic that inhibited growth, a buffer-subtracted value ≥ 0.1 was used as the threshold for bacterial growth in each well. The MIC_{amp} was determined to be the lowest concentration of ampicillin that inhibited growth of the *E. coli* cells.

B.4 Detailed Rosetta Methods

All calculations were carried out using Rosetta version:
442bff4fb7bf2ccb44655e8d15276c9bccfbbd0. The
following command line was used to minimize the total energy of the 1btl crystal structure from the
Protein Data Bank using the Rosetta relax protocol:
<Path to>/Rosetta/main/source/bin/relax.default.linuxgccrelease -s
<input_file> @<path to>/relax.flags
The contents of relax.flags was:
-nstruct 1
-relax:default_repeats 5
-relax:constrain_relax_to_start_coords
-relax:coord_constrain_sidechains
-relax:ramp_constraints false
-ex1
-ex2
-use_input_sc
-flip_HNQ
-ignore_unrecognized_res

-relax:coord_cst_stdev 0.5
The DesignAround protocol was initiated with the following command line:
<path to>/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease -
out:nstruct 25 -jd2:ntrials 50 -parser:protocol <path to>/design.xml -
packing:resfile <path to>[resfile] -database <path to>/Rosetta/main/database
-out::overwrite -s <input file>
@<path to>/general_design.flags
Where the contents of general_design.flags was:
-run:preserve_header
-output_virtual true
-use_input_sc
-no_his_his_pairE
-score::hbond_params correct_params
-lj_hbond_hdis 1.75
-lj_hbond_OH_donor_dis 2.6
-linmem_ig 10
-nblist_autoupdate true
-in:ignore_unrecognized_res
-out::overwrite
And the contents of design.xml was:
<ROSETTASCRIPTS>
<SCOREFXNS>
<ScoreFunction name=”ref2015” weights=”ref2015.wts”/>
</SCOREFXNS>
<TASKOPERATIONS>
<ReadResfile name=”read_res” filename<path_to_resfile>”/>
<DesignAround name=”des_aro” design_shell=”<desired_design_sphere>”
resnums=”<target_residue>” repack_shell=”<design_sphere+4>” allow_design=”1”
resnums_allow_design=”0”/>
</TASKOPERATIONS>
<MOVERS>
<PackRotamersMover name=”prm” scorefxn=”ref2015”
task_operations=”des_aro,read_res”/>
<MinMover name=”min” scorefxn=”ref2015” chi=”1” bb=”0” jump=”ALL”
type=”dfpmin_armijo_nonmonotone” tolerance=”0.001” max_iter=”1000”/>
<MinMover name=”min_bb” scorefxn=”ref2015” chi=”1” bb=”1” jump=”ALL”
type=”dfpmin_armijo_nonmonotone” tolerance=”0.001” max_iter=”1000”/>
<GenericMonteCarlo name=”multi_min” mover_name=”min_bb”
scorefxn_name=”ref2015” trials=”10” sample_type=”low” temperature=”0.6”
drift=”0” recover_low=”1” preapply=”0”/>
</MOVERS>
<PROTOCOLS>
<Add mover_name=”prm”/>

<Add mover_name=”multi_min”/>
<Add mover_name=”prm”/>
</PROTOCOLS>
</ROSETTASCRIPTS>
The content of the resfile was:
ALLAA EX 1 EX 2 USE_INPUT_SC
start
#1 A PIKAA H
2 A PIKAA P
#3 A PIKAA E
#4 A PIKAA T
#5 A PIKAA L
#6 A PIKAA V
#7 A PIKAA K
#8 A PIKAA V
#9 A PIKAA K
#10 A PIKAA D
#11 A PIKAA A
#12 A PIKAA E
#13 A PIKAA D
#14 A PIKAA Q
#15 A PIKAA L
#16 A PIKAA G
#17 A PIKAA A
#18 A PIKAA R
19 A PIKAA V #rigid resi
20 A PIKAA G #rigid resi
#21 A PIKAA Y
#22 A PIKAA I
#23 A PIKAA E
#24 A PIKAA L
#25 A PIKAA D
#26 A PIKAA L
#27 A PIKAA N
#28 A PIKAA S
#29 A PIKAA G
#30 A PIKAA K
#31 A PIKAA I
#32 A PIKAA L
#33 A PIKAA E
#34 A PIKAA S
#35 A PIKAA F
#36 A PIKAA R
37 A PIKAA P

#38 A PIKAA E
#39 A PIKAA E
#40 A PIKAA R
#41 A PIKAA F
42 A PIKAA P
#43 A PIKAA M
#44 A PIKAA M
45 A PIKAA S #Active site
#46 A PIKAA T
#47 A PIKAA F
48 A PIKAA K #Active site
#49 A PIKAA V
#50 A PIKAA L
51 A PIKAA L #rigid resi
#52 A PIKAA C
#53 A PIKAA G
#54 A PIKAA A
#55 A PIKAA V
#56 A PIKAA L
#57 A PIKAA S
#58 A PIKAA R
#59 A PIKAA I
#60 A PIKAA D
#61 A PIKAA A
#62 A PIKAA G
#63 A PIKAA Q
#64 A PIKAA E
#65 A PIKAA Q
#66 A PIKAA L
#67 A PIKAA G
#68 A PIKAA R
#69 A PIKAA R
#70 A PIKAA I
#71 A PIKAA H
#72 A PIKAA Y
#73 A PIKAA S
#74 A PIKAA Q
#75 A PIKAA N
#76 A PIKAA D
#77 A PIKAA L
#78 A PIKAA V
#79 A PIKAA E
#80 A PIKAA Y
#81 A PIKAA S

82 A PIKAA P
#83 A PIKAA V
#84 A PIKAA T
#85 A PIKAA E
#86 A PIKAA K
#87 A PIKAA H
#88 A PIKAA L
#89 A PIKAA T
#90 A PIKAA D
#91 A PIKAA G
#92 A PIKAA M
#93 A PIKAA T
#94 A PIKAA V
#95 A PIKAA R
#96 A PIKAA E
97 A PIKAA L #rigid resi
#98 A PIKAA C
#99 A PIKAA S
#100 A PIKAA A
#101 A PIKAA A
#102 A PIKAA I
#103 A PIKAA T
#104 A PIKAA M
105 A PIKAA S #Active site
#106 A PIKAA D
107 A PIKAA N #Active site
#108 A PIKAA T
#109 A PIKAA A
#110 A PIKAA A
#111 A PIKAA N
#112 A PIKAA L
#113 A PIKAA L
#114 A PIKAA L
#115 A PIKAA T
#116 A PIKAA T
#117 A PIKAA I
#118 A PIKAA G
#119 A PIKAA G
120 A PIKAA P
#121 A PIKAA K
#122 A PIKAA E
#123 A PIKAA L
#124 A PIKAA T
#125 A PIKAA A

#126 A PIKAA F
#127 A PIKAA L
#128 A PIKAA H
#129 A PIKAA N
#130 A PIKAA M
#131 A PIKAA G
#132 A PIKAA D
#133 A PIKAA H
#134 A PIKAA V
#135 A PIKAA T
#136 A PIKAA R
#137 A PIKAA L
#138 A PIKAA D
#139 A PIKAA R
#140 A PIKAA W
141 A PIKAA E #Active site
142 A PIKAA P #This proline is really important for folding stability
#143 A PIKAA E
#144 A PIKAA L
#145 A PIKAA N
#146 A PIKAA E
#147 A PIKAA A
#148 A PIKAA I
149 A PIKAA P
#150 A PIKAA N
#151 A PIKAA D
#152 A PIKAA E
#153 A PIKAA R
#154 A PIKAA D
#155 A PIKAA T
#156 A PIKAA T
#157 A PIKAA M
158 A PIKAA P
#159 A PIKAA V
#160 A PIKAA A
#161 A PIKAA M
#162 A PIKAA A
#163 A PIKAA T
#164 A PIKAA T
#165 A PIKAA L
#166 A PIKAA R
#167 A PIKAA K
#168 A PIKAA L
#169 A PIKAA L

#170 A PIKAA T
#171 A PIKAA G
#172 A PIKAA E
#173 A PIKAA L
#174 A PIKAA L
#175 A PIKAA T
#176 A PIKAA L
#177 A PIKAA A
#178 A PIKAA S
#179 A PIKAA R
#180 A PIKAA Q
#181 A PIKAA Q
#182 A PIKAA L
#183 A PIKAA I
#184 A PIKAA D
#185 A PIKAA W
#186 A PIKAA M
#187 A PIKAA E
#188 A PIKAA A
#189 A PIKAA D
#190 A PIKAA K
#191 A PIKAA V
#192 A PIKAA A
#193 A PIKAA G
194 A PIKAA P
#195 A PIKAA L
#196 A PIKAA L
#197 A PIKAA R
#198 A PIKAA S
#199 A PIKAA A
#200 A PIKAA L
201 A PIKAA P
#202 A PIKAA A
#203 A PIKAA G
#204 A PIKAA W
#205 A PIKAA F
#206 A PIKAA I
#207 A PIKAA A
#208 A PIKAA D
209 A PIKAA K #Active site
#210 A PIKAA S
#211 A PIKAA G
#212 A PIKAA A
#213 A PIKAA G

#214 A PIKAA E
#215 A PIKAA R
#216 A PIKAA G
#217 A PIKAA S
218 A PIKAA R #Active site
#219 A PIKAA G
#220 A PIKAA I
#221 A PIKAA I
#222 A PIKAA A
#223 A PIKAA A
#224 A PIKAA L
#225 A PIKAA G
226 A PIKAA P
#227 A PIKAA D
#228 A PIKAA G
#229 A PIKAA K
230 A PIKAA P
#231 A PIKAA S
#232 A PIKAA R
#233 A PIKAA I
#234 A PIKAA V
235 A PIKAA V #rigid resi
#236 A PIKAA I
#237 A PIKAA Y
#238 A PIKAA T
#239 A PIKAA T
#240 A PIKAA G
#241 A PIKAA S
#242 A PIKAA Q
#243 A PIKAA A
#244 A PIKAA T
#245 A PIKAA M
#246 A PIKAA D
#247 A PIKAA E
#248 A PIKAA R
#249 A PIKAA N
#250 A PIKAA R
#251 A PIKAA Q
#252 A PIKAA I
#253 A PIKAA A
#254 A PIKAA E
#255 A PIKAA I
#256 A PIKAA G
#257 A PIKAA A

#258 A PIKAA S
#259 A PIKAA L
#260 A PIKAA I
#261 A PIKAA K
#262 A PIKAA H
#263 A PIKAA W
Sequences of Designed Proteins in FASTA format

>Native β-lactamase signal peptide
MSIQHFRVALIPFFAAFCLPVFA
>Rdg262a
HPETLVKVKDAEDQLGARVGFQLTDLNSGKILEYFRAEERFPMMSTFKVLLCGA
VLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNT
AANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTQPKAMA
QTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIACKSGAGERG
SRGIIAALGPDGKPSRIVVIFTTGSQATMDERNRQIAEIGASLIKHW
>Rdg262b
HPETLVKVKDAEDQLGARVGFILLDLNSGKILESFRAEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTTPRAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGQIAALGPDGKPSRIVVIMTTGSQATMDERNRQIAEIGASLIKHW
>Rdg44a
HPETLVKVKDAVDQLGAPVGMIELDLNSGKILESYNPEERFPMMSTFKVLLCGA
VLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNT
AANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMA
TTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERG
SRGIIAALGPDGKPSRIVVIMMTGSQATMDERNRAIAEIGASLIKHW
>Rdg44b
HPETLVKVKKAVDDLGAPVGFIELDLNSGKILESYKPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGIIAALGPDGKPSRIVVTMTSGSQATMDERNRAIAEIGASLIKHW
>Rdg44c
HPETLVVVKQAEDKLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGSI
GIIAALGPDGKPSRIVVIYATGSQATMDELNRAIAEIGASLIKHW
>Flx226a
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT

TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGIIAALGPNGKPSRIVVIYTTGSQATMDERNRQIAEIGASLFKHW
>Flx226b
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPPGWFIADKSGAGERGS
RGIIAAFGPNGVPTRIVVIYTTGSQATMDERNRQIAEIGASLFKHY
>Flx226c
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPPGWFIADKSGAGERGS
RGIIAALGPNGVPSRIVVIYTTGSQATMDERNRQIAEIGASLFKHW
>Flx256
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMAADKVAGPLLRSALPPGWFIADKSGAGERGS
RGIIASLGPNGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW
>Flx55
HPETLVKVKDAEDQLGARVGYILLDADSGKILEAFRPEERFPMMSTFKVLLCGA
VLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNT
AANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPRAMA
ETLRKLLLGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERG
SRGIIAMLGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW

Table B.1: Mutations present in the computationally designed proteins and the distance of the nearest mutation to a catalytic residue in angstroms.

Designed Protein	Mutations	Distance from closest catalytic residue (Å)
Rdg44a	E37V, R43P, Y46M, F60Y, R61N, Y264M, T265M, Q278A	11.8
Rdg44b	D35K, E37V, Q39D, R43P, Y46F, F60Y, R61K, I263T, Y264M, T266S, Q278A	11.1
Rdg44c	K32V, D35Q, Q39K, R244I, T265A, R275L, Q278A	9.7
Rdg262a	Y46F, I47Q, E48L, L49T, S59Y, P62A, M182Q, V184K, T188Q, D233C, Y264F	3.9
Rdg262b	Y46F, E48L, P62A, M182T, V184R, I246Q, I246M	5.8
Flx226a	D254N, I287F	21.0
Flx226b	A227P, L250F, D254N, K256V, S258T, I287F, W290Y	12.1
Flx226c	A227P, D254N, K256V, I287F	17.5
Flx256	E212A, A227P, A249S, D254N	9.0
Flx55	E48L, L51A, N52D, S59A, V184R, T188E, T195L, A249M	9.8

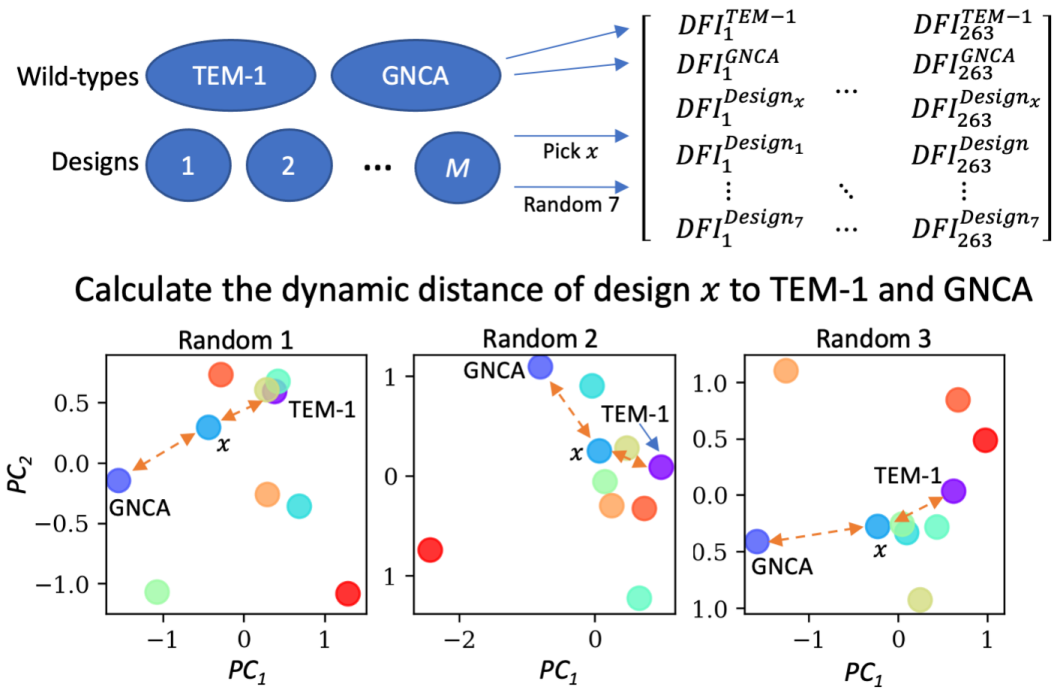


Figure B.1: Schematic of the dynamic distance calculation process. The dynamic profile of each design (using the dfi metric) is clustered using PCA in a set composed of TEM-1, GNCA, and seven randomly chosen designs. The dynamic distance of the design from TEM-1 and GNCA is calculated. Notably, the dynamic distance of the designed protein from TEM-1 and GNCA varies according to the set of proteins incorporated. To capture a statistically accurate distribution, this procedure is iterated a thousand times, each time varying the set of designed proteins.

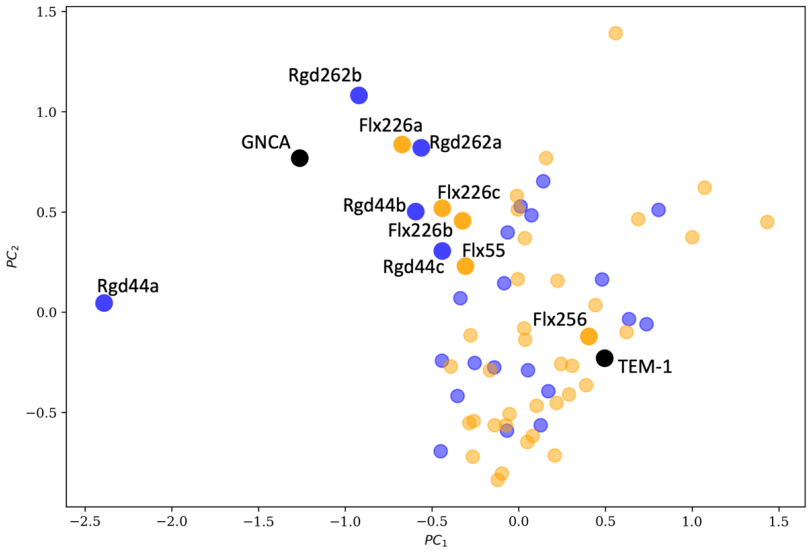


Figure B.2: PCA of a selection of the flexible and rigid designed proteins. The rigid designs with allosteric dynamic coupling to the active site are marked with blue dots. Uncoupled flexible designs are marked with orange dots. TEM-1 and GNCA are shown as black dots. For both rigid and flexible designs, the variants chosen for experimental characterization are named and highlighted with darker colors.

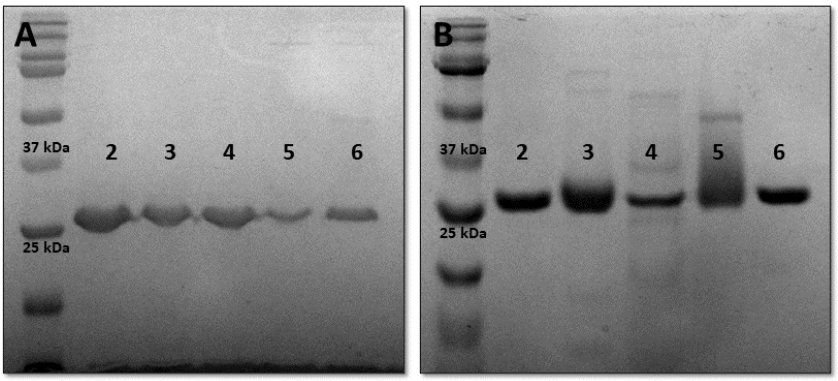


Figure B.3: 12% SDS PAGE gels of the purified designed proteins. The gels were stained with Coomassie Brilliant Blue G-250. For the gels, proteins were heat denatured. The protein standard (lane 1) is Bio-Rad Precision Plus Protein Kaleidoscope Prestained Protein Standards (A) Flx226a (lane 2) Flx226b (lane 3) Flx226c (lane 4) Flx256 (lane 5) Flx55 (lane 6) (B) TEM-1 (lane 2) GNCA (lane 3) Rdg44b (lane 4) Rdg44c (lane 5) Rdg262b (lane 6).

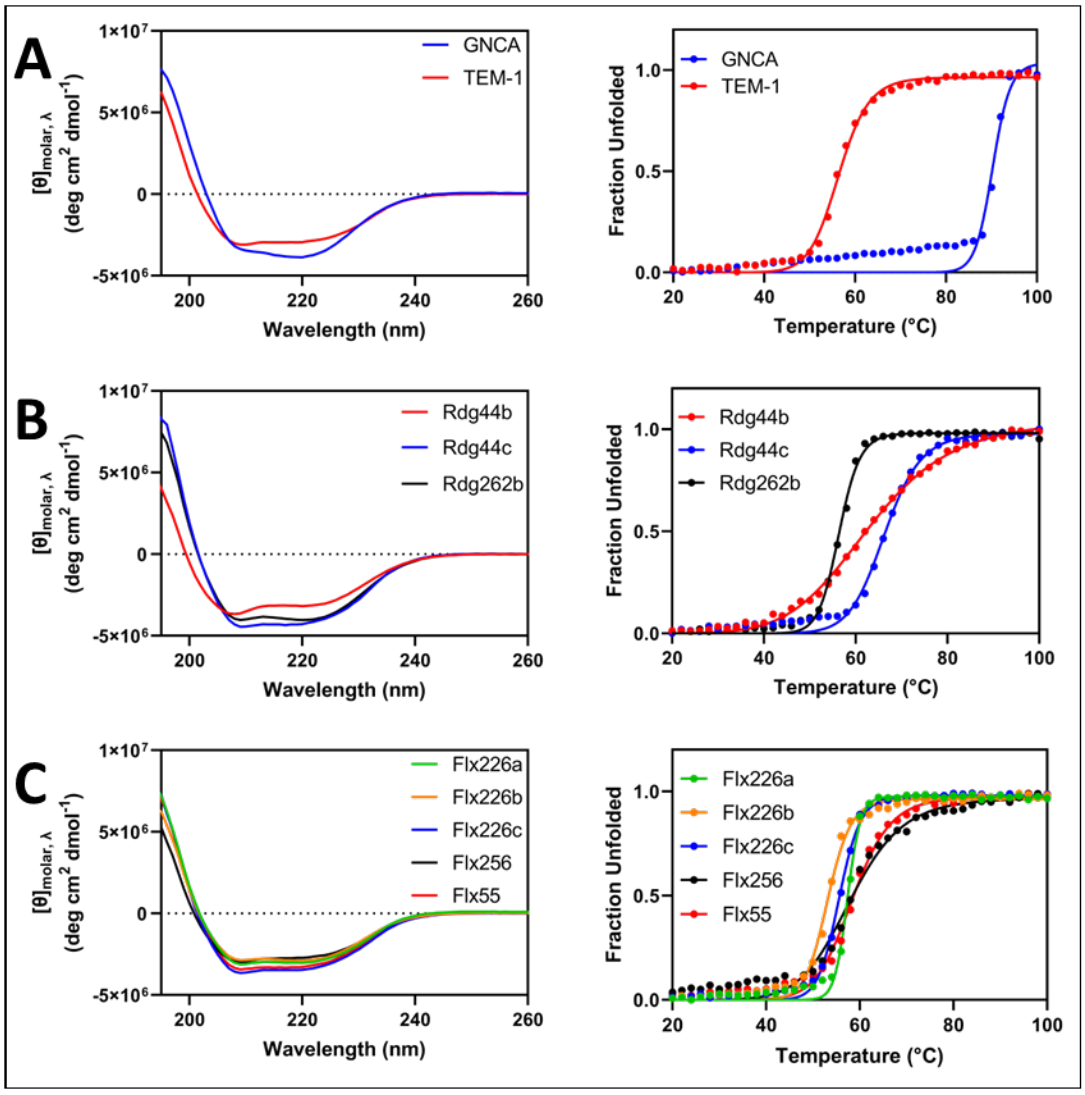


Figure B.4: Far-ultraviolet circular dichroism wavelength scans and thermal melts with fitted curves of (A) wild type GNCA and TEM-1 (B) protein designs targeting rigid residues and (C) protein designs targeting flexible residues. All measurements were performed in triplicate on a Jasco J-815 spectrophotometer and adjusted for protein concentration. Thermal melts were monitored by the absorption signal at 222 nm with a temperature slope of 5 °C/min. For wavelength scans and thermal melts, the purified protein was in TBS buffer (10mM Tris 50 mM NaCl, pH 7.0) in a cuvette with a 1 mm path length. Protein concentrations were calculated in triplicate using the absorbance at 280 nm and ranged between 0.18-0.25 mg/mL for all scans.

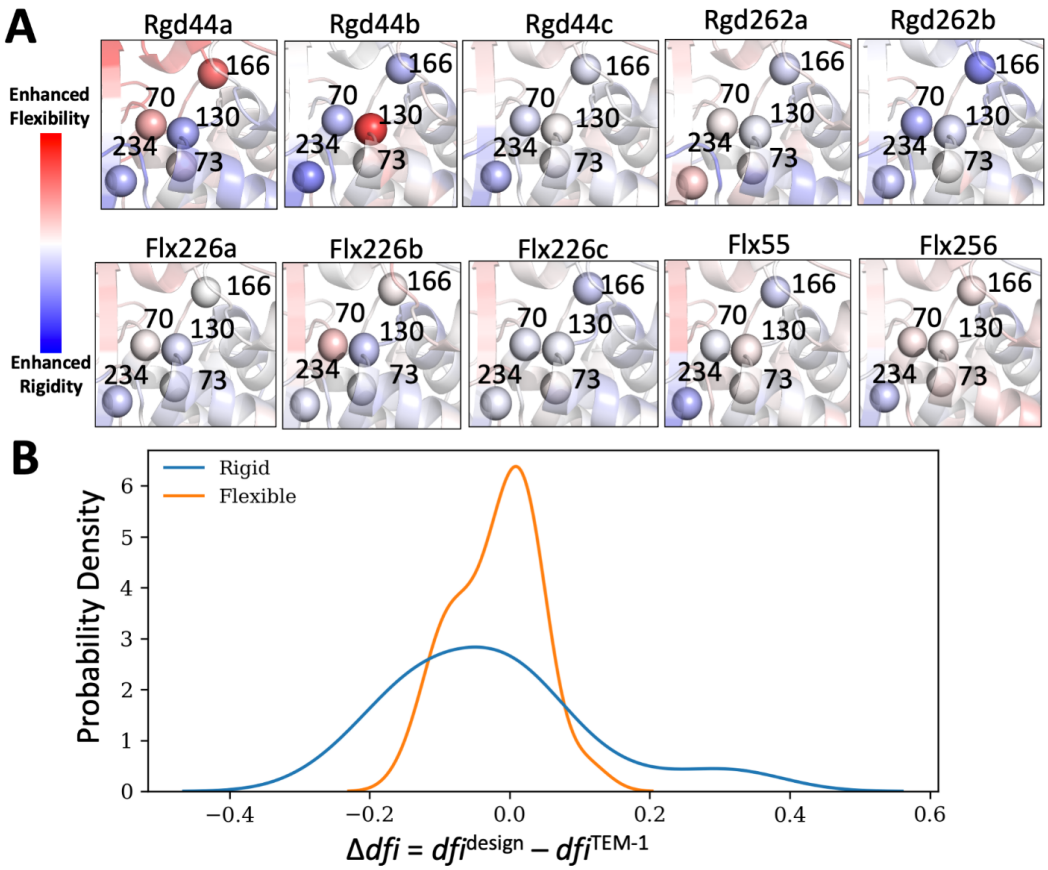


Figure B.5: The change in dynamics as measured by the Δ*dfi* mapped onto the catalytic residues of each experimentally characterized protein. A) Catalytic residues are modeled as spheres and color coded by their change in *dfi* score relative to TEM-1. B) The Δ*dfi* distribution of active site residues in the flexible and rigid designs. The flexible design distribution shows a low variance compared to that of the rigid designs. A change in *dfi* score of + 0.2 is noteworthy as it is indicative of a shift in flexibility. This analysis suggests that designing new interactions around a rigid residue that is dynamically coupled to the active site can allosterically modulate the flexibility/rigidity of the amino acids in the active site.

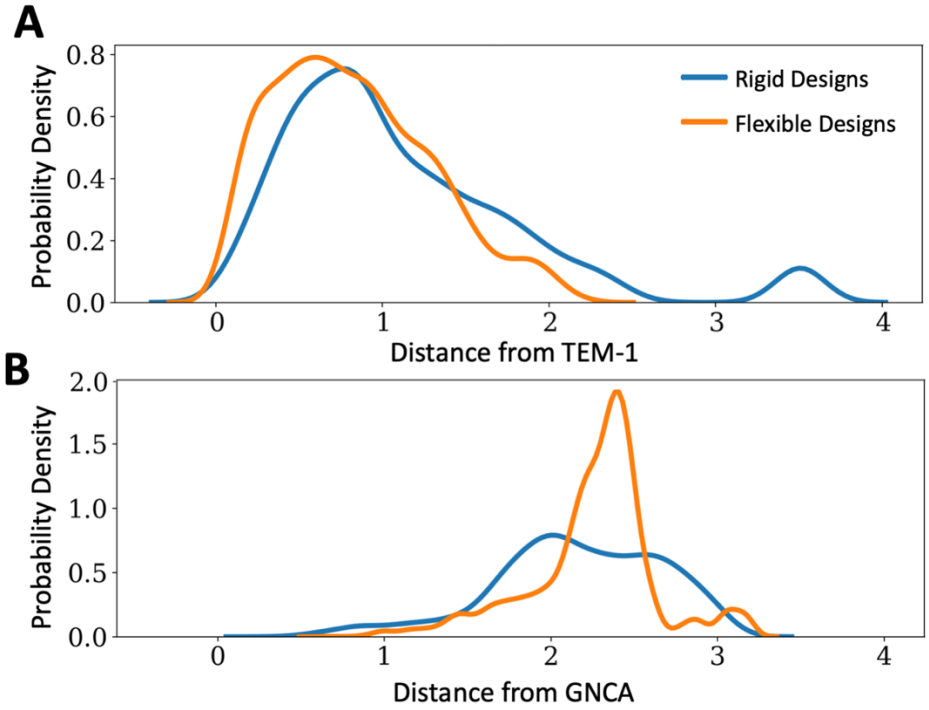


Figure B.6: Dynamic distance distribution from (A) TEM-1 and (B) GNCA for all experimentally characterized rigid (blue) and flexible designed proteins (orange). The distribution of the rigid designs shows a displacement moving away from TEM-1 and closer to GNCA. Inversely, the uncoupled flexible designs form a narrow distribution close to TEM-1 and further away from GNCA.

APPENDIX C

EXPERIMENTAL METHODS AND SUPPLEMENT DATA FOR DESIGN OF NOVEL CYANOVIRIN-N VARIANTS BY MODULATION OF BINDING DYNAMICS THROUGH DISTAL MUTATIONS

C.1 Mutant Proteins Cloning, Expression, and Purification

The genes for mutants (I34Y, I34K, and I34L) were generated by applying mutagenic primers to P51G-m4-gene sequence and amplifying by PCR. The constructs were subsequently cloned in pET26B vector between NdeI and XhoI sites and transformed in BL21(DE3) for expression and purification. The proteins were expressed from a 10 ml starter culture in LB broth overnight at 37°C, inoculated into 1 l LB medium. The culture was induced with 1 mM isopropyl thiogalactoside when OD reached 0.6 and grown for another 6–8 hr. Then, the cells were harvested by centrifugation, lysed in 6 M guanidine hydrochloride at pH 8.0, and sonicated for 10 min. The supernatant recovered after centrifugation was used to purify proteins with GE HisTrap HP column (GE Healthcare Bio-Sciences, Piscataway, NJ) and a Bio-Rad EconoPump (Bio-Rad, Richmond, CA) under denaturing conditions. In brief, the proteins were loaded on the column in Gu-HCl buffer, which was buffer exchanged by 8 M urea buffer. The nonspecific proteins were washed out by 4 M urea and 20 mM imidazole buffer, pH 8.0 and eluted with 2 M urea and 200 mM imidazole, pH 8.0 buffer before putting it for overnight dialysis against 10 mM Tris pH 8.0 and 100 mM NaCl buffer. The buffer was changed once during the night. The refolded protein was concentrated and re-purified to isolate the monomeric species by size exclusion chromatography using Sephadex 75 10/300 column on Agilent's Infinity 1260 system. The gel filtered protein was finally used for all the experiments.

C.2 CD Spectroscopy and T-melts

In CV-N family proteins, thermodynamic parameters like free energy of unfolding, enthalpy, and entropy cannot be extracted by thermal denaturation because the transition from folded to unfolded state is non-reversible (Patsalo et al., 2011), therefore melting temperatures are used. Far-UV CD spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a thermostatic cell holder, PTC 424S. Spectra were measured from 250 to 200 nm, using a scanning speed of 50 nm/min and a data pitch of 1.0 nm at 25°C. Samples concentration was approximately 15 µM in 10 mM Tris, pH 8.0, and 100 mM NaCl. For thermal denaturation experiments, the melting profile was monitored at 202 nm from 25°C to 90°C. The data points were plotted and fitted in Origin8.5 software to get apparent Tm.

C.3 Isothermal Titration Calorimetry (ITC)

ITC was performed at the Sanford-Burnham Medical Research Institute Protein Analysis Facility using ITC200 calorimeter from Microcal (Northampton, MA) at 23°C; 2.0 µl aliquots of solution containing between 3 and 10 mM Man2 were injected into the cell containing between 0.057 and 0.11 mM protein. Nineteen of 2.0 µl injections were made. The experiments were performed in 10 mM Tris, 100 mM NaCl, pH 8.0 buffer. ITC data were analyzed using Origin software provided by Microcal.

C.4 Chemical Denaturation Experiments

Chemical denaturation experiments were done by monitoring the shift in the intrinsic tryptophan fluorescence on Cary Eclipse instrument (Varian). Ten µM of protein samples

were incubated with increasing concentrations of guanidine hydrochloride in the range of 0–6 M in 50 mM Tris pH 8.0 buffer for 72 hr at 25°C. The emission spectra for the same were recorded by keeping the excitation wavelength at 295 nm and bandwidth of 1 nm. A ratio of fluorescence at 330 and 360 nm (I_{330}/I_{360}) was plotted at respective Gu-HCl concentrations, and the data points were fit to following sigmoidal equation to obtain C_m.

$$y = A2 + \frac{A1-A2}{1+e^{(x-x0)/dx}} \quad (C.1)$$

Where, A1and A2 are the initial and final 330/360 ratios and x0 is the concentration of Gu-HCl, where y= (A1 +A2) /2, or the point where 50 % of the population is unfolded. It is also denoted as C_m. The denaturation curve was used to calculate the free energy of the protein in the absence of denaturant (ΔG_{H2O}). Fraction unfolded (f_U) was calculated using the following formula:

$$f_U = (y_F - y_{obs})/(y_F - y_U) \quad (C.2)$$

where f_U, is the fraction unfolded, y_F is the value when there is no denaturant, y_{obs} is the value at each position and y_U is the value for unfolded protein. Since $f_{U+} f_F = 1$, the equilibrium constant, K, the free energy change can be calculated using

$$K = f_U / f_F \quad (C.3)$$

$$K = f_U / 1 - f_F \quad (C.4)$$

$$\Delta G = -RT\ln K \quad (C.5)$$

Where R is the gas constant whose value is 1.987 cal/mol.K and T is the temperature of incubation, which was 298K. The value of ΔG is linear over a limited range of Gu-HCl. The linear fit over that range was extrapolated to obtain ΔG_{H2O}.

C.5 Crystallization and Structure Determination

I34Y was purified as discussed previously and the monomeric gel filtered protein was concentrated to 8 mg/ml. We got the crystals in 2 M ammonium sulphate and 5% (v/v) 2-propanol after screening it in Index HT screen from Hampton Research. The protein crystals were reproduced using same condition in hanging drop method. For protein crystals with dimannose, the crystals were incubated in 1.2-fold molar excess of dimannose. Single needle-like crystals were picked up and cryo-preserved in 25% glycerol before freezing them for data collection at Synchrotron ALS, beamline 8.2.1. Single crystal diffraction was measured at wavelength of 0.999 A with ADSC quantum 315r detector. The data were evaluated to resolution of 1.25 A. The data acquired was indexed using XDS and scaled by the aimless package from CCP4i program suite. The structural coordinates and phase were determined by molecular replacement using 2RDK PDB code. The structure of I34Y of CV-N is deposited under PDB accession code 6X7H. The structure was further refined in Coot.

Table C.1: DFI, DCI, RaptorX, Evcoupling, and MISTIC metrics are used to identify residues in TEM-1 β-lactamase for the four unique categories.

		Dynamically Coupled to Binding Sites										Not Dynamically Coupled to Binding Sites									
		1										0									
		Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid	Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid
					Minimum	Average									Minimum	Average					
Coevolved with Binding Sites	1	44	0.20	0.94	14	19	0.92	0.87	0.32	V	AFILMV	51	0.56	0.35	17	24	0.77	0.77	0.92	L	ACEFGIKLM NPQRSTV
		45	0.10	0.89	15	19	0.93	0.99	0.67	G	AGR	121	0.55	0.63	13	17	0.66	0.91	0.91	E	ADEGHKMN QRSTV
		46	0.05	0.86	15	19	0.90	0.88	0.69	Y	AFGILMTVY	142	0.35	0.52	13	19	0.83	0.61	0.77	I	ACFHILMNQ RSTVY
		47	0.13	0.85	16	20	0.89	0.84	0.97	I	ACDEFGHIL MNQRSTVW Y	148	0.30	0.68	11	16	0.94	0.78	0.95	L	ADFILMQTV WY
		122	0.15	0.92	10	14	0.71	0.92	0.86	L	ACFHILMSV	155	0.58	0.51	17	23	0.68	0.91	0.91	M	ACEFGHIKL MNPQRSTV W
		137	0.13	0.93	10	15	0.93	0.95	0.95	L	AEFGIKLMQ RSVY	183	0.04	0.50	14	17	0.84	0.97	0.75	P	ALPS
		179	0.25	0.85	11	13	0.98	0.97	0.93	D	ADGNR	199	0.61	0.45	17	23	0.74	0.99	0.67	L	ILV
		181	0.19	0.90	12	15	0.92	0.99	0.71	T	ACILSTV	220	0.49	0.67	10	16	0.95	0.96	0.99	L	ACEFGHILM QRSTVWY
		182	0.10	0.70	15	18	0.66	0.98	0.65	M	ACKMNRSTV	222	0.61	0.56	11	18	0.93	0.99	0.74	R	ACGKLNPQR SV
		187	0.08	0.73	13	17	0.84	0.91	0.99	A	ACEGILMNQ RSTV	223	0.71	0.51	14	20	0.88	0.88	0.64	S	ADEGHKLPQ RSTV
		190	0.19	0.78	12	16	0.95	0.94	0.86	L	AFHILMNST VWY	224	0.69	0.59	16	22	0.83	0.68	0.75	A	ADEGHIKLNP RSTVY
		262	0.01	0.94	10	15	0.96	0.64	0.99	V	ACGILNSTV	225	0.74	0.45	15	22	0.84	0.76	0.99	L	AFIKLMQST VW
		263	0.01	0.90	11	15	0.94	0.80	0.99	I	ACFGILMST V	227	0.98	0.67	20	27	0.66	0.62	0.89	A	ADEFGHIKL MNPQRSTV
Not Coevolved with Binding Sites	0	83	0.74	0.89	16	22	0.16	0.41	0.44	R	ADEGHKLM NQRSTVY	35	0.80	0.81	21	27	0.52	0.09	0.53	D	ADEGHIKLM NPQRSTVW Y
		84	0.77	0.84	17	23	0.33	0.41	0.48	V	ACDEFGHIKL MNQRSTVW Y	52	0.75	0.28	21	27	0.44	0.49	0.50	N	ADEFGHIKM NPQRSTV
		93	0.92	0.92	19	24	0.16	0.35	0.51	R	ADEHKLNQR STVY	197	0.71	0.42	21	26	0.33	0.39	0.49	E	ACDEGHKLN PQRSTY
		94	0.90	0.93	19	23	0.16	0.29	0.40	R	ACDEFGHIKL MNPQRSTV W	201	0.79	0.43	20	25	0.33	0.09	0.32	L	ACDEFGHIKL MNPQRSTV Y
												289	0.85	0.72	23	30	0.52	0.10	0.50	H	ACDEFGHIKL MNPQRSTV WY

Table C.2: DFI, DCI, RaptorX, Evcoupling, and MISTIC metrics are used to identify residues in CV-N for the ICDC categories.

		Dynamically Coupled to Binding Sites										Not Dynamically Coupled to Binding Sites									
		1										0									
		Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid	Residue	DFI Score	DCI Score	Distance		RAPTORX Score	MISTIC Score	EVCOUPLING Score	Current Amino acid	Available Amino acid
					Minimum	Average									Minimum	Average					
Coevolved with Binding Sites	1	34	0.17	0.71	16	22	0.65	0.93	0.71	I	FIKLMVY	18	0.54	0.51	16	21	0.65	0.98	0.84	L	FILMV
		61	0.15	0.83	11	15	0.81	0.76	0.86	T	ACFIKLMQST VWY	66	0.95	0.27	21	26	0.27	0.91	0.94	S	ADEFGKLMN PQRST
		71	0.01	0.93	11	13	0.94	0.92	0.98	A	ACDEGQSTV										
Not Coevolved with Binding Sites	0	32	0.02	0.91	23	29	0.44	0.92	0.48	S	ACDHLNSTV	65	0.88	0.49	21	24	0.26	0.7	0.69	G	ADEFGHIKN PQRSTWY
		99	0.35	0.85	18	26	0.53	0.3	0.37	K	ADEFHIKLM NQRSTV	67	0.82	0.29	18	23	0.34	0.79	0.86	S	ADEGHIKLM NPQRSTVY
												88	0.74	0.55	16	23	0.59	0.72	0.73	D	ADEGHNRST

Table C.3: The complete TEM-1 dynamic flexibility index (DFI), dynamic coupling index (DCI), RaptorX, Evcoupling, and MISTIC metric data.

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
26	0.89	0.77	24	30	0.58	0.11	0.75
27	0.91	0.75	26	31	0.62	0.09	0.55
28	0.87	0.78	25	31	0.52	0.21	0.74
29	0.73	0.81	21	27	0.72	0.25	0.81
30	0.77	0.76	22	27	0.68	0.22	0.94
31	0.82	0.76	24	29	0.44	0.11	0.68
32	0.72	0.84	21	26	0.62	0.13	0.91
33	0.63	0.83	19	24	0.80	0.50	0.88
34	0.78	0.79	21	26	0.62	0.51	0.67
35	0.80	0.81	21	27	0.52	0.09	0.53
36	0.69	0.83	17	23	0.72	0.95	0.72
37	0.73	0.80	17	24	0.78	0.97	0.56
38	0.86	0.82	20	27	0.58	0.59	0.88
39	0.86	0.81	18	26	0.62	0.37	0.88
40	0.78	0.80	15	23	0.82	0.57	0.72
41	0.82	0.82	17	25	0.82	0.42	0.87
42	0.66	0.89	15	22	0.89	0.93	0.83
43	0.52	0.90	15	21	0.91	0.54	0.52
44	0.20	0.94	14	19	0.92	0.87	0.32
45	0.10	0.89	15	19	0.93	0.99	0.67
46	0.05	0.86	15	19	0.90	0.88	0.69
47	0.13	0.85	16	20	0.89	0.84	0.97
48	0.16	0.86	16	22	0.83	0.56	0.74
49	0.32	0.83	17	23	0.84	0.21	0.44
50	0.52	0.65	19	25	0.52	0.90	0.86
51	0.56	0.35	17	24	0.77	0.77	0.92
52	0.75	0.28	21	27	0.44	0.49	0.50
53	0.78	0.28	23	29	0.33	0.33	0.92
54	0.70	0.32	20	26	0.52	0.88	0.44
55	0.71	0.31	22	28	0.52	0.22	0.82
56	0.65	0.62	21	26	0.76	0.39	0.70
57	0.58	0.77	21	27	0.83	0.63	0.76
58	0.54	0.78	21	26	0.82	0.01	0.39
59	0.47	0.76	20	24	0.68	0.51	0.91
60	0.38	0.87	20	24	0.80	0.72	0.98

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
61	0.30	0.89	20	23	0.58	0.99	0.68
62	0.25	0.61	19	22	0.62	0.37	0.97
63	0.32	0.75	20	22	0.71	0.24	0.29
64	0.24	0.82	18	21	0.75	0.64	0.72
65	0.17	0.88	16	18	0.91	0.66	0.41
66	0.07	0.89	12	15	0.91	0.99	0.59
67	0.14	0.94	9	13	0.98	0.98	0.48
68	0.09	0.98	6	10	0.96	0.71	0.97
69	0.17	0.99	4	8	0.99	0.89	0.89
70	0.15	0.99	0	6	0.99	0.99	0.99
71	0.06	0.99	4	8	0.98	0.93	0.32
72	0.09	0.99	4	9	0.98	0.98	0.86
73	0.05	0.99	0	8	0.99	0.99	0.94
74	0.08	0.99	4	10	0.98	0.35	0.88
75	0.11	0.98	5	12	0.90	0.25	0.60
76	0.03	0.99	5	12	0.92	0.90	0.91
77	0.03	0.97	6	13	0.93	0.84	0.83
78	0.21	0.97	9	15	0.81	0.77	0.84
79	0.28	0.97	10	17	0.62	0.86	0.82
80	0.25	0.92	11	17	0.66	0.98	0.90
81	0.49	0.86	13	19	0.58	0.99	0.61
82	0.67	0.91	15	21	0.44	0.89	0.99
83	0.74	0.89	16	22	0.16	0.41	0.44
84	0.77	0.84	17	23	0.33	0.41	0.48
85	0.87	0.91	19	25	0.16	0.75	0.78
86	0.94	0.91	20	27	0.01	0.18	0.67
87	0.95	0.92	22	28	0.16	0.42	0.67
88	0.93	0.86	20	26	0.16	0.69	0.79
89	0.84	0.80	18	23	0.16	0.73	0.96
90	0.90	0.89	18	24	0.02	0.45	0.88
91	0.83	0.87	17	22	0.33	0.88	0.59
92	0.93	0.89	19	24	0.16	0.69	0.88
93	0.92	0.92	19	24	0.16	0.35	0.51
94	0.90	0.93	19	23	0.16	0.29	0.40
95	0.89	0.91	18	21	0.80	0.87	0.95
96	0.96	0.90	19	23	0.62	0.28	0.69

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
97	0.97	0.93	16	21	0.92	0.81	0.91
98	0.98	0.92	16	22	0.85	0.48	0.58
99	0.99	0.90	16	22	0.84	0.25	0.94
100	0.99	0.92	14	21	0.86	0.27	0.98
101	0.96	0.92	11	18	0.88	0.61	0.86
102	0.91	0.92	10	17	0.91	0.92	0.99
103	0.84	0.96	8	14	0.92	0.33	0.34
104	0.90	0.97	9	14	0.92	0.33	0.67
105	0.85	0.97	8	12	0.93	0.70	0.76
106	0.75	0.98	8	13	0.95	0.93	0.92
107	0.83	0.98	9	14	0.94	0.99	0.62
108	0.79	0.95	9	15	0.95	0.65	0.61
109	0.76	0.92	11	16	0.96	0.64	0.52
110	0.94	0.93	13	18	0.94	0.62	0.98
111	0.97	0.90	15	20	0.91	0.86	0.75
112	0.97	0.91	16	21	0.88	0.54	0.46
113	0.98	0.93	18	22	0.86	0.65	0.99
114	0.99	0.92	21	26	0.81	0.24	0.66
115	0.99	0.92	21	25	0.68	0.43	0.92
116	0.95	0.93	18	22	0.52	0.45	0.70
117	0.81	0.91	16	20	0.87	0.83	0.43
118	0.76	0.87	17	20	0.44	0.97	0.43
119	0.54	0.79	14	18	0.58	0.94	0.99
120	0.63	0.84	14	18	0.44	0.42	0.99
121	0.55	0.63	13	17	0.66	0.91	0.91
122	0.15	0.92	10	14	0.71	0.92	0.86
123	0.11	0.96	8	13	0.87	0.42	0.88
124	0.27	0.96	9	14	0.83	0.26	0.36
125	0.21	0.97	7	12	0.92	0.88	0.68
126	0.13	0.99	5	9	0.99	0.96	0.72
127	0.22	0.99	5	9	0.99	0.84	0.85
128	0.31	0.98	6	11	0.96	0.78	0.86
129	0.37	0.99	4	10	0.97	0.15	0.54
130	0.38	0.99	0	7	0.99	0.99	0.94
131	0.39	0.99	4	10	0.98	0.99	0.67
132	0.46	0.98	6	9	0.99	0.99	0.94
133	0.46	0.96	8	12	0.96	0.41	0.98

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
134	0.14	0.95	9	12	0.98	0.98	0.45
135	0.06	0.95	7	10	0.99	0.64	0.87
136	0.21	0.95	6	12	0.99	0.97	0.53
137	0.13	0.93	10	15	0.93	0.95	0.95
138	0.05	0.89	10	14	0.94	0.98	0.99
139	0.08	0.91	10	14	0.97	0.82	0.84
140	0.29	0.73	12	17	0.95	0.33	0.56
141	0.34	0.77	14	19	0.81	0.1	0.5
142	0.35	0.52	13	19	0.83	0.61	0.77
143	0.44	0.7	15	20	0.82	0.99	0.59
144	0.33	0.88	12	17	0.91	0.99	0.67
145	0.36	0.87	11	16	0.97	0.91	0.53
146	0.51	0.89	14	19	0.88	0.48	0.9
147	0.48	0.88	14	19	0.81	0.8	0.86
148	0.3	0.68	11	16	0.94	0.78	0.95
149	0.34	0.71	13	17	0.86	0.71	0.87
150	0.5	0.84	16	21	0.44	0.34	0.89
151	0.48	0.75	14	20	0.68	0.89	0.99
152	0.37	0.52	13	18	0.76	0.29	0.86
153	0.44	0.51	17	22	0.44	0.68	0.84
154	0.57	0.6	18	24	0.33	0.74	0.47
155	0.58	0.51	17	23	0.68	0.91	0.91
156	0.56	0.33	19	23	0.44	0.99	0.6
157	0.4	0.29	16	21	0.78	0.96	0.34
158	0.41	0.31	18	22	0.58	0.54	0.78
159	0.33	0.64	17	20	0.8	0.5	0.76
160	0.22	0.6	13	16	0.88	0.56	0.71
161	0.3	0.83	13	16	0.88	0.62	0.5
162	0.26	0.86	11	14	0.94	0.68	0.24
163	0.45	0.97	10	16	0.96	0.85	0.45
164	0.43	0.98	7	14	0.97	0.93	0.87
165	0.38	0.99	4	12	0.95	0.38	0.97
166	0.41	0.99	0	10	0.94	0.99	0.99
167	0.53	0.99	3	11	0.93	0.72	0.99
168	0.54	0.99	5	13	0.93	0.9	0.99
169	0.33	0.99	6	12	0.96	0.99	0.99

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
170	0.43	0.95	8	11	0.93	0.96	0.99
171	0.51	0.94	10	14	0.94	0.82	0.98
172	0.36	0.91	10	15	0.96	0.49	0.99
173	0.62	0.95	14	18	0.97	0.2	0.95
174	0.66	0.95	16	21	0.96	0.7	0.32
175	0.68	0.94	18	22	0.97	0.56	0.25
176	0.6	0.89	17	20	0.97	0.78	0.56
177	0.51	0.68	16	19	0.96	0.22	0.93
178	0.44	0.84	13	17	0.96	0.73	0.34
179	0.25	0.85	11	13	0.98	0.97	0.93
180	0.16	0.91	11	14	0.93	0.93	0.25
181	0.19	0.9	12	15	0.92	0.99	0.71
182	0.1	0.7	15	18	0.66	0.98	0.65
183	0.04	0.5	14	17	0.84	0.97	0.75
184	0.16	0.43	17	20	0.44	0.33	0.81
185	0.2	0.56	16	19	0.71	0.92	0.39
186	0.12	0.7	12	15	0.93	0.43	0.31
187	0.08	0.73	13	17	0.84	0.91	0.99
188	0.14	0.59	16	20	0.66	0.47	0.88
189	0.18	0.63	13	18	0.83	0.4	0.9
190	0.19	0.78	12	16	0.95	0.94	0.86
191	0.22	0.71	15	20	0.58	0.25	0.98
192	0.24	0.58	16	21	0.52	0.7	0.56
193	0.23	0.69	13	19	0.87	0.63	0.47
194	0.35	0.52	14	19	0.77	0.23	0.67
195	0.52	0.48	18	23	0.16	0.16	0.94
196	0.56	0.4	19	24	0.16	0.88	0.61
197	0.71	0.42	21	26	0.33	0.39	0.49
198	0.57	0.58	18	24	0.44	0.33	0.66
199	0.61	0.45	17	23	0.74	0.99	0.67
200	0.76	0.33	19	25	0.33	0.14	0.6
201	0.79	0.43	20	25	0.33	0.09	0.32
202	0.75	0.47	19	25	0.16	0.13	0.76
203	0.57	0.34	16	21	0.74	0.6	0.42
204	0.53	0.52	15	20	0.81	0.6	0.55
205	0.59	0.79	16	21	0.58	0.45	0.96

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
206	0.48	0.71	14	19	0.71	0.37	0.27
207	0.26	0.85	11	16	0.93	0.95	0.62
208	0.35	0.9	11	17	0.93	0.43	0.92
209	0.42	0.92	12	18	0.8	0.32	0.76
210	0.28	0.92	10	14	0.93	0.77	0.34
211	0.23	0.94	6	13	0.97	0.94	0.5
212	0.41	0.92	9	16	0.93	0.2	0.7
213	0.43	0.94	10	16	0.94	0.41	0.47
214	0.27	0.94	7	13	0.96	0.97	0.73
215	0.4	0.95	9	14	0.96	0.45	0.73
216	0.39	0.96	8	12	0.96	0.98	0.7
217	0.31	0.96	7	12	0.95	0.97	0.88
218	0.64	0.87	11	16	0.94	0.41	0.95
219	0.68	0.64	12	17	0.93	0.34	0.9
220	0.49	0.67	10	16	0.95	0.96	0.99
221	0.42	0.55	10	17	0.95	0.88	0.34
222	0.61	0.56	11	18	0.93	0.99	0.74
223	0.71	0.51	14	20	0.88	0.88	0.64
224	0.69	0.59	16	22	0.83	0.68	0.75
225	0.74	0.45	15	22	0.84	0.76	0.99
226	0.92	0.44	19	26	0.71	0.99	0.6
227	0.98	0.67	20	27	0.66	0.62	0.89
228	0.95	0.68	20	28	0.72	0.76	0.51
229	0.79	0.76	17	25	0.86	0.95	0.47
230	0.7	0.89	14	22	0.89	0.31	0.52
231	0.47	0.9	10	18	0.93	0.93	0.91
232	0.24	0.98	8	15	0.95	0.97	0.47
233	0.17	0.99	4	12	0.98	0.97	0.53
234	0.07	0.99	0	9	0.99	0.99	0.32
235	0.06	0.99	4	9	0.99	0.99	0.81
236	0.12	0.99	4	8	0.99	0.99	0.66
237	0.32	0.99	0	9	0.99	0.87	0.78
238	0.45	0.99	4	11	0.99	0.91	0.97
240	0.6	0.98	7	14	0.93	0.85	0.85
241	0.65	0.96	10	17	0.94	0.56	0.22
242	0.62	0.95	8	16	0.97	0.89	0.25
243	0.29	0.97	6	13	0.98	0.63	0.49

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
244	0.18	0.99	4	12	0.99	0.81	0.94
245	0.02	0.97	7	11	0.99	0.99	0.93
246	0.01	0.98	6	11	0.99	0.99	0.74
247	0.03	0.98	6	12	0.99	0.74	0.4
248	0.04	0.98	7	14	0.97	0.98	0.42
249	0.19	0.95	10	18	0.94	0.82	0.23
250	0.29	0.9	12	20	0.9	0.54	0.35
251	0.59	0.9	16	24	0.84	0.6	0.6
252	0.83	0.83	19	27	0.74	0.4	0.32
254	0.94	0.87	21	29	0.72	0.53	0.71
255	0.88	0.89	18	26	0.66	0.27	0.64
256	0.8	0.89	18	25	0.44	0.33	0.93
257	0.62	0.9	15	23	0.74	0.57	0.35
258	0.64	0.87	17	24	0.72	0.63	0.64
259	0.37	0.89	15	22	0.8	0.68	0.98
260	0.11	0.89	12	19	0.9	0.58	0.45
261	0.02	0.92	11	18	0.9	0.9	0.48
262	0.01	0.94	10	15	0.96	0.64	0.99
263	0.01	0.9	11	15	0.94	0.8	0.99
264	0.02	0.93	10	14	0.96	0.91	0.58
265	0.1	0.96	10	16	0.97	0.36	0.82
266	0.27	0.94	10	16	0.97	0.67	0.81
267	0.6	0.92	12	19	0.98	0.08	0.41
268	0.72	0.87	12	20	0.97	0.3	0.72
269	0.87	0.86	14	21	0.96	0.7	0.97
270	0.88	0.81	13	21	0.96	0.17	0.89
271	0.92	0.87	12	20	0.95	0.47	0.93
272	0.86	0.87	9	17	0.98	0.72	0.52
273	0.89	0.87	12	19	0.94	0.2	0.93
274	0.81	0.81	13	20	0.9	0.26	0.85
275	0.68	0.78	10	17	0.96	0.45	0.46
276	0.67	0.78	10	17	0.97	0.91	0.99
277	0.7	0.82	14	20	0.81	0.66	0.83
278	0.59	0.86	14	20	0.84	0.68	0.93
279	0.4	0.87	13	18	0.95	0.97	0.87
280	0.5	0.83	14	19	0.89	0.98	0.95

Table C.3: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
281	0.55	0.84	18	23	0.58	0.43	0.98
282	0.49	0.84	16	22	0.72	0.68	0.52
283	0.46	0.81	15	21	0.85	0.76	0.39
284	0.63	0.77	18	24	0.79	0.48	0.32
285	0.67	0.77	20	26	0.58	0.81	0.85
286	0.65	0.74	18	25	0.74	0.85	0.75
287	0.73	0.67	18	26	0.71	0.08	0.84
288	0.81	0.65	22	29	0.52	0.54	0.95
289	0.85	0.72	23	30	0.52	0.1	0.5
290	0.84	0.66	21	29	0.72	0.48	0.81

Table C.4: The predicted binding affinities of domain B and comparison with experimental ITC data for wild type, mutDB, and P51G-m4 benchmarking.

Protein	Predicted Binding Score (X-score energy unit)	ITC dimannose Kd (μM)	ITC dimannose ΔH (kcal/mol)	ITC dimannose TΔS (kcal/mol) (T=298K)	ITC dimannose ΔG (kcal/mol)
Wild Type	-7.08	16 ± 1	-12.5 ± 0.3	-6.00 ± 0.1	-6.50 ± 0.3
mutDB	-5.97	No-binding	No-binding	No-binding	No-binding
P51G-m4	-6.62	117 ± 3	-12.3 ± 0.3	-7.00 ± 0.3	-5.30 ± 0.3

Table C.5: The complete cyanovirin-N (CV-N) dynamic flexibility index (DFI), dynamic coupling index (DCI), RaptorX, Evcoupling, and MISTIC metric data used in this study.

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
1	0.98	0.72	20	27	0.69	0.30	0.38
2	0.91	0.73	18	25	0.73	0.30	0.60
3	0.86	0.58	20	27	0.72	0.30	0.61
4	0.43	0.67	21	28	0.84	0.30	0.64
5	0.80	0.36	22	28	0.84	0.30	0.28
6	0.79	0.42	25	31	0.86	0.89	0.45
7	0.66	0.59	27	33	0.85	0.88	0.29
8	0.37	0.22	25	31	0.88	0.91	0.36
9	0.52	0.14	26	31	0.81	0.90	0.64
10	0.48	0.36	26	30	0.73	0.77	0.40
11	0.41	0.35	22	26	0.75	0.75	0.58
12	0.72	0.69	20	24	0.63	0.88	0.70
13	0.67	0.68	17	21	0.69	0.85	0.89
14	0.73	0.66	16	20	0.61	0.30	0.58
15	0.58	0.76	12	17	0.69	0.85	0.65
16	0.31	0.77	12	16	0.73	0.82	0.80
17	0.44	0.62	15	19	0.65	0.94	0.90
18	0.54	0.51	16	21	0.65	0.98	0.84
19	0.33	0.80	20	25	0.56	0.79	0.81
20	0.06	0.54	22	27	0.48	0.83	0.40
21	0.10	0.44	26	31	0.38	0.85	0.82
22	0.22	0.48	26	32	0.38	0.81	0.57
23	0.76	0.62	29	35	0.38	0.92	0.85
24	0.90	0.55	30	37	0.39	0.78	0.75
25	0.97	0.52	32	39	0.39	0.94	0.84
26	1.00	0.40	35	42	0.35	0.77	0.75
27	0.99	0.43	35	42	0.35	0.93	0.90
28	0.94	0.50	34	40	0.32	0.92	0.91
29	0.87	0.60	31	37	0.33	0.95	0.84
30	0.70	0.55	28	34	0.28	0.75	0.82
31	0.18	0.84	26	32	0.35	0.84	0.78
32	0.02	0.91	23	29	0.44	0.92	0.48
33	0.11	0.90	19	25	0.53	0.79	0.76
34	0.17	0.71	16	22	0.65	0.93	0.71
35	0.26	0.82	13	19	0.76	0.61	0.97
36	0.27	0.76	10	16	0.90	0.98	0.98

Table C.5: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
37	0.23	0.78	8	14	0.90	0.91	0.72
38	0.59	0.79	8	16	0.87	0.69	0.67
39	0.56	0.89	7	15	0.90	0.68	0.91
40	0.09	0.99	4	12	0.94	0.93	0.95
41	0.16	1.00	0	9	0.92	0.60	0.61
42	0.14	1.00	0	7	0.96	1.00	0.98
43	0.57	0.97	4	8	0.92	0.80	0.88
44	0.69	0.98	6	8	0.98	0.93	0.97
45	0.47	0.93	6	8	0.96	0.99	0.99
46	0.36	0.95	5	10	0.89	0.83	0.83
47	0.19	0.99	4	10	0.99	0.94	0.92
48	0.39	0.95	5	12	0.99	0.94	0.95
49	0.45	0.91	6	14	0.98	0.85	0.91
50	0.83	0.95	5	13	0.98	0.98	0.97
51	0.81	0.92	5	13	0.93	0.99	0.91
52	0.71	0.96	5	11	0.88	0.97	0.93
53	0.40	0.92	6	11	0.91	0.81	0.96
54	0.29	0.97	5	9	0.94	0.91	0.98
55	0.20	0.96	5	11	0.90	0.80	0.98
56	0.38	0.98	4	10	0.89	0.99	0.97
57	0.42	1.00	0	7	0.96	1.00	0.98
58	0.08	0.99	4	9	0.95	0.98	0.93
59	0.34	0.95	7	12	0.84	0.74	0.75
60	0.32	0.93	11	15	0.76	0.88	0.87
61	0.15	0.83	11	15	0.81	0.76	0.86
62	0.55	0.86	15	18	0.53	0.94	0.94
63	0.63	0.70	17	20	0.52	0.84	0.91
64	0.75	0.68	18	21	0.36	0.89	0.93
65	0.88	0.49	21	24	0.26	0.70	0.69
66	0.95	0.27	21	26	0.27	0.91	0.94
67	0.82	0.29	18	23	0.34	0.79	0.86
68	0.65	0.63	16	20	0.41	0.94	0.95
69	0.49	0.79	13	17	0.70	0.96	0.82
70	0.13	0.91	13	16	0.56	0.93	0.92
71	0.01	0.93	11	13	0.94	0.92	0.98
72	0.04	0.88	9	11	0.79	0.95	0.95

Table C.5: *Continued*

Residue	DFI Score	DCI Score	Distance from Binding Sites (Å)		RAPTORX Score	MISTIC Score	EVCOUPLING Score
			Minimum	Average			
73	0.03	0.94	6	8	0.99	0.83	0.79
74	0.50	0.97	5	8	0.99	0.97	0.98
75	0.77	0.97	4	8	0.97	0.98	0.99
76	0.92	1.00	0	8	0.93	0.62	0.95
77	0.93	0.99	4	10	0.87	0.90	0.99
78	0.89	1.00	0	9	0.90	0.92	0.95
79	0.84	0.99	4	10	0.77	0.93	0.98
80	0.62	0.90	7	11	0.86	1.00	0.97
81	0.46	0.55	10	12	0.93	0.78	0.95
82	0.25	0.81	12	13	0.91	0.79	0.95
83	0.05	0.86	9	12	0.97	0.30	0.91
84	0.12	0.89	11	14	0.87	0.86	0.83
85	0.25	0.89	11	15	0.96	0.97	0.70
86	0.53	0.69	13	19	0.76	0.71	0.64
87	0.64	0.53	13	19	0.83	0.96	0.56
88	0.74	0.55	16	23	0.59	0.72	0.73
89	0.85	0.63	15	21	0.63	0.94	0.87
90	0.61	0.71	14	22	0.72	0.74	0.73
91	0.30	0.76	15	23	0.67	0.97	0.78
92	0.24	0.84	19	26	0.62	0.97	0.83
93	0.28	0.87	22	29	0.67	0.99	0.13
94	0.68	0.87	24	32	0.64	0.30	0.78
95	0.78	0.78	27	35	0.53	0.81	0.58
96	0.60	0.73	26	33	0.58	0.98	0.18
97	0.21	0.83	23	30	0.60	0.96	0.81
98	0.07	0.84	20	27	0.75	0.97	0.33
99	0.35	0.85	18	26	0.53	0.30	0.37
100	0.51	0.75	15	23	0.86	0.30	0.55
101	0.75	0.74	16	24	0.76	0.30	0.95

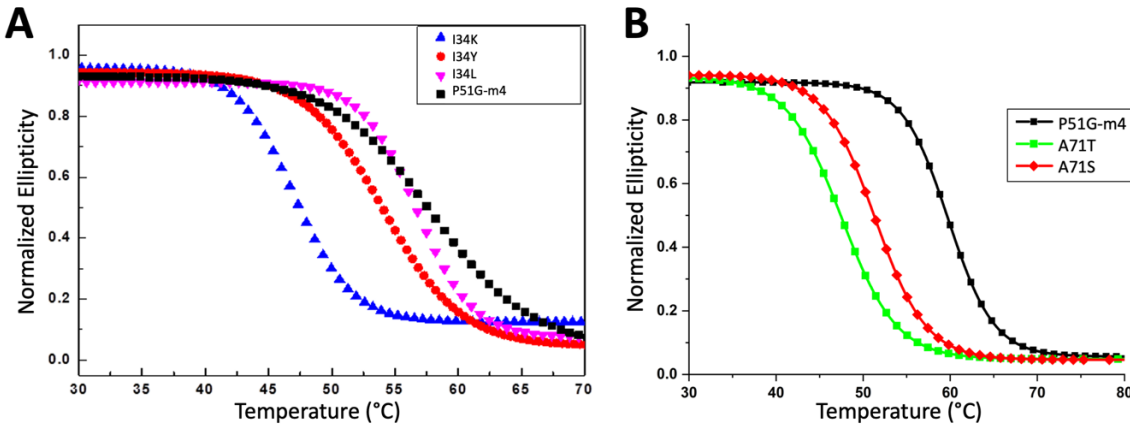


Figure C.1: Fits for thermal melts of the CV-N mutants A) I34 variants, and B) A71 variants.

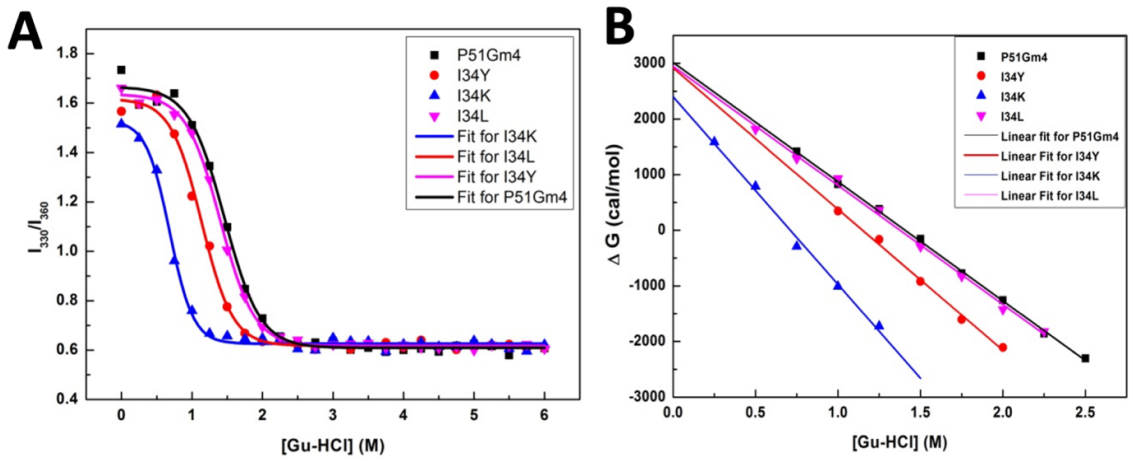


Figure C.2: Fits for the chemical denaturation experiments of the variants. (**A**) Chemical denaturation curve showing I_{330}/ I_{360} ratio as a function of Gu-HCl concentration. (B) ΔG_{H2O} versus Gu-HCl concentration plot for cyanovirin-N (CV-N) mutants.

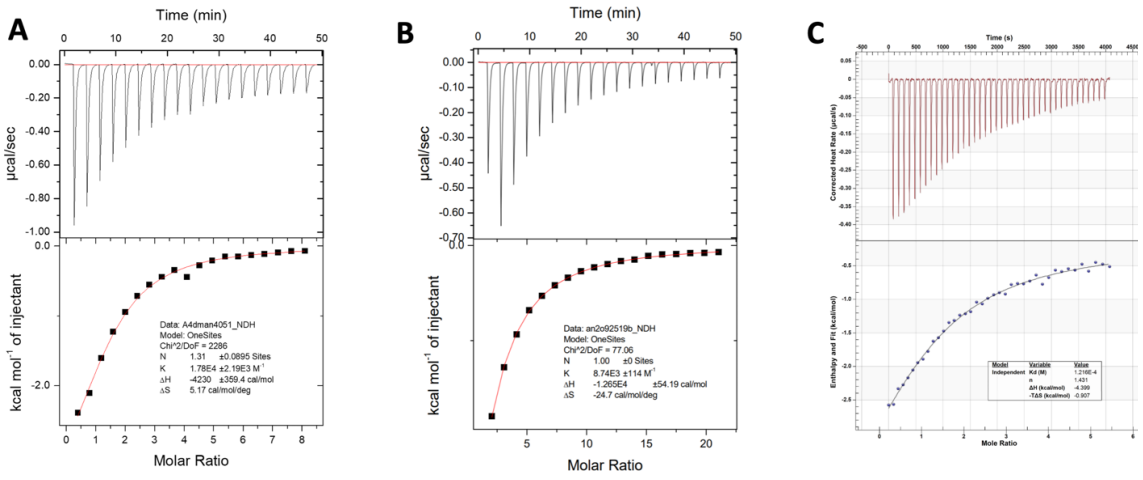


Figure C.3: Binding isotherms of CV-N mutants upon titration with dimannose: A) I34Y and B) P51G-m4 C) A71T.

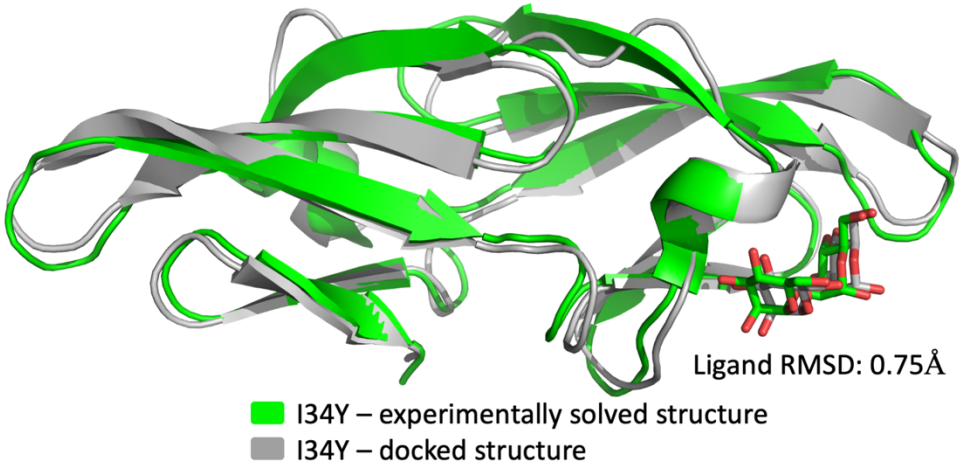


Figure C.4: Comparison of experimentally solved I34Y structure with docked pose from Adaptive BP dock algorithm. The RMSD of the ligand is calculated as 0.75 Å.

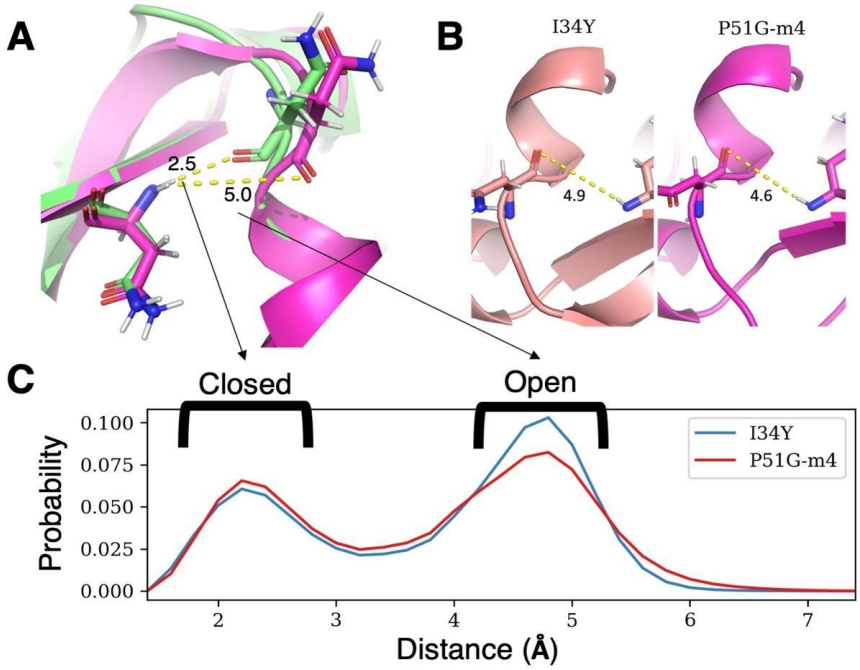


Figure C.5: The difference in accessibility of the binding pocket for P51G-m4 andI34Y. A) Structural difference of open vs closed conformation based on the hydrogen bond distance between residue N42 and N53 B) Hydrogen bond distance between residue N42 and N53 from crystal structures of P51G-m4, and P51G-m4-I34Y C) Frequencies of hydrogen bond distance between residue N42 and N53 from GROMACS production runs showing I34Y variant sampling more open conformation compared to P51G-m4.

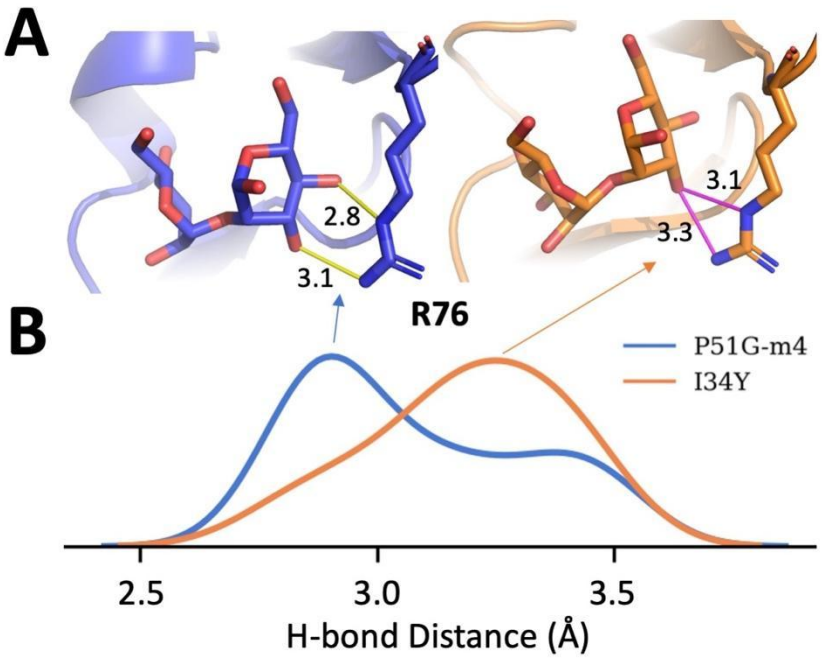


Figure C.6: We sampled 2000 different conformations from MD simulations for P51G-m4 CV-N and I34Y mutant and performed dimannose docking to obtained docked poses and then analyzed hydrogen bond patterns A) Hydrogen bonds (representing the peak of the distribution on panel B) and their distances are shown between dimannose and residue R76 for P51G-m4 (blue) and I34Y (orange) B) H-bond distance distribution between dimannose and residue R76.

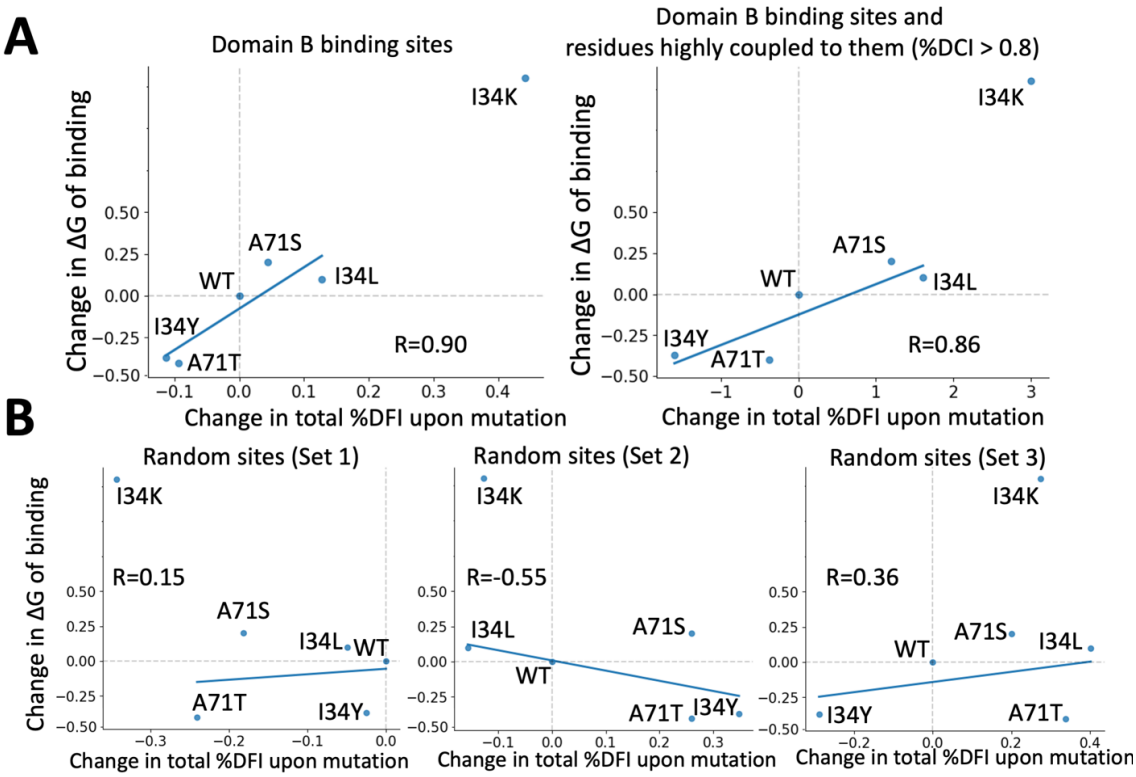


Figure C.7: Correlation between change in DFI profiles aind change in ΔG of binding. (A) Change in ΔG of binding ($\Delta G_{mut} - \Delta G_{wt}$) is compared with change in total dynamic flexibility index (DFI) scores (∑DFImut− DFIwt) for selected residues. The correlation with experimental binding scores is compared with the total sum of DFI values considering only domain B binding site residues first, and also summing over the domain B binding sites as well as the residues highly coupled (coupling greater than 0.8) to them. The observed high correlations indicate that these residues play an important role in the binding modulation upon mutations. (B) In addition, we randomly selected residues in domain B to calculate total DFI change over these positions upon mutations. Three different randomly selected residue sets all show poor correlation with change in experimental binding free energy.

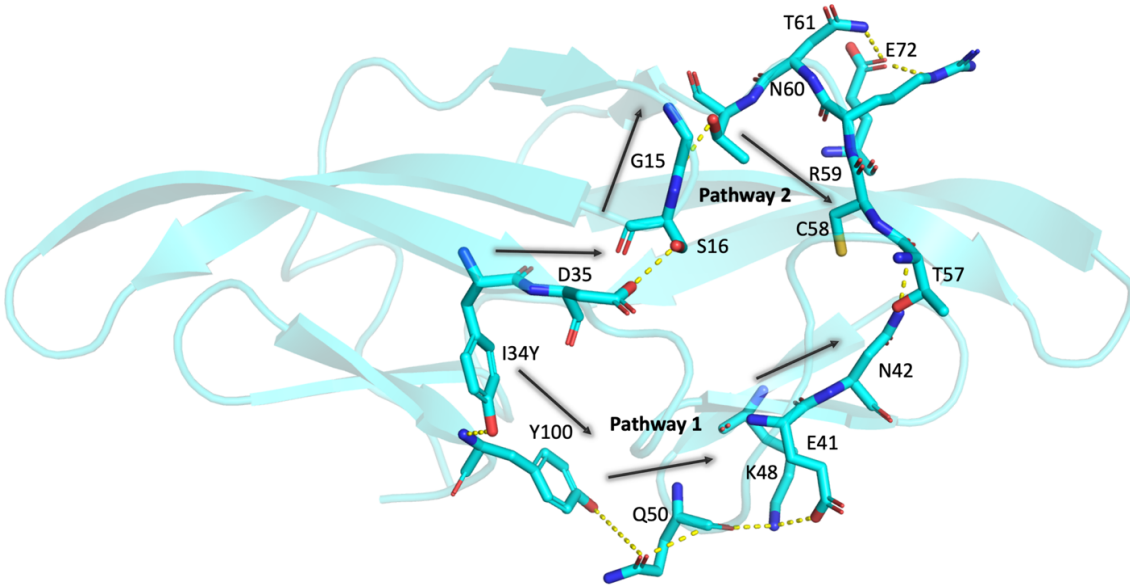


Figure C.8: Network of hydrogen bond interactions connecting residue location 34 to T57 is investigated in I34Y variant and P51G-m4 CV-N. Two Hydrogen bond pathways are found connecting residue 34 to 57. Pathway 1 is unique to I34Y. Pathway 2 is also observed in P51G-m4 CV-N but sampled much more frequently in I34Y variant.

APPENDIX D

STATEMENT OF CO-AUTHOR PERMISSIONS

For the use of published articles in chapters 3,4,5,6, and 7, the co-authors have granted their consent.