



Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations


Graduate School

2022

Protein Structure and Interaction: The Role of Aromatic Residues in Protein Structure and Interactions Between Pyridoxine 5'-Phosphate Oxidase/Dopa Decarboxylase

Mohammed H. AL Mughram
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>

 Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Bioinformatics Commons](#), [Enzymes and Coenzymes Commons](#), [Medicinal and Pharmaceutical Chemistry Commons](#), and the [Structural Biology Commons](#)

© Mohammed H. AL Mughram

Downloaded from

<https://scholarscompass.vcu.edu/etd/7136>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

Mohammed H. AL Mughram

2022

All Rights Reserved

**Protein Structure and Interaction: The Role of Aromatic Residues in
Protein Structure and Interactions Between Pyridoxine 5'-Phosphate
Oxidase/Dopa Decarboxylase**

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor
of Philosophy at Virginia Commonwealth University.

by

Mohammed H. AL Mughram

Master of Medicinal and Biological Chemistry, University of Toledo, USA, 2018
Bachelor of Pharmaceutical Science, King Khalid University, KSA, 2013

Directors: Martin K. Safo, Ph.D.
Professor,
Department of Medicinal Chemistry
Glen E. Kellogg, Ph.D.
Professor,
Department of Medicinal Chemistry

Virginia Commonwealth University
Richmond, Virginia
December 2022

Acknowledgment

This work is dedicated to everyone who has guided and supported me as I pursued my Ph.D. study. I would like to express my gratitude and sincere appreciation to my PhD advisors, Dr. Martin K. Safo and Dr. Glen E. Kellogg. I can't thank them enough for their never-ending guidance and support over the past years. This dissertation, along with other projects, were accomplished in large part due to the help of their invaluable experiences.

I am immensely grateful to Dr. Masahiro Sakagami, Dr. Youzhong Guo, and Dr. Ashton Cropp for serving on my graduate student committee. Thanks for taking the time out of their busy schedules to review and evaluate my research. I'm grateful for their precious insights and suggestions throughout my research.

This thesis would not have been possible without the guidance and help of several people who, in various ways, contributed to and provided invaluable assistance in the preparation and completion of this work. I would like to thank current and former members of Safo's group for sharing their research skills and experiences with me, in particular Dr. Mohini Ghatge, Dr. Faik Musayev, Dr. Anfal Aljahdali, Dr. Rana Al-Hashmi, Dr. Dina Alramdhani, Mohammed Al-Awadh, and Mona Alkhairi. I'd also want to thank Dr. Claudio Catalano and Dr. Noah Herrington, former members of Kellogg's group, for sharing their research talents and experiences with me. There are no adequate words to describe the impact they have had on my life and my career.

Sincere gratitude also goes out to King Khalid University for the scholarship that

enabled me to pursue my master's and doctoral degrees. I am grateful for their trust and looking forward to joining them. I would like to thank the Institute of Structural Biology and Drug Discovery and the Department of Medicinal Chemistry of the School of Pharmacy at Virginia Commonwealth University, who made it possible for me to pursue graduate studies.

Making this journey without my beloved wife, Samah, is unimaginable. Thank you for your unwavering support and love. It's been a wonderful journey with you, filled with unforgettable memories. Thanks very much, with all my heart, to my beloved daughter, Taraaf. She has been my source of motivation throughout my Ph.D. I want to sincerely thank all my friends in Richmond, especially Ali Algubaishi, Mohammed Alasmari, and Abdullah Alahmari, for making me feel like I had a family there. Everyone, thanks!

Above all, I owe a huge debt of gratitude to my father, Hassan, and mother, Saliha, who lived in every sense of this journey with me. I would have never gotten this far without their unconditional love and endless support. I also would like to thank my sisters, Kholoud and Ohoud, and brothers, Khalid, Majid, and Saud, who have been by my side throughout the years.

TABLE OF CONTENTS

Acknowledgments	ii
List of Tables	vi
List of Figures	vii
Abstract.....	x

Chapter 1

1. General Introduction	1
1.1 Experimental protein structure determination	3
1.1.1 Nuclear magnetic resonance (NMR).....	3
1.1.2 X-ray crystallography.....	4
1.1.3 Cryo-electron microscopy (Cryo-EM).....	6
1.2 Computational protein structure determination	7
1.2.1 Homology modeling	9
1.2.2 Protein threading	9
1.2.3 <i>Ab initio</i> approach	10
1.3 Protein-protein docking	11
1.4 Molecular dynamics simulations.....	14
1.5 Bioinformatics structural analysis of experimental protein structures.....	14
1.5.1 Aromatic amino acids.....	15
1.5.2 3D hydrophobic interaction maps for protein structure description .	18
1.6 Vitamin B ₆ (B6 vitamers and metabolic pathway)	20
1.6.1 PLP-dependent enzymes	21
1.6.2 PLP deficiency and neurological pathologies	23
1.6.3 DDC and AADC deficiency.....	25
1.7 Research plan	28

Chapter 2

2. 3D hydrophobic interaction maps as a new motif for understanding the role of aromatic amino acid residues in protein structure.	43
2.1.1 Introduction.....	43
2.1.2 Hypotheses and research plan	46
2.2 Methods.....	48
2.3 Results and Discussion.....	57
2.4 Summary and conclusion.....	110

Chapter 3

3. Computational, Biophysical, and Kinetic Approaches to Elucidate Interactions Between pyridoxine 5'-phosphate oxidase (PNPO) and Dopa Decarboxylase (DDC)	117
3.1 Introduction	117
3.1.1 PNPO allosteric characteristics and PLP binding	119
3.1.2 Hypotheses and research plan	120
3.2 Molecular docking and dynamics simulations to characterize the interaction of PNPO with apo- and holo-DDC	122
3.2.1 Introduction.....	122
3.2.2 Methods.....	125
3.2.3 Results and discussion	128
3.2.4 Conclusion.....	142
3.3 Biophysical and PLP transfer studies to probe the interaction between PNPO and DDC	143
3.3.1 Introduction.....	143
3.3.2 Methods.....	144
3.3.3 Results and discussion	153
3.3.4 Conclusion.....	164
3.4 Site directed mutagenesis studies to probe the interaction between PNPO and3DDC	166
3.4.1 Introduction.....	166
3.4.2 Methods.....	167
3.4.3 Results and discussion	168
3.4.4 Conclusion.....	183

Chapter 4

4. Overall conclusion.....	189
----------------------------	-----

LIST OF TABLES

1.1 Example of Protein-Protein docking tools (docking algorithm, description, and CAPRI validation).....	13
1.2 Classification PLP-dependent enzymes with representative examples of each class	22
1.3 PNPO gene mutations affecting enzyme structure and function.	25
2.1 Construction of Cylinders for PHE, TYR and TRP.....	56
2.2 Residue populations of chess squares/parses for PHE, TYR and TRP	61
2.3 Number of PHE, TYR and TRP clusters by chess square and χ_1 parse	62
2.4 Summary of results for clustering of maps in chess square/parse <i>b1.60</i> for PHE, TYR and TRP sidechains	64
2.5 Analysis of cylinder character for π - π and π -cation interactions	110
3.1 Theoretical R_{\max} values for the SPR experiment.....	150
3.2 Optimization of ITC parameters for PNPO and DDC titration.	151
3.3 ITC measured thermodynamic parameters between PNPO mutants and DDC in both holo (left) and apo (right) forms	181

LIST OF FIGURES

1.1	General scheme of model construction	8
1.2	Hydropathic interaction map representing the Gaussian-weighted average sidechain environment for a PHE sidechain cluster in our dataset.	18
1.3	Chemical structure of the B6 vitamers and derivatives	21
1.4	The PLP salvage pathway and forming the active holo-B6 enzyme.....	21
1.5	PLP-dependent enzymes superfamilies.	23
1.6	The metabolic pathway involves DDC for the synthesis of dopamine and serotonin neurotransmitters.	27
2.1	Ramachandran (ϕ versus ψ) plots with superimposed chess square schema in blue	48
2.2	Ramachandran chessboard displaying the chess square/parse population for PHE residue.	59
2.3	Ramachandran chessboard displaying the chess square/parse population for TYR residue.....	60
2.4	Ramachandran chessboard displaying the chess square/parse population for... TRP residue.....	61
2.5	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the b1 chess square ($\chi_1 = 60^\circ$ parse).....	67
2.6	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the b1 chess square ($\chi_1 = 180^\circ$ parse)....	68
2.7	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the b1 chess square ($\chi_1 = 300^\circ$ parse)....	68
2.8	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the c5 chess square ($\chi_1 = 60^\circ$ parse).....	69
2.9	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the c5 chess square ($\chi_1 = 180^\circ$ parse)....	69
2.10	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the c5 chess square ($\chi_1 = 300^\circ$ parse). ...	70
2.11	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the d5 chess square ($\chi_1 = 60^\circ$ parse).....	71
2.12	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the d5 chess square ($\chi_1 = 180^\circ$ parse)...	71
2.13	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the d5 chess square ($\chi_1 = 300^\circ$ parse) ...	72
2.14	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the f6 chess square ($\chi_1 = 300^\circ$ parse). ...	73
2.15	Hydropathic interaction map comparisons and putative π -cation features for phenylalanine.	76
2.16	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the b1 chess square ($\chi_1 = 60^\circ$ parse).	79
2.17	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the b1 chess square ($\chi_1 = 180^\circ$ parse). ..	79

2.18	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the b1 chess square ($\chi_1 = 300^\circ$ parse)....	80
2.19	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the c5 chess square ($\chi_1 = 60^\circ$ parse).....	81
2.20	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the c5 chess square ($\chi_1 = 180^\circ$ parse)....	81
2.21	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the c5 chess square ($\chi_1 = 300^\circ$ parse)....	82
2.22	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the d5 chess square ($\chi_1 = 60^\circ$ parse) ...	83
2.23	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the d5 chess square ($\chi_1 = 180^\circ$ parse) ...	83
2.24	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the d5 chess square ($\chi_1 = 300^\circ$ parse). ...	84
2.25	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the f6 chess square ($\chi_1 = 300^\circ$ parse). ...	84
2.26	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the b1 chess square ($\chi_1 = 60^\circ$ parse). ...	87
2.27	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the b1 chess square ($\chi_1 = 180^\circ$ parse). ...	87
2.28	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the b1 chess square ($\chi_1 = 300^\circ$ parse). ...	88
2.29	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the c5 chess square ($\chi_1 = 60^\circ$ parse).....	89
2.30	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the c5 chess square ($\chi_1 = 180^\circ$ parse)....	89
2.31	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the c5 chess square ($\chi_1 = 300^\circ$ parse)....	90
2.32	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the d5 chess square ($\chi_1 = 60^\circ$ parse) ...	91
2.33	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the d5 chess square ($\chi_1 = 180^\circ$ parse). ...	91
2.34	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the d5 chess square ($\chi_1 = 300^\circ$ parse)	92
2.35	Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the f6 chess square ($\chi_1 = 300^\circ$ parse)	93
2.36	Definition of map cylinders for PHE, TYR and TRP	97
2.37	Hydropathic character for clusters in PHE cylinders	99
2.38	Hydropathic interaction characters for clusters in PHE cylinder by secondary structure.....	102
2.39	Hydropathic character for clusters in TYR cylinders	104
2.40	Interaction types for modified volumes above and below phenol ring in TYR by secondary structure.....	105
2.41	Hydropathic character for clusters in TRP cylinders	107
2.42	Interaction types for modified volumes above and below phenol ring in TRP by secondary structure.....	108

3.1	Dimeric structures of apoDDC (open conformation) and holoDD (closed conformation).....	123
3.2	Dimeric structure of PNPO.....	124
3.3	Cluspro 2.0 docking results for PNPO and apoDDC complex.....	131
3.4	Cluspro 2.0 docking findings for PNPO and apoDDC including missing residues modeled (3RCH-AF)	132
3.5	Electrostatic map comparison for the top cluster of PNPO•apoDDC complex.....	133
3.6	Comparison of docking results for PNPO and holoDDC with the crystal structure (PDB: 1JS3) and modeled residues (PDB: 1JS3-model)	134
3.7	Superposition of crystal structures (PNPO and holoDDC) and the PNPO•holoDDC complex model	135
3.8	MD simulation of PNPO•holoDDC and PNPO•apoDDC complexes	136
3.9	Interaction profile analysis of hydrogen bonds and salt bridges at the domain interface for the PNPO•apoDDC complex model.....	138
3.10	Interaction analysis of hydrogen bonds and salt bridges at the domain interface for the PNPO•holoDDC complex model	138
3.11	Hierarchical cluster analysis of MD trajectories for the PNPO-apoDDC complex.....	140
3.12	Key interactions between PNPO and apoDDC as predicted by our model.....	141
3.13	Fluorescence polarization saturation binding curves	154
3.14	SPR analysis results.....	156
3.15	Enzymatic activity of recombinant human DDC in both apo and holo forms.....	158
3.16	ITC thermodynamic analysis of PNPO•apoDDC and PNPO•holoDDC binding.....	159
3.17	Schematic representation of the PLP transfer from PNPO•PLP to apoDDC	162
3.18	Rate of transfer of PLP from PNPO•PLP to apoDDC.....	162
3.19	Differential scanning fluorimetry measurements of PNPO wild-type and mutants	169
3.20	Isothermal titration calorimetry (ITC) characterization of PNPO mutants with holoDDC.....	172
3.21	Isothermal titration calorimetry (ITC) characterization of PNPO mutants with apoDDC.....	176
3.22	ITC thermodynamic binding data of PNPO mutants bound to apo- and holo-DDC	180
3.23	Predicted PNPO-apoDDC complex showing the proposed key residue (PNPO-E114) interaction at the protein-protein interface.....	182

ABSTRACT

Protein Structure and Interaction: The Role of Aromatic Residues in Protein Structure,
and Interactions between Pyridoxine 5'-Phosphate Oxidase/Dopa Decarboxylase

By Mohammed H. AL Mughram, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2022.

Advisors:

Dr. Martin K. Safo
Professor, Department of Medicinal Chemistry

Dr. Glen E. Kellogg
Professor, Department of Medicinal Chemistry

Naturally developed proteins are capable of carrying out a wide variety of molecular functions due to their highly precise three-dimensional structures, which are determined by their genetically encoded sequences of amino acids. A thorough knowledge of protein structures and interactions at the atomic level will enable researchers to get a deep foundational understanding of the molecular interactions and enzymatic processes required for cells, resulting in more effective therapeutic interventions. This dissertation intends to use structural knowledge from solved protein structures for two distinct objectives.

In the first project, we conducted a bioinformatics structural analysis of experimental protein structures using our novel paradigm "3D Interaction Homology". The three-dimensional structure of biological macromolecules, particularly proteins, provides

us with a better understanding of protein interactions and functions, enabling us to establish hypotheses about how to modulate, regulate, or modify their functions. Therefore, the Kellogg lab proposed a new paradigm for the building and refinement of protein structure models using 3D hydrophobic maps. To accomplish this goal, we have been characterizing the hydrophobic interaction residue environments by compiling a database of residue type- and backbone angle-dependent 3D maps. In this work, 3D hydrophobic interaction maps feature of the HINT (Hydrophobic INTeraction) program enabled calculation and visualization of the 3D hydrophobic environments of the three aromatic amino acid residues. We have shown that these 3D maps are information rich descriptors of preferred conformations, interaction types and energetics, and solvent accessibility. We calculated and analyzed sidechain-to-environment 3D maps for over 70,000 phenylalanine, tyrosine, and tryptophan residues. Moreover, significant and occurrence of some special noncovalent interactions (π - π and π -cation) were calculated and analyzed. This recognition of even these subtle interactions in the 3D hydrophobic environment maps is key support for our interaction homology paradigm of protein structure elucidation and possibly prediction.

In the second project, we aimed to investigate the physical interaction between a vitamin B₆-salvage enzyme, pyridoxine-5' phosphate oxidase (PNPO), and a vitamin B₆-dependent enzyme, dopa decarboxylase (DDC), employing different approaches, including molecular modeling, biophysical, enzyme kinetics, and site-directed mutagenesis studies. PLP, the active vitamer of B₆, serves as a cofactor for approximately 180 B₆-dependent (PLP-dependent) enzymes and play crucial roles on many of cellular Functions, e.g., heme, amino acid, neurotransmitter, DNA/RNA biosynthesis. Vitamin B₆ deficiency is suspected to contribute to several pathologies, e.g., seizures, autism,

schizophrenia, epilepsy, and Alzheimer's disease. High levels of vitamin B₆ are also linked to neurotoxic effects due in part to potential toxicity by free PLP in the cell. Therefore, the cellular content of free PLP is kept very low. Understanding the role of this vitamin in these pathologies requires knowledge on its metabolism and regulation, and subsequent transfer to dozens of apo-B₆ enzymes. We hypothesize that the transfer of PLP from the donor PNPO salvage enzyme to the acceptor apo-B₆ enzyme DDC requires that both enzymes form a complex to offer an efficient and protected means of delivery of the highly reactive PLP. Knowledge of the 3D protein structures of PNPO and DDC (in both active state or holo-form and inactive state or apo-form) enabled us to undertake protein-protein docking and molecular dynamics simulations studies to predict the most likely near-native structure of the complex. The physical binding between PNPO and DDC were experimentally characterized using fluorescence polarization (FP), surface plasmon resonance (SPR), and isothermal calorimetry (ITC) techniques. The dissociation constants (K_D) was observed to be in low micromolar range. Expectedly, interactions between PNPO and apoDDC was found to be about 3-fold stronger than interactions between PNPO and holoDDC, with K_D values of $0.92 \pm 0.07 \mu\text{M}$ and $2.59 \pm 0.11 \mu\text{M}$, respectively. PLP transfer studies were carried out to demonstrate that PLP is capable of transferring from PNPO and activating the apoDDC. Site mutation investigations of critical residues identified by computational/modeling studies to be important in protein-protein interaction were carried out but showed negligible effect on the complex formation.

CHAPTER 1

GENERAL INTRODUCTION

1. Introduction

Proteins are the workhorses of the cellular machinery, and a major structural feature of cells and cell membranes. They play a variety of critical cellular roles and are essential for the normal structure, function, and regulation of body tissues and organs.^{1,2} Proteins serve as the molecular foundation for numerous biological activities, including signal transduction, transport, metabolism, gene expression, cell development, and proliferation.¹⁻⁴ Furthermore, essential cellular processes including DNA replication, transcription, translation, and transmembrane signal transduction are all dependent on specific proteins. Proteins are distinguished according to their usual functions in the body such as enzymes, antibodies, messengers, structural components, and transport/storage proteins.^{5,6}

The biological activity of a protein is governed by the type and arrangement of the amino acid sequence in three dimensions, which governs the protein fold, appropriate arrangement of amino acid residues in an active site, or specific interactions with other proteins for structural, regulatory or other functions. The three-dimensional structure of biological macromolecules, and proteins in particular, gives us a better knowledge of

protein interactions and functions, allowing us to develop hypotheses to modulate, regulate, or modify their functions. Knowledge of how cells and organs work require an understanding of how proteins function, which is associated with understanding structure and interactions of proteins. When we consider the number of available protein sequences to the number of available protein structures, it is clear that new protein sequences are being acquired at a far faster rate than structures.⁷ Indeed, with increasing structural knowledge of 3D macromolecules, understanding of protein function would be expanded, leading to more successful structure-based drug design and therapeutic interventions. Three-dimensional (3D) structure of proteins can be determined experimentally, using methods such as X-ray crystallography, nuclear magnetic resonance (NMR), Cryo-electron microscopy (Cryo-EM), and neutron diffraction, or, computationally, such as homology modeling, *Ab Initio* and threading methods. Each of these methods has its own set of advantages and drawbacks. On the other hand, proteins are involved in a complex protein-protein interactions (PPIs) network within cells.^{8,9} PPIs is a large, complex network of interactions that are necessary for most cellular functions and regulations. The human interactome is estimated to have between 130,000 and 650,000 PPIs.¹⁰ Structural characterization of protein-protein complexes remains difficult, with one significant reason being the weak and transient associations of many biologically relevant protein-protein interactions, which makes it difficult to capture many complexes in crystals.¹¹ On the other hand, computational techniques, such as molecular docking and dynamics simulations, are widely applied to bridge this gap.¹²

1.1 Experimental protein structure determination

1.1.1 Nuclear magnetic resonance (NMR): Nuclear magnetic resonance (NMR) spectroscopy is an indispensable analytical technique for organic chemists. It also serves, under certain conditions, as a powerful tool to identify the 3D structure of proteins. NMR is a technique for measuring the absorbance of radio frequency (RF) radiation through specific nuclei such as ^1H , ^{13}C , and ^{15}N when subjected to strong magnetic fields.¹³ This works on the basis of the fact that molecules are magnetic, and their nucleus spin in the presence of magnetic field. So that, an energy transfer from the base energy to a higher energy level is achievable when an external magnetic field is applied. Protons exhibit such spinning states (α or β forms) when external magnetic fields are introduced, and the variation between these states is proportional to the intensity of the applied magnetic field. Through adding the required pulse frequency of electromagnetic radiation specific to atom, the spin state can be changed from α form to β form (excited form), a process known as resonance. This resonance spectrum can be achieved by varying either the electromagnetic radiation or magnetic field while holding one as constant.¹³

One of the most key features for determining molecular structure using NMR is chemical shift. The difference between the resonance frequency of the spinning protons and the signal of the reference molecule is defined as chemical shift. Each nucleus in the molecule, ideally, has a distinct electronic environment and hence a distinct chemical

shift that can be recognized. Details about internuclear distances and atom positions can be determined by the Nuclear Overhauser effect (NOE), i.e., the magnetization effect is shifted from one nuclei to another when they are less than 5 Å apart. This includes the position of an atom in three-dimensional space, and can be used to recreate the protein's structure. Several considerations, including inadequate data set, estimated distance information and necessity of proteins in solution state, all have an impact on the determination of protein structure by NMR. In 2020-2021, 17,011 3D structures were deposited in the PDB, including 486 structures attributed to the NMR technique.¹⁵

1.1.2 X-ray crystallography: Until now, X-ray crystallography technique has been the primary means to characterize 3D structures of proteins and other macromolecules.¹⁶ The use of this technique began in the late 1950s, culminating in the crystal structure determinations of myoglobin and hemoglobin by Sir John Cowdery Kendrew and Max Perutz, respectively, for which they shared the 1962 Nobel Prize in Chemistry.¹⁷ The ability to acquire a crystal is a prerequisite for a successful structure determination. In fact, protein crystallization is a complex process due to lack of understanding of the actual physics underlying it. Many factors influence the crystallization processes, i.e., pH, temperature, protein and precipitant concentration, as well as protein purity and stability.¹⁸ Once a crystal is obtained, it is placed on a goniometer and exposed to X-rays that are generated when accelerated electron bombards an anode metal. The X-rays are scattered by the atomic electrons in the crystal producing diffraction patterns that contain structural details of the molecule. Since

a crystal contains infinite numbers of orderly and repeated molecules, the diffraction patterns add up to be detected. Laue and W. L. Bragg developed the foundation for this experiment, which is known as Bragg's law (equation 1); as a sequence, X-ray diffraction is analogous to the reflection of atoms in a plane of crystals. Constructive coherent interference must arise in order to produce a diffraction pattern that occurs only when the specified equation is fulfilled.

$$n\lambda = 2d \sin\theta \quad (1)$$

Where λ is the X-ray wavelength, d is the interplanar distance, and θ is the angle between the incident and planes.

The scattered X-rays by the crystal are captured by a charged coupled device (CCD) detector or an image photographic plate, to give diffraction intensity or amplitude data. The diffraction data can be influenced by several factors, such as crystal quality, crystal symmetry, X-ray intensity, and detector performance. To produce electron density maps, the diffraction or intensity data is processed using several crystallographic softwares for crystal system recognition, indexing, intensity calculation, and scaling factor. One of the key bottlenecks in crystallography is the phase problem (information of the direction of the intensity amplitudes is lost during data collection), but can be obtained using molecular replacement, and isomorphous replacement approaches. Electron density maps are then computed using a fast Fourier transform method with the amplitude and phase information. The electron density map is subsequently used to build the 3D atomic structure of the molecule, and the crude structure is subsequently refined, while being monitored for accuracy by an R-factor. While X-ray crystallography yields

the highest atomic resolution of protein structures, it has several drawbacks, including the need for a pure, stable, crystallizable protein in sufficient amounts. Larger proteins are even more difficult to express and crystallize, so target proteins larger than 250 kDa are difficult to solve using this common approach.¹⁹ According to Proteopedia, the median resolution attained with X-ray crystallography in the Protein Data Bank is 2.05 Å,²⁰ while the highest resolution achieved on record is 0.48 Å.²¹

1.1.3 Cryo-electron microscopy (Cryo-EM): Cryo-EM is a groundbreaking method that has the potential to revolutionize the study of biology.²² It is a decades-old technology for determining the form of flash-frozen objects by shooting electrons at the sample and capturing the pictures that arise. The two main kinds of electron microscopes (EM) are scanning and transmission. Transmission EM has better resolution than scanning electron microscopes, therefore it is favored for protein structure determination.²³ The transmission electron microscope is comprised of three major systems, electron gun that generates the electron beam that is focused on the specimen sample via a condenser mechanism. An image producing system that consists of lenses to help focus the electrons, resulting in a magnified image. Finally, an image recording system, which is mostly made up of a fluorescent screen that aids in the production of an image that can be seen with the naked eye. As a result of technological and computational advancements in these technologies over the last decade, it is now possible to obtain higher resolution atomic structures. There are several steps that are essential in determining the structure of a targeted sample using Cryo-EM, including specimen optimization and grid preparation, image acquisition, and image processing.

Interestingly, Cryo-EM has already substituted X-ray crystallography for the study of membrane-proteins, i.e., a field of study that is fast growing. Many membrane-bound proteins are involved in illnesses and serve as therapeutic targets. As of 2020, most protein structures determined using cryo-EM are at a lower resolution of 3–4 Å,²⁴ nonetheless as the technique improves, more and more proteins are being solved at higher resolutions.

1.2 Computational protein structure determination

Protein structures that cannot be determined experimentally or using previously stated techniques, frequently due to time/cost limitations or experimental challenges, may be modeled using a variety of computational methodologies. During the past few years, deep-learning (DL)-based approaches for protein structure prediction have received much attention and are fast becoming very useful tool in predicting protein structures with high accuracy. These algorithms use the accumulated information from 50 years of experimental structures characterization deposited in the PDB database by combining numerous deep learning advances. Their results are enlightening, with best-in-class performances made in the two most recent CASP (Critical Assessment of PRediction of Interactions) competitions.⁴⁶ These approaches, e.g., AlphaFold¹⁵⁶ and new versions of Rosetta (RoseTTAFold),¹⁵⁷ certainly show amazing promise. In general, DL-based approaches are trained to extract rich structural information from evolutionary data to make structure predictions, rather than relying on individual amino acid

sequences. However, before the advent of AlphaFold and artificial intelligence based approaches, three main approaches for modeling "predicting" 3D structure of proteins were known, i.e., homology modeling (comparative modeling), protein threading (fold recognition), and *ab initio* (*de novo*) approaches. The essential idea behind these three approaches is depicted in **Figure 1.1** and briefly described below.

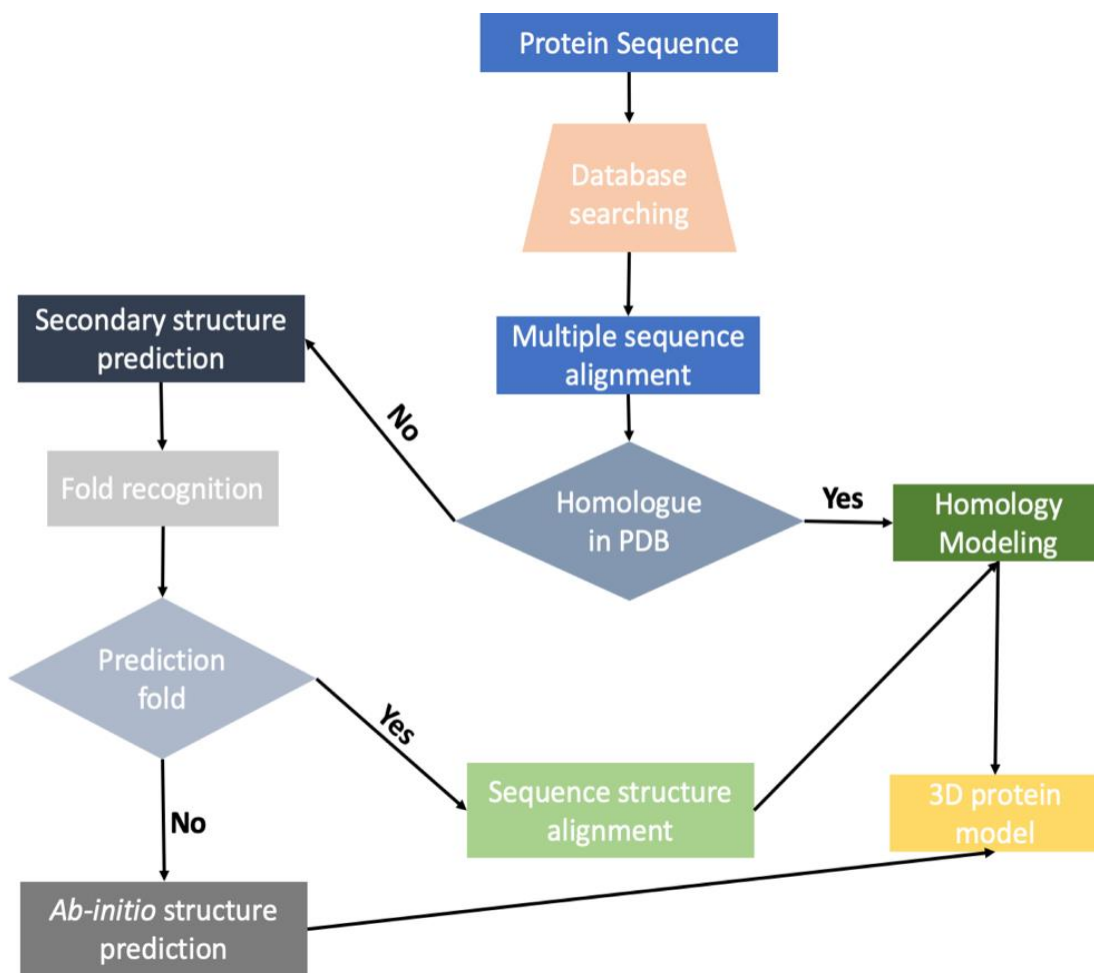


Figure 1.1 General scheme of model construction.

1.2.1 Homology modeling: This approach is commonly used to predict protein structures and aids in the construction of comparative models of unknown protein structures based on resolved structures of related homologous proteins.²⁵ This approach seeks one, or more, protein structures that are homologues to the target sequence (templates). The sequences can be aligned of both query and template proteins to fit them at its best. This method can also be used to predict unknown proteins using multiple templates for different parts of the query protein. The only necessity is that the query protein shares at least ~30% sequence identity with a template protein.²⁶ Once query and template proteins are aligned, an initial model is constructed using insertion, deletion, and substitution from templates. MODELLER,¹⁵³ YASARA,¹⁵⁴ and SWISS-MODEL¹⁵⁵ are three of the most used tools for predicting protein structure using the homology modeling approach.

1.2.2 Protein threading: Protein threading is close to homology modeling, but there is no requirement for homologues sequencing. The hypothesis, probably clear, that number of unique folds is smaller than the number of proteins in nature.²⁷ Protein domains, smaller subunits of protein, have independent folds from the protein. Parameters in this approach take into consideration not only the amino acids but also other properties such as secondary structure, solvent accessibility, and surrounding amino acids. I-TASSER^{28,29} (Iterative Threading ASSEmbly Refinement) and Phyre2³⁰ are two commonly used tools for threading-based protein structure determination.

1.2.3 *Ab initio* approach: This method varies from the previously described modeling and threading approaches. Rather than utilizing previously solved structure as a template, *ab initio* (*de novo*) methods seek to build three-dimensional structure model based on physical principles.³¹ The hypothesis is backed by many lines of evidence that the primary protein sequence comprises all the requires structural information for overall three-dimensional protein structure, i.e., almost all proteins fold into the structure with the lowest free energy state, which is dictated by the sequence of their amino acids. While a protein's overall shape may be encoded in its primary amino acid sequence, its folding may depend on chaperons for assistance. These methods, however, require substantial computational resources, and therefore are used mostly for small proteins.

A designed energy function can be used to perform a conformational search, which produces possible conformations (decoy structures) from which the most feasible ones can be identified and selected.³² *Ab initio* methods generally rely on three key aspects: energy function design, which can be used to sort and select the generated decoy structures based on physics- and knowledge-based energy functions; a conformational search engine which enables low-energy state identification through conformational search using Monte Carol (MC) and molecular dynamic (MD) methods; and finally, model selection strategy, which allows for selection of a near-native model from the structures created by the decoys. Despite the enormous computational hurdles, ongoing advances in computational sources have motivated the field of *de novo* protein structure prediction science.³³

1.3 Protein-protein docking

As mentioned earlier, protein-protein interactions (PPIs) are involved in many biological processes and functions. Due to this, characterizing the complex nature of interacting proteins have become an important approach for understanding biological processes and pathways, as well as designing effective PPI-targeting modulators. Most proteins, especially protein-protein complexes, do not have corresponding structures in the Protein Data Bank (PDB), raising the needs for precise and effective docking approaches for understanding and modulating their interactions. Identifying correct binding for protein-protein docked complexes is a difficult task due to the comparatively flat and large interfaces of protein-protein complexes.¹² A variety of docking algorithms have been developed over the years, with the majority of them consisting of two stages, i.e., docking (sampling) and ranking stages.^{34,35,35-39} During the docking stage, a large number of structures are generated, and potential conformations are sampled from the generated structures. The sampled conformations in the first stage are sorted by different scoring functions in the ranking stage. The computational cost of search algorithms is high in protein-protein docking due to the large number of degrees of freedom. In the rigid docking process, a number of search algorithms can be used to optimize rotations and translations of the docked protein, as well as the dihedral angles of residues in the refinement step. Fast Fourier transformation, geometric hashing and Monte Carlo searching algorithm are the three predominant search algorithms employed by different rigid-docking calculations.⁴⁰ Despite the difficulty of resolving structures of protein-

protein complexes, several docking methods, including easy-to-use servers, are available for performing protein-protein docking.^{34–36,40,41} In general, docking methodologies are commonly evaluated^{38,42–46} using the CAPRI (Critical Assessment of PRediction of Interactions) evaluation, it is community-wide experiments which evaluate the ability of current protein-docking methods to predict protein-protein interactions.^{42–45,47,48} Also, Weng's group has developed protein-protein docking benchmarks which can be used to develop and assess protein-protein docking methods.^{38,42,48,49} Cluspro 2.0, on the other hand, is one of the most successful Protein-Protein docking methods in the CAPRI last rounds.^{50,51} It uses a Fast Fourier Transform (FFT)-based docking with pairwise potential and low-energy clustering of top conformations. **Table 1.1** lists several well-known protein-protein docking algorithms that have been developed and widely used over the years.

Table 1.1– Example of Protein-Protein docking tools (docking algorithm, description, and CAPRI validation).

Server	Sampling algorithm	Algorithm description	CAPRI validated	Web address	Reference
Cluspro2.0	FFT	Using clustering to rank docked structures with better electrostatic and de-solvation energy scores	Yes	https://cluspro.bu.edu/login.php	34
HADDOCK	Monte Carlo	Based on information-driven docking approach to predict the complex structure	Yes	https://wenmr.science.uu.nl/haddock2.4/	41
GRAMMX	FFT	Including smoothed potentials, refinement, and knowledge-based scoring	Yes	http://vasser.compbio.ku.edu/resources/grammx/	37
ZDOCK	FFT	Evaluate binding modes with a scoring function that include shape complementary, ACE statistical potential and electrostatic terms.	NO	http://zdock.umassmed.edu/	52
RosettaDock	Monte Carlo	Employing rigid-body orientation optimization and side chain conformation to rank the docked solutions	Yes	https://rosie.rosettacommons.org/doc/king/	35
FireDock	Monte Carlo	Flexible docking for refining and scoring of docked complexes	NO	http://biointfo3d.cs.tau.ac.il/FireDock.php	36
IM_DISO (Docking and Interface Side-Chain Optimization)	Monte Carlo	Based on global energy optimization and entirely flexible sidechains	--	--	39

1.4 Molecular dynamics Simulation

Molecular dynamics (MD) simulation is on the verge of becoming a routine computational approach for deeper understanding of protein structure and function, as well as drug discovery and development efforts.⁵³ MD takes structural flexibility and entropic effects into consideration. As increasingly complex algorithms and hardware designs are implemented, this allows for more realistic estimation of the thermodynamics and kinetics of protein-protein and protein-ligand interaction and binding. Classical MD applies Newton's physics to the study of atom and molecule interactions and motions. It employs force fields to calculate and measure the forces between the system's interacting atoms.⁵⁴ Newton's laws are incorporated into MD simulations to generate successive configurations that provide trajectories for determining the position and velocity of particles over time. Many parameters, including free energy, kinetics measures, and other macroscopic variables, may be derived from MD trajectories and correlated to experimental observables.⁵⁵ The precision of the used physical model (force field) has a strong influence on the success of MD simulation performance. In general, simulations of biomolecules are performed using the CHARMM^{56,57}, AMBER^{58,59}, and GROMOS⁶⁰ forcefields.

1.5 Bioinformatics structural analysis of experimental protein structures

Our understanding of the structure and function of proteins continues to expand, assisted in large part by research that determines their three-dimensional structures.

These findings will continue to provide the groundwork for understanding how the structure of a protein influences its activity and for future drug discovery. The composition and character of each of the twenty amino acids is the determining factor in forming proteins, which contain a broad variety of forms and folds, and produce a phenomenal array of functions. This diverse collection of amino acid residues shares the same backbone, but each of the different sidechains vary in size, shape, polarity, and relative hydrophobicity. This variety in sidechain gives each unique roles and functional characteristics, with which they can participate in many different interactions that result in “proper” protein folding and function. Although amino acids were discovered in the early 1800s, and their roles and bonding to form proteins was revealed by Emil Fischer in the early 1900s,⁹⁸ since the advent of X-ray crystallography, sidechain conformational analysis has been the subject of many reports for understanding the precise conformational propensities^{99–104} and folding^{104–108} of amino acids in proteins because it is a window into structure and function.

1.5.1 Aromatic amino acids: Three amino acids: phenylalanine (PHE), tyrosine (TYR), and tryptophan (TRP) are referred to as “aromatic amino acids” as they possess an aromatic moiety in their sidechains. Altogether, these amino acids represent 8.5% of those observed in proteins (PHE, 3.9%), (TYR, 3.3%), and (TRP, 1.3%).¹⁰⁹ Their unique sidechain characters, with their conjugated planar rings, enable them to engage in multiple noncovalent hydrophobic and polar interactions, both as purely hydrophobic

moieties, but also in leveraging their aromatic character, such as π - π , π -CH, π -OH, π -NH, π -SH, π -cation, π -anion interactions, etc.^{110,111} The increasing availability of detailed structural information on proteins and macromolecules have increased our understanding of the roles of inter- and intra-aromatic sidechain interactions in many biological processes, including early folding events and stability,^{111,112} protein-nucleic acid interactions,^{113,114} carbohydrate-aromatic interactions,¹¹⁵ and protein-ligand recognitions.^{116–119} Thus, while their frequency of observation is a little more than half of average, their unique interaction properties, and with them being present in both buried and surface positions, have made PHE, TYR and TRP the subject of numerous protein structure and function studies.^{112,120–122} In contrast to simple hydrophobic amino acids, alanine (ALA), isoleucine (ILE), leucine (LEU), and valine (VAL), aromatic sidechain residues are also capable of participating in polar interactions as well: the surrounding H-donor residues and/or water molecules can make π -polar interactions with aromatic sidechains. Cationic sidechains of arginine (ARG) and lysine (LYS) can also engage in a “quat-binding” association known as the π -cation interaction that can occur between the face of an aromatic sidechain (its π -electrons) and the positively charged $-\text{NH}_3^+$ of LYS or the (three) charged $-\text{NH}_x^{\delta+}$ of ARG sidechains. The first and seminal observation of amine/aromatic interactions within proteins was made by Perutz,³² which was followed by Levitt and Perutz, where an H-bond phenomenon was suggested between the aromatic sidechains and charged amino groups and further structural analysis by Burley and Petsko.¹²³ Later reports by Mitchell and Thornton,¹²⁴ however, indicated that this interaction is “remarkably rare” in protein structures. In this

sense, many published reports have been dedicated to a better understanding of the occurrence and magnitude of π -cation interactions in many aspects including protein structure,^{110,125,126} molecular recognition and catalysis.^{126–128} Modeling and dissecting the π -cation interaction in terms of its origin and strength was the core subject of numerous computational^{129–134} and experimental^{135–137} studies. The quantum mechanical calculations of π -cation interactions have been found to be correlated with experimental studies in the gas phase. The latter is important because the “real” biological environment of proteins is substantially different and more complex than the gas phase, so the translation of quantitative estimates of QM calculated energetics to proteins is challenging. Nevertheless, the existence of π -cation interactions in aqueous solutions have been computationally^{138–140} and experimentally^{125,126,141} acknowledged. In addition, peptide models used to examine the stabilizing forces of π -cation interactions concluded that they provided no substantial contribution to peptide stabilization and that stability was most likely due to hydrophobic effects.^{142,143} Clearly, however, the understanding of the structural and functional contributions and roles of hydrophobic/polar interactions for aromatic amino acid sidechains in biological systems remains a topic of interest. Furthermore, in the current era of accelerating interest in de novo protein design and engineering that impacts multiple fields with, e.g., enzyme variants with improved catalytic activities, more accessible physiochemical attributes, altered substrate specificities and/or and stereoselectivities, it will be valuable to understand the underlying blueprint of all aspects of protein structure.

1.5.2 3D hydrophobic interaction maps for protein structure description: The diversity of amino acid sidechains provides each with distinct roles and functional features, allowing them to engage in a wide range of interactions that result in "proper" protein folding and function. One product of HINT (Hydrophobic INTERaction) program^{144–151} that has been utilized is its "3D interaction map" feature. For each residue of interest, four general map types are calculated: favorable and unfavorable hydrophobic and polar. These maps can define the hydrophobic environment surrounding each residue's sidechain and/or backbone in 3D space. These maps can be also contoured and displayed in 3D space to show the source, nature, and strength of interactions (see Figure 1.2).

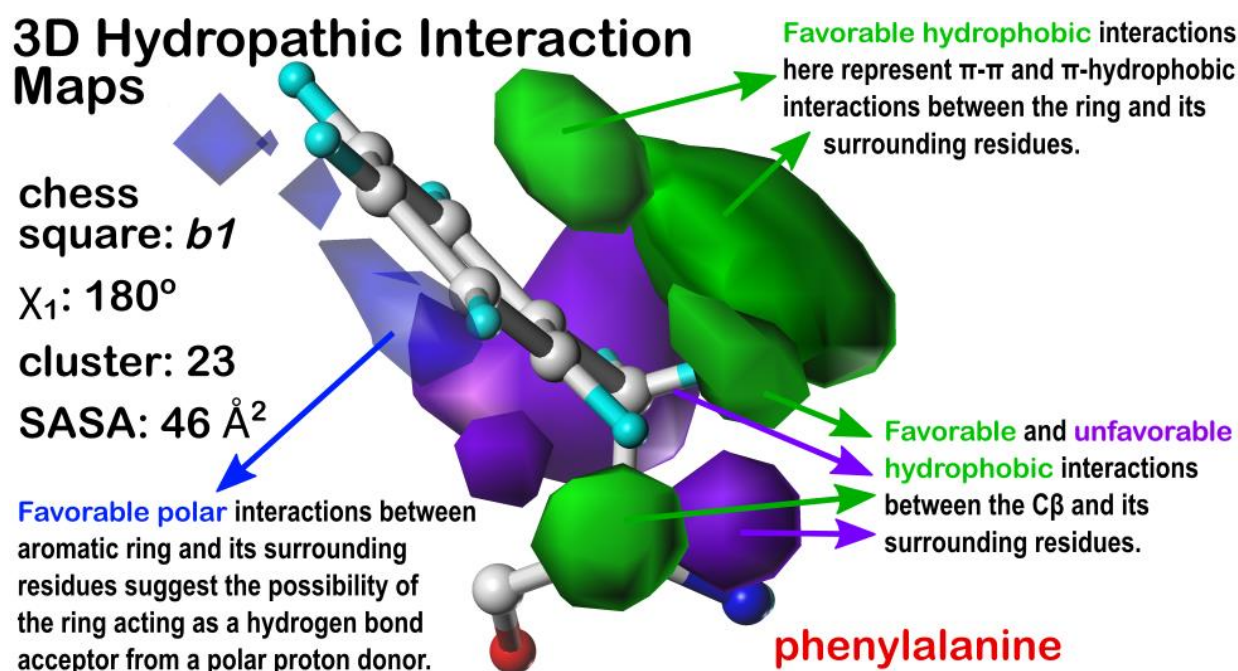


Figure 1.2 Hydrophobic interaction map representing the Gaussian-weighted average sidechain environment for a PHE sidechain cluster in our dataset. The green and purple

contours represent favorable and unfavorable hydrophobic interactions, respectively. The blue and red contours, respectively, reflect favorable and unfavorable polar interactions.

In the first contribution, tyrosine was the focused case by Kellogg's lab, in which (~30000) TYR residue maps in the dataset were compressed into 262 backbone-dependent environments represented by composite maps¹⁴⁶. Importantly, analysis describes residues by contextualizing the hydrophobic environments of the surrounding atoms and environment. For instance, a properly oriented HOH, a residue or a ligand can provide the same 'acidic' character as a TYR –OH and serine (SER) –OH or even a LYS –NH₃⁺ to satisfy its "*hydrophobic valence*". Thus, protein structure is not driven by sequence, but instead by the hydrophobic interactions that each residue makes in terms of its sidechain and backbone conformations.

In the second article, the focus was on alanine and its structural and functional roles. For nearly 57,000 alanine residues in the dataset, backbone-to-environment and sidechain-to-environment 3D maps were calculated and analyzed separately.¹⁴⁷ All ALA maps were compressed into backbone- and sidechain-dependent environments of 136 and 150, respectively. Following that, aromatic amino acid residues,¹⁵⁰ as described in this dissertation, cysteine and serine,¹⁴⁸ and lastly aspartic acid, glutamic acid, and histidine,¹⁴⁹ were studied and analyzed using this novel motif of protein structure description.

1.6 Vitamin B₆ (B₆ vitamers and metabolic pathway)

Vitamin B₆, arguably the most important vitamin, is comprised of six forms (**Figure 1.3**), including the primary forms pyridoxal (PL), pyridoxine (PN), and pyridoxamine (PM), and their phosphate derivatives PLP, PNP, and PMP, respectively. PLP serves as a cofactor for approximately 180 B₆-dependent (PLP-dependent) enzymes involved in glucose, lipid and amino acid metabolism, as well as heme, DNA/RNA and neurotransmitter syntheses.⁶¹ B₆-dependent enzymes are biosynthesized as apo-B₆ enzymes and then converted to the catalytically active holo-B₆ enzymes, mostly by covalently binding to PLP as an aldimine. To synthesize PLP from the primary vitamers PNP and PMP, and/or recycle PLP during protein turnover, humans depend on a salvage pathway involving pyridoxal kinase (PLKase), pyridoxine 5'-phosphate oxidase (PNPO), and pyridoxal phosphatase (**Figure 1.4**).^{62–65} PLKase phosphorylates the primary B₆, pyridoxine (PN), pyridoxamine (PM) and pyridoxal (PL), respectively, to PNP, PMP and PLP. PNPO converts PNP and PMP to the active PLP. During protein turnover, phosphatases convert PLP to PL.

and substitution.^{66–68} These processes are responsible for neurotransmitter production, heme biosynthesis, one carbon unit transfer, and hence nucleic acid production, sphingomyelin synthesis, and glucose metabolism.^{61,64} Since the structure of the first PLP-dependent enzyme, aspartate aminotransferase (AAT) was determined, the structures of numerous other PLP-dependent enzymes have been characterized, as well as efforts to categorize these enzymes. Based on their fold, active site, and amino acid sequence similarity, PLP-dependent enzymes are classified into five distinct subtypes⁶⁹, this categorization is shown in **Table 1.2** and **Figure 1.5**.

Table 1.2– Classification PLP-dependent enzymes with representative examples of each class.

Fold type I	Aspartate amino transferase (AAT)
	Serine hydroxymethyltransferase (SHMT)
	Dopa decarboxylase (DDC)
Fold type II	Serine dehydratase
	Cystathionine β -synthase
Fold type III	Alanine racemase
	Ornithine decarboxylase
Fold type IV	Branched-chain amino-acid amino transferase
Fold type V	Glycogen phosphorylase

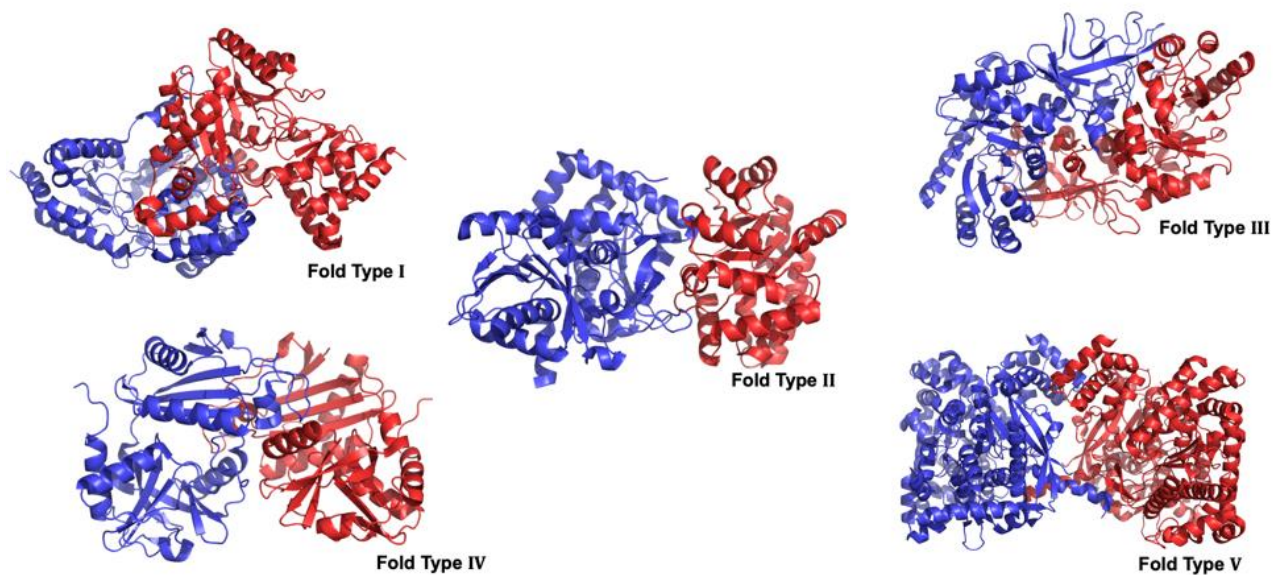


Figure 1.5 PLP-dependent enzymes superfamilies.⁷⁰ Fold-type I: aspartate transaminase (PDB: 5T4L); fold-type II: tryptophan synthase (PDB: 5CGQ); fold-type III: alanine racemase (PDB: 2VD8); fold-type IV: D-amino acid aminotransferase (PDB: 3LQS); and fold-type V: glycogen phosphorylase (PDB: 5OWY).

This categorization provides a good illustration of the variety of PLP-dependent enzymes and helps to explain and clarify the physiological relevance of these enzymes, as well as of PLP, a cofactor required for their functioning. Any change in PLP levels may have a detrimental effect on a variety of physiological processes through the failure of PLP-dependent enzymes functions.

1.6.2 PLP deficiency and neurological pathologies: Deficiency of cellular B₆ vitamers, or mutations in the B₆-salvage enzymes or key B₆-dependent enzymes, is

known or suspected to contribute to several pathologies, e.g., seizures, hallucination, headache, convulsions, autism, Down syndrome, schizophrenia, Neonatal Epileptic Encephalopathy (NEE), Parkinson's, Alzheimer's, and learning disabilities.^{65,71-76} High intake of vitamin B₆ is also linked to neurotoxic effects as a result of the PLP reactive aldehyde forming aldimines with free amino groups on non-B₆ proteins, disrupting their function.^{63,64,77,78} To avoid toxicity, the cell maintains a very low concentration of free PLP (unbound) through dephosphorylation to PL by pyridoxal phosphatase, as well as PLP feed-back inhibition of PNPO and PLKase. PLP-dependent enzymes, e.g., serine hydroxymethyltransferase (SHMT), DOPA decarboxylase (DDC/AADC), glutamate decarboxylase, serine racemase, are examples of B₆ enzymes involved in the synthesis of key neurotransmitters, e.g., D-serine, D-aspartate, glycine, GABA, serotonin, histamine, Dopamine, serotonin, norepinephrine (NE) and epinephrine. Low levels of GABA, dopamine and serotonin have been implicated in different neurological disorders as we already mentioned. Neonatal Epileptic Encephalopathy (NEE) is the most well-documented neuropathology linked with impaired B₆ activity.⁷⁹⁻⁸³ NEE patients does not respond to conventional anti-epileptic treatments, and without proper treatment eventually die as a result of progressive deterioration. NEE is caused by mutations in the human PNPO gene, in which there are currently 18 identified mutations (see **Table 1.3**).⁸¹ NEE patients exhibit decreased PLP levels and/or activity of multiple B₆ enzymes, resulting in a variety of clinical phenotypes, including seizures, fetal pain, stunted growth, hypoglycemia, anemia, asphyxia. PLP injection in large doses alleviate seizures and other,⁸⁴ demonstrating the critical role PNPO and PLKase in the production of sufficient PLP to meet the requirements of apo-B₆ enzymes.

Table 1.3– PNPO gene mutations affecting enzyme structure and function.

PNPO mutation	Location	PNPO mutation	Location
D33V	N-term	R141C	β -sheet 5
E50K	N-term	R161C	Loop9
L83W	β -sheet 1	P213S	β -sheet 6
R95C/H	β -sheet 2	R225C/L/H	Loop12
R116Q	Helix 2	R229W/Q	β -sheet 7
G118Q/R	Helix 2	P150R	Helix 3
E120K	Helix 2	X262Q	

1.6.3 DDC and AADC deficiency: DOPA decarboxylase (DDC) or aromatic L-amino acid decarboxylase (AADC) is a PLP-dependent enzyme that catalyzes the conversion of aromatic amino acids to their corresponding amines during the synthesis of a variety of essential neurotransmitters.⁸⁵ DDC catalyzes the decarboxylation of L-DOPA into Dopamine (DA), 5-HTP to serotonin, L-histidine to histamine, L-tyrosine to tyramine, and tryptophan to tryptamine (**Figure 1.6**). DDC is a major target for the research of Parkinson's disease, depression, and other neurological illnesses because of its essential involvement in neurotransmitter biosynthesis.^{86–91} One genetic disorder associated with DDC/AADC deficiency is called aromatic L-amino acid decarboxylase deficiency (AADCD) disorder, an inherited disorder affecting synthesis of neurotransmitters that results in serotonin and catecholamine deficiency.^{91,92} This disorder manifests in early childhood and results in significant developmental impairment as well as lifelong motor, behavioral, and autonomic symptoms such as oculogyric crises

(OGC), sleep disorders, and mood disorders.^{91,93} So far, more than 20 mutations in the AADC gene have been characterized and identified.⁹⁴ Clinical phenotypes of patients with AADC deficiency are well established; nevertheless, enzymatic phenotypes of AADC patients are mostly unknown and molecular mechanisms by which some of these mutations result in DDC deficiency remain unclear and ambiguous. Some of these mutations, i.e., S147R, G102S, F309L, and A275T, are located in the active site and most likely to impair the integrity of the active site, resulting in decreased PLP binding. In addition, solving the crystal structure of apoDDC (open conformation) has shown that some of these mutations in Loop1, i.e., Y79C, H70T, H72Y, and T69M, required for proper conformational change during the PLP-induced apo-holo transition.⁹⁵⁻⁹⁷ This effect was further studied by Giardina et al. and shown that the K_D of PLP is significantly affected, i.e., 100 nM, 510 nM, 2145 nM, 487 nM, 1520 nM, 390 nM, for T69M, H70T, H72Y, Y79C, F80A and P81L, respectively, compared to the wild-type DDC (43 nM). Further mutations, L38P and A110Q, are in the dimer interface and likely to perturb the dimerization configuration process. Other mutations, on the other hand, are in locations that are neither engaged in dimerization nor in catalytic site, and some of them may be ascribed to affecting PLP transfer through physical binding with salvage enzymes.

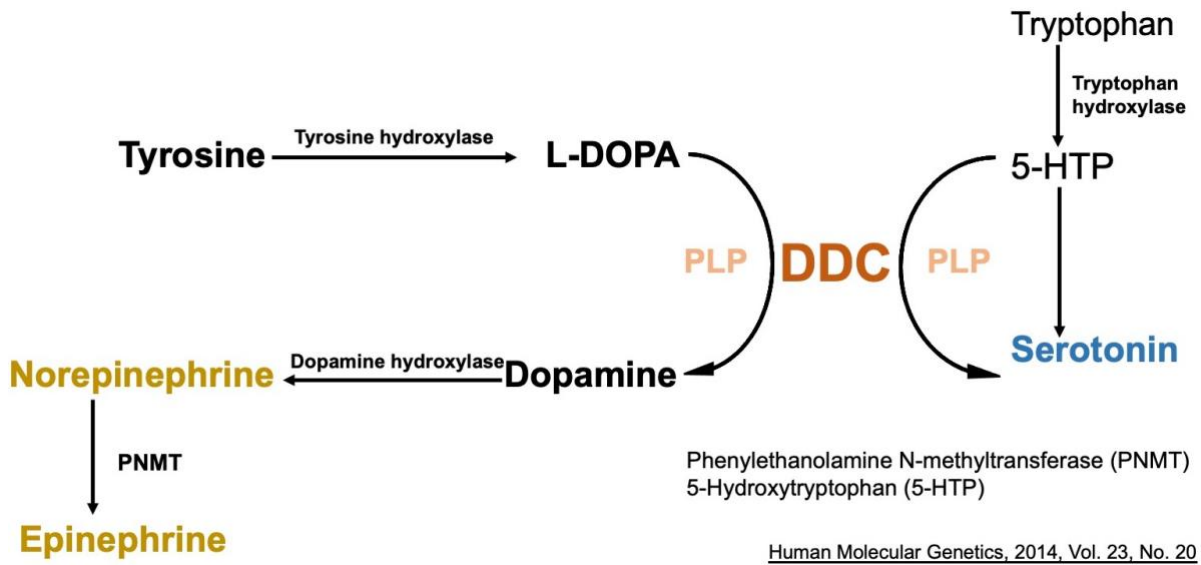


Figure 1.6 The metabolic pathway involves DDC for the synthesis of dopamine and serotonin neurotransmitters.

1.7 Research plan

Our understanding of protein structure and interaction is constantly evolving and expanding. These findings will continue to serve as the basis for understanding how protein structures influence their functions, allowing us to develop better therapeutic interventions. Many proteins, are engaged in a complex network of protein-protein interactions (PPI) within cells,¹⁰ that are essential for the majority of cellular function and regulation.^{8,152} Many biologically relevant protein-protein interactions are difficult to solve experimentally due to their weak and transient nature.⁸ However, *in silico* protein structure and interaction prediction have accelerated in the last decade, with improved data processing and prediction efficiency, which would significantly contribute to the field and narrow this gap.⁵¹ In this dissertation, the overall objective can be divided into two goals; 1) utilizing computational approaches to explore the structural and functional roles of aromatic amino acid residues in protein structure; and 2) investigating the nature of the protein-protein interaction between the vitamin B₆-salvage enzyme pyridoxine 5'-phosphate oxidase (PNPO) and the vitamin B₆-dependent enzyme dopa decarboxylase (DDC). The rational for these two projects are described below:

Project 1. 3D Interaction Homology: Hydrophobic Analyses of the “ π -Cation” and “ π - π ” Interaction Motifs in Phenylalanine, Tyrosine, and Tryptophan Residues.

Three-dimensional maps of the hydrophobic environments of protein amino acid residues are rich in information regarding preferred conformations, interaction types and energies,

and solvent accessibility. Each residue's interactions are the major determinant of rotamer selection as well as secondary, tertiary, and even quaternary protein structure. One product of the HINT (Hydropathic INTeraction) program^{144,145} that has been implemented is its 3D interaction maps feature. These maps can be used to calculate and determine the environment around protein amino acid residues with respect to their hydrophobic and polar interactions.^{146,147} These maps can be also contoured and displayed to illustrate types, strength, and loci of these interactions in 3D space. In **chapter 2**, we focused on the structural roles of the aromatic residues' phenylalanine, tyrosine, and tryptophan. In the first part, we calculated and analyzed sidechain-to-environment 3D maps for over 70,000 residues of these three types that reveal, with respect to hydrophobic and polar interactions, the environment around each. We also attempted to study the hydrophobic interaction character of the space above and below the aromatic rings of these residues. This was followed by calculating and analyzing of key non-covalent interactions, i.e., π -cation and π - π types of interactions, and evaluating the strength and prevalence of these interactions throughout our relatively large data set.

Project 2. Computational, biophysical, and kinetic approaches to elucidate interactions between PNPO and DDC. PLP serves as a cofactor for approximately 180 B₆-dependent (PLP-dependent) enzymes involved in glucose, lipid and amino acid metabolism, as well as heme, DNA/RNA and neurotransmitter syntheses.^{60,63} B₆-dependent enzymes are biosynthesized as apo- B₆ enzymes and then converted to the catalytically active holo-B₆ enzymes, by covalently binding to PLP as an aldimine. The

level of PLP in healthy people's plasma is reported to be low, around 40 nmol/L.⁸ High B₆ intake has been associated to neuropathy, apparently due to the PLP reactive aldehyde forming aldimines with non-B₆ proteins.^{1-2,6} To minimize toxicity, the cell keeps the content of free PLP (unbound) very low through dephosphorylation and PLP feedback inhibition of PNPO and PLKase. This raises the unanswered question of how a low concentration of free PLP converts dozens of competing apo-B₆ enzymes to their active holo-forms. Our group was able to characterize the physical complex between the B₆-salvage enzymes, PNPO and PLKase, and several B₆-dependent enzymes using a fluorescent polarization approach.⁷⁸ However, our interest in atomic-level understanding of the complex structure continues. This current study explores the mechanism by which apoDDC are converted to their catalytically active holo-form *in vitro*. We tested the central hypothesis that the transfer of PLP from the salvage enzyme (PNPO) to apoDDC involves channeling, which appears to offer an efficient and protected means of delivery of the highly reactive PLP. Channeling requires that the donor and acceptor enzymes form a physical contact, and inability of such complex formation may lead to detrimental outcome. In **Chapter 3.2**, we reported the findings of molecular docking and dynamic simulation studies of the PNPO•holoDDC and PNPO•apoDDC complexes. The goal of the modeling study is to gain insight into how these two enzymes potentially recognize each other and provide directions for testable hypothesis through site-directed mutagenesis. So, what are the key residues that have a high energy contribution to the formation of the complex? Following, experimental explorations involving biophysical and PLP transfer kinetic experiments were conducted (**Chapter 3.3**). Summarily, these studies attempt to establish experimentally if transfer of PLP from PNPO to apoDDC

involves complex formation! In **Chapter 3.4**, site-directed mutagenesis studies were carried out on key residues, as predicted by our modeling studies, to evaluate their ability to influence the complex formation.

References:

- (1) *Molecular Biology of the Cell*, 4th ed.; Alberts, B., Ed.; Garland Science: New York, 2002.
- (2) Zaretsky, J. Z.; Wreschner, D. H. Protein Multifunctionality: Principles and Mechanisms. *Transl Oncogenomics* **2008**, *3*, 99–136.
- (3) Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **1998**, *92* (3), 291–294. [https://doi.org/10.1016/S0092-8674\(00\)80922-8](https://doi.org/10.1016/S0092-8674(00)80922-8).
- (4) Brady, A. E.; Limbird, L. E. G Protein-Coupled Receptor Interacting Proteins: Emerging Roles in Localization and Signal Transduction. *Cellular Signalling* **2002**, *14* (4), 297–309. [https://doi.org/10.1016/S0898-6568\(01\)00239-X](https://doi.org/10.1016/S0898-6568(01)00239-X).
- (5) Materi, W.; Gombos, Z.; Wishart, D. S. Proteins: Hormones, Enzymes, and Monoclonal Antibodies-Background. In *Pharmaceutical Sciences Encyclopedia*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; p pse312. <https://doi.org/10.1002/9780470571224.pse312>.
- (6) Ur Rehman, H.; Azam, N.; Yao, J.; Benso, A. A Three-Way Approach for Protein Function Classification. *PLoS ONE* **2017**, *12* (2), e0171702. <https://doi.org/10.1371/journal.pone.0171702>.
- (7) Muhammed, M. T.; Aki-Yalcin, E. Homology Modeling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chem Biol Drug Des* **2019**, *93* (1), 12–20. <https://doi.org/10.1111/cbdd.13388>.
- (8) Perkins, J. R.; Diboun, I.; Dessailly, B. H.; Lees, J. G.; Orengo, C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure* **2010**, *18* (10), 1233–1243. <https://doi.org/10.1016/j.str.2010.08.007>.
- (9) Speer, S. L.; Zheng, W.; Jiang, X.; Chu, I.-T.; Guseman, A. J.; Liu, M.; Pielak, G. J.; Li, C. The Intracellular Environment Affects Protein–Protein Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (11), e2019918118. <https://doi.org/10.1073/pnas.2019918118>.
- (10) Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent Advances in the Development of Protein–Protein Interactions Modulators: Mechanisms and Clinical Trials. *Sig Transduct Target Ther* **2020**, *5* (1), 213. <https://doi.org/10.1038/s41392-020-00315-3>.
- (11) Nakajima, N.; Hayashida, M.; Jansson, J.; Maruyama, O.; Akutsu, T. Determining the Minimum Number of Protein-Protein Interactions Required to Support Known Protein Complexes. *PLoS ONE* **2018**, *13* (4), e0195545. <https://doi.org/10.1371/journal.pone.0195545>.
- (12) Vakser, I. A. Challenges in Protein Docking. *Current Opinion in Structural Biology* **2020**, *64*, 160–165. <https://doi.org/10.1016/j.sbi.2020.07.001>.
- (13) Silverstein, R. M.; Webster, F. X.; Kiemle, D. J.; Bryce, D. L. *Spectrometric Identification of Organic Compounds*, Eighth edition.; Wiley: Hoboken, NJ, 2015.
- (14) Bax, A. Multidimensional Nuclear Magnetic Resonance Methods for Protein Studies. *Current Opinion in Structural Biology* **1994**, *4* (5), 738–744. [https://doi.org/10.1016/S0959-440X\(94\)90173-2](https://doi.org/10.1016/S0959-440X(94)90173-2).
- (15) Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (16) Drenth, J.; Mesters, J. *Principles of Protein X-Ray Crystallography*, 3rd ed.; Springer: New York, 2007.
- (17) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. <https://doi.org/10.1038/181662a0>.

- (18) Parker, M. W. Protein Structure from X-Ray Diffraction. *Journal of Biological Physics* **2003**, 29 (4), 341–362. <https://doi.org/10.1023/A:1027310719146>.
- (19) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and Limitations of X-Ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem. Int. Ed.* **2003**, 42 (24), 2718–2736. <https://doi.org/10.1002/anie.200200539>.
- (20) Martz, E.; Namba, K.; Theis, K. *Resolution*. <https://proteopedia.org/wiki/index.php/Resolution> (accessed 2022-07-06).
- (21) Schmidt, A.; Teeter, M.; Weckert, E.; Lamzin, V. S. Crystal Structure of Small Protein Crambin at 0.48 Å Resolution. *Acta Crystallogr F Struct Biol Cryst Commun* **2011**, 67 (4), 424–428. <https://doi.org/10.1107/S1744309110052607>.
- (22) Shoemaker, S. C.; Ando, N. X-Rays in the Cryo-Electron Microscopy Era: Structural Biology's Dynamic Future. *Biochemistry* **2018**, 57 (3), 277–285. <https://doi.org/10.1021/acs.biochem.7b01031>.
- (23) Kisielowski, C.; Freitag, B.; Bischoff, M.; van Lin, H.; Lazar, S.; Knippels, G.; Tiemeijer, P.; van der Stam, M.; von Harrach, S.; Stekelenburg, M.; Haider, M.; Uhlemann, S.; Müller, H.; Hartel, P.; Kabius, B.; Miller, D.; Petrov, I.; Olson, E. A.; Donchev, T.; Kenik, E. A.; Lupini, A. R.; Bentley, J.; Pennycook, S. J.; Anderson, I. M.; Minor, A. M.; Schmid, A. K.; Duden, T.; Radmilovic, V.; Ramasse, Q. M.; Watanabe, M.; Erni, R.; Stach, E. A.; Denes, P.; Dahmen, U. Detection of Single Atoms and Buried Defects in Three Dimensions by Aberration-Corrected Electron Microscope with 0.5-Å Information Limit. *Microsc Microanal* **2008**, 14 (5), 469–477. <https://doi.org/10.1017/S1431927608080902>.
- (24) Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-Resolution Protein Structure Determination by Cryo-EM. *Nature* **2020**, 587 (7832), 157–161. <https://doi.org/10.1038/s41586-020-2833-4>.
- (25) Vyas, V.; Ukawala, R.; Chintha, C.; Ghate, M. Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. *Indian J Pharm Sci* **2012**, 74 (1), 1. <https://doi.org/10.4103/0250-474X.102537>.
- (26) Xiang, Z. Advances in Homology Protein Structure Modeling. *CPPS* **2006**, 7 (3), 217–227. <https://doi.org/10.2174/138920306777452312>.
- (27) Lemer, C. M.-R.; Rooman, M. J.; Wodak, S. J. Protein Structure Prediction by Threading Methods: Evaluation of Current Techniques. *Proteins* **1995**, 23 (3), 337–355. <https://doi.org/10.1002/prot.340230308>.
- (28) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat Methods* **2015**, 12 (1), 7–8. <https://doi.org/10.1038/nmeth.3213>.
- (29) Zheng, W.; Zhang, C.; Wuyun, Q.; Pearce, R.; Li, Y.; Zhang, Y. LOMETS2: Improved Meta-Threading Server for Fold-Recognition and Structure-Based Function Annotation for Distant-Homology Proteins. *Nucleic Acids Research* **2019**, 47 (W1), W429–W436. <https://doi.org/10.1093/nar/gkz384>.
- (30) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat Protoc* **2015**, 10 (6), 845–858. <https://doi.org/10.1038/nprot.2015.053>.
- (31) Lee, J.; Wu, S.; Zhang, Y. Ab Initio Protein Structure Prediction. In *From Protein Structure to Function with Bioinformatics*; Rigden, D. J., Ed.; Springer Netherlands: Dordrecht, 2009; pp 3–25. https://doi.org/10.1007/978-1-4020-9058-5_1.
- (32) Guan, W.; Ozakin, A.; Gray, A.; Borreguero, J.; Pandit, S.; Jagielska, A.; Wroblewska, L.; Skolnick, J. Learning Protein Folding Energy Functions. In *2011 IEEE 11th International*

Conference on Data Mining; IEEE: Vancouver, BC, Canada, 2011; pp 1062–1067.

<https://doi.org/10.1109/ICDM.2011.88>.

(33) Jothi, A. Principles, Challenges and Advances in Ab Initio Protein Structure Prediction. *PPL* **2012**, 19 (11), 1194–1204. <https://doi.org/10.2174/092986612803217015>.

(34) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein–Protein Docking. *Nat Protoc* **2017**, 12 (2), 255–278. <https://doi.org/10.1038/nprot.2016.169>.

(35) Lyskov, S.; Gray, J. J. The RosettaDock Server for Local Protein-Protein Docking. *Nucleic Acids Research* **2008**, 36 (Web Server), W233–W238.

<https://doi.org/10.1093/nar/gkn216>.

(36) Mashiach, E.; Schneidman-Duhovny, D.; Andrusier, N.; Nussinov, R.; Wolfson, H. J. FireDock: A Web Server for Fast Interaction Refinement in Molecular Docking. *Nucleic Acids Research* **2008**, 36 (Web Server), W229–W232. <https://doi.org/10.1093/nar/gkn186>.

(37) Tovchigrechko, A.; Vakser, I. A. GRAMM-X Public Web Server for Protein-Protein Docking. *Nucleic Acids Research* **2006**, 34 (Web Server), W310–W314.

<https://doi.org/10.1093/nar/gkl206>.

(38) Hwang, H.; Pierce, B.; Mintseris, J.; Janin, J.; Weng, Z. Protein-Protein Docking Benchmark Version 3.0. *Proteins* **2008**, 73 (3), 705–709. <https://doi.org/10.1002/prot.22106>.

(39) Fernández-Recio, J.; Totrov, M.; Abagyan, R. ICM-DISCO Docking by Global Energy Optimization with Fully Flexible Side-Chains: ICM Protein-Protein Docking and Refinement. *Proteins* **2003**, 52 (1), 113–117. <https://doi.org/10.1002/prot.10383>.

(40) Kanguene, P.; Nilofar, C. Protein-Protein Docking: Methods and Tools. In *Protein-Protein and Domain-Domain Interactions*; Springer Singapore: Singapore, 2018; pp 161–168.

https://doi.org/10.1007/978-981-10-7347-2_14.

(41) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; C. Schmitz; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **2016**, 428 (4), 720–725.

<https://doi.org/10.1016/j.jmb.2015.09.014>.

(42) Hwang, H.; Vreven, T.; Janin, J.; Weng, Z. Protein-Protein Docking Benchmark Version 4.0: Protein-Protein Docking Benchmark Version 4.0. *Proteins* **2010**, 78 (15), 3111–3114.

<https://doi.org/10.1002/prot.22830>.

(43) Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J. E.; Vajda, S.; Vakser, I.; Wodak, S. J. CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins* **2003**, 52 (1), 2–9. <https://doi.org/10.1002/prot.10381>.

(44) Lensink, M. F.; Méndez, R.; Wodak, S. J. Docking and Scoring Protein Complexes: CAPRI 3rd Edition. *Proteins* **2007**, 69 (4), 704–718. <https://doi.org/10.1002/prot.21804>.

(45) Lensink, M. F.; Velankar, S.; Wodak, S. J. Modeling Protein-Protein and Protein-Peptide Complexes: CAPRI 6th Edition: Modeling Protein-Protein and Protein-Peptide Complexes. *Proteins* **2017**, 85 (3), 359–377. <https://doi.org/10.1002/prot.25215>.

(46) Lensink, M. F.; Brysbaert, G.; Mauri, T.; Nadzirin, N.; Velankar, S.; Chaleil, R. A. G.; Clarence, T.; Bates, P. A.; Kong, R.; Liu, B.; Yang, G.; et al. Prediction of Protein Assemblies, the next Frontier: The CASP14-CAPRI Experiment. *Proteins* **2021**, 89 (12), 1800–1823.

<https://doi.org/10.1002/prot.26222>.

(47) Kastritis, P. L.; Bonvin, A. M. J. J. Are Scoring Functions in Protein–Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. *J. Proteome Res.* **2010**, 9 (5), 2216–2225. <https://doi.org/10.1021/pr9009854>.

- (48) Vreven, T.; Moal, I. H.; Vangone, A.; Pierce, B. G.; Kastiris, P. L.; Torchala, M.; Chaleil, R.; Jiménez-García, B.; Bates, P. A.; Fernandez-Recio, J.; Bonvin, A. M. J. J.; Weng, Z. Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *Journal of Molecular Biology* **2015**, *427* (19), 3031–3041. <https://doi.org/10.1016/j.jmb.2015.07.016>.
- (49) Mintseris, J.; Wiehe, K.; Pierce, B.; Anderson, R.; Chen, R.; Janin, J.; Weng, Z. Protein-Protein Docking Benchmark 2.0: An Update. *Proteins* **2005**, *60* (2), 214–216. <https://doi.org/10.1002/prot.20560>.
- (50) Vajda, S.; Yueh, C.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Kozakov, D. New Additions to the ClusPro Server Motivated by CAPRI. *Proteins* **2017**, *85* (3), 435–444. <https://doi.org/10.1002/prot.25219>.
- (51) Desta, I. T.; Porter, K. A.; Xia, B.; Kozakov, D.; Vajda, S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* **2020**, *28* (9), 1071-1081.e3. <https://doi.org/10.1016/j.str.2020.06.006>.
- (52) Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. ZDOCK Server: Interactive Docking Prediction of Protein–Protein Complexes and Symmetric Multimers. *Bioinformatics* **2014**, *30* (12), 1771–1773. <https://doi.org/10.1093/bioinformatics/btu097>.
- (53) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol* **2011**, *9* (1), 71. <https://doi.org/10.1186/1741-7007-9-71>.
- (54) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106* (5), 1589–1615. <https://doi.org/10.1021/cr040426m>.
- (55) Childers, M. C.; Daggett, V. Validating Molecular Dynamics Simulations against Experimental Observables in Light of Underlying Conformational Ensembles. *J. Phys. Chem. B* **2018**, *122* (26), 6673–6689. <https://doi.org/10.1021/acs.jpcc.8b02144>.
- (56) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. <https://doi.org/10.1021/jp973084f>.
- (57) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. <https://doi.org/10.1002/jcc.23354>.
- (58) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197. <https://doi.org/10.1021/ja00124a002>.
- (59) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package: Amber Biomolecular Simulation Package. *WIREs Comput Mol Sci* **2013**, *3* (2), 198–210. <https://doi.org/10.1002/wcms.1121>.
- (60) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676. <https://doi.org/10.1002/jcc.20090>.
- (61) Parra, M.; Stahl, S.; Hellmann, H. Vitamin B6 and Its Role in Cell Metabolism and Physiology. *Cells* **2018**, *7* (7), 84. <https://doi.org/10.3390/cells7070084>.

- (62) Salvo, M. L.; Safo, M. K.; Contestabile, R. Biomedical Aspects of Pyridoxal 5'-Phosphate Availability. *Front Biosci (Elite Ed)* **2012**, *4*, 897–913.
- (63) Ghatge, M. S.; Contestabile, R.; Salvo, M. L.; Desai, J. V.; Gandhi, A. K.; Camara, C. M.; Florio, R.; González, I. N.; Parroni, A.; Schirch, V.; Safo, M. K. Pyridoxal 5'-Phosphate Is a Slow Tight Binding Inhibitor of E. Coli Pyridoxal Kinase. *PLoS ONE* **2012**, *7* (7), e41680. <https://doi.org/10.1371/journal.pone.0041680>.
- (64) di Salvo, M. L.; Contestabile, R.; Safo, M. K. Vitamin B6 Salvage Enzymes: Mechanism, Structure and Regulation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011**, *1814* (11), 1597–1608. <https://doi.org/10.1016/j.bbapap.2010.12.006>.
- (65) Ghatge, M. S.; Al Mughram, M.; Omar, A. M.; Safo, M. K. Inborn Errors in the Vitamin B6 Salvage Enzymes Associated with Neonatal Epileptic Encephalopathy and Other Pathologies. *Biochimie* **2021**, *183*, 18–29. <https://doi.org/10.1016/j.biochi.2020.12.025>.
- (66) Mehta, P. K.; Christen, P. The Molecular Evolution of Pyridoxal-5'-Phosphate-Dependent Enzymes. In *Advances in Enzymology - and Related Areas of Molecular Biology*; Purich, D. L., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006; pp 129–184. <https://doi.org/10.1002/9780470123201.ch4>.
- (67) Christen, P.; Mehta, P. K. From Cofactor to Enzymes. The Molecular Evolution of Pyridoxal-5'-Phosphate-Dependent Enzymes. *Chem. Record* **2001**, *1* (6), 436–447. <https://doi.org/10.1002/tcr.10005>.
- (68) Percudani, R.; Peracchi, A. The B6 Database: A Tool for the Description and Classification of Vitamin B6-Dependent Enzymatic Activities and of the Corresponding Protein Families. *BMC Bioinformatics* **2009**, *10*, 273. <https://doi.org/10.1186/1471-2105-10-273>.
- (69) Eliot, A. C.; Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **2004**, *73* (1), 383–415. <https://doi.org/10.1146/annurev.biochem.73.011303.074021>.
- (70) Rocha, J. F.; Pina, A. F.; Sousa, S. F.; Cerqueira, N. M. F. S. A. PLP-Dependent Enzymes as Important Biocatalysts for the Pharmaceutical, Chemical and Food Industries: A Structural and Mechanistic Perspective. *Catal. Sci. Technol.* **2019**, *9* (18), 4864–4876. <https://doi.org/10.1039/C9CY01210A>.
- (71) Coburn, S. P.; Mahuren, J. D.; Schaltenbrand, W. E. Increased Activity of Pyridoxal Kinase in Tongue in Down's Syndrome. *J Ment Defic Res* **1991**, *35* (Pt 6), 543–547.
- (72) Song, H.; Ueno, S.; Numata, S.; Iga, J.; Shibuya-Tayoshi, S.; Nakataki, M.; Tayoshi, S.; Yamauchi, K.; Sumitani, S.; Tomotake, T.; Tada, T.; Tanahashi, T.; Itakura, M.; Ohmori, T. Association between PNPO and Schizophrenia in the Japanese Population. *Schizophr. Res.* **2007**, *97* (1–3), 264–270. <https://doi.org/10.1016/j.schres.2007.08.004>.
- (73) Nishino, N.; Fujiwara, H.; Noguchi-Kuno, S. A.; Tanaka, C. GABAA Receptor but Not Muscarinic Receptor Density Was Decreased in the Brain of Patients with Parkinson's Disease. *Jpn. J. Pharmacol.* **1988**, *48* (3), 331–339.
- (74) Aoyagi, T.; Wada, T.; Nagai, M.; Kojima, F.; Harada, S.; Takeuchi, T.; Takahashi, H.; Hirokawa, K.; Tsumita, T. Increased Gamma-Aminobutyrate Aminotransferase Activity in Brain of Patients with Alzheimer's Disease. *Chem. Pharm. Bull.* **1990**, *38* (6), 1748–1749.
- (75) Herrmann, W.; Lorenzl, S.; Obeid, R. Review of the role of hyperhomocysteinemia and B-vitamin deficiency in neurological and psychiatric disorders--current evidence and preliminary recommendations. *Fortschr Neurol Psychiatr* **2007**, *75* (9), 515–527. <https://doi.org/10.1055/s-2007-980112>.
- (76) Hassel, B.; Rogne, A. G.; Hope, S. Intellectual Disability Associated With Pyridoxine-Responsive Epilepsies: The Need to Protect Cognitive Development. *Front. Psychiatry* **2019**, *10*, 116. <https://doi.org/10.3389/fpsy.2019.00116>.

- (77) Safo, M. K.; Mathews, I.; Musayev, F. N.; Salvo, M. L.; Thiel, D. J.; Abraham, D. J.; Schirch, V. X-Ray Structure of Escherichia Coli Pyridoxine 5'-Phosphate Oxidase Complexed with FMN at 1.8 Å Resolution. *Structure* **2000**, *8* (7), 751–762.
- (78) Ghatge, M. S.; Karve, S. S.; David, T. M. S.; Ahmed, M. H.; Musayev, F. N.; Cunningham, K.; Schirch, V.; Safo, M. K. Inactive Mutants of Human Pyridoxine 5'-Phosphate Oxidase: A Possible Role for a Noncatalytic Pyridoxal 5'-Phosphate Tight Binding Site. *FEBS Open Bio* **2016**, *6* (5), 398–408. <https://doi.org/10.1002/2211-5463.12042>.
- (79) Musayev, F. N.; Di Salvo, M. L.; Saavedra, M. A.; Contestabile, R.; Ghatge, M. S.; Haynes, A.; Schirch, V.; Safo, M. K. Molecular Basis of Reduced Pyridoxine 5'-Phosphate Oxidase Catalytic Activity in Neonatal Epileptic Encephalopathy Disorder. *Journal of Biological Chemistry* **2009**, *284* (45), 30949–30956. <https://doi.org/10.1074/jbc.M109.038372>.
- (80) Mills, P. B.; Surtees, R. A. H.; Champion, M. P.; Beesley, C. E.; Dalton, N.; Scambler, P. J.; Heales, S. J. R.; Briddon, A.; Scheimberg, I.; Hoffmann, G. F.; Zschocke, J.; Clayton, P. T. Neonatal Epileptic Encephalopathy Caused by Mutations in the PNPO Gene Encoding Pyridox(Am)lne 5'-Phosphate Oxidase. *Human Molecular Genetics* **2005**, *14* (8), 1077–1086. <https://doi.org/10.1093/hmg/ddi120>.
- (81) Mills, P. B.; Camuzeaux, S. S. M.; Footitt, E. J.; Mills, K. A.; Gissen, P.; Fisher, L.; Das, K. B.; Varadkar, S. M.; Zuberi, S.; McWilliam, R.; Stodberg, T.; Plecko, B.; Baumgartner, M. R.; Maier, O.; Calvert, S.; Riney, K.; Wolf, N. I.; Livingston, J. H.; Bala, P.; Morel, C. F.; Feillet, F.; Raimondi, F.; Del Giudice, E.; Chong, W. K.; Pitt, M.; Clayton, P. T. Epilepsy Due to PNPO Mutations: Genotype, Environment and Treatment Affect Presentation and Outcome. *Brain* **2014**, *137* (5), 1350–1360. <https://doi.org/10.1093/brain/awu051>.
- (82) Barile, A.; Nogués, I.; di Salvo, M. L.; Bunik, V.; Contestabile, R.; Tramonti, A. Molecular Characterization of Pyridoxine 5'-Phosphate Oxidase and Its Pathogenic Forms Associated with Neonatal Epileptic Encephalopathy. *Sci Rep* **2020**, *10* (1), 13621. <https://doi.org/10.1038/s41598-020-70598-7>.
- (83) Alghamdi, M.; Bashiri, F. A.; Abdelhakim, M.; Adly, N.; Jamjoom, D. Z.; Sumaily, K. M.; Alghanem, B.; Arold, S. T. Phenotypic and Molecular Spectrum of Pyridoxamine-5'-phosphate Oxidase Deficiency: A Scoping Review of 87 Cases of Pyridoxamine-5'-phosphate Oxidase Deficiency. *Clinical Genetics* **2021**, *99* (1), 99–110. <https://doi.org/10.1111/cge.13843>.
- (84) Hoffmann, G. F.; Schmitt, B.; Windfuhr, M.; Wagner, N.; Strehl, H.; Bagci, S.; Franz, A. R.; Mills, P. B.; Clayton, P. T.; Baumgartner, M. R.; Steinmann, B.; Bast, T.; Wolf, N. I.; Zschocke, J. Pyridoxal 5'-Phosphate May Be Curative in Early-Onset Epileptic Encephalopathy. *J. Inher. Metab. Dis.* **2007**, *30* (1), 96–99. <https://doi.org/10.1007/s10545-006-0508-4>.
- (85) Bertoldi, M. Mammalian Dopa Decarboxylase: Structure, Catalytic Activity and Inhibition. *Archives of Biochemistry and Biophysics* **2014**, *546*, 1–7. <https://doi.org/10.1016/j.abb.2013.12.020>.
- (86) Opladen, T.; Hoffmann, G. F. Disorders of Monoamine Metabolism. In *Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases*; Blau, N., Dionisi Vici, C., Ferreira, C. R., Vianey-Saban, C., van Karnebeek, C. D. M., Eds.; Springer International Publishing: Cham, 2022; pp 313–329. https://doi.org/10.1007/978-3-030-67727-5_19.
- (87) Meiser, J.; Weindl, D.; Hiller, K. Complexity of Dopamine Metabolism. *Cell Commun Signal* **2013**, *11* (1), 34. <https://doi.org/10.1186/1478-811X-11-34>.
- (88) Kow, R. L.; Sikkema, C.; Wheeler, J. M.; Wilkinson, C. W.; Kraemer, B. C. DOPA Decarboxylase Modulates Tau Toxicity. *Biological Psychiatry* **2018**, *83* (5), 438–446. <https://doi.org/10.1016/j.biopsych.2017.06.007>.

- (89) Cellini, B. Biochemical and Computational Approaches to Improve the Clinical Treatment of Dopa Decarboxylase-Related Diseases: An Overview. *TOBIOCJ* **2012**, 6 (1), 131–138. <https://doi.org/10.2174/1874091X01206010131>.
- (90) Devos, D.; Lejeune, S.; Cormier-Dequaire, F.; Tahiri, K.; Charbonnier-Beaupel, F.; Rouaix, N.; Duhamel, A.; Sablonnière, B.; Bonnet, A.-M.; Bonnet, C.; Zahr, N.; Costentin, J.; Vidailhet, M.; Corvol, J.-C. Dopa-Decarboxylase Gene Polymorphisms Affect the Motor Response to L-Dopa in Parkinson's Disease. *Parkinsonism & Related Disorders* **2014**, 20 (2), 170–175. <https://doi.org/10.1016/j.parkreldis.2013.10.017>.
- (91) Wassenberg, T.; Molero-Luis, M.; Jeltsch, K.; Hoffmann, G. F.; Assmann, B.; Blau, N.; Garcia-Cazorla, A.; Artuch, R.; Pons, R.; Pearson, T. S.; Leuzzi, V.; Mastrangelo, M.; Pearl, P. L.; Lee, W. T.; Kurian, M. A.; Heales, S.; Flint, L.; Verbeek, M.; Willemsen, M.; Opladen, T. Consensus Guideline for the Diagnosis and Treatment of Aromatic L-Amino Acid Decarboxylase (AADC) Deficiency. *Orphanet J Rare Dis* **2017**, 12 (1), 12. <https://doi.org/10.1186/s13023-016-0522-z>.
- (92) Pearson, T. S.; Gilbert, L.; Opladen, T.; Garcia-Cazorla, A.; Mastrangelo, M.; Leuzzi, V.; Tay, S. K. H.; Sykut-Cegielska, J.; Pons, R.; Mercimek-Andrews, S.; Kato, M.; Lücke, T.; Oppebøen, M.; Kurian, M. A.; Steel, D.; Manti, F.; Meeks, K. D.; Jeltsch, K.; Flint, L. AADC Deficiency from Infancy to Adulthood: Symptoms and Developmental Outcome in an International Cohort of 63 Patients. *Jrnl of Inher Metab Disea* **2020**, 43 (5), 1121–1130. <https://doi.org/10.1002/jimd.12247>.
- (93) Pearson, T. S.; Gupta, N.; San Sebastian, W.; Imamura-Ching, J.; Viehoveer, A.; Grijalvo-Perez, A.; Fay, A. J.; Seth, N.; Lundy, S. M.; Seo, Y.; Pampaloni, M.; Hyland, K.; Smith, E.; de Oliveira Barbosa, G.; Heathcock, J. C.; Minnema, A.; Lonser, R.; Elder, J. B.; Leonard, J.; Larson, P.; Bankiewicz, K. S. Gene Therapy for Aromatic L-Amino Acid Decarboxylase Deficiency by MR-Guided Direct Delivery of AAV2-AADC to Midbrain Dopaminergic Neurons. *Nat Commun* **2021**, 12 (1), 4251. <https://doi.org/10.1038/s41467-021-24524-8>.
- (94) Wen, Y.; Wang, J.; Zhang, Q.; Chen, Y.; Bao, X. The Genetic and Clinical Characteristics of Aromatic L-Amino Acid Decarboxylase Deficiency in Mainland China. *J Hum Genet* **2020**, 65 (9), 759–769. <https://doi.org/10.1038/s10038-020-0770-6>.
- (95) Montioli, R.; Borri Voltattorni, C. Aromatic Amino Acid Decarboxylase Deficiency: The Added Value of Biochemistry. *IJMS* **2021**, 22 (6), 3146. <https://doi.org/10.3390/ijms22063146>.
- (96) Montioli, R.; Cellini, B.; Borri Voltattorni, C. Molecular Insights into the Pathogenicity of Variants Associated with the Aromatic Amino Acid Decarboxylase Deficiency. *J Inherit Metab Dis* **2011**, 34 (6), 1213–1224. <https://doi.org/10.1007/s10545-011-9340-6>.
- (97) Montioli, R.; Dindo, M.; Giorgetti, A.; Piccoli, S.; Cellini, B.; Voltattorni, C. B. A Comprehensive Picture of the Mutations Associated with Aromatic Amino Acid Decarboxylase Deficiency: From Molecular Mechanisms to Therapy Implications. *Human Molecular Genetics* **2014**, 23 (20), 5429–5440. <https://doi.org/10.1093/hmg/ddu266>.
- (98) Wieland, T.; Bodanszky, M. *The World of Peptides: A Brief History of Peptide Chemistry*; Springer-Verlag: Berlin ; New York, 1991.
- (99) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. Conformation of Amino Acid Side-Chains in Proteins. *Journal of Molecular Biology* **1978**, 125 (3), 357–386. [https://doi.org/10.1016/0022-2836\(78\)90408-4](https://doi.org/10.1016/0022-2836(78)90408-4).
- (100) Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins Application to Side-Chain Prediction. *Journal of Molecular Biology* **1993**, 230 (2), 543–574. <https://doi.org/10.1006/jmbi.1993.1170>.

- (101) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins* **2000**, *40* (3), 389–408. [https://doi.org/10.1002/1097-0134\(20000815\)40:3<389::AID-PROT50>3.0.CO;2-2](https://doi.org/10.1002/1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2).
- (102) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved Prediction of Protein Side-Chain Conformations with SCWRL4: Side-Chain Prediction with SCWRL4. *Proteins* **2009**, *77* (4), 778–795. <https://doi.org/10.1002/prot.22488>.
- (103) Miao, Z.; Cao, Y. Quantifying Side-Chain Conformational Variations in Protein Structure. *Sci Rep* **2016**, *6* (1), 37024. <https://doi.org/10.1038/srep37024>.
- (104) Zhang, J.; Liu, J. S. On Side-Chain Conformational Entropy of Proteins. *PLoS Comput Biol* **2006**, *2* (12), e168. <https://doi.org/10.1371/journal.pcbi.0020168>.
- (105) Onuchic, J. N.; Wolynes, P. G. Theory of Protein Folding. *Current Opinion in Structural Biology* **2004**, *14* (1), 70–75. <https://doi.org/10.1016/j.sbi.2004.01.009>.
- (106) Li, B.; Fooksa, M.; Heinze, S.; Meiler, J. Finding the Needle in the Haystack: Towards Solving the Protein-Folding Problem Computationally. *Critical Reviews in Biochemistry and Molecular Biology* **2018**, *53* (1), 1–28. <https://doi.org/10.1080/10409238.2017.1380596>.
- (107) Miyazawa, S.; Jernigan, R. L. Residue – Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *Journal of Molecular Biology* **1996**, *256* (3), 623–644. <https://doi.org/10.1006/jmbi.1996.0114>.
- (108) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338* (6110), 1042–1046. <https://doi.org/10.1126/science.1219021>.
- (109) Trinquier, G.; Sanejouand, Y. H. Which Effective Property of Amino Acids Is Best Preserved by the Genetic Code? *Protein Engineering Design and Selection* **1998**, *11* (3), 153–169. <https://doi.org/10.1093/protein/11.3.153>.
- (110) Chakrabarti, P.; Bhattacharyya, R. Geometry of Nonbonded Interactions Involving Planar Groups in Proteins. *Progress in Biophysics and Molecular Biology* **2007**, *95* (1–3), 83–137. <https://doi.org/10.1016/j.pbiomolbio.2007.03.016>.
- (111) Pace, C. J.; Gao, J. Exploring and Exploiting Polar- π Interactions with Fluorinated Aromatic Amino Acids. *Acc. Chem. Res.* **2013**, *46* (4), 907–915. <https://doi.org/10.1021/ar300086n>.
- (112) Makwana, K. M.; Mahalakshmi, R. Implications of Aromatic-Aromatic Interactions: From Protein Structures to Peptide Models: Aromatic Interactions in Protein Structures. *Protein Science* **2015**, *24* (12), 1920–1933. <https://doi.org/10.1002/pro.2814>.
- (113) Baker, C. M.; Grant, G. H. Role of Aromatic Amino Acids in Protein-Nucleic Acid Recognition. *Biopolymers* **2007**, *85* (5–6), 456–470. <https://doi.org/10.1002/bip.20682>.
- (114) Wilson, K. A.; Wetmore, S. D. A Survey of DNA-Protein π -Interactions: A Comparison of Natural Occurrences and Structures, and Computationally Predicted Structures and Strengths, Noncovalent Forces. In Scheiner, S., Ed., “Noncovalent Forces.” *Springer International Publishing Switzerland*, **2015**, 501–532.
- (115) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 946–954. <https://doi.org/10.1021/ar300024d>.
- (116) Dougherty, D. A. Cation- π Interactions Involving Aromatic Amino Acids. *The Journal of Nutrition* **2007**, *137* (6), 1504S-1508S. <https://doi.org/10.1093/jn/137.6.1504S>.
- (117) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine Binding to Brain Receptors Requires a Strong Cation- π Interaction. *Nature* **2009**, *458* (7237), 534–537. <https://doi.org/10.1038/nature07768>.

- (118) Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chem. Int. Ed.* **2011**, *50* (21), 4808–4842. <https://doi.org/10.1002/anie.201007560>.
- (119) Ferreira de Freitas, R.; Schapira, M. A Systematic Analysis of Atomic Protein–Ligand Interactions in the PDB. *Med. Chem. Commun.* **2017**, *8* (10), 1970–1981. <https://doi.org/10.1039/C7MD00381A>.
- (120) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi Contacts Are an Overlooked Protein Feature Relevant to Phase Separation. *eLife* **2018**, *7*, e31486. <https://doi.org/10.7554/eLife.31486>.
- (121) Hu, J.-S.; Grzesiek, S.; Bax, A. Two-Dimensional NMR Methods for Determining χ_1 Angles of Aromatic Residues in Proteins from Three-Bond JCC γ and JNC γ Couplings. *J. Am. Chem. Soc.* **1997**, *119* (7), 1803–1804. <https://doi.org/10.1021/ja963625z>.
- (122) Lanzarotti, E.; Biekofsky, R. R.; Estrin, D. A.; Marti, M. A.; Turjanski, A. G. Aromatic–Aromatic Interactions in Proteins: Beyond the Dimer. *J. Chem. Inf. Model.* **2011**, *51* (7), 1623–1633. <https://doi.org/10.1021/ci200062e>.
- (123) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* **1985**, *229* (4708), 23–28. <https://doi.org/10.1126/science.3892686>.
- (124) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. Amino/Aromatic Interactions in Proteins: Is the Evidence Stacked Against Hydrogen Bonding? *Journal of Molecular Biology* **1994**, *239* (2), 315–331. <https://doi.org/10.1006/jmbi.1994.1370>.
- (125) Prajapati, R. S.; Sirajuddin, M.; Durani, V.; Sreeramulu, S.; Varadarajan, R. Contribution of Cation– π Interactions to Protein Stability. *Biochemistry* **2006**, *45* (50), 15000–15010. <https://doi.org/10.1021/bi061275f>.
- (126) Dougherty, D. A. The Cation– π Interaction. *Acc. Chem. Res.* **2013**, *46* (4), 885–893. <https://doi.org/10.1021/ar300265y>.
- (127) Kennedy, C. R.; Lin, S.; Jacobsen, E. N. The Cation– π Interaction in Small-Molecule Catalysis. *Angew. Chem. Int. Ed.* **2016**, *55* (41), 12596–12624. <https://doi.org/10.1002/anie.201600547>.
- (128) Liang, Z.; Li, Q. X. π –Cation Interactions in Molecular Recognition: Perspectives on Pharmaceuticals and Pesticides. *J. Agric. Food Chem.* **2018**, *66* (13), 3315–3323. <https://doi.org/10.1021/acs.jafc.8b00758>.
- (129) Rodham, D. A.; Suzuki, S.; Suenram, R. D.; Lovas, F. J.; Dasgupta, S.; Goddard, W. A.; Blake, G. A. Hydrogen Bonding in the Benzene–Ammonia Dimer. *Nature* **1993**, *362* (6422), 735–737. <https://doi.org/10.1038/362735a0>.
- (130) Ma, J. C.; Dougherty, D. A. The Cation– π Interaction. *Chem. Rev.* **1997**, *97* (5), 1303–1324. <https://doi.org/10.1021/cr9603744>.
- (131) Feller, D.; Dixon, D. A.; Nicholas, J. B. Binding Enthalpies for Alkali Cation–Benzene Complexes Revisited. *J. Phys. Chem. A* **2000**, *104* (48), 11414–11419. <https://doi.org/10.1021/jp002631l>.
- (132) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. The Magnitude of the CH/ π Interaction between Benzene and Some Model Hydrocarbons. *J. Am. Chem. Soc.* **2000**, *122* (15), 3746–3753. <https://doi.org/10.1021/ja993972j>.
- (133) Xu, Y.; Shen, J.; Zhu, W.; Luo, X.; Chen, K.; Jiang, H. Influence of the Water Molecule on Cation– π Interaction: Ab Initio Second Order Møller–Plesset Perturbation Theory (MP2) Calculations. *J. Phys. Chem. B* **2005**, *109* (12), 5945–5949. <https://doi.org/10.1021/jp044568w>.

- (134) Marshall, M. S.; Steele, R. P.; Thanthiriwatte, K. S.; Sherrill, C. D. Potential Energy Curves for Cation- π Interactions: Off-Axis Configurations Are Also Attractive. *J. Phys. Chem. A* **2009**, *113* (48), 13628–13632. <https://doi.org/10.1021/jp906086x>.
- (135) Armentrout, P. B.; Rodgers, M. T. An Absolute Sodium Cation Affinity Scale: Threshold Collision-Induced Dissociation Experiments and Ab Initio Theory. *J. Phys. Chem. A* **2000**, *104* (11), 2238–2247. <https://doi.org/10.1021/jp991716n>.
- (136) Amicangelo, J. C.; Armentrout, P. B. Absolute Binding Energies of Alkali-Metal Cation Complexes with Benzene Determined by Threshold Collision-Induced Dissociation Experiments and Ab Initio Theory. *J. Phys. Chem. A* **2000**, *104* (48), 11420–11432. <https://doi.org/10.1021/jp002652f>.
- (137) Ruan, C.; Rodgers, M. T. Cation- π Interactions: Structures and Energetics of Complexation of Na⁺ and K⁺ with the Aromatic Amino Acids, Phenylalanine, Tyrosine, and Tryptophan. *J. Am. Chem. Soc.* **2004**, *126* (44), 14600–14610. <https://doi.org/10.1021/ja048297e>.
- (138) Xie, N.-Z.; Du, Q.-S.; Li, J.-X.; Huang, R.-B. Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design. *PLoS ONE* **2015**, *10* (9), e0137113. <https://doi.org/10.1371/journal.pone.0137113>.
- (139) Orabi, E. A.; Lamoureux, G. Cation- π and π - π Interactions in Aqueous Solution Studied Using Polarizable Potential Models. *J. Chem. Theory Comput.* **2012**, *8* (1), 182–193. <https://doi.org/10.1021/ct200569x>.
- (140) Orabi, E. A.; Lamoureux, G. Cation- π Interactions between Quaternary Ammonium Ions and Amino Acid Aromatic Groups in Aqueous Solution. *J. Phys. Chem. B* **2018**, *122* (8), 2251–2260. <https://doi.org/10.1021/acs.jpcc.7b11983>.
- (141) Zhu, D.; Herbert, B. E.; Schlautman, M. A.; Carraway, E. R. Characterization of Cation- π Interactions in Aqueous Solution Using Deuterium Nuclear Magnetic Resonance Spectroscopy. *J. environ. qual.* **2004**, *33* (1), 276–284. <https://doi.org/10.2134/jeq2004.2760>.
- (142) Slutsky, M. M.; Marsh, E. N. G. Cation- π Interactions Studied in a Model Coiled-Coil Peptide. *Protein Sci.* **2004**, *13* (8), 2244–2251. <https://doi.org/10.1110/ps.04702104>.
- (143) Riemen, A. J.; Waters, M. L. Design of Highly Stabilized β -Hairpin Peptides through Cation- π Interactions of Lysine and *N*-Methyllysine with an Aromatic Pocket. *Biochemistry* **2009**, *48* (7), 1525–1531. <https://doi.org/10.1021/bi801706k>.
- (144) Sarkar, A.; Kellogg, G. Hydrophobicity - Shake Flasks, Protein Folding and Drug Discovery. *CTMC* **2010**, *10* (1), 67–83. <https://doi.org/10.2174/156802610790232233>.
- (145) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Computer-Aided Mol Des* **1991**, *5* (6), 545–552. <https://doi.org/10.1007/BF00135313>.
- (146) Ahmed, M. H.; Koparde, V. N.; Safo, M. K.; Neel Scarsdale, J.; Kellogg, G. E. 3d Interaction Homology: The Structurally Known Rotamers of Tyrosine Derive from a Surprisingly Limited Set of Information-Rich Hydropathic Interaction Environments Described by Maps: Tyrosine Hydropathic Interaction Environments Define Its Rotamers. *Proteins* **2015**, *83* (6), 1118–1136. <https://doi.org/10.1002/prot.24813>.
- (147) Ahmed, M. H.; Catalano, C.; Portillo, S. C.; Safo, M. K.; Neel Scarsdale, J.; Kellogg, G. E. 3D Interaction Homology: The Hydropathic Interaction Environments of Even Alanine Are Diverse and Provide Novel Structural Insight. *Journal of Structural Biology* **2019**, *207* (2), 183–198. <https://doi.org/10.1016/j.jsb.2019.05.007>.
- (148) Catalano, C.; AL Mughram, M. H.; Guo, Y.; Kellogg, G. E. 3D Interaction Homology: Hydropathic Interaction Environments of Serine and Cysteine Are Strikingly Different and Their

- Roles Adapt in Membrane Proteins. *Current Research in Structural Biology* **2021**, *3*, 239–256. <https://doi.org/10.1016/j.crstbi.2021.09.002>.
- (149) Herrington, N. B.; Kellogg, G. E. 3D Interaction Homology: Computational Titration of Aspartic Acid, Glutamic Acid and Histidine Can Create PH-Tunable Hydropathic Environment Maps. *Front. Mol. Biosci.* **2021**, *8*, 773385. <https://doi.org/10.3389/fmolb.2021.773385>.
- (150) AL Mughram, M. H.; Catalano, C.; Bowry, J. P.; Safo, M. K.; Scarsdale, J. N.; Kellogg, G. E. 3D Interaction Homology: Hydropathic Analyses of the “ π -Cation” and “ π - π ” Interaction Motifs in Phenylalanine, Tyrosine, and Tryptophan Residues. *J. Chem. Inf. Model.* **2021**, *61* (6), 2937–2956. <https://doi.org/10.1021/acs.jcim.1c00235>.
- (151) AL Mughram, M. H.; Herrington, N. B.; Catalano, C.; Kellogg, G. E. Systematized Analysis of Secondary Structure Dependence of Key Structural Features of Residues in Soluble and Membrane-Bound Proteins. *Journal of Structural Biology: X* **2021**, *5*, 100055. <https://doi.org/10.1016/j.yjsbx.2021.100055>.
- (152) Deng, Y.; Gam, J.; French, J. B.; Zhao, H.; An, S.; Benkovic, S. J. Mapping Protein-Protein Proximity in the Purinosome. *J. Biol. Chem.* **2012**, *287* (43), 36201–36207. <https://doi.org/10.1074/jbc.M112.407056>.
- (153) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54* (1). <https://doi.org/10.1002/cpbi.3>.
- (154) Land, H.; Humble, M. S. YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. In *Protein Engineering*; Bornscheuer, U. T., Höhne, M., Eds.; *Methods in Molecular Biology*; Springer New York: New York, NY, **2018**, 43–67. https://doi.org/10.1007/978-1-4939-7366-8_4.
- (155) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* **2018**, *46*, 296–303. <https://doi.org/10.1093/nar/gky427>.
- (156) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Research* **2022**, *50* (D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
- (157) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373* (6557), 871–876. <https://doi.org/10.1126/science.abj8754>.

CHAPTER 2*

3D Hydrophobic Interaction Maps as a New Motif for Describing the Role of Aromatic Amino Acid Residues in Protein Structure

*This chapter is adapted from "3D Interaction Homology: Hydrophobic Analyses of the π -Cation" and π - π Interaction Motifs in Phenylalanine, Tyrosine, and Tryptophan Residues" by AL Mughram M. H. et al., published in *J. Chem. Inf. Model.* 2021.

2.1.1 Introduction

The composition and character of each of the twenty amino acids is the determining factor in forming proteins, which contain a broad variety of forms and folds, and produce a phenomenal array of functions.^{1,2} This diverse collection of amino acid residues shares the same backbone, but each of the different sidechains vary in size, shape, polarity, and relative hydrophobicity. This variety in sidechain gives each unique roles and functional characteristics, with which they can participate in many different interactions that result in "proper" protein folding and function.³⁻⁵ Three amino acids: phenylalanine (PHE), tyrosine (TYR), and tryptophan (TRP) are referred to as "aromatic amino acids" as they possess an aromatic moiety in their sidechains. Altogether, these amino acids represent 8.5% of those observed in proteins (PHE, 3.9%), (TYR, 3.3%), and (TRP, 1.3%).⁶ Their unique sidechain characters, with their conjugated planar rings, enable them to engage in multiple noncovalent hydrophobic and polar interactions, both as purely hydrophobic moieties, but also in leveraging their aromatic character, such as π - π , π -CH, π -OH, π -NH, π -SH, π -cation, π -anion interactions, etc.^{7,8} Thus, while their

frequency of observation is a little more than half of average, their unique interaction properties, and with them being present in both buried and surface positions, have made PHE, TYR and TRP the subject of numerous protein structure and functions studies.^{9–15} In general, aromatic amino acids are hydrophobic in nature and tend to be more often buried in protein cores than exposed to solvent. This hydrophobic nature and their packing preferences have been shown to assist in early folding steps and support the stability of protein cores.^{16,17} Peptide models were also used to investigate aromatic sidechain contributions to protein stability and have afforded a better understanding of the fundamental driving forces of aromatic sidechain residues in protein folding, function and molecular recognition.^{10,15,18,19}

Of particular interest, in contrast to simple hydrophobic amino acids, alanine (ALA), isoleucine (ILE), leucine (LEU), and valine (VAL), aromatic sidechain residues are also capable of participating in polar interactions as well: the surrounding H-donor residues and/or water molecules can make π -polar interactions with aromatic sidechains. In this sense, many published reports have been dedicated to a better understanding of the occurrence and magnitude of π -cation interactions in many aspects including protein structure,^{7,20,21} molecular recognition and catalysis.^{21–23} Modeling and dissecting the π -cation interaction in terms of its origin and strength was the core subject of numerous computational^{24–30} and experimental^{31–33} studies. The quantum mechanical calculations of π -cation interactions have been found to be correlated with experimental studies in the gas phase. The latter is important because the “real” biological environment of proteins is substantially different and more complex than the gas phase, so the translation of quantitative estimates of QM calculated energetics– i.e., largely

performed in the gas phase— to proteins is challenging. Nevertheless, the existence of π -cation interactions in aqueous solutions have been computationally^{34–36} and experimentally^{20,21,37} acknowledged. In addition, peptide models used to examine the stabilizing forces of π -cation interactions concluded that they provided no substantial contribution to peptide stabilization and that stability was most likely due to hydrophobic effects.^{38,39} Clearly, however, the understanding of the structural and functional contributions and roles of hydrophobic/polar interactions for aromatic amino acid sidechains in biological systems remains a topic of interest. Furthermore, in the current era of accelerating interest in *de novo* protein design and engineering that impacts multiple fields with, for example, enzyme variants with improved catalytic activities, more accessible physiochemical attributes, altered substrate specificities and/or and stereoselectivities, it will be valuable to understand the underlying blueprint of all aspects of protein structure.

The general roles and magnitudes of noncovalent interactions as they effect biomolecular assemblies and associations have been studied since the first structures of proteins were defined. While each amino acid residue in each protein is unique, and their inter-residue interactions determine the structure—from the sidechain rotamer to the protein's fold—there are commonalities in terms of interaction types, *loci* and strengths. Recently, our group utilized 3D hydrophobic interaction maps to probe the hydrophobic environments surrounding amino acids within proteins.^{40–45} This approach relies on calculating the 3D “hydrophobic interaction” maps of the residue of interest with respect to all other elements within structures. Perhaps obvious, our hypothesis is that unique constellations of interactions are comparatively few in number and conserved for each

residue type and secondary structure. Importantly, our analysis describes residues by contextualizing the hydrophobic environments of the surrounding atoms and environment. For instance, a properly oriented HOH, a residue or a ligand can provide the same ‘acidic’ character as a TYR –OH and serine (SER) –OH or even a LYS –NH₃⁺ to satisfy its “hydrophobic valence”. Thus, protein structure is not driven by sequence, but instead by the hydrophobic interactions that each residue makes in terms of its sidechain and backbone conformations.

2.1.2 Hypotheses and research plan

In this chapter, we analyzed the hydrophobic environments of the three aromatic residues derived from the same relatively large dataset of X-ray crystallographic protein structures. The local secondary structure for each of the extracted residue was assigned based on its backbone conformation (**Figure 2.1**). Furthermore, the buriedness was described using calculated solvent-accessible surface area for each of the extracted residues. In previous analyses, each cluster-derived set of extracted residues was shown to have a mostly uniform SASA, suggesting that buriedness is also an environmental factor defining hydrophobic environments and conformation. The framework for our analysis in this work is hydrophobic interactions, i.e., energetics based on the relative solubility of an “atom” with respect to water and 1-octanol. We have shown that this simple free energy-based thermodynamic parameter is very revealing of structure in the biological environment. In that regard, interactions are defined as four types: favorable polar (e.g., hydrogen bond or acid-base), unfavorable polar (acid-acid

or base-base), favorable hydrophobic (hydrophobic-hydrophobic) and unfavorable hydrophobic (hydrophobic-polar or desolvation).

One of our interests in this chapter is to determine if the hydropathic environment maps for the aromatic residues also encode the π -cation and π - π types of interactions and to assess the strength and prevalence of these over our fairly large data set. Since the amount of data analyzed during this work is so extensive, the discussion and presentation items within this article will concentrate on only a few chosen segments of the Ramachandran "chessboard" schema^{40,46} (**Figure 2.1**). To address the questions raised above, this chapter will primarily investigate the following:

- 1- 3D similarity and clustering of the hydropathic environments of the three aromatic amino acid residues using a large dataset of experimental protein structures.
- 2- Sampling non-covalent interactions above and below the sidechain ring space to evaluate the magnitude and occurrence of key non-covalent interactions such as π -cation, π - π interactions.

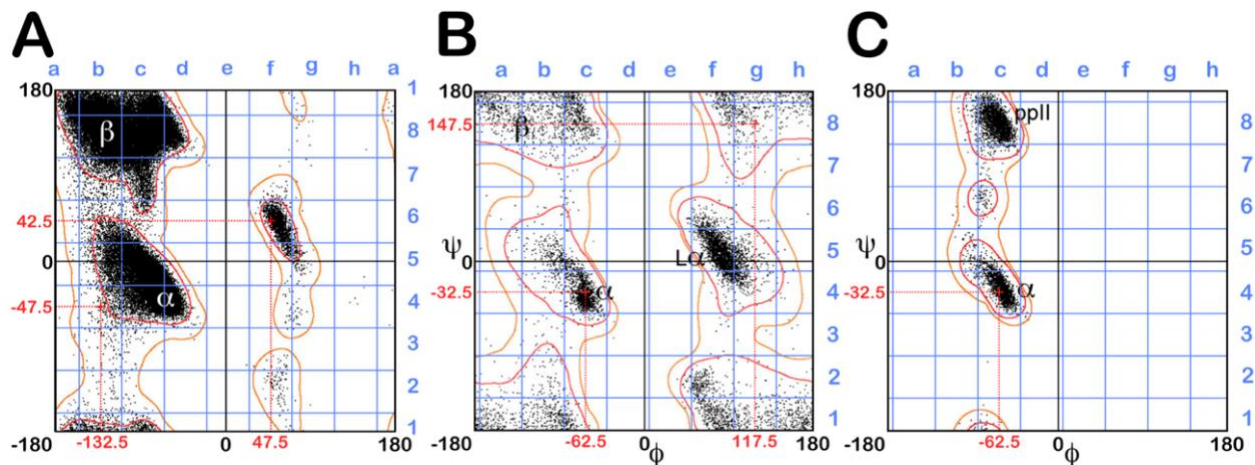


Figure 2.1 Ramachandran (ϕ versus ψ) plots with superimposed chess square schema

in blue. A) For all residues (except glycine and proline); B) glycine; and C) proline. Two centroids of each plot with red dotted lines are illustrated, e.g., b4 ($\phi = -132.5$, $\psi = -47.5$), g8 ($\phi = 117.5$, $\psi = 147.5$), and c4 ($\phi = -62.5$, $\psi = -32.5$) for 1A, 1B, and 1C, respectively. The three aromatic residues involved in this chapter (PHE, TYR, and TRP) were investigated using

Schema A.

2.2 Methods

2.2.1 Dataset. Our dataset consists of 2703 randomly chosen proteins from the RCSB Protein Data Bank,⁴⁷ each of which has no ligand or cofactor in their original structures. Our selection criteria were defined and described earlier.⁴⁰ This protein structure collection was chosen to emphasize several simultaneous factors, e.g., population-based sampling of fold and chain identities (whether or not proteins with very similar or even identical sequences were included in the data set), and resolution-blind selection to ensure that unique structures are included – not just the easier to crystallize and analyze high-resolution structures. The dataset size should make sure that all or

nearly all possible environments of PHE, TYR, and TRP are represented. By assigning initial positions of hydrogen atoms to their heavy atoms, i.e., to which they are covalently bonded, with idealized geometry, all-atom models were created and optimized via conjugate gradient minimization.

2.2.2 Alignment calculations. We recast the standard Ramachandran ϕ vs ψ plot^{40,46} as 45° by 45° ($\pi/4$ by $\pi/4$) chess squares in the $\phi - \psi$ space. The boundaries for the squares in ϕ and ψ were shifted by -20 and -25° , respectively, so that the highest density regions would fall more squarely within matrix squares. These were labeled as **a1** through **h8** (see **Figure 2.1**). The ϕ , ψ , and ω torsion angles for each nonterminal PHE, TYR, and TRP residue within the proteins of the dataset were calculated, and every residue was then binned into the suitable chess square based on its ϕ and ψ . We then sub-binned (or “parsed”) each collection of residues in the chess square by their χ_1 (N-CA-CB-CG) angle, group A ($0^\circ \leq \chi_1 < 120^\circ$), group B ($120^\circ \leq \chi_1 < 240^\circ$), and group C ($240^\circ \leq \chi_1 < 360^\circ$), which will be referred to as the “.60”, “.180”, and “.300” parses, respectively (**Table 2.2**). Complete listings of all PHE, TYR, and TRP residues in the dataset, their ϕ , ψ , and ω torsions, and also the backbone atom numbers in their respective PDB files, are accessible as Supporting Information in the published version.⁴² Additionally, these data are key to identify specific PHE, TYR, and TRP residues in the dataset. For simplicity in notation, this work will commonly just allude to the PHE’s, TYR’s, and TRP’s sequence numbers within the chess square/parse.

A model residue (ALA) template was constructed at the centroid of each chess

square, such that it had ϕ and ψ angles for that centroid. In terms of orientation, CA was defined to be at the origin and CB was on the z-axis, and therefore, the CA–HA bond vector is in the $-y$, $-z$ quadrant of the yz-plane. As each aromatic residue was transformed to align on the centroid template, the rotation and translation matrices were calculated by least squares fitting of the C_{-1} , N, CA, CB, C, O, N_{+1} , and CA_{+1} atoms to the template. In other words, the whole protein was frame-shifted such that these atoms of the aromatic residue superimposed onto the centroid template of its associated chess square.

2.2.3 HINT basis interaction maps. Interatomic interactions were evaluated using the HINT force field and score model^{48,49} that uses atom-centered parameters, a_i and S_i , the hydrophobic atom constant-derived from experimental determinations of $\log P_{o/w}$, the partition constant for 1-octanol and water and the SASA, for atom i . Generally, $a_i > 0$ for a hydrophobic atom and $a_i < 0$ for a polar atom. S_i is larger for solvent-exposed frontier atoms but near zero for atoms at the center of functional groups. The score between atoms i and j is

$$b_{ij} = a_i S_i a_j S_j T_{ij} e^{-r} + L_{ij},$$

where r is the distance (\AA) between atoms i and j , T_{ij} has values of -1 , 0 and 1 to account for the acid, base, etc. character of the interacting atoms and thus properly sign their interaction scores, and L_{ij} is an implementation of the Lennard-Jones potential function.⁴⁸ By convention, $b_{ij} > 0$ indicates favorable interactions, e.g., hydrophobic-hydrophobic or

Lewis acid-Lewis base, and $b_{ij} < 0$ indicates unfavorable interactions, e.g., hydrophobic-polar or Lewis acid-Lewis acid.

A three-dimensional box with an additional 5 Å buffer margin at all sides, with the CA atom at the origin, was constructed to explore and calculate a map. The overall box dimensions are: for PHE, $-10.0 \text{ \AA} \leq x \leq 10.0 \text{ \AA}$; $-10.0 \text{ \AA} \leq y \leq 10.0 \text{ \AA}$; $-7.5 \text{ \AA} \leq z \leq 10.5 \text{ \AA}$, (62,197 points, 7200 \AA^3); for TYR, $-11.0 \text{ \AA} \leq x \leq 11.0 \text{ \AA}$; $-11.0 \text{ \AA} \leq y \leq 11.0 \text{ \AA}$; $-7.5 \text{ \AA} \leq z \leq 10.5 \text{ \AA}$, (74,925 points, 8712 \AA^3); and for TRP, $-11.0 \text{ \AA} \leq x \leq 11.0 \text{ \AA}$; $-11.0 \text{ \AA} \leq y \leq 11.0 \text{ \AA}$; $-7.5 \text{ \AA} \leq z \leq 12.5 \text{ \AA}$, (83,025 points, 9680 \AA^3); all with a point spacing of 0.5 Å. The HINT basis interaction map formalism⁴⁰ was used to calculate 3D interaction grids that represent the 3D interaction environment associated for a residue with respect to its environment. Briefly, the HINT pairwise interaction list,^{48,49} i.e., b_{ij} s, were translated into a 3D map object encoding interaction type and strength and also the loci of those interactions in 3D space. We computed the value for every grid point within the map using:

$$\rho_{xyz} = \sum b_{ij} \exp \{ -[(x - x_{ij})^2 + (y - y_{ij})^2 + (z - z_{ij})^2] / \sigma \},$$

where ρ_{xyz} is the map value at a given grid point (x, y, z) , x_{ij} , y_{ij} and z_{ij} are the midpoint coordinates of the vector between atoms i and j , and σ is the width of the Gaussian map peak, for which we are using 0.5.⁴⁰ As indicated in the above equation, sums were computed for every grid point over all pairs of interactions; separate maps for PHE, TYR, and TRP sidechain atoms for each of the four general map-types (favorable and unfavorable, polar and hydrophobic) were calculated.

2.2.4 Calculation of map-map correlation metrics. The grid maps, \mathbf{m} and \mathbf{n} , are 3D arrays of real-valued points. First, each raw map datum (G_t , where t is the index for map points) was transformed into \log_{10} space while holding its sign with $(G_t/|G_t|)$; each map point value was also standardized such that 1 was a predefined floor value.

$$\text{if } |G_t|/F > 1.0, A_t = (G_t/|G_t|) \log_{10} (|G_t|/F); \text{ else, } A_t = 0.$$

The similarity, $D(\mathbf{m}, \mathbf{n})$, between pairs of grid maps was quantitated with the previously described correlation coefficient-based metric,⁴⁰ i.e.,

$$D(\mathbf{m}, \mathbf{n}) = \sum \{ 1 - (|A_t(\mathbf{m}) - A_t(\mathbf{n})|)^2 / [(|A_t(\mathbf{m})| + |A_t(\mathbf{n})|) \cdot (|A(\mathbf{m})|_{\max} + |A(\mathbf{n})|_{\max})] \}.$$

In this case, $A_t(\mathbf{m})$ and $A_t(\mathbf{n})$ are values for the same points in map \mathbf{m} and map \mathbf{n} , respectively, and $|A|_{\max}$ is the absolute value of the maximum for the map's grid points. Overall, $D(\mathbf{m}, \mathbf{n})$ ranges between 0 and 1, where two identical maps display a value of 1. Our map boxes are intended to encompass all possible environments and always include a large proportion (> 60%) of zero-valued points, which if taken into account would bias $D(\mathbf{m}, \mathbf{n})$ such that all map pairs would seem very similar. We thus only use points where $|A_t(\mathbf{m})| \geq 8 |A(\mathbf{m})_{\text{stddev}}|$ or $|A_t(\mathbf{n})| \geq 8 |A(\mathbf{n})_{\text{stddev}}|$ (A_{stddev} is the standard deviation of the average value of all map points) in calculating $D(\mathbf{m}, \mathbf{n})$.⁵⁶ Normally, a correlation metric such as $D(\mathbf{m}, \mathbf{n})$ should range from 0 and 1; however, pairs of maps which are exact mirror images of each other, i.e., with exactly the same patterns and values but with the opposite sign, which would be required for $D(\mathbf{m}, \mathbf{n}) = 0$, clearly are not possible. Also, even if the two maps were completely nonoverlapping, D would be ~ 0.5 . Therefore, the

minimum D is more on the order of 0.6. For maps encoding the four individual classes of interactions for the residue sidechain, map-map correlation scores were calculated separately. We calculated overall map-map correlations between two structures/environments by weighted averages of these four terms,

$$D(\mathbf{m},\mathbf{n})_{all} = \{ 4[D(\mathbf{m},\mathbf{n})_{hydro(+)}] + 2[D(\mathbf{m},\mathbf{n})_{hydro(-)}] + [D(\mathbf{m},\mathbf{n})_{polar(+)}] + [D(\mathbf{m},\mathbf{n})_{polar(-)}] \} / 8.$$

As we expected that hydrophobic contributions to be the most diverse and information-rich, favorable and unfavorable hydrophobic contributions are scaled by 4 and 2, respectively.⁵⁶ Map calculations and map-map correlation scores were computed on our GPU-based compute server with locally coded programs, including work-reduction algorithms such as a first-pass filter to expedite the calculation of matrix components when two maps are identical or near identical.

2.2.5 Clustering and validation. For cluster analysis, we used the freely accessible R software programming language and environment.⁵⁰ We operated on the pairwise Euclidean distances similarity matrices between vectors of the calculated map correlation coefficients as above using the k-means method⁴⁰ with largely default values for all of the clustering method parameters.

Because the k-means clustering algorithm does not determine the optimum number of clusters for matrices of map $D_{(\mathbf{m},\mathbf{n})}$ values, this critical parameter is user-specified. With experience gained through previous reports of our lab and further preliminary studies, we settled on a maximum number of cluster values that allowed for both significant map diversity and inter-chess square/inter-residue type comparisons. For the three residue types in this study, we set the maximum number of clusters to 12.

Most chess square/parses, however, because they are less populated, had fewer than 12 clusters in their optimal solutions (**Table 2.2**). Another feature of k-means clustering is that it will not form singleton clusters, i.e., if a single member of the set is so unique that no other member will join it, that member is left out of the clustering solution. While this is fairly rare (~5%), such maps are potentially interesting, and our protocols recover them by reconstructing the cluster solutions with the missing singletons. Any chess square/parse with four or less maps was not clustered, but instead, averaged, to create what is, in effect, a 1 cluster case. Rarely, a set with five or more maps failed to cluster, in which case it was also treated as a 1 cluster case.

2.2.6 Average map, RMSD and solvent-accessible surface area calculations. Calculation of average maps requires consideration of a variety of subtle variables. Importantly, to avoid what we call “brown mapping”,⁴⁰ only maps that have shown strong similarity to each other should be merged. Average maps for cluster sets are constructed by Gaussian weighting of each map’s contribution with respect to its Euclidian distance from each cluster’s centroid to more favorably account for maps closer to the center of a cluster. We borrowed the term “exemplar” from affinity propagation clustering to describe the datum closest to the centroid of each cluster generated by the k-means algorithm. For the PHE, TYR, and TRP structures, RMSDs (root-mean square distances) were calculated by first averaging, using the same weighting as above, the atomic positions of all residues within a cluster set to establish an average residue model for the set, followed by computing the RMSD from that average model for each atom (by name); the atomic values were then averaged to obtain

the reported residue RMSD for the set. With the GETAREA algorithm⁵¹ default settings, we calculated the solvent-accessible surface areas (SASAs) for all PHE, TYR, and TRP sidechains. The protein coordinates were used as originally reported in the PDB files and only the first conformation was used if the PDB file residue records reported more than one.

2.2.7 Ring hydrophobic environment calculations. The map contents of cylinders above and below the rings of phenylalanine, tyrosine and tryptophan were collected to determine the types of interactions between the rings and their environments. First, the centroid of each ring system was constructed from CG, CD1, CD2, CE1, CE2, and CZ (for PHE and TYR) and CG, CD1, CD2, NE1, CE2, CE3, CZ2, CH2 and CZ3 (for TRP). Next, trigonal prisms were constructed perpendicular to triangles in the ring system as defined in **Table 2.1**, with heights of 4.0 Å (*vide infra*). These triangles extended to the protons attached to each ring carbon (except for CG, where CB defined the corresponding triangle vertices). Also, to avoid the hydrogen bonding influence of tyrosine's hydroxyl group, the two triangles and associated prisms involving that group were omitted. Lastly, the distances from the CA atom to the centroid's two normals (+4.0 Å and -4.0 Å) were calculated; the "side" of the ring with the longer distance was defined as "above" and the side with the shorter distance was defined as "below".

Table 2.1—Construction of Cylinders for PHE, TYR and TRP.

Residue	Prism	Atoms
PHE	1	Centroid-CB-HD1
	2	Centroid-HD1-HE1
	3	Centroid-HE1-HZ
	4	Centroid-HZ-HE2
	5	Centroid-HE2-HD2
	6	Centroid-HD2-CB
TYR	1	Centroid-CB-HD1
	2	Centroid-HD1-HE1
	3	Centroid-HE2-HD2
	4	Centroid-HD2-CB
TRP	1	Centroid-CB-HD1
	2	Centroid-HD1-HE1
	3	Centroid-HE1-HZ2
	4	Centroid-HZ2-HH2
	5	Centroid-HH2-HZ3
	6	Centroid-HZ3-HE3
	7	Centroid-HE3-CB

2.3 Results and Discussion

2.3.1 Dataset

We extracted 30,932 non-terminal phenylalanines, 28,886 non-terminal tyrosines and 11,573 non-terminal tryptophans using the dataset of 2703 protein structures from the RCSB Protein Data Bank documented and thoroughly described earlier.^{40,41} We overlaid a standard Ramachandran-type plot with an 8 by 8 chessboard (**Figure 2.1**) to control the effects on the structure of the ϕ and ψ angles of the protein backbone, where each chess square represents a 45° by 45° sample of $\phi - \psi$ space, with boundaries that have been slightly shifted to help focus the highest occupied regions inside rather than straddling individual squares. Furthermore, we parsed the residues in each chess square by their χ_1 into angles three groups corresponding to those normally observed in rotamer libraries:⁵²⁻⁵⁵ a group around 60° , a group around 180° , and a group near 300° . Our group showed previously that map-based clustering was able to identify this low level of detail nearly flawlessly, except in cases of high solvent-accessible surface area –presumably arising in surface exposed residues where there were few if any interactions.^{40,41}

However, such failures were problematical in calculating averaged maps and residue coordinates. Furthermore, splitting the chess square members increased computational efficiency. (Many calculations scale as n^2 : $3 \times (n/3)^2 < n^2$).

The occupancies of the chess squares range from 0 to 6361 for PHE, 5376 for TYR and 2622 for TRP (**Table 2.2**). For PHE, a total of 32 (of 64) chess squares contain 10 or more residues and 116 chess square/parses (of 192) are occupied at all (**Table**

2.3 and **Figure 2.2**); for TYR, a total of 32 chess squares contain 10 or more residues, and 117 chess square/parses are occupied at all (**Table 2.3** and **Figure 2.3**); and for TRP, a total of 25 chess squares contain 10 or more residues and 100 chess square/parses are occupied at all (**Table 2.3** and **Figure 2.4**). Throughout this work, chess square names, e.g., ***d4***, will be given in bold italics. The χ_1 parses will be denoted by the suffixes ***.60***, ***.180*** and ***.300***. Also, we are using a numerical scheme to simplify nomenclature of residues in this article, wherein the sequential number of that residue is its name in its particular chess square/parse. Thus, PHE 100 in chess square ***a1.60*** is the 100th PHE listed for that chess square/parse in Supporting Information (**Table S1**) in the published version,⁴² within which its actual pdbid, chain and residue name can be found.

In the Ramachandran plot, four regions have been associated with secondary structure motifs. In our schema, fifteen chess squares (***a1***, ***a6***, ***a7***, ***a8***, ***b1***, ***b2***, ***b7***, ***b8***, ***c1***, ***c2***, ***c6***, ***c7***, ***c8***, ***d1*** and ***d8***) are within the β -pleat motif, seven chess squares (***b4***, ***b5***, ***b6***, ***c4***, ***c5***, ***d4*** and ***d5***) are within the α -helix motif, five chess squares (***f5***, ***f6***, ***f7***, ***g5*** and ***g6***) correspond to the left-hand α -helix motif, and lastly, a single chess square (***f2***) is in a small and isolated region that would correspond to a more common “loop” motif (**Figure 2.2 – 2.4**). The other thirty-six chess squares include a handful that are probably mixed β -pleat motif and α -helix motifs. Although calculations have been made for all Ramachandran chess squares, in this dissertation we are concentrating the discussion on four here: three of the common secondary structure elements are being sampled –

β -pleat with **b1**, right-hand α -helix with **c5** and **d5** (comparing environmental sets within a secondary structure motif), and left-hand α -helix with **f6**.

Each residue was superimposed on a stub placed at the centroid of its respective chess square. The entire protein was oriented to superimpose the environments surrounding each residue in a common frame: the CA-CB vector for each is defined as the z-axis; hence the environmental features and variations of the sidechain would be a consequence of the residue's interactions and not likely to be due to backbones misalignments. The average root-mean square distances (RMSDs) for the backbone atom sets are generally around 0.15 Å, indicating a very small error with this alignment scheme, and that the 45° by 45° binning is adequate to maintain common sets of residue secondary structures (**Table 2.4**).

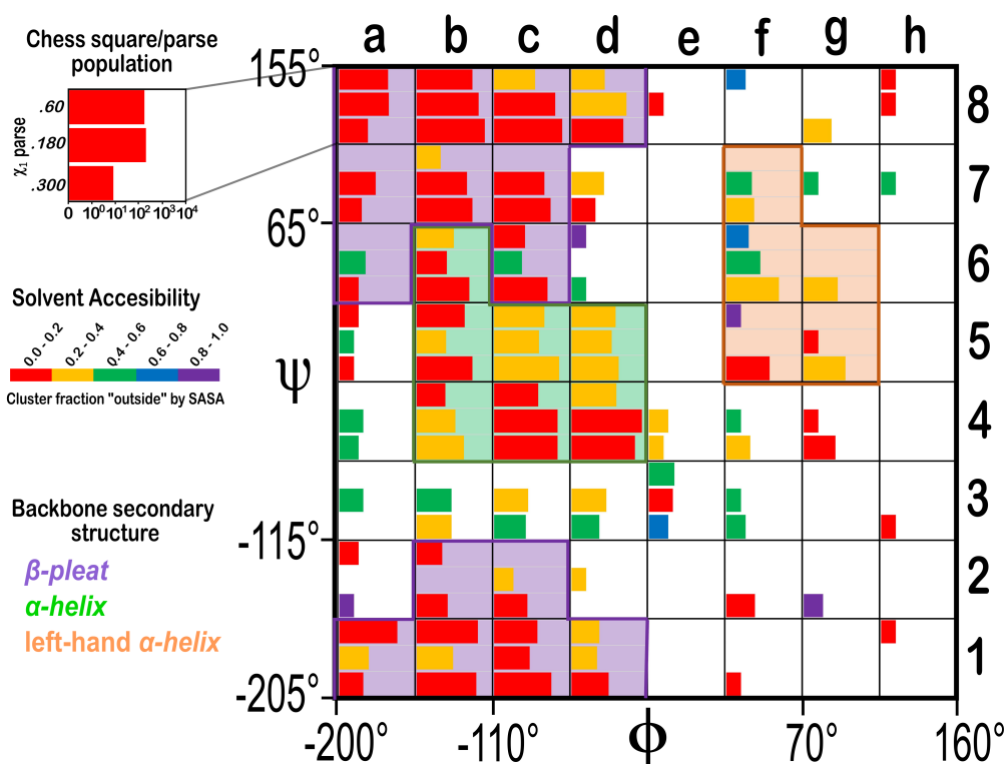


Figure 2.2 Ramachandran chessboard displaying the chess square/parse population for PHE residue. The Ramachandran ϕ vs ψ plot is rendered into sixty-four 45° by 45° ($\pi/4$ by $\pi/4$) chess squares. The (χ_1) parse populations are represented in the \log_{10} scale with the colored bars. Their colors reflect the average weighted fraction outside or solvent-exposed, that is, “ f_{outside} ” a measure of solvent accessibility (see text for definition). The ϕ vs ψ regions associated with β -pleat, α -helix, and left-hand α -helix secondary structure motifs are shaded in light purple, light green, and light orange, respectively.

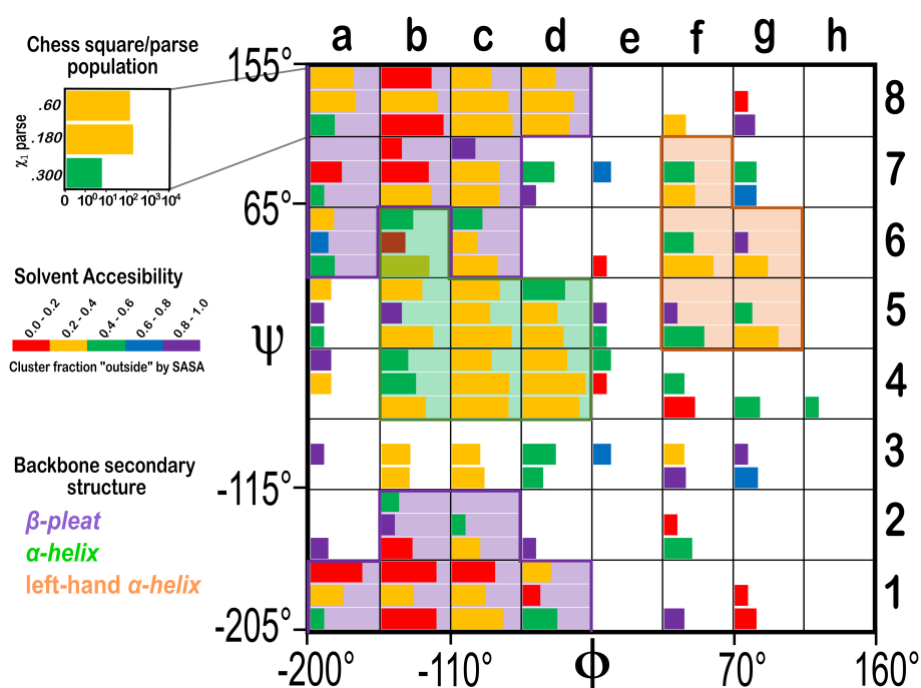


Figure 2.3 Ramachandran chessboard displaying the chess square/parse population for TYR residue. See caption for Figure 2.2.

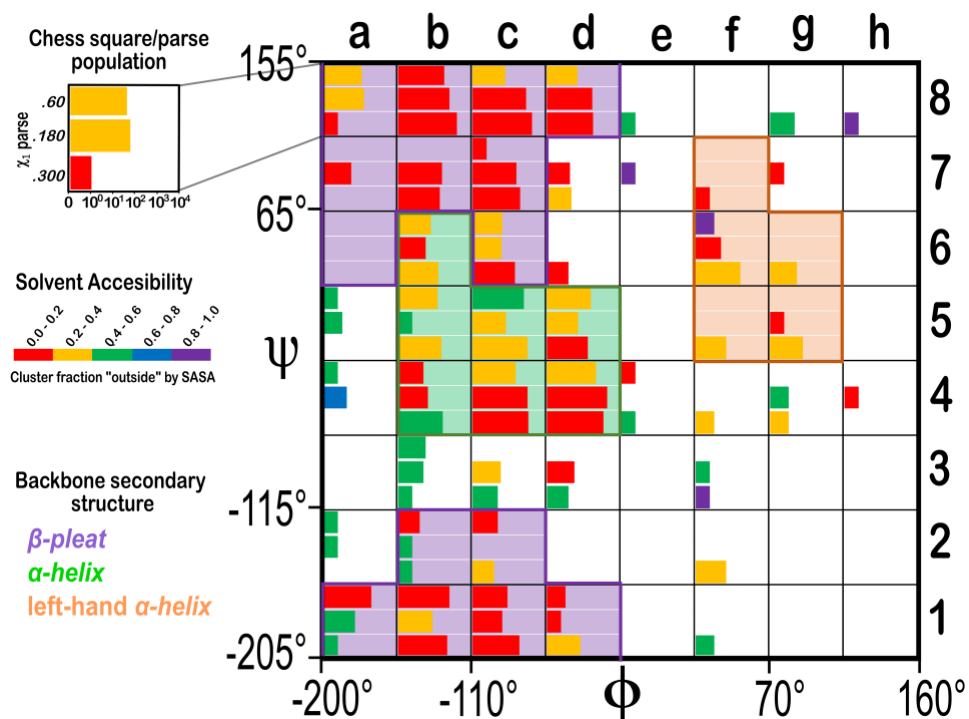


Figure 2.4 Ramachandran chessboard displaying the chess square/parse population for TRP residue. See caption for Figure 2.2.

Table 2.2– Residue populations of chess squares/parses for PHE, TYR, and TRP.

Number of phenylalanines in parses 60 / 180 / 300								
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
1	722/9/4	1217/27/905	76/23/624	7/5/31	0/0/0	0/0/1	0/0/0	1/0/0
2	2/0/1	5/0/11	0/2/17	0/1/0	0/0/0	0/0/8	0/0/2	0/0/0
3	0/4/0	0/22/20	1/18/13	0/20/7	5/4/2	0/1/2	0/0/0	0/0/1
4	0/4/2	8/36/146	85/1695/1656	93/4680/1589	0/2/1	0/1/4	0/1/13	0/0/0
5	2 /1/1	148/9/522	231/93/2085	83/50/143	0/0/0	1/0/80	0/1/63	0/0/0
6	0/6/2	29/10/333	12/7/343	1/0/1	0/0/0	3/18/326	0/0/19	0/0/0
7	0/28/3	4/222/518	0/212/564	0/14/4	0/0/0	0/5/7	0/1/0	0/1/0
8	179/198/8	488/1273/3355	51/1191/3357	17/478/292	0/1/0	2/0/0	0/2/7	1/1/0

Number of <u>tyrosines</u> in parses 60 / 180 /300								
	<i>a</i>	<i>b</i>	<i>C</i>	<i>D</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>H</i>
1	606/23/1	1111/24/1105	144/26/543	12/2/32	0/0/0	0/0/3	0/1/4	0/0/0
2	0/0/2	2/1/19	0/1/12	0/0/1	0/0/0	0/1/12	0/0/0	0/0/0
3	0/1/0	0/13/11	0/12/24	0/28/3	0/2/0	0/3/4	0/1/5	0/0/0
4	3/3/0	9/32/187	80/1425/1383	172/3949/1255	2/1/0	0/3/18	0/0/7	0/0/1
5	3/1/1	101/3/596	305/64/2080	120/35/95	0/1/1	0/1/84	0/2/155	0/0/0
6	5/2/6	21/6/318	17/7/208	0/0/0	0/0/1	0/14/399	0/1/25	0/0/0
7	0/19/1	3/269/438	5/279/301	0/20/1	0/2/0	0/17/19	0/4/4	0/0/0
8	136/190/6	483/1350/3468	81/1323/2601	26/567/261	0/0/0	0/0/4	0/1/3	0/0/0
Number of <u>tryptophans</u> in parses 60 / 180 /300								
	<i>a</i>	<i>b</i>	<i>C</i>	<i>D</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>H</i>
1	189/15/1	376/24/253	27/11/183	2/1/20	0/0/0	0/0/2	0/0/0	0/0/0
2	1/1/0	3/1/1	6/0/3	0/0/0	0/0/0	0/0/13	0/0/0	0/0/0
3	0/0/0	8/6/1	0/9/6	0/8/3	0/0/0	0/1/1	0/0/0	0/0/0
4	1/4/0	6/12/123	95/646/765	246/1506/870	0/1/1	0/0/2	0/2/2	0/1/0
5	1/2/0	54/1/99	351/22/627	112/14/71	0/0/0	0/0/13	0/1/19	0/0/0
6	0/0/0	19/8/62	11/10/89	0/0/3	0/0/0	2/6/129	0/0/7	0/0/0
7	0/8/0	0/114/74	1/108/191	0/4/5	0/1/0	0/0/1	0/1/0	0/0/0
8	41/64/1	160/371/1124	19/501/1264	13/142/160	0/1/1	0/0/0	0/0/5	0/0/1

Table 2.3– Number of PHE, TYR and TRP clusters by chess square and χ_1 parse.

Number of <u>phenylalanine</u> clusters in parses 60 / 180 /300								
	<i>a</i>	<i>B</i>	<i>c</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>g</i>	<i>H</i>
1	12/4/1	12/5/12	9/4/12	2/2/5	0/0/0	0/0/1	0/0/0	1/0/0
2	1/0/1	2/0/2	0/1/5	0/1/0	0/0/0	0/0/1	0/0/1	0/0/0
3	0/1/0	0/7/7	0/6/5	0/4/2	2/1/1	0/1/1	0/0/0	0/0/1

4	0/1/1	3/7/10	12/12/12	11/12/12	0/1/1	0/1/1	0/1/4	0/0/0
5	1/1/1	12/3/12	12/9/12	7/8/12	0/0/0	1/0/10	0/1/8	0/0/0
6	0/2/1	6/3/12	4/2/12	1/0/1	0/0/0	1/5/12	0/0/5	0/0/0
7	0/6/1	1/12/12	0/12/12	0/4/1	0/0/0	0/1/2	0/1/0	0/1/0
8	11/11/3	12/12/12	5/12/12	5/12/11	0/1/0	1/0/0	0/1/2	1/1/0
Number of tyrosine clusters in parses 60 / 180 / 300								
	a	b	c	d	E	F	g	H
1	12/4/1	12/5/12	9/5/12	3/1/7	0/0/0	0/0/1	0/1/1	0/0/0
2	0/0/1	1/1/4	0/1/4	0/0/1	0/0/0	0/1/3	0/0/0	0/0/0
3	0/1/0	0/4/3	0/3/5	0/5/1	0/1/0	0/1/1	0/1/2	0/0/0
4	1/1/0	3/7/12	8/12/12	12/12/12	1/1/0	0/1/4	0/0/2	0/0/1
5	1/1/1	12/1/12	12/9/12	11/7/9	0/1/1	0/1/10	0/1/8	0/0/0
6	1/1/2	4/2/12	5/2/12	0/0/0	0/0/1	1/4/12	0/1/5	0/0/0
7	0/5/1	1/12/12	2/12/12	0/6/1	0/1/0	0/3/5	0/1/1	0/0/0
8	11/10/2	12/12/12	9/12/12	6/12/12	0/0/0	0/0/1	0/0/2	0/0/0
Number of tryptophan clusters in parses 60 / 180 / 300								
	a	b	c	d	E	F	g	H
1	12/3/1	12/6/12	5/4/8	1/1/5	0/0/0	0/0/1	0/0/0	0/0/0
2	1/1/0	1/1/1	1/0/1	0/0/0	0/0/0	0/0/3	0/0/0	0/0/0
3	0/0/0	1/2/1	0/3/1	0/2/1	0/0/0	0/1/1	0/0/0	0/0/0
4	1/1/0	2/5/12	8/12/12	12/12/12	1/0/1	0/0/1	0/1/1	0/1/0
5	1/1/0	7/1/8	12/5/12	12/3/8	0/0/0	0/0/3	0/1/5	0/0/0
6	0/0/0	5/2/9	3/3/12	0/0/1	0/0/0	1/1/10	0/0/2	0/0/0
7	0/1/0	0/10/9	1/8/6	0/1/2	0/1/0	0/0/1	0/1/0	0/0/0
8	7/8/1	9/12/12	4/12/12	3/9/9	0/0/1	0/0/0	0/0/1	0/0/1

Table 2.4– Summary of results for clustering of maps in chess square/parse *b1.60* for PHE, TYR and TRP sidechains.

Residue	Cluster Set ^{a,b}	Member Count ^c	Composite Similarity ^d	RMSD (Å) ^e	1 ^f	Average sidechain SASA (Å ²) ^g	<i>foutside</i> (fraction outside) ^h	Fractional character in "cylinder" ⁱ			
								<i>Above ring</i> ^j		<i>Below ring</i> ^j	
								+Hydro ^k	+Polar ^l	+Hydro ^k	+Polar ^l
PHE	88	35	0.82572	0.559	63 ± 10	74 ± 49	0.5429	0.1035	0.4099	0.3737	0.3229
	148	89	0.85690	0.552	64 ± 9	31 ± 28	0.1932	0.3014	0.3262	0.6085	0.2164
	516	118	0.90518	0.476	61 ± 8	8 ± 19	0.0424	0.9207	0.0452	0.8028	0.0978
	553	116	0.85948	0.474	62 ± 9	7 ± 11	0.0129	0.9175	0.0517	0.5340	0.2410
	650	28	0.84215	0.579	65 ± 15	81 ± 38	0.6607	0.5348	0.2310	0.2191	0.3535
	725	165	0.90332	0.472	65 ± 6	1 ± 4	0.0000	0.9811	0.0159	0.7597	0.1234
	761	73	0.86562	0.571	64 ± 15	16 ± 16	0.0634	0.7951	0.0871	0.2381	0.3733
	825	63	0.85881	0.584	62 ± 11	33 ± 32	0.2143	0.8195	0.0898	0.7436	0.1299
	872	195	0.90175	0.528	65 ± 7	4 ± 14	0.0205	0.9214	0.0491	0.6136	0.2277
	1027	92	0.87258	0.516	62 ± 7	11 ± 11	0.0109	0.7542	0.1257	0.7758	0.1048
	1082	98	0.84830	0.564	61 ± 12	29 ± 26	0.2010	0.4666	0.2788	0.619	0.1965
	1192	145	0.90421	0.540	62 ± 8	4 ± 7	0.0069	0.9852	0.0082	0.822	0.0910
TYR	57	167	0.89272	0.227	68 ± 7	13 ± 13	0.0210	0.8424	0.0488	0.2903	0.3125
	166	9	0.88245	0.770	60 ± 23	119 ± 42	0.8333	0.3637	0.2183	0.3150	0.2767
	244	36	0.87764	0.328	52 ± 8	24 ± 23	0.1250	0.4656	0.2079	0.4754	0.2122
	392	79	0.85620	0.408	68 ± 12	37 ± 27	0.2278	0.7105	0.1157	0.3455	0.2894
	316	80	0.86962	0.337	54 ± 9	16 ± 18	0.0750	0.6554	0.1504	0.2707	0.3119
	337	78	0.87343	0.720	63 ± 7	25 ± 21	0.1346	0.8545	0.0462	0.2280	0.3526
	396	104	0.86374	0.284	64 ± 7	57 ± 29	0.4087	0.1477	0.3291	0.3310	0.2834
	403	47	0.85460	0.412	58 ± 11	88 ± 27	0.6915	0.3464	0.2500	0.4630	0.2479
	484	141	0.90495	0.262	60 ± 7	10 ± 14	0.0284	0.8303	0.0554	0.2943	0.3221
	581	151	0.88951	0.260	71 ± 7	15 ± 19	0.0662	0.7933	0.0678	0.2426	0.3042
	714	80	0.87024	0.312	66 ± 8	36 ± 21	0.2000	0.3844	0.2438	0.4082	0.2827
	914	139	0.90620	0.250	66 ± 6	21 ± 15	0.0540	0.5530	0.2190	0.4317	0.2816

TRP	23	54	0.91179	0.282	61 ± 7	21 ± 28	0.0833	0.6478	0.1748	0.2634	0.4213
	32	49	0.85943	0.617	54 ± 1	50 ± 45	0.2959	0.4833	0.2476	0.5628	0.2457
	41	30	0.91433	0.279	68 ± 6	51 ± 33	0.3333	0.1396	0.4619	0.6321	0.1834
	106	46	0.89061	0.419	62 ± 1	6 ± 10	0.0000	0.7974	0.0632	0.4796	0.2938
	193	15	0.95680	0.130	72 ± 3	62 ± 9	0.5000	0.0696	0.4390	0.3181	0.4132
	224	52	0.92056	0.238	57 ± 5	26 ± 28	0.0962	0.7202	0.1334	0.1216	0.4682
	236	4	0.95663	0.185	67 ± 2	63 ± 5	0.5000	0.3047	0.3336	0.3238	0.3700
	259	28	0.92735	0.928	63 ± 4	43 ± 53	0.1786	0.4767	0.2568	0.3889	0.3180
	273	23	0.91954	0.463	73 ± 9	20 ± 22	0.0435	0.6690	0.1679	0.5624	0.2148
	292	38	0.90149	0.335	60 ± 6	7 ± 12	0.0000	0.7127	0.1143	0.5086	0.2447
	327	17	0.97374	0.879	60 ± 0	11 ± 13	0.0294	0.3879	0.3359	0.5159	0.2333
	348	20	0.94817	0.253	66 ± 6	16 ± 30	0.0750	0.2362	0.3543	0.3004	0.3763

a Cluster sets are simply named for the exemplar residue, which in turn is named for its sequential position in a list ordered by pdbid and residue number. (See Supporting Material, Table S1).⁴² **b** Detailed lists of the cluster set memberships are set out in Tables S2-S4.⁴² **c** Counts listed are for all maps. **d** Similarities are calculated as described in the text. Here, the average map of all members of the cluster set (including those in the associated similarity sets) is compared to each individual in the set and the averages of these similarities are reported. **e** RMSD for heavy (non-hydrogen) set members based on clustering of the sidechain interaction maps. **f** average χ_1 of the residues in the cluster confirms that the pre-analysis parsing was appropriate. **g** Average solvent-accessible surface area for sidechain of atom set from GETAREA. **h** Fraction of residues outside (*foutside*) adapts the *i/o* criteria in GETAREA as follows: inside = 0.0, indeterminate = 0.5, outside = 1.0 (see text). **i** Cylinders as defined in text were superimposed on the map quartet and the map values of each thus enclosed grid point were summed [$\Sigma(-\text{hydro})$, $\Sigma(+\text{hydro})$, $\Sigma(-\text{polar})$, and $\Sigma(+\text{polar})$] and analyzed. **j** Above the ring is defined as the face furthest from the residue's CA carbon, while below is defined as the face closest. **k** $|\Sigma v_{\text{hydro}(+)})| / \{|\Sigma v_{\text{hydro}(-)}| + |\Sigma v_{\text{hydro}(+)})| + |\Sigma v_{\text{polar}(-)}| + |\Sigma v_{\text{polar}(+)})|\}$.

2.3.2 Clustering results in 3D maps. The results described in **Table 2.4** demonstrate that the cluster sets we generated for the *b1.60* chess squares of PHE, TYR and TRP are fairly robust, mathematically. However, our intent is to ascertain whether these clusters encode information-rich sets of interaction environments that illuminate protein structure and can be exploited.

2.3.2.1 Sidechain maps of phenylalanine. The contour plots of the average maps for PHE sidechain clusters of *b1*, *c5*, *d5*, *f6* for the **.60**, **.180** and **.300** parses are set out in **Figures 2.5 – 2.14**, respectively. Two views are shown for each map to help visualization: the left element of each pair is rotated such that the x-axis points to the right and the z-axis (the CA-CB bond) points up, while the second orientation (a rotation around the x-axis) brings the z-axis to the front. The maps are superimposed on the exemplar structure for the map. The contour levels chosen for all map pairs are identical to allow visual comparisons of relative interaction strengths: favorable polar (+12); unfavorable polar (−12); favorable hydrophobic (+3); and unfavorable hydrophobic (−6). In some maps, to illustrate the presence of weak polar interactions, translucent blue and red contours at ± 6 were also plotted. Noted on each map are the cluster name, the contribution of that cluster to the overall chess square/parse population and the average solvent accessible area (SASA, *S*) for that cluster.

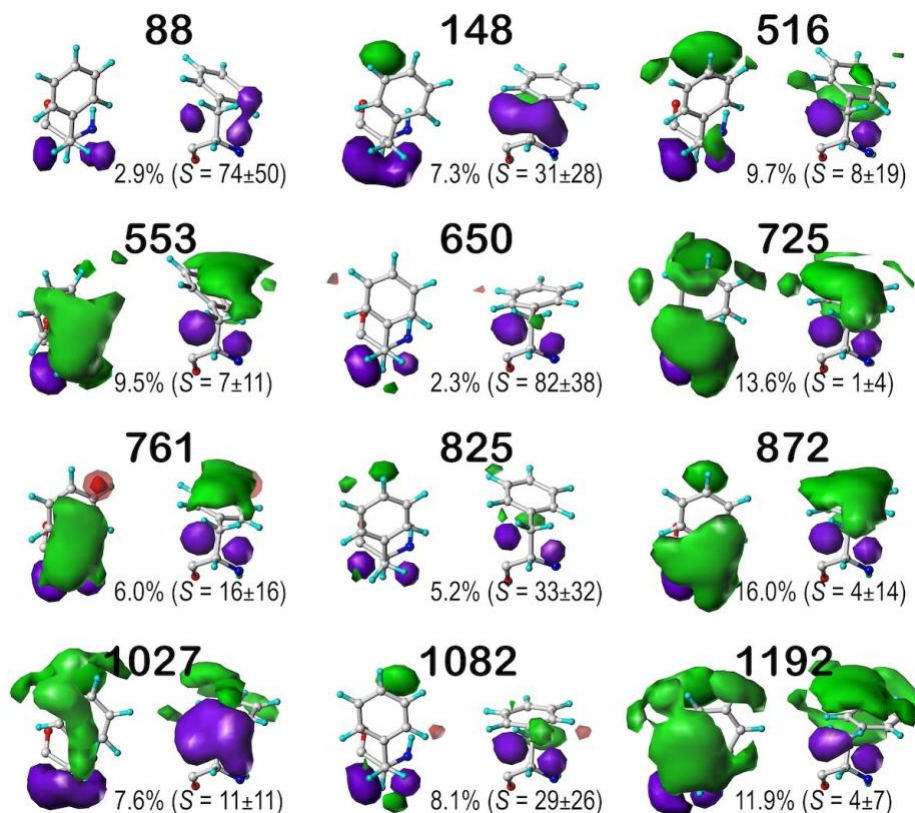


Figure 2.5 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *b1* chess square ($\chi_1 = 60^\circ$ parse). Two views are shown for each map: left) the CA-CB z-axis points up, right) the CA-CB axis points out of the page. The x-axis of both views point right, and the y-axis point back on the left and up on the right. The green contours represent favorable hydrophobic-hydrophobic interactions between the CB methyl, the sidechain, and the neighboring residues. Purple contours reflect unfavorable hydrophobic-polar interactions between the PHE sidechain and the neighboring residues. Translucent blue and red contours, when present, are plotted at one-half the map density of the solid corresponding contours

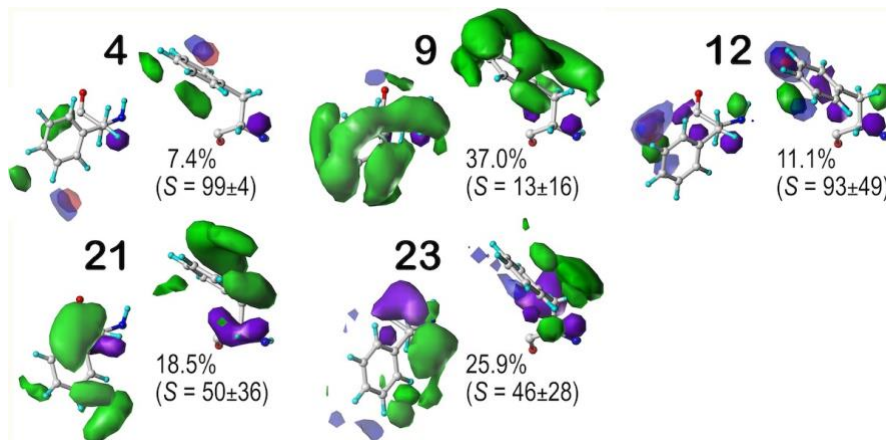


Figure 2.6 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *b1* chess square ($\chi_1 = 180^\circ$ parse).

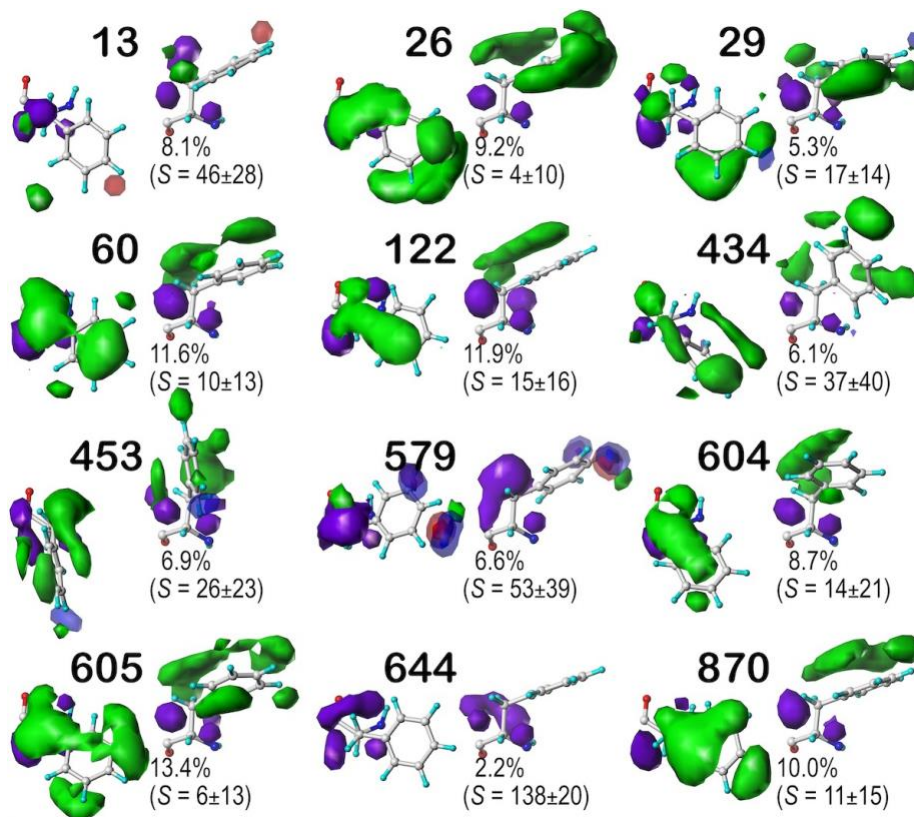


Figure 2.7 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *b1* chess square ($\chi_1 = 300^\circ$ parse).

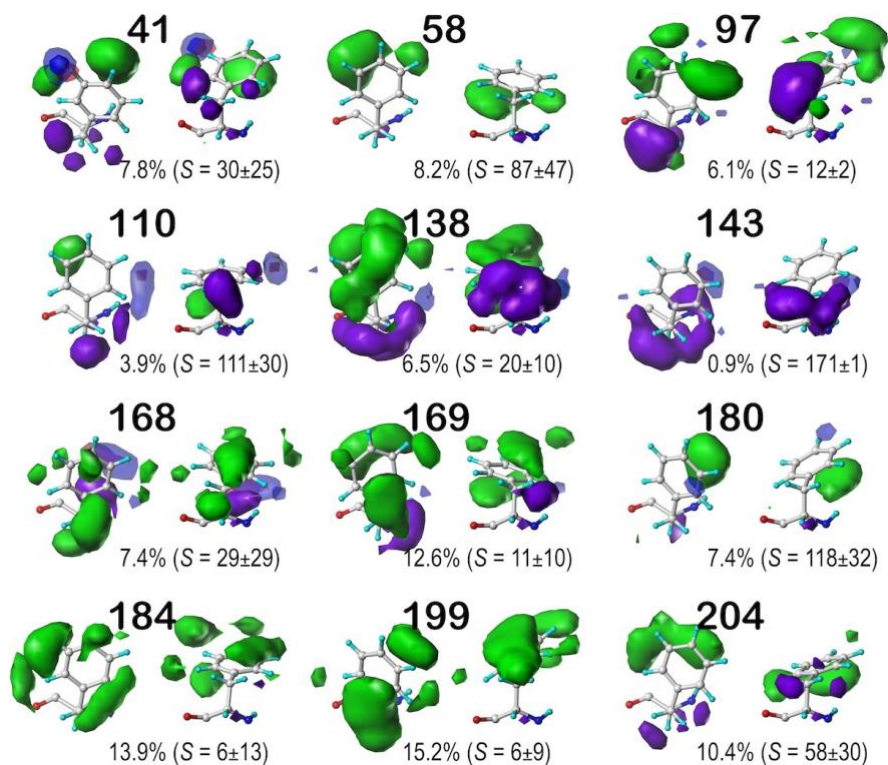


Figure 2.8 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the $c5$ chess square ($\chi_1 = 60^\circ$ parse).

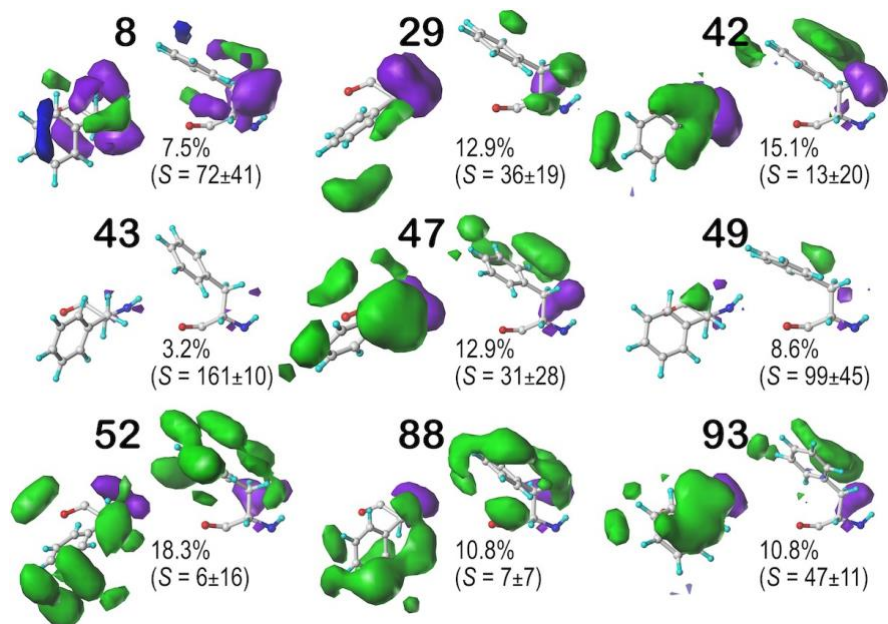


Figure 2.9 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the $c5$ chess square ($\chi_1 = 180^\circ$ parse).

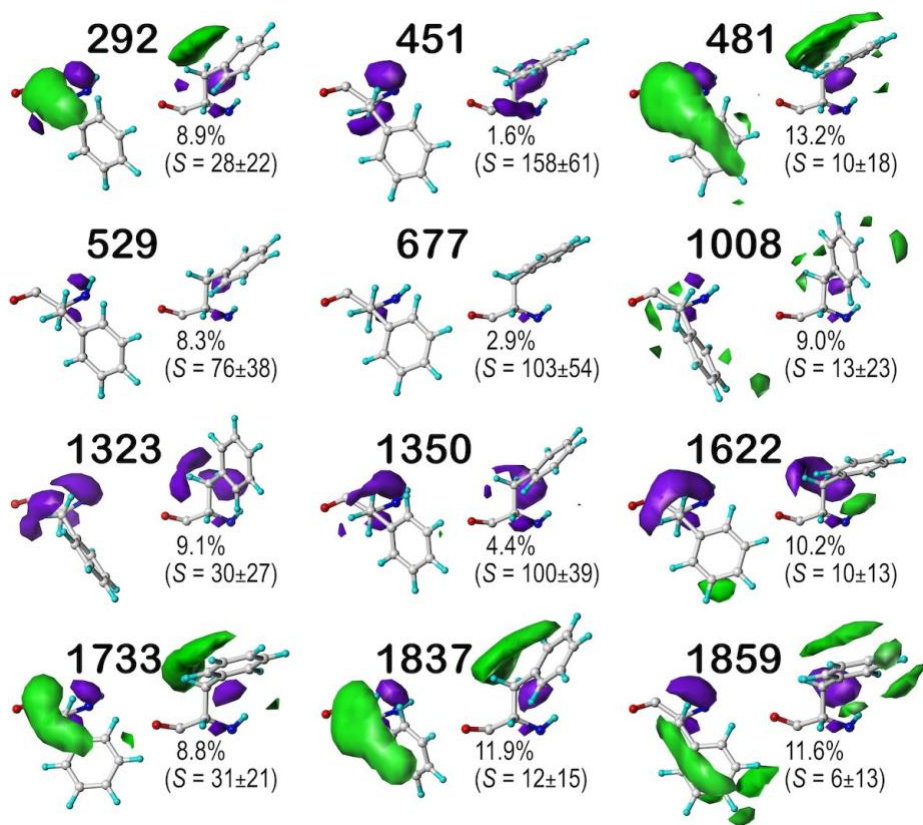


Figure 2.10 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *c5* chess square ($\chi_1 = 300^\circ$ parse).

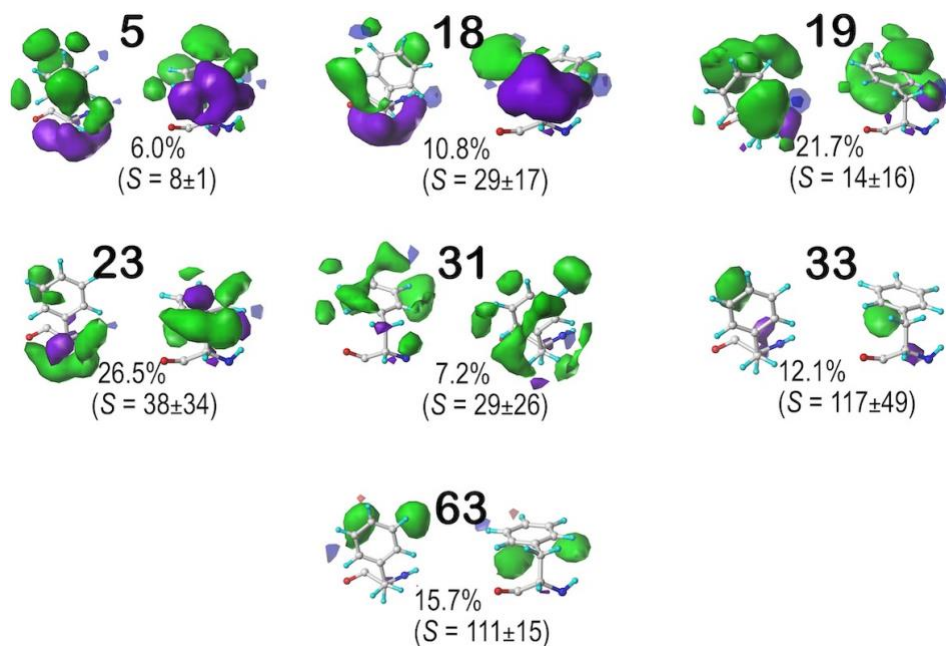


Figure 2.11 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *d5* chess square ($\chi_1 = 60^\circ$ parse).

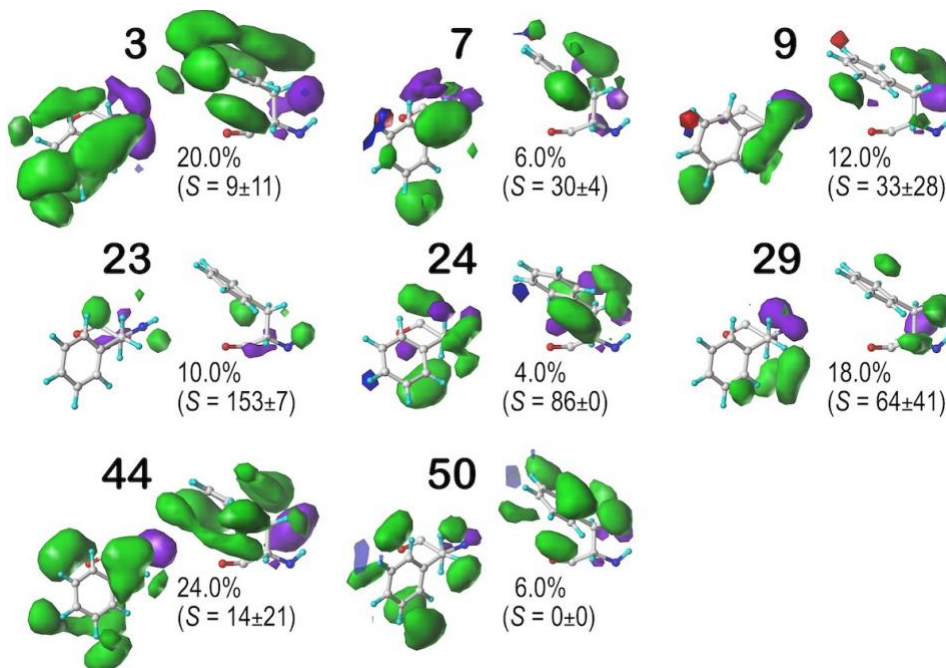


Figure 2.12 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *d5* chess square ($\chi_1 = 180^\circ$ parse).

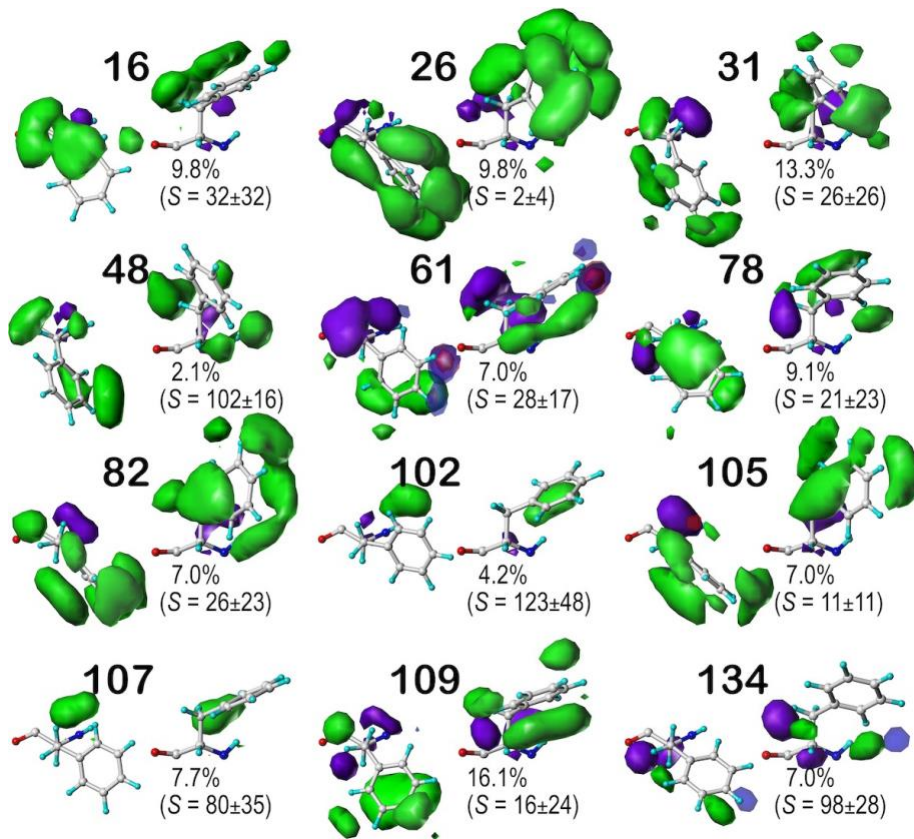


Figure 2.13 Hydrophathic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *d5* chess square ($\chi_1 = 300^\circ$ parse).

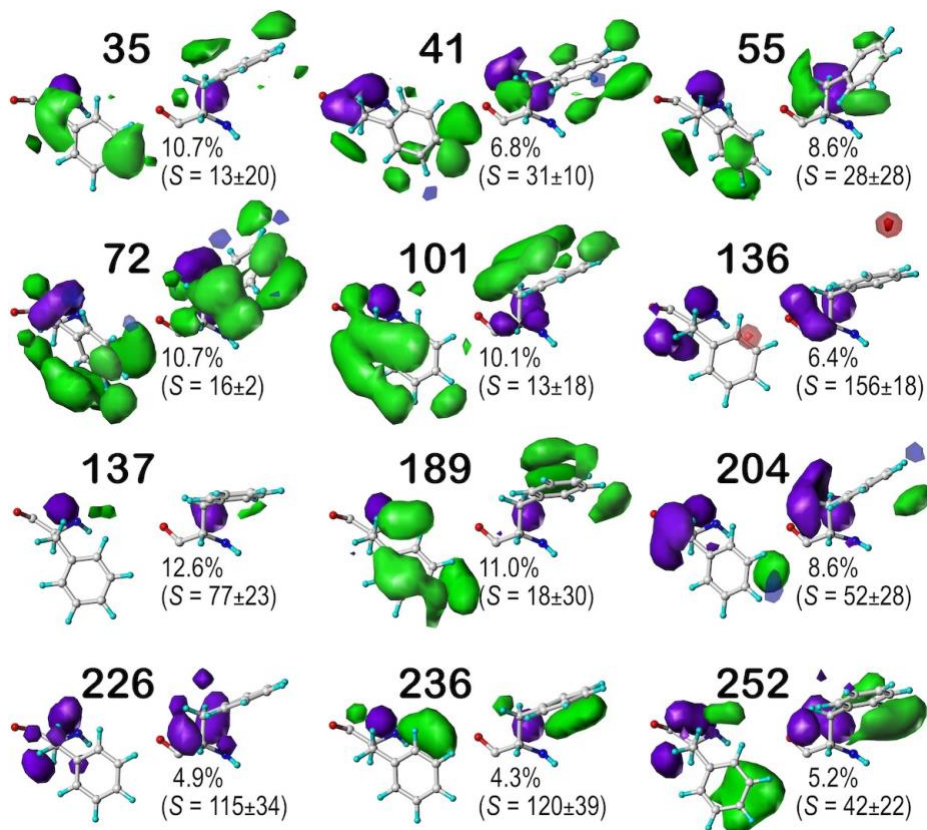


Figure 2.14 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered PHE sidechain environments for the *f6* chess square ($\chi_1 = 300^\circ$ parse).

The *b1* chess square represents a subset of the β -pleat conformation and contains 2149 maps of the overall PHEs in our dataset. The maps were binned into their **.60**, **.180** and **.300** parses, i.e., 1217, 27 and 905 PHE maps were then clustered into 12, 5 and 12 unique cluster environment sets, respectively. At the most basic level, the PHE sidechain consists of the CB methylene and the aromatic phenyl ring. The former is in all cases interacting with polar backbone atoms (N, O) on adjacent residues, and thus demonstrates unfavorable hydrophobic interactions (purple contours) generally seen as spheroids, one on each side of the CB (Figures 2.5–2.7). Since these interaction contours are nearly always in the same place in this chess square, they must be

backbone angle dependent. Some favorable hydrophobic interactions with the CB are shown as small (or merged) green contours, most often in cases with low solvent accessible surface area (*S*), such as clusters **516**, **553**, **725**, **872** and **1192** (***b1.60***, Figure 2.5), **9** (***b1.180***, Figure 2.6) and **26**, **29**, **60**, **122**, **604**, **605** and **870** (***b1.300***, Figure 2.7). Similarly, favorable hydrophobic interactions between the PHE ring and its environment are seen in the maps, but again, this is largely dependent on *S*: the more encompassing green contours, whether above the ring, below the ring or both, are found on the maps of low solvent accessibility. Correspondingly, maps **553** and **761** of ***b1.60*** (Figure 2.5), **9**, **21** and **23** of ***b1.180*** (Figure 2.6) and **60**, **122** and **870** of ***b1.300*** (Figure 2.7) have hydrophobic interactions mostly above the ring; maps **516** of ***b1.60*** and **29** of ***b1.300*** have hydrophobic interactions mostly below the ring; and maps **725**, **872**, **1027** and **1192** of ***b1.60*** and **26**, **453** and **605** of ***b1.300*** have hydrophobic interactions both above and below the ring. Several of these profiles are certainly suggestive of π - π stacking interactions.

Cluster maps with unfavorable hydrophobic (purple) interactions surrounding the ring were rarely seen, which means that, to the first order, aromatic rings are *not* found in polar environments. However, there are some exceptions and caveats: 1) some PHEs are very solvent exposed, such as **88** and **650** (***b1.60***, Figure 2.5), **4** and **12** (***b1.180***, Figure 2.6) and **644** (***b1.300***, Figure 2.7) and potentially available for interactions with water, etc.; and 2) the π -cation interaction (to first order) *is* a hydrophobic-polar interaction, but it is a special case, and a key topic of this work is the relationship between this so-called π -cation and solvent exposure. The HINT scoring rubric recognizes π systems as Lewis bases and thus scores their interactions with hydrogen bond donors as favorable polar.

There is evidence of this type of interaction (blue contours near the rings) in a few **b1** maps: cluster **12** of **b1.180** (Figure 2.6) and clusters **453** and **579** of **b1.300** (Figure 2.7). The “classic” π -cation scenario would also have the PHE sidechain mostly buried, which is the case for the above two clusters in **b1.300**. Examination of the maps for the three χ_1 parses of the **b1** chess squares shows our rationale for further dividing the data this way: 1) the sets of underlying molecular structures are clearly distinct from each other; and 2) high S cases where the phenyl ring shows few interactions would be difficult (or impossible) to differentiate by purely *map-based* clustering. Our description of the three map sets for **b1** above, will serve as guidelines for viewing and interpretation of the other map sets for **c5**, **d5** and **f6** which are shown in Figures 2.8-2.10 (for PHE **c5.60**, **c5.180** and **c5.300**, respectively), Figures 2.11-2.13 (for PHE **d5.60**, **d5.180** and **d5.300**) and Figure 2.14 (for PHE **f6.300**). A few selected maps will be displayed and discussed here. Since the **c5** and **d5** chess squares are both within the Ramachandran plot’s α -helix region, we are interested in comparing their observed interaction profiles to both the **b1** chess square and to each other. In earlier report on ALA,⁴¹ similar map profiles were found in these two chess squares. However, here the situation is far more complex: first, there are three distinct χ_1 rotamer groups, and second, there are many more potential interactions of diverse types between the PHE sidechain and its environment than those of the ALA’s methyl. To assess cluster map similarity, we performed the same type of calculation used in creating the cluster matrix by using the cluster averaged map quartets. High similarity, which in this case is ~ 0.9 , suggests map sets representing the same or very compatible hydrophobic environments. Three such cases are shown in

Figure 2.15: map **169** of **c5.60** with map **19** of **d5.60** (similarity = 0.8933, Figure 2.15A), map **1622** of **c5.300** with map **61** of **d5.300** (similarity = 0.8996, Figure 2.15B), and map **1837** of **c5.300** with map **16** of **d5.300** (similarity = 0.8989, Figure 2.15C). For the **f6** chess square, which is within the left-hand α -helix region of the Ramachandran plot, we only plotted the **f6.300** maps because the other two **f6** χ_1 rotamer groups are essentially unpopulated. This group somewhat surprisingly shows similar map profiles to the right hand α -helix chess squares, especially **c5.300**. For example, cluster **1859** of **c5.300** and cluster **35** of **f6.300** (Figure 2.15D) have a similarity of 0.9113. Also of note are two particularly good examples of the aromatic ring acting as a Lewis base are given by map **168** of **c5.60** (Figure 2.15E) and map **8** of **c5.180** (Figure 2.15F).

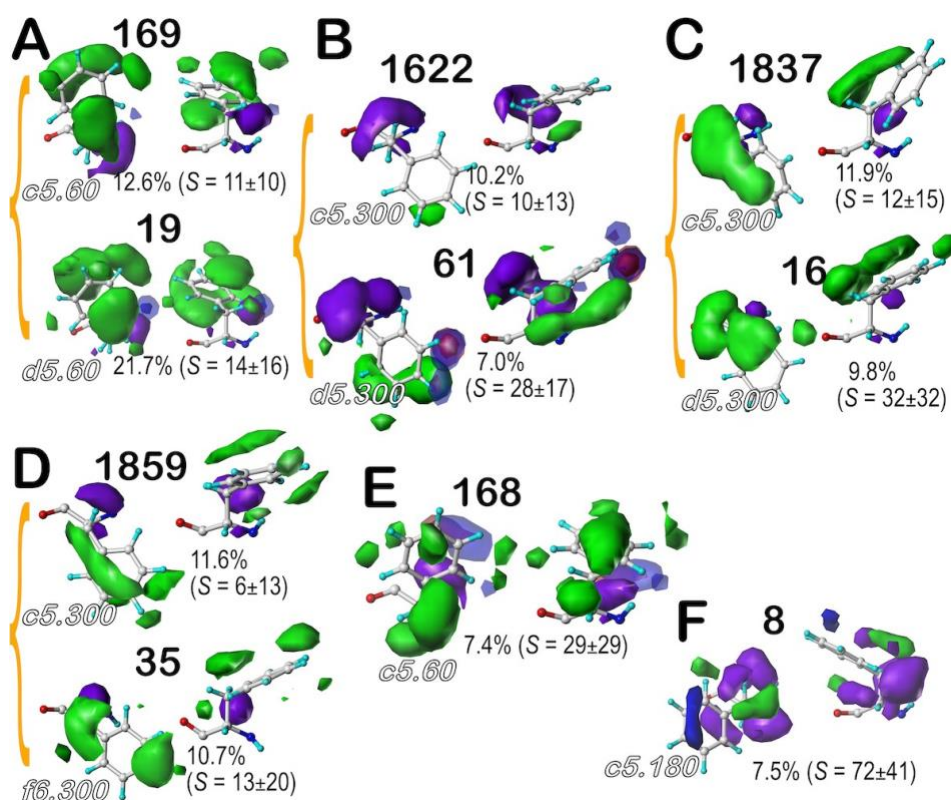


Figure 2.15 Hydropathic interaction map comparisons and putative π -cation features for phenylalanine. A) Maps 169 of c5.60 and 19 of d5.60 have similarity = 0.8933; B) maps 1622 of c5.300 and 61 of d5.300 have similarity = 0.8996; C) maps 1837 of c5.300 and 16 of d5.300 have similarity = 0.8989; D) maps 1859 of c5.300 and 35 of f6.300 have a similarity of 0.9113; E) map 168 of c5.60 and F) map 8 of c5.180 clearly show what may be π -cation interactions (blue contours near phenyl ring). See also caption of **Figure 2.5**.

One rather obvious but interesting observation is the relationship between the map profiles and solvent-accessible surface area (S). There is a backstory as well: S also varies rather significantly with chess square and χ_1 parse. In **b1**, the average S values are 15 Å², 39 Å², 22 Å², respectively for the **.60**, **.180** and **.300** parses (overall S = 18 Å²). The **.180** parse itself is rare for PHE **b1** (1.25%), which suggests a somewhat unusual structural role for these phenylalanines.

The average S for **c5** is 31 Å² (37 Å², 44 Å² and 30 Å² for the three parses); for **d5**, S = 41 Å² (50 Å², 41 Å², 36 Å²) and for **f6**, S = 49 Å² (93 Å², 60 Å², 48 Å²). The relationship amongst protein secondary structure, solvent-accessible surface area and map characteristics, particularly hydrophobic enclosure and polar interactions of the phenyl ring, will be explored in a later section.

2.3.2.2 Sidechain maps of tyrosines. TYR maps in b1 were binned into their .60, .180 and .300 parses, i.e., 1111, 24 and 1105 maps (**Figure 2.3**), were then clustered into 12, 5 and 12 unique cluster environment sets, respectively. The contour levels chosen for all map pairs in both TYR and TRP are identical to allow visual

comparisons of relative interaction strengths: favorable polar (+18); unfavorable polar (-18); favorable hydrophobic (+3); and unfavorable hydrophobic (-6). In some maps, to illustrate the presence of weak polar interactions, translucent blue and red contours at +/-6 were also plotted.

Our description will focus on the interaction profiles and contributions of **b1.60** maps (Figure 2.16), however, the remaining maps in TYR **b1** and the other selected chess squares (**c5**, **d5**, and **f6**) are shown in Figures 2.17–2.18 (for TYR **b1.60** and **b1.180**), Figures 2.19–2.21 (for TYR **c5.60**, **c5.180**, and **c5.300**, respectively), Figures 2.22–2.24 (for TYR **d5.60**, **d5.180**, and **d5.300**, respectively), and Figure 2.25 (for TYR **f6.300**). In the first contribution,⁴⁰ tyrosine hydroxyl was found to be essential in that all environment maps featured favorable polar interactions – hydrogen bonds – with the OH, likely as either/both a donor and acceptor. Simultaneously, the presence of the hydroxyl produced some unfavorable polar and unfavorable hydrophobic interactions. Although our calculation protocols have been somewhat refined since the first report of our group in 2015 (notably, calculating independent sidechain maps and parsing by χ_1), the features seen in **Figure 2.16** and **Figures 2.17–2.26** are consistent. Favorable polar (blue) interactions between the phenolic –OH and its surroundings are observed, and as noted, tend to be ubiquitous in TYR map cluster sets. Excepting the hydroxyl, in general, the interaction profiles of tyrosines tend to be similar to those found above in phenylalanines, displaying the usual observed unfavorable (purple) and favorable (green) hydrophobic interactions between the CB methyl and the lower half of the TYR ring and their surrounding residues.

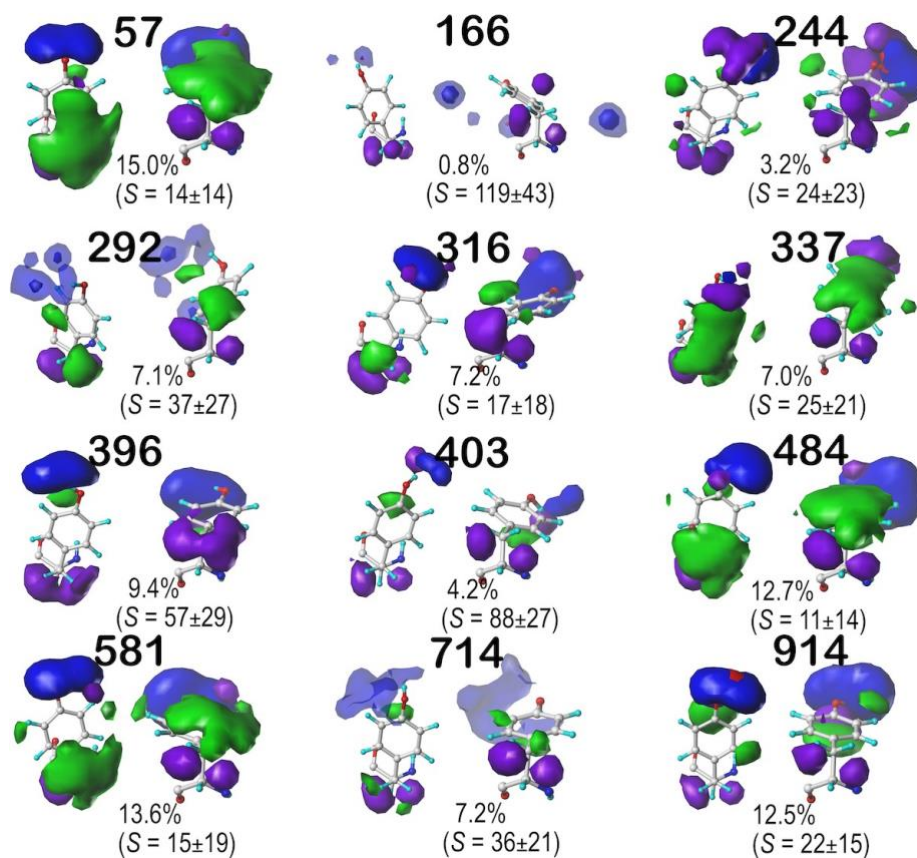


Figure 2.16 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *b1* chess square ($\chi_1 = 60^\circ$ parse).

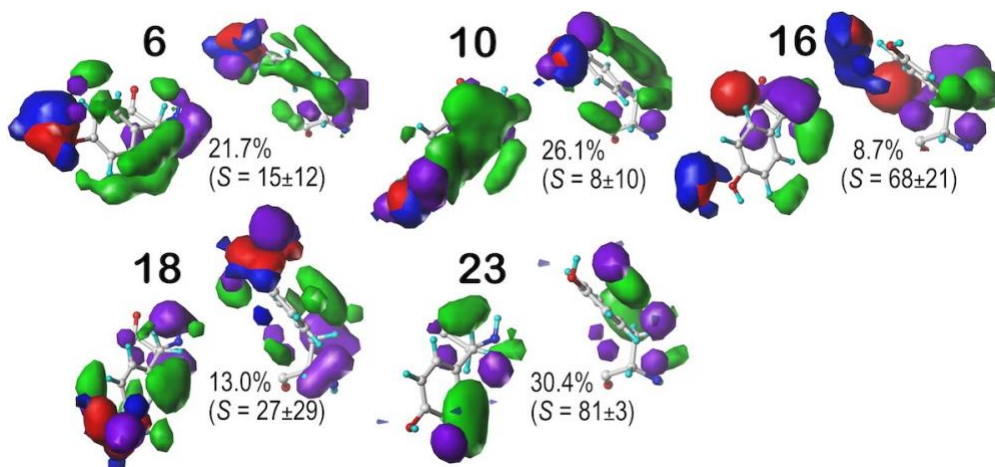


Figure 2.17 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *b1* chess square ($\chi_1 = 180^\circ$ parse).

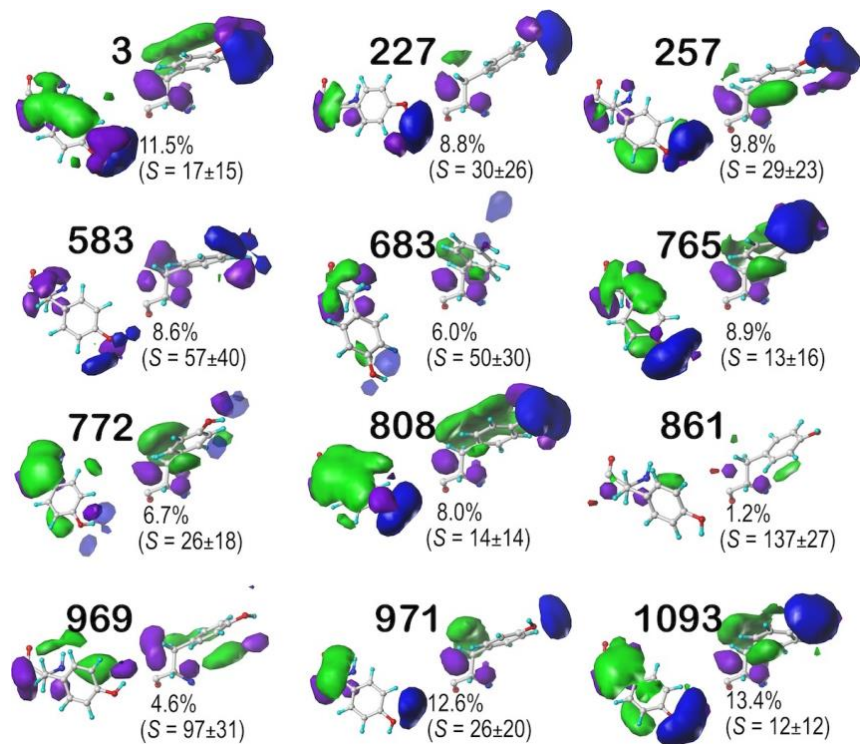


Figure 2.18 Hydrophathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *b1* chess square ($\chi_1 = 300^\circ$ parse).

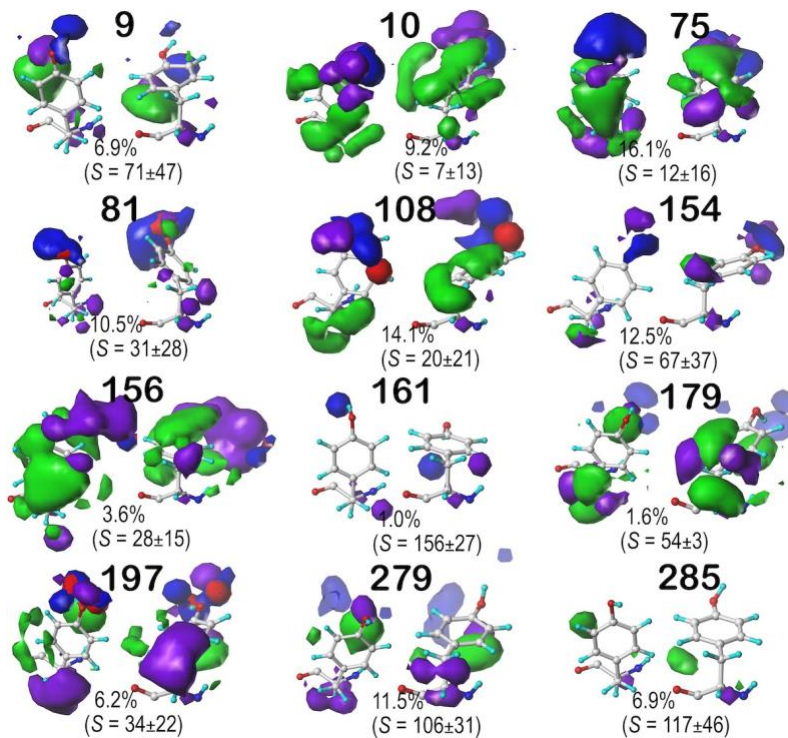


Figure 2.19 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the c5 chess square ($\chi_1 = 60^\circ$ parse).

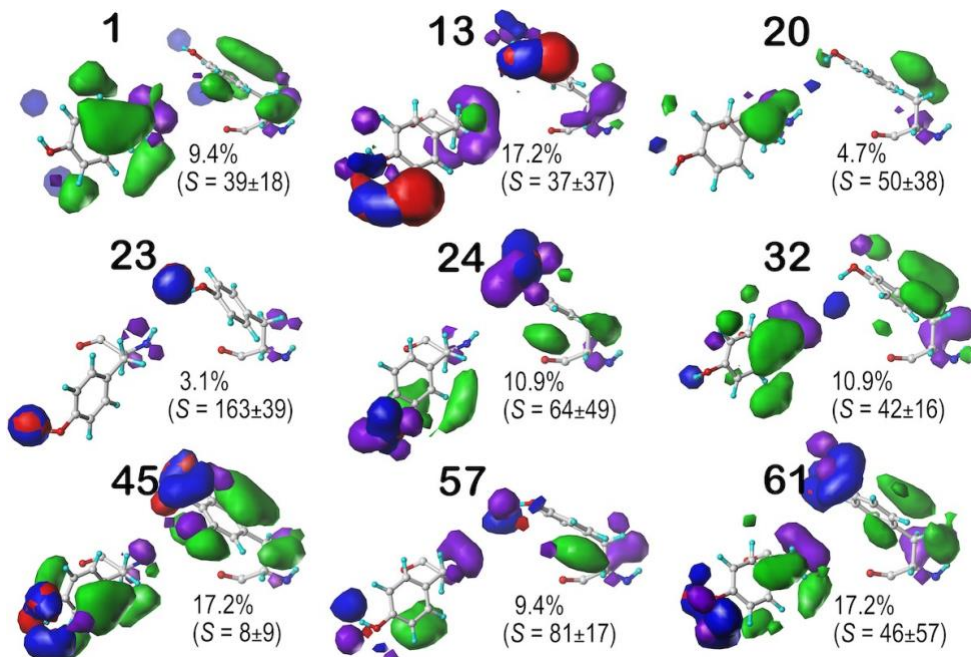


Figure 2.20 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the c5 chess square ($\chi_1 = 180^\circ$ parse).

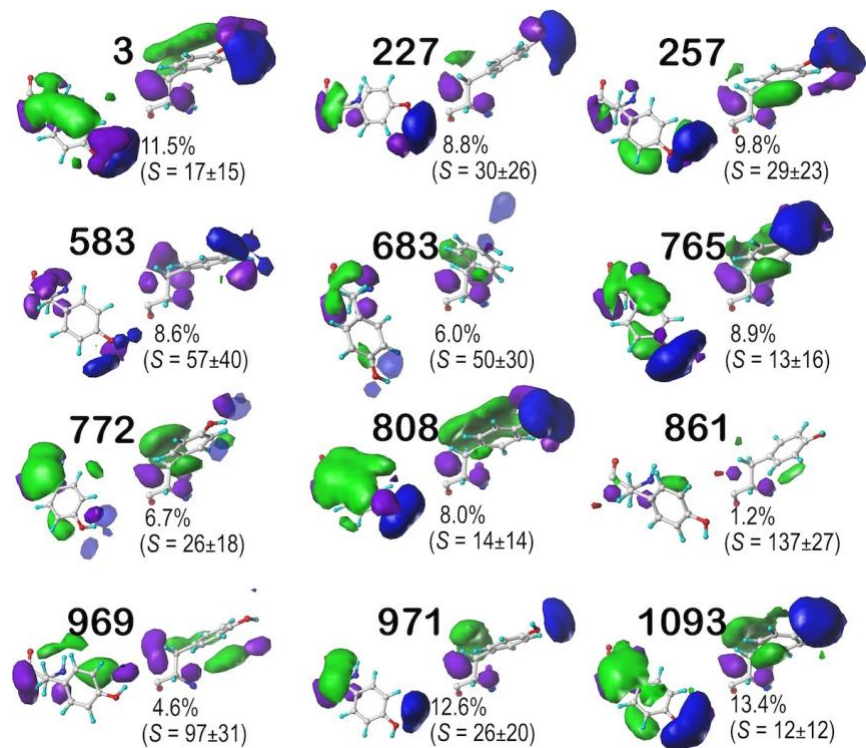


Figure 2.21 Hydrophathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *c5* chess square ($\chi_1 = 300^\circ$ parse).

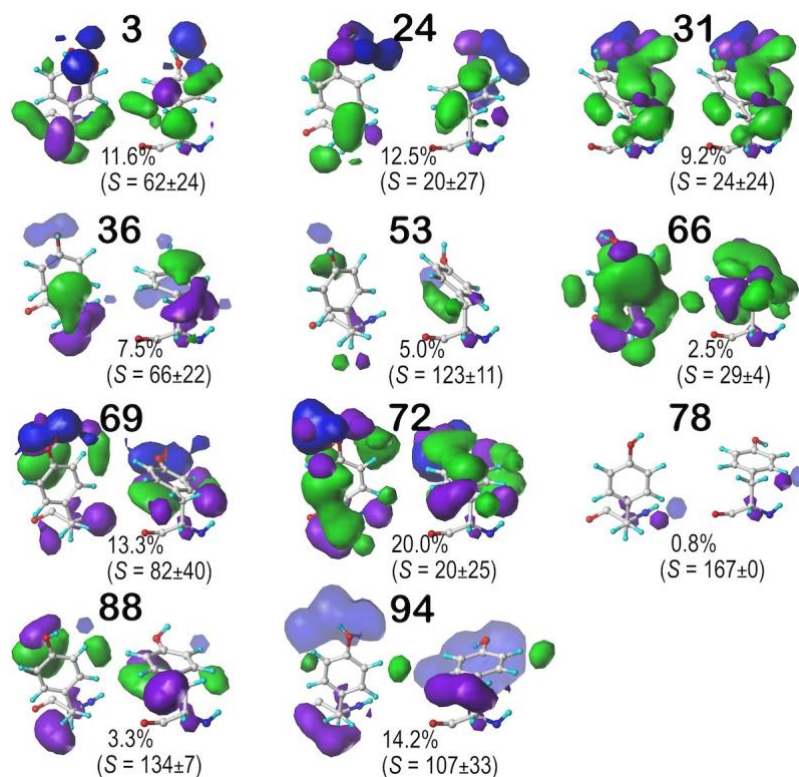


Figure 2.22 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *d5* chess square ($\chi_1 = 60^\circ$ parse).

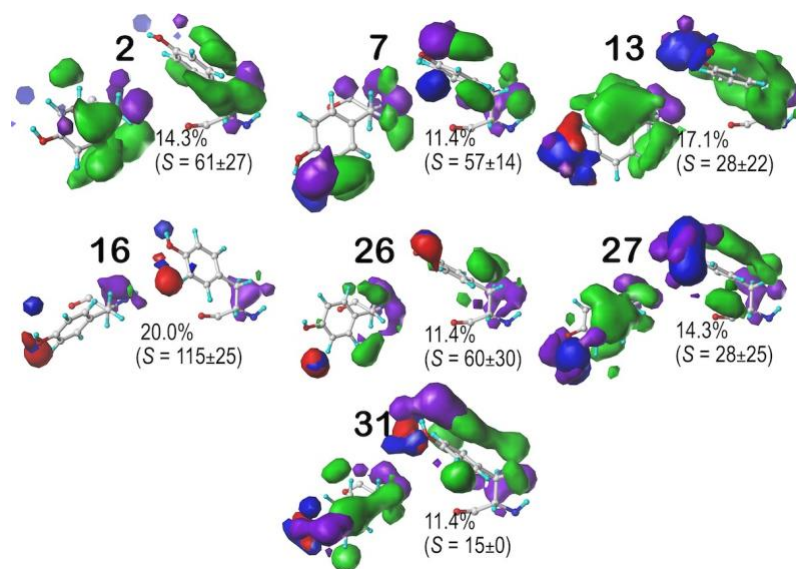


Figure 2.23 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *d5* chess square ($\chi_1 = 180^\circ$ parse).

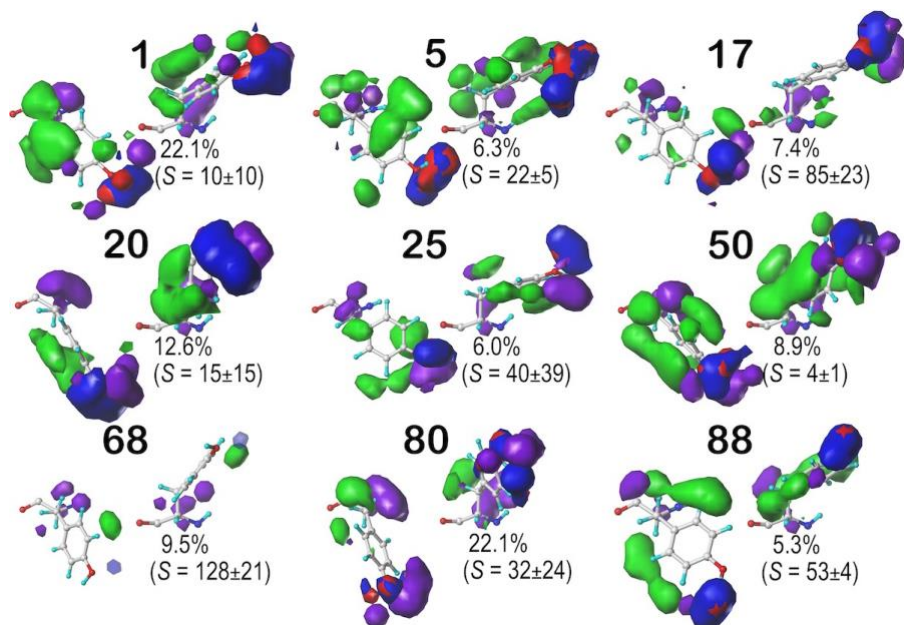


Figure 2.24 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *d5* chess square ($\chi_1 = 300^\circ$ parse).

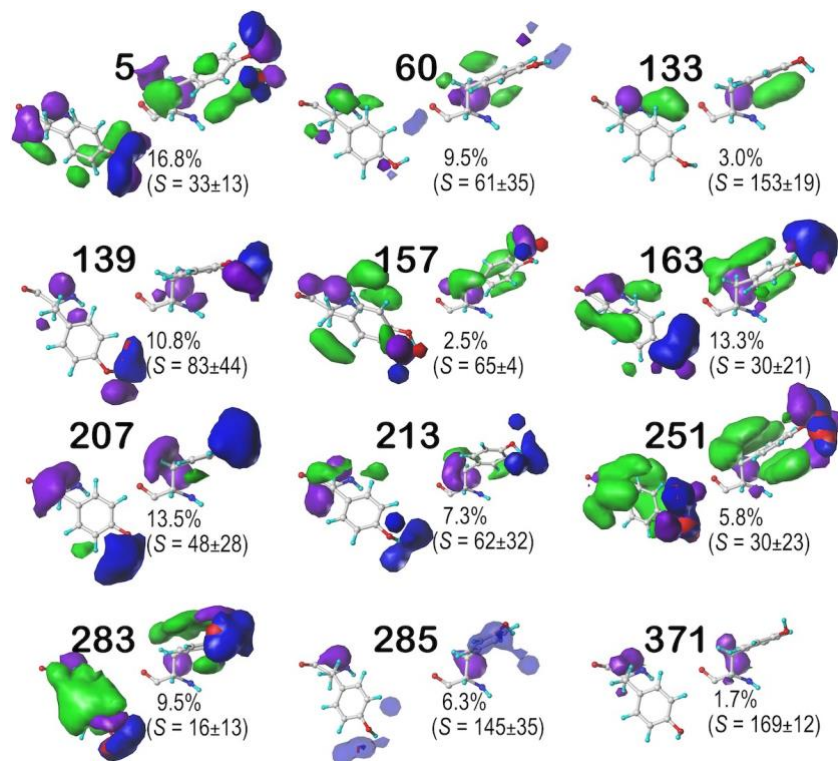


Figure 2.25 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TYR sidechain environments for the *f6* chess square ($\chi_1 = 300^\circ$ parse).

In **b1.60** (Figure 2.16), interaction patterns where the phenolic –OH is engaged in strong favorable polar (blue) interactions, and the aromatic ring is involved in favorable hydrophobic (green) interactions exclusively above the ring, are typified by maps **57**, **316**, **337**, **484**, and **581**, or below the ring by maps **396**, **403**, and **914**. The remaining maps within the **b1.60** chess square/parse reveal somewhat weaker favorable (blue) interactions surrounding the OH: maps **292**, **337** and **714** have average to very sparse hydrophobic interactions around the ring. Cluster **166** is *highly* solvent exposed with comparatively very limited population; TYRs within this group tend to participate in unfavorable hydrophobic (around CB) and favorable polar (around OH) interactions. Identifying the “ π -cation” interaction in these maps is complicated by the presence of the TYR hydroxyl.

As in PHE, there are similarities in map profiles between the **c5** and **d5** chess squares of TYR, rising from both chess squares possessing similar backbone conformations. Compare, for example, map **9** in **c5.60** (Figure 2.19) and map **69** in **d5.60** (Figure 2.22) – similarity = 0.9163, map **108** of **c5.60** (Figure 2.19) and map **24** of **d5.60** (Figure 2.22) – similarity = 0.8943, and map **471** of **c5.300** (Figure 2.21) and map **1** of **d5.300** (Figure 2.24) – similarity = 0.9054. Also, as in PHE, some maps in **f6** have profiles similar to members of the **c5** group: map **1498** of **c5.300** (Figure 2.21) and map **207** of **f6.300** (Figure 2.25) – similarity = 0.9244, and map **1640** of **c5.300** (Figure 2.21) and map **283** of **f6.300** (Figure 2.25) – similarity = 0.9012.

Other than, or probably because of, the –OH of TYR, a key reason for differences between PHE and TYR is that, on average, TYR is more solvent exposed, but again, this is secondary structure-dependent. Thus, tyrosines are mostly found in partially polar

environments, i.e., where the phenolic OH is involved in an H-bond interaction with its surroundings, which would be more likely in a surface or near surface locus. In terms of the relation between map profiles and solvent accessibility, the average S value for **b1** is 27 \AA^2 , 39 \AA^2 , 30 \AA^2 , respectively for the **.60**, **.180** and **.300** parses and overall $S = 29 \text{ \AA}^2$. **c5** shows a $S = 46 \text{ \AA}^2$ (48 \AA^2 , 46 \AA^2 , 45 \AA^2); for **d5**, $S = 50$ (60 \AA^2 , 56 \AA^2 , 37 \AA^2); and for **f6**, $S = 57 \text{ \AA}^2$ (16 \AA^2 , 83 \AA^2 , 57 \AA^2). These are about 10 \AA^2 larger than the analogous values for PHE.

2.3.2.3 Sidechain maps of tryptophanes. TRPs in the **b1** region were binned into 376, 24, and 253 TRP maps (Figure 2.4) and separated into 12, 5, and 12 unique clustered environment sets according to their respective **.60**, **.180**, and **.300** parses. As for TYR, we will focus on the unique interaction profiles of TRP maps in **b1.60** (Figure 2.26). However, the remaining maps in **TRP b1** and other selected chess squares (**c5**, **d5**, and **f6**) are illustrated in Figures 2.27-2.28 (for TRP **b1.300**, and **b1.180**, respectively), Figures 2.29-2.31 (for TRP **c5.60**, **c5.180**, and **c5.300**), Figures 2.32-2.34 (for TRP **d5.60**, **d5.180**, and **d5.300**), and Figure 2.35 (for TRP **f6.300**). In many respects, the interaction profiles of TRP are reminiscent of PHE and TYR. The fused indole aromatic ring system allows more extensive hydrophobic interactions, both above and below it, which is especially evident in maps **106** and **224** (Figure 2.26). Unfavorable hydrophobic contours attributable to the CB methylene are seen, again as in PHE and TYR. The new feature is the favorable (blue) polar interactions between the indolic HE1 and its neighbors seen on the NE1-HE1 vector in nearly all maps.

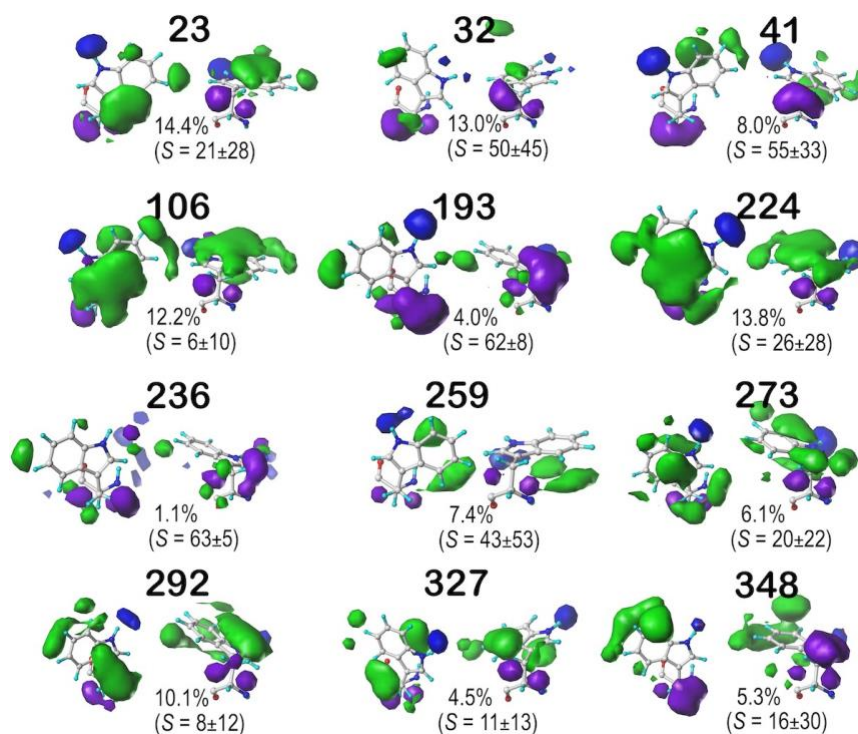


Figure 2.26 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *b1* chess square ($\chi_1 = 60^\circ$ parse).

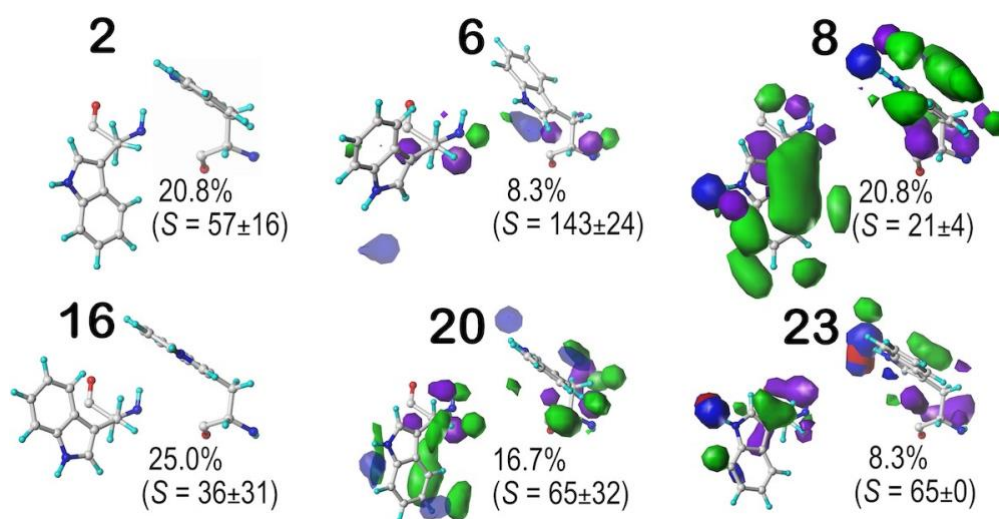


Figure 2.27 Hydrophobic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *b1* chess square ($\chi_1 = 180^\circ$ parse).

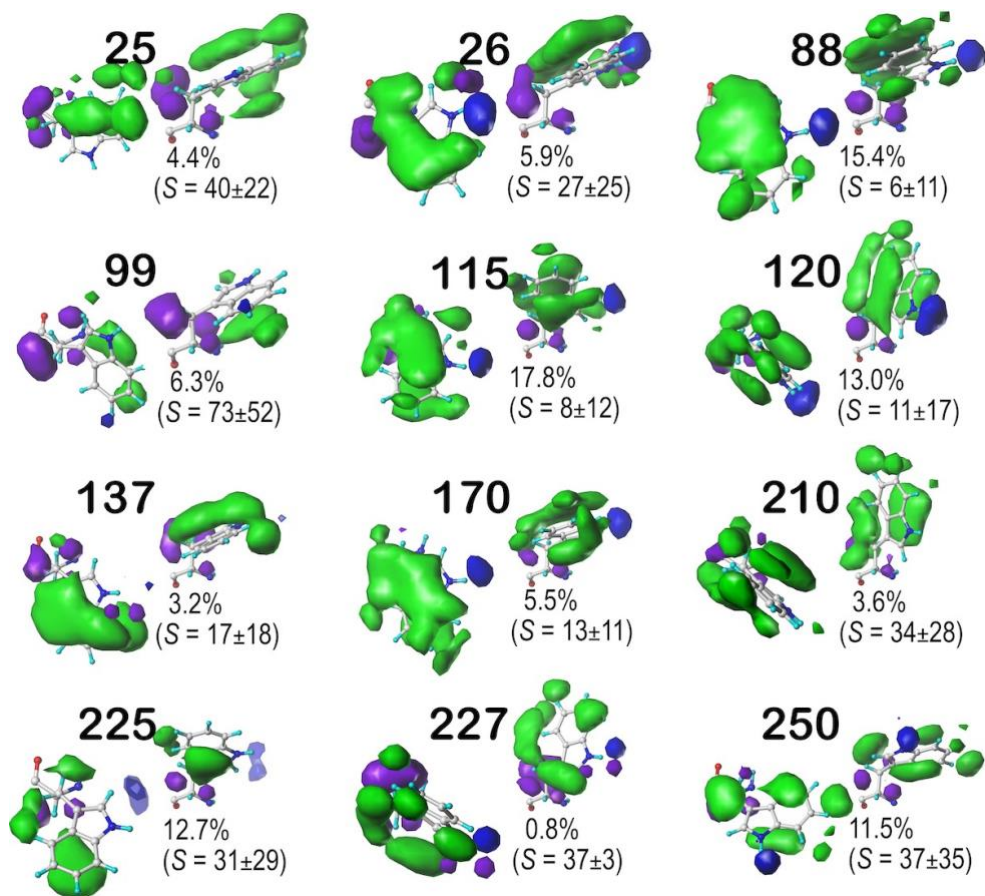


Figure 2.28 Hydrophathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *b1* chess square ($\chi_1 = 300^\circ$ parse).

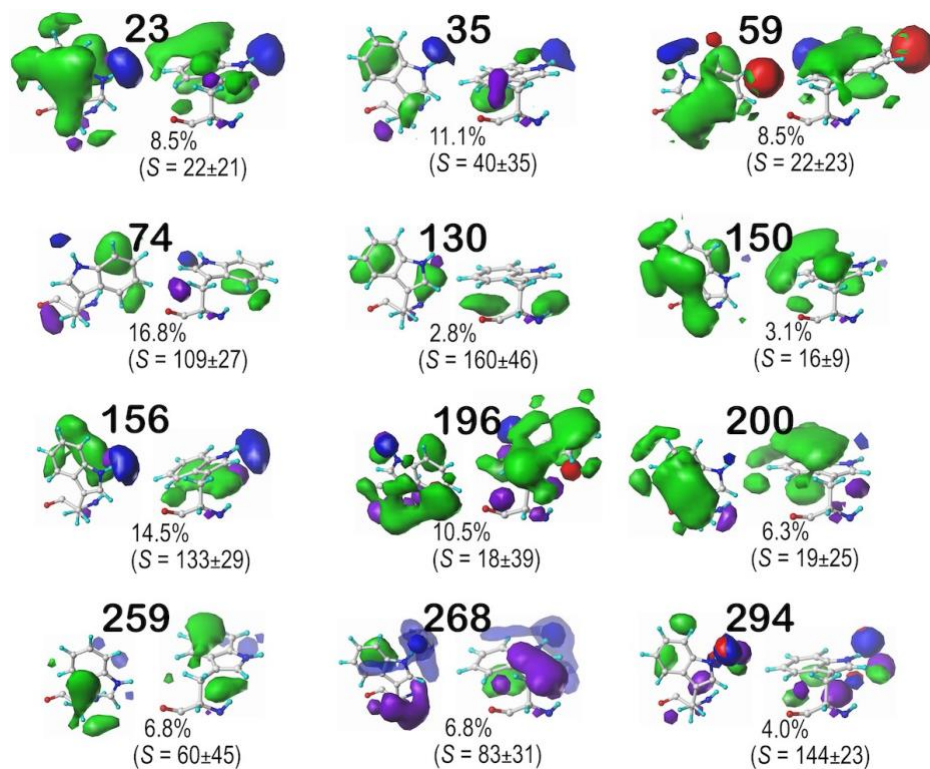


Figure 2.29 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *c5* chess square ($\chi_1 = 60^\circ$ parse).

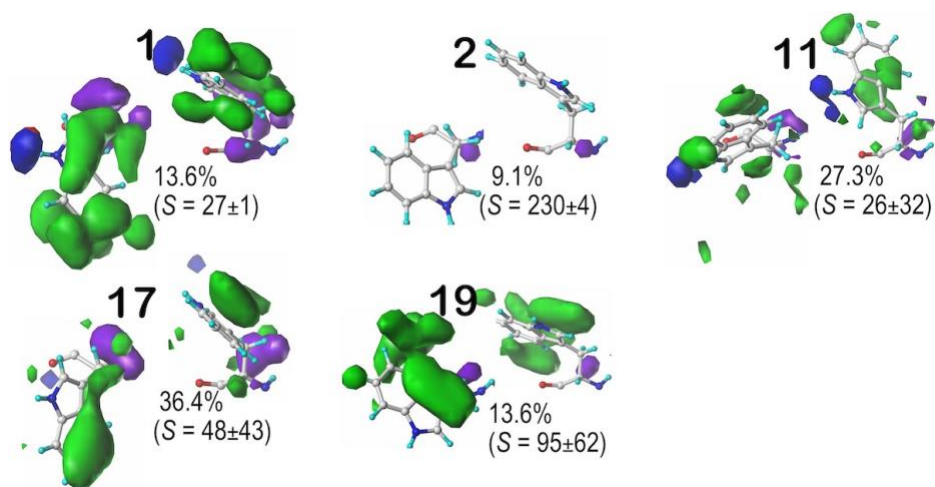


Figure 2.30 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *c5* chess square ($\chi_1 = 180^\circ$ parse).

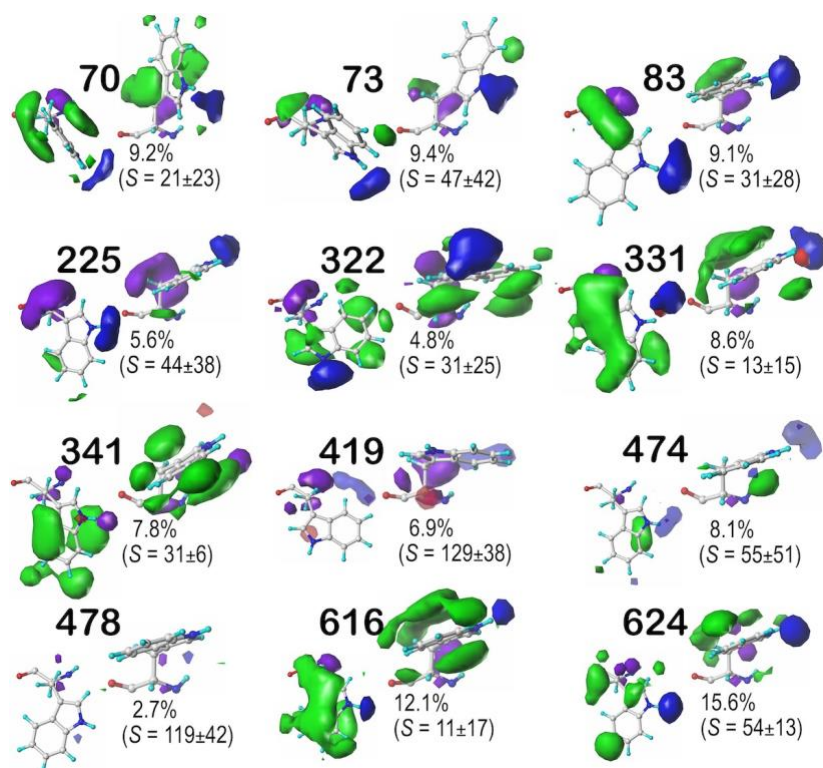


Figure 2.31 Hydrophathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *c5* chess square ($\chi_1 = 300^\circ$ parse).

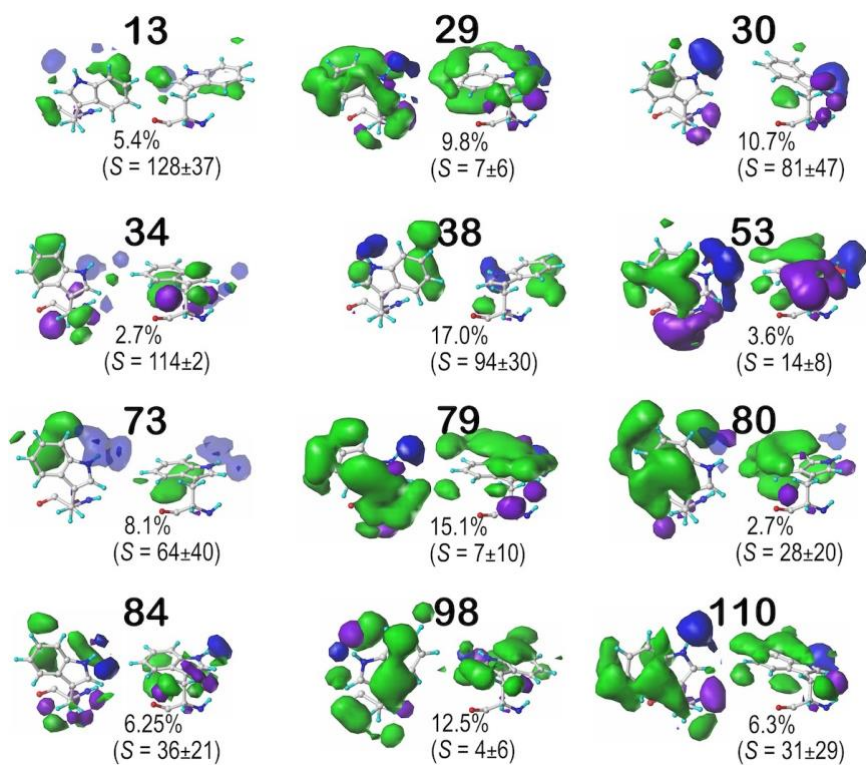


Figure 2.32 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *d5* chess square ($\chi_1 = 60^\circ$ parse).

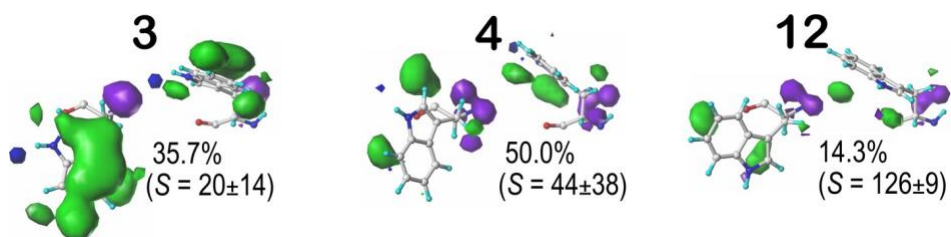


Figure 2.33 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *d5* chess square ($\chi_1 = 180^\circ$ parse).

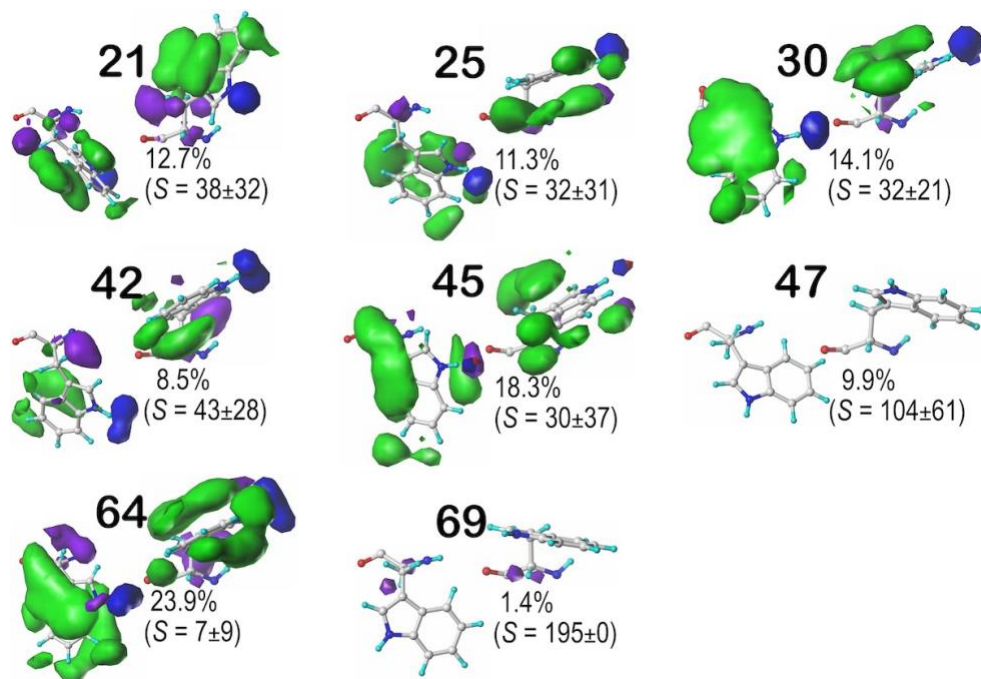


Figure 2.34 Hydropathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *d5* chess square ($\chi_1 = 300^\circ$ parse).

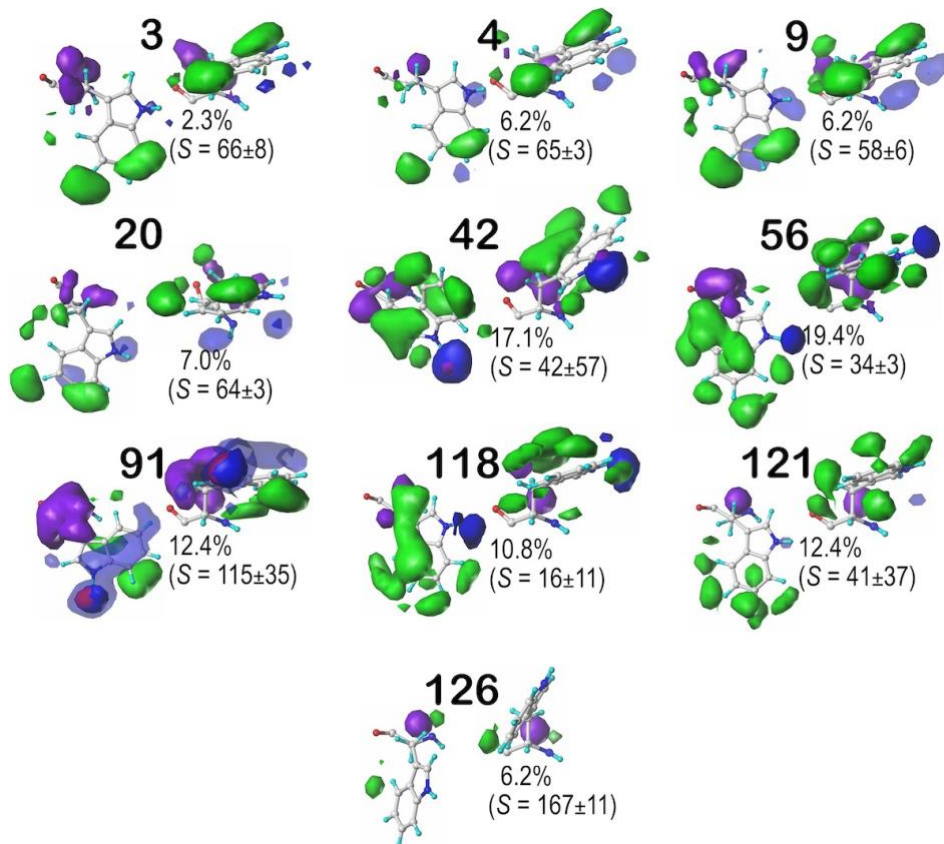


Figure 2.35 Hydrophathic interaction maps illustrating the Gaussian-weighted average clustered TRP sidechain environments for the *f6* chess square ($\chi_1 = 300^\circ$ parse).

We calculated intercluster similarities between maps in *c5* and *d5* to identify map pairs that share compatible hydrophathic environments and S within the same secondary structure motif: map **616** of *c5.300* (Figure 2.31) and map **64** of *d5.300* (Figure 2.34)– similarity = 0.903 and map **331** of *c5.300* (Figure 2.31) and map **30** of *d5.300* (Figure 2.34)– similarity = 0.876 are examples. While we expect that the *f6.300* chess square/parse (Figure 2.35) would show similarity to members of *c5.300*, this is a rare motif for a rare residue type, so we do not have enough data to be confident in these comparisons. TRP has a similar accessibility to TYR: the average S value for *b1* is,

overall, 27 Å² (28 Å², 53 Å², 23 Å², respectively for the **.60**, **.180** and **.300** parses), **c5** has $S = 53 \text{ Å}^2$ (69 Å², 62 Å², 43 Å²), for **d5** has $S = 44$ (48 Å², 47 Å², 37 Å²), and for **f6** has $S = 58 \text{ Å}^2$ (171 Å², 9 Å², 59 Å²).

2.3.3 Solvent-accessible surface areas for the aromatic residues. In the data above, we found that the solvent accessible surface areas for the three residues appeared to show some dependence on chess square and χ_1 parse. This suggested to us that there might be a discernable pattern that could be revealing of structure, particularly within our chessboard schema, which was by design a rational system to organize structural data as a function of backbone angles. To explore this relationship, we used our $f_{outside}$ metric based on GETAREA⁵¹ as described above.

The $f_{outside}$ for each chess square/parse was calculated and illustrated in **Figures 2.2-2.4** with the colors of the bars (that represent parse populations by their lengths). Chess square/parses within the β -pleat region of the Ramachandran plot for phenylalanine (Figure 2.2), as expected, show the lowest $f_{outside}$ (more buried) relative to the right- and left-hand α -helix, i.e., averaged $f_{outside}$ of all PHE parses within the β -pleat ranges from 0.0-0.2 (red) to 0.2-0.4 (yellow), only two outliers, **a6.180** and **c6.180**, displaying a higher (0.4–0.6) exposed fraction. Phenylalanines in the α -helix region are somewhat more solvent exposed, with $f_{outside}$ ranging between 0.0 to 0.6, and the left-hand α -helix is still more exposed. The blueprint for PHE residues in protein structure seems to suggest that full burial is preferred, as is seen in its most populated chess squares/parses. Larger solvent exposure correlates with rarer (and often unusual)

backbone $+\chi_1$ conformations. For example, the two β -pleat outliers in **a6** and **c6** are $\chi_1 = 180^\circ$ parses, where the PHE sidechain does not interact with, or reinforce, other elements of the β -sheet. Clearly because of its *para* hydroxyl, tyrosine is found in more solvent exposed loci. This is evident in **Figure 2.3**, where most chess squares/parses report $f_{outside}$ values in the 0.2-0.4 range. Some backbone $+\chi_1$ conformations in the β -pleat region remain fully buried, which is consistent with the expected stabilization of β -sheets via hydrophobic interactions. Tryptophan data (see **Figure 2.4**) are more similar to those of phenylalanine. As in PHE, both α -helix regions of both TYR and TRP are marginally more solvent accessible than in their β -pleat motif regions. In fact, $f_{outside}$ values are dependent on the relative hydrophobicities of the sidechains (PHE < TRP < TYR). Generally, PHE has the least exposed fractions: the average $f_{outside}$ for the β -pleat region is 0.12, for the α -helix $f_{outside} = 0.15$, and for the left-hand α -helix $f_{outside} = 0.30$; for TYR, the average $f_{outside}$ for the β -pleat region is 0.20, for the α -helix $f_{outside} = 0.26$, and for the left-hand α -helix $f_{outside} = 0.38$; and for TRP, the averaged $f_{outside}$ for the β -pleat region is 0.14, for the α -helix $f_{outside} = 0.21$, and for the left-hand α -helix $f_{outside} = 0.32$.

The $f_{outside}$ values for all three aromatic residues on a cluster-by-cluster basis have been published and made available (Supporting Information Tables S2-S4)⁴², and were sampled in **Table 2.4** for the **b1.60** parses of PHE, TYR and TRP. Clear from these data is that the aggregate averages are not the complete story: some clusters in **b1.60** are > 50% exposed even though the overall $f_{outside}$ for these residues suggests nearly complete burial.

2.3.4 The special roles of aromatic residues. It is likely that no class of amino acid residues has received more attention than the three with aromatic sidechains. As noted, two interaction motifs of special interest are “ π -cation” and “ π - π ” between residues in the same protein, residues between interaction proteins and between proteins and small molecules. These interaction types are clearly intriguing and, even after decades, remain controversial. Even recent articles, e.g., by Dougherty^{9,21} and Vernon et al.,¹⁰ in continuing efforts to illuminate the functional contributions of aromatic amino acids in various biological processes, do not yield a fully satisfactory understanding of these interactions with respect to structure.

In contrast, our simple HINT model, upon which the present work is based, recognizes these interactions at their basic level: a) the π -cation interaction is the result of aromatic carbons being hydrogen bond acceptors, and b) π - π is a special case of the hydrophobic interaction. On an atom-atom pair basis neither of these terms are particularly noteworthy, but with idealized geometry and the energetics of, e.g., a lysine's $-\text{NH}_3^+$ perched over a phenylalanine ring, or a pair of neatly stacked phenylalanines, the effects of those atom-atom interactions could be significantly multiplied. The three-dimensional hydrophobic interaction maps we have calculated for the three aromatic residues thus should, in principle, encode these interactions. The π -cation interaction should be observed as favorable polar interactions more or less perpendicular to the ring and the π - π interactions should be observed as favorable hydrophobic more or less parallel to the ring. Note that these are less rigid definitions of the two phenomena: a) we are allowing the possibility of a neutral hydrogen bond donor (even a water) to interact

with the ring, and b) any hydrophobic group, not just those that are aromatic, could interact with the ring.

2.3.5 Calculating interaction character near aromatic rings. In order to evaluate the character of interactions we sampled in “cylinders” perpendicular to the rings with a height of 4.0 Å (**Figure 2.36**). We defined two cases, above the ring and below the ring, to see if there were any differences. We presupposed that there would be a fairly limited effect from the somewhat more sterically constricted region on the CA side of the ring, i.e., “below the ring”. The raw interaction character is the sum of all grid point values enclosed within each cylinder for the four map types in the quartet, as described in the Methods. The resulting hydro(+) and polar(+) characters are reported for PHE, TYR and TRP for all clusters of *b1.60* in Table 2.4, and in detail in Tables S2-S4 (published version).⁴²

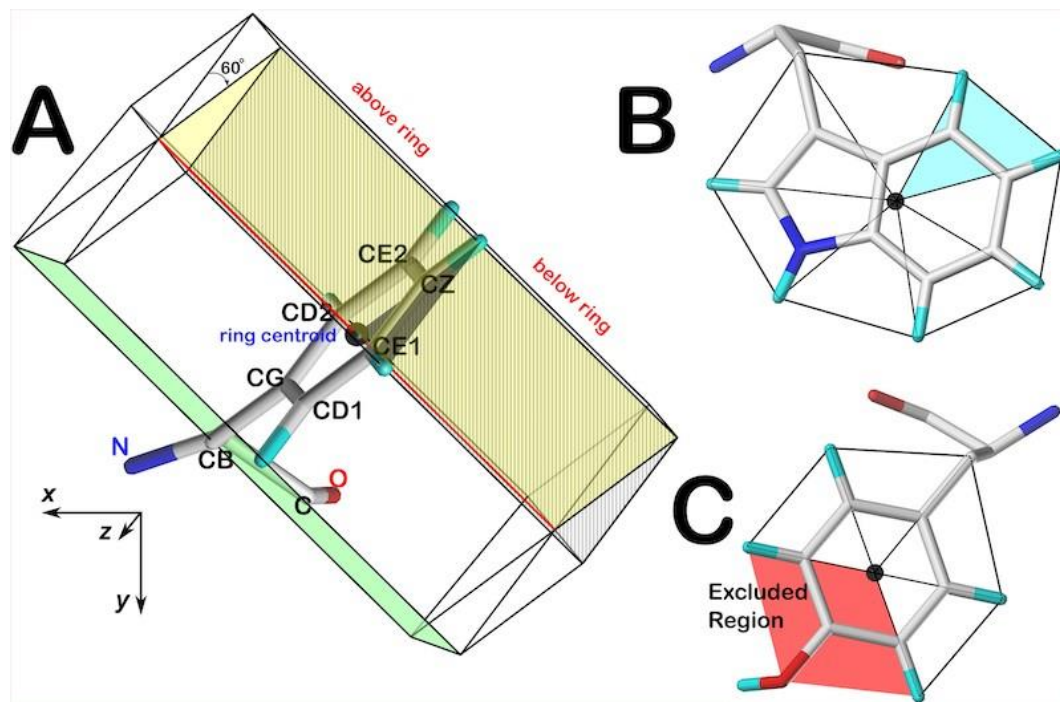


Figure 2.36 Definition of map cylinders for PHE, TYR and TRP. A) The 6 carbons of the phenyl ring are averaged to define its centroid. Seven normals passing through the centroid

and the indicated atoms are used to define six triangular prisms where regions above and below the ring can be differentiated. The sums of the six prisms defines the cylinders. B) The triangular prisms and cylinders for TRP are as indicated. C) To remove from consideration most interactions due to the phenol -OH of TYR, four trigonal prisms and the resulting cylinders are defined as shown.

2.3.6 Relationship between $f_{outside}$ and interaction character. For PHE, TYR, and TRP, **Figures 2.37–2.39**, respectively, set out plots of these interaction characters as functions of the SASA parameter $f_{outside}$ that we defined above, for each hydrophobic interaction map cluster. We show both “above the ring” (top) and “below the ring” (bottom) cases. Each marker in the plots is color-coded by its interaction type and sized based on the population of that cluster. Further, weighted by $1+2\log_{10}(\text{cluster population})$, least squares fit to the functional form $ax^3 + bx^2 + cx + d = 0$, are shown as the solid lines. We do not advocate that these fits provide any meaningful analytical data – indeed their statistical metrics are poor – but they are useful for visualization.

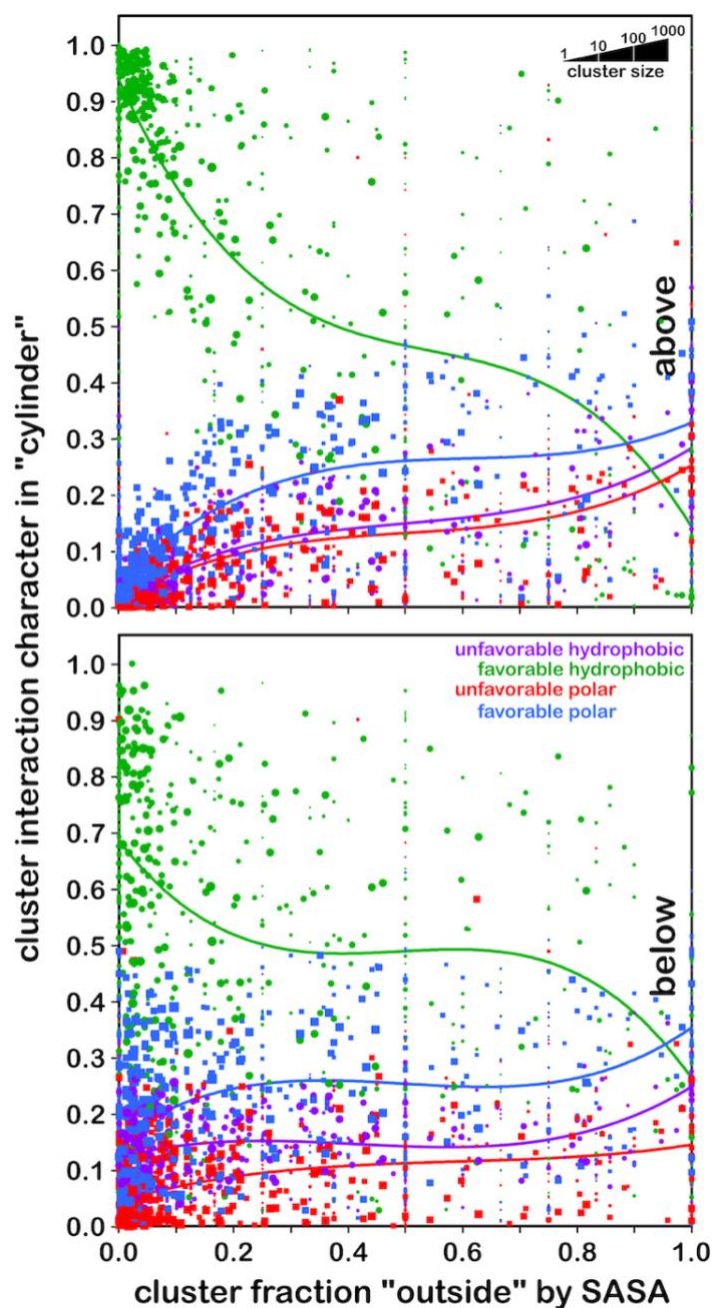


Figure 2.37 Hydrophobic character for clusters in PHE cylinders. Top) interaction character for cylinders above the phenyl ring, i.e., on the side farthest from the C_A atom; Bottom) interaction character for cylinders below the phenyl ring, i.e., on the side closest to the C_A atom. The green markers and line represent favorable hydrophobic interactions, purple represents unfavorable hydrophobic, blue represents favorable polar and red represents unfavorable polar. Each data point plots interaction character (summed from grid points as

described in the text) as a function of f_{outside} for a map environment cluster. The sizes of the markers are \log_{10} -scaled by the number of members of the cluster. The fit lines are the result of weighted least squares analyses as described in the text.

The most significant interaction character, obviously, is the favorable hydrophobic interaction, prominent both above and below the PHE ring. These observations are in accord with the averaged 3D PHE maps (**Figures 2.5-2.14**) where most clusters with low S showed hydrophobic interactions in one or both of these regions. Strikingly, but probably be expected, the character of interactions above the ring in the low f_{outside} range (0.0- 0.2) are 65-95% favorable hydrophobic. The below the ring case shows a less dramatic trend, probably due to steric effects of the backbone, and interactions between its polar atoms and the ring. With the more exposed map clusters, we see a decrease in favorable hydrophobic character coupled with increasing favorable polar interaction character within the cylinderspace. The unfavorable hydrophobic interaction character tracks closely with that of the favorable polar, likely in part due to interactions between the same polar atoms and the nonpolar protons on the ring.

Clearly, at the higher f_{outside} range, e.g., >0.50 , polar interactions dominate, and the associated interaction profiles may even involve water molecules. Favorable and unfavorable polar interactions under these conditions are expected, as hydrophobic environments are by definition not solvent exposed. In contrast, favorable polar interactions at relatively low f_{outside} , for $0.20 < f_{\text{outside}} \leq 0.50$, are much more interesting, and account for ~20-25% of the overall interactions. These interactions would seem to fit the classic “ π -cation” definition, but regardless if they are true to that definition, they

contribute significantly to the stabilization and conformation preferences for PHE residues.

We also explored whether there are secondary structure-related differences in these relationships. **Figure 2.38** illustrates these effects. We found interesting variations amongst the three secondary structure motifs, β -pleat, and right- and left hand- α -helix, with respect to interaction character. We already noted that the below the ring cases likely include polar interactions from the backbone, and that is evident here as well. The β -pleat case displays a rapid and mostly linear loss of favorable hydrophobic character as solvent accessibility increases compared to the other two secondary structure motifs, which are surprisingly nearly identical to each other. These plots suggest that even at $f_{outside} \sim 0.75$, there is a significant fraction of favorable hydrophobic interactions in the two α -helix motifs ($\sim 60\%$) compared to the β -pleat motif ($\sim 30\%$), but it is important to recall that overall, especially in the β -pleat region of the Ramachandran plot, most PHEs are buried (with $f_{outside} < 0.40$, **Figure 2.2**). Thus, the high $f_{outside}$ cases are themselves less common – a point also evident in **Figure 2.37** in that more large-sized markers are found on the left side of the plot.

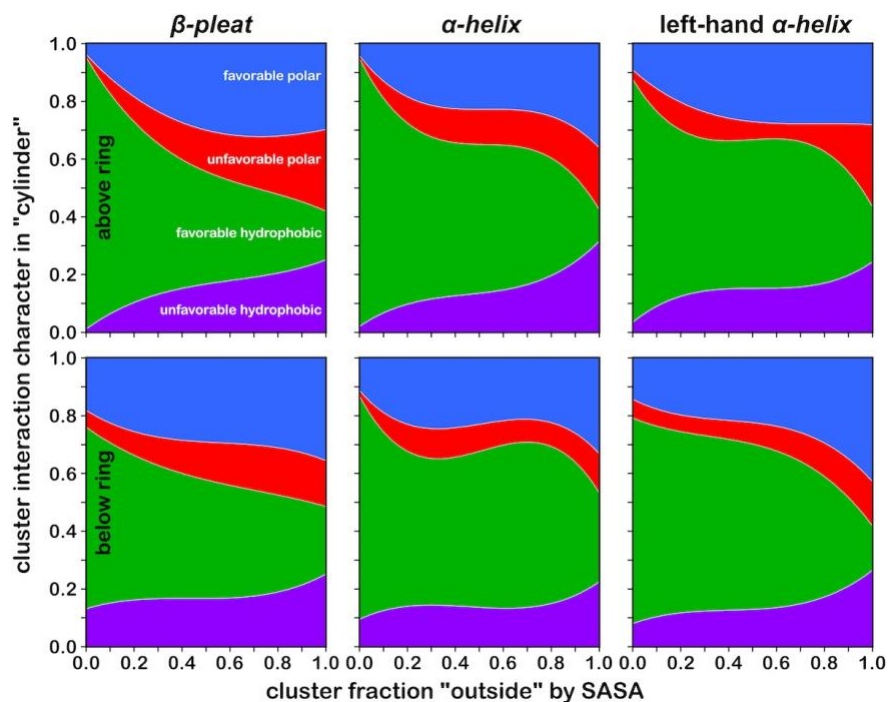


Figure 2.38 Hydrophathic interaction characters for clusters in PHE cylinder by Secondary Structure. Secondary structure-related differences for the four interaction characters within cylinders above (top row) and below (bottom row) the phenyl ring. Fractions are based on the weighted least squares fits calculated as described in the text. Character fractions are colored as shown.

Figure 2.39 shows the relationship between interaction character and solvent accessibility within the 3D cylinders above and below the TYR ring. Like PHE, the favorable hydrophobic character seems to be more significant and SASA-dependent above the ring than below. In fact, the below the ring plot is monotonic with all four of the character type contributions varying little over the entire $f_{outside}$ range. These behaviors are likely due to two related factors: 1) TYR residues are, in general, more solvent exposed than PHE (**Figure 2.3**); and 2) the $-OH$ produces interactions with other residues within the cylinders, despite our efforts to exclude them in our cylinder design

(Figure 2.36). Again, favorable polar interactions for $0.20 < f_{outside} \leq 0.50$ account for ~20-25% of the overall, suggesting the presence of π -cation interactions. We also explored whether there are secondary structure-related differences in the hydrophobic interaction characters above and below TYR rings: **Figure 2.40** shows these effects. Briefly: 1) for the β -pleat motif, the TYR behavior above the ring is very similar to that of PHE (Figure 2.37); 2) for the two α -helix motifs, probably due to increased solvent exposure, the behavior above the ring is more similar to that of the β -pleat motif; and 3) similar to the overall result, the behavior below the ring for all three motifs is largely featureless.

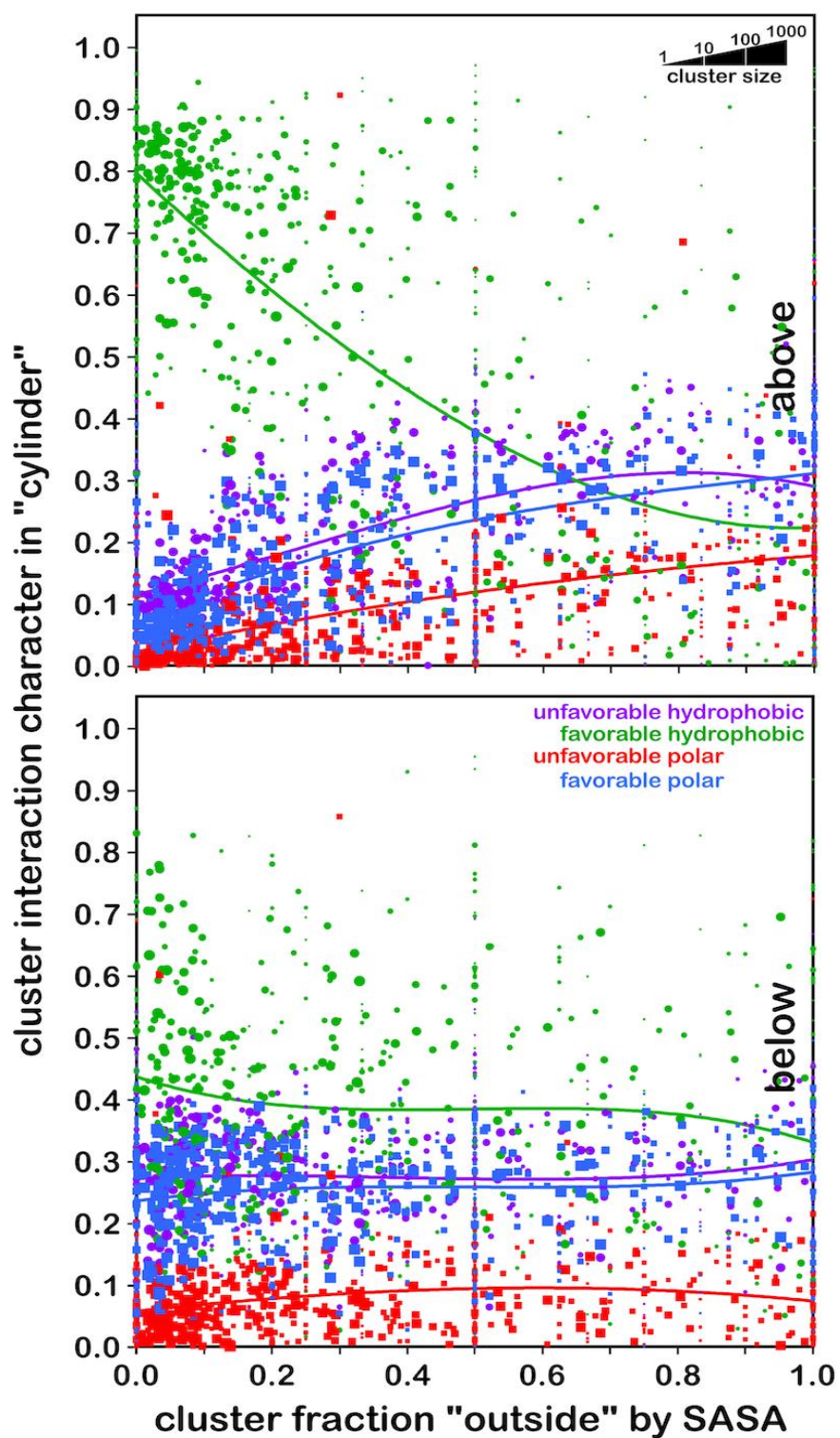


Figure 2.39 Hydrophobic character for clusters in TYR cylinders. Top) interaction character for cylinders above the tyrosine ring, i.e., on the side farthest from the C_A atom; Bottom)

interaction character for cylinders below the tyrosine ring, i.e., on the side closest to the C_A atom. See caption for **Figure 2.37**.

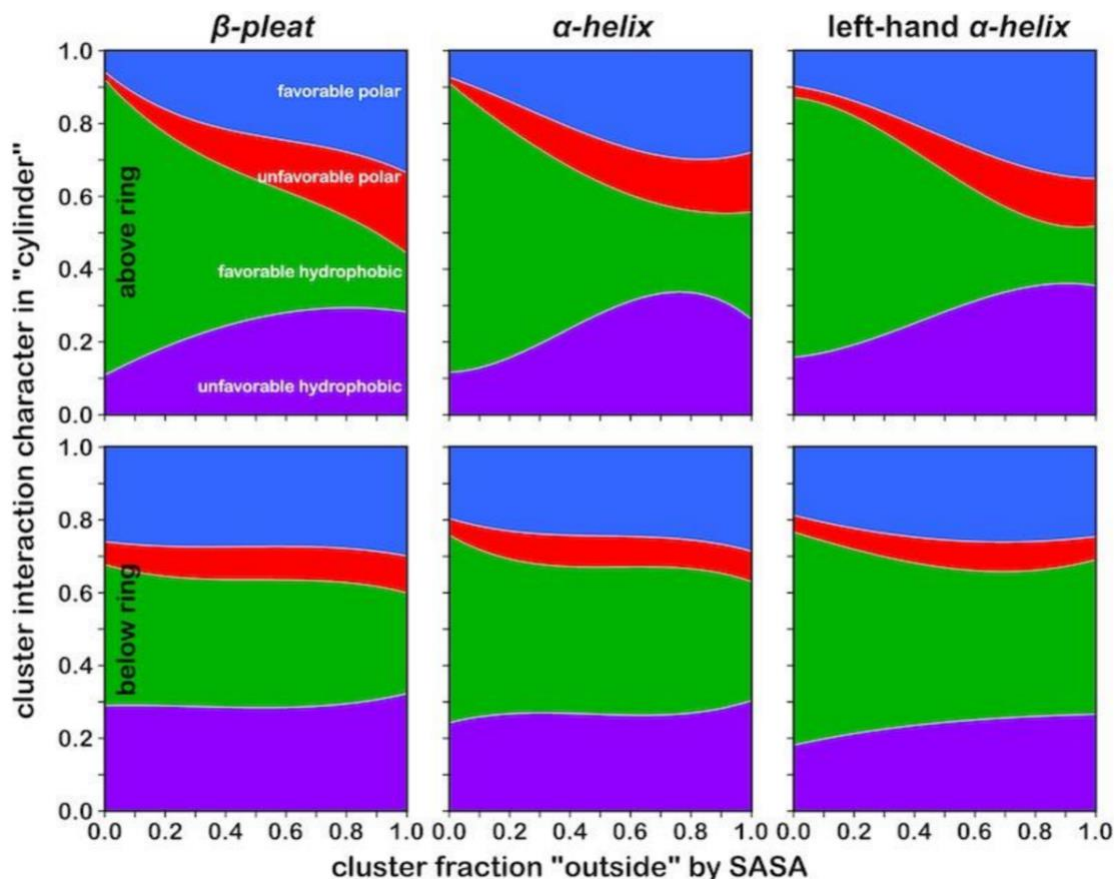


Figure 2.40 Interaction types for modified volumes above and below phenol ring in TYR by secondary structure. Results from weighted least squares fits as described in text; colors as indicated.

The graph in **Figure 2.41** illustrates the relationship between character and $f_{outside}$ in cylinders below and above the tryptophan ring. These plots are qualitatively similar to those of tyrosine (Figure 2.39), except that there is a somewhat larger favorable polar character above the ring for $0.20 < f_{outside} \leq 0.50$, suggesting, as has been previously

suggested,⁹ that TRP is more effective for forming π -cation interactions. We also studied the secondary structure-related profiles for tryptophan (Figure 2.42) which are generally similar to those of TYR.

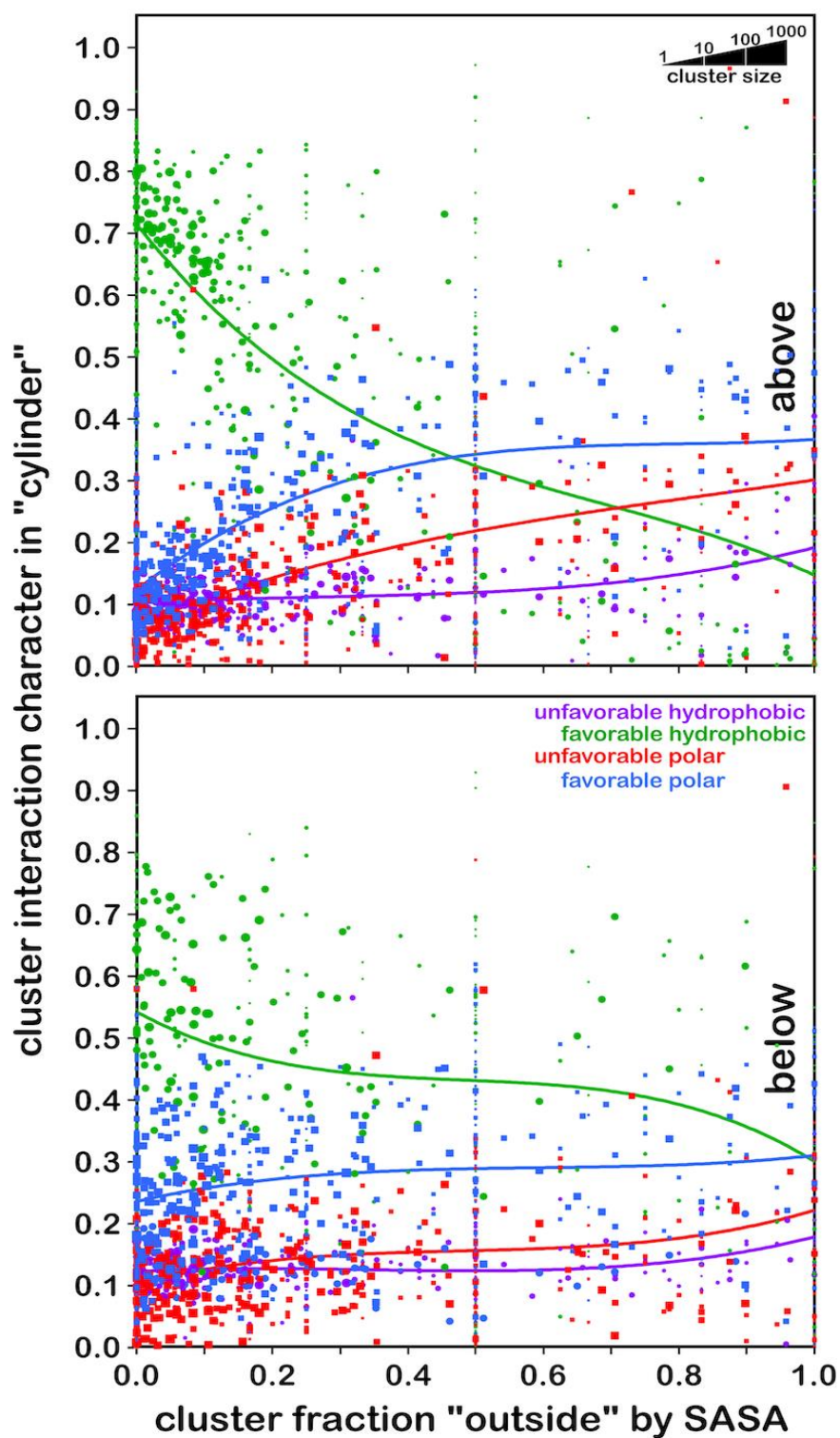


Figure 2.41 Hydropathic character for clusters in TRP cylinders. Top) interaction character for cylinders above the tyrosine ring, i.e., on the side farthest from the C_A atom; Bottom)

interaction character for cylinders below the tyrosine ring, i.e., on the side closest to the C_A atom. See caption for **Figure 2.37**.

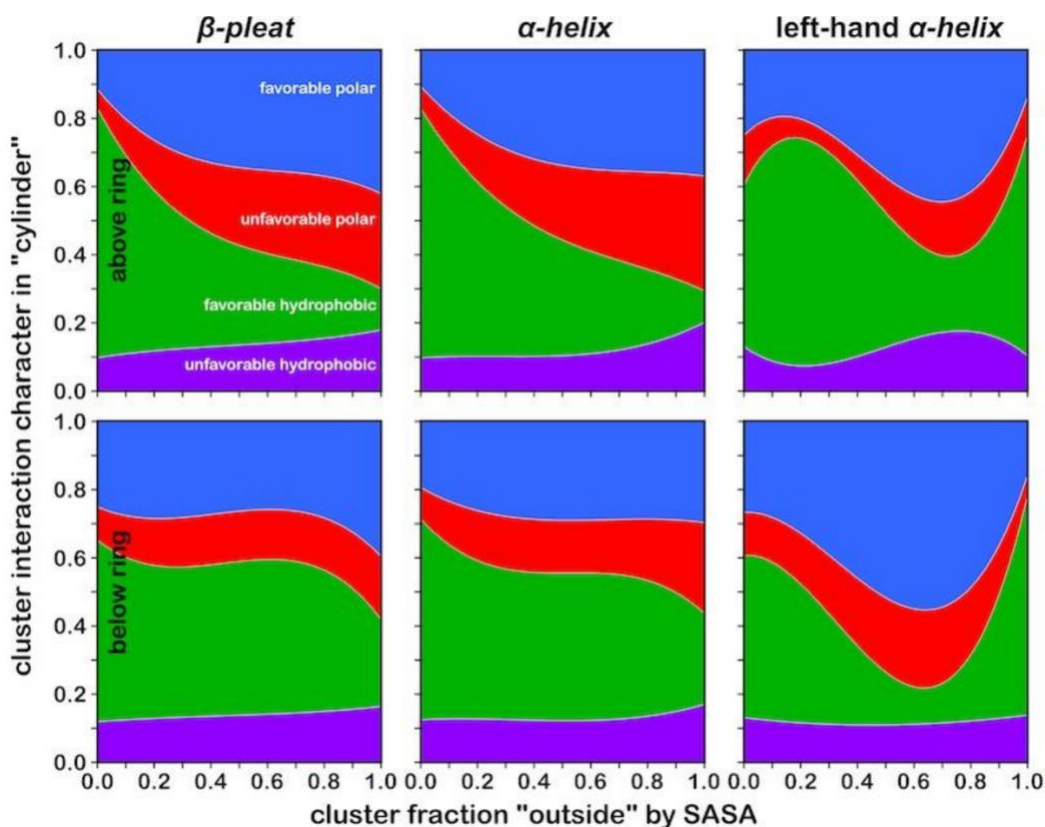


Figure 2.42 Interaction types for modified volumes above and below phenol ring in TRP by secondary structure. Results from weighted least squares fits as described in text; colors as indicated.

2.3.7 Identifying “True” π -cation and π - π stacking interactions. In the results above, we broadly evaluated favorable polar and favorable hydrophobic interactions within the cylinder volumes above and below the aromatic rings. It is reasonable to

inquire what fraction of these interactions fit the “classic” π -cation and π - π definitions. Thus, we reanalyzed our data with a different approach: we surveyed the lists of interactions used for map generation by: 1) filtering for those interactions with their midpoint within the cylinders, 2) filtered that set for interactions involving one of the aromatic atoms of the residue, and then, 3a) for the favorable polar interaction set, filtered for interactions with HZ1, HZ2 and HZ3 of a LYS or with HE, HH11, HH12, HH21 and HH22 of an ARG (i.e., the cationic protons), or 3b) for the favorable hydrophobic interaction set, filtered for interactions with one of the aromatic atoms of another PHE, TYR or TRP. **Table 2.5** tabulates this data in two ways: 1) by score, i.e., the fraction of the total favorable hydrophobic interactions attributable to “true” π - π interactions and the fraction of the total favorable polar interactions attributable to “true” π -cation interactions (both as defined above); and 2) by presence, i.e., whether interactions meeting those criteria were found in the cylinders above or below the ring. While π - π interactions are found in a large fraction (between one-third and more than one-half) of the cylinders for the three aromatic residues, energetically they contribute less than 5%. Similarly, the presence of π -cation interactions (~8-27% of the residue sidechain cylinders) is much larger than their ~1% contribution to score (i.e., energetics). As also reported by Gallivan and Dougherty,⁹ we found that TRP is the most likely residue to form π -cation interactions, i.e., 27% (**Table 2.5**) of total TRPs in our dataset seem to engage in this interaction while they report 26%. The very recent study by Vernon et al.¹⁰ more than qualitatively agrees with our assessment of π - π interactions: our fractions by presence counted above and below the ring cylinders as separate occurrences – their

reported frequencies are thus about half of ours, 28%, 20% and 29%, for PHE, TYR and TRP, respectively.

Table 2.5– Analysis of cylinder character for π - π and π -cation interactions.

Residue Type	Fraction in cylinder π - π		Fraction in cylinder π -cation	
	by score	by presence	by score	by presence
PHE	0.039	0.533	0.012	0.141
TYR	0.043	0.341	0.005	0.083
TRP	0.025	0.557	0.014	0.268

2.4 Summary and Conclusion

In this chapter, we demonstrated that the hydrophobic environment surrounding an amino acid residue within a structured protein, in terms of its interactions, could be mapped, and that these maps can be clustered and averaged into a fairly limited set. Underlying each average map is an average prototype residue structure. Thus, these data provide what is truly a *backbone-dependent* library of not only sidechain rotamers, but also 3D residue interaction preferences. In other words, the presence of a feature, such as a favorable polar interaction in one of these maps, e.g., a TYR in the **b1.60** (β -pleat) chess square parse cluster **581**, where the –OH is acting as a hydrogen bond donor, should have a complementary acceptor on a neighboring residue. Accordingly, that residue’s map should have a similar feature in its hydrophobic interaction map, and the 3D alignment of these features—and others in a collection of residue-by-residue maps

–would be a hallmark of a well-organized hydrophobic interaction network. Significantly, it is not only favorable interactions that compose the hydrophobic network. Clearly, from the sampling of maps illustrated here, and in our other studies,^{40–45} unfavorable polar and especially unfavorable hydrophobic interactions are found in the maps because these interactions are found in protein structure. Some of the latter are unavoidable: even the most polar atoms on sidechains are covalently bonded to hydrophobic atoms and “background noise” of unfavorable hydrophobic interactions quite often accompanies strong favorable polar interactions. However, there are many unfavorable hydrophobic interactions that are significant and functional contributors to structure, e.g., to add flexibility, sequester water or perform other roles.

We analyzed the interaction environments of about 70,000 aromatic, i.e., phenylalanine, tyrosine and tryptophan, amino acid residues in a diverse collection of protein structures. While there is not a tremendous variety in sidechain conformations for these residues – rotamer libraries show three preferred χ_1 s and two χ_2 s –there is tremendous diversity in interactions, solvent accessibility, and their structural roles.

In the previous report,⁴¹ it was stated that full understanding of the individual environment maps for alanine would first require completing the analysis for all residue types. This current work is a status update on that task – for PHE, TYR and TRP. The remaining residues have been completed^{43–45} or are in different stages of completion and analysis, and additional studies are estimated in the near future. As with alanine, our evaluation of interactions of the aromatic residues with three-dimensional maps supports our interaction homology paradigm as a tool for exploring and perhaps

predicting protein structure. In these cases, the hydrophobic valence is much more complex, involving additional terms such as hydrophobic interactions with aromatic carbons that may be of π - π character and polar interactions that include hydrogen bonding with the OH/HH of tyrosine and the NE1/HE1 of tryptophan and aromatic atoms acting as Lewis bases. Recording these effects as we have with 3D maps enables the second tenet of the interaction homology paradigm: it is the hydrophobic “field” of the atoms surrounding a residue that directs its conformation or other properties, including secondary structure. This pattern defines a residue’s hydrophobic valence, which when represented by a numerically limited set of maps, is a conserved structural property of amino acid residues in a protein. Lastly, biological structure is clearly a delicate balance of effects, and assembling this puzzle in the absence of high-quality structural data, e.g., by homology modeling^{56–58} or *de novo* structure prediction,^{59–61} involves many pieces and interactions; our maps and strategy inherently include the special features of the aromatic residues in the model.

References:

- (1) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. Conformation of Amino Acid Side-Chains in Proteins. *Journal of Molecular Biology* **1978**, *125* (3), 357–386. [https://doi.org/10.1016/0022-2836\(78\)90408-4](https://doi.org/10.1016/0022-2836(78)90408-4).
- (2) Onuchic, J. N.; Wolynes, P. G. Theory of Protein Folding. *Current Opinion in Structural Biology* **2004**, *14* (1), 70–75. <https://doi.org/10.1016/j.sbi.2004.01.009>.
- (3) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338* (6110), 1042–1046. <https://doi.org/10.1126/science.1219021>.
- (4) Li, B.; Fooksa, M.; Heinze, S.; Meiler, J. Finding the Needle in the Haystack: Towards Solving the Protein-Folding Problem Computationally. *Critical Reviews in Biochemistry and Molecular Biology* **2018**, *53* (1), 1–28. <https://doi.org/10.1080/10409238.2017.1380596>.
- (5) Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (45), 15873–15880. <https://doi.org/10.1073/pnas.1411798111>.
- (6) Trinquier, G.; Sanejouand, Y. H. Which Effective Property of Amino Acids Is Best Preserved by the Genetic Code? *Protein Engineering Design and Selection* **1998**, *11* (3), 153–169. <https://doi.org/10.1093/protein/11.3.153>.
- (7) Chakrabarti, P.; Bhattacharyya, R. Geometry of Nonbonded Interactions Involving Planar Groups in Proteins. *Progress in Biophysics and Molecular Biology* **2007**, *95* (1–3), 83–137. <https://doi.org/10.1016/j.pbiomolbio.2007.03.016>.
- (8) Pace, C. J.; Gao, J. Exploring and Exploiting Polar- π Interactions with Fluorinated Aromatic Amino Acids. *Acc. Chem. Res.* **2013**, *46* (4), 907–915. <https://doi.org/10.1021/ar300086n>.
- (9) Gallivan, J. P.; Dougherty, D. A. Cation- π Interactions in Structural Biology. *Proceedings of the National Academy of Sciences* **1999**, *96* (17), 9459–9464. <https://doi.org/10.1073/pnas.96.17.9459>.
- (10) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. π - π Contacts Are an Overlooked Protein Feature Relevant to Phase Separation. *eLife* **2018**, *7*, e31486. <https://doi.org/10.7554/eLife.31486>.
- (11) Ferreira de Freitas, R.; Schapira, M. A Systematic Analysis of Atomic Protein–Ligand Interactions in the PDB. *Med. Chem. Commun.* **2017**, *8* (10), 1970–1981. <https://doi.org/10.1039/C7MD00381A>.
- (12) Grauffel, C.; Yang, B.; He, T.; Roberts, M. F.; Gershenson, A.; Reuter, N. Cation- π Interactions As Lipid-Specific Anchors for Phosphatidylinositol-Specific Phospholipase C. *J. Am. Chem. Soc.* **2013**, *135* (15), 5740–5750. <https://doi.org/10.1021/ja312656v>.
- (13) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate–Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 946–954. <https://doi.org/10.1021/ar300024d>.
- (14) Baker, C. M.; Grant, G. H. Role of Aromatic Amino Acids in Protein-Nucleic Acid Recognition. *Biopolymers* **2007**, *85* (5–6), 456–470. <https://doi.org/10.1002/bip.20682>.
- (15) Hong, H.; Park, S.; Flores Jiménez, R. H.; Rinehart, D.; Tamm, L. K. Role of Aromatic Side Chains in the Folding and Thermodynamic Stability of Integral Membrane Proteins. *J. Am. Chem. Soc.* **2007**, *129* (26), 8320–8327. <https://doi.org/10.1021/ja068849o>.
- (16) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* **1985**, *229* (4708), 23–28. <https://doi.org/10.1126/science.3892686>.

- (17) Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M. S.; Chakrabarti, P. CH- π Hydrogen Bonds in Biological Macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16* (25), 12648–12683. <https://doi.org/10.1039/C4CP00099D>.
- (18) Makwana, K. M.; Mahalakshmi, R. Implications of Aromatic-Aromatic Interactions: From Protein Structures to Peptide Models: Aromatic Interactions in Protein Structures. *Protein Science* **2015**, *24* (12), 1920–1933. <https://doi.org/10.1002/pro.2814>.
- (19) Wu, L.; McElheny, D.; Takekiyo, T.; Keiderling, T. A. Geometry and Efficacy of Cross-Strand Trp/Trp, Trp/Tyr, and Tyr/Tyr Aromatic Interaction in a β -Hairpin Peptide. *Biochemistry* **2010**, *49* (22), 4705–4714. <https://doi.org/10.1021/bi100491s>.
- (20) Prajapati, R. S.; Sirajuddin, M.; Durani, V.; Sreeramulu, S.; Varadarajan, R. Contribution of Cation- π Interactions to Protein Stability. *Biochemistry* **2006**, *45* (50), 15000–15010. <https://doi.org/10.1021/bi061275f>.
- (21) Dougherty, D. A. The Cation- π Interaction. *Acc. Chem. Res.* **2013**, *46* (4), 885–893. <https://doi.org/10.1021/ar300265y>.
- (22) Kennedy, C. R.; Lin, S.; Jacobsen, E. N. The Cation- π Interaction in Small-Molecule Catalysis. *Angew. Chem. Int. Ed.* **2016**, *55* (41), 12596–12624. <https://doi.org/10.1002/anie.201600547>.
- (23) Liang, Z.; Li, Q. X. π -Cation Interactions in Molecular Recognition: Perspectives on Pharmaceuticals and Pesticides. *J. Agric. Food Chem.* **2018**, *66* (13), 3315–3323. <https://doi.org/10.1021/acs.jafc.8b00758>.
- (24) Rodham, D. A.; Suzuki, S.; Suenram, R. D.; Lovas, F. J.; Dasgupta, S.; Goddard, W. A.; Blake, G. A. Hydrogen Bonding in the Benzene-Ammonia Dimer. *Nature* **1993**, *362* (6422), 735–737. <https://doi.org/10.1038/362735a0>.
- (25) Ma, J. C.; Dougherty, D. A. The Cation- π Interaction. *Chem. Rev.* **1997**, *97* (5), 1303–1324. <https://doi.org/10.1021/cr9603744>.
- (26) Feller, D.; Dixon, D. A.; Nicholas, J. B. Binding Enthalpies for Alkali Cation-Benzene Complexes Revisited. *J. Phys. Chem. A* **2000**, *104* (48), 11414–11419. <https://doi.org/10.1021/jp002631l>.
- (27) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. The Magnitude of the CH/ π Interaction between Benzene and Some Model Hydrocarbons. *J. Am. Chem. Soc.* **2000**, *122* (15), 3746–3753. <https://doi.org/10.1021/ja993972j>.
- (28) Xu, Y.; Shen, J.; Zhu, W.; Luo, X.; Chen, K.; Jiang, H. Influence of the Water Molecule on Cation- π Interaction: Ab Initio Second Order Møller-Plesset Perturbation Theory (MP2) Calculations. *J. Phys. Chem. B* **2005**, *109* (12), 5945–5949. <https://doi.org/10.1021/jp044568w>.
- (29) Marshall, M. S.; Steele, R. P.; Thanthiriwatte, K. S.; Sherrill, C. D. Potential Energy Curves for Cation- π Interactions: Off-Axis Configurations Are Also Attractive. *J. Phys. Chem. A* **2009**, *113* (48), 13628–13632. <https://doi.org/10.1021/jp906086x>.
- (30) Caldwell, J. W.; Kollman, P. A. Cation- π Interactions: Nonadditive Effects Are Critical in Their Accurate Representation. *J. Am. Chem. Soc.* **1995**, *117* (14), 4177–4178. <https://doi.org/10.1021/ja00119a037>.
- (31) Armentrout, P. B.; Rodgers, M. T. An Absolute Sodium Cation Affinity Scale: Threshold Collision-Induced Dissociation Experiments and Ab Initio Theory. *J. Phys. Chem. A* **2000**, *104* (11), 2238–2247. <https://doi.org/10.1021/jp991716n>.
- (32) Amicangelo, J. C.; Armentrout, P. B. Absolute Binding Energies of Alkali-Metal Cation Complexes with Benzene Determined by Threshold Collision-Induced Dissociation Experiments and Ab Initio Theory. *J. Phys. Chem. A* **2000**, *104* (48), 11420–11432. <https://doi.org/10.1021/jp002652f>.

- (33) Ruan, C.; Rodgers, M. T. Cation- π Interactions: Structures and Energetics of Complexation of Na⁺ and K⁺ with the Aromatic Amino Acids, Phenylalanine, Tyrosine, and Tryptophan. *J. Am. Chem. Soc.* **2004**, *126* (44), 14600–14610. <https://doi.org/10.1021/ja048297e>.
- (34) Xie, N.-Z.; Du, Q.-S.; Li, J.-X.; Huang, R.-B. Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design. *PLoS ONE* **2015**, *10* (9), e0137113. <https://doi.org/10.1371/journal.pone.0137113>.
- (35) Orabi, E. A.; Lamoureux, G. Cation- π and π - π Interactions in Aqueous Solution Studied Using Polarizable Potential Models. *J. Chem. Theory Comput.* **2012**, *8* (1), 182–193. <https://doi.org/10.1021/ct200569x>.
- (36) Orabi, E. A.; Lamoureux, G. Cation- π Interactions between Quaternary Ammonium Ions and Amino Acid Aromatic Groups in Aqueous Solution. *J. Phys. Chem. B* **2018**, *122* (8), 2251–2260. <https://doi.org/10.1021/acs.jpcc.7b11983>.
- (37) Zhu, D.; Herbert, B. E.; Schlautman, M. A.; Carraway, E. R. Characterization of Cation- π Interactions in Aqueous Solution Using Deuterium Nuclear Magnetic Resonance Spectroscopy. *J. environ. qual.* **2004**, *33* (1), 276–284. <https://doi.org/10.2134/jeq2004.2760>.
- (38) Slutsky, M. M.; Marsh, E. N. G. Cation- π Interactions Studied in a Model Coiled-Coil Peptide. *Protein Sci.* **2004**, *13* (8), 2244–2251. <https://doi.org/10.1110/ps.04702104>.
- (39) Riemen, A. J.; Waters, M. L. Design of Highly Stabilized β -Hairpin Peptides through Cation- π Interactions of Lysine and *N*-Methyllysine with an Aromatic Pocket. *Biochemistry* **2009**, *48* (7), 1525–1531. <https://doi.org/10.1021/bi801706k>.
- (40) Ahmed, M. H.; Koparde, V. N.; Safo, M. K.; Neel Scarsdale, J.; Kellogg, G. E. 3d Interaction Homology: The Structurally Known Rotamers of Tyrosine Derive from a Surprisingly Limited Set of Information-Rich Hydrophobic Interaction Environments Described by Maps: Tyrosine Hydrophobic Interaction Environments Define Its Rotamers. *Proteins* **2015**, *83* (6), 1118–1136. <https://doi.org/10.1002/prot.24813>.
- (41) Ahmed, M. H.; Catalano, C.; Portillo, S. C.; Safo, M. K.; Neel Scarsdale, J.; Kellogg, G. E. 3D Interaction Homology: The Hydrophobic Interaction Environments of Even Alanine Are Diverse and Provide Novel Structural Insight. *Journal of Structural Biology* **2019**, *207* (2), 183–198. <https://doi.org/10.1016/j.jsb.2019.05.007>.
- (42) AL Mughram, M. H.; Catalano, C.; Bowry, J. P.; Safo, M. K.; Scarsdale, J. N.; Kellogg, G. E. 3D Interaction Homology: Hydrophobic Analyses of the “ π -Cation” and “ π - π ” Interaction Motifs in Phenylalanine, Tyrosine, and Tryptophan Residues. *J. Chem. Inf. Model.* **2021**, *61* (6), 2937–2956. <https://doi.org/10.1021/acs.jcim.1c00235>.
- (43) AL Mughram, M. H.; Herrington, N. B.; Catalano, C.; Kellogg, G. E. Systematized Analysis of Secondary Structure Dependence of Key Structural Features of Residues in Soluble and Membrane-Bound Proteins. *Journal of Structural Biology: X* **2021**, *5*, 100055. <https://doi.org/10.1016/j.yjsbx.2021.100055>.
- (44) Catalano, C.; AL Mughram, M. H.; Guo, Y.; Kellogg, G. E. 3D Interaction Homology: Hydrophobic Interaction Environments of Serine and Cysteine Are Strikingly Different and Their Roles Adapt in Membrane Proteins. *Current Research in Structural Biology* **2021**, *3*, 239–256. <https://doi.org/10.1016/j.crstbi.2021.09.002>.
- (45) Herrington, N. B.; Kellogg, G. E. 3D Interaction Homology: Computational Titration of Aspartic Acid, Glutamic Acid and Histidine Can Create PH-Tunable Hydrophobic Environment Maps. *Front. Mol. Biosci.* **2021**, *8*, 773385. <https://doi.org/10.3389/fmolb.2021.773385>.
- (46) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology* **1963**, *7* (1), 95–99. [https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6).

- (47) Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (48) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: Is LogPo/w More than the Sum of Its Parts? *European Journal of Medicinal Chemistry* **2000**, *35* (7–8), 651–661. [https://doi.org/10.1016/S0223-5234\(00\)00167-7](https://doi.org/10.1016/S0223-5234(00)00167-7).
- (49) Sarkar, A.; Kellogg, G. Hydrophobicity - Shake Flasks, Protein Folding and Drug Discovery. *CTMC* **2010**, *10* (1), 67–83. <https://doi.org/10.2174/156802610790232233>.
- (50) R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL [Http://Www.R-Project.Org/](http://www.R-Project.Org/).
- (51) Fraczkiwicz, R.; Braun, W. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comput. Chem.* **1998**, *19* (3), 319–333. [https://doi.org/10.1002/\(SICI\)1096-987X\(199802\)19:3<319::AID-JCC6>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1096-987X(199802)19:3<319::AID-JCC6>3.0.CO;2-W).
- (52) Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins Application to Side-Chain Prediction. *Journal of Molecular Biology* **1993**, *230* (2), 543–574. <https://doi.org/10.1006/jmbi.1993.1170>.
- (53) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins* **2000**, *40* (3), 389–408. [https://doi.org/10.1002/1097-0134\(20000815\)40:3<389::AID-PROT50>3.0.CO;2-2](https://doi.org/10.1002/1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2).
- (54) Scouras, A. D.; Daggett, V. The Dynameomics Rotamer Library: Amino Acid Side Chain Conformations and Dynamics from Comprehensive Molecular Dynamics Simulations in Water: The Dynameomics Rotamer Library. *Protein Science* **2011**, *20* (2), 341–352. <https://doi.org/10.1002/pro.565>.
- (55) van der Kamp, M. W.; Schaeffer, R. D.; Jonsson, A. L.; Scouras, A. D.; Simms, A. M.; Toofanny, R. D.; Benson, N. C.; Anderson, P. C.; Merkley, E. D.; Rysavy, S.; Bromley, D.; Beck, D. A. C.; Daggett, V. Dynameomics: A Comprehensive Database of Protein Dynamics. *Structure* **2010**, *18* (4), 423–435. <https://doi.org/10.1016/j.str.2010.01.012>.
- (56) Laughton, C. A. Prediction of Protein Side-Chain Conformations from Local Three-Dimensional Homology Relationships. *Journal of Molecular Biology* **1994**, *235* (3), 1088–1097. <https://doi.org/10.1006/jmbi.1994.1059>.
- (57) Eisenmenger, F.; Argos, P.; Abagyan, R. A Method to Configure Protein Side-Chains from the Main-Chain Trace in Homology Modelling. *Journal of Molecular Biology* **1993**, *231* (3), 849–860. <https://doi.org/10.1006/jmbi.1993.1331>.
- (58) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved Prediction of Protein Side-Chain Conformations with SCWRL4: Side-Chain Prediction with SCWRL4. *Proteins* **2009**, *77* (4), 778–795. <https://doi.org/10.1002/prot.22488>.
- (59) Alley, E. C.; Khimulya, G.; Biswas, S.; AlQuraishi, M.; Church, G. M. Unified Rational Protein Engineering with Sequence-Based Deep Representation Learning. *Nat Methods* **2019**, *16* (12), 1315–1322. <https://doi.org/10.1038/s41592-019-0598-1>.
- (60) Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved Protein Structure Prediction Using Predicted Interresidue Orientations. *Proc Natl Acad Sci USA* **2020**, *117* (3), 1496–1503. <https://doi.org/10.1073/pnas.1914677117>.
- (61) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>.

CHAPTER 3

Computational, Biophysical, and Kinetic Approaches to Elucidate Interactions Between Pyridoxine 5'-Phosphate Oxidase (PNPO) and Dopa Decarboxylase (DDC)

3.1 Introduction

Unlike bacteria and plants, humans cannot synthesize pyridoxal 5'-phosphate (PLP) *de novo* and must rely on a salvage pathway for PLP synthesis involving pyridoxal kinase (PL kinase), pyridoxine 5'-phosphate oxidase (PNPO), and phosphatases to provide enough PLP as a co-factor to hundreds or PLP-dependent enzymes for their biological activities.¹⁻³ Vitamin B₆-dependent neonatal encephalopathy (NEE) and other vitamin B₆-dependent disorders associated with a cellular PLP deficiency can be induced by mutations in the B₆ salvage and/or PLP-dependent enzymes. Due to the fact that several apo-B₆ enzymes compete for available PLP in the brain, it cannot be assumed that PLP is distributed evenly among them. Since PLP deficiency usually manifests first in DOPA decarboxylase and kynureninase, their activities have been used to monitor B₆ deficiency.^{4,5} Excessive B₆ intake is linked to neurotoxic consequences as a result of the PLP reactive aldehyde forming aldimines (Schiff base) with amino groups on non-B₆ proteins; altering their function.^{1,6,7} PLP is found in plasma in healthy individuals at low concentrations, i.e., around 40 nmol/L.⁸ To prevent toxicity, the free PLP (unbound) content in cells is therefore maintained very low.¹ Due to the need for PLP by hundreds of enzymes during protein turnover, an unanswered issue exists! How can dozens of

competing (apo) B₆-enzymes be converted to their active holo-forms when only a low concentration of free PLP is available in the cell? Our hypothesis is that most of the PLP required by the PLP-dependent enzymes are made available directly from the salvage enzyme to the PLP-dependent enzyme, which may involve direct contact between the donor and acceptor enzymes. In this context, this chapter aims to provide a better knowledge of the interaction between the B₆-salvage enzyme PNPO, and the B₆-dependent enzyme, DDC, by employing computational and experimental approaches. Further evidence for our key premise that "direct channeling" is required for cellular activation of apo to holo B₆-enzyme is also among our primary goals.

The traditional paradigm for PLP transfer to PLP-dependent enzymes is for PLP to be released from the salvage enzymes PLKase or PNPO into the bulk solvent and then acquired by PLP-dependent enzymes. However, this proposal lacks an explanation for why **free PLP** is always reported to be at **low levels** *in vivo*. Additionally, releasing free PLP into the bulk solvent allows it to be degraded by PLP phosphatase. Moreover, labile free PLP is known to be toxic. Therefore, a second possibility is that substrates are channeled between enzymes that synthesize PLP and enzymes that are PLP dependent.^{2,9} Numerous contributions from our group and collaborators^{6,10-14} have supported the second hypothesis where PLP are more likely to be transferred directly from B₆-salvage enzymes to B₆-dependent enzymes, to produce the active holo-B₆ enzymes. In the absence of structural evidence, *in silico* studies were carried out using molecular-based docking and dynamics simulation to predict the most likely near-native complex between the PNPO and DDC, in both apo- and holo-forms.

3.1.1 PNPO allosteric characteristics and PLP binding

PLP, the active form of vitamin B₆, serves crucial physiological functions in all living organisms.^{1,15} PNPO is a flavin mononucleotide (FMN)-dependent enzyme that catalyzes the terminal step in the *de novo* biosynthesis of vitamin B₆ and is involved in the salvage route in both *Escherichia coli* (*E. coli*) and mammalian cells.^{2,22} PLP, or active form of vitamin B₆, is formed by PNPO oxidizing the 4'-hydroxyl group of PNP or the 4'-amino group of PMP at the active site (**Figure 1.3**). This catalytic process involves two electrons transfer to FMN, resulting in formation of FMNH₂ (**Figure 1.4**). These electrons are finally transported to molecular oxygen, resulting in the formation of hydrogen peroxide and the regeneration of FMN.^{2,7,12} In addition to acting as catalysis, PLP has been revealed to fulfill other biological functions, such as reactive oxygen species scavenger in plants,^{16,17} transcriptional regulator in Eubacteria,¹⁸ and virulence factor in *Helicobacter pylori* (*H. pylori*) and *Mycobacterium tuberculosis* (*M. tuberculosis*).^{19,20} PNPO is known to have a strong affinity for PLP at a secondary binding site (allosteric site) distinct from the active site.^{1,10,21} *In vitro* experiments revealed that this tightly bound PLP is solvent-protected and easily transferred to apo PLP-dependent enzymes.^{11,22} In addition, it has been proven that, at least in *E. coli*, the inhibition of PNPO is induced by PLP binding to the allosteric site.²³ However, structural position for the PLP allosteric-bound site has eluded researchers for decades. That is until recently, when the Contestabile group at the University of Rome, Sapienza was able to characterize the allosteric PLP binding site of *E. coli* PNPO using site-directed mutagenesis and crystallographic studies (**Figure 3.2**).²⁴ A critical component of our investigation in this chapter is to test the hypothesis that PLP

transfer to apo B₆-enzymes involves channeling from this recently characterized allosteric PLP binding site.

3.1.2 Hypotheses and research plan

Dopa decarboxylase,²⁵ which has been determined in both the apo²⁶ and holo²⁷ forms, could be used as a starting point to obtain a better understanding of the activation of B₆-enzymes. DDC is required for the biosynthesis of key neurotransmitters, i.e., dopamine, serotonin, norepinephrine, and epinephrine neurotransmitters. Our group has already demonstrated that B₆-salvage enzymes form physical complexes, in both apo and holo forms, with several B₆-dependent enzymes.¹¹ Furthermore, using kinetic studies, the complex with SHMT was found to be involved in PLP transfer and activation of apoSHMT, a B₆-dependent enzyme.¹¹ In this work, we carried out several studies including molecular modeling, biophysical, PLP transfer studies, and site directed mutagenesis studies to investigate the interactions and PLP transfer between PNPO and DDC. The overall goal of this study was to provide a better and comprehensive understanding of how the B₆-salvage enzyme (PNPO) interacts with the B₆-dependent enzyme (DDC). Interactions between PNPO and DDC were investigated using molecular modeling, biophysical, and PLP kinetic transfer studies, as discussed in the subsequent sections and listed below:

3.2 Molecular docking and dynamics simulation studies to elucidate interactions between PNPO and DDC

3.3 Biophysical and PLP transfer studies to probe interactions between PNPO and DDC

3.4 Site-directed mutagenesis and biophysical studies to validate the PNPO•DDC complex model

3.2 Molecular Docking and Dynamics Simulations to Characterize the Interactions between PNPO with Apo-and Holo-DDC

3.2.1 Introduction

Protein-protein interaction (PPI) complexes are key components of a wide range of biological events that need an atomic-level description to fully understand. Because of the nature of the formed interaction, characterizing the 3D structure of transitory and low affinity binding PPI complexes remains a difficult challenge.²⁸⁻³⁰

Both crystal structures of DDC in the apo- and holo-conformations, i.e., holoDDC (pig kidney DDC with 90% sequence similarity to human DDC; PDB: 1JS3)²⁷ and human apoDDC crystal structures (PDB: 3RCH, 3RBF, and 3RBL)²⁶ have been determined. PLP binding to the open conformation of apoDDC induces a substantial structural rearrangement, resulting in the closed (holo) conformation (**Figure 3.1**). In apoDDC, the active site is totally exposed to solvent, and the two large domains (residues 86-360) move by up to 20 Å compared to the closed (holo) conformation (**Figure 3.1**). With crystal structural knowledge of DDC in both conformations, this would provide more assistance in modeling the complex interaction between PNPO with DDC in both states, resulting in a better understanding of the nature of their association.

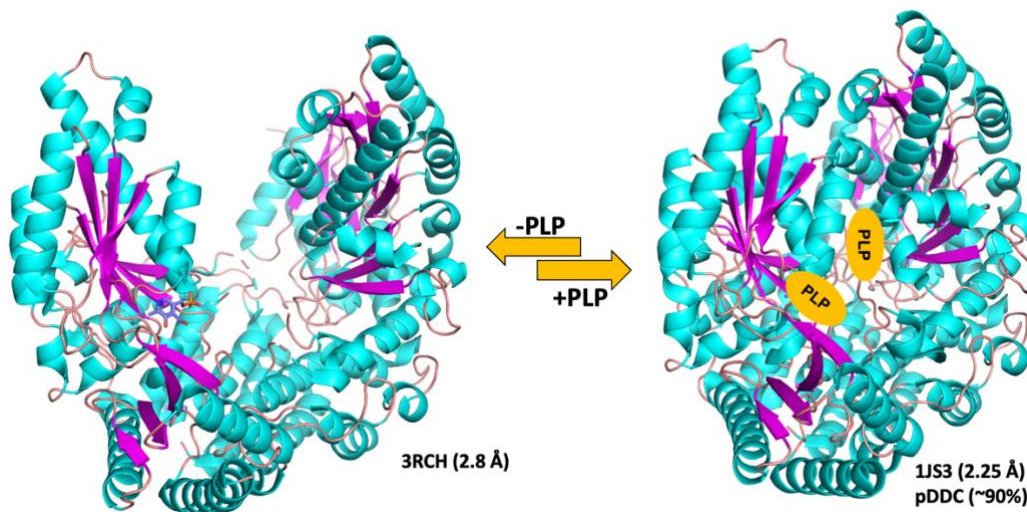


Figure 3.1 Dimeric crystal structures of apoDDC (open conformation) and holoDDC (closed conformation). β -sheet is shown in magenta, α -helix in cyan, and loops in salmon.

The crystal structure of PNPO from human, *E. coli* and yeast have been determined, and form a homodimer in which each subunit binds one FMN molecule/subunit (**Figure 3.2A**). The human PNPO enzyme has been shown to have an allosteric PLP binding site, which is found to be important for the enzyme allosteric regulation and, as a result, for the regulation of vitamin B₆ metabolism in humans. The PLP allosteric binding site (PLP tightly binding site) was recently identified crystallographically; **Figure 3.2B**), and stoichiometry of PLP tight binding revealed to be two PLP molecules per human PNPO dimer.^{24,31} The crystal structures of PNPO from human, *E. coli*, and yeast lack density in the N-terminal region, missing the first 48 residues due to disorder (**Figure 3.2A**).

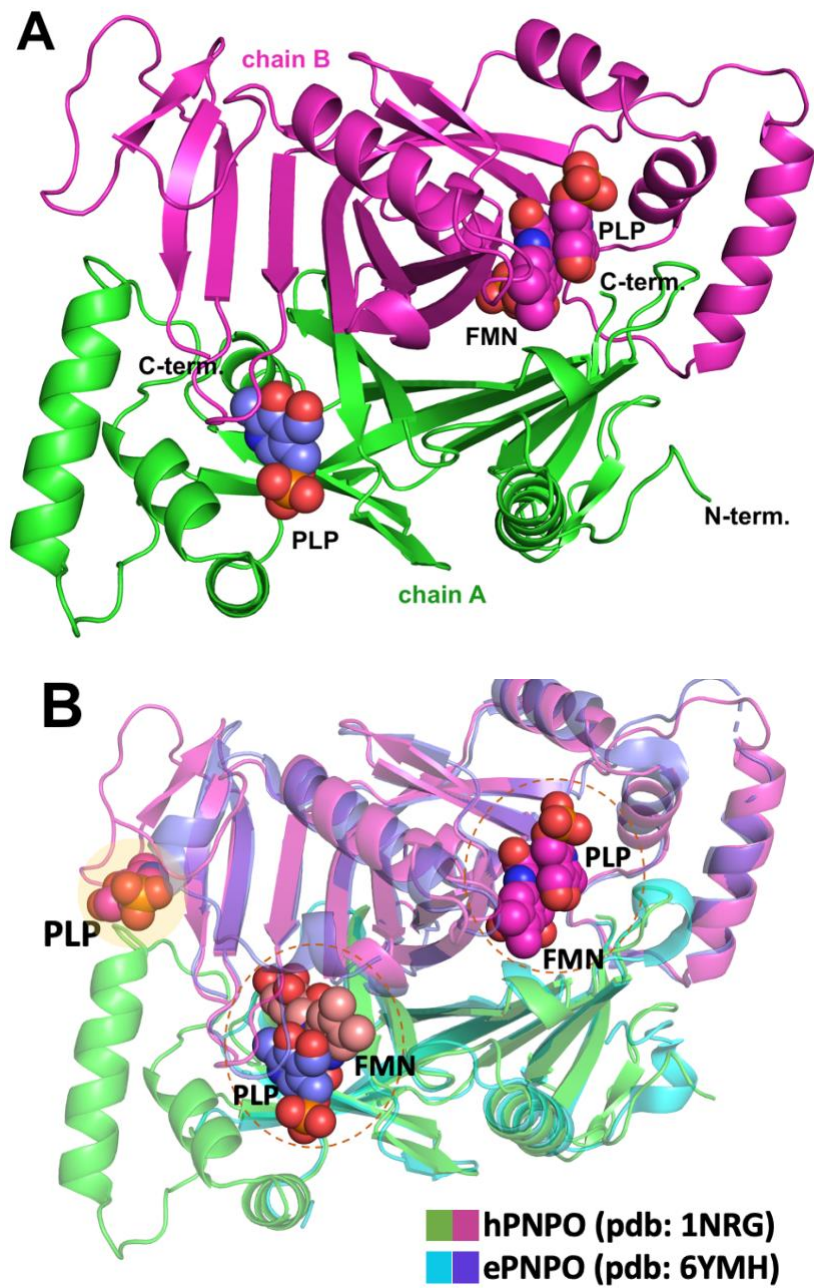


Figure 3.2 Dimeric structure of PNPO. A) human PNPO structure (PDB: 1nrg); B) superposition of human PNPO (green and magenta) and E. coli PNPO (cyan and purple).

Yellow shading indicates PLP binding to the tightly bound (allosteric) site of E. coli

PNPO.²⁴

Our group and others have investigated the interaction and activation of B₆-dependent enzymes utilizing biophysical and kinetic approaches.^{10,11,32,33} Despite the proposition that the salvage and B₆-dependent enzymes form complex for productive transfer of PLP from the former to the later, there is no report of the crystallographic complex structure or even computationally predicted complex structure of the enzyme complex. Hence, we aimed to conduct molecular docking and dynamics simulation studies to elucidate the most likely interactions between PNPO and DDC. Further, these investigations will serve as a guide to identify key residues that can be mutated and tested experimentally.

3.2.2 Methods

3.2.2A Preparation of the crystal structures of PNPO, holoDDC, and apoDDC:

All crystal structures were retrieved in pdb format from the Protein Data Bank (PDB).³⁴ MODELLER (MODELLER-II-Chimera GUI interface)³⁵ and the AlphaFold protein database³⁶ were utilized to model missing residues from the apo and holo-DDC crystal structures, with final structures prepared in Sybylx.2.1.1 (Tripos International, USA).

Gaps due to missing electron density in both holoDDC (Leu328-Gly339) and apoDDC (Gly102-Ala107, Thr323-Arg355) were modeled using MODELLER and AlphaFold, respectively. For holoDDC, a total of 100 models were generated, and best models were inspected using DOPE (Discrete Optimized Protein Energy) and GA341 scores, top 10 models were selected for further refinement using the GalaxyLoop

server,³⁷ where modeled regions were specified for refinement. MolProbity was utilized to assess the quality of models. For apoDDC, AlphaFold³⁶ prediction was found to generate the most reliable model for missing regions. Therefore, missing residues were extracted and inserted into the apoDDC crystal structure (3RCH), hydrogens were added, and final structure (3RCH-AF) was minimized in Sybylx.2.1.1 with the Tripos forcefield.

3.2.2B Molecular docking: The molecular docking study was performed with ClusPro server³⁸ that uses the PIPER docking algorithm to accomplish sampling. The receptor protein (apo- and holoDDC) is set at the coordinate system's origin, and the various rotational and translational positions of the ligand are assessed at the specified level of discretization. The interaction energy between two proteins is represented by PIPER using the expression $E = w_1E_{\text{rep}} + w_2E_{\text{attr}} + w_3E_{\text{elec}} + w_4E_{\text{DARS}}$, in which E_{rep} and E_{attr} are the repulsive and attractive contributions to the van der Waals interaction energy, respectively, while E_{elec} is an electrostatic energy factor. E_{DARS} is a pairwise structure-based potential that is built using 'decoys as the reference state' (DARS). From previous step, 1000 rotation/translation combinations with the lowest scores are considered, which would be clustered with a C-alpha RMSD radius of 9 Å, resulting in 30 clusters for the 1000 poses with the lowest energy. From the preceding stage, 1000 rotation/translation combinations with the lowest scores are considered, which are clustered using a C-alpha RMSD radius of 9 Å, resulting in 30 clusters for the 1000 poses with the lowest energy.

3.2.2C Molecular dynamics simulations: The chosen complexes of PNPO•apoDDC and PNPO•holoDDC were refined using an all-atom Molecular Dynamics (MD) simulation. MD simulations were performed using the NAMD 2.9 package developed by the Theoretical and Computational Biophysics Group at the University of Illinois in Urbana-Champaign's Institute for Advanced Science and Technology.³⁹ The force field used was CHARMM (Charmm-36).⁴⁰ VMD 1.9.3 was used to prepare the system for MD simulations.⁴¹ Using the psfgen module, coordinate (.pdb) and connectivity (.psf) files were generated for the final complexes followed by solvation in a cubic box of TIP3P water molecules with a minimum spacing of 9 Å between the box boundaries and the nearest solute atoms. The solvated system was then ionized with 0.15 M NaCl using the VMD autoionize plugin. All MD simulations were conducted in five sequential steps: i) water minimization, ii) entire system minimization, iii) heating stage, iv) NPT pre-production simulation, and v) production simulation. A 2000-step of conjugate gradient minimization was applied on solvent molecules to ensure that water molecules were distributed evenly throughout the complex's surface, which was then followed by 20,000 steps of energy minimization of the entire system. Subsequently, gradual heating to 310 K was performed by over a duration of 100 ps. In the first two rounds of equilibration, harmonic restraints were applied using restraint force constants of (5 kcal/(mol·Å²)) under isothermal isobaric (NPT) ensemble for 1 ns, placed on proteins and the cofactor, in case of holoDDC, and all other parameters remaining unchanged. In the following stage, the harmonic constraint was released, and the whole system was equilibrated for a further 1 ns using the NPT ensemble. The final production run was performed under the NPT ensemble, which simulated the whole system for 20 ns. All

simulations were run on our GPU-based compute server, and molecular dynamic trajectories were analyzed using VMD, HINT score,^{42,43} Bio3D⁴⁴, and MDAnalysis packages.⁴⁵ All figures were generated using PyMOL (Version 2.0 Schrödinger, LLC).

3.2.3 Results and Discussion

3.2.3A Molecular docking and scoring: Protein-Protein docking is a readily available method for predicting the most likely interaction between two proteins. In this work, we used a blind docking approach to sample the near-native complex between PNPO and DDC. As described in the Methods section (**3.2.2B**), docking solutions were clustered into distinct structures, with the assumption that large clusters would have near-native structures. In the case of apoDDC, the three available crystal structures were docked to PNPO and showed no noticeable differences amongst them; therefore, the structure with the best resolution, 3RCH (2.85 Å), was chosen for this study. Molecular docking of PNPO to DDC was conducted using the PDB structures of apo- and holo-DDC as well as structures with the missing residues added. **Figures 3.3** and **3.4** represent the docking results of the PNPO•apoDDC complex with the structures 3RCH and 3RCH-AF, respectively. Importantly, our docking studies clearly suggested that there is no direct interaction between the active sites of the two proteins (**Figure 3.3C**), indicating that a direct transfer of PLP from the active site of PNPO is computationally unfeasible. Interestingly, both apo and holo formed a multimeric interface with PNPO at the PLP allosteric binding site, i.e., clusters 0 and 1 in case of PNPO•apoDDC (Site A),

and for PNPO•holoDDC, clusters 0, 1, 4, and 6 (PLP site A), as well as clusters 2, 3, 5, 7, 8, and 9 (PLP site B). Importantly, PNPO was found to recognize just one site of the asymmetric dimeric structure of apoDDC (**Figure 3.3D-E**), whereas holoDDC (symmetric dimer) revealed a favorable binding for both PLP sites (**Figure 3.6B**). This observation can explain the partial activation of apo B₆-enzymes (apoSHMT), as seen in previous work,^{10,11} and apoDDC in this work.

The modeling of missing residues in both holo- and apo-DDC was found to be necessary for the improvement of protein-protein interface sampling, as seen in **Figure 3.5C and 3.6**). In the case of PNPO•holoDDC, cluster 0 and 1 account for 15.2% (of total generated solutions) after modeling missing residues (**Figure 3.6F**), compared to only 11.3% when docking to the crystal structure (**Figure 3.6E**). Similarly, the HINT scores for cluster 0 and 1 were improved from 2222 and 2401 to 4746 and 4756, respectively, without affecting the generated interface (**Figure 3.6**). In the same manner, all crystal structures of apoDDC were missing a significant portion on loop3 (residues: Thr323-Arg355). Several approaches were carried out to obtain a suitable model for the missing region., The AlphaFold³⁶ structure from the AlphaFold protein structure database was found to be the most realistic model for this missing portion of the protein (**Figure 3.4A**). Therefore, missing residues were extracted from the AF model and inserted into the crystal structure of apoDDC to create the final structure (**3RCH-AF**), a model with an entire amino acid chain. Docking PNPO to the 3RCH-AF model revealed a significant improvement in the formed interface and shape complementarity, and consequently, an improvement in the HINT scores of complex interface residues, i.e., from 2032 to 8650 for cluster 0 (**Figures 3.4C**). One of the objective of this study was to determine whether

this interaction would include PLP transfer from the active site of PNPO or the recently discovered PLP allosteric site.²⁴ Our modeling studies strongly support the latter hypothesis, specifically when superimposing crystal structures of PNPO and holoDDC on the top candidate PNPO•holoDDC complex model, we observed a strong alignment of the co-crystallized phosphate molecules (at the surface) on both PNPO (at the PLP allosteric site) and holoDDC. This could be the first site of recognition for PLP transfer from the B₆-salvage enzyme PNPO (**Figure 3.7**). At this stage, the best models for PNPO•holoDDC and PNPO•apoDDC complexes were chosen for further refinement and analysis using MD simulation

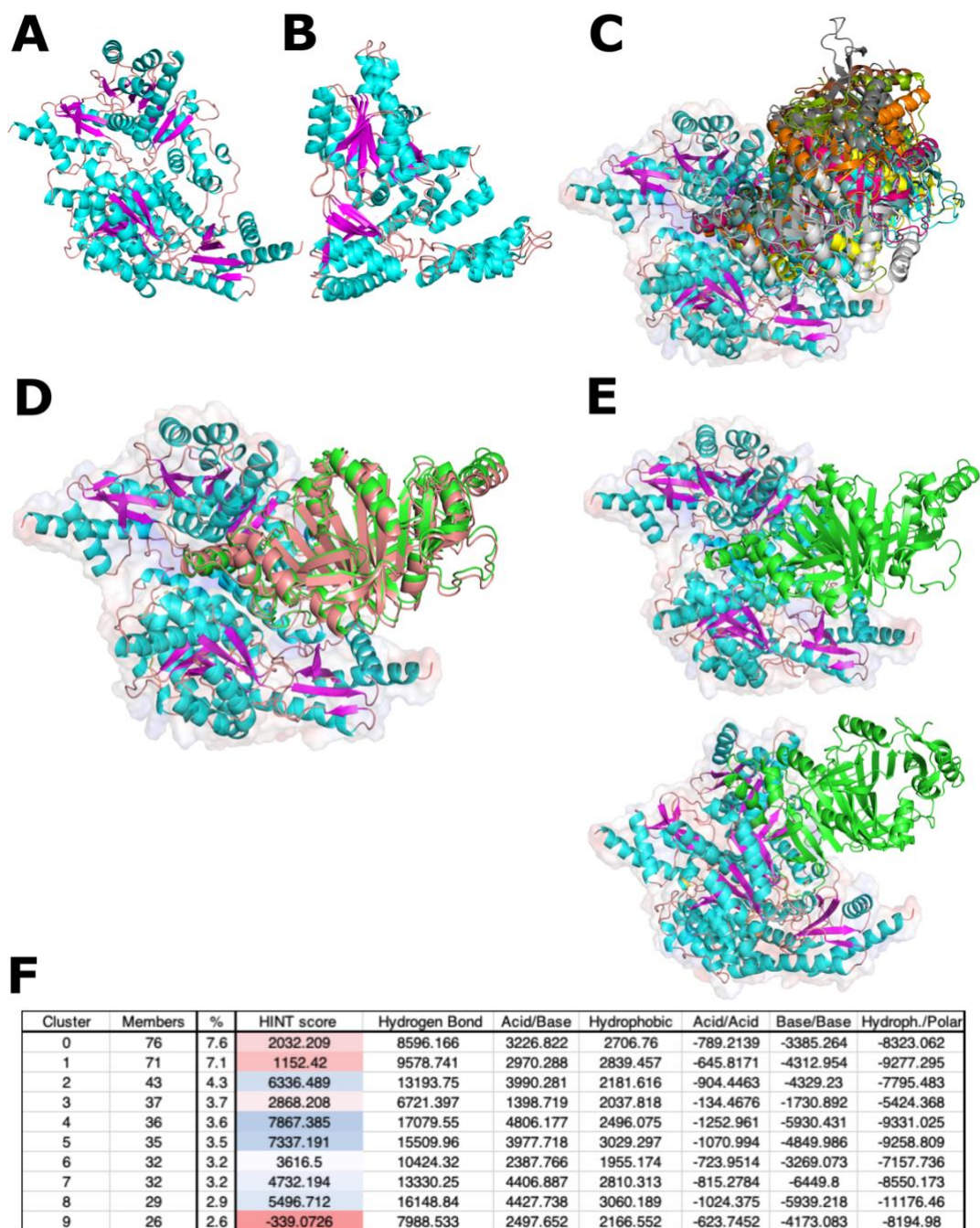


Figure 3.3 Cluspro2.0 docking results for PNPO and apoDDC complex. A) A dimeric crystal structure of apoDDC (PDB: 3RCH); B) A superposition of the two asymmetric monomers of apoDDC (RMSD= 1.16 Å); C) The top 10 cluster models of the PNPO•apoDDC complex are displayed; D) Top two clusters (most populated) of PNPO docked into apoDDC; E) The top model of PNPO-

apoDDC complex of both posterior (top) and inferior (below) sides; F) Tabulated results for the top 10 clusters using Cluspro2.0, along with calculated HINT scores.

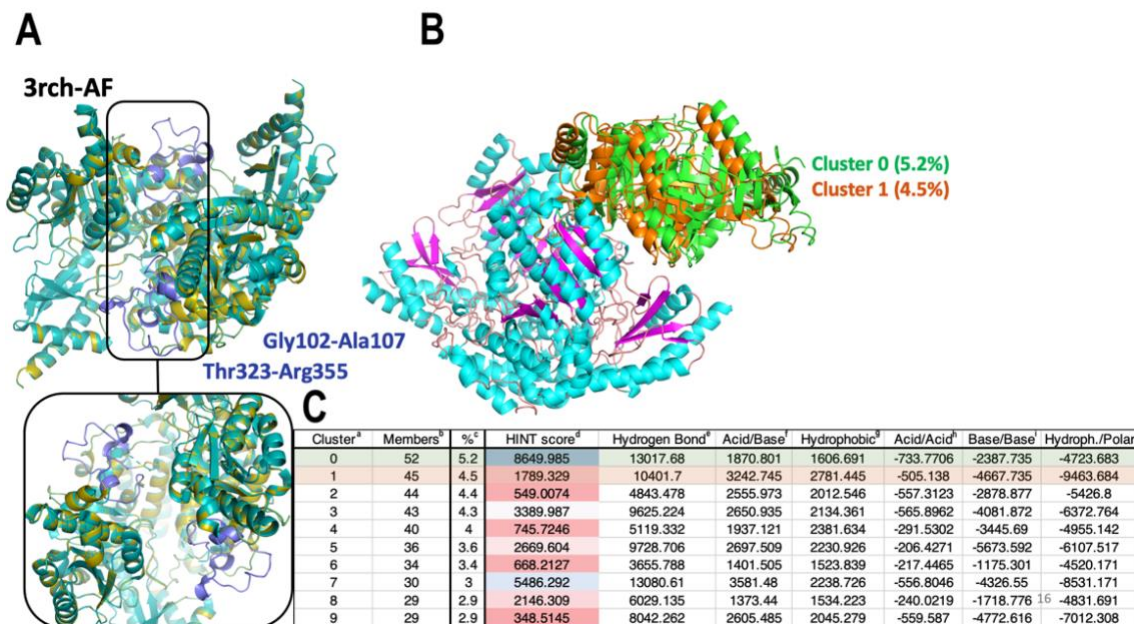


Figure 3.4 Cluspro2.0 docking findings for PNPO and apoDDC complex including missing residues modeled (3RCH-AF). A) Superimposition of the apoDDC structure with modeled missing residues (3RCH-AF) and the apoDDC crystal structure (PDB: 3RCH); B) Model for top 2 clusters (most populous) docked into 3RCH-AF model; and C) Table displaying top 10 clusters, cluster sizes, and HINT scores.

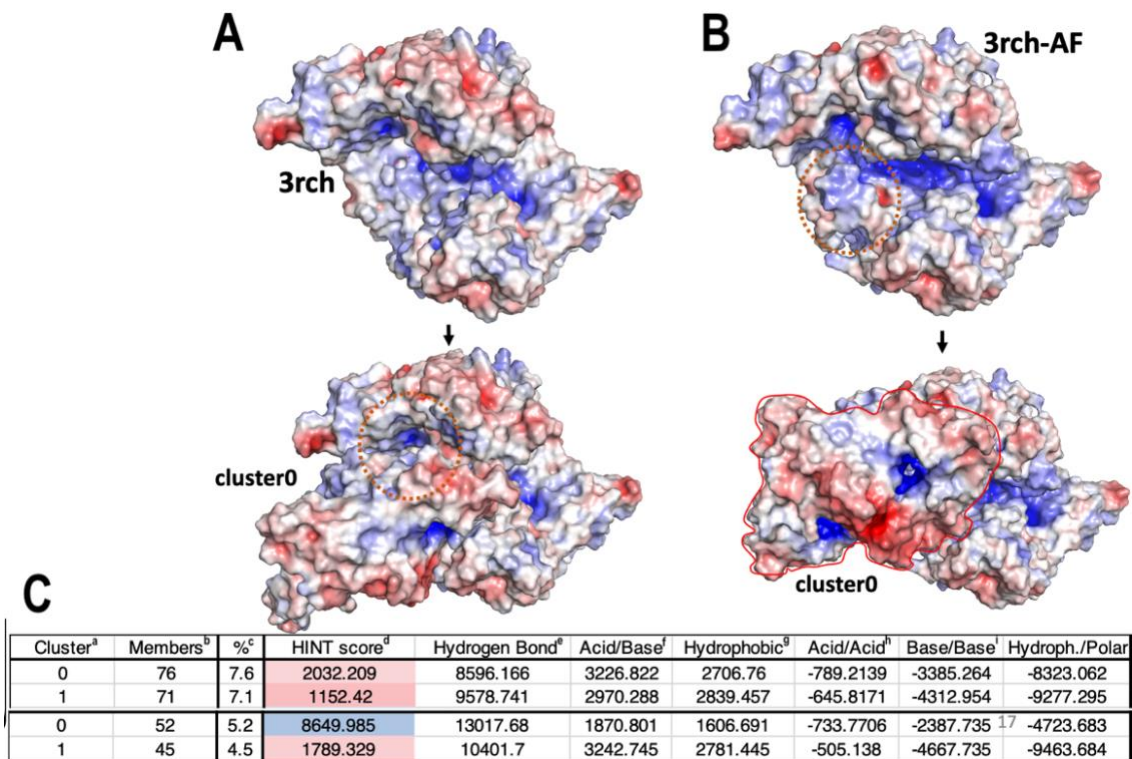


Figure 3.5 Electrostatic map comparison for the top cluster of PNPO•apoDDC complex. A) PNPO docked into the apoDDC crystal structure; B) PNPO docked into the apoDDC model with missing residues modeled; and C) Clustering and score comparison of the top two clusters before (above) and after (below) modeling missing residues.

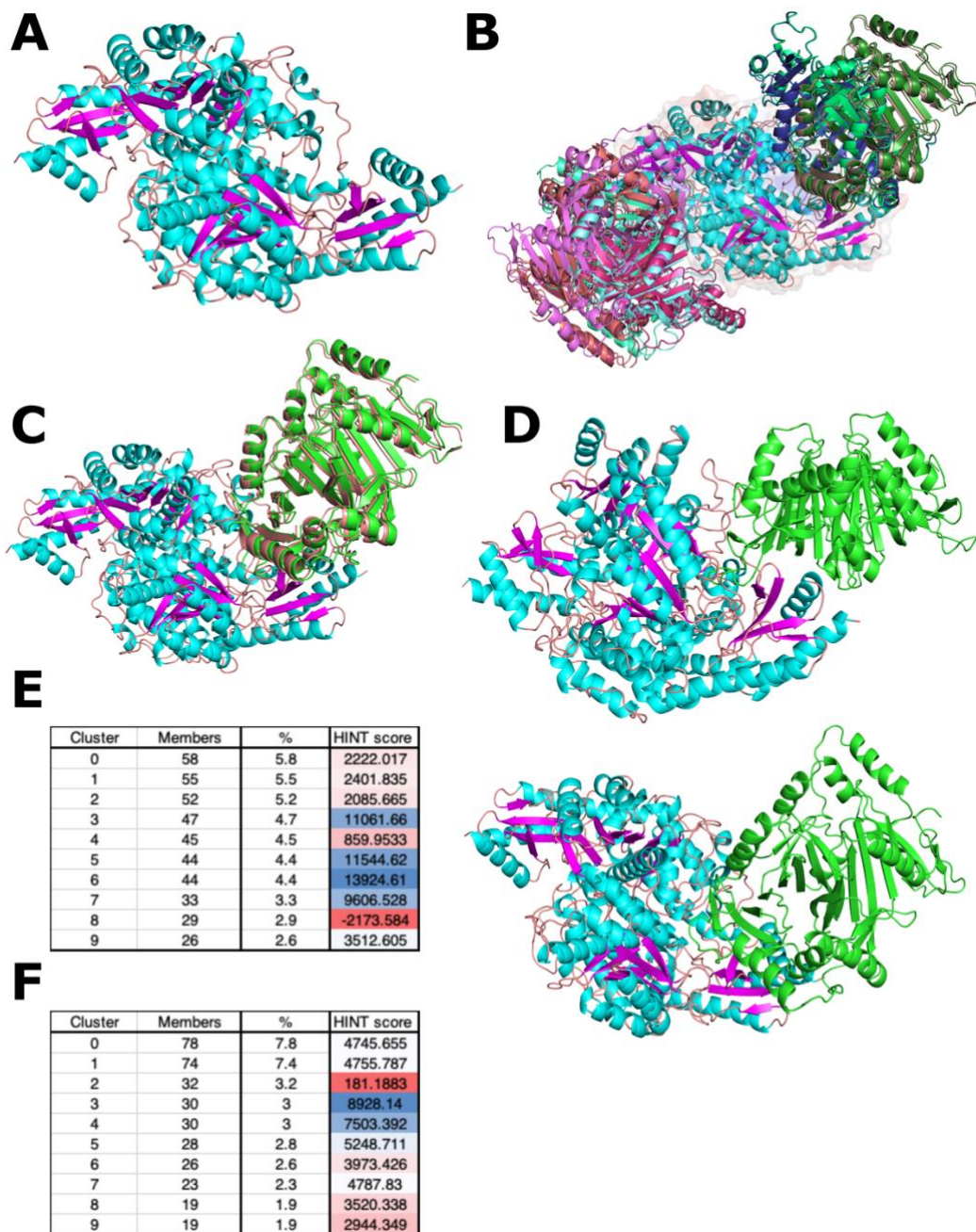
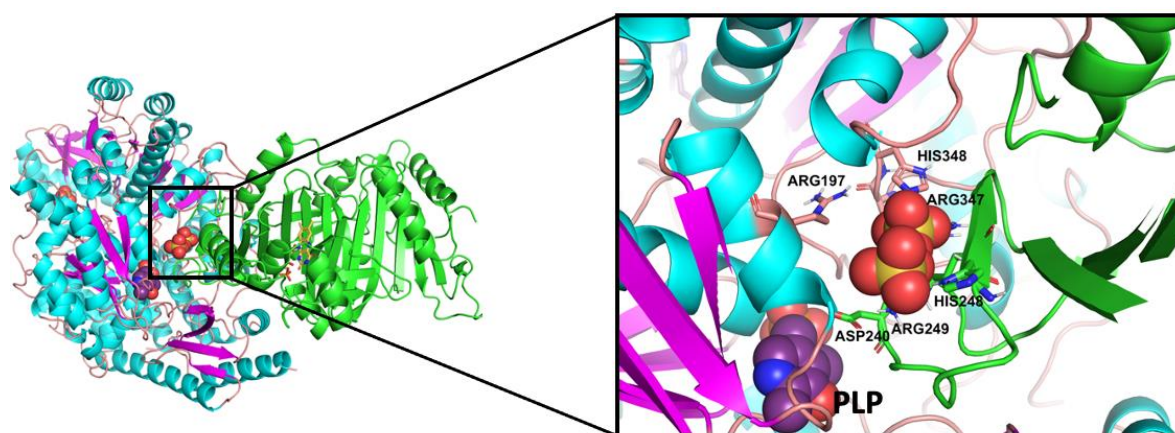


Figure 3.6 Comparison of docking results for PNPO and holoDDC complex with the crystal structure (PDB: 1JS3) and missing residues modeled (PDB: 1JS3-model). A) A dimeric crystal structure of the holoDDC enzyme (PDB: 1JS3); B) The top 10 clusters and calculated HINT scores for the PNPO•holoDDC complex models; C) PNPO•holoDDC complex model for most populated clusters , i.e., clusters 0 (green) and 1 (orange); D) The top PNPO•holoDDC complex model of

both posterior (top) and inferior (below) sides; E-F) Tabulated results for top 10 clusters using Cluspro2.0, along with calculated HINT scores, for PNPO docked into holoDDC crystal structure (E) and holoDDC with missing residues modeled (F).



Arg347Gln: Aromatic L-amino acid decarboxylase deficiency (AADCD)

Figure 3.7 Superposition of crystal structures (PNPO and holoDDC) and the PNPO•holoDDC complex model. Superposition of the crystal structures of PNPO (PDB: 6H00) and holoDDC (PDB: 1JS3) on the molecular model of the PNPO•holoDDC complex reveals the alignment of the two co-crystallized phosphate molecules. The box illustrates a close-up view of the interacting amino acids with the phosphate moiety. The amino acids ARG197, HIS348, and ARG347 (salmon) can be seen interacting with the phosphate moiety in holoDDC, whereas the corresponding amino acids in PNPO are ASP240, HIS248, and ARG249 (green).

3.2.3A Molecular dynamics simulations: an All-atom Molecular Dynamics (MD) simulations were used to refine the selected PNPO•apoDDC and PNPO•holoDDC complexes. MD is a frequently used computational approach that integrates Newton's equations of motion. It can sample the conformational space of a molecular system

reliably, allowing time-dependent behavior and evolution studies. The ability to calculate thermodynamic and kinetic parameters of molecular systems is particularly useful for examining local motion and conformational changes in proteins, DNA, and other biological systems. Because biological systems are usually large, simulation time has usually been restricted to nanoseconds. All molecular modeling simulations were performed using NAMD 2.9 as previously described in the Methods section (3.2.2C). To measure the quality and convergence of the MD trajectories, backbone root mean square deviation (RMSD) and pairwise RMSD values relative to the initial structures were determined as shown in **Figure 3.8**.

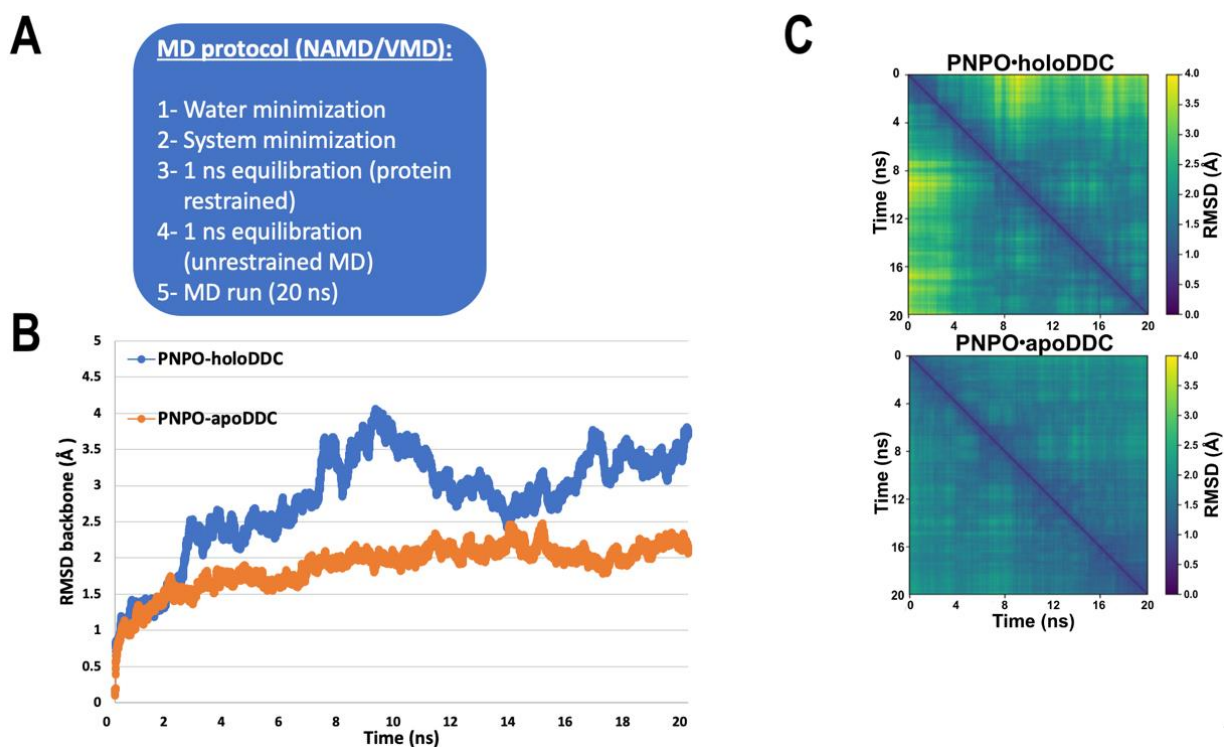


Figure 3.8 MD simulation of PNPO•holoDDC and PNPO•apoDDC complexes. A) Overview of MD simulation steps in this study; B) RMSD analysis (Å) for the PNPO•holoDDC (blue) and PNPO•apoDDC (orange) complexes during MD simulation; and C) heat maps representing

pairwise RMSD (Å) calculated for the PNPO•holoDDC (above) and PNPO•apoDDC (below) complexes.

Both complexes were shown to be well maintained over the duration of the 20 ns simulation, i.e., following an increase in RMSD within the first 2 ns, MD trajectories stabilized with average RMSD values of 1.96 ± 0.21 Å and 2.99 ± 0.48 Å for PNPO•apoDDC and PNPO•holoDDC, respectively (**Figure 3.8**). Hydrogen bonds and salt bridges at the protein-protein interface for both complexes were also computed and analyzed using VMD1.9.3 (**Figure 3.9** and **Figure 3.10**). Stable salt bridges were observed during the stimulation between the residues PNPO–E114 and apoDDC–K207, PNPO–R88 and apoDDC–D442, as well as PNPO–R88 and apoDDC–E421, with distances ranging between 2–4 Å, 3–5 Å, and 3–6 Å, respectively (**Figure 3.9**). Compared to PNPO•apoDDC complex, the PNPO•holoDDC complex revealed less stable salt bridges, i.e., PNPO–Lys91 and holoDDC–E425, and PNPO–D253 and holoDDC–R347 (**Figure 3.10**).

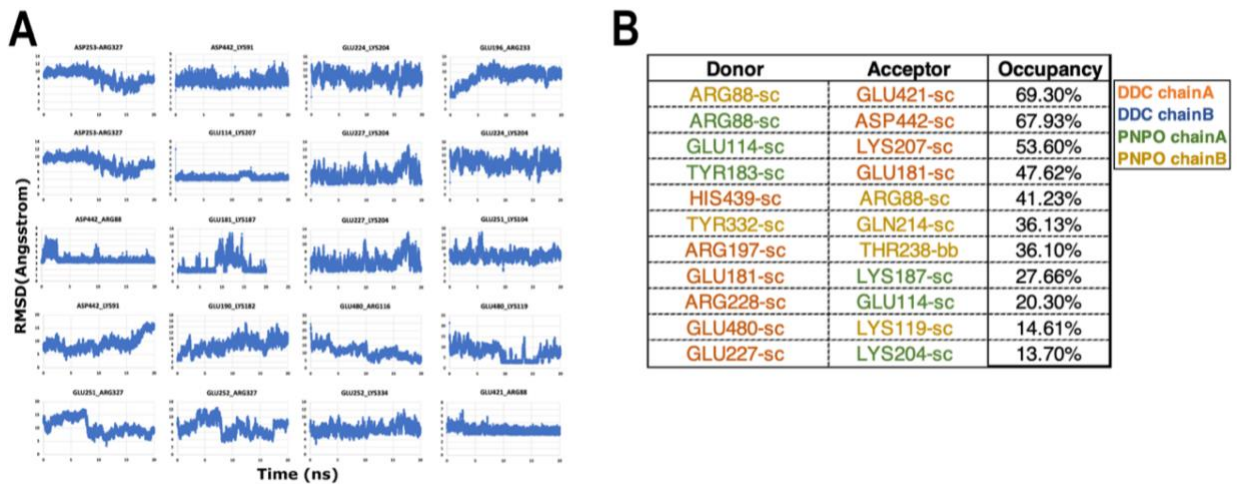


Figure 3.9 Interaction profile analysis of hydrogen bonds and salt bridges at the domain interface for the PNPO•apoDDC complex model. A) Details of the observed salt bridges (Å) at the domain interface of PNPO and apoDDC during MD simulation; B) details of the probability of forming hydrogen bonds between residues at the domain interface of PNPO and apoDDC during the MD simulation.

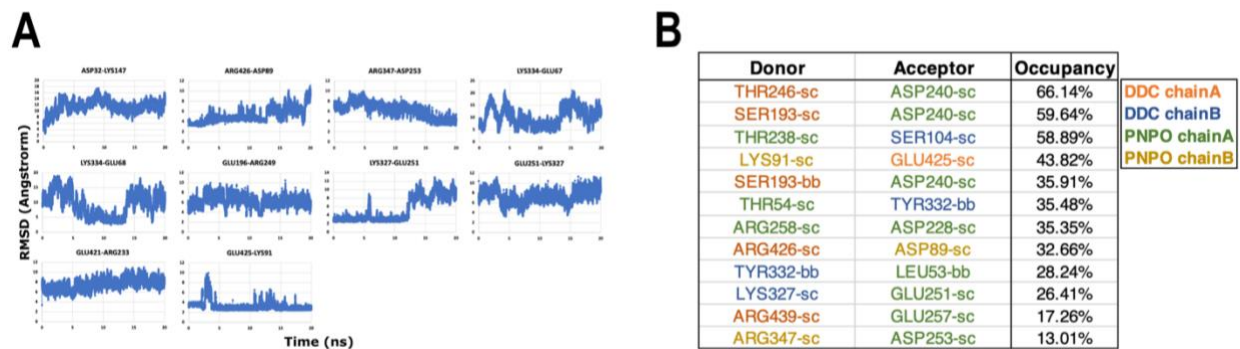


Figure 3.10 Interaction analysis of hydrogen bonds and salt bridges at the domain interface for the PNPO•holoDDC complex model. A) Details of the observed salt bridges (Å) at the domain interface of PNPO and holoDDC complex during MD simulation; B) details of the probability of forming hydrogen bonds between residues at the domain interface of PNPO and holoDDC complex during the MD simulation.

MD trajectories of the last 10 ns (500 frames) of the PNPO•apoDDC complex were clustered into unique structural conformations (**Figure 3.11A**) using the RMSD dendrogram with clustering annotation (Bio3D package in R).⁴⁴ This allows the energetic contributions of interfacial residues to be evaluated using the HINT score function (**Figure 3.11B**), as well as *in silico* alanine mutagenesis studies to estimate the impact of interfacial residues alanine mutations on the binding free energy ($\Delta\Delta G$) of the complex. Alanine-scanning mutagenesis is an effective and common approach for identifying hotspot residues: important contributions are made by key residues whose replacement with alanine results in a binding energy loss of $\Delta\Delta G \geq 1$ kcal mol⁻¹.^{46,47} **Figure 3.11** revealed that E114 and R88 from PNPO seemed to have strong interactions with K207/R228 and E421/D442 from apoDDC, respectively. Also, *in silico* alanine scanning (Robetta server)⁴⁷ showed that Arg88 and Glu114 would have the largest impact on stability of the complex (**Figure 3.11C**), i.e., the average $\Delta\Delta G_{\text{calc}}$ values were 2.35 ± 1.23 kcal mol⁻¹ and 1.1 ± 0.79 kcal mol⁻¹ for PNPO–R88 and PNPO–E114, respectively. These residues are possible candidates for site-directed mutagenesis experimental testing **Figure 3.12** displays the secondary structure of the PNPO•apoDDC complex model, including the putative salt bridges formed between PNPO–R88 and apoDDC–E421/apoDDC–D442, and PNPO–E114 and apoDDC–R228 (initial structure).

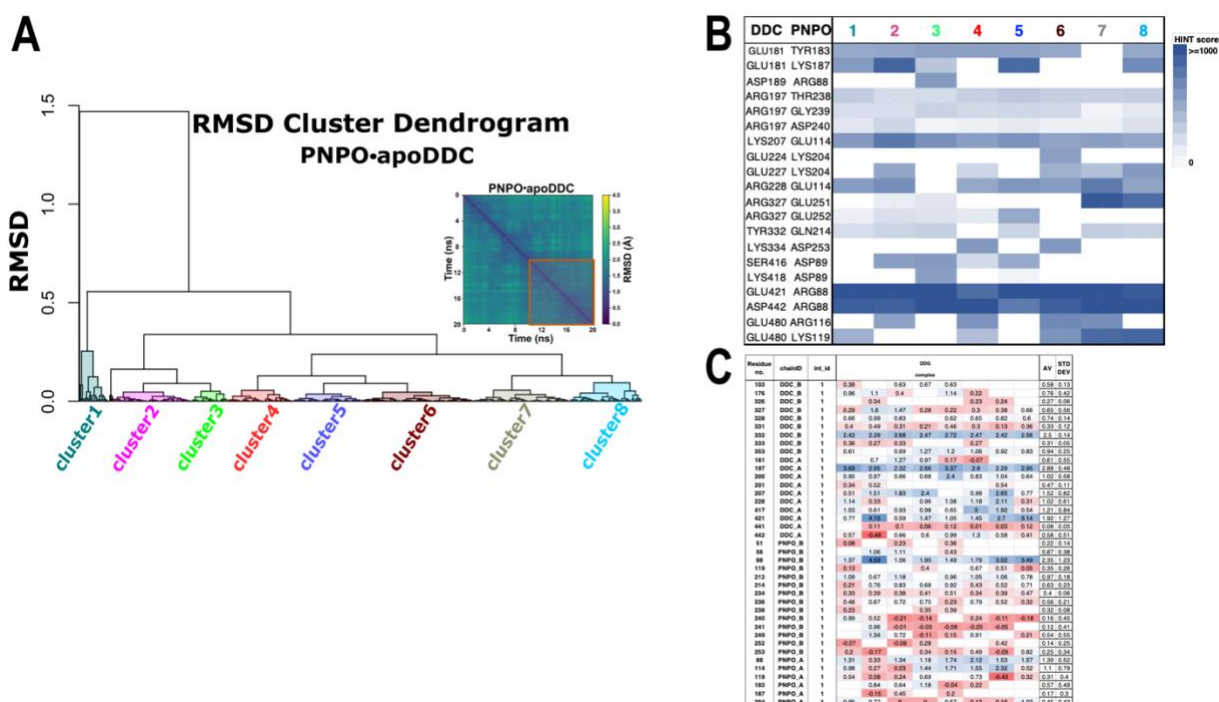


Figure 3.11 Hierarchical cluster analysis of MD trajectories for the PNPO-apoDDC complex. A) Hierarchical clustering analysis for the last 10ns of trajectories (500 frames); B) energetic scoring (HINT score) of key residues at the interface domain of the 8 distinct structures obtained by clustering; and C) binding energy changes ($\Delta\Delta G = \Delta G_{\text{wild-type}} - \Delta G_{\text{ALA}}$) for interface residues of the 8 unique structures obtained by clustering, where negative $\Delta\Delta G$ values (kcal mol^{-1}) indicate unfavorable replacement.

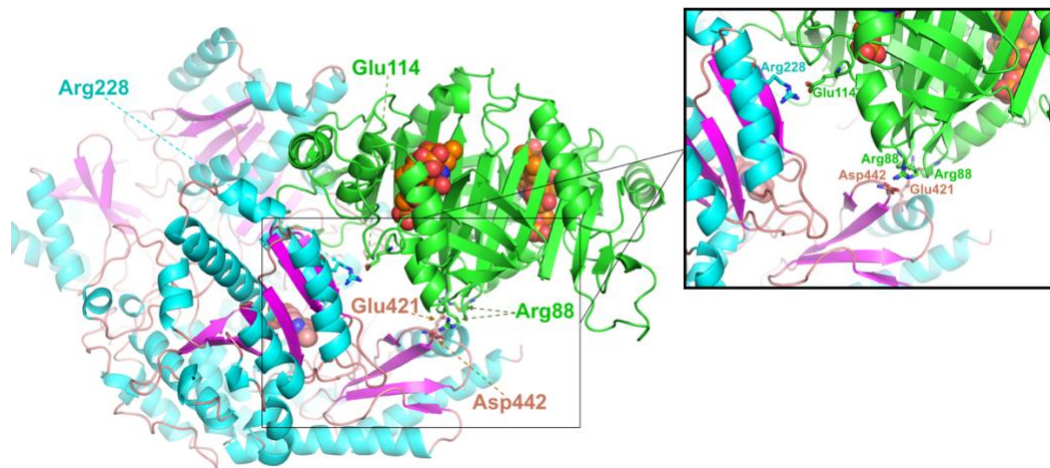


Figure 3.12 Key interactions between PNPO and apoDDC as predicted by our model. Key salt bridges formed between PNPO and apoDDC, as indicated by HINT score and *in silico* alanine mutation scanning analysis.

3.2.4 Conclusion

Our *in silico* studies showed that PNPO can form a favorable complex with both apo- and holo DDC. Interestingly, the model showed that the binding of PNPO to apo- or holo-DDC does not involve the active site of PNPO at the complex interface, but rather involves the PNPO allosteric PLP tightly binding site of PLP, which was recently characterized.²⁴ The HINT scores for the two complexes indicated that the PNPO•apoDDC complex had a stronger affinity than the PNPO•holoDDC complex, with values of 8650 (~16.8 kcal mol⁻¹) vs 4746 (~9.21 kcal mol⁻¹), respectively.^{48,49} The molecular docking investigations also suggested that PNPO predominantly recognizes one PLP site of the apoDDC dimeric structure (asymmetric dimeric), rather than both sites, as observed for the holoDDC dimeric structure (symmetric dimer). This might explain the consistent observation of partial activation (30-40%) of apo B₆-dependent enzymes during PLP transfer experiments.¹¹ Further analysis, including MD trajectories clustering, HINT scoring, and *in silico* alanine scanning mutagenesis analysis, proposed some key putative interacting residues, i.e., R88–PNPO and E421–apoDDC/D442–apoDDC as well as E114–PNPO and R228–apoDDC/K207–apoDDC. These residues can be mutated to ALA and tested experimentally for their ability to affect the complex stability.

3.3 Biophysical and PLP Transfer Studies to Probe Interactions Between PNPO and DDC

3.3.1 Introduction

PLP transfer studies and biophysical techniques, including fluorescent polarization (FP) and microscale thermophoresis (MST), have been employed by our group to investigate mechanism of PLP transfer from PNPO or PLKase and B₆-salvage enzymes.¹¹ Apo serine hydroxymethyltransferase (apoSHMT) has been the standard B₆-dependent enzyme used for studying PLP transfer activity due to its stability and relatively simple and straightforward assay, in which the formation of holo-SHMT can be directly monitored by the formation of the abortive holo-SHMT•Gly•tetrahydrofolate complex at 495 nm.^{1,10,11,32} In the previous chapter, *in silico* studies demonstrated that PNPO and DDC (apo- and holo-form) can form a favorable complex, allowing us to predict key residues that may have significant contributions in the formation of the complex. This chapter, focuses on experimental characterization of this complex interaction, as well as the ability of this biological association to activate apoDDC into its holoDDC form. All proteins were expressed and purified recombinantly, as described in the Methods section.

3.3.2 Methods

3.3.2A Subcloning, expression and purification of DDC: DDC was produced by subcloning the coding sequence of the human wild-type dopa decarboxylase (DDC-OHu25359, GenScript) into the pET21d(+) plasmid between the *NcoI* and *HindIII* restriction sites. The construct was confirmed by sequence analysis. The plasmid was inserted into BL21(DE3)pLysS competent cells of *E. coli*. The recombinant cells were grown in LB broth with ampicillin (100 µg/mL) at 37 °C until the OD600 was around 0.6-0.7, then induced with 100 µM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were grown at 30 °C for a further 14 hours before being harvested and resuspended in a lysis buffer containing 300 mM NaCl, 10 mM imidazole, 50 µM PLP, and 0.2 mM PMSF in 50 mM sodium phosphate pH 8. The clear lysate was diluted and loaded on a 5 mL HisTrap column equilibrated with 10 column volume of Buffer A (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, and 10 mM imidazole). Following, a linear gradient of buffer B (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, and 1 M imidazole) was applied (0-100% in 500 ml). DDC eluted as a single peak from the Ni²⁺ resin with imidazole at concentrations ranging from 100 to 300 mM. The purified DDC was dialyzed against 150 mM NaCl and 100 µM PLP in 50 mM potassium phosphate pH 7.4 overnight. A second dialysis in the same buffer, without PLP, was performed for additional 5 hours. Purified DDC was concentrated using Amicon Ultra 10 concentrators (Millipore), and the concentration was determined using a ϵ_M of $1.3 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ at 280 nm.⁵⁰

3.3.2B Expression and purification of PNPO: PNPO subcloning, expression and purification was carried out as previously described by our group.⁷ An overnight culture was used to inoculate 6 L of LB medium containing (40 µg/mL) of kanamycin. Cells were grown at 37 °C with rotary shaking until OD₆₀₀ reached 0.6–0.7, then induced with 100 µM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were incubated at 18 °C for a further 14 hours before being harvested and resuspended in a lysis buffer containing 300 mM NaCl, 10 mM imidazole, 10 µM FMN, and 0.2 mM PMSF in 50 mM sodium phosphate pH 8. The clear lysate was diluted and loaded on a 5 mL HisTrap column equilibrated with 10 column volume of Buffer A (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, and 10 mM imidazole) after centrifugation. Following, a linear gradient of buffer B (50 mM potassium phosphate buffer pH 8, 300 mM NaCl, and 500 mM imidazole) was applied (0-100% in 500 ml). PNPO eluted as a single peak (yellow fractions) from the Ni²⁺ resin with imidazole at concentrations ranging from 100 to 300 mM. The purified PNPO was dialyzed against 150 mM NaCl and 10 µM FMN in 50 mM potassium phosphate pH 7.4 overnight. A second dialysis in the same buffer, without FMN, was performed for additional 5 hours. Purified PNPO was concentrated using Amicon Ultra 10 concentrators (Millipore), and the concentration was determined using the molar extinction coefficient of 76,760 M⁻¹cm⁻¹ at 280 nm for the holoform.⁷

3.3.2C Preparation of apoDDC: The enzyme holoDDC was incubated at room temperature for 3 hours with 5 mM hydroxylamine in 0.5 M potassium phosphate buffer,

pH 6.9,^{26,53} followed by overnight dialyzing against 150 mM NaCl, 5% glycerol in 50 mM potassium phosphate buffer, pH 7.4. Dialyzes 2 was continued for an additional 5 hours in the same buffer without glycerol. The total conversion of holoDDC to apoDDC was validated using the DDC enzymatic activity assay (see **3.3.2D**) in the presence and absence of the cofactor (PLP).

3.3.2D Activity assay of DDC: The enzymatic activity of dopa decarboxylase was determined as previously described by Sherald et al.,⁵¹ with some modifications by Charteris and John⁵², and recently reported by Voltattorni's group.^{26,53} The enzyme DDC (0.5 μ M) was incubated with 2 mM L-Dopa and 10 μ M PLP in a final volume of 250 μ L, the reaction was stopped by heating to 100 °C for 1 minute. To enable extraction of the formed dopamine, 1 mL of 2,4,6-trinitrobenzenesulfonic acid (TNBSA) at 4.3 mM and 1.5 mL of benzene were added to the reaction mixture to enable conjugation of the formed dopamine with TNBSA. The trinitrophenylamine derivative was extracted into the benzene layer with one hour of continuous shaking at 42 °C (275 RPM). The benzene layer was then collected, and the formed product measured at 340 nm.

3.3.2E Fluorescence polarization (FP): FP measurements have been utilized in analytical and clinical chemistry, as well as biomedical research to explore membrane mobility, protein domain motions, and interactions on molecular level. As previously described by our group,¹¹ human PNPO was designated as the molecule to be labeled with the fluorescent tag fluorescein 5-maleimide (FMI) (Invitrogen), which demonstrated

no inhibitory activity for PNPO. FMI is a fluorescent molecule that interacts primarily with sulfhydryls between pH 6.5 and 7.5, creating stable thioether linkages.

50 μ M PNPO was mixed with 1 mM FMI in 50 mM sodium HEPES pH 7.55 containing 150 mM KCl and 0.01% Triton. The reaction was allowed to proceed overnight at 4 °C in the dark. After centrifuging the reaction mixture, the supernatant was dialyzed against the same buffer in the dark overnight to eliminate any excess dye. Equation 1 was used to calculate the degree of labeling.

$$\frac{\text{Moles flour per mole protein}}{= \frac{A_{\text{max}} \text{ of the labeled protein}}{\epsilon \times \text{protein concentration (M)}} \times \text{DF} , \quad (1)$$

The wavelength of A_{max} is A_{495} nm, and ϵ' is the molar extinction coefficient of FMI, which is $68,000 \text{ M}^{-1} \text{ cm}^{-1}$. Vitamin B₆ enzymes (DDC and PNPO) and a control (lysozyme) were dialyzed overnight at 4 °C against 50 mM sodium HEPES buffer, pH 7.55, containing 150 mM KCl and 0.01 % Triton X-100.

The assay was carried out using a BMG Labtech CLARIOstar Microplate Reader. Excitation was set at 495 nm, and emission to 535 nm. DDC was prepared in a series of dilutions ranging from 500 μ M to 0.021 μ M and applied to the wells of a Corning Inc. 96-well round bottom polystyrene opaque plate (Model #3792), followed by 0.25 μ M FMI-PNPO. The plate was incubated at room temperature, and measurements taken at various time points at a temperature of 25 °C. All the data provided here were obtained following a four-hour incubation period. Polarization data were collected using a Microsoft excel spreadsheet interfaced to the Calorimeter. The polarization values (mP) were

plotted versus DDC and lysozyme concentrations (Control). Binding data was fitted and analyzed in Excel using the Solver algorithm. K_D values were calculated using (equation 3) on the basis that the observed polarization signal in mP is a weighted sum of the free ligand and bound ligand signals (equation 2). Binding curves of lysozyme, a non B_6 -dependent enzyme, were also obtained as a control.

$$P_{obs} = \frac{([FMI-PNPO]P_{FMI-PNPO} + [FMI-PNPO-E]P_{FMI-PNPO-E})}{([FMI-PNPO] + [FMI-PNPO-E])}, \quad (2)$$

Where P_{obs} is the observed polarization value; $P_{FMI-PNPO}$ is the polarization value of tagged PNPO; $P_{FMI-PNPO-E}$ is the polarization value of tagged PNPO with DDC when all the PNPO was complexed; and $[FMI-PNPO]$ and $[FMI-PNPO-E]$ are the concentrations of uncomplexed and complexed tagged PNPO, respectively.

$$P_{obs} = \frac{(P_0 * K_D + 2P_{max} * [E])}{(K_D + 2[E])}, \quad (3)$$

Where P_{obs} is the observed polarization value; P_0 is the polarization value in the absence of DDC; P_{max} is the maximum polarization value in presence of DDC; K_D is the dissociation constant of the interaction between the proteins and $[E]$ is the concentration of DDC in the sample.

3.3.2F Surface plasmon resonance (SPR): SPR is a biophysical technique used for studying molecular interactions including protein-protein interactions (PPIs).⁵⁴

For this purpose, Biacore T200, at the Biacore Molecular Interaction Shared Resource BMISR (Georgetown University, USA), was utilized to conduct binding experiments of PNPO and DDC. Three enzymes, apoDDC, holoDDC, and holoSHMT (rabbit), were employed as ligands to be immobilized on the CM5 chip surface using standard amine coupling chemistry. Samples were placed in flow cells (FCs) 2 to 4, with FC1 serving as a reference for FC2-4, and analyses were carried at 25 °C. The proteins to be immobilized were diluted (1:50 dilution, 0.02 mg/mL diluted concentration) in 10 mM sodium acetate buffer at pH 5.5 and immobilized onto flow cell 2 (FC2) to a level of ~12800 RU, flow cell 3 (FC3) to a level of ~10100 RU, and flow cell 4 (FC4) to a level of ~9800 RU for apoDDC, holoDDC, and holoSHMT, respectively. PBS-P (10 mM phosphate buffer, pH 7.4, 140 mM NaCl, 3 mM KCl, 0.05% v/v surfactant P20) was used as the immobilization running buffer. Based on the ligand immobilized response value, theoretical R_{max} values were calculated and presented in **Table 3.1**. The R_{max} values assume 1:1 interaction mechanism. Overnight kinetics were performed for analytes (PNPO and albumin) in the presence of PBS-P. One 15 s pulse of 1 M NaCl was injected for the surface regeneration and flow rate of all analytes were maintained at 50 μ L/min. The contact and dissociation times used for all analytes were 60 s and 300 s, respectively. The concentrations of the analytes used ranged from 100 μ M to 1.5625 μ M (two-fold dilutions), and each analyte was injected in triplicate.

Table 3.1– Theoretical max response (R_{max}) values for analytes binding using SPR.

Ligand	FC	Analyte	Ligand Binding (RU)	MWL (Da)	MWA(Da)	Stoichiometry	R_{max}
apoDDC	2	PNPO	12800	110000	51000	1:1	5934.5
		Albumin			67000		7796.4
holoDDC	3	PNPO	10100	110000	51000	1:1	4682.7
		Albumin			67000		6151.8
rSHMT	4	PNPO	9800	205000	51000	1:1	2438
		Albumin			67000		3202.9

3.3.2G Isothermal Calorimetry (ITC): Isothermal titration calorimetry (ITC) is a biophysical technique that can be used to determine the thermodynamic properties of solution interactions. ITC provides quantitative information on the binding affinity (K_a), enthalpy changes (ΔH), and stoichiometry (n) of the interaction between two or more molecules in solution.^{55,56} The Gibbs free energy changes (ΔG) and entropy changes (ΔS) may be estimated from these initial observations using the following equation:

$$\Delta G = RT \ln K_D = \Delta H - T\Delta S$$

All ITC studies were conducted using a MICROCAL PEAQ-ITC automated system housed at the Institute for Structural Biology, Drug Discovery and Development (ISB3D) at VCU. Both cell and titrant samples were prepared in identical solutions to avoid large heat of dilution. Samples were co-dialyzed in 50 mM potassium phosphate (pH 7.4)

containing 150 mM NaCl. For PNPO titration to apoDDC, apoDDC prepared as described in section **3.3.2D** was dialyzed in 50 mM potassium phosphate solution (pH 7.4) containing 150 mM NaCl and 5% glycerol for 12 hours, followed by a second dialysis for 4 hours in the same buffer without glycerol. **Table 3.2** illustrates the ITC parameters for this experiment.

Table 3.2 – Optimization of ITC parameters for PNPO•DDC titration

ITC parameters	Description
Total no. of injection	19
Cell Temperature °C	25
Reference power ($\mu\text{cal/s}$)	5
Stir speed	750
Volume (μL)	2
Duration (s)	4
Spacing (s)	150

3.3.2H PNPO-PLP Complex Preparation: In this experiment, the PNPO enzyme and PLP (Sigma-Aldrich, St. Louis, MO, USA) were incubated for 30 min. at 30 °C. The PNPO/PLP ratio was always maintained at 1:5 when preparing the complex. After incubation, the mixture was loaded onto a Sephadex G50 (0.6 X 45 cm) gel filtration column that had been pre-equilibrated with a 100 mM potassium phosphate buffer pH 7.4 solution containing 5 mM 2-mercaptoethanol, and the column rinsed with equilibration

buffer to elute any unbound PLP from the complex. The mixture was dialyzed overnight at 4 °C in the same buffer, with two buffer changes. The concentration of PNPO was determined by measuring the UV absorbance at 280 nm.

3.3.2I PNPO-PLP is able to transfer the tightly bound PLP to activate apoDDC: PLP transfer from PNPO-PLP to activate apoDDC into the active holoDDC was studied using a spectrophotometric assay. Free PLP was studied as a positive control. First, in 100 mM potassium phosphate (pH 7.4) at 25 °C, an equal amount of free PLP (50 µM) or PNPO•PLP (50 µM) was mixed with apoDDC (50 µM) at various times (2.5, 5, 10, 30, 60, and 90 min). Subsequently, the mixture was diluted (100-fold dilution) into the same solution containing 2 mM of L-Dopa (substrate) to a final volume of 1 mL, then 250 µL aliquots were collected at different time intervals (0, 2, 6, and 10 min) and resuspended in 0.1M perchloric acid solution (10% v/v) to stop the reaction. Formed dopamine (product) was determined by extracting and measuring the absorbance at 340 nm, as previously described in section **3.3.2D**. All experiments were run in at least triplicate.

3.3.3 Results and Discussion

3.3.3A Fluorescence polarization (FP): Following the computational investigation of a potential physical complex between PNPO and DDC (**Chapter 3.2**), we conducted several biophysical studies to quantify the binding affinity between the putative interacting proteins. One of the techniques used for the study was fluorescence polarization (FP). FP is a biophysical approach that enables fast and quantitative investigations of a variety of molecular interactions and enzyme activity. It is based on the change in polarization of emitted light when a fluorescent molecule is excited by plane-polarized light. If one of the two proteins is small and fluorescent, as detailed in the Methods, this approach may be used to quantify the binding and dissociation of two molecules (Section **3.3.2E**). Thus, binding of the fluorescently labeled small protein (FMI-PNPO) to the relatively large protein (DDC) can be monitored and measured through a change in polarization and an increase in the FP signal.¹¹ Degree of FMI tagging of PNPO were determined to be 1.1 moles per mole of protein using the equation 1. Our lab has previously studied the enzyme activity of human PNPO, and PNPO labeling had no effect on the catalytic activity of PNPO.¹¹ The saturation binding curves of tagged PNPO with holoDDC clearly suggested a physical interaction between the two proteins, as shown by a rise in polarization value (FP) with increasing DDC concentrations (**Figure 3.13**). Lysozyme, used a negative control showed no significant changes in the polarization of tagged human PNPO at increasing lysozyme concentrations, i.e., showed no binding to PNPO as expected. **Figure 3.13** shows the FP result of PNPO and holoDDC, which yielded a K_D of $57.4 \pm 0.32 \mu\text{M}$.

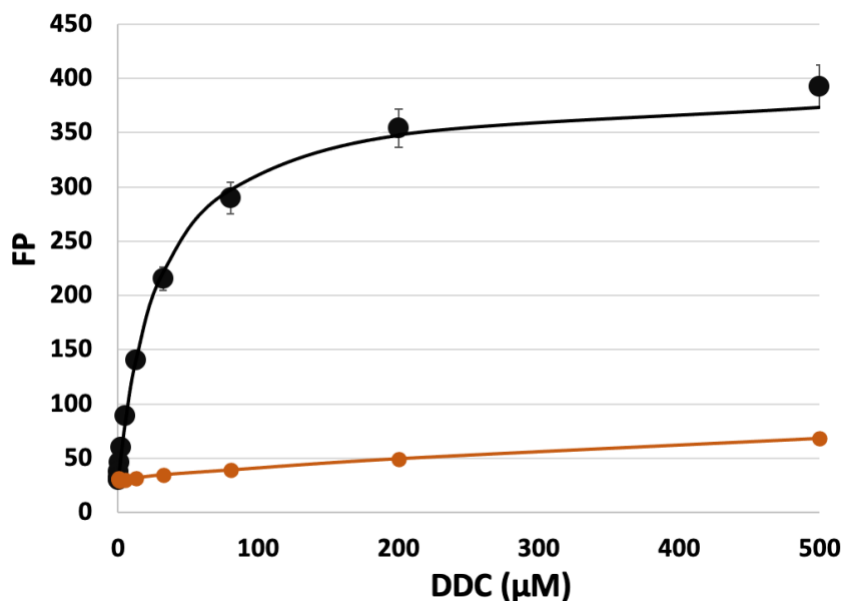


Figure 3.13 Fluorescence polarization saturation binding curves. Titration of FMI-tagged PNPO with increasing concentrations of holoDDC (black) and lysozyme (orange).

3.3.3B Surface plasmon resonance (SPR): Surface plasmon resonance was also used to characterize the dissociation constant of the PNPO•DDC complex. A control experiment involving PNPO-SHMT complex was also conducted. SPR experiments consist of molecules immobilized to the biosensor surface, i.e., ligand sample, and another in solution, i.e., analyte sample. The enzymes DDC and PNPO were expressed and purified as described in sections **3.2.2A-B**, and SPR experiments were carried out as described in section **3.3.2F**. Because the refractive index in SPR is sensitive to the composition of the chip surface, binding or dissociation of PNPO is expected to cause a

change in the refractory index, which is defined as a "response unit". A typical SPR sensorgram involves three stages: an associative phase in which PNPO binds to the immobilized protein, e.g., DDC; a steady state phase in which the rate of association is offset by dissociation; and a dissociative phase that occurs at the end of the injection when a PNPO-free buffer passes over the chip and bound PNPO dissociates from the chip surface. The binding of PNPO to holoDDC and holoSHMT was determined to be 3.7 μM and 15.4 μM , respectively, using seven concentrations of PNPO, as shown in **Figure 3.14A-B**. SPR was incapable of determining the binding affinity of PNPO to apoDDC. While the signals from the control flow cell (without protein) were normal, the signals from the flow cell where apoDDC immobilized were rapidly declined. This observation is likely due to the fact that apoDDC underwent a significant structural rearrangement after binding to PNPO, leading to PLP transfer to apoDDC with a significant conformational change to the holo-form (open-closed transition). The negative control analyte, albumin, showed no influence, i.e., no binding, on the generated response unit (see **Figure 3.14C-D**).

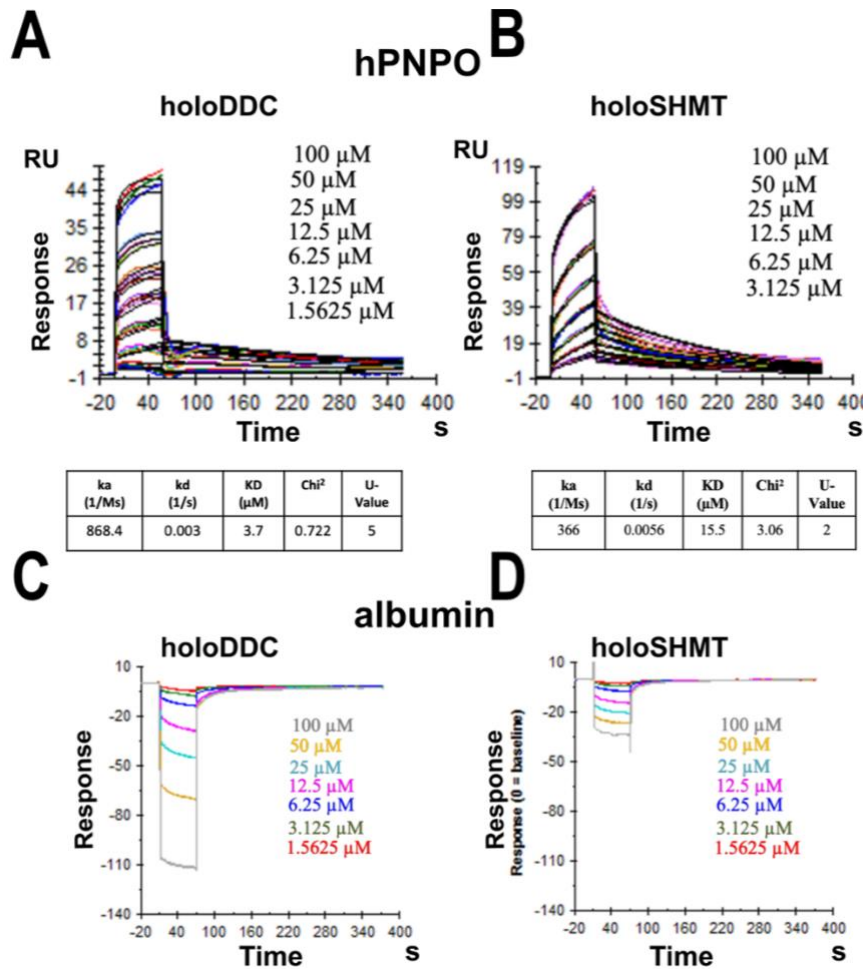


Figure 3.14 SPR binding results. A) Binding of PNPO to holoDDC; B) Binding of PNPO to holoSHMT (positive control); and C-D) Binding of Lysozyme to holoDDC and holoSHMT (negative control), respectively. Flow rate of all analytes were maintained at 50 $\mu\text{L}/\text{min}$, and injections of 100 μM to 1.5625 μM (from top to bottom). The contact and dissociation times used for all analytes were 60s and 300s, respectively. One 15s pulse of 1 M NaCl was injected for surface regeneration. All analytes were injected in triplicate. RU= Resonance Unit.

The FP and SPR studies of PNPO and holoDDC showed K_D of 57.4 μM and 3.7 μM , respectively, which compare with earlier studies for B6 salvage and B6 enzymes

with K_D values ranging from 0.3 to 12.3 μM . As a result, a third biophysical binding techniques ITC was used to determine the most reliable K_D for this complex.

3.3.3C Isothermal Calorimetry (ITC): ITC is a technique for measuring the amount of heat released or absorbed during a chemical reaction. It can be used to evaluate the nature of any chemical process involving spontaneous changes in temperature, without requiring any modification or tagging of the proteins being evaluated.^{55,56} Thermodynamic parameters such as free energy (ΔG), dissociation constant (K_D), change in enthalpy (ΔH), and stoichiometry (N) of the interaction between two or more molecules in solution can be determined. ITC binding titration for the purified recombinant PNPO and holoDDC enzymes were prepared and carried out as detailed in the Methods section (see **3.3.2G**). Also, the apoDDC enzyme was prepared as described in section **3.3.2C**. We found that the majority of holoDDC was converted to apoDDC ($\geq 97\%$) during the preparation process (**Figure 3.15**). Both PNPO•holoDDC and PNPO•apoDDC complexes were satisfactorily characterized by ITC in terms of binding interactions (**Figure 3.16A** and **3.16B**, respectively). The dissociation constants (K_D) and other parameters for the binding were determined using the equation ($\Delta G = RT \ln K_D = \Delta H - T\Delta S$) shown below (**Figure 3.16C**). The thermodynamical parameters ΔG , ΔH , and ΔS for the interaction between PNPO and holoDDC were -7.62 ± 0.08 , -2.33 ± 0.10 , and -5.21 ± 0.05 , respectively, -8.22 ± 0.06 , -3.03 ± 0.10 , and -5.19 ± 0.11 for PNPO and apoDDC, respectively.

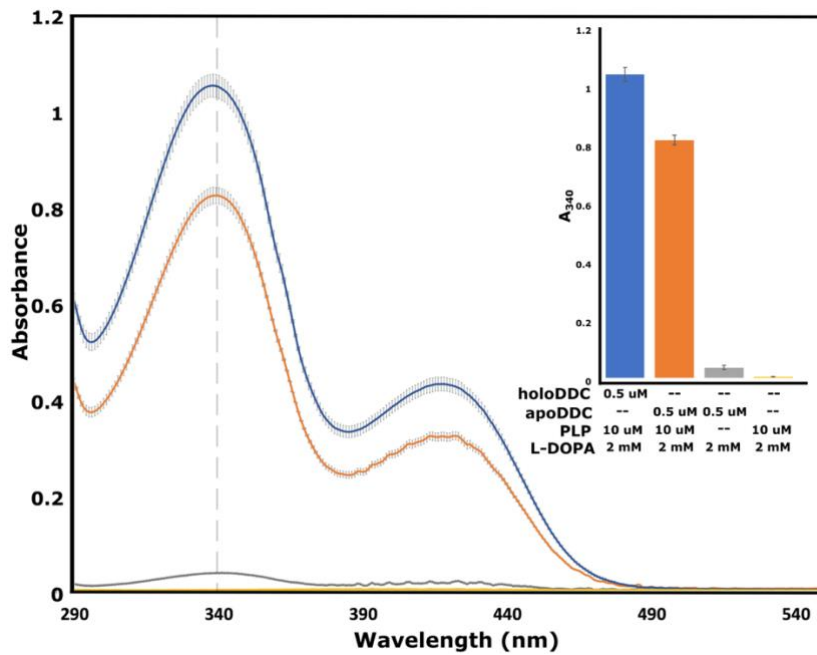


Figure 3.15 Enzymatic activity of recombinant human DDC in both apo and holo forms.

Activity assay performed using 0.5 μM of holoDDC (blue) and apoDDC (orange and grey) in the presence (orange) and absence (grey) of the cofactor PLP.

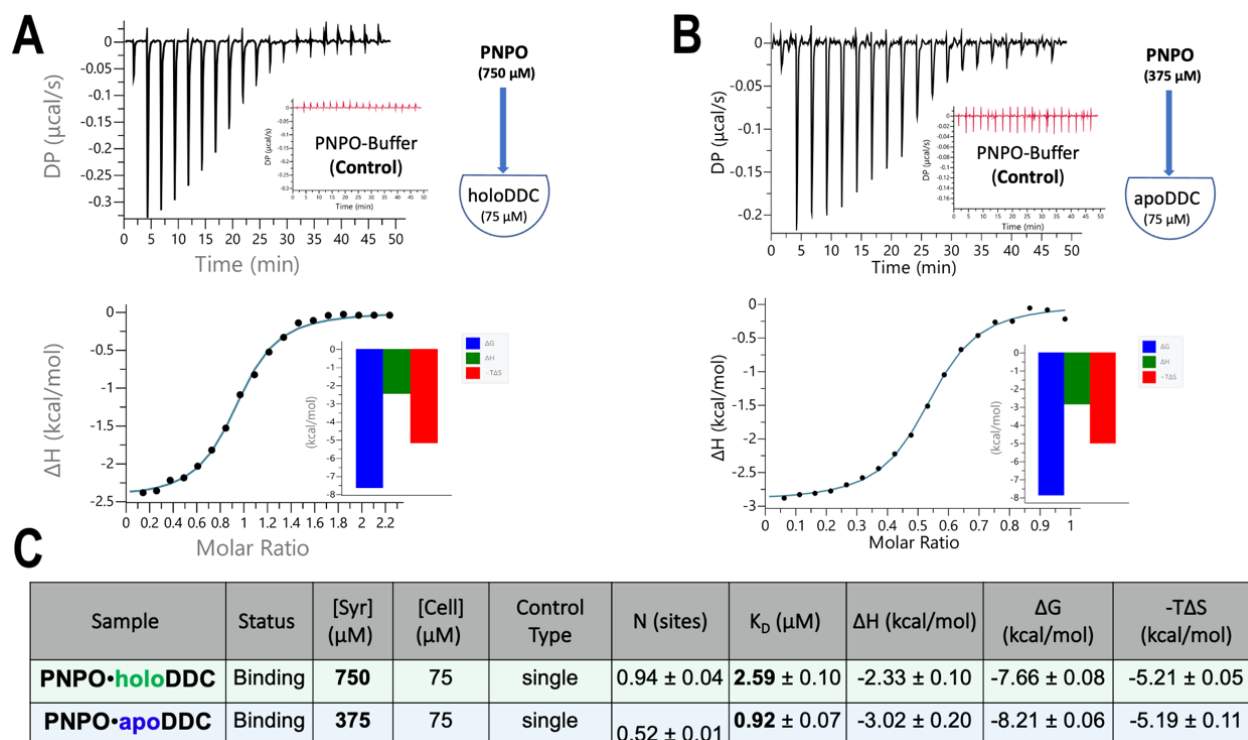


Figure 3.16 ITC thermodynamic analysis of PNPO•holoDDC and PNPO•apoDDC complexes binding. Injection heats for titration of PNPO into (A) holoDDC and (B) apoDDC. C) Averaged thermodynamic data of PNPO•holoDDC and PNPO•apoDDC bindings. All titrations were run in at least triplicate.

Both titration studies provided an exothermic binding isotherm, with the enthalpy change being more negative for apoDDC than for holoDDC; likely due to the larger protein-protein interface in the case of apoDDC. Furthermore, PNPO displayed more affinity toward apoDDC than holoDDC, i.e., $0.92 \pm 0.07 \mu\text{M}$ versus $2.59 \pm 0.11 \mu\text{M}$. This compares to $57.4 \mu\text{M}$ and $3.7 \mu\text{M}$ for the PNPO•holoDDC complex using FP and SPR, respectively. The PNPO•holoDDC complex showed a stoichiometry close to 1 (0.94 ± 0.04), i.e., one PNPO to one DDC, whereas the PNPO•apoDDC complex revealed a

stoichiometry of 0.5 (0.52 ± 0.01), i.e., one monomer of PNPO to a dimeric DDC. These observations are consistent with the modeling/docking studies described above that showed only one molecule of PNPO is capable of binding to the asymmetric apoDDC dimer structure, while two molecules of PNPO demonstrated an equivalent recognition of both sites of the symmetric dimer of the holoDDC.

Importantly, we were able to show that ITC can reliably characterize the binding of PNPO and DDC in both apo and holo forms, which was not with SPR and FP. Based on the computational and binding studies, site-directed mutagenesis studies were performed to validate the model.

3.3.3D PLP transfer from PNPO-PLP to activate apoDDC: Our group has focused on studying PLP transfer kinetics using a well-established assay with PNPO and PL kinase as a PLP source, and *E. coli* or rabbit serine hydroxyethyltransferase (apoSHMT) as a PLP acceptor.^{10,11} Our key premise is that channeling will provide a safe and efficient pathway for the transfer of cellular PLP from B₆-salvage enzymes to activate apo-B₆ enzymes. In this work, our computational and biophysical investigations showed that PNPO and DDC (both apo-and holo-forms) can form a physical complex. Our hypothesis is that this complex has a biological function that involves the transfer of PLP for productive activation of B₆-dependent enzymes. Therefore, a PLP transfer assay was developed and used to study the capability of PLP transfer from PNPO to activate apoDDC to its holoDDC form. The apoDDC preparation and PNPO saturation with PLP were prepared as described in the Methods section (**3.3.2C** and **3.3.2H**, respectively).

The transfer assay is a spectrophotometric assay that monitors formation/activation of holoDDC indirectly by measuring product (dopamine) formation, as described in the Methods section (see **3.3.2I**) and illustrated in **Figure 3.17**. Briefly, two steps are required for the assay. First, PLP binding, in which the enzyme apoDDC (50 μ M) was mixed with equivalent amount of free PLP (50 μ M) or PNPO•PLP (50 μ M) and incubated at 25 $^{\circ}$ C (dry bath incubator) for 0, 2.5, 5, 10, 30, 60, and 90 min. This step allows the formation of the enzymatic active holoDDC. Second, aliquots of these mixtures were diluted (100-fold) in a solution containing 2 mM L-Dopa in 100 mM KPO₄, pH 7.4 (substrate mixture). The L-Dopa is converted to dopamine by holoDDC, and the formed dopamine was extracted and measured at 2, 6, and 10 minute intervals, as described in the Methods section (see **3.3.2I**).

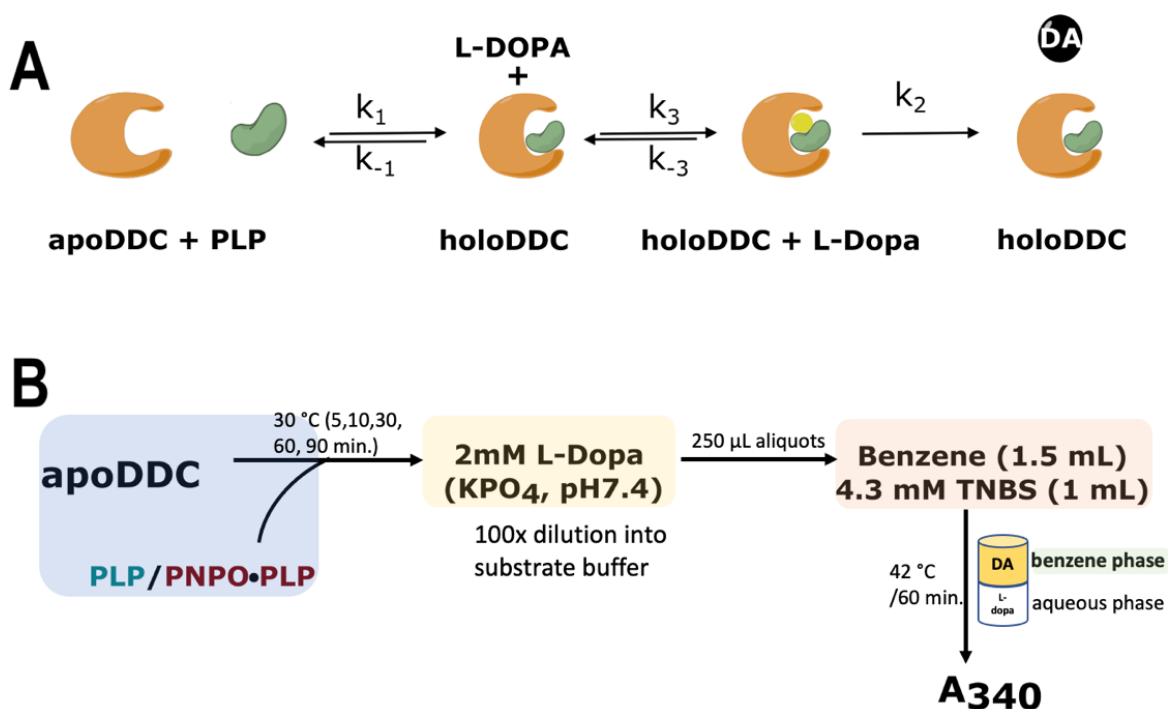


Figure 3.17 Schematic representation of the PLP transfer from PNPO•PLP to apoDDC. A) Enzyme kinetics for the cofactor (PLP) activation of apoDDC, substrate binding (L-DOPA) and conversion (dopamine). B) PLP transfer assay diagram for assessing apoDDC activation.

Dopamine formation at different time intervals for both free PLP and PNPO•PLP was fitted into a linear equation, as shown in **Figure 3.18B-C**, respectively, with R^2 ranging from 0.993 to 0.9997 for free PLP and 0.9878 to 0.9979 for PNPO•PLP complex. The final PLP transfer plot revealed that in the presence of PNPO•PLP complex, holoDDC activity reached ~30% of its activity when compared to equal amount of free PLP (**Figure 3.18D**). Importantly, we further demonstrated that the observed interaction between PNPO and apoDDC results in cofactor transfer and enzyme activation.

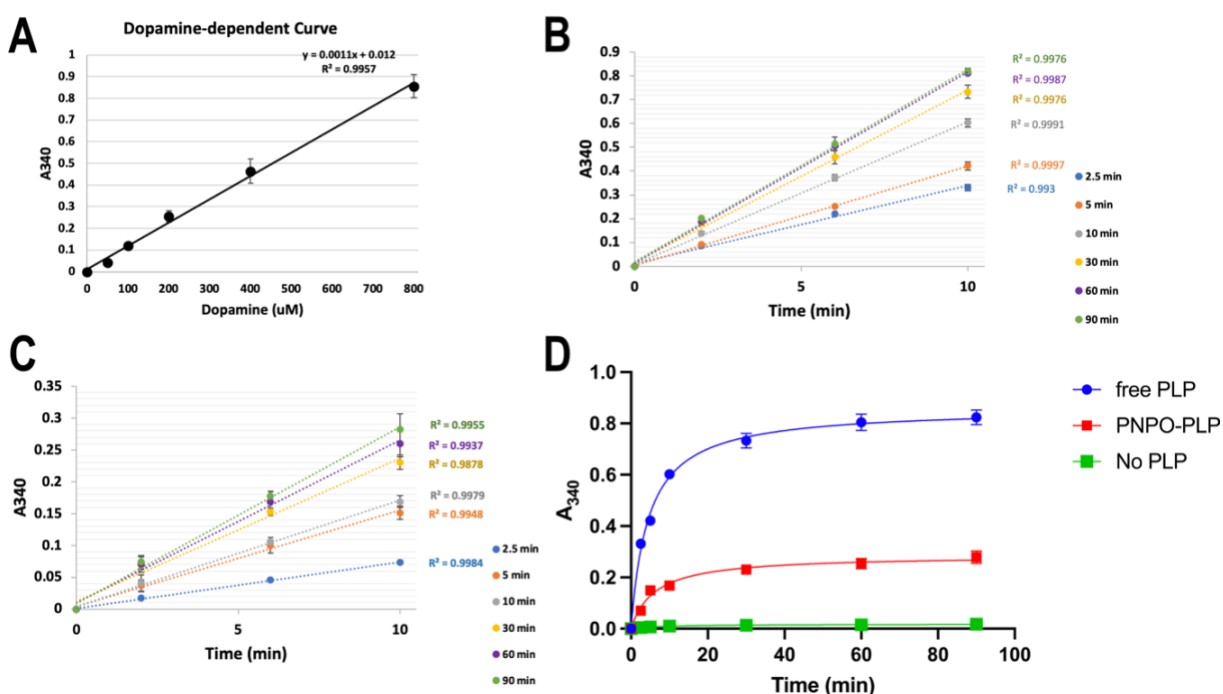


Figure 3.18 PLP transfer rate from PNPO•PLP to apoDDC. 50 μ M PNPO was 1:1 mixed with PLP or PNPO•PLP in 100 mM potassium phosphate buffer (pH 7.4) at 25 $^{\circ}$ C. PLP activation of

apoDDC was determined at different time intervals (2.5, 5, 10, 30, 60, and 90 min). At each time point, each sample was diluted in a solution containing 2 mM L-DOPA, and dopamine formation (A_{340}) was determined at three-time points (2, 6, and 10 min). A) Calibration curve of extracted dopamine at concentrations ranging from 0 to 800 μM ($r = 0.9957$); B-C) Activation of apoDDC after 2.5, 5, 10, 30, 60, and 90 min incubation with free PLP or PNPO•PLP, respectively; D) Final plot displaying dopamine formation at different activation time points, data were fitted and analyzed using prism GraphPad software. The PLP transfer rate with free PLP, PNPO•PLP, or no PLP source is depicted by the blue, red, and green lines of fitted data, respectively.

3.3.4 Conclusion

Our previous reported studies,¹¹ show that the B6 salvage enzymes PNPO or PLKase and the B₆-dependent enzyme SHMT are capable of forming a complex with each other with subsequent transfer of PLP from the former to activate the latter. The study in this chapter using computational, FP, SPR, and ITC techniques also demonstrated that PNPO forms a physical complex with both apo- and holoDDC. Both SPR and ITC yielded a similar K_D value for the PNPO•holoDDC complex, with K_D of 3.7 μ M and 2.59 μ M, respectively, but different from the K_D of 57.4 μ M from FP. The different K_D value by FP can be attributed to the labeled protein (PNPO-FMI), where labeling may have interfered with the native complex of PNPO and holoDDC. SPR and ITC, on the other hand, are label-free methods that would yield the most reliable affinity data.

Characterization of PNPO•apoDDC complex binding was only achievable with ITC, where the binding affinity was about 3-fold stronger than that of PNPO•holoDDC (0.93 μ M vs 3.7 μ M). Interestingly, while the PNPO•holoDDC complex showed a stoichiometry close to 1, (i.e., one PNPO to one DDC), PNPO•apoDDC complex had a stoichiometry of 0.5, (i.e., one PNPO monomer to each DDC dimer). The 1:1 stoichiometric ratio of PNPO•holoDDC complex implies that the PNPO binds to both symmetric sites in the dimeric holoDDC, which is consistent with previous findings (REF). An observed stoichiometric ratio of 1:0.5 for apoDDC•PNPO complex implies that the PNPO binds to only one of the asymmetric sites in the dimeric apoDDC, which is corroborated by the computational docking of PNPO into the apoDDC. Although there are very few structures for apo B6-enzymes to examine, the crystal structures of apo tyrosine

decarboxylase from *Papaver somniferum* (PDB: 6LIU) and apo glycine decarboxylase from *Thermus thermophilus* (PDB: 1WYT) have been solved as asymmetric structures. Furthermore, glutamate-1-semialdehyde aminomutase, a PLP dependent enzyme, was previously investigated for its asymmetry in structure and active site reactivity.⁶³

An assay for monitoring and characterizing PLP transfers from B₆-salvage enzyme, i.e., PNPO, to the DDC enzyme was successfully developed and optimized. PLP transfer studies indicated that ~30% of the tightly bound PLP from PNPO•PLP is transferred to activate apoDDC into the holo-form. In contrast, we observed 100% activation of apoDDC with equal amounts of free PLP. Similar observations have been reported for apoSHMT, where ~25-40% of the tightly bound PLP is transferred when equal concentrations of PNPO or PLKase (PLP donor) and apoSHMT (PLP acceptor) are present.^{10,11} Partial PNPO binding to one site of the asymmetric apo B₆ enzymes helps explain why apo B₆ enzymes are only partially activated. These findings, as well as the results of PLP transfer experiments, indicate that cellular PLP channeling from PNPO to DDC, or presumably other B₆-dependent enzymes, is likely to be physiologically relevant.

3.4 Site directed mutagenesis studies to probe interactions between PNPO and DDC

3.4.1 Introduction

In protein engineering, site-directed mutagenesis studies are a common approach that can be used to generate cloned DNAs with altered sequences in order to investigate the significance of certain amino acid residues on protein structure and function.⁵⁷ In the context of protein–protein interactions, "hot spot" residues refer to a residue or cluster of residues that contributes significantly to the binding free energy.⁵⁸ Computationally, a combination of machine learning, structure- and energy-based approaches has been developed to predict hot spot residues at protein-protein interfaces.^{47,59–62} *In silico* alanine-scanning mutagenesis is an effective approach for finding hotspot residues: important contributions are made by residues whose replacement with alanine results in a binding energy loss of $\Delta\Delta G \geq 1$ kcal mol⁻¹.^{46,47} Our computational study has identified a potential complex formed between PNPO and DDC (both apo- and holo-forms). Moreover, dynamics simulation studies on the putative complex model suggests several **putative strong** and **stable** salt-bridges, including R88–D442, E114–K207, and R88–E421, could form between PNPO and DDC, respectively. HINT scoring and *in silico* ALA scanning mutagenesis investigations suggested that R88–PNPO and E114–PNPO are candidate "hotspot" residues for site mutations studies. The site-mutations experiment was expected to abrogate interactions between the two enzymes and prevent activation

of apoDDC, and thus validating our computational model and, more importantly, our hypothesis.

3.4.2 Methods

3.4.2A Subcloning, expression and purification of PNPO mutants (R88A, E114A, R88A/E114A): Mutations of the PNPO coding region were carried out by Genscript (NJ, USA), using the NM_018129_pET-28b(+) construct as a template. Constructed plasmids (R88A_NM_018129_pET-28b(+), E114A_NM_018129_pET-28b(+), and R88A/E114A_NM_018129_pET-28b(+)) were confirmed by DNA sequence analysis followed by transformation into *E. coli* HMS174 (λ DE3) pLysS competent cells. Transformants were cultured on LB-agar medium containing kanamycin (40 μ g/mL) at 37 °C. Protein expressions and purifications were carried out as previously described for wild-type PNPO (section **3.3.2B**).

3.4.2B Thermal Stability of PNPO variants and DDC: The PNPO mutants (R88A, E114A, and R88A/E114A) were obtained using site directed mutagenesis as described earlier. Thermal stability study was conducted to evaluate the overall stability of PNPO mutants using the Prometheus NT.48 NanoDSF instrument (NanoTemper Technologies). The instrument, which uses Differential Scanning Fluorimetry (DSF) technology, detects the smallest changes in the fluorescence of intrinsic tryptophan and tyrosine found in all proteins (free label). All PNPO mutants were prepared in a 50 mM potassium phosphate buffer (pH 7.4) containing 150 mM NaCl at 1 mg/mL. Each sample

was measured in triplicate. During heating on a linear thermal ramp (2 °C/min, 25-95 °C), unfolding was detected with low detector sensitivity and 10% excitation power. The ratio of tryptophan emission at 330 and 350 nm (350/330 nm ratio), which depicts the shift in tryptophan emission upon unfolding, was used to determine the unfolding transition points. Thermal stability of DDC in both holo- and apo-forms was also determined at the same concentration (1 mg/mL).

3.4.2C Binding analysis of PNPO variants to apo- and holo-DDC by ITC:

Isothermal titration calorimetric measurements of the binding of PNPO mutants to the wild-type holo- and apo-DDC were performed. All samples were prepared and carried out as previously described for the ITC titration of wild-type PNPO and DDC (under section **3.3.2G**). PNPO variants were titrated to holoDDC at a molar ratio of 1:10, i.e., 750 µM to 75 µM, whereas PNPO variants were titrated to apoDDC at a molar ratio of 1:5, i.e., 375 µM to 75 µM.

3.4.3 Results and Discussion

3.4.3A Thermal Stability of PNPO variants and DDC: Wild-type PNPO and PNPO mutants (R88A, E114A, and R88A/E114) were expressed and purified recombinantly as previously described in sections **3.3.2B** and **3.4.2A**, respectively. Thermal stability is such a useful tool for assessing the overall stability of proteins. Therefore, thermal stability of PNPO mutants was evaluated and compared to that of wild-

type PNPO. The experiment was carried out using the Prometheus NT.48 NanoDSF instrument, as described in the Methods section (3.4.2B). As seen in **Figure 3.19**, PNPO mutants exhibited a lower melting temperature (T_m) than wild-type PNPO, making them relatively less stable. Values of T_m obtained for R88A, E114A, R88A/E114A were 61.29 ± 0.10 °C, 61.09 ± 0.06 °C, and 57.95 ± 0.08 °C, respectively, compared to 62.67 ± 0.15 °C for wild-type PNPO. (**Figure 3.19**). The Prometheus DSF experiment proved the structural integrity of the PNPO mutants by demonstrating that the melting temperatures (T_m) of PNPO mutants did not vary considerably. As a result, their binding affinities to the DDC enzyme can be measured.

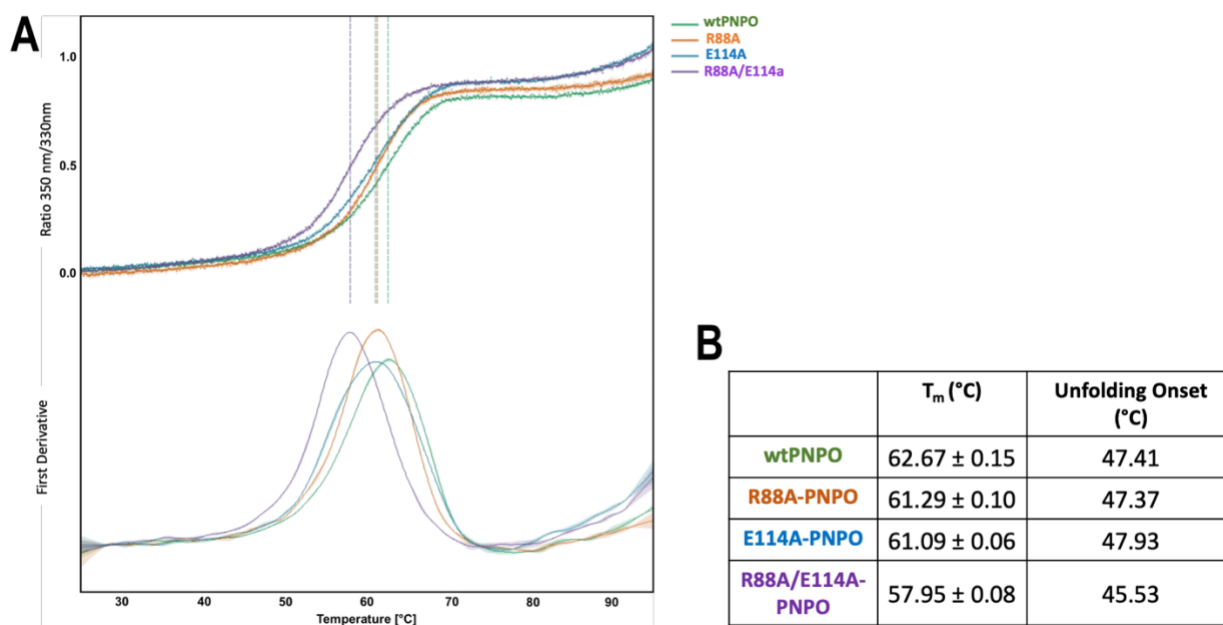


Figure 3.19 Differential scanning fluorimetry measurements of PNPO wild-type and mutants. A) The fluorescence change (above) and the first derivative of fluorescence emission (below) are shown in green, orange, cyan, and purple for wild-type PNPO, R88A, E114A, and R88A/E114A PNPO mutants, respectively. B) Tabulated obtained DSF measurements for

melting temperatures (T_m) and unfolding onsets for both the wild-type PNPO and PNPO mutants.

3.4.3B ITC titration analysis of PNPO variants and DDC: Isothermal titration calorimetric measurements of PNPO mutant binding to wild type holo- and apo-DDC were carried out using the parameters and conditions previously described (**3.3.2G**) and the results shown in **Table 3.3**. For titration of PNPO mutants to holoDDC, the dissociation constants (K_D) of R88A–PNPO, E114A–PNPO, and R88A/E114A–PNPO were found to be $4.59 \pm 0.57 \mu\text{M}$, $2.21 \pm 0.02 \mu\text{M}$, and $2.28 \pm 0.04 \mu\text{M}$, respectively, as compared to $2.59 \pm 0.11 \mu\text{M}$ for wild-type PNPO (**Figure 3.20**). For titration of PNPO mutants to apoDDC, the K_D values for R88A–PNPO, E114A–PNPO, and R88A/E114A–PNPO were $1.52 \pm 0.27 \mu\text{M}$, $0.82 \pm 0.14 \mu\text{M}$, and $1.03 \pm 0.08 \mu\text{M}$, respectively, compared to $0.93 \pm 0.07 \mu\text{M}$ for wild-type PNPO (**Figure 3.21**). The thermodynamical parameters ΔG , ΔH , and ΔS for binding of PNPO to DDC (both holo- and apo-forms) are summarized in **Table 3.3** and illustrated in **Figure 3.22**. For holoDDC, free energy changes (ΔG) shown to be $-7.62 \pm 0.08 \text{ kcal mol}^{-1}$, $-7.29 \pm 0.06 \text{ kcal mol}^{-1}$, $-7.72 \pm 0.01 \text{ kcal mol}^{-1}$, and $-7.71 \pm 0.04 \text{ kcal mol}^{-1}$, for wild-type PNPO, R88A–PNPO, E114A–PNPO, and R88A/E114A–PNPO, respectively. The enthalpy change (ΔH) found to be $-2.33 \pm 0.10 \text{ kcal mol}^{-1}$, $-2.11 \pm 0.05 \text{ kcal mol}^{-1}$, $-2.58 \pm 0.02 \text{ kcal mol}^{-1}$, and $-2.85 \pm 0.05 \text{ Kcal mol}^{-1}$, this is associated with an entropy change of $-5.21 \pm 0.05 \text{ kcal mol}^{-1}$, $-5.15 \pm 0.05 \text{ kcal mol}^{-1}$, $-5.13 \pm 0.04 \text{ kcal mol}^{-1}$, and $-4.86 \pm 0.04 \text{ kcal mol}^{-1}$, for wild-type PNPO, R88A–PNPO, E114A–PNPO, and R88A/E114A–PNPO, respectively (see **Figures 3.20** and **3.22**). Clearly, PNPO mutants

showed no significant changes in their ability to bind to the holoDDC when compared to the wild-type PNPO, except perhaps R88A-PNPO that showed 2-fold less affinity than WT-PNPO (4.59 μ M vs 2.59 μ M, respectively). In contrast E114A-PNPO (2.21 μ M) and R88A/E114A-PNPO (2.28 μ M) show similar binding affinity as the WT-PNPO (2.59 μ M). However, these mutants were proposed on the basis of the apoDDC structure, thus we were interested in evaluating their binding affinities to the apoDDC, as described below.

R88A-PNPO•holoDDC

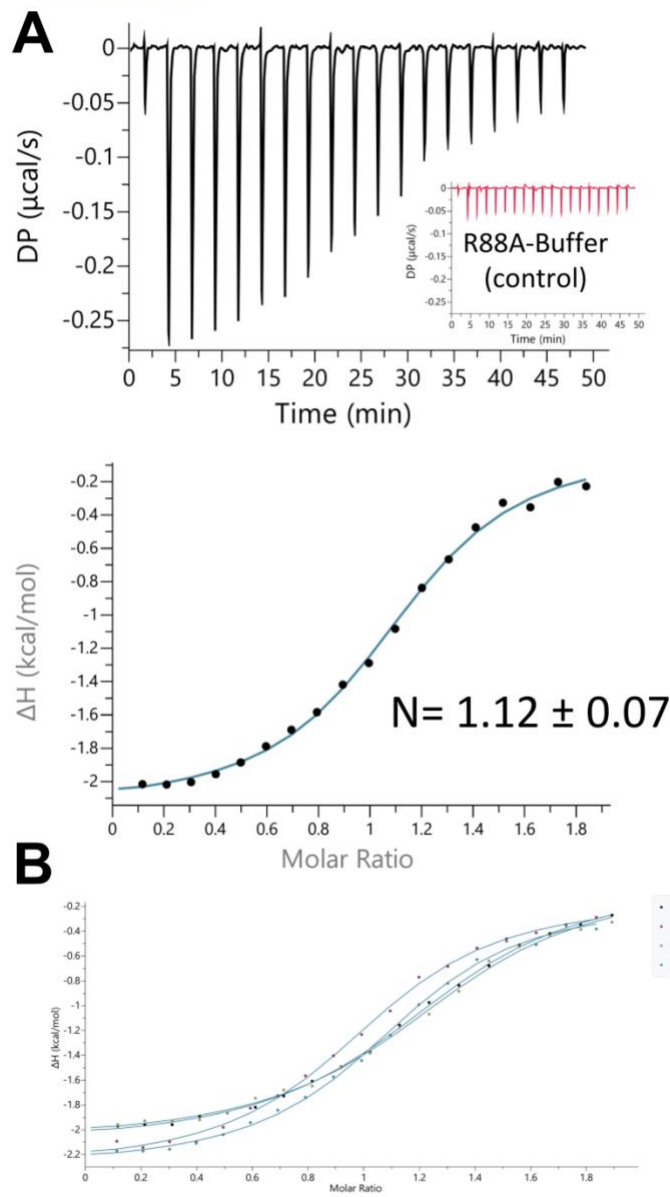


Figure 3.20A Isothermal titration calorimetry (ITC) characterization of R88A-PNPO mutant with holoDDC. A) ITC binding isotherm of R88A-PNPO with holoDDC; B) ITC thermogram obtained from four independent measurements.

E114A-PNPO•holoDDC

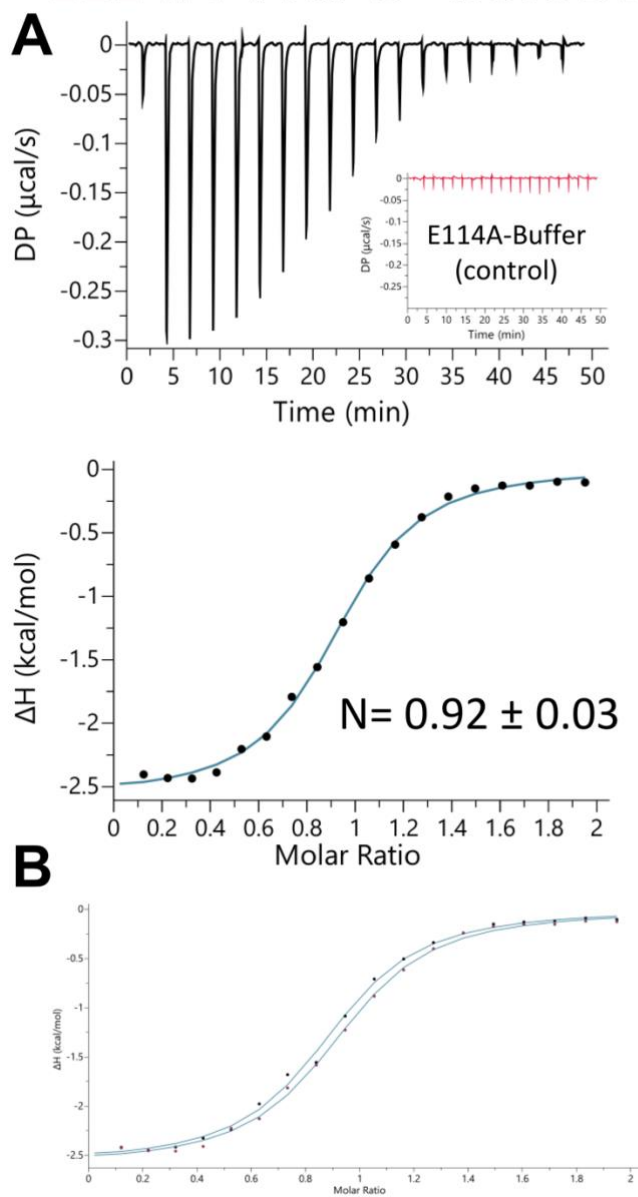


Figure 3.20B Isothermal titration calorimetry (ITC) characterization of E114A-PNPO mutant with holoDDC. A) ITC binding isotherm of E114A-PNPO with holoDDC; B) ITC thermogram obtained from duplicate measurements.

R88A/E114A-PNPO•holoDDC

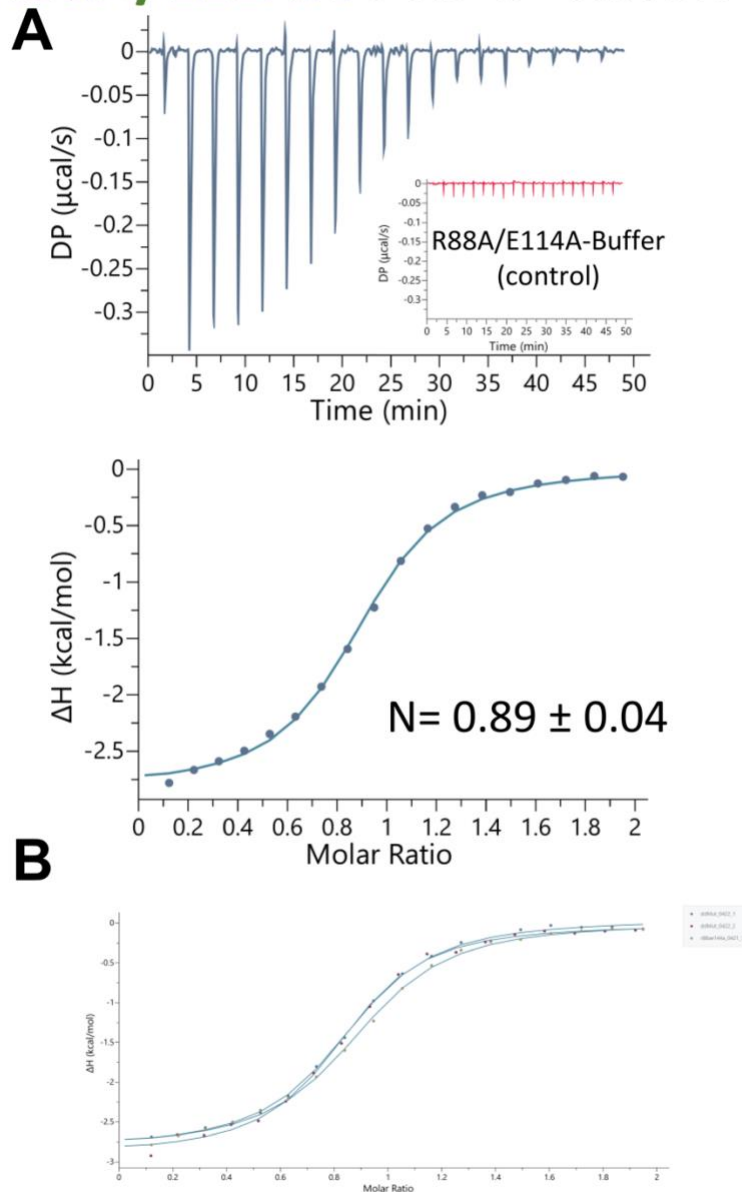


Figure 3.20C Isothermal titration calorimetry (ITC) characterization of R88A/E114A-PNPO mutant with holoDDC. A) ITC binding isotherm of R88A/E114A-PNPO with holoDDC; B) ITC thermogram obtained from triplicate measurements.

As previously demonstrated, apoDDC exhibited more affinity for wild-type and PNPO mutants than holoDDC (**Figure 3.16; Table 3.3**). Free energy changes (ΔG) were -8.22 ± 0.06 kcal mol⁻¹, -7.95 ± 0.11 kcal mol⁻¹, -8.31 ± 0.11 kcal mol⁻¹, and -8.17 ± 0.05 kcal mol⁻¹, for wild-type PNPO, R88A, E114A, and R88A/E114A, respectively. The enthalpy change (ΔH) shown to be -3.03 ± 0.10 kcal mol⁻¹, -2.81 ± 0.17 kcal mol⁻¹, -3.17 ± 0.14 kcal mol⁻¹, and -3.47 ± 0.07 kcal mol⁻¹, this is associated with an entropy change of -5.19 ± 0.11 kcal mol⁻¹, -5.01 ± 0.21 kcal mol⁻¹, -5.20 ± 0.16 kcal mol⁻¹, and -4.70 ± 0.11 kcal mol⁻¹, for wild-type PNPO, R88A, E114A, and R88A/E114A, respectively (see **Figures 3.21 and 3.22 ; Table 3.3**).

R88A-PNPO•apoDDC

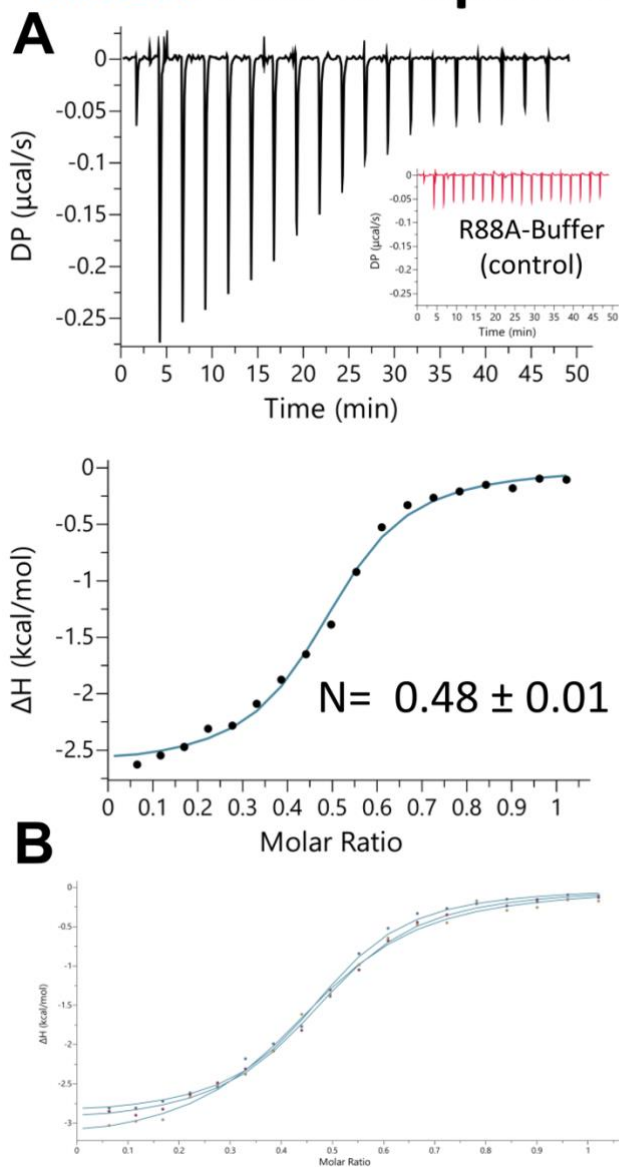


Figure 3.21A Isothermal titration calorimetry (ITC) characterization of R88A-PNPO mutant with apoDDC. A) ITC binding isotherm of R88A-PNPO with apoDDC; B) ITC thermogram obtained from triplicate measurements.

E114A-PNPO•apoDDC

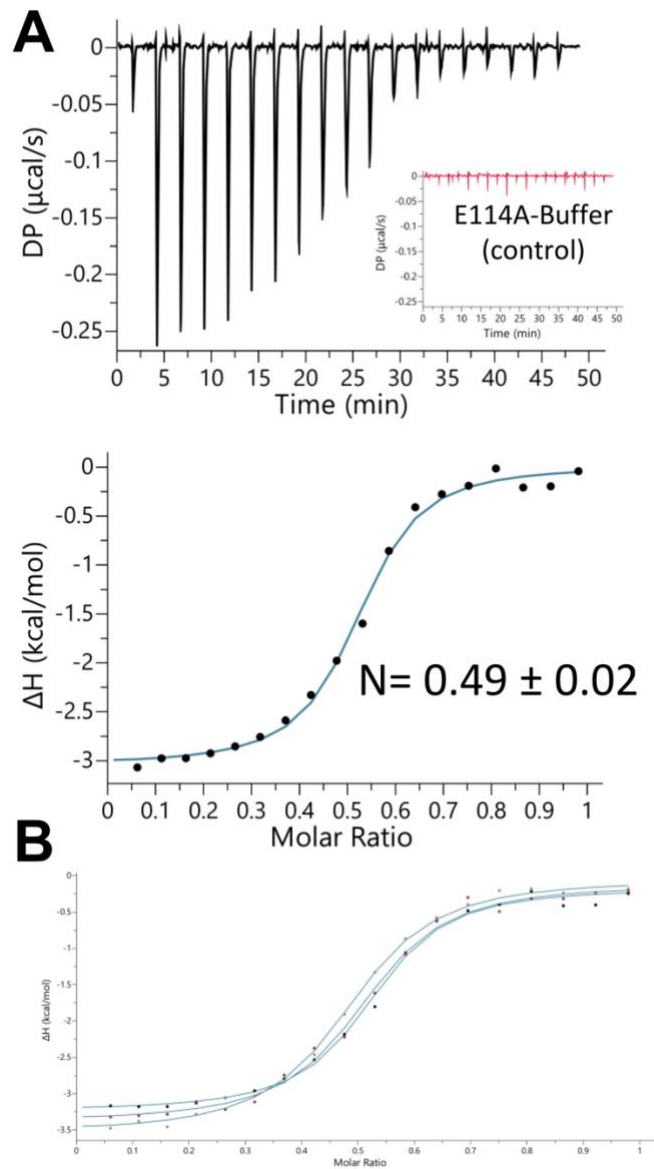


Figure 3.21B Isothermal titration calorimetry (ITC) characterization of E114A-PNPO mutant with apoDDC. A) ITC binding isotherm of E114A-PNPO with apoDDC; B) ITC thermogram obtained from triplicate measurements.

R88A/E114A-PNPO•apoDDC

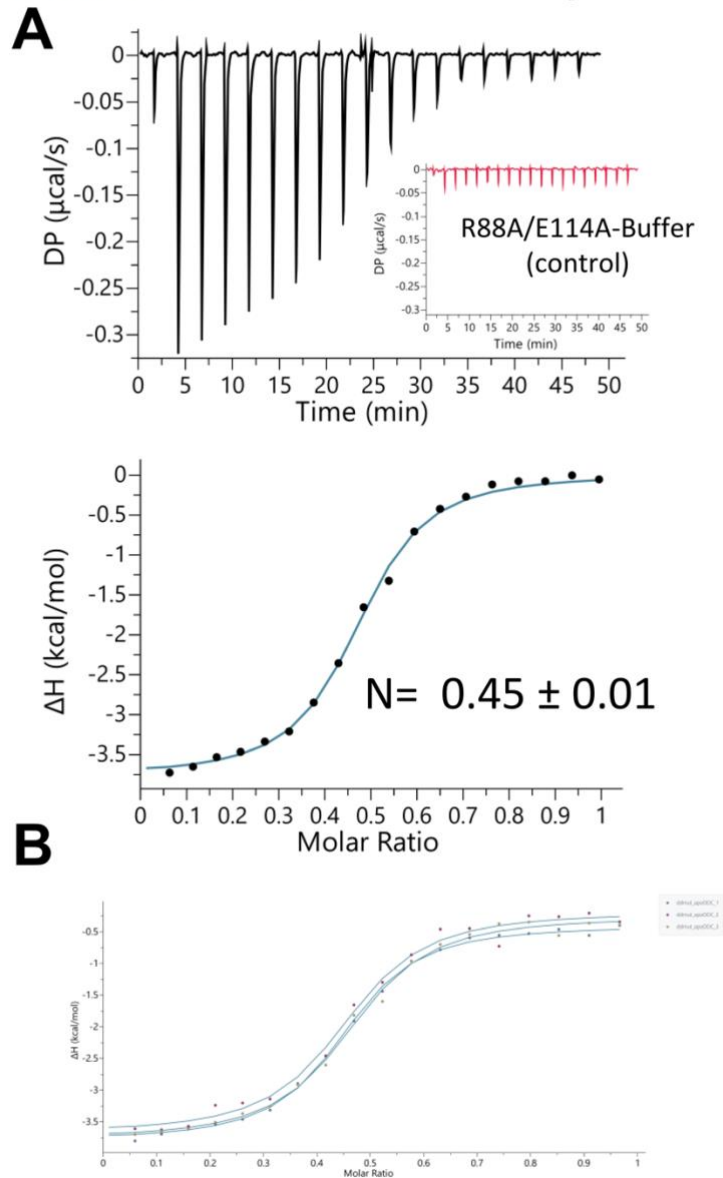


Figure 3.21C Isothermal titration calorimetry (ITC) characterization of R88A/E114A-PNPO mutant with apoDDC. A) ITC binding isotherm of R88A/E114A-PNPO with apoDDC; B) ITC thermogram obtained from triplicate measurements.

Apparently, the PNPO mutants proposed by our modeling studies had almost no effect on the formation of PNPO•apoDDC and PNPO•holoDDC complexes. For the mutant R88A-PNPO, the enthalpy change was slightly affected and was 0.23 ± 0.06 and 0.22 ± 0.18 kcal mol⁻¹ less favorable than the wild-type values for holoDDC and apoDDC, respectively (**Figure 3.22; Table 3.3**). As a result, the dissociation rate constant for the single mutant R88A-PNPO was approximately two-fold weaker than that of the wild type, with values of 4.59 and 1.52 μM compared to 2.50 and 0.92 μM for holoDDC and apoDDC, respectively. The single mutant E114A–PNPO and the double mutant R88A/E114A–PNPO, on the other hand, displayed unexpectedly more favorable enthalpic contribution. Both E114A and R88A/E114A mutants showed more favorable enthalpy changes of -0.25 ± 0.05 kcal mol⁻¹ and -0.52 ± 0.06 kcal mol⁻¹, respectively, when bound to holoDDC (**Figure 3.22; Table 3.3**). The mutants E114A and R88A/E114A, on the other hand, demonstrated similar behavior when titrated to apoDDC, with enthalpy changes of -0.14 ± 0.13 kcal mol⁻¹ and -0.44 ± 0.06 kcal mol⁻¹, respectively (**Figure 3.22; Table 3.3**). Overall, the PNPO mutants provided by our modeling investigations had no significant effects on preventing the formation of the complex, and possible explanations for this observation is provided below, which suggest additional site-directed mutagenesis investigations may be necessary to locate other key residues to validate the proposed model of the complex.

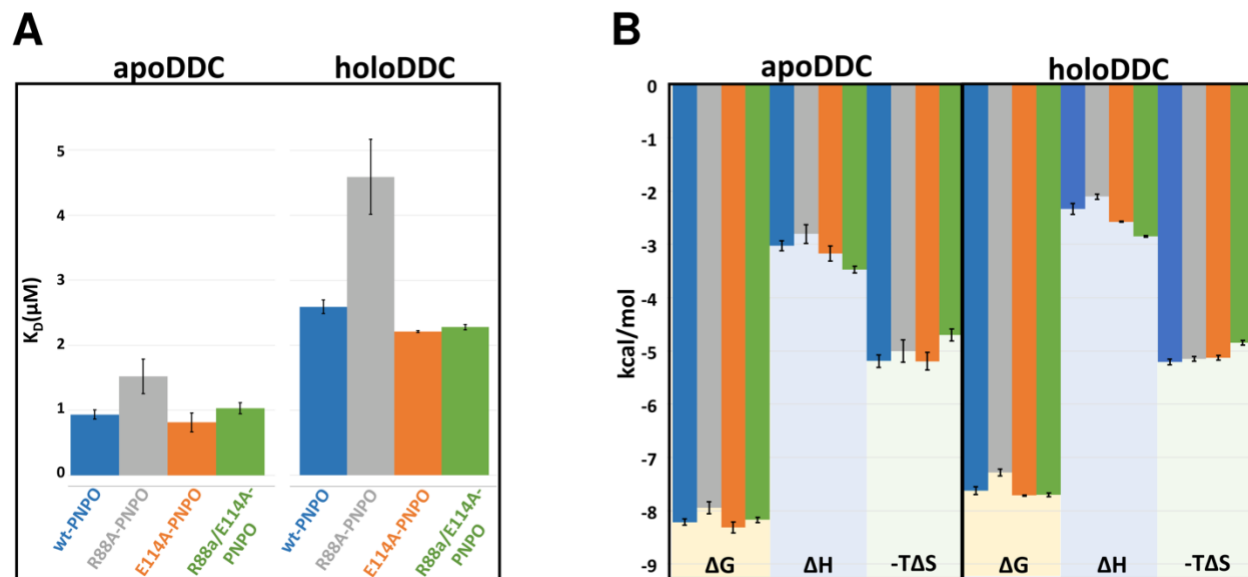


Figure 3.22 ITC thermodynamic binding data of PNPO mutants bound to apo- and holoDDC. A) The dissociation constants (K_D) of wild-type and mutant PNPO bound to apoDDC (left) and holoDDC (right); B) ITC parameters for free energy change (ΔG), enthalpy change (ΔH), and entropy change ($-T\Delta S$) are represented by bars and shaded in light yellow, light blue, and light green, respectively.

Table 3.3 ITC measured thermodynamic parameters between PNPO mutants and DDC in both holo (left) and apo (right) forms.

	holoDDC				apoDDC			
	K_D (μM)	ΔG (kcal mol ⁻¹)	ΔH (kcal mol ⁻¹)	$-T\Delta S$ (kcal mol ⁻¹)	K_D (μM)	ΔG (kcal mol ⁻¹)	ΔH (kcal mol ⁻¹)	$-T\Delta S$ (kcal mol ⁻¹)
wt-PNPO	2.59 ± 0.11	-7.62 ± 0.08	-2.33 ± 0.10	-5.21 ± 0.05	0.93 ± 0.07	-8.22 ± 0.06	-3.03 ± 0.10	-5.19 ± 0.11
R88A-PNPO	4.59 ± 0.57	-7.29 ± 0.06	-2.11 ± 0.05	-5.15 ± 0.05	1.52 ± 0.27	-7.95 ± 0.11	-2.81 ± 0.17	-5.01 ± 0.21
E114A-PNPO	2.21 ± 0.02	-7.72 ± 0.01	-2.58 ± 0.02	-5.13 ± 0.04	0.82 ± 0.14	-8.31 ± 0.11	-3.17 ± 0.14	-5.20 ± 0.16
R88A/E114A-PNPO	2.28 ± 0.04	-7.71 ± 0.04	-2.85 ± 0.05	-4.86 ± 0.04	1.03 ± 0.08	-8.17 ± 0.05	-3.47 ± 0.07	-4.7 ± 0.11

Based on our model, the residue E114–PNPO of one chain of the dimer suggested a stable salt bridge formation with K207–DDC, ranging between 2–4Å over simulation time; hence, the slight increase in hydrophilic interaction was unexpected. However, the residue E114 on the other chain of the PNPO dimer is stabilized by intramolecular contact with K119, which appears to be involved in forming a salt bridge with E180 at the DDC C-terminus (see **Figure 3.23**). One explanation is that the loss of negatively charged side chain of E114 may allow a better electrostatic interaction between K119–PNPO and E480–DDC (**Figure 3.23**). Therefore, K119–PNPO is a candidate for further site-mutation studies (K119A) to determine its affect in abrogating the PNPO•DDC complex formation.

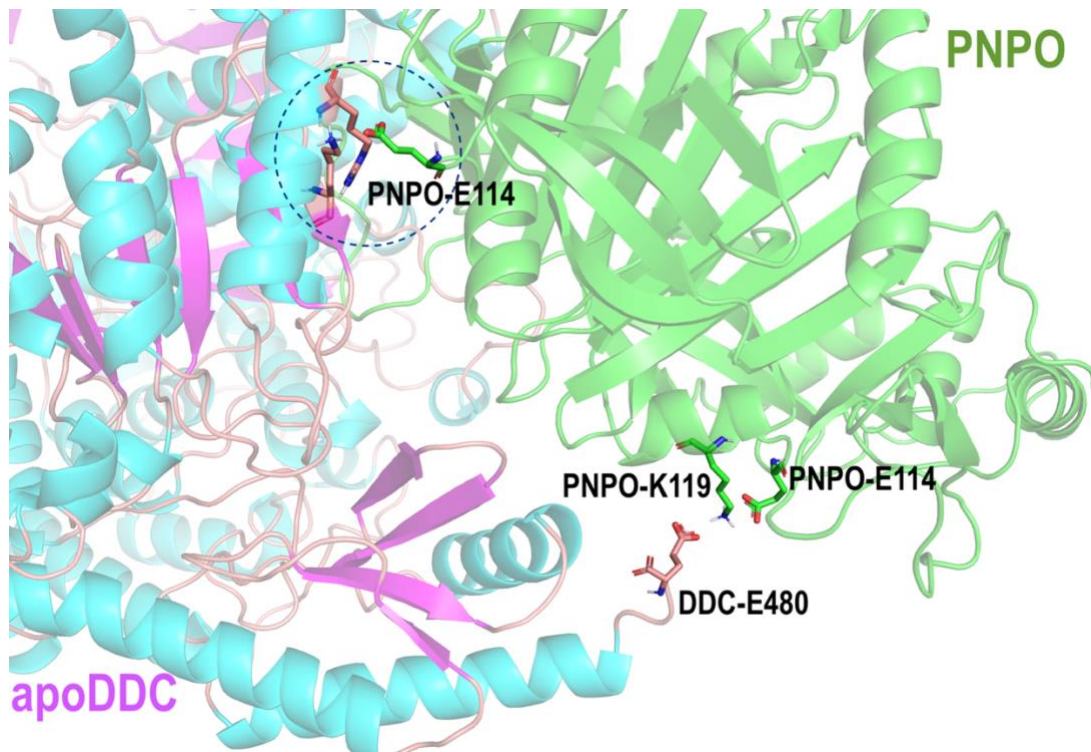


Figure 3.23 Predicted PNPO·apoDDC complex showing the proposed key residue (PNPO-E114) interaction at the protein-protein interface.

3.4.4 Conclusion

Our computational and biophysical investigations revealed that PNPO is capable of forming a complex with DDC, in both its apo- and holo-forms. On the basis of our computational model, we proposed key residues that are amenable to mutation and experimental testing for their potential to affect the formation of the complex. Three PNPO mutants, R88A, E114A, and R88A/E114A, were successfully subcloned, expressed, and purified. The thermal stability of the three mutants was then assessed, with T_m values of 61.29 ± 0.10 °C, 61.09 ± 0.06 °C, and 57.95 ± 0.08 °C for R88A, E114A, and R88A/E114A, respectively, compared to 62.67 ± 0.15 °C for the wild-type PNPO. The melting temperature (T_m) of the PNPO mutants did not vary significantly, therefore their binding affinities towards the DDC enzyme were assessed. ITC measurements were carried out on PNPO mutants and DDC (in both holo- and apo-forms). ITC binding studies showed no substantial impact of PNPO mutants on the stability of the complex formation. Because of this, additional molecular modeling and site-directed mutagenesis investigations are necessary to gain a comprehensive understanding of the near-native complex.

Reference:

- (1) di Salvo, M. L.; Contestabile, R.; Safo, M. K. Vitamin B6 Salvage Enzymes: Mechanism, Structure and Regulation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011**, *1814* (11), 1597–1608. <https://doi.org/10.1016/j.bbapap.2010.12.006>.
- (2) Salvo, M. L.; Safo, M. K.; Contestabile, R. Biomedical Aspects of Pyridoxal 5'-Phosphate Availability. *Front Biosci (Elite Ed)* **2012**, *4*, 897–913.
- (3) Bowling, F. G. Pyridoxine Supply in Human Development. *Semin. Cell Dev. Biol.* **2011**, *22* (6), 611–618. <https://doi.org/10.1016/j.semcdb.2011.05.003>.
- (4) Ciorba, M. A. Kynurenine Pathway Metabolites: Relevant to Vitamin B-6 Deficiency and Beyond. *Am. J. Clin. Nutr.* **2013**, *98* (4), 863–864. <https://doi.org/10.3945/ajcn.113.072025>.
- (5) Midttun, Ø.; Ulvik, A.; Ringdal Pedersen, E.; Ebbing, M.; Bleie, Ø.; Schartum-Hansen, H.; Nilsen, R. M.; Nygård, O.; Ueland, P. M. Low Plasma Vitamin B-6 Status Affects Metabolism through the Kynurenine Pathway in Cardiovascular Patients with Systemic Inflammation. *The Journal of Nutrition* **2011**, *141* (4), 611–617. <https://doi.org/10.3945/jn.110.133082>.
- (6) Musayev, F. N.; Di Salvo, M. L.; Saavedra, M. A.; Contestabile, R.; Ghatge, M. S.; Haynes, A.; Schirch, V.; Safo, M. K. Molecular Basis of Reduced Pyridoxine 5'-Phosphate Oxidase Catalytic Activity in Neonatal Epileptic Encephalopathy Disorder. *J. Biol. Chem.* **2009**, *284* (45), 30949–30956. <https://doi.org/10.1074/jbc.M109.038372>.
- (7) Musayev, F. N.; Di Salvo, M. L.; Ko, T.-P.; Schirch, V.; Safo, M. K. Structure and Properties of Recombinant Human Pyridoxine 5'-Phosphate Oxidase. *Protein Sci.* **2003**, *12* (7), 1455–1463. <https://doi.org/10.1110/ps.0356203>.
- (8) Meyers, L. D.; Hellwig, J. P.; Otten, J. J. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements.; National Academies Press., 2006.
- (9) Pan, P.; Woehl, E.; Dunn, M. F. Protein Architecture, Dynamics and Allostery in Tryptophan Synthase Channeling. *Trends in Biochemical Sciences* **1997**, *22* (1), 22–27. [https://doi.org/10.1016/S0968-0004\(96\)10066-9](https://doi.org/10.1016/S0968-0004(96)10066-9).
- (10) Ghatge, M. S.; Contestabile, R.; Salvo, M. L.; Desai, J. V.; Gandhi, A. K.; Camara, C. M.; Florio, R.; González, I. N.; Parroni, A.; Schirch, V.; Safo, M. K. Pyridoxal 5'-Phosphate Is a Slow Tight Binding Inhibitor of E. Coli Pyridoxal Kinase. *PLoS ONE* **2012**, *7* (7), e41680. <https://doi.org/10.1371/journal.pone.0041680>.
- (11) Ghatge, M. S.; Karve, S. S.; David, T. M. S.; Ahmed, M. H.; Musayev, F. N.; Cunningham, K.; Schirch, V.; Safo, M. K. Inactive Mutants of Human Pyridoxine 5'-Phosphate Oxidase: A Possible Role for a Noncatalytic Pyridoxal 5'-Phosphate Tight Binding Site. *FEBS Open Bio* **2016**, *6* (5), 398–408. <https://doi.org/10.1002/2211-5463.12042>.
- (12) Salvo, M. L.; Safo, M. K.; Musayev, F. N.; Bossa, F.; Schirch, V. Structure and Mechanism of Escherichia Coli Pyridoxine 5'-Phosphate Oxidase. *Biochim. Biophys. Acta* **2003**, *1647* (1–2), 76–82.
- (13) Gandhi, A. K.; Ghatge, M. S.; Musayev, F. N.; Sease, A.; Aboagye, S. O.; Salvo, M. L.; Schirch, V.; Safo, M. K. Kinetic and Structural Studies of the Role of the Active Site Residue Asp235 of Human Pyridoxal Kinase. *Biochem. Biophys. Res. Commun.* **2009**, *381* (1), 12–15. <https://doi.org/10.1016/j.bbrc.2009.01.170>.
- (14) Gandhi, A. K.; Desai, J. V.; Ghatge, M. S.; Salvo, M. L.; Di Biase, S.; Danso-Danquah, R.; Musayev, F. N.; Contestabile, R.; Schirch, V.; Safo, M. K. Crystal Structures of Human Pyridoxal Kinase in Complex with the Neurotoxins, Ginkgotoxin and Theophylline: Insights into Pyridoxal Kinase Inhibition. *PLoS ONE* **2012**, *7* (7), e40954. <https://doi.org/10.1371/journal.pone.0040954>.

- (15) Rocha, J. F.; Pina, A. F.; Sousa, S. F.; Cerqueira, N. M. F. S. A. PLP-Dependent Enzymes as Important Biocatalysts for the Pharmaceutical, Chemical and Food Industries: A Structural and Mechanistic Perspective. *Catal. Sci. Technol.* **2019**, *9* (18), 4864–4876. <https://doi.org/10.1039/C9CY01210A>.
- (16) Bilski, P.; Li, M. Y.; Ehrenshaft, M.; Daub, M. E.; Chignell, C. F. Symposium-in-Print Vitamin B6 (Pyridoxine) and Its Derivatives Are Efficient Singlet Oxygen Quenchers and Potential Fungal Antioxidants. *Photochem Photobiol* **2000**, *71* (2), 129. [https://doi.org/10.1562/0031-8655\(2000\)071](https://doi.org/10.1562/0031-8655(2000)071).
- (17) Ehrenshaft, M.; Bilski, P.; Li, M. Y.; Chignell, C. F.; Daub, M. E. A Highly Conserved Sequence Is a Novel Gene Involved in *de Novo* Vitamin B6 Biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (16), 9374–9378. <https://doi.org/10.1073/pnas.96.16.9374>.
- (18) Tramonti, A.; Nardella, C.; Salvo, M. L.; Pascarella, S.; Contestabile, R. The MocR-like Transcription Factors: Pyridoxal 5'-phosphate-dependent Regulators of Bacterial Metabolism. *FEBS J* **2018**, *285* (21), 3925–3944. <https://doi.org/10.1111/febs.14599>.
- (19) Grubman, A.; Phillips, A.; Thibonnier, M.; Kaparakis-Liaskos, M.; Johnson, C.; Thiberge, J.-M.; Radcliff, F. J.; Ecobichon, C.; Labigne, A.; de Reuse, H.; Mendz, G. L.; Ferrero, R. L. Vitamin B₆ Is Required for Full Motility and Virulence in *Helicobacter Pylori*. *mBio* **2010**, *1* (3), e00112-10. <https://doi.org/10.1128/mBio.00112-10>.
- (20) Dick, T.; Manjunatha, U.; Kappes, B.; Gengenbacher, M. Vitamin B6 Biosynthesis Is Essential for Survival and Virulence of Mycobacterium Tuberculosis: Vitamin B6 Biosynthesis of the Tubercle Bacillus. *Molecular Microbiology* **2010**, *78* (4), 980–988. <https://doi.org/10.1111/j.1365-2958.2010.07381>.
- (21) Zhao, G.; Winkler, M. E. Kinetic Limitation and Cellular Amount of Pyridoxine (Pyridoxamine) 5'-Phosphate Oxidase of Escherichia Coli K-12. *J. Bacteriol.* **1995**, *177* (4), 883–891.
- (22) Yang, E. S.; Schirch, V. Tight Binding of Pyridoxal 5'-Phosphate to Recombinant Escherichia Coli Pyridoxine 5'-Phosphate Oxidase. *Arch. Biochem. Biophys.* **2000**, *377* (1), 109–114. <https://doi.org/10.1006/abbi.2000.1737>.
- (23) Barile, A.; Tramonti, A.; di Salvo, M. L.; Nogués, I.; Nardella, C.; Malatesta, F.; Contestabile, R. Allosteric Feedback Inhibition of Pyridoxine 5'-Phosphate Oxidase from Escherichia Coli. *Journal of Biological Chemistry* **2019**, *294* (43), 15593–15603. <https://doi.org/10.1074/jbc.RA119.009697>.
- (24) Barile, A.; Battista, T.; Fiorillo, A.; di Salvo, M. L.; Malatesta, F.; Tramonti, A.; Ilari, A.; Contestabile, R. Identification and Characterization of the Pyridoxal 5'-Phosphate Allosteric Site in Escherichia Coli Pyridoxine 5'-Phosphate Oxidase. *Journal of Biological Chemistry* **2021**, *296*, 100795. <https://doi.org/10.1016/j.jbc.2021.100795>.
- (25) Bertoldi, M. Mammalian Dopa Decarboxylase: Structure, Catalytic Activity and Inhibition. *Archives of Biochemistry and Biophysics* **2014**, *546*, 1–7. <https://doi.org/10.1016/j.abb.2013.12.020>.
- (26) Giardina, G.; Montioli, R.; Gianni, S.; Cellini, B.; Paiardini, A.; Voltattorni, C. B.; Cutruzzola, F. Open Conformation of Human DOPA Decarboxylase Reveals the Mechanism of PLP Addition to Group II Decarboxylases. *Proceedings of the National Academy of Sciences* **2011**, *108* (51), 20514–20519. <https://doi.org/10.1073/pnas.1111456108>.
- (27) Burkhard, P.; Dominici, P.; Borri-Voltattorni, C.; Jansonius, J. N.; Malashkevich, V. N. Structural Insight into Parkinson's Disease Treatment from Drug-Inhibited DOPA Decarboxylase. *Nat. Struct. Biol.* **2001**, *8* (11), 963–967. <https://doi.org/10.1038/nsb1101-963>.

- (28) Lanzarotti, E.; Biekofsky, R. R.; Estrin, D. A.; Marti, M. A.; Turjanski, A. G. Aromatic–Aromatic Interactions in Proteins: Beyond the Dimer. *J. Chem. Inf. Model.* **2011**, *51* (7), 1623–1633. <https://doi.org/10.1021/ci200062e>.
- (29) Smits, A. H.; Vermeulen, M. Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends in Biotechnology* **2016**, *34* (10), 825–834. <https://doi.org/10.1016/j.tibtech.2016.02.014>.
- (30) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-Level Characterization of Protein–Protein Association. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (10), 4244–4249. <https://doi.org/10.1073/pnas.1815431116>.
- (31) Barile, A.; Nogués, I.; di Salvo, M. L.; Bunik, V.; Contestabile, R.; Tramonti, A. Molecular Characterization of Pyridoxine 5'-Phosphate Oxidase and Its Pathogenic Forms Associated with Neonatal Epileptic Encephalopathy. *Sci Rep* **2020**, *10* (1), 13621. <https://doi.org/10.1038/s41598-020-70598-7>.
- (32) Malerba, F.; Bellelli, A.; Giorgi, A.; Bossa, F.; Contestabile, R. The Mechanism of Addition of Pyridoxal 5'-Phosphate to *Escherichia Coli* Apo-Serine Hydroxymethyltransferase. *Biochemical Journal* **2007**, *404* (3), 477–485. <https://doi.org/10.1042/BJ20061681>.
- (33) Tramonti, A.; Nardella, C.; di Salvo, M. L.; Barile, A.; Cutruzzolà, F.; Contestabile, R. Human Cytosolic and Mitochondrial Serine Hydroxymethyltransferase Isoforms in Comparison: Full Kinetic Characterization and Substrate Inhibition Properties. *Biochemistry* **2018**, *57* (51), 6984–6996. <https://doi.org/10.1021/acs.biochem.8b01074>.
- (34) Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (35) Yang, Z.; Lasker, K.; Schneidman-Duhovny, D.; Webb, B.; Huang, C. C.; Pettersen, E. F.; Goddard, T. D.; Meng, E. C.; Sali, A.; Ferrin, T. E. UCSF Chimera, MODELLER, and IMP: An Integrated Modeling System. *Journal of Structural Biology* **2012**, *179* (3), 269–278. <https://doi.org/10.1016/j.jsb.2011.09.006>.
- (36) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein–Sequence Space with High-Accuracy Models. *Nucleic Acids Research* **2022**, *50* (D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
- (37) Park, H.; Lee, G. R.; Heo, L.; Seok, C. Protein Loop Modeling Using a New Hybrid Energy Function and Its Application to Modeling in Inaccurate Structural Environments. *PLoS ONE* **2014**, *9* (11), e113811. <https://doi.org/10.1371/journal.pone.0113811>.
- (38) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein–Protein Docking. *Nat Protoc* **2017**, *12* (2), 255–278. <https://doi.org/10.1038/nprot.2016.169>.
- (39) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802. <https://doi.org/10.1002/jcc.20289>.
- (40) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. <https://doi.org/10.1002/jcc.23354>.
- (41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (42) Eugene Kellogg, G.; Abraham, D. J. Hydrophobicity: Is LogP_{o/w} More than the Sum of Its Parts? *European Journal of Medicinal Chemistry* **2000**, *35* (7–8), 651–661. [https://doi.org/10.1016/S0223-5234\(00\)00167-7](https://doi.org/10.1016/S0223-5234(00)00167-7).

- (43) Sarkar, A.; Kellogg, G. Hydrophobicity - Shake Flasks, Protein Folding and Drug Discovery. *CTMC* **2010**, *10* (1), 67–83. <https://doi.org/10.2174/156802610790232233>.
- (44) Grant, B. J.; Skjærven, L.; Yao, X. The BIO3D Packages for Structural Bioinformatics. *Protein Science* **2021**, *30* (1), 20–30. <https://doi.org/10.1002/pro.3923>.
- (45) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319–2327. <https://doi.org/10.1002/jcc.21787>.
- (46) Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein–Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121* (36), 8133–8143. <https://doi.org/10.1021/ja990935j>.
- (47) Kortemme, T.; Kim, D. E.; Baker, D. Computational Alanine Scanning of Protein-Protein Interfaces. *Sci. STKE* **2004**, *2004* (219). <https://doi.org/10.1126/stke.2192004pl2>.
- (48) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free Energy of Ligand Binding to Protein: Evaluation of the Contribution of Water Molecules by Computational Methods. *CMC* **2004**, *11* (23), 3093–3118. <https://doi.org/10.2174/0929867043363929>.
- (49) Marabotti, A.; Spyraakis, F.; Facchiano, A.; Cozzini, P.; Alberti, S.; Kellogg, G. E.; Mozzarelli, A. Energy-Based Prediction of Amino Acid-Nucleotide Base Recognition. *J. Comput. Chem.* **2008**, *29* (12), 1955–1969. <https://doi.org/10.1002/jcc.20954>.
- (50) Dominici, P.; Moore, P. S.; Borri Voltattorni, C. Dissociation, Unfolding and Refolding Trials of Pig Kidney 3,4-Dihydroxyphenylalanine (Dopa) Decarboxylase. *Biochemical Journal* **1993**, *295* (2), 493–500. <https://doi.org/10.1042/bj2950493>.
- (51) Sherald, A. F.; Sparrow, J. C.; Wright, T. R. F. A Spectrophotometric Assay for Drosophila Dopa Decarboxylase. *Analytical Biochemistry* **1973**, *56* (1), 300–305. [https://doi.org/10.1016/0003-2697\(73\)90194-2](https://doi.org/10.1016/0003-2697(73)90194-2).
- (52) Charteris, A.; John, R. An Investigation of the Assay of Dopamine Using Trinitrobenzenesulphonic Acid. *Analytical Biochemistry* **1975**, *66* (2), 365–371. [https://doi.org/10.1016/0003-2697\(75\)90604-1](https://doi.org/10.1016/0003-2697(75)90604-1).
- (53) Montioli, R.; Cellini, B.; Borri Voltattorni, C. Molecular Insights into the Pathogenicity of Variants Associated with the Aromatic Amino Acid Decarboxylase Deficiency. *J Inherit Metab Dis* **2011**, *34* (6), 1213–1224. <https://doi.org/10.1007/s10545-011-9340-6>.
- (54) Douzi, B. Protein–Protein Interactions: Surface Plasmon Resonance. In *Bacterial Protein Secretion Systems*; Journet, L., Cascales, E., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2017; Vol. 1615, pp 257–275. https://doi.org/10.1007/978-1-4939-7033-9_21.
- (55) Pierce, M. M.; Raman, C. S.; Nall, B. T. Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods* **1999**, *19* (2), 213–221. <https://doi.org/10.1006/meth.1999.0852>.
- (56) Liang, Y. Applications of Isothermal Titration Calorimetry in Protein Science. *Acta Biochimica et Biophysica Sinica* **2008**, *40* (7), 565–576. <https://doi.org/10.1111/j.1745-7270.2008.00437.x>.
- (57) Walker, K. W. Site-Directed Mutagenesis. In *Encyclopedia of Cell Biology*; Elsevier, 2016; pp 122–127. <https://doi.org/10.1016/B978-0-12-394447-4.10010-0>.
- (58) Zerbe, B. S.; Hall, D. R.; Vajda, S.; Whitty, A.; Kozakov, D. Relationship between Hot Spot Residues and Ligand Binding Hot Spots in Protein–Protein Interfaces. *J. Chem. Inf. Model.* **2012**, *52* (8), 2236–2244. <https://doi.org/10.1021/ci300175u>.
- (59) Kruger, D. M.; Gohlke, H. DrugScorePPI Webserver: Fast and Accurate in Silico Alanine Scanning for Scoring Protein–Protein Interactions. *Nucleic Acids Research* **2010**, *38* (suppl_2), W480–W486. <https://doi.org/10.1093/nar/gkq471>.

- (60) Ramadoss, V.; Dehez, F.; Chipot, C. AlaScan: A Graphical User Interface for Alanine Scanning Free-Energy Calculations. *J. Chem. Inf. Model.* **2016**, *56* (6), 1122–1126. <https://doi.org/10.1021/acs.jcim.6b00162>.
- (61) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Research* **2005**, *33* (Web Server), W382–W388. <https://doi.org/10.1093/nar/gki387>.
- (62) Segura Mora, J.; Assi, S. A.; Fernandez-Fuentes, N. Presaging Critical Residues in Protein Interfaces-Web Server (PCRPI-W): A Web Server to Chart Hot Spots in Protein Interfaces. *PLoS ONE* **2010**, *5* (8), e12352. <https://doi.org/10.1371/journal.pone.0012352>.
- (63) Hennig, M.; Grimm, B.; Contestabile, R.; John, R. A.; Jansonius, J. N. Crystal Structure of Glutamate-1-Semialdehyde Aminomutase: An α_2 -Dimeric Vitamin B₆-Dependent Enzyme with Asymmetry in Structure and Active Site Reactivity. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (10), 4866–4871. <https://doi.org/10.1073/pnas.94.10.4866>.

CHAPTER 4

Conclusions

The various types of proteins in cells carry out a wide variety of complicated molecular functions due to their highly precise three-dimensional structures, which are determined by their genetically encoded sequences of amino acids.¹ A thorough knowledge of protein structures and interactions at the atomic level enables researchers to get a deep foundational understanding of the potential functions and enzymatic processes required for cells, resulting in more effective therapeutic interventions. Two main projects were the focus of this dissertation, both of which aimed to make use of structural data obtained from determined protein structures.

In the first project, we adopt a new motif known as "3D hydrophobic interaction maps" motif for describing protein structure and interaction.^{2,3} Many therapeutic research endeavors, such as protein structure-function analyses and structure-based drug discovery, require atomic-resolution protein structural models. Unfortunately, atomic-resolution data for some of the most therapeutically relevant proteins is frequently unavailable. As a result, accurate structural model construction (particularly from less-than-ideal experimental data) is required. 3D hydrophobic maps are information-rich descriptors of preferred conformations, interaction types and energetics, and solvent accessibility. These calculated maps can be utilized to define and characterize the hydrophobic environment around protein amino acid residues with respect to their hydrophobic and polar interactions. It can also be contoured and displayed to illustrate the types, strengths, and loci of these interactions in 3D space. In this project, we focused

on the structural roles of phenylalanine (PHE), tyrosine (TYR), and tryptophan (TRP) residues.⁴ These residues are referred to as “aromatic amino acids” as they possess an aromatic moiety in their sidechains. Their unique sidechain characters, with their conjugated planar rings, enable them to engage in multiple noncovalent hydrophobic and polar interactions. Here, we calculated and analyzed sidechain-to-environment 3D maps for over 70,000 residues of these three types that reveal, with respect to hydrophobic and polar interactions, the environment around each. Calculated maps were clustered using the k-means algorithm based on map-map pair similarities. As a result, the ~31,000 phenylalanines, ~29,000 tyrosines and ~12,000 tryptophans in the data set were clustered, respectively, into 607, 609, and 457 unique sidechain-dependent environments. A common map motif of unfavorable hydrophobic-hydrophobic interaction along the CA-CB axis was found to be shared by the three aromatic residues. The phenylalanines showed a common motif of a favorable hydrophobic-hydrophobic interaction surrounding the aromatic ring, which is more ubiquitous in the β -pleat region than in the left-hand and right-hand α -helix regions. A common motif in the tyrosines showed a favorable hydrophobic-hydrophobic interaction of tyrosine rings and a favorable polar-polar interaction of phenolic hydroxyls with their surrounding environments. The latter motif was also observed to be common in tryptophans, where hydrophobic-hydrophobic interactions of the indole ring and favorable-favorable interactions of the indole NH with their surroundings were consistently observed. Furthermore, interactions above and below the space of the sidechain rings were sampled to evaluate the magnitude/occurrence of key noncovalent interactions ,i.e., π -cation, π - π stacking, or other classes of interactions. Around half of the residues studied showed evidence of π -

π interactions (PHE: 53%, TYR: 34%, TRP: 56%); however, this accounted for only 4% of the total energy contribution. Evidence for π -cation interactions was seen in 14% of PHE, 8% of TYR and 27% of TRP residues, but this contributed only ~1% of the total energy between the residue and environment. In this project, we showed that these 3D maps are able to assemble the protein structure as combinations of backbone-oriented 3D maps selected from the clustered set by optimizing the overlaps of interaction features. We also demonstrated that these maps encode not only traditional hydrogen bonding and hydrophobic interactions, but also noncovalent interactions such as π -cation hydrogen bonding and π - π stacking. The recognition of even these subtle interactions in the 3D hydrophobic environment maps is key support for our interaction homology paradigm of protein structure elucidation and possibly prediction. Other residues, such as ALA,³ SER/CYS⁵ and ASP/GLU/HIS,⁶ were defined and characterized using our described motif. Furthermore, studies comparing membrane-based vs soluble-based residues have been carried out,⁷ with additional contributions expected in the near future.

The goal of the second project was to investigate interactions between pyridoxine-5' phosphate oxidase (PNPO), a vitamin B6-salvage enzyme, and dopa decarboxylase (DDC), a vitamin B6-dependent enzyme, using different approaches. Deficiency of cellular B6 vitamers, or mutations in the B6-salvage enzymes or key B6-dependent enzymes, is known or suspected to contribute to several pathologies, such as seizures,^{8,9} hallucination,¹⁰ convulsions,¹¹ autism,¹² Down syndrome,^{13,14} schizophrenia,¹⁵ Neonatal Epileptic Encephalopathy (NEE),¹⁶⁻²⁰ Parkinson's,^{21,22} Alzheimer's,²³⁻²⁵ and learning disabilities.^{26,27} High intake of vitamin B6 is also linked to neurotoxic effects because of

the PLP reactive aldehyde forming aldimines with free amino groups on non-B6 proteins, disrupting their function. To avoid toxicity, the cell maintains a very low concentration of free PLP.²⁸ This raises the unresolved question of how a low cellular concentration of free PLP converts hundreds of competing apo-B6 enzymes to their active holo-forms? In fact, as we learn about the health problems involving vitamin B6 deficiency, it is also clear that we have an inadequate knowledge of how this important vitamin is made available to newly synthesized apo-B6 salvage enzymes. Therefore, this project explored the mechanism by which apoDDC are converted to their catalytically active holo-form. We tested the hypothesis that the transfer of PLP from the PNPO salvage enzyme to apoDDC involves channeling, which appears to offer an efficient and protected means of delivery of the highly reactive PLP. Channeling requires that the donor and acceptor enzymes form physical contact, and the inability of such complex formation may lead to detrimental outcome.^{29,30} Due to the lack of a crystallographic structure for the PNPO•DDC complex or any other B6-salvage enzyme/B6-dependent enzyme complex, several approaches were employed, including molecular modeling, biophysical, PLP transfer kinetics, and site-directed mutagenesis studies. With the known 3D structures of hPNPO³¹ and DDC (in both apo-and holo-forms)^{32,33}, protein-protein docking and molecular dynamics simulation studies were carried out to predict the most likely near-native structure of the complex. We showed that PNPO can form a favorable complex with apo-and holo-DDC. Moreover, HINT scoring demonstrated that PNPO has a more favorable binding to apoDDC than holoDDC (8650 vs. 4746). Furthermore, this modeling studies supported the fact that B6-activation would involve PLP transfer from the allosteric binding site of PNPO, with no potential recognition of the PNPO active site (catalytic site) at the interface

of the complex was observed. The final models for PNPO•apoDDC and PNPO•holoDDC complexes were subjected to an all-atom MD simulation for further refinement. The energetics and binding mode analyses suggested that R88 and E114 of PNPO may play significant roles (large energy contributions) in this association. Later, these residues were mutated and evaluated experimentally. Following that, physical binding studies were carried out using FP, SPR, and ITC techniques. The dissociation constant (K_D) between PNPO and DDC was determined in the low micromolar range, with SPR and ITC yielding similar K_D values for the PNPO•holoDDC complex, 3.7 μM and 2.59 μM , respectively. This is within the range of K_D values reported by our groups for PNPO with several B6-dependent enzymes, ranging from 0.3 to 12.3 μM .^{29,30} The ITC was used to characterize interactions between PNPO and apoDDC, which were shown to be about three times stronger than interactions between PNPO and holoDDC, with K_D values of 0.92 ± 0.07 μM and 2.59 ± 0.11 μM , respectively. To the best of our knowledge, no binding studies have been published between any of the apo-B6 enzymes and the B6-salvage enzymes. A PLP transfer kinetic study was also accomplished in this study. The goal was to demonstrate that PNPO forms a physical complex with apoDDC for the productive transfer of the bound PLP. We showed that ~30% of the tightly-bound PLP was transferred to apoDDC to form the catalytically active holo-form. Similar findings have previously been reported by our lab, in which 30–50% of the tightly bound PLP is transferred when equal amounts of PNPO or PLKase (PLP donor) and apoSHMT (PLP acceptor) are present.^{29,30} Site mutation investigations of key residues identified by modeling studies, R88A–PNPO, E114A–PNPO, and R88A/E114A–PNPO, were carried out but showed no significant effects on the complex formation. When compared to the

wild-type PNPO, PNPO mutants showed comparable binding to apo-and holo-DDC. Therefore, further site-mutation studies are required. In this dissertation, we proposed K119–PNPO as a promising candidate for further site-directed mutagenesis study.

Overall, this multidisciplinary work aimed to integrate computational and experimental methodologies—from describing the role of aromatic residues in protein structures using 3D maps to exploring interactions between the two B6-related enzymes, PNPO and DDC. The findings of this research will contribute to a better understanding of the structural and functional roles of aromatic amino acid residues in protein structures, as well as a step forward in revealing the mechanism of apo-B6 activation.

References:

- (1) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338* (6110), 1042–1046. <https://doi.org/10.1126/science.1219021>.
- (2) Ahmed, M. H.; Koparde, V. N.; Safo, M. K.; Neel Scarsdale, J.; Kellogg, G. E. 3d Interaction Homology: The Structurally Known Rotamers of Tyrosine Derive from a Surprisingly Limited Set of Information-Rich Hydrophobic Interaction Environments Described by Maps: Tyrosine Hydrophobic Interaction Environments Define Its Rotamers. *Proteins* **2015**, *83* (6), 1118–1136. <https://doi.org/10.1002/prot.24813>.
- (3) Ahmed, M. H.; Catalano, C.; Portillo, S. C.; Safo, M. K.; Neel Scarsdale, J.; Kellogg, G. E. 3D Interaction Homology: The Hydrophobic Interaction Environments of Even Alanine Are Diverse and Provide Novel Structural Insight. *Journal of Structural Biology* **2019**, *207* (2), 183–198. <https://doi.org/10.1016/j.jsb.2019.05.007>.
- (4) AL Mughram, M. H.; Catalano, C.; Bowry, J. P.; Safo, M. K.; Scarsdale, J. N.; Kellogg, G. E. 3D Interaction Homology: Hydrophobic Analyses of the “ π -Cation” and “ π - π ” Interaction Motifs in Phenylalanine, Tyrosine, and Tryptophan Residues. *J. Chem. Inf. Model.* **2021**, *61* (6), 2937–2956. <https://doi.org/10.1021/acs.jcim.1c00235>.
- (5) Catalano, C.; AL Mughram, M. H.; Guo, Y.; Kellogg, G. E. 3D Interaction Homology: Hydrophobic Interaction Environments of Serine and Cysteine Are Strikingly Different and Their Roles Adapt in Membrane Proteins. *Current Research in Structural Biology* **2021**, *3*, 239–256. <https://doi.org/10.1016/j.crstbi.2021.09.002>.
- (6) Herrington, N. B.; Kellogg, G. E. 3D Interaction Homology: Computational Titration of Aspartic Acid, Glutamic Acid and Histidine Can Create PH-Tunable Hydrophobic Environment Maps. *Front. Mol. Biosci.* **2021**, *8*, 773385. <https://doi.org/10.3389/fmolb.2021.773385>.
- (7) AL Mughram, M. H.; Herrington, N. B.; Catalano, C.; Kellogg, G. E. Systematized Analysis of Secondary Structure Dependence of Key Structural Features of Residues in Soluble and Membrane-Bound Proteins. *Journal of Structural Biology: X* **2021**, *5*, 100055. <https://doi.org/10.1016/j.yjsbx.2021.100055>.
- (8) Jang, H.-S.; Roh, S. Y.; Jeong, E. H.; Kim, B.-S.; Sunwoo, M. K. Ginkgotoxin Induced Seizure Caused by Vitamin B6 Deficiency. *J Epilepsy Res* **2015**, *5* (2), 104–106. <https://doi.org/10.14581/jer.15018>.
- (9) Mills, P. B.; Struys, E.; Jakobs, C.; Plecko, B.; Baxter, P.; Baumgartner, M.; Willemsen, M. A. A. P.; Omran, H.; Tacke, U.; Uhlenberg, B.; Weschke, B.; Clayton, P. T. Mutations in Antiquitin in Individuals with Pyridoxine-Dependent Seizures. *Nat. Med.* **2006**, *12* (3), 307–309. <https://doi.org/10.1038/nm1366>.
- (10) Toriumi, K.; Miyashita, M.; Suzuki, K.; Yamasaki, N.; Yasumura, M.; Horiuchi, Y.; Yoshikawa, A.; Asakura, M.; Usui, N.; Itokawa, M.; Arai, M. Vitamin B6 Deficiency Hyperactivates the Noradrenergic System, Leading to Social Deficits and Cognitive Impairment. *Transl Psychiatry* **2021**, *11* (1), 262. <https://doi.org/10.1038/s41398-021-01381-z>.
- (11) McCormick, D. B.; Chen, H. Update on Interconversions of Vitamin B-6 with Its Coenzyme. *J. Nutr.* **1999**, *129* (2), 325–327. <https://doi.org/10.1093/jn/129.2.325>.
- (12) Adams, J. B.; George, F.; Audhya, T. Abnormally High Plasma Levels of Vitamin B6 in Children with Autism Not Taking Supplements Compared to Controls Not Taking Supplements. *J Altern Complement Med* **2006**, *12* (1), 59–63. <https://doi.org/10.1089/act.2006.12.59>.
- (13) Coburn, S. P.; Mahuren, J. D.; Schaltenbrand, W. E. Increased Activity of Pyridoxal Kinase in Tongue in Down’s Syndrome. *J Ment Defic Res* **1991**, *35* (Pt 6), 543–547.

- (14) Share, J. B. Review of Drug Treatment for Down's Syndrome Persons. *Am J Ment Defic* **1976**, *80* (4), 388–393.
- (15) Song, H.; Ueno, S.; Numata, S.; Iga, J.; Shibuya-Tayoshi, S.; Nakataki, M.; Tayoshi, S.; Yamauchi, K.; Sumitani, S.; Tomotake, T.; Tada, T.; Tanahashi, T.; Itakura, M.; Ohmori, T. Association between PNPO and Schizophrenia in the Japanese Population. *Schizophr. Res.* **2007**, *97* (1–3), 264–270. <https://doi.org/10.1016/j.schres.2007.08.004>.
- (16) Balasubramaniam, S.; Bowling, F.; Carpenter, K.; Earl, J.; Chaitow, J.; Pitt, J.; Mornet, E.; Sillence, D.; Ellaway, C. Perinatal Hypophosphatasia Presenting as Neonatal Epileptic Encephalopathy with Abnormal Neurotransmitter Metabolism Secondary to Reduced Co-Factor Pyridoxal-5'-Phosphate Availability. *J. Inherit. Metab. Dis.* **2010**, *33 Suppl 3*, S25–33. <https://doi.org/10.1007/s10545-009-9012-y>.
- (17) Barile, A.; Nogués, I.; di Salvo, M. L.; Bunik, V.; Contestabile, R.; Tramonti, A. Molecular Characterization of Pyridoxine 5'-Phosphate Oxidase and Its Pathogenic Forms Associated with Neonatal Epileptic Encephalopathy. *Sci Rep* **2020**, *10* (1), 13621. <https://doi.org/10.1038/s41598-020-70598-7>.
- (18) Musayev, F. N.; Di Salvo, M. L.; Saavedra, M. A.; Contestabile, R.; Ghatge, M. S.; Haynes, A.; Schirch, V.; Safo, M. K. Molecular Basis of Reduced Pyridoxine 5'-Phosphate Oxidase Catalytic Activity in Neonatal Epileptic Encephalopathy Disorder. *Journal of Biological Chemistry* **2009**, *284* (45), 30949–30956. <https://doi.org/10.1074/jbc.M109.038372>.
- (19) Ghatge, M. S.; Al Mughram, M.; Omar, A. M.; Safo, M. K. Inborn Errors in the Vitamin B6 Salvage Enzymes Associated with Neonatal Epileptic Encephalopathy and Other Pathologies. *Biochimie* **2021**, *183*, 18–29. <https://doi.org/10.1016/j.biochi.2020.12.025>.
- (20) Alghamdi, M.; Bashiri, F. A.; Abdelhakim, M.; Adly, N.; Jamjoom, D. Z.; Sumaily, K. M.; Alghanem, B.; Arold, S. T. Phenotypic and Molecular Spectrum of Pyridoxamine-5'-phosphate Oxidase Deficiency: A Scoping Review of 87 Cases of Pyridoxamine-5'-phosphate Oxidase Deficiency. *Clinical Genetics* **2021**, *99* (1), 99–110. <https://doi.org/10.1111/cge.13843>.
- (21) Burkhard, P.; Dominici, P.; Borri-Voltattorni, C.; Jansonius, J. N.; Malashkevich, V. N. Structural Insight into Parkinson's Disease Treatment from Drug-Inhibited DOPA Decarboxylase. *Nat. Struct Biol.* **2001**, *8* (11), 963–967. <https://doi.org/10.1038/nsb1101-963>.
- (22) Alonso-Navarro, H.; Jimenez-Jimenez, F.; Garcia-Martin, E.; Agundez, J. Genomic and Pharmacogenomic Biomarkers of Parkinson's Disease. *CDM* **2014**, *15* (2), 129–181. <https://doi.org/10.2174/138920021502140327175404>.
- (23) Aoyagi, T.; Wada, T.; Nagai, M.; Kojima, F.; Harada, S.; Takeuchi, T.; Takahashi, H.; Hirokawa, K.; Tsumita, T. Increased Gamma-Aminobutyrate Aminotransferase Activity in Brain of Patients with Alzheimer's Disease. *Chem. Pharm. Bull.* **1990**, *38* (6), 1748–1749.
- (24) Mielech, A.; Puścion-Jakubik, A.; Markiewicz-Żukowska, R.; Socha, K. Vitamins in Alzheimer's Disease—Review of the Latest Reports. *Nutrients* **2020**, *12* (11), 3458. <https://doi.org/10.3390/nu12113458>.
- (25) An, Y.; Feng, L.; Zhang, X.; Wang, Y.; Wang, Y.; Tao, L.; Qin, Z.; Xiao, R. Dietary Intakes and Biomarker Patterns of Folate, Vitamin B6, and Vitamin B12 Can Be Associated with Cognitive Impairment by Hypermethylation of Redox-Related Genes NUDT15 and TXNRD1. *Clin Epigenet* **2019**, *11* (1), 139. <https://doi.org/10.1186/s13148-019-0741-y>.
- (26) Hassel, B.; Rogne, A. G.; Hope, S. Intellectual Disability Associated With Pyridoxine-Responsive Epilepsies: The Need to Protect Cognitive Development. *Front. Psychiatry* **2019**, *10*, 116. <https://doi.org/10.3389/fpsy.2019.00116>.

- (27) Belardo, A.; Gevi, F.; Zolla, L. The Concomitant Lower Concentrations of Vitamins B6, B9 and B12 May Cause Methylation Deficiency in Autistic Children. *The Journal of Nutritional Biochemistry* **2019**, *70*, 38–46. <https://doi.org/10.1016/j.jnutbio.2019.04.004>.
- (28) di Salvo, M. L.; Contestabile, R.; Safo, M. K. Vitamin B6 Salvage Enzymes: Mechanism, Structure and Regulation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011**, *1814* (11), 1597–1608. <https://doi.org/10.1016/j.bbapap.2010.12.006>.
- (29) Ghatge, M. S.; Contestabile, R.; Salvo, M. L.; Desai, J. V.; Gandhi, A. K.; Camara, C. M.; Florio, R.; González, I. N.; Parroni, A.; Schirch, V.; Safo, M. K. Pyridoxal 5'-Phosphate Is a Slow Tight Binding Inhibitor of E. Coli Pyridoxal Kinase. *PLoS ONE* **2012**, *7* (7), e41680. <https://doi.org/10.1371/journal.pone.0041680>.
- (30) Ghatge, M. S.; Karve, S. S.; David, T. M. S.; Ahmed, M. H.; Musayev, F. N.; Cunningham, K.; Schirch, V.; Safo, M. K. Inactive Mutants of Human Pyridoxine 5'-Phosphate Oxidase: A Possible Role for a Noncatalytic Pyridoxal 5'-Phosphate Tight Binding Site. *FEBS Open Bio* **2016**, *6* (5), 398–408. <https://doi.org/10.1002/2211-5463.12042>.
- (31) Musayev, F. N.; Di Salvo, M. L.; Ko, T.-P.; Schirch, V.; Safo, M. K. Structure and Properties of Recombinant Human Pyridoxine 5'-Phosphate Oxidase. *Protein Sci.* **2003**, *12* (7), 1455–1463. <https://doi.org/10.1110/ps.0356203>.
- (32) Burkhard, P.; Dominici, P.; Borri-Voltattorni, C.; Jansonius, J. N.; Malashkevich, V. N. Structural Insight into Parkinson's Disease Treatment from Drug-Inhibited DOPA Decarboxylase. *Nat. Struct Biol.* **2001**, *8* (11), 963–967. <https://doi.org/10.1038/nsb1101-963>.
- (33) Giardina, G.; Montioli, R.; Gianni, S.; Cellini, B.; Paiardini, A.; Voltattorni, C. B.; Cutruzzola, F. Open Conformation of Human DOPA Decarboxylase Reveals the Mechanism of PLP Addition to Group II Decarboxylases. *Proceedings of the National Academy of Sciences* **2011**, *108* (51), 20514–20519. <https://doi.org/10.1073/pnas.1111456108>.

VITA

Mohammed H. AL Mughram was born on September 30, 1990, in Billasmar City, Saudi Arabia, and is a Saudi citizen. He graduated from King Abdullah High School, Abha, Saudi Arabia in 2008. In 2013, he received his Bachelor of Science in Pharmaceutical Science from King Khalid University in Abha, Saudi Arabia, and in 2018, he received his Master of Science in Biological and Medicinal Chemistry from the University of Toledo in Toledo, Ohio, USA. In August 2018, he started his graduate doctoral studies at the School of Pharmacy at Virginia Commonwealth University in Richmond, VA, with a specialization in Medicinal Chemistry.