



Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2011

Water Molecules: A Closer Look at Their Behavior at Protein-Protein Interfaces and Their Contributions to the Docked Model of Pyridoxal Kinase - Serine Hydroxymethyltransferase Complex

Mostafa H. Ahmed
Virginia Commonwealth University

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

 Part of the [Pharmacy and Pharmaceutical Sciences Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/2578>

This Thesis 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.

© Mostafa H. Ahmed 2011
All Rights Reserved

WATER MOLECULES: A CLOSER LOOK AT THEIR BEHAVIOR AT PROTEIN-
PROTEIN INTERFACES AND THEIR CONTRIBUTIONS TO THE DOCKED
MODEL OF PYRIDOXAL KINASE – SERINE HYDROXYMETHYLTRANSFERASE
COMPLEX

A thesis submitted in partial fulfillment of the requirements for the degree of Master of
Science at Virginia Commonwealth University.

By

MOSTAFA AHMED

B.Sc. Misr International University, Cairo, Egypt 2006

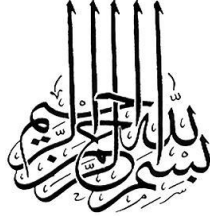
Advisors: GLEN EUGENE KELLOGG, Ph.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY &
INSTITUTE FOR STRUCTURAL BIOLOGY AND DRUG DISCOVERY

MARTIN K. SAFO, Ph.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY &
INSTITUTE FOR STRUCTURAL BIOLOGY AND DRUG DISCOVERY

Virginia Commonwealth University
Richmond, Virginia
August 2011



أَوَلَمْ يَرِ الَّذِينَ كَفَرُوا أَنَّ السَّمَوَاتِ
وَالْأَرْضَ كَانَتَا رَتْقًا فَفَتَقْنَاهُمَا ۖ وَجَعَلْنَا
مِنَ الْمَاءِ كُلَّ شَيْءٍ حَيٍّ ۖ أَفَلَا
يُؤْمِنُونَ ﴿٣١﴾

القرآن الكريم
سورة الأنبياء آية ٣١

In the name of God, Most Gracious, Most Merciful
Do not the disbelievers see that the heavens and the earth were *a* closed-up *mass*, then
We opened them out? And We made from water every living thing. Will they not then
believe?

Holy Quran
Al-Anbiya “The Prophets” Chapter 21: Verse 31

Acknowledgment

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, my utmost gratitude to my advisors Dr. Glen E. Kellogg, Ph.D. Associate Professor, Department of Medicinal Chemistry & Institute of Structural Biology and Drug Discovery and Dr. Martin K. Safo, Ph.D. Associate Professor, Department of Medicinal Chemistry & Institute of Structural Biology and Drug Discovery whose sincerity and encouragement I will never forget. They have been my inspiration as I hurdle all the obstacles I faced in the completion of this research work and I learned a lot from them. They provided me with their support all the way.

Dr. J. Neel Scarsdale, Ph.D. for serving as a member of my graduate student committee.

Dr. Philip Mosier, Ph.D., for the valuable insights he has shared.

The Fulbright commission who financially supported my M.Sc. study that produced this work. All the people there were very helpful.

My wife Esraa, who went through a lot with me and who always kept encouraging me to go on. She transformed my life for the better. I would also like to thank her family for all of their support.

The teaching assistants of the advanced molecular modeling class Mr. Chengxiao Da and Ms. May Abdel Aziz.

My fellow colleagues in Dr. Kellogg's and Dr. Safo's groups for their constant support.

My Friend Mr. Osama Shoair, my life in Richmond would have been harder if it were not for him and Mr. Osama El-Mahdy, whom have always been there for me. I value their friendship.

My father and mother, as I would not be who I am now if it was not for them.

The Institute of Structural Biology and Drug Discovery and the Medicinal Chemistry Department, School of Pharmacy, Virginia Commonwealth University who gave me this opportunity to pursue my graduate studies.

Last but not the least, the one above us all, God, for answering my prayers and for giving me the strength to continue on. I would not have accomplished everything if it were not for God in the first place. Thank you so much Dear Lord.

TABLE OF CONTENTS

List of Tables.....	ix
List of Figures.....	x
List of Abbreviations.....	xi
Abstract.....	xii
Chapter	Page
1 General Introduction.....	1
1.1 Role of water in protein-protein complex interface.....	1
1.1.1 Water the vivacious molecule.....	1
1.1.2 Different roles of water in biological processes.....	1
1.1.2.1 Role of water molecules in protein-ligand binding...	2
1.1.2.2 Role of water molecules in protein-DNA binding....	3
1.1.2.3 Role of water molecules in protein-protein binding..	4
1.2 Pyridoxal kinase (PLK) – Serine hydroxymethyltransferase (SHMT) complex.....	8
1.2.1 Pyridoxal 5'-Phosphate (PLP)-dependent enzymes.....	8
1.2.2 Pyridoxal Kinase (PLK).....	8
1.2.3 Pyridoxal 5'-Phosphate (PLP) metabolic pathways.....	10
1.2.4 Mechanism of transfer of PLP from PLK to SHMT and other	

PLP-dependent enzymes.....	10
1.3 HINT (Hydropathic INteractions).....	14
1.4 Rational and specific aims.....	14
1.4.1 Analysis of water molecules at protein-protein interfaces.....	14
1.4.2 Developing a model for PLK-SHMT interaction.....	16
References.....	18
2. Water molecules at protein-protein interface: a detailed analysis and quantification of their contributions with respect to different residue types	27
2.1 Introduction.....	27
2.2 Materials and Methods.....	28
2.2.1 Data set.....	28
2.2.2 Hydropathic Analysis.....	32
2.2.3 Rank Algorithm.....	34
2.2.4 Relevance.....	35
2.3 Results and discussion.....	35
2.3.1 The Water Relevance Metric.....	35
2.3.2 Residue Preferences for Interfacial H ₂ O.....	43
2.3.3 Sidechain and Backbone Preferences for Interface water.....	47
2.3.4 Residue-Pair Preferences for Interface H ₂ O.....	52

2.3.5 Residue-Pair Roles in Water Interactions.....	56
2.3.6 Waters Relevant to Multiple Proteins.....	59
2.3.7 Waters not Relevant to either protein.....	63
2.3.8 Predictions of water roles.....	67
2.4 Conclusion.....	68
References.....	70
3. Pyridoxal kinase-serine hydroxymethyltransferase complex model.....	79
3.1 Introduction.....	79
3.1.1 Different forms of vitamin B ₆	79
3.1.2 Toxic effects of high concentrations of Pyridoxal 5'- Phosphate (PLP).....	81
3.2 Materials and Methods.....	82
3.2.1 Preparation of crystal structures.....	82
3.2.2 Predicting the active residues for the protein-protein complex.	83
3.2.3 SASA calculation.....	83
3.2.4 Protein-protein docking.....	84
3.2.5 Candidate model refinement.....	85
3.2.6 Hotspot prediction.....	86
3.3 Results and Discussion.....	86

3.3.1 Predicted PLK and SHMT tunnels.....	86
3.3.2 Predicted active residues for PLK-SHMT complex.....	91
3.3.3 Protein-protein docking results.....	96
3.3.4 HINT analysis and water relevance.....	98
3.4 Conclusion.....	105
References.....	106
4. Conclusions.....	112
Vita.....	116

LIST OF TABLES

Table 2.1 Protein complexes examined in study with interface parameters and water roles.....	29
Table 2.2 Water metrics for human placental RNase inhibitor (hRI)- human angiogenin (hAng) complex (PDB 1a4y, 2.00 Å).....	39
Table 2.3 Frequencies and HINT scores of water molecules at protein-protein interfaces with respect to interacting amino acid residues.....	45
Table 2.4 Frequencies and HINT scores of water molecules at protein-protein interfaces with respect to backbones and sidechains of interacting amino acid residues.....	49
Table 3.1 Different forms of vitamin B ₆	80
Table 3.2 Residues forming the tunnel walls for PLK and SHMT.....	90
Table 3.3 Global Energy and HINT scores of 8 candidate models.....	97
Table 3.4 Calculated HINT scores for model 7_2.....	99
Table 3.5 HINT water rank for the water molecules added to model 7_2.....	102
Table 3.6. The effect of water molecules relevant to both PLK and SHMT in model 7_2.....	105

LIST OF FIGURES

Figure 1.1 Pyridoxal Kinase catalyzed reactions.....	9
Figure 1.2 Tryptophan synthase channel.....	12
Figure 2.1 Molecular model of human placental RNase inhibitor (hRI) (red)- human angiogenin (hAng) (blue) complex (1a4y).....	37
Figure 2.2 Relative fractions of waters with Relevance to neither (green), one (red) and both (blue) proteins.....	42
Figure 2.3 Average HINT interaction scores for waters at protein-protein interfaces	51
Figure 2.4 Color heat maps depicting Res1-H ₂ O-Res2 interactions for water molecules found at protein-protein interfaces.....	55
Figure 2.5 Dendograms indicating clustering of residues with respect to average HINT score (normalized by weighted count) in Res1-H ₂ O-Res2 interaction.....	57
Figure 2.6 Water as a nano-scale buffer.....	62
Figure 2.7 Water in chain of three water molecules.....	64
Figure 2.8 Interaction type scores for waters with Relevance to zero, one and two proteins.....	65
Figure 3.1 Sketch of the computational algorithm implemented in CAVER.....	88
Figure 3.2 Predicted tunnels connecting PLP active site and the bulk solvent.....	89
Figure 3.3 Predicted active residues for PLK-SHMT complex.....	92
Figure 3.4 Electrostatic maps of both posterior and anterior sides.....	95
Figure 3.5 The predicted tunnel connecting PLK and SHMT.....	98

LIST OF ABBREVIATIONS

HINT	Hydrophathic INTeractions a group of tools that employs a natural force field based on empirical energetic terms developed by Glen E. Kellogg
MD	Molecular Dynamics
PA	4-Pyridoxic Acid
PL	Pyridoxal
PLK	Pyridoxal Kinase enzyme
PLP	Pyridoxal 5'-Phosphate
PM	Pyridoxamine
PMP	Pyridoxamine 5'-Phosphate
PN	Pyridoxine
PNP	Pyridoxine 5'-Phosphate
PNPOx	Pyridoxine 5'-Phosphate Oxidase enzyme
SASA	Solvent Accessible Surface Area
SHMT	Serine Hydroxymethyltransferase

Abstract

WATER MOLECULES: A CLOSER LOOK AT THEIR BEHAVIOR AT PROTEIN- PROTEIN INTERFACES AND THEIR CONTRIBUTIONS TO THE DOCKED MODEL OF PYRIDOXAL KINASE – SERINE HYDROXYMETHYLTRANSFERASE COMPLEX

By Mostafa H. Ahmed, M.Sc.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of
Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2011.

Advisors: GLEN EUGENE KELLOGG, Ph.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY &
INSTITUTE FOR STRUCTURAL BIOLOGY AND DRUG DISCOVERY

MARTIN K. SAFO, Ph.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY &
INSTITUTE FOR STRUCTURAL BIOLOGY AND DRUG DISCOVERY

The work in this thesis is divided into two aims. The first aim is to provide a detailed analysis of water molecules at protein-protein interfaces as well as quantifying their contributions with respect to different residue types. To achieve this aim a data set of 4741 water molecules abstracted from 179 high-resolution (≤ 2.30 Å) X-ray crystal structures of protein-protein complexes was analyzed with a suite of modeling tools

based on HINT. The second aim is to observe the effect of adding interfacial water molecules in developing a model for the protein-protein interaction between pyridoxal kinase and serine hydroxymethyltransferase. This model was created to explore the possibility of the formation of a channel between the two proteins upon interaction providing a safe way to transport the substrate pyridoxal 5'-phosphate (active form of vitamin B₆). This work demonstrates a substantial progress in the understanding of the role of water molecules in protein-protein binding.

CHAPTER 1

GENERAL INTRODUCTION

1.1 Role of water in protein-protein complex interface:

1.1.1 Water the vivacious molecule:

Water is a vital component in all living organisms. It plays various roles in different biochemical processes. For macromolecules, water is crucial for maintaining structure and mediating molecular recognition, it provides a way of communication across membranes and between the inside and outside of proteins [1]. Although it may seem that water's chemical and physical properties are similar to those of other polar solvents, it is very hard to imagine any other solvent that could fulfill all of its roles, especially in biology. Water molecules are unique in their ability to engage in four directional hydrogen bonds in a way that allows for easy and rapid reorientation and reconfiguration into different three-dimensional structures.

1.1.2 Different roles of water in biological processes:

When it comes to biological models, water is often described as an inactive constituent. However, water plays a central role in many life processes. Water molecules are known to mediate protein folding [2,3]. In addition, the presence of water molecules was found to enhance and tune functions of proteins. Ohno *et al.* found by using quantum-chemical methods that water molecules helped to enhance the catalytic activity of ribonuclease T₁ in addition to maintaining

its structure [4]. Okada *et al.* also showed that two water molecules play a central role in tuning the central chromophore of rhodopsin, retinal, to different wavelengths in the red, green, and blue cone cells of the retina [5]. Moreover, water molecules were found to be directly involved in the catalytic action of some enzymes. For example, a water molecule in the bacterial enzyme zinc lactamase, acts as a nucleophile to initiate splitting of the lactam ring, a mechanism whereby bacteria resists lactam antibiotics [6]. In addition, water molecules are involved in electron transfer between proteins and other macromolecules as evident by two ordered water molecules bound at the interface between the redox centers of cross-linked azurin proteins, which appeared to assist significantly in electron transfer [7].

1.1.2.1 Role of water molecules in protein-ligand binding:

Generally, small molecule modulators are designed to replace water molecules in the protein active site. However, these waters might also have other roles. Retained water molecules in the active site can make the binding surface highly adaptable and can act as extensions to residues for assisting in the specificity of substrate binding. To prove the importance of including water molecules in ligand binding calculations, Kellogg and co-workers analyzed the interactions between 23 ligands and HIV-1 protease. They found that the inclusion of bridging water molecules results a significant improvement of the correlation between HINT scores (*vide infra*) and the experimentally determined binding constants (r^2 improved from 0.30 to 0.61) [8-14].

In a subsequent study, Kellogg and co-workers used HINT score and Rank to predict the role of water molecules in protein active sites. They were particularly interested in three categories of water molecules: waters with high Rank and HINT scores, waters with moderate Rank and high HINT scores and waters with low Rank and HINT scores. Water molecules from

the first category are unlikely to make additional interactions with the ligand and are mostly irrelevant to the binding process, while water molecules in the second category are available to interact with ligands. Water molecules in the third category were found to be easily displaced from the protein's active site due to steric reasons [12]. These results emphasized the importance of mapping water contribution to the energetics of ligand-protein binding. The same principles could also apply to bio-macromolecular associations as in protein-protein and protein-DNA recognition.

1.1.2.2 Role of water molecules in protein-DNA binding:

In another analysis by Kellogg and co-workers done on a dataset of 100 high-resolution protein-DNA structures using HINT, it was found that about 22% of water molecules mediating the protein-DNA recognition (located within 4.0 Å of both protein and DNA at the complex interface) act as protein-DNA linkers. In addition, it was found that water-mediated interaction between Adenine and positively charged or H-bond donor residues like Arg, Lys, Asn, Tyr, His, and Ser likely relieve electrostatic repulsion between H-bond donor groups. Cytosine was typically involved in water-mediated interactions with Asp and Glu, in which water act as an extension arm for the short amino acid side. The probability of an interaction between Thr-A and Lys-A were observed to exhibit about tenfold and fivefold increases, respectively, when there is a water molecule involved in the interaction [10].

The above observations were in agreement with others observed in earlier studies. Luscombe *et al.* analyzed 129 protein-DNA complexes and found that water molecules mediate approximately 16% of the protein-DNA interactions. In addition, the authors indicated that the distribution of water-mediated hydrogen bonds and direct hydrogen bonds are comparable. They

also observed the extensive use of aspartate and glutamate, for which the unfavorable electrostatic charge is minimized by water interaction [15]. Reddy and co-workers analyzed 109 unique protein-DNA complexes. Their analysis was based on the chemical identity of macromolecular atoms proximal to the interfacial water molecules. They inferred that most of water molecules serve to buffer electrostatic repulsions between electronegative atoms of the DNA and the protein and that only 2 % of the observed water molecules act as linkers to form hydrogen bonds that compensate for the lack of a direct hydrogen bond. They also indicated that water molecules at the interface of protein-DNA complexes can play additional roles such as: mediating interaction and specific recognition by contacting both protein and DNA, acting as solvating agents or buffering electrostatic repulsions between protein polar residues and DNA phosphate groups, and waters contact only with other water molecules, thus forming a hydrogen-bond network [16].

1.1.2.3 Role of water molecules in protein-protein binding:

Similarly, water proved to have a central role in protein-protein complex formation. The stability of protein-protein complexes depends on a complicated network of non-covalent bonds such as ionic interactions, hydrophobic interactions and hydrogen bonds. The energies of these types of bonds are in the range of 2–6 kcal/mol, which is considered weak [17]. Therefore, a large number of non-covalent interactions are essential for the formation of stable protein-protein complexes. Water molecules within cavities formed between two bounded proteins play a fundamental role for the formation and stabilization of the protein-protein complexes [18]. 10%–20% of the interface areas of protein complexes are made up of cavities in which at least one water molecule was observed [19,20].

In a review by Lazaridis *et al.*, the authors concluded that water molecules greatly influence the thermodynamic properties of binding of biomolecules. Interfacial waters involved in hydrogen bonds make a negative contribution to the entropy, enthalpy, and heat capacity of binding, while, waters that do not form hydrogen bonds can have higher entropy than in the bulk. In addition, they noted that water-mediated interactions can be as strong as direct interactions [18]. Papoian, Ulander and Wolynes applied energy landscape theory to evaluate water-mediated recognition [21]. Keskin and Nussinov have described water inclusion as an alternative strategy for proteins to achieve optimum association [22,23].

A small number of research groups have worked on understanding the importance of interfacial waters to protein-protein complexes over the last several years. Notably, Baker and co-workers described a simple model for the energetics of water-mediated hydrogen bonds, which improved the prediction of free-energy changes upon mutation at protein-protein interfaces. They also described a “solvated rotamer” approach for the prediction of water molecules positions, at protein-protein interfaces and in monomeric proteins [24]. In another study, Backer and colleagues reengineered the protein-protein interface of colicin E7 and DNase-Im7, which improved their specificity by 30-fold. However, when they designed a de novo hydrogen bond network by mutating some of the residues at the interface to displace highly conserved water, it resulted in a 300-fold increase in specificity. These results were further confirmed by solving the crystallographic structure for this reengineered complex [25].

Janin and co-workers pointed out that bridging protein-water-protein H-bonds are nearly as abundant as direct protein-protein H-bonds [26]. In another study, Janin and co-workers did an analysis of the water molecules trapped at the protein-protein interfaces of 115 homodimeric

proteins and 46 protein–protein complexes, and compared them with 173 large crystal packing interfaces representing nonspecific interactions. They observed different patterns of hydrations: packed interfaces have an average of 15 waters per 1000 Å² of interface area while homodimeric interfaces have 10–11 waters per 1000 Å². They also observed that water molecules permeate the majority of packed interfaces, which they termed “wet” interfaces, whereas in homodimers the majority of water molecules form a ring around the interface, in which case they termed the interface “dry”. Also worth noting was that water molecules at interfaces prefer to form hydrogen bonds with the main-chain carbonyl and the charged sidechains of Glu, Asp, and Arg that are more or less the same in homodimers and packed interfaces. These interactions are similar to those observed on other parts of the protein surface [27].

Pisabarro and co-workers performed an MD study on 17 protein complexes from two families of different interfacial nature. They showed that water molecules in protein interfaces contribute to the conservation of protein interactions by allowing sequence variability in the interacting partners. They have also shown that interfacial residues interacting through water are more mobile than directly interacting residues, but less mobile than solvent exposed residues. They also observed that water molecules involved in protein-water-protein interactions have significantly longer residence time than those on the protein surface [28]. In another analysis, Pisabarro and co-workers divided a dataset of protein-protein interfaces into obligate and transient interfaces. Obligate being defined as proteins that bind and fold simultaneously while transient referring to proteins that fold separately and then bind. They found that 40.1% of the interfacial residues are interacting through water and that an average of five water molecules per 1000 Å² mediate interactions between the two interfaces. Moreover, they found that 14.5% of the interfacial residues only interact through a water molecule and termed them “wet spots”. They

also observed that interacting residue pairs vary whether the residue is interacting directly “dual” or not “wet spot”. Dually interacting residues interact more frequently by their long polar sidechains, whereas wet spots prefer to interact mainly by their main-chain and short polar sidechains. It was also noted that the contribution of wet spots is quite significant. They concluded that the role of water molecules in transient interfaces is mainly to hydrate the charged sidechains, whereas in obligate interfaces water molecules tend to mediate a broad range of main-chain interactions to complement the hydrophobic interactions forming the interface [29].

Kellogg and colleagues used HINT to model the free energy of dimer–tetramer association in several deoxy-hemoglobin double mutants that have been solved crystallographically and characterized thermodynamically. Initially estimated free energies for these mutants were conducted without including crystallographically conserved water molecules, which resulted in an underestimation of the experimentally calculated loss in free energy observed for each mutant dimer–tetramer association. Conversely, when crystallographic waters interacting at the dimer–dimer interface of each mutant were included, free energies that are more accurate were estimated with respect to experimental data. This study showed that differences in the stability of bound water molecules among other things contribute to free energy changes observed for each mutant structure. Surprisingly they found that bound waters may account for up to 100% of observed free energy changes, and on average accounted for approximately 15% of the total estimated free energy change in those mutants [11].

1.2 Pyridoxal kinase (PLK) – Serine hydroxymethyltransferase (SHMT) complex:

1.2.1 Pyridoxal 5'-phosphate (PLP)-dependent enzymes:

Vitamin B₆ refers to six interconvertible compounds: pyridoxine (PN), pyridoxamine (PM), pyridoxal (PL) and their 5'-phosphorylated forms (PNP, PMP and PLP, respectively). PLP is the biologically active and arguably the most important vitamin in nature, since it is used as enzyme cofactor by several enzymes. PLP-dependent enzymes catalyze several important biochemical reactions such as amino acid and lipid metabolism, carbohydrate breakdown, neurotransmitter synthesis, heme synthesis, nucleic acid synthesis. In recent years, an additional function of B₆ vitamers (different vitamin B₆ forms) as reactive oxygen species (ROS) scavengers and factors able to increase resistance to biotic and abiotic stress has been demonstrated in plants [30, 31]. PLP and PN may also function as regulators of membrane ion transporters [32-33], and have been found to bind to steroid receptors [34] and to modulate transcription factors [35,36]. Although all living beings rely on vitamin B₆ for their existence, only microorganisms and plants are able to synthesize it *de novo*. All other organisms, including mammals acquire vitamin B₆ from nutrients and interconvert its different forms to PLP. The enzymes involved in the biosynthesis of the B₆ into PLP are pyridoxal kinase and pyridoxine 5'-phosphate oxidase via B₆ salvage pathway [37-40].

1.2.2 Pyridoxal Kinase (PLK):

PLK phosphorylates the 5' alcohol group of PN, PL and PM to form PNP, PLP and PMP respectively. This reaction occurs by the transfer of γ -phosphate from ATP to the 5'-methyl hydroxyl group, as shown in Figure 1.1. PLK is found in most organisms. It is encoded by the gene *pdxK*, which is highly homologous among prokaryotes and eukaryotes [38]. Some

organisms possess an additional PLK beside the one coded by the *pdxK* gene. This is termed PLK 2 and coded by the gene *pdxY* [41]. PLK 2 shares very low sequence identity (~30%) with PL kinase [41]. This protein also functions in the salvage pathway. However, it has a much lower activity than PLK, which makes its exact role in vitamin B₆ metabolism unclear. The PNP and PMP synthesized by PLK are further oxidized to PLP by a flavin mononucleotide (FMN)-dependent pyridoxine 5'-phosphate (or pyridoxamine 5'-phosphate) oxidase (PNPOx) [42]. PLP synthesized by these two enzymes is then transferred to apo B₆ (PLP-dependent) enzymes to form the holo B₆ enzymes for their catalytic activities. In a salvage pathway, PLP, PNP and PMP that are ingested or already in vivo are dephosphorylated by phosphatase class of enzymes to PL, PN and PM respectively, and then recycled to form PLP as described above.

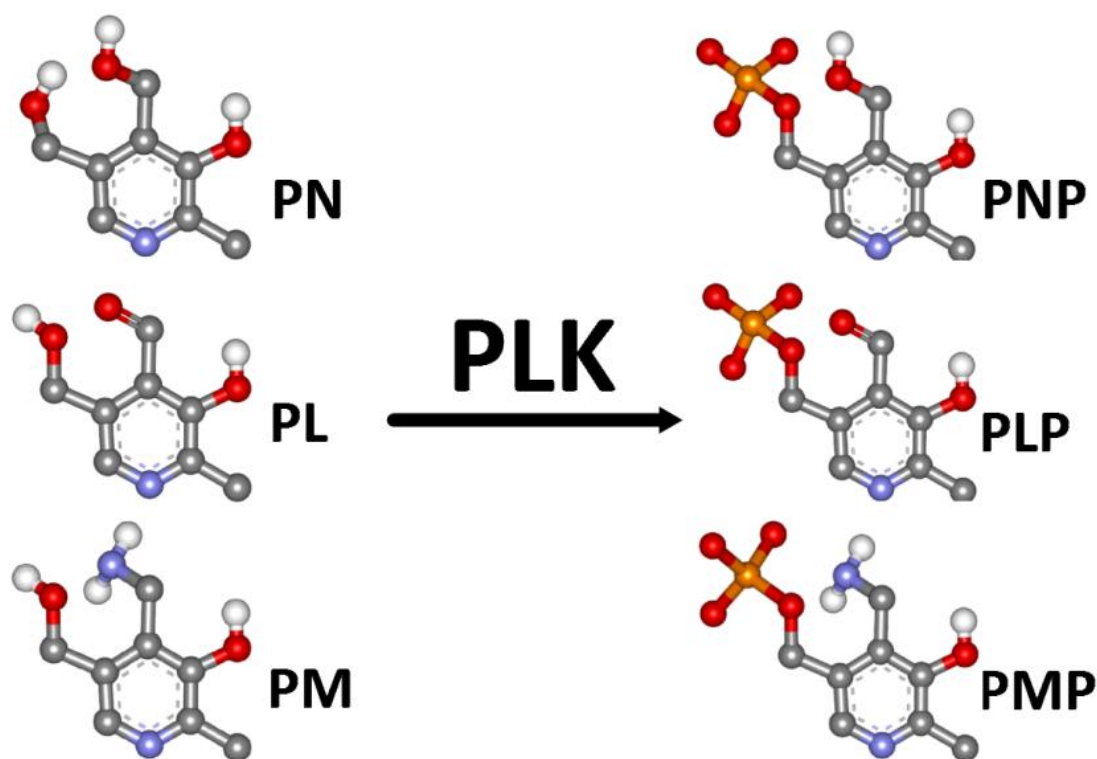


Figure 1.1. *Pyridoxal Kinase catalyzed reactions:* PLK phosphorylates the 5' alcohol group of PN, PL and PM to form PNP, PLP and PMP respectively.

1.2.3 Pyridoxal 5'-phosphate (PLP) metabolic pathways:

Until now the exact mechanisms of the control of PLP concentration in the cells is not yet fully established. However, what is certain is that to achieve the important task of neutralizing excess PLP, the cell utilizes more than one metabolic pathway. Zhao and Winkler observed feedback inhibition of pyridoxine 5'-phosphate oxidase, a key enzyme in vitamin B₆ biosynthesis, by its product PLP with a K_i of 8 μM [43,44]. Safo's group also observed that PLP is involved in feedback inhibition of PLK through the formation of a ternary complex with MgATP thus shutting down a major pathway for synthesis of metabolically active PLP [39,44]. These feedback inhibition mechanisms ensure that PLP synthesis does not exceed the cell needs. However to metabolize any unused PLP that may exist, the cell employs a special phosphatase to dephosphorylate PLP to its harmless form PL [45]. Compared to 30 μM of PLK, PLP phosphatase has a K_m of 2.5 μM which explains its importance to the control of PLP levels within the cells [37,45]. Furthermore, PLP phosphatase was found to be in various species with a wide-spread distribution throughout different tissues in mammals especially the brain [45]. This could be attributed to the fact that nerve cells are the most susceptible to PLP toxicity. This tight regulation of PLP, plus the fact that free PLP level is maintained at a very low concentration in the body (1 μM in eukaryotic cells) raises a very important question of how the PLP-dependent enzymes manage to get sufficient PLP to maintain their activities [44].

1.2.4 Mechanism of transfer of PLP from PLK to SHMT and other PLP-dependent enzymes:

The traditional model of how PLP is transferred to PLP-dependent enzymes is through the release of PLP from PLK or PNPOx into the bulk solvent, which is then acquired by PLP-dependent enzymes [44]. However, a shortcoming of this model is that it does not explain why

free PLP is always found to be scarce in vivo. Moreover, release of free PLP into the bulk solvent would render it available for destruction by PLP phosphatase. A second alternative model is substrate channeling between PLP synthesizing enzymes and PLP dependent enzymes [44]. In a study done by Hutchmacker *et al.*, it was found that enzymes catalyzing reactions that share one or more metabolites are more likely to interact allowing the transfer of substrates from one active site to the next without releasing it into the bulk solvent. This micro-compartmentalization of substrates results in more efficient metabolism by decreasing the transit time and increasing local substrate concentration with respect to the whole cell [46]. As an example for channeling, Michael F. Dunn and co-workers were among the first to observe substrate channeling. They described the channeling of indole between the alpha- and beta-subunits of tryptophan synthase in the last two steps of L-tryptophan synthesis (Figure 1.2) [47]. Others also described substrate channeling (Moriguchi *et al.* and Hakobyan *et al.*) [48,49]. Y.-H. Percival Zhang recently described methods for the biotechnological utilization of substrate channeling in areas such as multi-functional fusion proteins, metabolic engineering, synthetic cellulosomes and recombinant cellulolytic microorganisms and co-immobilization of multiple enzymes [50].

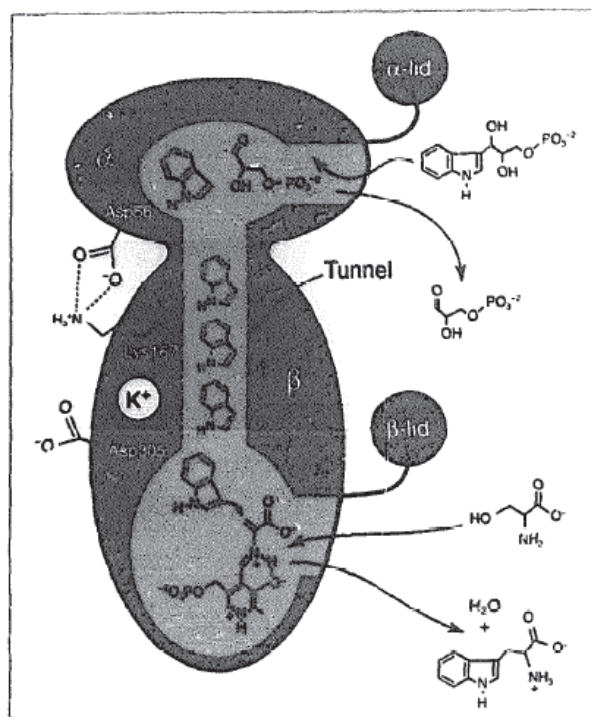


Figure 1.2. Tryptophan synthase channel: Indole channeling between alpha- and beta- subunits of tryptophan synthase [47].

The Substrate Channeling model might explain how PLP is transported securely and without causing damage to the cell. Although it is challenging to prove this theory, there are studies suggesting that PLP channeling between these enzymes might be the case here. In a study by Schirch's group using *E. coli* cell extract, it was shown that activation of the apo PLP-dependent enzyme, serine hydroxymethyltransferase (apo SHMT) into the holo form (holo SHMT-PLP) was more efficient using a PNPOx-PLP complex compared to the use of free PLP. They hypothesized that free PLP in the extracts was forming non-specific aldimines with other cell components like proteins [51]. In another study, Churchich's group used fluorescence spectroscopy, affinity chromatography and a trapping agent (alkaline phosphatases) to show that PLK forms a complex with aspartate aminotransferase (AAT), another PLP-dependent enzyme, with a K_d of 3 μ M, and that the trapping agent did not inhibit the transfer of PLP to the PLP-

dependent enzyme [52]. This was also confirmed by a recent study by Cheung *et al.*, who used fluorescence polarization and surface plasmon resonance biosensor analyses and showed that PLK can bind to AAT and glutamate decarboxylase with affinity constants in the low μM range [53]. All these studies suggest that PLP is likely transported by channeling.

In several unpublished studies by Safo's group, the activation of apo-SHMT or apo-AAT with free PLP or an equivalent amount of tightly bound PLP on PNPOx and PLK, were carefully monitored in the presence and absence of PLP-phosphatase. It was observed that activation of apo-B₆ enzymes is severely compromised when free PLP was used, while the phosphatases had no effect on the transfer if the activation is conducted using PLK or PNPOx with tightly bound PLP. The group also used fluorescence polarization techniques to study the binding interactions between *E. coli* PLK or PNPOx and several B₆ enzymes, including *E. coli* SHMT, AAT and l-threonine aldolase, as well as glycogen phosphorylase. The results showed that both PLK and PNPOx form specific interactions with every B₆ enzyme tested, with K_d ranging from 0.3 to 56 μM . The strongest affinity was between AAT and both PLK and PNPOx ($\sim 0.3 \mu\text{M}$), while glycogen phosphorylase showed the weakest interaction, 56 μM with PLK and 25 μM with PNPOx. It is worth noting that the most abundant B₆ enzyme in the body is glycogen phosphorylase, which compensate for its weak interactions with PLK and PNPOx, this also ensures that it does not outcompete other B₆ enzymes for PLP [44]. To confirm the specificity of these interactions, control experiments with several non-B₆ enzymes, including PLP-phosphatase, lysozyme, lactate dehydrogenase, and bovine serum albumin were used and did not show any specific binding with either B₆ salvage enzyme. In addition, affinity pull down chromatography experiments were performed and further confirmed the specificity of the interactions between the B₆ salvage enzymes and PLP-dependent enzymes.

1.3 HINT (Hydropathic INTeractions):

The cornerstone of analysis presented here in this thesis is HINT (Hydropathic INTeractions), a force field that describes and quantifies all interactions in the biological environment through the exploitation of the interaction information implicit in Log $P_{o/w}$. It is considered a "natural" force field because it is totally based on empirical energetic terms, which are defined by real experiments. Encoded within it, are all the types of interactions including coulombic, hydrogen bond and hydrophobic interactions, which are expected to be found between molecules in the biological environment. Thus, it also encodes a free energy force field and by including entropy and solvation/desolvation besides the other enthalpic terms [8-14]. The HINT score (H_{TOTAL}) is a double sum over all atom-atom pairs of the product (b_{ij}) of the hydrophobic atom constants (a_i , partial log $P_{octanol/water}$) and atomic solvent accessible surface areas (S_i) for the interacting atoms, mediated by a function of the distance between the atoms:

$$H_{TOTAL} = \sum_i \sum_j b_{ij} = \sum_i \sum_j (a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}) \quad (1.1)$$

where R_{ij} is a simple exponential function, e^{-r} [13], r_{ij} is an adaptation of the Lennard-Jones function [54,55], and T_{ij} is a logic function assuming +1 or -1 values, depending on the polar (Lewis acid or base) nature of interacting atoms.

1.4 Rational and specific aims:

1.4.1 Analysis of water molecules at protein-protein interfaces:

Studies discussed above provide compelling evidence that understanding protein-water-protein interactions is an important part of understanding protein-protein complexes and their biological roles. It is not simply the case that water molecules can bridge two proteins: such

contacts could be encoding for significant information that allows the interactions to be discriminating. Harnessing such information will be valuable in order to deepen our understanding of the rules governing the formation as well as the dissociation of macromolecules. This knowledge may be significantly important in the field of designing small molecule inhibitors for protein-protein complexes. Protein-protein complexes are under intense scrutiny as possible targets for new therapies, particularly in cancer and amyloidogenic diseases [56-60]. It has proven difficult to design molecules that can inhibit specific protein-protein associations due to the relative paucity of structural data on relevant complexes, although the number of such structures is growing [61].

Another area where this knowledge will be of great importance is in the development of computational approaches for building reliable models of protein-protein complexes, which is currently hindered due to the lack of knowledge. In the absence of specific knowledge, there are nearly an infinite number of ways to dock two irregularly shaped objects with a relatively small surface contact area. This contrasts to the better-defined and easier problem of small molecule docking in pockets of proteins. Even there, however, no universal scoring function has emerged that can confidently predict either the docked conformation or the free energy of binding [62-64]. Despite these major issues, computational algorithms and protocols are being developed for macromolecular docking [65-69].

Looking at protein-water-protein interaction, the highest-level view is that there are three distinct roles for waters at these interfaces: bridging, i.e., having significant interactions with both proteins; non-bridging, i.e., having significant interactions with only one of the two proteins; or simply trapped without significant interactions with either protein. More detailed analyses may reveal additional details such as whether these classifications are dependent on the

resolution of the underlying X-ray crystallographic experiment, e.g., are trapped waters more or less likely to be detected at high-resolution? Do different residue types have differences in interaction preferences for waters in these three categories, e.g., what residue types are most often involved in interactions with bridging waters? Water is unique in its ability to simultaneously provide two hydrogen-bond acceptor sites and two donor sites. Thus, it can effectively bridge in every way possible: donor-to-donor, donor-to-acceptor and acceptor-to-acceptor.

1.4.2 Developing a model for PLK-SHMT interaction:

PLP has a reactive aldehyde function which interacts with almost all nucleophiles, including proteins other than vitamin B₆ enzymes, which may cause neurological as well as non-neurological toxicities. The *in vivo* concentration of free PLP is thus maintained very low (~1 μM) by hydrolysis of free PLP back into PL by phosphatases and conversion of free PLP into 4-pyridoxic acid, and possibly by mechanisms such as feedback inhibition of PLK and PNPO by PLP. It is, therefore, very interesting to determine how, in spite of the low level of *in vivo* free PLP, as well as the activity of phosphatases, sufficient PLP gets transferred from PNPO and PLK to the vitamin B₆ dependent enzymes. The hypothesis here is that PLK or PNPOx specifically bind with PLP-dependent enzyme and channel the PLP from the former enzymes to the latter enzymes. This chapter focuses on developing a model for the protein-protein interaction and channel formation between PLK and SHMT as well as the identification of water molecules and their relevance to this protein-protein complex. This model might answer the question of whether these interactions are specific to each enzyme or whether they share a common binding site. In addition, this model might be useful for guiding site directed mutagenesis studies to further

confirm channeling; furthermore, this model might be useful in developing inhibitors for that protein-protein interaction providing a new drug target for cancer chemotherapy.

This study is divided into two specific aims:

1. A detailed analysis of water molecules at protein-protein interfaces as well as quantifying their contributions with respect to different residue types.
2. Analyze the effect of adding interfacial water molecules in developing a model for the protein-protein complex of PLK-SHMT

REFERENCES

REFERENCES

1. Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108*, 74-108.
2. Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. Water in protein structure prediction. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 3352-3357.
3. Zong, C.; Papoian, G. A.; Ulander, J.; Wolynes, P. G. Role of topology, nonadditivity, and water-mediated interactions in predicting the structures of alpha/beta proteins. *J. Am. Chem. Soc.* **2006**, *128*, 5168-5176.
4. Ohno, K.; Kamiya, N.; Asakawa, N.; Inoue, Y.; Sakurai, M. Effects of Hydration on the Electronic Structure of an Enzyme: Implications for the Catalytic Function. *J. Am. Chem. Soc.* **2001**, *123*, 8161-8162.
5. Okada, T.; Fujiyoshi, Y.; Silow, M.; Navarro, J.; Landau, E. M.; Shichida, Y. Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5982-5987.
6. Krauss, M.; Gilson, H. S. R.; Gresh, N. Structure of the First-Shell Active Site in Metallolactamase: Effect of Water Ligands. *J. Phys. Chem. B* **2001**, *105*, 8040-8049.
7. van Amsterdam, I. M.; Ubbink, M.; Einsle, O.; Messerschmidt, A.; Merli, A.; Cavazzini, D.; Rossi, G. L.; Canters, G. W. Dramatic modulation of electron transfer in protein complexes by crosslinking. *Nat. Struct. Biol.* **2002**, *9*, 48-52.
8. Fornabaio, M.; Spyraakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. Simple, Intuitive Calculations of Free Energy of Binding for Protein-Ligand Complexes. 3. The Free Energy Contribution of Structural Water Molecules in HIV-1 Protease Complexes. *J. Med. Chem.* **2004**, *47*, 4507-4516.

9. Spyrakis, F.; Cozzini, P.; Bertoli, C.; Marabotti, A.; Kellogg, G. E.; Mozzarelli, A. Energetics of the protein-DNA-water interaction. *BMC Struct. Biol.* **2007**, 7, 4.
10. Marabotti, A.; Spyrakis, F.; Facchiano, A.; Cozzini, P.; Alberti, S.; Kellogg, G. E.; Mozzarelli, A. Energy-based prediction of amino acid-nucleotide base recognition. *Journal of Computational Chemistry* **2008**, 29, 1955-1969.
11. Burnett, J. C.; Kellogg, G. E.; Abraham, D. J. Computational Methodology for Estimating Changes in Free Energies of Biomolecular Association upon Mutation. The Importance of Bound Water in Dimer-Tetramer Assembly for beta 37 Mutant Hemoglobins. *Biochemistry (N. Y.)* **2000**, 39, 1622-1633.
12. Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the Energetics of Water-Protein and Water-Ligand Interactions with the “Natural” HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules. *J. Mol. Biol.* **2006**, 358, 289-309.
13. Eugene Kellogg, G.; Abraham, D. J. Hydrophobicity: is LogP(o/w) more than the sum of its parts? *Eur. J. Med. Chem.* **2000**, 35, 651-661.
14. Sarkar, A.; Kellogg, G. E. Hydrophobicity--shake flasks, protein folding and drug discovery. *Curr. Top. Med. Chem.* **2010**, 10, 67-83.
15. Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **2001**, 29, 2860-2874.
16. Reddy, C. K.; Das, A.; Jayaram, B. Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.* **2001**, 314, 619-632.

17. Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry* University Science: 2005; , pp 1104.
18. Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Physical Chemistry Chemical Physics* **2007**, *9*, 573-581.
19. Larsen, T. A.; Olson, A. J.; Goodsell, D. S. Morphology of protein–protein interfaces. *Structure* **1998**, *6*, 421-427. Hubbard, S. J.; Argos, P. Cavities and packing at protein interfaces. *Protein Sci.* 1994, *3*, 2194-2206.
20. Sonavane, S.; Chakrabarti, P. Cavities and atomic packing in protein structures and interfaces. *PLoS Comput. Biol.* **2008**, *4*, e1000188.
21. Papoian, G. A.; Uler, J.; Wolynes, P. G. Role of Water Mediated Interactions in Protein-Protein Recognition Landscapes. *J. Am. Chem. Soc.* **2003**, *125*, 9170-9178.
22. Keskin, O.; Ma, B.; Nussinov, R. Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *J. Mol. Biol.* **2005**, *345*, 1281-1294.
23. Keskin, O.; Nussinov, R. Similar Binding Sites and Different Partners: Implications to Shared Proteins in Cellular Pathways. *Structure* **2007**, *15*, 341-354.
24. Jiang, L.; Kuhlman, B.; Kortemme, T.; Baker, D. A “solvated rotamer” approach to modeling water-mediated hydrogen bonds at protein–protein interfaces. *Proteins: Structure, Function, and Bioinformatics* **2005**, *58*, 893-904.
25. Joachimiak, L. A.; Kortemme, T.; Stoddard, B. L.; Baker, D. Computational Design of a New Hydrogen Bond Network and at Least a 300-fold Specificity Switch at a Protein-Protein Interface. *J. Mol. Biol.* **2006**, *361*, 195-208.

26. Dey, S.; Pal, A.; Chakrabarti, P.; Janin, J. The Subunit Interfaces of Weakly Associated Homodimeric Proteins. *J. Mol. Biol.* **2010**, *398*, 146-160.
27. Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of protein–protein interfaces. *Proteins: Structure, Function, and Genetics* **2005**, *60*, 36-45.
28. Samsonov, S.; Teyra, J.; Pisabarro, M. T. A molecular dynamics approach to study the importance of solvent in protein interactions. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73*, 515-525.
29. Teyra, J.; Pisabarro, M. T. Characterization of interfacial solvent in protein complexes and contribution of *wet spots* to the interface description. *Proteins: Structure, Function, and Bioinformatics* **2007**, *67*, 1087-1095.
30. 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*, 129-134.
31. 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*, 9374-9378.
32. Lambrecht, G.; Braun, K.; Damer, M.; Ganso, M.; Hildebrandt, C.; Ullmann, H.; Kassack, M. U.; Nickel, P. Structure-activity relationships of suramin and pyridoxal-5'-phosphate derivatives as P2 receptor antagonists. *Curr. Pharm. Des.* **2002**, *8*, 2371-2399.
33. K. Dakshinamurti, K.; Lal, P.; Ganguly, Hypertension, calcium channel and pyridoxine (vitamin B₆). *Molecular and cellular biochemistry* **1998**, *188*, 137-148.
34. Salhany, J. M.; Rauenbuehler, P. B.; Sloan, R. L. Characterization of pyridoxal 5'-phosphate affinity labeling of band 3 protein. Evidence for allosterically interacting transport inhibitory subdomains. *J. Biol. Chem.* **1987**, *262*, 15965-15973.
35. Oka, T. Modulation of gene expression by vitamin B₆. *Nutrition Research Reviews* **2001**, *14*, 257-265.

36. Huq, M. D.; Tsai, N. P.; Lin, Y. P.; Higgins, L.; Wei, L. N. Vitamin B₆ conjugation to nuclear corepressor RIP140 and its role in gene regulation *Nat. Chem. Biol.* **2007**, *3*, 161-165.
37. di Salvo, M. L.; Hunt, S.; Schirch, V. Expression, purification, and kinetic constants for human and *Escherichia coli* pyridoxal kinases. *Protein Expr. Purif.* **2004**, *36*, 300-306.
38. Safo, M. K.; Musayev, F. N.; di Salvo, M. L.; Hunt, S.; Claude, J. B.; Schirch, V. Crystal structure of pyridoxal kinase from the *Escherichia coli* pdxK gene: implications for the classification of pyridoxal kinases *J. Bacteriol.* **2006**, *188*, 4542-4552.
39. Musayev, F. N.; diSalvo, M. L.; Ko, T.; Gandhi, A. K.; Goswami, A.; Schirch, V.; Safo, M. K. Crystal structure of human pyridoxal kinase: Structural basis of M⁺ and M²⁺ activation. *Protein Sci.*, 2007, *16*, 2184-2194
40. McCormick, D. B.; Gregory, M. E.; Snell, E. E. Pyridoxal phosphokinases. I. Assay, distribution, I. Assay, distribution, purification, and properties *J. Biol. Chem.* **1961**, *236*, 2076-2084.
41. Yang, Y.; Tsui, H. C.; Man, T. K.; Winkler, M. E. Identification and function of the pdxY gene, which encodes a novel pyridoxal kinase involved in the salvage pathway of pyridoxal 5'-phosphate biosynthesis in *Escherichia coli* K-12 *J. Bacteriol.* **1998**, *180*, 1814-1821.
42. Safo, M. K.; Mathews, I.; Musayev, F. N.; di 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 Fold. Des.* **2000**, *8*, 751-762.
43. 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*, 883-891.
44. di Salvo, M.; Contestabile, R.; Safo M. K. Vitamin B₆ salvage enzymes: Mechanism, structure and regulation. *Biochim Biophys Acta.* **2010**, Epub ahead of print.
45. Jang, Y. M.; Kim, D. W.; Kang, T. C.; Won, M. H.; Baek, N. I.; Moon, B. J.; Choi, S. Y.; Kwon, O. S. Human pyridoxal phosphatase. Molecular cloning, functional expression, and tissue distribution *J. Biol. Chem.* **2003**, *278*, 50040-50046.

46. Huthmacher, C.; Gille, C.; Holzhutter, H. G. A computational analysis of protein interactions in metabolic networks reveals novel enzyme pairs potentially involved in metabolic channeling. *J. Theor. Biol.* **2008**, *252*, 456-464.
47. Pan, P.; Woehl, E.; Dunn, M. F. Protein architecture, dynamics and allostery in tryptophan synthase channeling. *Trends Biochem. Sci.* **1997**, *22*, 22-27.
48. Moriguchi, T.; Ida, K.; Hikima, T.; Ueno, G.; Yamamoto, M.; Suzuki, H. Channeling and conformational changes in the heterotetrameric sarcosine oxidase from *Corynebacterium* sp. U-96. *The Journal of Biochemistry* **2010**, *148*, 491-505.
49. Hakobyan, D.; Nazaryan, K. Molecular dynamics study of interaction and substrate channeling between neuron-specific enolase and B-type phosphoglycerate mutase. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1691-1704.
50. Zhang, Y. H. Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol. Adv.* **2011**, .
51. Yang, E. S.; Schirch, V. Tight Binding of Pyridoxal 5'-Phosphate to Recombinant *Escherichia coli* Pyridoxine 5'-Phosphate Oxidase. *Arch. Biochem. Biophys.* **2000**, *377*, 109-114.
52. Kim, Y. T.; Kwok, F.; Churchich, J. E. Interactions of pyridoxal kinase and aspartate aminotransferase emission anisotropy and compartmentation studies *J. Biol. Chem.* **1988**, *263*, 13712-13717.
53. Cheung, P.; Fong, C.; Ng, K.; Lam, W.; Leung, Y.; Tsang, C.; Yang, M.; Wong, M. Interaction between Pyridoxal Kinase and Pyridoxal-5-phosphate-Dependent Enzymes. *The Journal of Biochemistry* **2003**, *134*, 731-738.
54. Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **1983**, *168*, 595-617.
55. Levitt, M.; Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **1988**, *201*, 751-754.

56. Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **2004**, *3*, 301-317.
57. L. Garner, A.; D. Janda, K. Protein-Protein Interactions and Cancer: Targeting the Central Dogma. *Current Topics in Medicinal Chemistry* **2011**, *11*, 258-280.
58. Matallanas, D.; Crespo, P. New druggable targets in the Ras pathway? *Curr. Opin. Mol. Ther.* **2010**, *12*, 674-683.
59. Castillo, V.; Ventura, S. Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases. *PLoS Comput. Biol.* **2009**, *5*, e1000476.
60. Wolfe, K. J.; Cyr, D. M. Amyloid in neurodegenerative diseases: Friend or foe? *Semin. Cell Dev. Biol.* **2011**, .
61. Dutta, S.; Berman, H. M. Large Macromolecular Complexes in the Protein Data Bank: A Status Report. *Structure* **2005**, *13*, 381-388.
62. Jain, A. N. Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.* **2006**, *7*, 407-420.
63. Englebienne, P.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 4. Are Popular Scoring Functions Accurate for this Class of Proteins? *Journal of Chemical Information and Modeling* **2009**, *49*, 1568-1580.
64. Spyraakis F.; Cozzini P.; Kellogg G. E. Docking and scoring in drug discovery. In: *Burger's Medicinal Chemistry and Drug Discovery*. Abraham, D. J.; Rotella, D., Eds; John Wiley & Sons: Hoboken, New Jersey, 2010; pp. 601-684.
65. Gray JJ, Moughon S, Wang C, Schueler-Furman O, et al. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 331: 281–299.
66. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2195-2199.
67. Totrov, M.; Abagyan, R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy *Nat. Struct. Biol.* **1994**, *1*, 259-263.

68. Ritchie, D. W.; Kozakov, D.; Vajda, S. Accelerating and focusing protein–protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics* **2008**, *24*, 1865-1873.
69. Chen, R.; Li, L.; Weng, Z. ZDOCK: An initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Genetics* **2003**, *52*, 80-87.

CHAPTER 2

WATER MOLECULES AT PROTEIN-PROTEIN INTERFACES: A DETAILED ANALYSIS AND QUANTIFICATION OF THEIR CONTRIBUTIONS WITH RESPECT TO DIFFERENT RESIDUE TYPES.

2.1 Introduction:

Over the last decade, there has been a growing interest in understanding and exploiting protein-protein interactions as potential new routes to disease therapeutics [1-8]. It is believed that if one or more critical, but often transient, protein-protein interactions could be inhibited by a peptidic or small molecule agent, this could lead to a novel and specific approach for treatment of a wide variety of human diseases. The understanding of numerous cell cycle pathways that we have developed has been nothing short of revolutionary, and these pathways repeatedly invoke protein-protein interactions, but to date few therapeutics have resulted from this knowledge [2]. One reason is that our structural knowledge of protein-protein complexes is lagging, largely because experimental X-ray crystallographic structure determinations of these complexes are demanding [9,10], principally due to the difficulties in co-crystallizing the involved proteins [11], which *in vivo* are often transiently associated and disordered [12], in diffraction-quality crystals.

Nonetheless, the RCSB Protein Data Bank [13] contains several hundred protein-protein complexes [9], although the collection is somewhat biased towards a few classes, e.g., antigen-antibody complexes. Whether the interactions at these interfaces differ from the interactions between ligands and proteins, between polynucleotides and proteins, or within a protein is a

widely explored issue. To this aim, an in-depth assessment of the role of water molecules located at protein-protein interfaces is particularly relevant.

This chapter describes a detailed analysis of protein-protein interfaces in 179 high-resolution (better than 2.30 Å) X-ray crystal structures of protein-protein complexes extracted from the RCSB Protein Data Bank [13]. All water molecules within 4.0 Å of both proteins, 4741 unique waters, comprised the data set.

2.2 Materials and Methods:

2.2.1 Data set:

The protein-protein complexes data set was obtained from the RCSB Protein Data Bank [13] by applying search filters for several structural criteria. First, the structures were required to have at least two separate protein entities where each was at least 100 amino acids in length. Structures with either DNA or RNA were excluded as were structures with sequence identity similarity > 50% to another protein complex in the data set. The data set was restricted to structures with resolutions 2.3 Å and better. This set (1331) of PDB structures consisted of both homo and hetero protein complexes. Further screening of the structures' description isolated protein-protein complexes (861) for individual inspection where only structures comprised of completely different proteins, i.e., not subunits or chains of the same protein, were retained. Finally, 179 structures (Table 2.1) were randomly selected from this set for analysis.

Table 2.1. Protein complexes examined in study with interface parameters and water roles.

PDB ID	Res. (Å)	All H ₂ O	Interface	Interface H ₂ O	Relevant to:		
					0	1	2
1a4y ^a	2.00	133	A / B	16	3	3 / 3	7
1ava ^a	1.90	748	A / C	38	8	9 / 9	12
1avw	1.75	142	A / B	11	1	2 / 6	2
1blx	1.90	294	A / B	33	8	7 / 10	8
1d2z ^a	2.00	266	AC / B	30	9	5 / 9	7
1eer	1.90	298	A / BC	33	8	11 / 9	5
1ev2 ^a	2.20	263	C / FGH	14	2	2 / 4	6
1f3v	2.00	208	A / B	12	3	4 / 3	2
1fns	2.00	636	A / LH	21	5	1 / 5	10
1fyh ^a	2.04	481	AD / B	24	6	8 / 4	6
1g4y	1.60	199	B / R	19	4	4 / 5	6
1ghq	2.04	666	A / BC	20	4	7 / 2	7
1gpq ^a	1.60	635	B / CD	38	9	12 / 6	11
1he1 ^a	2.00	712	A / C	28	10	2 / 15	1
1hx1	1.90	359	A / B	28	6	6 / 10	6
1i2m ^a	1.76	475	A / B	28	2	5 / 11	10
1i7w ^a	2.00	669	A / B	24	7	8 / 2	7
1iqd	2.00	477	AB / C	25	8	5 / 9	3
1jiw	1.74	592	I / P	29	7	4 / 14	4
1jyo ^a	1.90	629	AC / F	37	10	15 / 5	7
1ksh	1.80	124	A / B	13	2	4 / 3	4
1ktz	2.15	163	A / B	17	1	3 / 5	8
1kxp	2.10	388	A / D	42	11	13 / 9	9
1kxq ^a	1.60	2807	BD / E	45	4	11 / 14	16
1lk3 ^a	1.91	1203	LHM / B	23	3	6 / 6	8
1nf3 ^a	2.10	423	A / C	16	2	6 / 6	2
1nmb	2.20	83	LH / N	8	2	2 / 1	3

1o94 ^a	2.00	2149	A / CD	20	2	9 / 4	5
1okk	2.05	568	A / B	43	7	16 / 14	6
1ors	1.90	403	A / C	6	1	3 / 0	2
1osp	1.95	328	HL / O	23	2	8 / 9	4
1ow3	1.80	374	A / B	39	9	7 / 9	14
1oy3	2.05	248	BC / D	24	8	5 / 4	7
1pxv ^a	1.80	458	A / C	27	7	8 / 5	7
1q40 ^a	1.95	331	A / BD	29	7	10 / 7	5
1r8s	1.46	350	A / E	31	11	9 / 5	6
1rew ^a	1.86	185	ABD / C	16	1	4 / 7	4
1slu	1.80	137	A / B	5	4	0 / 1	0
1sq2	1.45	189	L / N	20	1	11 / 1	7
1t6g ^a	1.80	1074	AB / C	55	11	15 / 18	11
1ta3	1.70	785	A / B	52	12	7 / 12	21
1tue ^a	2.10	969	M / LQ	16	4	4 / 3	5
1tx4	1.65	497	A / B	44	12	7 / 17	8
1tx6 ^a	2.20	492	ABC / I	20	5	3 / 6	6
1unn ^a	1.90	947	AB / C	41	11	15 / 7	8
1usu	2.15	227	A / B	13	2	2 / 6	3
1v7p	1.90	463	AB / C	37	8	10 / 8	11
1vg0	2.20	438	A / B	32	8	11 / 7	6
1wa5	2.00	472	AC / B	28	8	10 / 7	3
1www ^a	2.20	262	VW / X	21	7	4 / 6	4
1wxc ^b	1.20	393	A / B	25	7	2 / 8	8
1xg2	1.90	453	A / B	41	13	3 / 14	11
1xkp	1.70	179	A / BC	26	9	9 / 1	7
1xx9 ^a	2.20	260	AB / C	15	2	6 / 4	3
1yar ^a	1.90	3470	DEF / O	19	6	5 / 5	3
1yca	2.20	275	A / B	7	0	2 / 2	3
1yro	1.90	830	BD / C	46	10	10 / 13	13
1yu6	1.55	329	B / D	17	6	4 / 3	4
1z5y	1.94	254	D / E	17	5	2 / 8	2
1zc3 ^a	2.00	413	A / D	23	9	7 / 4	3
1ze3	1.84	510	CH / D	27	5	4 / 12	6
1zhh	1.94	333	A / B	32	4	9 / 9	10
2a2q	1.80	722	HL / T	57	10	14 / 15	18
2a9k	1.73	210	A / B	21	4	8 / 7	2

2aq2	1.80	231	A / B	21	4	5 / 6	6
2arp	2.00	171	A / F	19	5	5 / 6	3
2b2x ^a	2.20	226	A / HL	8	0	3 / 1	4
2bcg	1.48	923	G / Y	48	12	14 / 13	9
2bex ^a	1.99	616	AB / C	36	4	14 / 4	14
2bkk ^a	2.15	347	AC / B	16	7	3 / 3	3
2bo9 ^a	1.60	1153	AC / D	56	14	13 / 17	12
2cio	1.50	161	A / B	4	1	1 / 1	1
2co7	1.80	274	A / B	29	10	7 / 5	7
2dfk ^a	2.15	581	AC / B	47	14	11 / 17	5
2e2d	2.00	323	A / C	34	8	14 / 9	3
2eke ^a	1.90	375	A / B	21	3	4 / 7	7
2es4 ^a	1.85	843	AB / D	55	18	14 / 13	10
2f2l	2.10	195	A / X	9	5	0 / 3	1
2f93	2.00	68	A / B	3	1	0 / 1	1
2f95	2.20	30	A / B	3	1	0 / 2	0
2fd6 ^a	1.90	336	ALH / U	13	4	6 / 0	3
2fdb ^a	2.28	125	MN / P	19	4	6 / 4	5
2fm8	2.20	612	AB / C	53	15	16 / 12	10
2fu5 ^a	2.00	301	A / C	12	2	2 / 3	5
2g2u	1.60	331	A / B	39	11	4 / 13	11
2gc7 ^a	1.90	1234	ABDE / C	17	7	4 / 3	3
2gh0 ^a	1.92	236	A / C	8	3	1 / 2	2
2goo ^a	2.20	327	A / BC	36	8	7 / 18	3
2hqs ^a	1.50	2812	DF / G	53	13	19 / 12	9
2iaa ^a	1.95	989	ABD / C	13	8	1 / 4	0
2j12	1.50	287	A / B	37	17	9 / 9	2
2j59 ^a	2.10	1656	ABCD / F	45	11	4 / 19	11
2jjs ^a	1.85	540	AB / D	39	8	10 / 18	3
2npt ^a	1.75	320	A / D	15	1	8 / 4	2
2nqd	1.75	564	A / B	36	11	7 / 14	4
2ns1	1.96	498	A / B	20	7	6 / 3	4
2nxy ^a	2.00	806	BCD / A	44	9	14 / 11	10
2nz8	2.00	242	A / B	42	6	19 / 7	10
2ode ^a	1.90	655	A / B	44	11	8 / 13	12
2omz	1.60	800	A / B	61	9	25 / 12	15
2ot3	2.10	463	A / B	25	9	5 / 6	5

2oul	2.20	171	A / B	18	9	6 / 1	2
2p45	1.10	319	A / B	21	5	8 / 3	5
2q0o ^a	2.00	484	AB / C	43	12	12 / 12	7
2q4g ^a	1.95	854	WY / X	40	7	16 / 6	11
2r25	1.70	238	A / B	21	3	7 / 9	2
2sic	1.80	258	E / I	17	4	5 / 6	2
2v9t	1.70	385	A / B	34	11	2 / 17	4
2vol	1.95	241	A / B	21	2	4 / 10	5
2vsm	1.80	705	A / B	54	8	18 / 19	9
2vxt	1.49	593	HL / I	33	5	11 / 4	13
2wel	1.90	405	A / D	14	2	4 / 7	1
2wwx	1.50	117	A / B	21	2	6 / 8	5
2wy3 ^a	1.80	639	A / B	32	9	18 / 4	1
2xg5	2.00	202	A / B	32	12	6 / 8	6
2xgy	1.80	367	A / B	30	15	2 / 10	3
2xna	2.10	249	AB / C	10	1	3 / 3	3
2xqy ^a	2.05	952	E / JK	27	6	3 / 11	7
2yvj	1.90	78	A / B	1	1	0 / 0	0
2z0d	1.90	314	A / B	46	12	17 / 10	7
2z3q ^a	1.85	228	ACD / B	24	5	12 / 5	2
2zd1	1.80	626	A / B	69	28	21 / 15	5
2zfd	1.20	236	A / B	24	6	7 / 5	6
3a4u	1.84	189	A / B	16	5	3 / 3	5
3a8k ^a	1.95	1495	AB / E	33	9	4 / 14	6
3a98 ^a	2.10	133	AC / D	18	2	2 / 6	8
3bh7	1.90	219	A / B	26	11	6 / 5	4
3bn3	2.10	226	A / B	22	6	4 / 9	3
3bn9 ^a	2.17	824	A / EF	32	5	9 / 10	8
3bwu	1.76	641	CD / F	46	17	6 / 17	6
3bx1 ^a	1.85	574	AB / C	28	8	5 / 11	4
3bx7	2.10	186	A / C	31	16	9 / 5	1
3cbj	1.80	243	A / B	27	11	7 / 5	4
3cip	1.60	461	A / G	28	9	13 / 2	4
3cx8	2.00	298	A / B	33	10	10 / 7	6
3d85 ^a	1.90	906	ABD / C	30	6	11 / 3	10
3d9a	1.20	683	LH / C	33	4	8 / 4	17
3ddc	1.80	115	A / B	18	7	7 / 1	3

3dlq	1.90	309	I / R	26	6	8 / 8	4
3egg ^a	1.85	521	AB / C	39	9	17 / 5	8
3egv	1.75	415	A / B	42	13	13 / 9	7
3evs	2.10	51	B / C	4	1	0 / 1	2
3f62	2.00	120	A / B	13	5	5 / 2	1
3f75	1.99	167	A / P	28	4	8 / 8	8
3ffd	2.00	204	AB / P	11	2	4 / 1	4
3fhi	2.00	159	A / B	18	4	11 / 3	0
3g5o ^a	2.00	173	AD / C	17	5	6 / 5	1
3gew ^a	2.00	299	AD / C	24	8	6 / 4	6
3gmw ^a	2.10	372	A / B	14	2	4 / 4	4
3grw	2.10	274	A / LH	29	10	8 / 3	8
3hct	2.10	195	A / B	12	3	4 / 3	2
3hei ^a	2.00	2756	CGIO / D	37	10	13 / 11	3
3hg0 ^a	2.10	446	ABC / D	15	3	3 / 6	3
3hh2 ^a	2.15	349	AB / C	23	6	5 / 7	5
3hy2 ^a	2.10	321	AB / X	30	14	5 / 8	3
3hzh	1.96	158	A / B	20	9	4 / 5	2
3jza	1.80	246	A / B	37	8	9 / 12	8
3k2m ^a	1.75	286	A / CD	23	4	6 / 5	8
3kdf ^a	1.98	260	BD / C	15	4	3 / 6	2
3kdj	1.88	170	A / B	9	3	1 / 3	2
3kf6	1.65	191	A / B	22	7	5 / 4	6
3kld	2.00	415	A / B	21	4	7 / 5	5
3kmu	1.80	298	A / B	15	3	4 / 5	3
3kyj	1.40	273	A / B	16	8	6 / 0	2
3l9j	2.10	246	C / T	16	0	2 / 5	9
3liz ^c	1.80	870	A / HL	43	13	11 / 14	5
3lxr	1.68	484	A / F	48	16	11 / 17	4
3m18	1.95	261	A / B	34	3	12 / 9	10
3m7f	2.00	141	A / B	11	5	3 / 1	2
3ma2 ^a	2.05	145	AD / B	12	4	6 / 0	2
3ma9	2.05	442	A / LH	23	5	4 / 10	4
3mc0 ^a	2.00	438	A / BD	22	7	8 / 3	4
3mdy ^a	2.05	687	AC / B	21	4	3 / 7	7
3n3a ^a	1.99	280	A / D	15	4	5 / 3	3
3nce	2.00	452	A / B	33	11	8 / 8	6

3og6	2.10	266	A / B	23	7	1 / 8	7
3oky	2.19	389	A / B	26	9	6 / 1	10
3orv ^a	1.91	1301	A / CDF	31	11	11 / 5	4
3q3j	1.97	38	A / B	3	0	1 / 1	1

Notes: ^aThese complexes have multiple protein-protein interfaces – only one (as indicated) was selected for this study;

^bThree waters (HOH254, HOH281 and HOH282) were deleted because of steric clashes; ^cOne water (HOH412) was deleted because of steric clashes

The downloaded coordinate files were prepared by first removing ligands or cofactors other than water. Then, using Sybyl 8.1 [14], hydrogen atoms were added and minimized (Tripos forcefield, with Gasteiger-Hückel charges and distance-dependent dielectric) to a gradient of $0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ while the non-hydrogen atoms were treated as an aggregate. Water molecules that were within 4.0 \AA from atoms on both of the interacting proteins were retained with each protein-protein complex. Together, the water data set is comprised of 4741 unique water molecules, which is 5.4% of all waters in these complexes (ranging from 0.5% to 17.9%).

2.2.2 Hydropathic Analysis:

Each model contains two proteins and an array of solvents, and was analyzed with HINT [15,16] by computing intermolecular scores between the proteins and the interfacial solvent arrays. HINT parameters and controls were as in previous studies [15,19-21]: partition calculations were performed with the “dictionary” method for the proteins with ‘essential hydrogens’, where polar hydrogens are treated explicitly and non-polar hydrogens are ‘united’ with their parent non-polar heavy atom; the HINT option that corrects the S_i terms for backbone amide nitrogens by adding 30 \AA^2 was used in this study to improve the relative energetics of inter- and intramolecular hydrogen bonds involving these nitrogens. Water molecules were partitioned as a “solvent set” with

analogous HINT parameters. Previous work [21,22] has suggested that approximately 500 HINT score units correspond to $-1.0 \text{ kcal mol}^{-1}$ of free energy.

Each crystallographically observed water molecule's orientation was optimized by an exhaustive protocol [23] that maximizes the HINT score with respect to its surrounding environment by evaluating its interactions with a “receptor” created from atoms within 6.0 \AA . For water molecules, this optimization rewards hydrogen bond and acid/base interactions while penalizing acid/acid and base/base interactions and those with hydrophobic entities on either of the two protein surfaces. Hydropathic interaction analysis was then performed with HINT for each of the optimized water molecules with respect to the two proteins with which it interacts. The resulting data were tabulated by frequency and strength of interactions with each amino acid residue type. In cases where a water molecule had significant interactions ($> |10|$ HINT score units, approximately $|0.02| \text{ kcal mol}^{-1}$) with more than one residue on a protein, that water's count was fractionally distributed to interacting residues based on the absolute values of the relative HINT scores for those residues that interact with it, i.e.,

$$W_i = \sum_n \{ |A_i^c| / \sum_i |A_i| \} \quad (2.1)$$

where A_i^c are the interaction HINT scores by residue type (i) interacting with water n. Similarly, the fractions of interactions with interfacial water molecules arising from backbone and sidechain atoms were calculated by weighted counts with A_i^c representing

the interaction HINT scores by i , separated into $c = \text{sidechain}$ or $c = \text{backbone}$ subsets. Heat maps for frequency and interaction scores and map clustering were calculated and drawn with gplots package within R [24].

2.2.3 Rank Algorithm:

Rank represents the weighted number of potential hydrogen bonds for each water molecule with respect to a pseudo-receptor of atoms from the target molecule(s) surrounding the water. Rank is calculated as:

$$\text{Rank} = \sum_n \{ (2.80 \text{ \AA}/r_n) + [\sum_m \cos (\theta_{Td} - \theta_{nm})]/6 \} \quad (2.2)$$

where r_n is the distance between the water's oxygen and the target's heavy atom n (n is the targets up to a maximum of 4). This is scaled relative to 2.8 Å, the presumed ideal hydrogen bond length. θ_{Td} is the optimum tetrahedral angle (109.5°) and θ_{nm} is the angle between targets n and m ($m = n$ to number of valid targets). The algorithm thus allows a maximum number of 4 targets (≤ 2 donors and ≤ 2 acceptors). To properly weight the geometrical quality of hydrogen bonds, targets that have an angle less than 60° with respect to other (higher quality) targets are rejected [23].

2.2.4 Relevance:

Relevance is a synthesis of HINT score and Rank [25]. Specifically,

$$\text{Relevance} = \{P_R(|W_R| + 1)^2 + P_H(|W_H| + 1)^2\} / \{(|W_R| + 1)^2 + (|W_H| + 1)^2\} \quad (2.3)$$

where P_R is the percent probability for water conservation based on Rank and P_H the probability based on HINT score. W_R and W_H are the weights for these probabilities, respectively. The values for P_R , P_H , W_R and W_H are as shown in Figure 2 of reference [25]. This relationship was derived with the expectation that water molecules with $\text{Relevance} \geq 0.5$ would be conserved and those with $\text{Relevance} < 0.5$ would be non-conserved because the waters analyzed in developing the training set were, by their nature, binary – either conserved and present in the ligand-bound complex or non-conserved and absent in the complex.

2.3 Results and discussion:

2.3.1 The Water Relevance Metric:

As described above, water Relevance [25] is a descriptor combining two metrics of structure: Rank [23] and HINT score [16], where each orientation-optimized water is scored against its environment. Others [26,27] suggested the crystallographic B-factor as a predictor of water conservation, but it was found to be not useful for this data set [25]. While Relevance was initially trained on and for protein-ligand complexes, *the role(s)*

that water molecules can play are independent of the stage: water will interact favorably with up to two hydrogen bond donors and up to two hydrogen bond acceptors, and will generally avoid interaction with hydrophobic functional groups, regardless of whether these groups are in small organic molecules or in proteins.

The Relevance algorithm was applied to the set of water molecules at protein-protein interfaces to understand their roles in these complexes. The water set for each complex was comprised of all water molecules that were within 4.0 Å of atoms in both proteins. This set, from 179 proteins, was comprised of 4741 unique water molecules, with between 1 and 69 waters (average 27) at the protein-protein interfaces. Rodier *et al.* [28] reported 20 per interface in their study of 46 protein-protein complexes. Figure 2.1 illustrates the set of 16 unique water molecules for the human placental RNase inhibitor (hRI)- human angiogenin (hAng) complex (PDB 1a4y, 2.00 Å) [29]. The training and derivation of the Relevance metric specified that $\text{Relevance} \geq 0.5$ corresponds to a water molecule that is conserved and largely static within a ligand binding pocket [25]. It is believed that this same Relevance score would also identify a water conserved at a protein-protein interface, and of the 4741 waters in this study, 37% (1741) have total $\text{Relevance} \geq 0.5$.

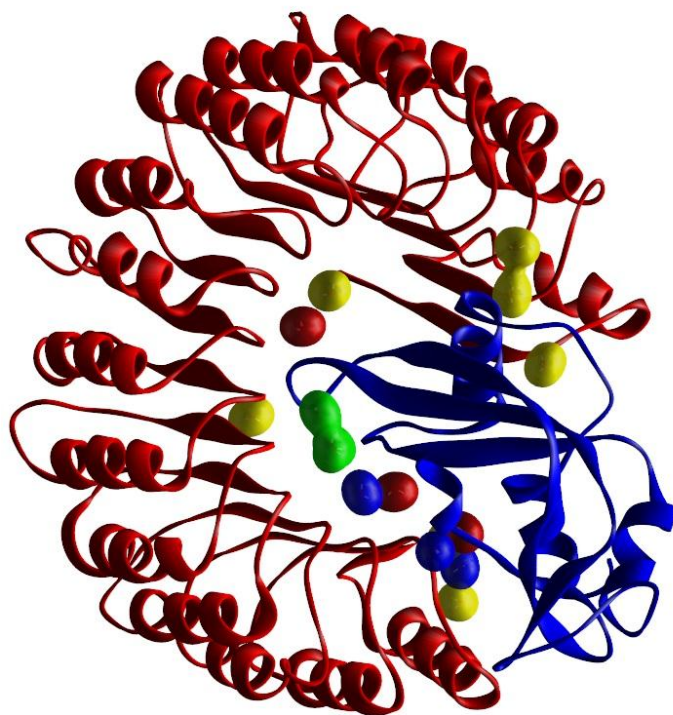


Figure 2.1. *Molecular model of human placental RNase inhibitor (hRI) (red)- human angiogenin (hAng) (blue) complex (1a4y):* Interface region; water molecules colored red are Relevant (≥ 0.25) with respect to hRI, blue with respect to hAng, yellow with respect to both hRI and hAng, and green with respect to neither (see Table 2.2). Of particular interest is the “hydrophobic bubble” enclosing the non-Relevant waters HOH59, HOH71 and HOH72 (green). Note that these three waters are encompassed within a region of the cavity that is of hydrophobic character.

More interesting are the evaluations of Relevance with respect to the partner proteins of the complexes. Applying this original definition of Relevance (≥ 0.5 for each

protein) identifies only 43 waters (< 1%) as bridging. Rodier *et al.* reported that 30% of waters at protein-protein interfaces are bridging, and while their definition of interaction is loose – the water must only be within 3.5 Å of a polar (N, O, S) protein atom to be counted as bridging [28] – here it is proposed that using an intermediate value of Relevance, such as 0.25, to flag association (or Relevance) with respect to a single protein, makes pragmatic sense. Thus, after exhaustive optimization of all waters' orientations (*vide supra*), the Rank, HINT score and Relevance for each were calculated with respect to each protein and in total. These data for 1a4y are listed in Table 2.2.

Table 2.2. Water metrics for human placental RNase inhibitor (hRI)- human angiogenin (hAng) complex (PDB 1a4y, 2.00 Å).

Water name	With hRI:			With hAng:			Total Rank	Total HINT score	Total Relevance	Relevance (≥ 0.25) w/ respect to:
	Rank	HINT score	Relevance	Rank	HINT score	Relevance				
HOH1	1.29	409	0.566	2.13	-96	0.205	3.41	313	0.778	hRI
HOH2	3.67	-64	0.481	1.18	70	0.333	4.85	6	0.640	Both
HOH19	3.51	-26	0.495	1.24	92	0.360	4.74	66	0.687	Both
HOH25	3.72	-25	0.529	1.31	68	0.347	5.03	44	0.682	Both
HOH52	2.34	358	0.687	1.09	-174	-0.137	3.43	184	0.727	hRI
HOH54	3.62	111	0.639	1.25	21	0.295	4.87	132	0.772	Both
HOH56	1.05	335	0.419	0.95	30	0.264	2.00	365	0.678	Both
HOH59	0.00	-35	-0.039	2.21	-236	-0.280	2.21	-271	-0.362	Neither
HOH60	3.78	316	0.822	1.46	-40	0.230	5.24	275	0.924	hRI
HOH61	2.30	271	0.627	2.60	141	0.563	4.90	412	0.948	Both
HOH68	0.98	80	0.305	1.03	24	0.273	2.01	105	0.441	Both
HOH70	1.05	-90	0.186	2.24	134	0.508	3.29	44	0.534	hAng
HOH71	0.72	-7	0.196	0.00	-255	-0.299	0.72	-262	-0.342	Neither
HOH72	0.89	-39	0.201	1.05	-321	-0.487	1.94	-360	-0.586	Neither
HOH73	0.91	22	0.251	1.12	62	0.315	2.03	84	0.418	Both
HOH74	1.32	-197	-0.191	2.38	105	0.490	3.70	-92	0.459	hAng

Only 21% (1018) of the interface waters have Relevance ≥ 0.25 with respect to *both* proteins, 53% (2514) have Relevance ≥ 0.25 with one member of the protein pair and 26% (1209) are not Relevant with respect to either (see Figure 2.2). This suggests that one-fifth of the waters found at a protein-protein interface are truly bridging, while one-fourth are merely trapped at the interface. More than half of the waters are strongly

associated with one protein, and while they provide steric constraints for the protein-protein association, they do not provide significant favorable energetic contributions to the association. This is an important distinction, as these waters still likely influence the association in more subtle ways (*vide infra*). While the choice of 0.25 as a threshold to determine the Relevance/non-Relevance of a water molecule with respect to a single protein is somewhat arbitrary, values smaller than 0.25 indicate a paucity of potential favorable interactions arguing against the water's conservation and values larger than 0.25 would suggest even fewer bridging waters than reported by Rodier *et al.* [28].

Here a data set comprised of protein X-ray crystal structures with resolutions better than 2.30 Å was used to construct a representative set of high-quality water molecules. The number of water molecules located and placed by crystallographers during refinement has been shown to be dependent on the resolution of the reflection data [30,31]. Thus whether, given the categories of waters defined here, there is a resolution-dependence in the relative ratios of water molecules Relevant to zero, one or two proteins was investigated. The hypothesis is that at poorer resolutions fewer non-Relevant water molecules would be located and placed in the electron density – presumably because they would be less ordered or conserved – and that the fraction of non-Relevant waters would decrease. However, the relative ratios of water molecules throughout different resolutions for this data set plus a second small data set of 16 poorer resolution complexes (2.4 – 3.5 Å) are relatively the same (Figure 2.2). Calculations performed for

waters in a second small data set of 16 poorer resolution complexes ($2.4 - 3.5 \text{ \AA}$), where 109 water molecules were located at the interfaces, revealed essentially the same fractions: 23 waters relevant to zero (21%), 62 waters relevant to one (57%) and 24 waters Relevant to two (22%). Crystallographic waters are seldom located in X-ray structures with resolutions poorer than 3.5 \AA , and water placements from structures with resolutions between 2.5 and 3.5 \AA may be considered somewhat unreliable. Assuming that all of these low-resolution waters are not crystallographic mistakes or artifacts [10], these data pose an interesting question: can water molecules without a stabilizing role at an interface be “conserved”?

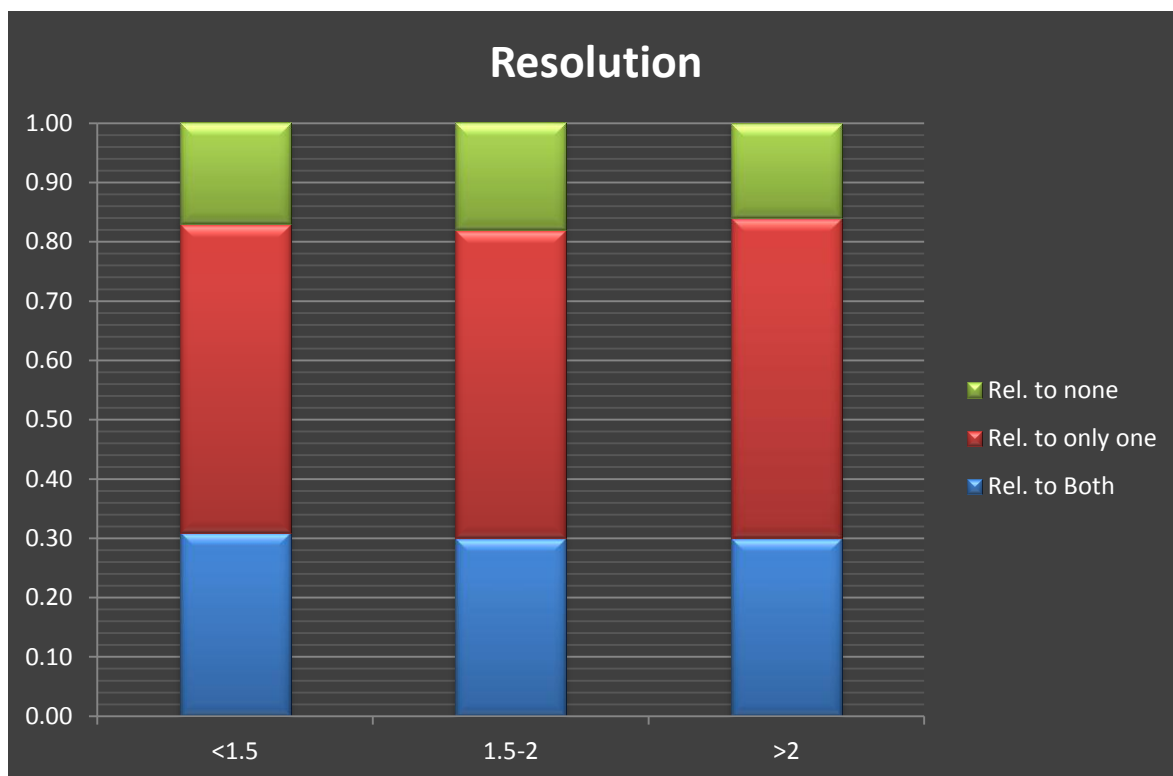


Figure 2.2. *Relative fractions of waters with Relevance to neither (green), one (red) and both (blue) proteins for:* full data set of 4741 waters from 179 protein X-ray structures of resolutions ≤ 2.3 Å; plus 109 waters from 16 structures with resolutions between 2.4 Å and 3.5 Å.

2.3.2 Residue Preferences for Interfacial H_2O :

Given the three general categories of interface waters we have described, the preferences these water molecules show for the types of amino acid residues within the interfaces were examined. First, for all interface waters, the preferences are tabulated by interaction counts (Table 2.3). As expected, the more polar residues, in particular Asp (11.9%) and Glu (11.3%), appear most often in interactions involving water at protein-protein interfaces. Cys (0.7%) is most rarely found. However, the aliphatic hydrophobic residues (Ala, Gly, Ile, Leu, Pro and Val) are surprisingly prevalent with 4.5 – 7.8% frequency, notably more so than His, Met, Phe or Trp (< 2.3 %). Glaser *et al.* [32] reported contact counts (within certain C_β - C_β cutoffs) at protein-protein interfaces that are generally similar except that Asp and Glu appear more than twice as frequently and Cys and Phe appear less than half as frequently in our water-mediated observations. Likewise, our results are in qualitative agreement with the report of Teyra and Pisabarro for “dual” and “wet” interactions between residues at protein-protein interfaces [33]. In their nomenclature, dual refers to an interaction that has both direct residue-residue interaction and water-mediated interaction, while wet refers to an interaction that is only water-mediated. When examining these preferences for waters having productive and Relevant interactions with both proteins, the fraction arising from residue sidechains carrying hydrogen bond donors or acceptors is enhanced (Arg, 9.6%; Asp, 18.4%; Glu, 17.0%) relative to those arising from hydrophobic sidechains. For the cases where the

waters are Relevant with respect to neither protein, the opposite is true – as expected (Ala, 11.0%; Ile, 6.9%; Leu, 13.0%; Pro, 9.9%; Thr, 8.8%; Val, 9.1%). However, as described by Teyra and Pisabarro [33], water interactions with non-polar residues may in some cases be energetically favorable from interactions involving backbone atoms (*vide infra*).

Table 2.3. Frequencies and HINT scores of water molecules at protein-protein interfaces with respect to interacting amino acid residues.

Residue Type	All Waters			Waters Relevant to 0			Waters Relevant to 1			Waters Relevant to 2		
	Wtd. Count ^a	Average HINT score ^b		Wtd. Count ^a	Average HINT score ^b		Wtd. Count ^a	Average HINT score ^b		Wtd. Count ^a	Average HINT score ^b	
		For All	For Type		For All	For Type		For All	For Type		For All	For Type
Ala	320	-28.51	-422.3	133	-48.90	-444.7	158	-27.42	-436.8	29	-6.95	-242.5
Arg	279	15.04	255.9	42	4.32	124.3	139	12.99	235.6	98	32.83	341.3
Asn	229	9.96	205.9	37	2.20	71.6	125	12.07	242.0	67	13.97	213.3
Asp	564	63.32	532.7	49	7.12	176.6	328	72.88	558.2	187	106.60	580.7
Cys	32	0.29	42.7	7	0.36	65.4	17	0.11	16.2	8	0.64	79.1
Gln	201	6.66	156.9	34	0.57	20.2	120	8.14	171.0	48	10.26	219.8
Glu	535	54.91	486.8	50	4.73	114.2	312	64.03	515.6	173	92.11	542.9
Gly	212	-9.89	-221.1	53	-11.15	-254.3	113	-10.87	-241.3	46	-5.98	-132.7
His	75	2.71	170.8	13	1.36	130.6	42	2.38	142.6	21	5.15	251.9
Ile	212	-21.49	-481.7	84	-36.71	-529.1	107	-20.31	-478.3	21	-6.31	-308.1
Leu	369	-35.54	-456.8	157	-62.71	-484.2	179	-33.34	-467.7	33	-8.65	-267.3
Lys	220	0.86	18.5	57	-5.16	-110.4	110	0.67	15.3	54	8.49	160.6
Met	107	-9.63	-425.2	41	-16.30	-483.4	54	-9.63	-448.8	13	-1.68	-135.7
Phe	75	1.25	79.7	16	0.16	12.2	40	1.83	113.9	18	1.14	63.3
Pro	278	-22.61	-385.3	120	-36.76	-369.9	137	-22.18	-408.0	21	-6.86	-327.3
Ser	260	-5.69	-103.8	69	-9.28	-162.8	137	-6.12	-112.3	54	-0.37	-7.0
Thr	307	-19.10	-294.5	106	-34.25	-390.5	158	-18.75	-298.6	44	-1.93	-45.2
Trp	52	1.62	148.1	10	0.55	67.2	26	1.48	143.9	16	3.21	205.0
Tyr	147	6.13	198.0	23	1.64	87.3	81	6.92	214.3	43	9.52	225.6
Val	267	-27.61	-489.7	110	-45.96	-503.8	131	-26.26	-503.2	26	-9.12	-360.7

Notes: ^aWeighted count is calculated as $\sum_n \{ |A_i| / \sum_i |A_i| \}$, where A_i are the interaction

HINT scores by residue type (i) interacting with water n; ^bHINT scores are averaged two ways: first, over all waters in set or Relevance subset, second, by frequency (weighted count) of that residue type in set or Relevance subset.

While optimizing and scoring, each water molecule in the present report was treated as a small ligand in a site defined by neighboring residues. The average HINT score for the waters in the entire data set is -17 ($\Delta G \sim +0.03 \text{ kcal mol}^{-1}$); thus, the average interaction of a water with only one of its neighboring proteins would be half of that value, i.e., essentially negligible. Table 2.3 lists the HINT score values for each of the twenty amino acid types, first by averaging over all waters in the data set, and second by averaging over all waters interacting (by weighted count) with that residue type. The first average, over all waters, reveals the reason for the near zero value for the average interaction energy of an interfacial water with its environment: there is a complex mix of favorable and unfavorable interactions with water, depending on the residue type. The latter average, weighted instead by the frequency of that particular water-residue interaction, represents the score that would be expected if a water interacted with only that residue and thus reveals the specific benefits of interacting with some residue types, e.g., Asp ($-1.03 \text{ kcal mol}^{-1}$), Glu ($-0.95 \text{ kcal mol}^{-1}$), or Arg ($-0.50 \text{ kcal mol}^{-1}$), vs. the cost of interacting with others, e.g., Pro ($+0.75 \text{ kcal mol}^{-1}$), Ala ($+0.82 \text{ kcal mol}^{-1}$), Met ($+0.83 \text{ kcal mol}^{-1}$), Leu ($+0.89 \text{ kcal mol}^{-1}$), Ile ($+0.94 \text{ kcal mol}^{-1}$) or Val ($+0.95 \text{ kcal mol}^{-1}$). The biggest surprise here is that Lys, while responsible for 4.6% of interactions with interface waters has, on average, a minimal contribution to the water score. This is partly because Lys, if NZ is protonated as expected, is only a hydrogen bond donor and is unable to accept from water, but also, the long hydrophobic polymethylene sidechain of Lys may

be interacting unfavorably with some water molecules compared with the other “basic” residue Arg that has multiple polar atoms and can act as an acceptor through its sidechain π system. Also, Lys with its flexible sidechain is more likely to be disordered and its atomic coordinates are thus less certain. Furthermore, Jones and Thornton [34] noted that Lys frequency is depleted at protein-protein interfaces relative to protein surfaces.

The differences in interactions between water molecules Relevant to zero, one and two proteins are instructive. First, these waters have average HINT scores of -284 (+0.55 kcal mol⁻¹), 9 (-0.02 kcal mol⁻¹) and 236 (-0.46 kcal mol⁻¹), respectively. Also, as calculated with the averages over all waters that are Relevant to zero, one or two proteins (Table 2.3), the interactions are dominated by Ala, Ile, Leu, Pro, Thr and Val (generally unfavorable, with negative HINT scores) for the waters Relevant to neither protein, and dominated by favorable interactions with Arg, Asp and Glu for the waters Relevant to both proteins.

2.3.3 Sidechain and Backbone Preferences for Interface water:

Teyra and Pisabarro [33] showed that a significant fraction of interface water molecules appear to be interacting with backbone atoms on one of both of the proteins. Rodier *et al.* calculate that 12% of water interactions at protein-protein interfaces are with backbone NH and 33% with CO [28]. Our analysis of backbone and sidechain interactions reveals interesting details: the average interaction score for a water with a

backbone atom [C, O, (OXT), CA, HA, N, HN, (HN2, HN3)] is favorable (57, -0.11 kcal mol⁻¹), while on average the interaction with sidechain atoms is unfavorable (-74, +0.14 kcal mol⁻¹). Obviously, this can be explained by the ability, although usually shielded by the sidechain, of the backbone to be both a hydrogen bond donor (via NH) and acceptor (via O). Table 2.4 lists the weighted counts and average scores for backbone and sidechain interactions with water by residue type. Calculations of weighted interaction counts, which are based on HINT scores of H-bond optimized structures and not simple distance metrics, suggest (Table 2.4) that only 21.5% of the water-protein interactions involve backbone atoms, and that the remaining 78.5% arise from sidechain atoms. Thus, while the backbone interactions are mostly favorable, they play a lesser role in describing the protein-protein interface than do the sidechain interactions. The average scores, when weighted by the frequency of interactions for the residue types for either the backbone or sidechain (Table 2.4), clearly show that the backbone interactions are remarkably consistent and independent of residue identity. These scores represent how strongly a single water would interact with a residue backbone (or sidechain) isolated from all other interactions.

Table 2.4. Frequencies and HINT scores of water molecules at protein-protein interfaces with respect to backbones and sidechains of interacting amino acid residues.

Residue Type	All	Interacting with backbone			Interacting with sidechain		
	Wtd. Count ^a	Wtd. Count ^b	Average HINT score ^c		Wtd. Count ^b	Average HINT score ^c	
			For All	For Type		For All	For Type
Ala	320	63	5.39	403.2	257	-33.90	-625.9
Arg	279	54	4.27	376.2	225	10.76	227.1
Asn	229	54	4.44	388.6	175	5.52	149.4
Asp	564	65	5.29	388.4	499	58.03	551.3
Cys	32	18	1.47	390.3	14	-1.18	-393.1
Gln	201	39	3.27	395.7	162	3.39	99.2
Glu	535	51	4.13	384.7	484	50.79	497.5
Gly ^d	212	212	-9.89	-221.1	0	0.00	---
His	75	20	1.71	410.0	55	1.00	85.5
Ile	212	29	2.60	430.5	183	-24.09	-624.3
Leu	369	53	4.78	429.2	316	-40.32	-604.9
Lys	220	45	3.80	398.7	175	-2.94	-79.7
Met	107	20	1.82	440.2	88	-11.44	-618.2
Phe	75	33	2.81	409.7	42	-1.56	-175.3
Pro	278	45	2.81	298.7	234	-25.43	-516.0
Ser	260	71	5.39	358.2	189	-11.09	-278.5
Thr	307	59	5.03	407.0	249	-24.13	-459.7
Trp	52	20	1.84	443.6	32	-0.22	-32.8
Tyr	147	36	2.90	380.2	111	3.23	138.4
Val	267	36	3.33	439.1	231	-30.95	-634.3

Notes: ^aSame as Table 2.3; ^bWeighted count is calculated as $\sum_n \{ |A_i| / \sum_i |A_i| \}$, where A_i

are the interaction HINT scores for the backbone or sidechain by residue type (i)

interacting with water n; ^cHINT scores are averaged two ways: first, over all waters in set

or Relevance subset, second, by frequency of the backbone or sidechain contribution (weighted count) of that residue type in set or Relevance subset; ^dFor Gly (and all other residues) the CA atom is considered part of the backbone, thus Gly has no sidechain.

However, the total score only tells part of the story and obscures the operational details on how the waters actually interact with the proteins. Figure 2.3 displays (A) backbone and (B) sidechain interactions by residue type and interaction class, averaged over all water molecules in the data set. In particular, favorable polar (hydrogen bonds and acid/base) interactions are plotted as positive contributions, while unfavorable polar (acid/acid and base/base) and unfavorable hydrophobic (i.e., interacting with polar) interactions are plotted as negative contributions. Figure 2.3A (backbone) and 2.3B (sidechain) illustrates the average scores for each residue type, i.e., weighted by the number of water interactions of those types in the data set. These charts emphasize the similar role of backbone interactions for nearly all residue types, excluding Gly. This contribution is largely independent of the Relevance of the water involved, increasing only modestly from 49 (-0.10 kcal mol⁻¹) to 57 (-0.11 kcal mol⁻¹) and 67 (-0.13 kcal mol⁻¹) for waters Relevant to zero, one and both proteins, respectively. At the same time, the average sidechain interaction scores respond dramatically, increasing from -333 (+0.65 kcal mol⁻¹) to -48 (-0.09 kcal mol⁻¹) and 169 (-0.33 kcal mol⁻¹).

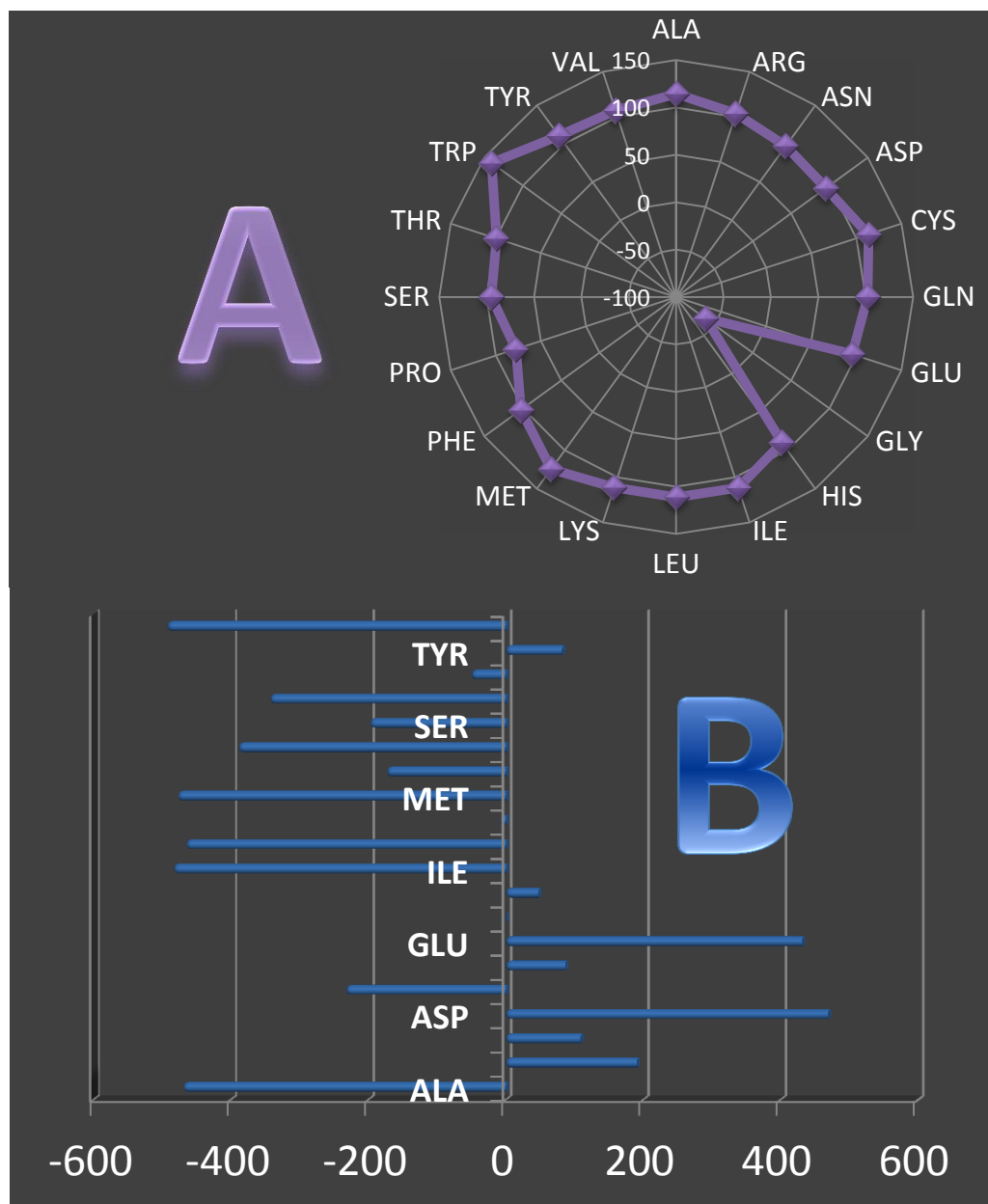


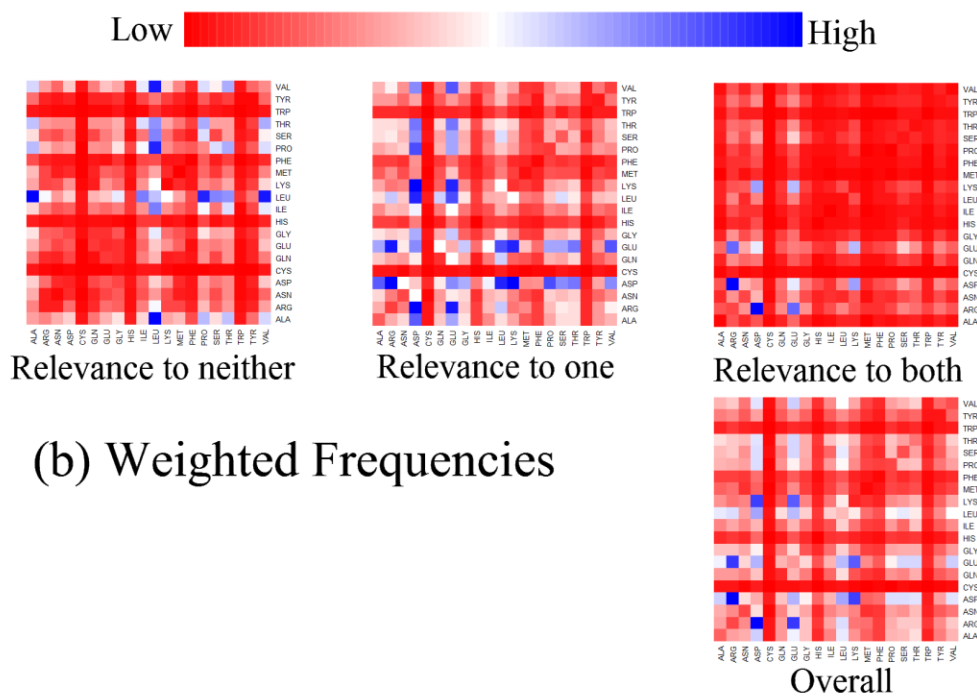
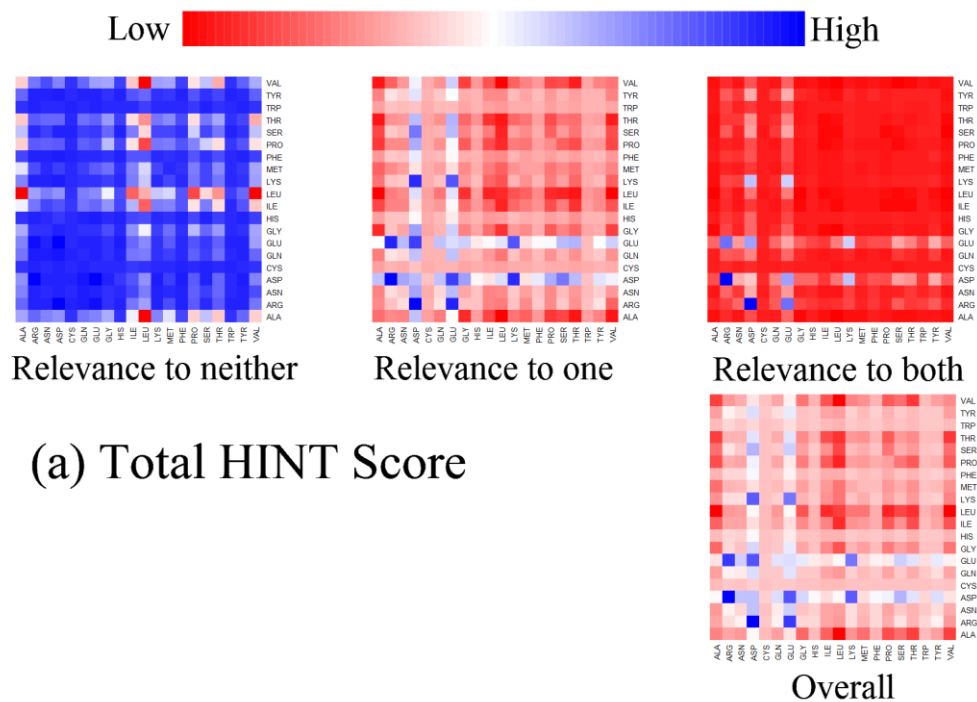
Figure 2.3. Average *HINT* interaction scores for waters at protein-protein interfaces: (A) scores normalized by weighted count of residue types (Table 2.3) with protein backbone atoms; and (B) scores normalized by weighted count of residue types with protein sidechain atoms.

2.3.4 Residue-Pair Preferences for Interface H₂O:

By definition, waters found at the interface should interact with residues on both proteins. The floor value for interactions of $|10|$ HINT score units, or about $|0.02|$ kcal mol⁻¹, excludes a small number of waters (< 1.5%) from having any recorded interaction with one (or in rare cases both) of the proteins. As shown above, in Table 2.3, there is a residue identity preference for water-mediated interactions at protein-protein interfaces and this differs depending on the role the water plays at the interface. More specifically, it is shown here that there are distinct residue identity preferences for mediated residue pairs. Consider first the total gross sum of HINT scores for each pair of amino acid residue types as graphically illustrated with color heat maps in Figure 2.4A for all waters, and those Relevant to neither, one and both proteins. This depiction combines both the strength of interaction and frequency of interaction for the residue pairs. Overall, in Figure 2.4A, the most energetically favorable pairs for interface water involve one of the polar residues, especially the hydrogen bond acceptors Asp and Glu. These can partner with each other – intriguingly Asp-H₂O-Glu scores higher than Asp-H₂O-Asp or Glu-H₂O-Glu – or partner extensively with the hydrogen bond donor or amphiprotic residues (Arg, Asn, Gln, Lys, Ser, Thr, Tyr), but not significantly with His or Trp. The most unfavorable pairings involve the most hydrophobic and aliphatic residues Ala, Ile, Leu, Pro and Val. The intermediate effect of Phe may be due to its aromatic ring being a potential hydrogen bond acceptor. The scores for waters with Relevance to neither

protein (Figure 2.4A) are dominated by strongly unfavorable interactions with hydrophobic residues, especially Leu and Ile, while the scores for waters with Relevance to both proteins (Figure 2.4A) are most favorable for interactions involving Asp and Glu, particularly when partnered with Arg. However, it must be noted that the total HINT score shown here reflects both the frequency of these residue pairings as well as their relative strengths.

Frequencies weighted as described in Materials and Methods are set out in Figure 2.4B. Overall, water-mediated interactions involving Asp, Glu, Lys, Arg and surprisingly Leu are clearly dominant while those involving Cys, His, Phe and Trp are most infrequent. Waters not relevant to either protein (Figure 2.4B) generally interact with hydrophobic residues. For waters relevant to both proteins (Figure 2.4B), the most frequent pairs are Asp and Glu with Arg and Lys. Also, Asp and Glu are found fairly frequently in water-bridged interactions with Asn, Gln, Ser and Tyr. Note that the color pattern here is strikingly similar to that of the overall score for the doubly relevant case (Figure 2.4A), which indicates that frequency of pair interactions is a key factor. Finally (Figure 2.4C), the score normalized by weighted frequency reveals the relative average energetic importance of each interaction pair ranging between -602 score units (+1.17 kcal mol⁻¹) and 541 score units (-1.05 kcal mol⁻¹).



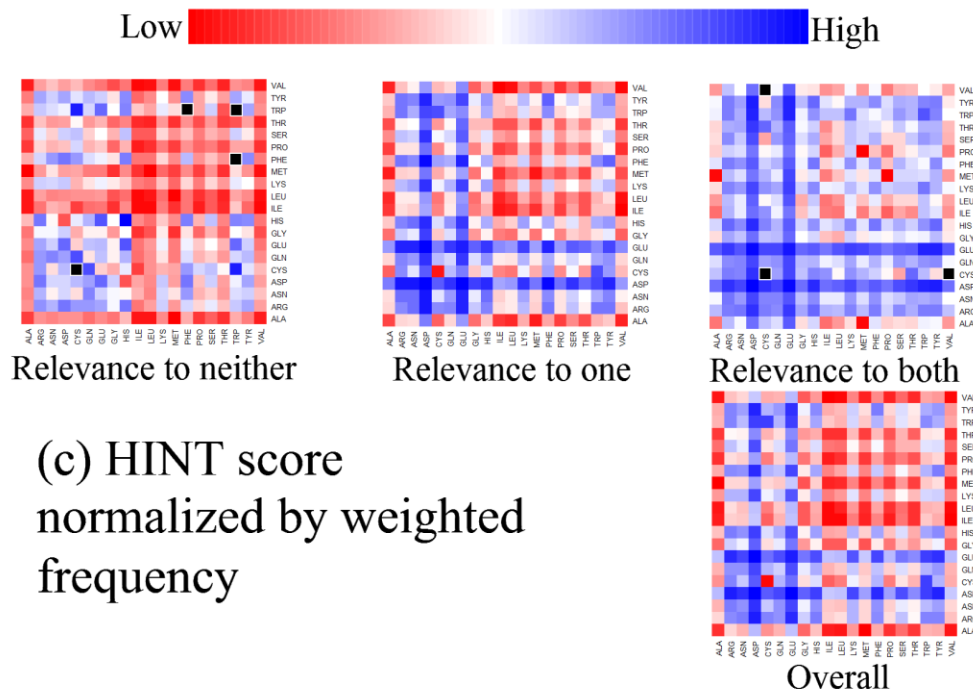


Figure 2.4. Color heat maps depicting Res_1 - H_2O - Res_2 interactions for water molecules found at protein-protein interfaces: All maps are linearly scaled over the maximum range of values for that data set. (A) Total HINT score between waters and Res_1/Res_2 : all waters in data set (minimum score -71,358, maximum score 114,632); waters in set with Relevance to neither protein (minimum -41,868, maximum 3,685); waters in set with Relevance to one protein (minimum -26,470, maximum 50,220); waters in set with Relevance to both proteins (minimum -3,534, maximum 60,727). (B) Weighted count of Res_1/Res_2 with water interactions: all waters in data set (minimum count 0.1, maximum count 242.7); waters in set with Relevance to neither protein (minimum 0.0, maximum 74.0); waters in set with Relevance to one protein (minimum 0.1, maximum 113.3); waters in set with Relevance to both proteins (minimum 0.0, maximum 114.5). (C)

Average HINT score (normalized by weighted count) between waters and Res₁/Res₂: all waters in data set (minimum average score -601.6, maximum average score 540.5); waters in set with Relevance to neither protein (minimum -624.3, maximum 483.0); waters in set with Relevance to one protein (minimum -633.7, maximum 499.7); waters in set with Relevance to both proteins (minimum -875.1, maximum 680.9). Cells colored black represent cases where the weighted count was zero, and the HINT score normalization yields an undefined value.

2.3.5 Residue-Pair Roles in Water Interactions:

Cluster analysis of the matrices behind the heat maps of Figure 2.4 provide additional insight into the roles that residues play in interacting with waters. Figure 2.5 sets out dendograms of average HINT score for all waters (A), waters Relevant to either protein (B), waters Relevant to one protein (C) and waters Relevant to both proteins (D). The Relevant to zero case is most different from the others. Generally, the most hydrophobic aliphatic residues (Ala, Ile, Leu, Met, Pro, Thr and Val) are clustered together with Thr (except for the case of Relevant to both, Figure 2.5D). At the opposite extreme, Asp and Glu are clustered, save the Relevant to zero case, far from all other clusters. The ability of water to be equally proficient as both a hydrogen bond donor and an acceptor somewhat blurs the distinction between residues that are formally acids or bases when they interact with it. The remaining residue types divide into two clusters

with somewhat variable membership. Because the aromatic ring of Phe can act as a hydrogen bond acceptor, it clusters with an eclectic group of residues: Ser, Gly, Gln, Lys, Trp and/or Thr, but surprisingly not Tyr. For waters Relevant to neither protein, there are typically few favorable interactions, regardless of the character of the residues interacting with the water. The patterns in the associated dendrogram (Figure 2.5B), other than the large distance separating the hydrophobic residues from the polar residues, are difficult to discern; here, Asp and Glu are not clustered together. A likely determinant defining these clusters may involve residue size.

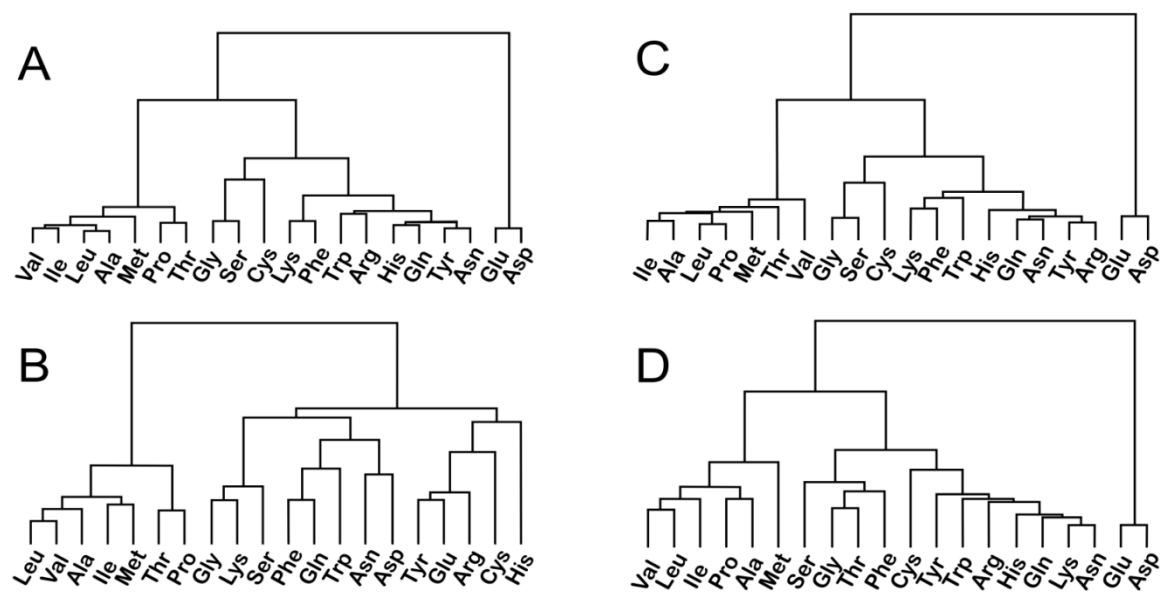


Figure 2.5. *Dendrograms indicating clustering of residues with respect to average HINT score (normalized by weighted count) in Res_1 - H_2O - Res_2 interactions:* (A) for all waters; (B) for waters with Relevance to neither protein; (C) for waters with Relevance to one protein; and (D) for waters with Relevance to both proteins.

This analysis of 4741 water molecules at 179 protein-protein interfaces has revealed new information about the various roles that water can play at interfaces. This analysis was anchored by the HINT free energy forcefield and the Relevance metric. The former characterizes the types and qualities of interactions between the interface waters and proteins, while the latter is a simple parameter that was previously shown to identify water molecules conserved/non-conserved in ligand binding sites [25]. Relevance was shown in the present report to be a useful classifier for identifying the roles and partner proteins and residues for interfacial waters.

Previous studies of water in the interface between interacting proteins have generally relied solely on interatomic distances in non-protonated crystallographic models to mark interactions between waters and proteins. This approach, however, often poorly represents the complex and subtle energetics and geometric preferences of hydrogen bonding. Thus, this study was performed with all atoms after exhaustive optimization of all water orientations [23] to surmount local minima in our models. The hydropathic minimization procedure rewards favorable polar interactions, i.e., hydrogen bonds and acid/base, and penalizes unfavorable polar, i.e., acid/acid and base/base, and hydrophobic-polar interactions.

2.3.6 Waters Relevant to Multiple Proteins: How important is the energetic contribution of water to protein-protein associations?

This is an important question since most protein-protein docking utilities ignore the actual (and potential) presence of water at putative interfaces. Unfortunately, it is difficult to determine *de novo* which water molecules are or will be energetically important. Only 59 (33%) of the protein-protein complexes have an overall favorable water contribution considering all interface waters, but 145 (81%) have a favorable contribution from waters Relevant to one/both proteins and nearly all, 173 (97%), have a favorable contribution from waters that are Relevant to both (the other 6 protein pairs have no waters of this class). The average scores are: -2072 (+4.02 kcal mol⁻¹), -84 (+0.16 kcal mol⁻¹) and 1297 (-2.52 kcal mol⁻¹) for the water sets at these interfaces Relevant to 0, 1 and 2 proteins, respectively. While each water at each protein-protein interface should be evaluated for its own specific environment and role, the overall analysis shows that the total water contribution can be quite important: ranging up to 5845 (-11.35 kcal mol⁻¹) per protein pair for the water sets Relevant to both proteins and presumably “bridging”. Also, the Relevance-based classification scheme we have proposed certainly has merit for facilely selecting waters that should be considered in modeling protein-protein complexes.

The energetic role of bridging water molecules at interfaces is clear and well understood, although difficult to experimentally quantify [35-38]. Reichmann *et al.* [36] performed double mutant cycle analysis on eight residue pairs (all with SASA < 10 Å²) that appeared to be bridged by waters at the TEM1/BLIP (1jtg) interface; only six of the eight pairs are truly bridged by water (residue-residue distance > 3.8 Å), yielding an average $\Delta\Delta G_{KA}$ [36] for these water-mediated hydrogen bonds of -0.003 kcal mol⁻¹, i.e., essentially having an energetically neutral effect on interface stability much as shown above (+0.03 kcal mol⁻¹) for an average interface water in our analysis. Only four waters support these six pairs because two of the waters interact with more than one residue on one of the partner proteins (one highly Relevant to both proteins and the other Relevant to only BLIP), and it is thus impossible to isolate the specific energetic contribution from experimental double mutant data for these two waters. Of the remaining two waters, HINT analysis showed that one (HOH72) is Relevant to only TEM1 and the other (HOH111) is not Relevant to either protein, supporting the view that the former is strongly associated with TEM1's Glu104 and weakly associated with BLIP's Ser146, while the latter is only weakly associated with Gln99 and repulsive with respect to Ser128. Even here, interpretation is not straightforward: mutating these residues to Ala may or may not excise the putative bridging waters, just change their environment. In fact, there may even be space for more than one water in some of the double mutant complexes.

Another, more subtle, role is that bridging waters also serve as nano-scale pH buffers (see Figure 2.6). By simply re-orienting, individual water molecules can swap between acting as donors and acceptors as necessary to maintain a mediated (wet) interaction and the integrity of the entire interface. In contrast, direct hydrogen-bonded (dry) interactions between proteins may be weakened by changes in pH. Of course, hydrophobic interactions between protein surfaces are largely unaffected by changes in pH. Evidence for this role of waters was given in the cluster dendogram of Figure 2.5D. Other than the distinct clustering of Asp with Glu and the aliphatic hydrophobic residues with Met, the remaining twelve residues cluster together regardless of their hydrogen bond donor or acceptor character.

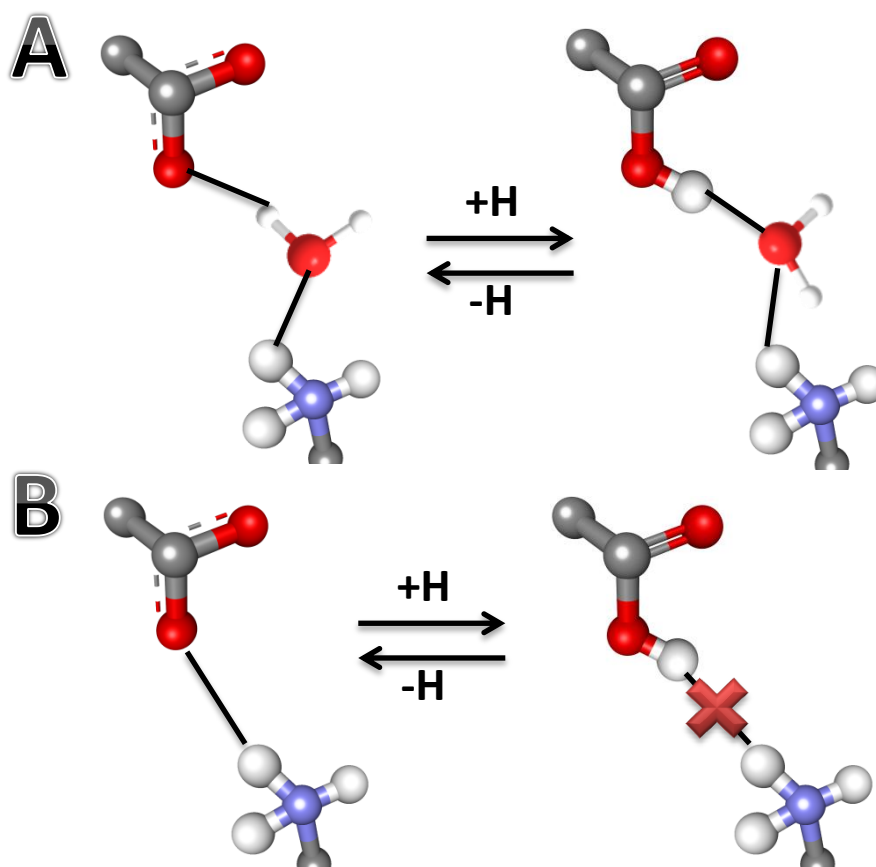


Figure 2.6. *Water as a nano-scale buffer:* (A) increasing the pH of the system is compensated by a reorientation of the bridging water molecule; (B) direct unmediated interactions are less able to compensate for changes in pH.

2.3.7 Waters not Relevant to either protein: Why are there so many waters that are seemingly non-Relevant?

There are a large number of water molecules that do not appear to have a role in structure. A brief survey of moderate-resolution complex structures revealed essentially the same fraction of waters that lacked favorable interactions with their protein pairs as did the much more extensive high-resolution set. These results suggest that this type of water is a conserved phenomenon as only the most ordered water molecules will have interpretable experimental electron density for resolutions poorer than 2.5 Å.

The analysis described above did not attempt to detect water molecules that are involved in water network chains, i.e., waters that are strongly and favorably interacting with two or more other waters that are themselves Relevant to a protein. To investigate this possibility (for an example, see Figure 2.7), we added the water molecules that were Relevant to one or both proteins to their partners of highest Relevance and examined the remaining (i.e., initially Relevance zero) waters with respect to these “hydrated” protein entities. Only 326 (27%) of the remaining waters were found to have Relevance (≥ 0.25) with one and 30 (2.5%) were found to have Relevance to both hydrated proteins. The latter represent water molecules networked in three-water chains. It is a surprisingly low number, but the Relevance-based definition of networking is fairly stringent, and these waters are already constrained to be within the confines of the interface region while not

already interacting favorably with other protein residues. It is therefore unlikely that significant numbers of these water molecules would turn up to be involved in higher order chains.

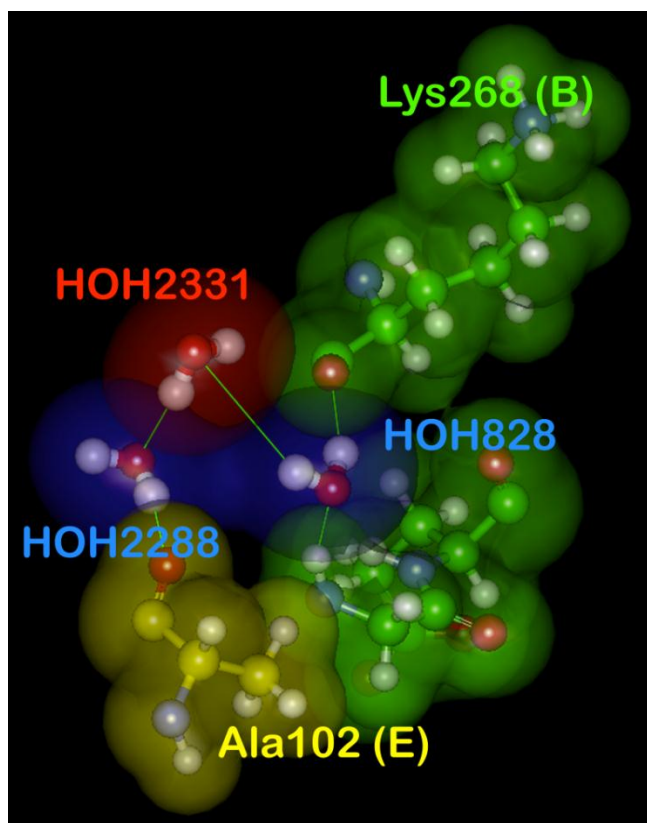


Figure 2.7. *Water in chain of three water molecules:* HOH2331 (red) from protein complex 1kxq is Relevant with respect to waters HOH828 and HOH2288 (blue), which are each, in turn, Relevant to the proteins in the complex.

As discussed above, Relevance zero waters have overall unfavorable interactions with their partner proteins, which largely arises from interactions with the protein's

sidechains. It can be seen in Figure 2.8 that the dominant unfavorable interaction type for these waters is hydrophobic-polar; the favorable polar interactions shown in Figure 2.8 are due to interactions with the backbone (see Table 2.4). Most of the Relevance zero water molecules within the interface are trapped in hydrophobic environments as “hydrophobic bubbles”.

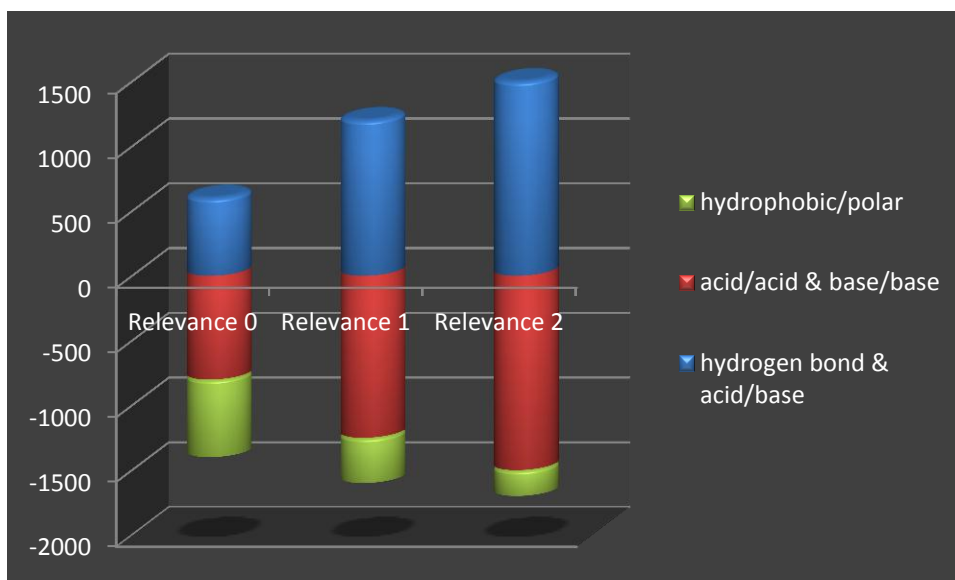


Figure 2.8. *Interaction type scores for waters with Relevance to zero, one and two proteins.*

It would appear that these hydrophobic bubbles represent a conserved motif. One intriguing possibility is that a certain amount of instability is required in protein-protein interfaces to ensure that the associations are dynamic. Meenan *et al.* described the role of

some waters found at the 1.77 Å structure of the colicin E9 endonuclease-immunity protein 2 interface as “aggravating” the binding between the two proteins [39]. Sundaralingham and Sekharudu [40] proposed that water may be considered a “lubricant” in dynamic protein folding and interaction. Teyra and Pisabarro [33] classified the complexes in their analysis as “obligate” meaning that the association is permanent as these interfaces were formed concurrent with chain folding and “transient” where the component proteins fold independently of their association [34,41]. The latter of course includes proteins involved in regulation of biochemical pathways and signal transduction. Similar concentrations (10 vs. 11 water-bridged residues/1000 Å² surface contact area) of waters were found in the two groups [33]. The primary data set used here is composed entirely of transient proteins. However, for comparison, a set of 12 homo-dimers, predominantly obligate [42] was examined, containing 546 water molecules selected as described above. In the obligate set, there were 113 waters (21%) Relevant to neither protein, 302 waters (55%) Relevant to one protein, and 131 (24%) Relevant to both. As would be expected, there are somewhat higher fractions of waters with Relevance to both one and two proteins, and a smaller fraction that are non-Relevant. It appears that protein-protein interfaces, independent of the longevity of their association, commonly include water molecules that do not have favorable interactions with either protein, although the possibility that some or maybe even many of these waters are incorrectly

assigned electron density or other crystallographic artifacts cannot be completely discounted [10].

2.3.8 Predictions of water roles:

The principle of correlated mutations is that interface contacts co-evolve to maintain or enhance biologically important associations [43-46]. Using this principle, Samsonov *et al.* recently reported [47] that including solvent matrices in contact predictions [48,49] of protein-protein interfaces improve these predictions by 20-30%. However, no residue level information was reported. It is noted above (Figure 2.4B) that the observed frequency of Asp-H₂O-Glu interactions, in waters Relevant to one or both proteins, is notably higher than Asp-H₂O-Asp or Glu-H₂O-Glu interactions. This suggests that water molecules may act as spacers to effectively lengthen Asp sidechains to mimic Glu sidechains. We observed a similar role for Asp+H₂O in protein/DNA interactions [19,20]. Whether this is a consequence of correlated mutations is difficult to say, but it is an intriguing possibility.

Water Relevance may be used as a metric to predict the locations of water molecules computationally. Kellogg's group previously described [50] an algorithm for generating water solvent arrays around proteins or in binding pockets that is superficially similar to the GRID algorithm proposed by Goodford [51]. This protocol can easily be adapted to use Relevance-based criteria for water placement; for this purpose it is

especially significant that Relevance is calculated independent of (experimentally-determined) crystallographic data like B-factors. However, this present study indicates that the presence of as many as one-in-four energetically unfavorable water molecules is an apparently conserved motif. Their positions and orientations will almost certainly be difficult to predict! Nevertheless, common structural features such as hydrophobic bubbles may aid in this understanding and in developing algorithms for computationally orienting and locating these waters. In previous studies, Kellogg's group proposed that these "unfavorable" water molecules may actually have an important biological purpose [21]. It is fair to say that we will not be able to completely model or exploit protein-protein interfaces until we can properly deal with all of the water molecules that are present.

2.4 Conclusion:

In this chapter it was shown that only about 21% of all waters at protein-protein interfaces are truly bridging while 26% are seemingly only trapped at the interface. While it was probably not surprising that Asp and Glu residues appeared most frequently in interactions with bridging waters, it was somewhat surprising that bridging is dominated by Asp-H₂O-Arg and Glu-H₂O-Arg interactions but Asp-H₂O-Asp or Glu-H₂O-Glu interactions are relatively infrequent, even compared to Asp-H₂O-Glu. Also of note is that certain unfavorable interaction motifs are conserved. The results from this

work have implications for the design of compounds that can break protein-protein interactions.

REFERENCES

REFERENCES

1. Zacharias, M. *Protein, Protein Complexes: Analysis, Modeling and Drug Design* Imperial College Press: 2010; , pp 400.
2. Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **2004**, *3*, 301-317.
3. Huang, R.; Martinez-Ferrando, I.; Cole P.A. Enhanced interrogation: emerging strategies for cell signaling inhibition. *Nat. Struct. Mol. Bio.* **2010**, *17*, 646-649.
4. Wilson A.J. Inhibition of protein-protein interactions using designed molecules. *Chem. Soc. Rev.* **2009**, *38*, 3289-3300.
5. Betzi, S.; Guerlesquin, F.; Morelli, X. Protein-protein interaction inhibition (2P2I): find fewer undruggable targets. *Comb. Chem. High Throughput Screen.* **2009**, *12*: 968-983.
6. Bidwell, G. L.,3rd; Raucher, D. Therapeutic peptides for cancer therapy. Part I - peptide inhibitors of signal transduction cascades. *Expert Opin. Drug Deliv.* **2009**, *6*, 1033-1047.
7. Wells, J.A.; McClendon, C.L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001-1009.

8. Hardcastle, I. R.; Liu, J.; Valeur, E.; Watson, A.; Ahmed, S. U.; Blackburn, T. J.; Bennaceur, K.; Clegg, W.; Drummond, C.; Endicott, J. A.; Golding, B. T.; Griffin, R. J.; Gruber, J.; Haggerty, K.; Harrington, R. W.; Hutton, C.; Kemp, S.; Lu, X.; McDonnell, J. M.; Newell, D. R.; Noble, M. E. M.; Payne, S. L.; Reville, C. H.; Riedinger, C.; Xu, Q.; Lunec, J. Isoindolinone Inhibitors of the Murine Double Minute 2 (MDM2)-p53 Protein-Protein Interaction: Structure-Activity Studies Leading to Improved Potency. *J. Med. Chem.* **2011**, *54*, 1233-1243.

9. Dutta, S.; Berman, H. M. Large Macromolecular Complexes in the Protein Data Bank: A Status Report. *Structure* **2005**, *13*, 381-388.

10. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal* **2008**, *275*, 1-21.

11. Mueller, M.; Jenni, S.; Ban, N. Strategies for crystallization and structure determination of very large macromolecular assemblies. *Curr. Opin. Struct. Biol.* **2007**, *17*, 572-579.

12. Cho, S.; Swaminathan, C. P.; Bonsor, D. A.; Kerzic, M. C.; Guan, R.; Yang, J.; Kieke, M. C.; Andersen, P. S.; Kranz, D. M.; Mariuzza, R. A.; Sundberg, E. J. Assessing Energetic Contributions to Binding from a Disordered Region in a Protein-Protein Interaction. *Biochemistry (N. Y.)* **2010**, *49*, 9256-9268.

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

14. Tripos, L.P. www.tripos.com. St. Louis, MO, USA.
15. Amadasi, A.; Spyraakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the Energetics of Water–Protein and Water–Ligand Interactions with the “Natural” HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules. *J. Mol. Biol.* **2006**, *358*, 289-309.
16. Eugene Kellogg, G.; Abraham, D. J. Hydrophobicity: is LogP(o/w) more than the sum of its parts? *Eur. J. Med. Chem.* **2000**, *35*, 651-661.
17. Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **1983**, *168*, 595-617.
18. Levitt, M.; Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **1988**, *201*, 751-754.
19. Spyraakis, F.; Cozzini, P.; Bertoli, C.; Marabotti, A.; Kellogg, G. E.; Mozzarelli, A. Energetics of the protein-DNA-water interaction. *BMC Struct. Biol.* **2007**, *7*, 4.
20. Marabotti, A.; Spyraakis, F.; Facchiano, A.; Cozzini, P.; Alberti, S.; Kellogg, G. E.; Mozzarelli, A. Energy-based prediction of amino acid-nucleotide base recognition. *Journal of Computational Chemistry* **2008**, *29*, 1955-1969.

21. 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. *Curr. Med. Chem.* **2004**, *11*, 3093-3118.
22. Burnett, J. C.; Botti, P.; Abraham, D. J.; Kellogg, G. E. Computationally accessible method for estimating free energy changes resulting from site-specific mutations of biomolecules: Systematic model building and structural/hydropathic analysis of deoxy and oxy hemoglobins. *Proteins: Structure, Function, and Genetics* **2001**, *42*, 355-377.
23. Kellogg, G. E.; Chen, D. L. The Importance of Being Exhaustive. Optimization of Bridging Structural Water Molecules and Water Networks in Models of Biological Systems. *Chemistry & Biodiversity* **2004**, *1*, 98-105.
24. The R Project for Statistical Computing. <http://www.R-project.org>, Vienna, Austria.
25. Amadasi, A.; Surface, J. A.; Spyraakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust Classification of “Relevant” Water Molecules in Putative Protein Binding Sites. *J. Med. Chem.* **2008**, *51*, 1063-1067.
26. García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *Journal of Molecular Modeling* **2003**, *9*, 172-182.

27. Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445-464.
28. Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of protein–protein interfaces. *Proteins: Structure, Function, and Genetics* **2005**, *60*, 36-45.
29. Papageorgiou, A. C.; Shapiro, R.; Acharya, K. R. Molecular recognition of human angiogenin by placental ribonuclease inhibitor--an X-ray crystallographic study at 2.0 Å resolution. *EMBO J.* **1997**, *16*, 5162-5177.
30. Karplus, P. A.; Faerman, C. Ordered water in macromolecular structure *Curr. Opin. Struct. Biol.* **1994**, *4*, 770-776.
31. Levitt, M.; Park, B. H. Water: now you see it, now you don't. *Structure* **1993**, *1*, 223-226.
32. Glaser, F.; Steinberg, D. M.; Vakser, I. A.; Ben-Tal, N. Residue frequencies and pairing preferences at protein–protein interfaces. *Proteins: Structure, Function, and Genetics* **2001**, *43*, 89-102.
33. Teyra, J.; Pisabarro, M. T. Characterization of interfacial solvent in protein complexes and contribution of *wet spots* to the interface description. *Proteins: Structure, Function, and Bioinformatics* **2007**, *67*, 1087-1095.

34. Jones, S.; Thornton, J. M. Principles of protein-protein interactions *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13-20.
35. Papoian, G. A.; Uler, J.; Wolynes, P. G. Role of Water Mediated Interactions in Protein-Protein Recognition Landscapes. *J. Am. Chem. Soc.* **2003**, *125*, 9170-9178.
36. Reichmann, D.; Phillip, Y.; Carmi, A.; Schreiber, G. On the Contribution of Water-Mediated Interactions to Protein-Complex Stability. *Biochemistry (N. Y.)* **2008**, *47*, 1051-1060.
37. Jang, D. S.; Cha, H. J.; Cha, S. S.; Hong, B. H.; Ha, N. C.; Lee, J. Y.; Oh, B. H.; Lee, H. S.; Choi, K. Y. Structural double-mutant cycle analysis of a hydrogen bond network in ketosteroid isomerase from *Pseudomonas putida* biotype B. *Biochem. J.* **2004**, *382*, 967-973.
38. Langhorst, U.; Backmann, J.; Loris, R.; Steyaert, J. Analysis of a Water Mediated Protein-Protein Interactions within RNase T1. *Biochemistry (N. Y.)* **2000**, *39*, 6586-6593.
39. Meenan, N. A.; Sharma, A.; Fleishman, S. J.; Macdonald, C. J.; Morel, B.; Boetzel, R.; Moore, G. R.; Baker, D.; Kleanthous, C. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10080-10085.
40. Sundaralingam, M.; Sekharudu, Y. C. Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science* **1989**, *244*, 1333-1337.

41. Nooren, I. M.; Thornton, J. M. Structural Characterisation and Functional Significance of Transient Protein-Protein Interactions. *J. Mol. Biol.* **2003**, 325, 991-1018.
42. Zhu, H.; Domingues, F. S.; Sommer, I.; Lengauer, T. NOXclass: prediction of protein-protein interaction types. *BMC Bioinformatics* **2006**, 7, 27.
43. Gregoret, L. M.; Sauer, R. T. Additivity of mutant effects assessed by binomial mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, 90, 4246-4250.
44. Lee, C.; Levitt, M. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature* **1991**, 352, 448-451.
45. Mintseris, J.; Weng, Z. Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102, 10930-10935.
46. Lee, B.; Kim, D. A new method for revealing correlated mutations under the structural and functional constraints in proteins. *Bioinformatics* **2009**, 25, 2506-2513.
47. Samsonov, S. A.; Teyra, J.; Anders, G.; Pisabarro, M. T. Analysis of the impact of solvent on contacts prediction in proteins. *BMC Struct. Biol.* **2009**, 9, 22.

48. Gobel, U.; Sander, C.; Schneider, R.; Valencia, A. Correlated mutations and residue contacts in proteins. *Proteins* **1994**, *18*, 309-317.
49. Halperin, I.; Wolfson, H.; Nussinov, R. Correlated mutations: Advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families. *Proteins: Structure, Function, and Bioinformatics* **2006**, *63*, 832-845.
50. Kellogg, G.E.; Fornabaio, M.; Chen, D.L.; Abraham, D.J. New application design for a 3D hydropathic map based search for potential water molecules bridging between protein and ligand. *Internet Electr. J. Mol. Design* **2005**, *4*, 194-209.
51. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.

CHAPTER 3

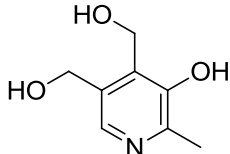
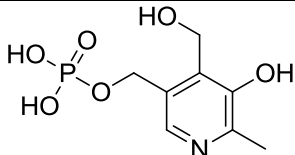
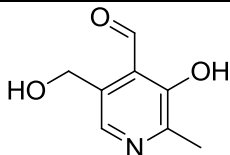
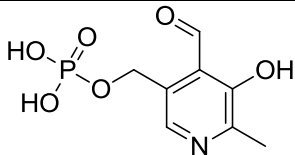
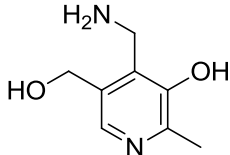
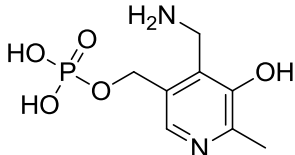
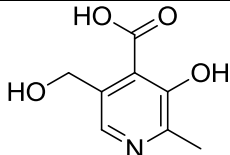
PYRIDOXAL KINASE – SERINE HYDROXYMETHYLTRANSFERASE COMPLEX MODEL

3.1 Introduction:

3.1.1 Different forms of vitamin B₆:

Vitamin B₆ has six different forms: pyridoxine (PN), pyridoxamine (PM), pyridoxal (PL) and their 5'-phosphorylated forms (PNP, PMP and PLP, respectively) (Table 3.1). PLP is the biologically active and arguably, the most important form of vitamin B₆ in nature, as it is used as enzyme cofactor by several B₆ enzymes (PLP-dependent enzymes) that include oxidoreductases, transferases, isomerases, lyases and hydrolases, with functions of many others unknown. PLP mainly functions as an electron sink during catalysis. It is estimated that there are more than 140 different enzymes utilizing PLP as a cofactor, a number that amounts to approximately 4% of all known catalytic activities [1].

Table 3.1. Different forms of vitamin B₆

Name	Structure	Notes
Pyridoxine (PN)		This is the form that is given as a vitamin B ₆ supplement
Pyridoxine 5'-phosphate (PNP)		
Pyridoxal (PL)		
Pyridoxal 5'-phosphate (PLP)		Metabolically active form
Pyridoxamine (PM)		
Pyridoxamine Phosphate (PMP)		
4-Pyridoxic acid (PA)		Metabolite excreted in the urine

3.1.2 Toxic effects of high concentrations of Pyridoxal 5'-phosphate (PLP):

Very high levels of vitamin B₆ in its active co-factor form PLP may have toxic effects [2-9]. This is due to the fact that PLP contains a very reactive aldehyde group at the 4' position, which easily forms aldimines with primary and secondary amines, and for this reason is often used as a protein labelling agent. Toxicities resulting from overconsumption of vitamin B₆ are well documented in the literature. Schaumburg *et al.* reported six different cases of vitamin B₆ overdose. Each of these cases experienced neurotoxicity with slightly different manifestations. Progressive sensory ataxia and profound distal limb impairment of position and vibration sense were commonly found between these cases. All the symptoms were reversed upon discontinuation of vitamin B₆ [10]. The exact mechanism of neurotoxicity is unclear, however in a study done on rats by Perry *et al.*, found that this neuropathy is characterized by necrosis of dorsal root ganglion sensory neurons and degeneration of the peripheral and central sensory projections [5]. High doses of PLP were also found to cause tonic-clonic convulsions in mice by Ishioka *et al.*[11]. Vermeersch *et al.*, also found that PLP inhibits DNA topoisomerase IB interfering with the process of winding and unwinding of DNA, that in turn impacts protein synthesis [12]. Bartzatt and Bechmann discovered that phenol sulfotransferase's ability to process phenols and other toxic substances is inhibited by PLP [8].

Due to the toxicity of PLP, the cell manages to keep the free PLP concentration very low. This fact leads to an important question which is how, despite the very low PLP concentration, do all the 140 PLP-dependent enzymes get sufficient amounts of PLP for their normal functioning? This chapter focuses down on developing a model for the protein-protein interaction and channel formation between PLK and SHMT, which may serve as a possible explanation for the safe PLP transportation between these two proteins avoiding its harmful effects to the cell. In addition, the identification and analysis of the interfacial water molecules and their relevance to this protein-protein complex is carried out. This model might be useful for guiding site directed mutagenesis. Furthermore, this model might also be useful in developing inhibitors for that protein-protein interaction providing a new drug target for cancer chemotherapy.

3.2 Materials and Methods:

3.2.1 Preparation of crystal structures:

For the PLK-SHMT protein-protein complex model, the crystal structures of pyridoxal kinase (PLK) and serine hydroxymethyltransferase (SHMT) were retrieved from the PDB database (PDB code 2DDM and 1DFO respectively) [13,14]. All ligands and non-protein atoms were removed including water. Then, using Sybyl 8.1, hydrogen atoms were added and minimized (Tripos forcefield, Gasteiger-Hückel charges, distance-dependent dielectric) to a gradient of $0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ [15]. To determine which

residues to be used as constraints for the protein-protein docking process, three computational methods were used for their determination.

3.2.2 Predicting the active residues for the protein-protein complex:

First, the tunnels whereby the PLP moves through from PLK to SHMT was calculated using the CAVER algorithm. CAVER was developed using Dijkstra's algorithm as a plug-in for PyMol [16]. Then, the Adaptive Poisson-Boltzmann Solver (APBS) software package was used to calculate a solvent surface map, colored by the electrostatic potential that was viewed in PyMol for each structure [17]. This proved to be useful in determining the compatibility of interacting surfaces of both proteins.

3.2.3 SASA calculation:

The last step in determining interface residues was through calculating the solvent accessible surface area (SASA) of the residues forming the tunnel, calculated by CAVER. Each sidechain solvent accessible surface area was calculated using GETAREA program developed by Robert Fraczekiewicz and Werner Braun [18]. The radius of water probe used was 1.4 Å. Residues with high sidechain SASA, were used to provide constraints for the protein-protein docking process.

3.2.4 Protein-protein docking:

Protein-protein docking was performed using HADDOCK (High Ambiguity Driven DOCKing) algorithm, which consists of two main stages: In the first stage, HADDOCK randomly orients the two proteins and performs a rigid body energy minimization with rotation and translation of each molecule. In the last stage, the top solutions resulting from the preceding energy minimization are then refined with three steps of simulated annealing refinements. In the first step the orientation of the proteins are optimized; the second step enhances the configuration of side chains at the interface; and the third permits some conformational rearrangements, where both backbone and side chains are allowed to move. The default parameters supplied by HADDOCK were used. Residues deduced from the previous steps were used to guide the docking procedure. The resulting structures were analyzed and ranked according to their average interaction energies (sum of E_{elec} , E_{vdw} , E_{ACS}) and their average buried surface area. Then these structures were clustered according to their pairwise backbone RMSD at the interface [19]. Docking results were individually inspected after which good models were submitted for refinement using FireDock.

3.2.5 Candidate model refinement:

For the refinement of the docked structures the Fast Interaction Refinement in Molecular Docking (FireDock) algorithm was used. The FireDock refinement process consisted of three main steps: (1) rearrangement of the interface side chains; (2) adjustment of the relative orientation of the molecules; and (3) scoring and ranking, which attempts to identify the near-native refined solutions. The ranking produces a score that includes a variety of energy terms including desolvation energy (atomic contact energy, ACE), Van der Waals interactions, partial electrostatics, hydrogen and disulfide bonds, π -stacking and aliphatic interactions, rotamer probabilities, etc. This binding score for the candidates ranking is an approximation of the binding-free energy function [20,21]. After that, candidate models were minimized (Tripos forcefield, Gasteiger-Hückel charges, distance-dependent dielectric) to a gradient of $0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. HINT was the main tool used to evaluate and choose the best models. Explicit water molecules were added in Sybyl 8.1, followed by minimization (Tripos forcefield, Gasteiger-Hückel charges, distance-dependent dielectric) to a gradient of $0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ and, finally evaluation by the HINT solvent accounting score [15, 22-27].

3.2.6 Hotspot prediction:

Finally, for the prediction of hotspots two different algorithms were used. Hot spots are defined as the residues for which the change in free energy of the complex is increased by 2 kcal mol⁻¹ when it is mutated to alanine. The first algorithm, HSPred, developed by Lise *et al.*, predicts hot spots based on Support Vector Machines (SVM) and on calculated energy potentials. It relies on the structure of the complex as input [28]. The second algorithm, Hotpoint, predicts hot spots based on conservation, solvent accessible surface area (SASA) and statistical pairwise residue potentials of the interface residues [29] followed by a BLAST search was done to see if predicted hotspot residues are conserved across species.

3.3 Results and Discussion:

3.3.1 Predicted PLK and SHMT tunnels:

The CAVER algorithm [16] provides rapid, accurate and fully automated calculation of channels (tunnels) in static structures. When given a starting point typically located inside the molecule, the algorithm searches for the easiest path from that point to the surface of the molecule. The algorithm automatically explores a grid constructed over the molecule. Nodes are evaluated using a cost function that determines the amount of free space around the node. The grid search algorithm is used to find the lowest-cost centerline path between a given starting point and the exterior of the molecule. The

identified path resembles a tunnel that connects protein residues in pockets or cavities with the surrounding bulk solvent. The tunnel characteristics, e.g. length, mean radius and gorge radius are determined and can be further analyzed (Figure 3.1). The tunnel gorge radius r_{gorge} is one of the most important tunnel characteristics because the tunnel gorge can form a bottleneck for substrate access or product release to and from the active site of a protein [16].

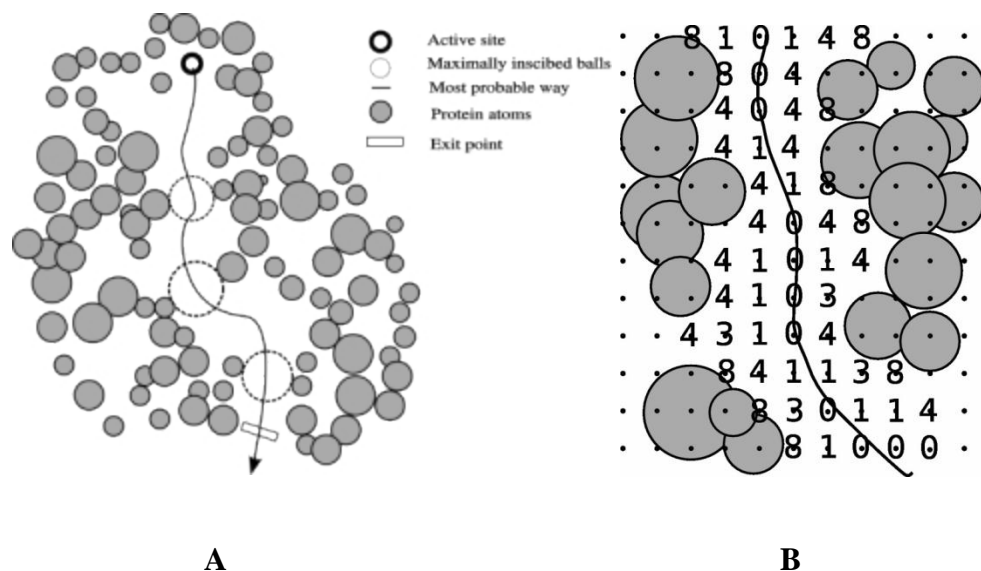


Figure 3.1. *Sketch of the computational algorithm implemented in CAVER:* (A)

The black bold circle represents the starting point. The protein is visualized by gray circles with Van der Waals atom radii mapped on a discrete grid (black dots). The solid line represents the boundary between the protein interior and its surroundings. Empty circles represent the maximally inscribed balls on the probable route (dashed line). (B) Evaluation of grid nodes by a cost function. The line represents the optimal centerline (path) [16].

The CAVER algorithm identifies by default the three largest tunnels for each protein. Only one tunnel was selected to be analyzed. The selection of tunnels was mainly based on the width of the tunnel radius, the larger the better; however, the length and the straightness of each tunnel were also taken into consideration. Ser 23 and Lys 229 were

used as starting points for PLK and SHMT respectively. Ser23 is one of the residues that bind to PL in PLK, while Lys229 covalently binds to PLP in SHMT. Because both structures are in closed conformers, the average radii of the tunnels were in the range of 2-4 Å, which too narrow for the substrate to pass through. Nevertheless, since proteins are dynamic, the tunnels have to open for the substrate to be transported either via protein-protein interaction or from the bulk solvent. Although, several functional studies show these proteins can assume an open conformation, unfortunately, the open conformers have not been crystallized. These tunnels are shown in Figure 3.2. The residues forming the wall of the tunnels for each protein are shown in Table 3.2 along with their sidechains' SASA.

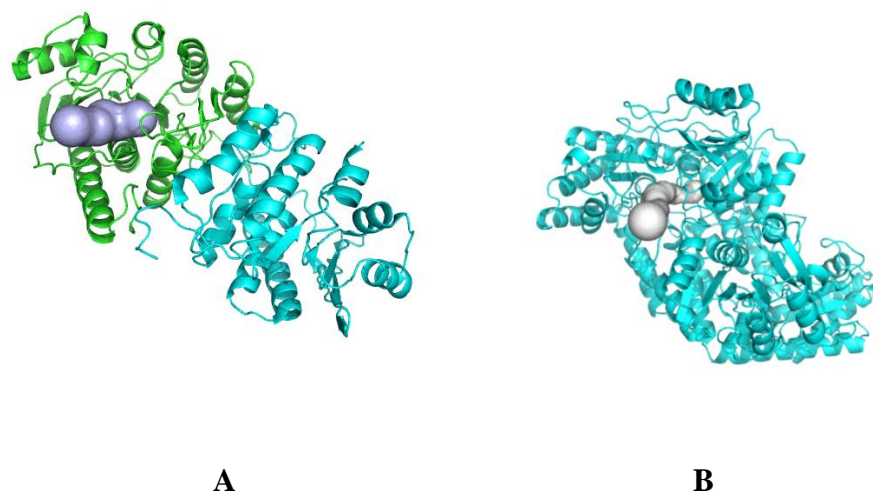


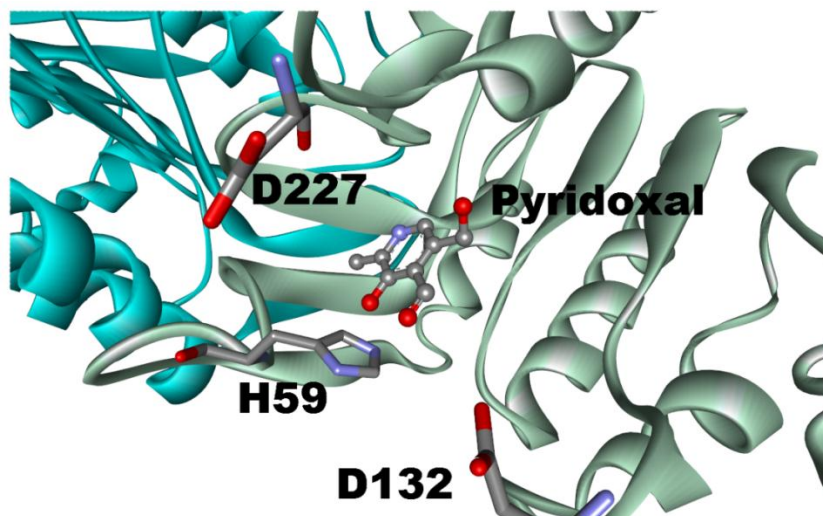
Figure 3.2. Predicted tunnels connecting PLP active site and the bulk solvent: (A) PLK and (B) SHMT.

Table 3.2: Residues forming the tunnel walls for PLK and SHMT

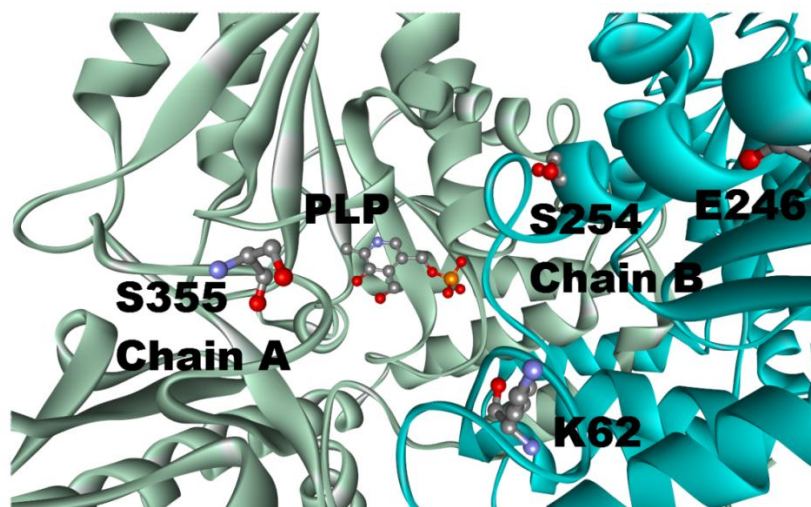
Rank	Residue Type	Residue Number	Chain	SASA
PLK	His	59	A	82.27
	Tyr	96		57.21
	Asp	130		53.63
	Ile	131		92.83
	Asp	132		68.70
	Tyr	136		99.70
	Thr	226		11.66
	Asp	227		115.53
	Leu	228		10.02
	Gly	230		0.00
	Thr	231		17.00
	Gly	232		0.00
	Asp	233		15.69
SHMT	His	126	A	43.40
	Leu	127		26.57
	Val	133		73.83
	Ser	175		19.01
	Asp	200		8.62
	Ala	202		11.68
	His	203		5.23
	Ser	355		53.79
	Pro	356		73.81
	Tyr	55	B	17.00
	Glu	57		15.34
	Lys	62		47.48
	Tyr	65		18.14
	Glu	246		75.12
	Ser	254		56.43
	Asn	347		20.61

3.3.2 Predicted active residues for PLK-SHMT complex:

For the active residue determination for PLK, four residues that are a part of the tunnel were found to be on the surface of the protein as indicated by their SASA. These residues, His59, Asp132, Tyr136 and Asp227 are shown in Figure 3.3a. Hence, if there is any protein-protein interaction that may occur for PLP channeling, those three residues are likely to be part of the contact residues. Because the active site of SHMT is formed by two monomers, the active residues Lys62, Glu246, Ser254 from chain A, and Ser355 from chain B shown in Figure 3.3b are part of the tunnel that are exposed on the surface of the protein. These active residues from both PLK and SHMT were used to guide the docking process.



A

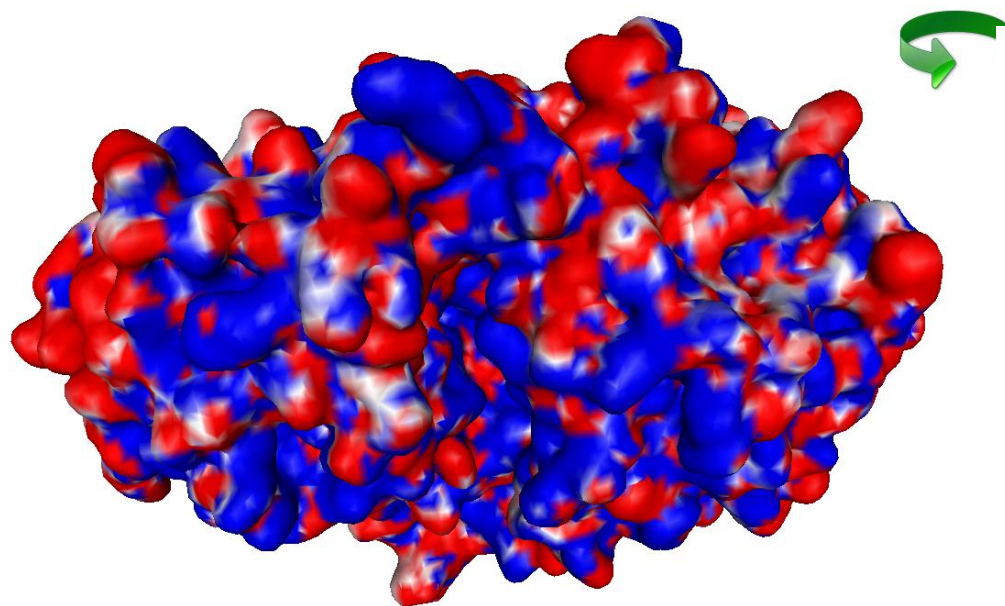
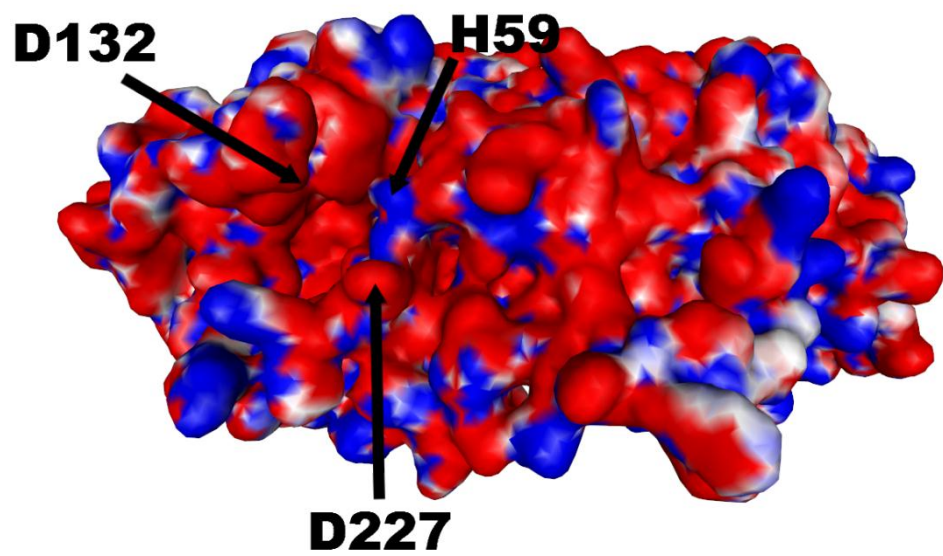


B

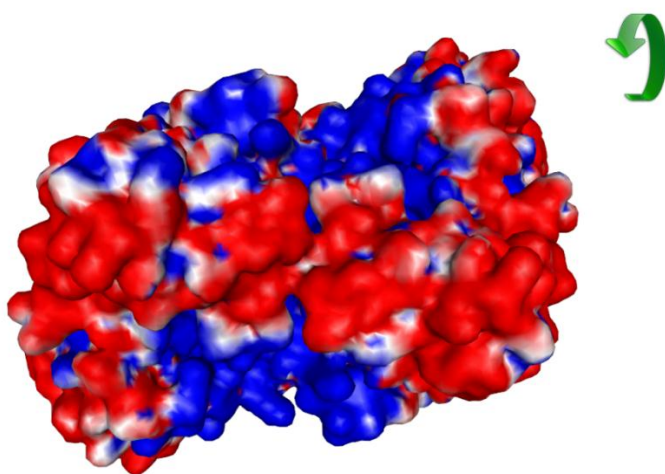
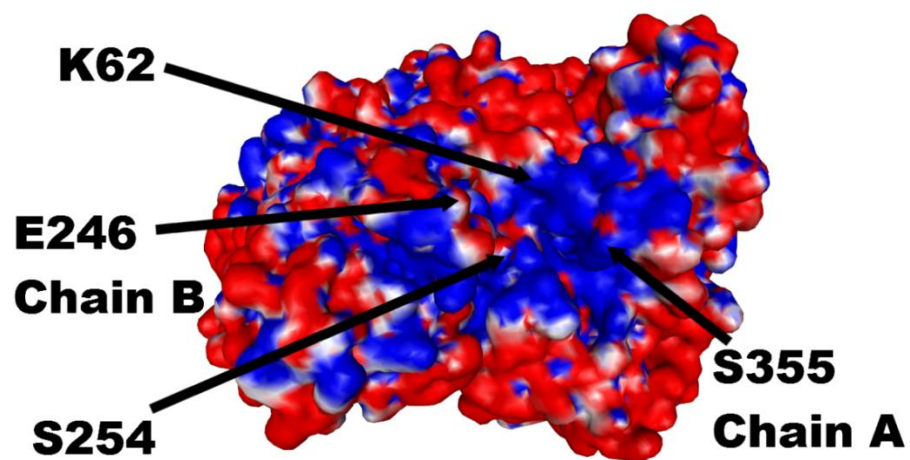
Figure 3.3. *Predicted active residues for PLK-SHMT complex:* (A) PLK and (B) SHMT.

The APBS calculation showed the compatibility of both proteins' surfaces in forming the protein-protein interface. Adaptive Poisson-Boltzmann Solver (APBS) is a program designed by Baker and co-workers [17] for modeling biomolecular solvation through solution of the Poisson-Boltzmann equation (PBE), which is a continuum model for describing electrostatic interactions between molecular solutes in aqueous media.

As shown in Figure 3.4a, PLK's interacting surface is mainly negatively charged, while SHMT's interacting surface is mainly positively charged as in Figure 3.4b. Closer examination of the rest of the surface of PLK revealed that the largest negatively charged surface is found at the opening of the tunnel to the two active sites of the dimer. Similarly, the largest positively charged surface of SHMT is found at the opening of the tunnel to the two active sites of the dimer (Figure 3.4).



A



B

Figure 3.4. *Electrostatic maps of both posterior and anterior sides: (A) PLK and (B) SHMT*

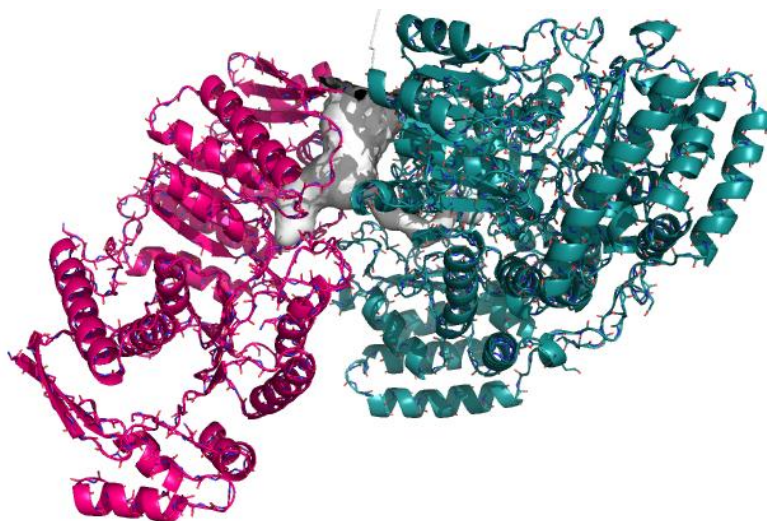
3.3.3 Protein-protein docking results:

After using the active residues to drive the docking procedure, the top 40 models were individually inspected and clustered into 10 groups according to their pairwise RMSDs. Their HADDOCK scores ranged from -18.4 to 52 (where the lower the score, the better). Of these 40, 8 models showed promising results with HADDOCK scores ranging from -18.4 to 15.6. These models were then submitted to the FireDock server for refinement. An additional round of minimization was carried out for these 8 models. Then to choose the best model, HINT scores were calculated for the 8 models as shown in Table 3.3. The model 7_2 appeared to have both the highest HINT score and the lowest global energy.

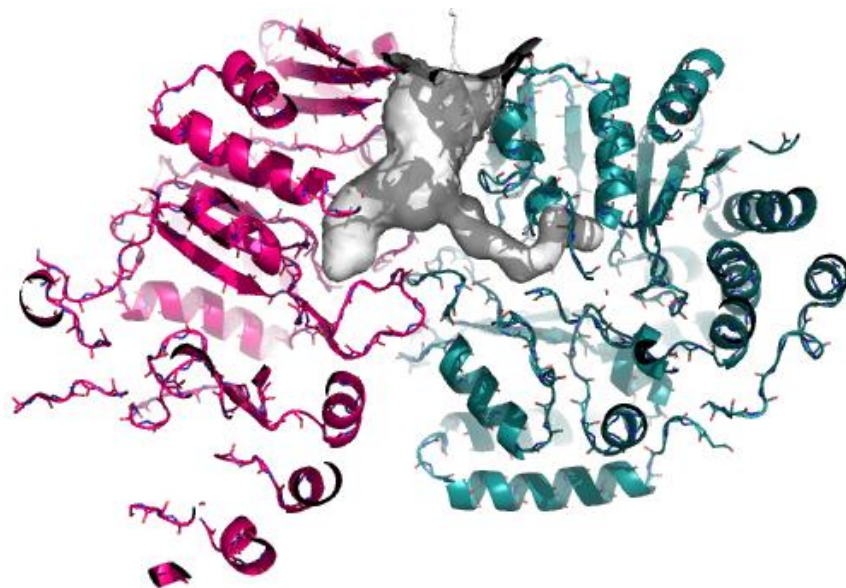
Table 3.3. Global Energy and HINT scores of 8 candidate models.

Rank	Model No	Global Energy	Attractive VdW	Repulsive VdW	ACE	HB	HINT score
1	7_2	-50.16	-32.89	12.98	2.2	-6.79	5.94e+03
2	7_1	-37.83	-26.88	9.13	-1.82	-1.58	3.32e+03
3	2_1	-31	-25.73	6.37	15.96	-4.45	4.70e+03
4	2_2	-15.5	-26.11	6.75	17.41	-4.45	3.41e+03
5	7_3	-5.26	-33	13.1	6.31	-2.95	4.83e+03
6	2_4	-2.03	-25.09	5.76	21.61	-3.06	3.84e+03
7	7_4	10.02	-29.96	13.18	14.67	-2.67	3.93e+03
8	2_3	21.62	-24.43	10.44	19.58	-2.33	5.55e+03

The CAVER algorithm was applied to model 7_2 to find the channel connecting the active sites of both PLK and SHMT and the result is shown in Figure 3.5.



A



B

Figure 3.5. *The predicted tunnel connecting PLK and SHMT:* (A) The PLK-SHMT complex in (B) parts of the complex were removed to reveal the tunnel.

3.3.4 HINT analysis and water relevance:

The calculated HINT scores for this complex are found in Table 3.4. It appears from HINT analysis that Glu 246, Lys 250, Glu 111, Lys 62, Asn 394, Glu 164, His 165, Lys 331, Ser 355 and Lys 251 from SHMT contributes the most to PLK-SHMT binding in the model as indicated by their high HINT scores. In a BLAST search, Lys 250, Lys 62, Ser 355 and Lys 251 appeared to be highly conserved across 500 different species. The Hot spot prediction algorithm predicted that SHMT Val 143 is a hot spot for this complex, which was also found to be highly conserved in the same BLAST search. For

PLK, Arg 12, Lys 225, Asp 273, Asp 61, Lys 10, Asp 227, Asn 170, Gln 204, Arg 172, Pro 58 and Glu 71 have the largest contribution to PLK-SHMT complex indicated by their HINT scores. Val 137 was identified as a hot spot and was found to be highly conserved in a non-redundant blast search in 500 different species.

Table 3.4. Calculated HINT scores for model 7_2.

PLK Name	SHMT Name	TOTAL Score	H-Bond Score	Acid/Base Score	Hydroph. Score	Acid/Acid Score	Base/Base Score	Hydr/Polar Score
Pro58	Ser355	246	314	5	16	0	-23	-65
Pro58	Val358	-19	0	0	7	0	0	-27
His59	Ser355	17	0	20	4	-1	0	-6
Tyr60	Lys331	13	0	13	0	0	0	0
Tyr60	Ser355	28	0	35	0	-1	-2	-5
Tyr60	Phe357	-21	0	0	1	0	-7	-16
Tyr60	Val358	-43	0	0	1	0	0	-44
Asp61	Lys62	738	676	140	9	0	-3	-84
Asp61	Lys331	239	238	21	2	0	0	-22
Asp61	Phe357	-80	0	1	25	0	-35	-71
Asp61	Val358	-78	0	0	4	0	-3	-80
Phe63	Glu332	-54	0	2	0	0	-55	-1
Phe63	Lys354	29	0	1	31	0	0	-4
Phe63	Ser355	31	0	23	9	0	0	-1
Phe63	Val358	59	0	0	87	0	-1	-28
Tyr64	Thr329	13	0	8	7	-2	0	-1
Tyr64	Glu332	-75	0	17	6	0	-18	-79
Tyr96	Pro353	19	0	42	1	0	-12	-12
Ile131	Asn120	19	0	26	0	0	-5	-3
Ile131	Ile142	-22	0	0	0	0	-4	-18

Ile131	Pro144	29	0	0	74	0	-2	-43
Asp132	Asn120	104	175	18	0	0	-81	-9
Asp132	Ala122	-12	0	0	0	0	0	-13
Asp132	Gly130	-22	0	1	0	0	-21	-2
Asp132	Pro132	-42	0	0	35	0	-1	-77
Asp132	Pro356	-15	0	0	1	0	0	-16
Ser133	Asn120	-118	49	24	9	-4	-139	-56
Ser133	Ala122	-27	0	3	51	0	-19	-62
Ser133	His123	-22	0	7	2	0	-2	-29
Ser133	Pro144	-10	0	0	1	0	-2	-10
Ser133	Asn351	-15	0	3	0	0	-17	-2
Gly134	Asn351	-58	0	9	5	0	-29	-42
Ile135	Asn351	-220	9	9	40	0	-111	-167
Ile135	Asp352	-79	0	2	0	0	-75	-6
Tyr136	Asp352	-38	0	1	3	0	-17	-25
Tyr136	Pro353	-43	0	2	11	0	-38	-18
Val137	Asn351	-20	0	1	1	0	-19	-3
Val137	Asp352	-18	0	1	0	0	-16	-3
Val137	Pro353	-59	0	1	12	0	-8	-64
Pro139	Val324	60	0	0	61	0	0	-2
Pro139	Asn351	-13	0	0	3	0	0	-16
Pro139	Pro353	31	0	0	31	0	0	0
Glu164	Glu164	-48	0	2	0	0	-49	-1
Glu164	His165	-44	0	6	0	0	-17	-34
Lys169	Glu164	-14	0	0	0	0	-12	-2
Asn170	Glu164	400	516	154	12	0	-120	-162
Asn170	His165	54	46	60	6	-25	-9	-24
Arg172	Leu117	-12	0	0	0	0	0	-12
Arg172	Val143	-111	0	6	4	0	0	-121
Arg172	Pro144	18	0	28	0	0	0	-10
Arg172	Gln161	58	0	75	0	-3	-1	-14
Arg172	Glu164	23	0	29	1	0	-1	-6
Arg172	His165	325	264	58	16	-9	3	-7
Gln204	Glu111	405	562	84	3	0	-189	-55
Lys225	Glu111	998	1171	2	8	0	-4	-178

Lys225	Glu247	-68	0	2	0	0	-65	-5
Lys225	Lys251	208	215	0	3	0	0	-10
Thr226	Glu247	12	0	25	5	0	-3	-15
Asp227	Lys250	717	672	133	10	0	-2	-95
Lys229	Lys250	-40	0	0	0	-38	0	-1
Gln268	Glu247	-30	0	2	1	0	-25	-8
Glu271	Glu246	-533	0	23	1	0	-537	-20
Asp273	Lys250	829	680	185	2	0	0	-38
Asp9	Lys83	21	0	21	0	0	0	0
Lys10	Glu246	734	919	2	10	0	-54	-143
Ser11	Glu246	-14	0	1	0	0	-15	-1
Arg12	Glu246	1579	1695	78	13	0	-26	-182
Glu71	Leu328	-213	0	0	9	0	0	-222
Glu71	Glu332	-130	0	8	3	0	-113	-28
Glu71	Ile393	-46	0	0	1	0	-18	-29
Glu71	Asn394	598	851	21	6	0	-156	-124
Glu82	Lys62	18	0	19	0	0	0	-1
Arg83	Lys62	-12	0	2	0	-14	0	0
Total HINT Score					5944			

Finally, a water layer was added on the interface of the PLK-SHMT complex, minimized and then evaluated using HINT. The total HINT score after addition of water was 7588. The water contribution is shown in Table 3.5.

Table 3.5. HINT water rank for the water molecules added to model 7_2.

HOH Name	O Atom	PLK			SHMT			TOTAL		Total Relevance	Relevance (≥ 0.25) w/ respect to:
		Rank	score	Relev	Rank	score	Relev	Rank	score		
HOH341	1	1.242	126.9	0.389	1.108	196.9	0.358	2.35	323.7	0.747	Both
HOH652	4	1.244	63.1	0.336	0.95	52.5	0.28	2.193	115.6	0.616	Both
HOH690	7	0	-42.2	-0.039	1.068	31.8	0.285	1.068	-10.3	0.246	SHMT
HOH726	10	0.898	-106	0.147	0	-197.2	-0.155	0.898	-303.2	-0.008	Neither
HOH731	13	1.169	32.8	0.3	1.305	390.8	0.564	2.475	423.6	0.864	Both
HOH770	16	0	-105.3	-0.04	1.15	422.3	0.497	1.15	317	0.457	SHMT
HOH789	19	2.595	35.9	0.445	1.213	147.2	0.392	3.808	183.1	0.837	Both
HOH797	22	1.263	34.3	0.31	1.362	124.5	0.409	2.625	158.8	0.719	Both
HOH899	25	2.254	105.5	0.474	1.107	-356.8	-0.578	3.362	-251.3	-0.104	PLK
HOH923	28	1.235	228.5	0.422	1.047	-16.4	0.245	2.282	212.1	0.667	Both
HOH925	31	1.15	534.2	0.534	0	-95.2	-0.04	1.15	439	0.494	PLK
HOH934	34	2.358	349.4	0.683	1.305	50.4	0.329	3.663	399.8	1.012	Both
HOH939	37	2.31	2.7	0.368	1.445	310.3	0.555	3.754	313	0.923	Both
HOH940	40	1.053	101.4	0.332	1.325	217	0.442	2.379	318.3	0.774	Both
HOH962	43	1.027	285.2	0.388	0.847	-13	0.213	1.874	272.2	0.601	Both
HOH963	46	2.214	441.8	0.729	0	-132.1	-0.044	2.214	309.7	0.685	Both
HOH982	49	0.99	245.7	0.355	1.199	78.1	0.343	2.189	323.8	0.698	Both
HOH1438	52	0.867	-130.1	-0.041	0.874	-27.7	0.207	1.741	-157.8	0.166	Neither
HOH1444	55	1.122	-128.4	-0.036	1.337	502.3	0.634	2.459	373.9	0.598	Both
HOH1445	58	0.877	99.8	0.3	0	75.3	-0.038	0.877	175.1	0.262	PLK
HOH1474	61	0	-349.5	-0.548	0.837	-35.7	0.194	0.837	-385.2	-0.354	Neither
HOH1491	64	1.125	-366.3	-0.603	2.711	171.5	0.616	3.836	-194.8	0.013	SHMT
HOH1497	67	2.531	-1.3	0.397	1.104	-151.3	-0.087	3.635	-152.6	0.31	PLK
HOH1498	70	0	-174.6	-0.106	1.968	103.5	0.435	1.968	-71.1	0.329	SHMT
HOH1500	73	1.135	-62.7	0.216	3.873	-18	0.56	5.008	-80.7	0.776	Both
HOH1533	76	0	-149.9	-0.065	0	-136.9	-0.048	0	-286.8	-0.113	Neither

HOH1534	79	0.815	-144.1	-0.071	0	-128.8	-0.041	0.815	-272.9	-0.112	Neither
HOH1539	82	0	-134.9	-0.046	0.838	-26.1	0.202	0.838	-161	0.156	Neither
HOH1540	85	1.141	5.3	0.274	0.855	-24	0.206	1.996	-18.7	0.48	PLK
HOH1544	88	2.137	-91	0.215	1.341	-1.5	0.275	3.478	-92.5	0.49	SHMT
HOH1545	91	1.139	-157.8	-0.101	0.921	-47.2	0.199	2.06	-205	0.098	Neither
HOH1552	94	0	-119.5	-0.034	0.776	-59.7	0.164	0.776	-179.2	0.13	Neither
HOH1560	97	2.423	80.6	0.469	0	-68.7	-0.039	2.423	11.8	0.43	PLK
HOH1563	100	2.521	435.6	0.746	0	-231.2	-0.237	2.521	204.4	0.509	Both
HOH1565	103	0.965	84.8	0.305	1.286	144.2	0.408	2.251	229	0.713	Both
HOH1572	106	0.865	255.3	0.334	0.861	-129.8	-0.04	1.726	125.5	0.294	PLK
HOH1574	109	2.238	-94.9	0.233	1.463	190.7	0.462	3.701	95.9	0.695	Both
HOH1701	112	0.809	404.1	0.4	0.939	-205.1	-0.208	1.748	199.1	0.192	PLK
HOH1721	115	1.286	325.5	0.515	0	-153.3	-0.069	1.286	172.3	0.446	PLK

Looking closely at water molecules in Table 3.5 and comparing them to water molecules included in the original starting crystal structures, it was found that waters 726, 731, and 1444 correspond to waters 166, 31, and 141 respectively from SHMT crystal structure (PDB ID 1DFO). The model suggests that waters 31 and 141 from SHMT crystal structure (PDB ID 1DFO) become relevant to both proteins in the complex and surprisingly water 166 is relevant to neither under a part of a hydrophobic bubble (as discussed in chapter 2). The reason for this is that this water was trapped originally near hydrophobic residues Gly 130, Pro 132, Gly 137 and Ile 142 then in the complex this water became more trapped by the addition of Ile 131 from PLK to the hydrophobic environment around this water. For PLK, water 982 from Table 3.5 was found to correlate to water 368 in the crystal structure, which the model suggests is relevant to both.

A detailed analysis of Relevance 2 waters (Chapter 2) in this model shows how significant this type of water molecules can be to the interactions between the two proteins especially in relieving repulsive interactions between similarly charged residues. Table 3.6 shows the interaction scores between pairs of residues from the two proteins before and after the addition of water to the model. As can be seen these waters improved substantially the HINT scores with an average of 366 ($-0.7 \text{ kcal mol}^{-1}$). In the model water molecule 341 relieved the unfavorable interaction between Glu 271 from PLK and Glu 246 from SHMT as indicated by their HINT score which was -533 ($1.066 \text{ kcal mol}^{-1}$) before the addition of water molecule 341 and became -103 ($0.206 \text{ kcal mol}^{-1}$) after its addition which is considered negligible. Also in the model, water molecule 1565 greatly improved the interaction between Glu 71 from PLK and Glu 332 SHMT as indicated by the 635 ($-1.27 \text{ kcal mol}^{-1}$) increase in HINT score.

Table 3.6. The effect of water molecules relevant to both PLK and SHMT in model 7_2.

PLK residue	SHMT residue	HINT score before water added (ΔG in kcal mol ⁻¹)	Water ID	HINT score after water added (ΔG in kcal mol ⁻¹)	Hint score difference (ΔG difference in kcal mol ⁻¹)
GLU271	GLU246	-533 (1.066)	HOH341	-103 (0.206)	430 (-0.86)
ASN170	HIS165	54 (-0.108)	HOH652	199 (-0.398)	145 (-0.29)
LYS225	GLU111	998 (-1.996)	HOH731	1562 (-3.124)	564 (-1.128)
GLN204	GLU111	405 (-0.81)	HOH789	607 (-1.214)	202 (-4.04)
LYS225	LYS251	208 (-0.416)	HOH797	563 (-1.126)	355 (-0.71)
GLU108	LYS62	18 (-0.036)	HOH934	210 (-0.42)	192 (-0.384)
GLU108	LYS62	18 (-0.036)	HOH940	333 (-0.666)	333 (-0.666)
LYS10	GLU246	734 (-1.468)	HOH939	1138 (-2.276)	404 (-0.808)
ASP9	LYS83	21 (-0.042)	HOH982	423 (-0.846)	402 (-0.804)
GLU71	GLU332	-130 (0.26)	HOH1565	505 (-1.01)	635 (-1.27)

3.4 Conclusion:

Channel formation between PLK and PLP-dependent enzymes explains how PLP can be transferred safely without causing any damages in a manner that is sufficient for satisfying the demand for this cofactor by approximately 140 different enzymes. Here, a model for the protein-protein interaction of PLK and SHMT is proposed and might be utilized in further confirming the theory of channeling by site-directed mutagenesis. This model might also be utilized in developing of small molecule inhibitors for this protein-protein interaction useful in anti-cancer research.

REFERENCES

REFERENCES

1. Jr., G. F. C. *The Vitamins, Second Edition: Fundamental Aspects in Nutrition and Health* Academic Press: 1998; , pp. 618.
2. Chung, J. Y.; Choi, J. H.; Hwang, C. Y.; Youn, H. Y. Pyridoxine induced neuropathy by subcutaneous administration in dogs. *J. Vet. Sci.* **2008**, 9, 127-131.
3. Gdynia, H.; Müller, T.; Sperfeld, A.; Kühnlein, P.; Otto, M.; Kassubek, J.; Ludolph, A. C. Severe sensorimotor neuropathy after intake of highest dosages of vitamin B₆. *Neuromuscular Disorders* **2008**, 18, 156-158.
4. Scott, K.; Zeris, S.;Kothari, M.J. Elevated B₆ levels and peripheral neuropathies. *Electromyography and clinical neurophysiology* **2008**, 48, 219-223.
5. Perry, T. A.; Weerasuriya, A.; Mouton, P. R.; Holloway, H. W.; Greig, N. H. Pyridoxine-induced toxicity in rats: a stereological quantification of the sensory neuropathy. *Exp. Neurol.* **2004**, 190, 133-144.
6. Salazar, P.; Tapia, R. Seizures induced by intracerebral administration of pyridoxal-5'-phosphate: effect of GABAergic drugs and glutamate receptor antagonists. *Neuropharmacology* **2001**, 41, 546-553.

7. Albin, R. L.; Albers, J. W.; Greenberg, H. S.; Townsend, J. B.; Lynn, R. B.; Burke, J. M., Jr; Alessi, A. G. Acute sensory neuropathy-neuronopathy from pyridoxine overdose. *Neurology* **1987**, *37*, 1729-1732.

8. Bartzatt, R.; Beckmann, J. D. Inhibition of phenol sulfotransferase by pyridoxal phosphate. *Biochem. Pharmacol.* **1994**, *47*, 2087-2095.

9. Schaeffer, M. C. Excess dietary vitamin B-6 alters startle behavior of rats. *J. Nutr.* **1993**, *123*, 1444-1452.

10. Schaumburg, H.; Kaplan, J.; Windebank, A.; Vick, N.; Rasmus, S.; Pleasure, D.; Brown, M. J. Sensory neuropathy from pyridoxine abuse. A new megavitamin syndrome. *N. Engl. J. Med.* **1983**, *309*, 445-448.

11. Ishioka, N.; Sato, J.; Nakamura, J.; Ohkubo, T.; Takeda, A.; Kurioka, S. In vivo modification of GABA_A receptor with a high dose of pyridoxal phosphate induces tonic-clonic convulsion in immature mice. *Neurochem. Int.* **1995**, *26*, 369-373.

12. Vermeersch, J. J.; Christmann-Franck, S.; Karabashyan, L. V.; Femandjian, S.; Mirambeau, G.; Der Garabedian, P. A. Pyridoxal 5'-phosphate inactivates DNA topoisomerase IB by modifying the lysine general acid. *Nucleic Acids Res.* **2004**, *32*, 5649-5657.

13. Safo, M. K.; Musayev, F. N.; di Salvo, M. L.; Hunt, S.; Claude, J. B.; Schirch, V. Crystal structure of pyridoxal kinase from the *Escherichia coli* pdxK gene: implications for the classification of pyridoxal kinases. *J. Bacteriol.* **2006**, *188*, 4542-4552.

14. Scarsdale, J. N.; Radaev, S.; Kazanina, G.; Schirch, V.; Wright, H. T. Crystal structure at 2.4 Å resolution of *E. coli* serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. *J. Mol. Biol.* **2000**, *296*, 155-168.
15. Tripos, L.P. www.tripos.com. St. Louis, MO, USA.
16. Beneš, P.; Chovancová, E.; Kozlíková, B.; Pavelka, A.; Strnad, O.; Brezovský, J.; Šustr, V.; Klvaňa, M.; Szabó, T.; Gora, A.; Zamborský, M.; Biedermannová, L.; Medek, P.; Damborský, J.; Sochor, J. CAVER 2.1, software, 2010.
17. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037-10041.
18. Fraczkiewicz, R.; Braun, W.; Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecule. *J. Comp. Chem.* **1998**, *19*, 319-333.
19. de Vries, S. J.; van Dijk, M.; Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **2010**, *5*, 883-897.
20. Andrusier, N.; Nussinov, R.; Wolfson, H. J. FireDock: Fast interaction refinement in molecular docking. *Proteins: Structure, Function, and Bioinformatics* **2007**, *69*, 139-159.

21. 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 Res.* **2008**, *36*, W229-W232.
22. Spyrakis, F.; Cozzini, P.; Bertoli, C.; Marabotti, A.; Kellogg, G. E.; Mozzarelli, A. Energetics of the protein-DNA-water interaction. *BMC Struct. Biol.* **2007**, *7*, 4.
23. Marabotti, A.; Spyrakis, F.; Facchiano, A.; Cozzini, P.; Alberti, S.; Kellogg, G. E.; Mozzarelli, A. Energy-based prediction of amino acid-nucleotide base recognition. *Journal of Computational Chemistry* **2008**, *29*, 1955-1969.
24. Burnett, J. C.; Kellogg, G. E.; Abraham, D. J. Computational Methodology for Estimating Changes in Free Energies of Biomolecular Association upon Mutation. The Importance of Bound Water in Dimer-Tetramer Assembly for beta 37 Mutant Hemoglobins. *Biochemistry (N. Y.)* **2000**, *39*, 1622-1633.
25. Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the Energetics of Water-Protein and Water-Ligand Interactions with the “Natural” HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules. *J. Mol. Biol.* **2006**, *358*, 289-309.
26. Eugene Kellogg, G.; Abraham, D. J. Hydrophobicity: is LogP(o/w) more than the sum of its parts? *Eur. J. Med. Chem.* **2000**, *35*, 651-661.
27. Sarkar, A.; Kellogg, G. E. Hydrophobicity--shake flasks, protein folding and drug discovery. *Curr. Top. Med. Chem.* **2010**, *10*, 67-83.

28. Lise, S.; Archambeau, C.; Pontil, M.; Jones, D. T. Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. *BMC Bioinformatics* **2009**, *10*, 365.
29. Tuncbag, N.; Gursoy, A.; Keskin, O. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics* **2009**, *25*, 1513-1520.

CHAPTER 4

CONCLUSIONS

The first aim of this work was to provide a detailed analysis of water molecules at protein-protein interfaces as well as quantifying their contributions with respect to different residue types. So, a data set of 4741 water molecules abstracted from 179 high-resolution (≤ 2.30 Å) X-ray crystal structures of protein-protein complexes was analyzed with a suite of modeling tools based on the HINT forcefield and hydrogen-bonding geometry. A metric termed Relevance was used to classify the roles of the water molecules.

Water molecules were found to be involved in: a) (bridging) interactions with both proteins (21%), b) favorable interactions with only one protein (53%), and c) no interactions with either protein (26%). This trend is shown to be independent of the crystallographic resolution. Interactions with residue backbones are consistent across all classes and account for 21.5% of all interactions. Interactions with polar residues are significantly more common for the first group and interactions with non-polar residues dominate the last group. Waters interacting with both proteins stabilize on average the

proteins' interaction by $-0.46 \text{ kcal mol}^{-1}$, but the overall average contribution of a single water to the protein-protein interaction energy is negligible ($+0.03 \text{ kcal mol}^{-1}$).

This research could be continued in various future directions. Although the work in this thesis answers many questions, it leads one to ask several new questions. While the role of bridging waters has already been established, the roles of water molecules that have favorable interactions with only one protein and water molecules with no interactions with either protein are not yet fully understood and require further investigation. Also, a more deeper investigation of the energetic role of the water molecules at protein-protein interfaces is important. In addition, the information extracted from this analysis could be employed in developing an algorithm to incorporate water molecules in the process of protein-protein docking. Another future direction is to use this information to find ways to exploit water molecules at protein-protein in developing small molecule inhibitors to these complexes, which is of a great advantage to medicinal chemistry and drug discovery.

The second aim was to observe the effect of adding interfacial water molecules in developing a model for the protein-protein interaction between pyridoxal kinase and serine hydroxymethyltransferase. This model was also created to explore the possibility of the formation of a channel between the two proteins upon interaction providing a safe way to transport the substrate pyridoxal 5'-phosphate. The crystal structures of the two

proteins were docked together and the results were refined and ranked according to their HINT scores. The highest ranking model was used to construct a channel using CAVER. Waters were then added on the interface of this PLK-SHMT model and evaluated using HINT's Rank algorithm.

The model showed that it is possible for a channel connecting the two active sites of pyridoxal kinase and serine hydroxymethyltransferase to be formed upon the interaction of the two proteins. The model showed favorable interactions formed between pyridoxal kinase and serine hydroxymethyltransferase as reflected by HINT analysis. The model had a HINT score of 5944, which was improved to 7588 upon adding and optimizing interfacial water molecules.

Although this model is not proven to be correct, it will be useful in guiding site-directed mutagenesis, which is the next step. Site-directed mutagenesis combined with kinetic studies could be carried out to a) the residues on the interface of the two proteins make sure the two proteins bind and b) to the residues surrounding the channel proposed in this model to observe whether the PLP transfer is affected. If this model is validated, it can be used to design small molecular inhibitors of this protein-protein interaction, which will be potentially useful as anti-cancer drugs.

Overall, this analysis produced considerable information that helps to deepen our understanding of the ever-growing field of biomolecular interactions. Although

computational analysis might not provide meticulously accurate information, it provides a good approximation to the reality of these interactions. It has always been a guide for researchers in their quest of uncovering scientific discoveries in the fields of medicinal and biological chemistry among various fields.

VITA

Mostafa Hassan Mohamed was born on August 23, 1984, in Kuwait, and is an Egyptian citizen. He received his Bachelors of Science in Pharmaceutical Sciences from the School of Pharmacy, Misr International University in Cairo, Egypt in 2006. Due to his outstanding performance, he was awarded a scholarship for excellence each year for five consecutive times during the whole program from the university to support his undergraduate study. Subsequently, he worked as a teaching assistant in the Pharmaceutical Chemistry Department, School of Pharmacy, Misr International University in Cairo, Egypt for three years. There he taught practical (Laboratory) Courses of Organic Chemistry, Analytical Chemistry and Instrumental analysis. In addition, he worked as an academic advisor in the Student Advising Center. In 2009, he was awarded a Fulbright scholarship for a Master of Science degree at Virginia Commonwealth University.