THE STRUCTURAL BASIS OF ELETRIPTAN AND SUMATRIPTAN TRANSPORT BY P-GLYCOPROTEIN

by

LAURA A. WILT

(Under the Direction of Arthur G. Roberts)

ABSTRACT

P-glycoprotein (Pgp) is an efflux transporter that is a member of the ATP binding cassette (ABC) superfamily. The transporter is found on the apical surface of cells and hydrolyzes ATP to expel a diverse range of xenobiotics into the extracellular space. The ubiquitous expression of Pgp dramatically effects drug disposition of many therapeutics and overexpression in cancerous tumors leads to multidrug resistance. As a result, Pgp-mediated transport represents a formidable hurdle to drug development. Another complicating factor is that drug transport rates can be different between drugs of the same class. To better understand this phenomenon, we investigated the triptan drugs, eletriptan (ETT) and sumatriptan (STT), with Pgp because the transport rate of ETT is 5- to 22-fold higher than STT. Despite their structural similarity, competition experiments revealed that they occupy distinct sites on the transporter. Fluorescence experiments revealed that the drugs shift Pgp into different conformations, while NMR experiments uncovered well defined drug interactions with the transporter. With this experimental information, a simplified conformationally-driven transport model was proposed. However, to understand drug transport on a molecular level, the drug locations need to be pinpointed on the transporter. Therefore, a hybrid molecular docking software wrapper called HADDOCK-Vina was developed, which fuses

HADDOCK protocols with AutoDock Vina. With the wrapper, we demonstrated that we can accurately reproduce the ligand positions of three ligand-bound X-ray crystal structures using experimental data to drive docking, as well as correct parametrization of a paramagnetic spin label. We used experimental data to drive the docking of ETT and STT onto Pgp by HADDOCK-Vina. The docking reveals distinct interactions between the drugs and Pgp, but some ambiguity still exists in their positions. Our preliminary ETT- and STT-bound models reveal the structural basis of triptan transport and demonstrates a rapid method to screen Pgp substrates for drug discovery. Future studies include incorporating a paramagnetic label onto Pgp to pinpoint drug binding sites and modeling these sites by HADDOCK-Vina.

INDEX WORDS: P-glycoprotein, ABC transporter, NMR spectroscopy, fluorescence spectroscopy, molecular docking, molecular dynamics, data-driven, HADDOCK-Vina

THE STRUCTURAL BASIS OF ELETRIPTAN AND SUMATRIPTAN TRANSPORT BY P-GLYCOPROTEIN

by

LAURA A. WILT

B.S., Winthrop University, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

THE STRUCTURAL BASIS OF ELETRIPTAN AND SUMATRIPTAN TRANSPORT BY P-GLYCOPROTEIN

by

LAURA A. WILT

Major Professor:	Arthur G. Roberts
Committee:	Shelley Hooks
	Scott Pegan
	Jeffrey Urbauer
	Jason Zastre

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

DEDICATION

I would like to dedicate my dissertation to my best friend and fiancé, Dennis J. Partyka Jr. Your constant love and support has never gone unnoticed. I could not have done this without you.

ACKNOWLEDGEMENTS

"I don't have to get a plaque. I don't have to get awards. I just walk out the door and all the girls will applaud. All the girls will commend as long as they understand that I'm fighting for the girls that never thought they could win because before they could begin, you told them it was the end but I am here to reverse the curse that they live in." –Onika Maraj

To the love of my life, Dennis Partyka Jr. Words cannot even begin to express my gratitude and appreciation of you during these past 5 years. For the late nights driving back and forth from Rock Hill to Athens while we maintained a 2-year long distant relationship. For working from home so you could move to Athens. The late nights at the CCRC, my qualifying exams, listening to me practice presentations, explaining my research to you, the tears and frustration of lab and the joy of my first publication, thank you. Thank you for your patience, your guidance, your unconditional love and your endless support. Without you, this would not be possible.

To my loving parents, Daniel and Lisa Wilt. Mom, thank you for the nonnegotiable homework time, the advanced school workbooks over the summer, forcing the summer book list, quizzing me with flashcards, helping me with my science fair project in 6th grade, making sure I got to my college exams on time and fostering my curiosity, even if it leads to opening doors in a moving vehicle, climbing Home Depot ladders or collecting rocks. Thank you for always being my cheerleader. Dad, thank you for teaching me the value of hard work, the value of playing a musical instrument and, most importantly, the value of an education. I will never forget the day

you told me to switch my major to chemistry or the night you told me you were glad that I no longer wanted to go to pharmacy school because my talents lie beyond that. Your encouragement and interest in science helped me reach and achieve my goals.

To my oldest brother, Michael Wilt. Your excitement and enthusiasm of math and science fills my heart with joy. I will never forget the day you asked me to explain my research. You were the first family member to ask me about my dissertation work and it absolutely thrilled me. Thank you for being my on-call math advisor, and most importantly, thank you for your love and support.

To my brother, Danny Wilt. You are the inspiration that thrived this scientific adventure. Thank you for your deep discussions on antiepileptic medications, improved therapies and general curiosity in my research. You always showed unending love, support and knew how to spark my curiosity. And most importantly, P'u.

To my best friend and sister, Leah Wilt. Thank you for bringing a smile to my face. Your visits to Athens uplifted my heart and helped me on my worst days to keep pursuing this degree. You never minded coming to lab with me on those weekends and you were a kid in a candy shop, taking pictures and saying it reminded you of potions class in Harry Potter. Thank you for your love, support and laughs. And remember, its NMR not edamame.

To my sister-in-law, Nichole Wilt. Thank you for your text messages and phone calls to check up on me. They truly lifted my heart and helped me through tough days. I always look forward to spending time together on our Christmas and summer vacations, which was always desperately needed. Thank you for your love and support.

To my friends, Hunter Connell, Leah Costyn, Courtney Daczkowski, Morgan Gibbs, Samantha Hord, Sudeepti Kuppa, Haley Roberson and Kristy Zera. Thank you for your laughs,

listening to my rants, your hugs, the nights out, the football games and, most importantly, your support. I could not do this without you all.

To my undergraduate advisor, Jason Hurlbert. Thank you for believing in me so many years ago and laying the groundwork for my scientific career. Your advisement never ended at Winthrop.

To my PhD advisor, Audie Roberts. Thank you for your constant support and patience these past five years. There was never a dull moment and I thank you for all the opportunities you have given me during my graduate career.

To my committee members. Thank you for your advice, your encouragement and your valued input. The impact you have made on my graduate career is priceless.

And lastly, I would like to acknowledge all the individuals who did not believe I could do this. To those who judged my intelligence by my looks. To the unending blonde jokes and the inexcusable sexual comments. This is *especially* for you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

This dissertation describes the research on the structural basis of eletriptan and sumatriptan transport by P-glycoprotein. The tables and figures are found at the end of each corresponding chapter. Chapter 1 contains the introduction and literature review. The literature review describes the expression, structure and function of the multidrug resistant transporter P-glycoprotein. Chapter 2 is the manuscript "Insights into the molecular mechanism of triptan transport by P-glycoprotein" published in the *Journal of Pharmaceutical Sciences*. Chapter 3 is the manuscript "HADDOCK-Vina: a hybrid approach for experimentally-driven receptor-small molecule docking" submitted in the journal *PROTEINS: structure, function and bioinformatics*. Chapter 4 is the manuscript "A NMR-derived model of the acetylsalicylic acid high affinity binding site on bovine serum albumin using the hybrid molecular docking wrapper HADDOCK-Vina" that is currently in preparation and will be submitted to the journal *PROTEINS: structure, function and bioinformatics*. The last chapter summarizes the findings of chapters 2-4 and discusses a preliminary docking model. This chapter also includes future work of this dissertation and a research outlook, which describes the larger impact of this research on the pharmaceutical hurdle of P-glycoprotein.

1.2 LITERATURE REVIEW

1.2.1 The functional role of P-glycoprotein

1.2.1.1 The expression of the ABC transporter P-glycoprotein

The ATP-binding cassette (ABC) superfamily of transporters is characterized by transmembrane helices (TM) spanning the membrane and two nucleotide-binding domains (NBDs) that bind and hydrolyze ATP. However, the structure of ABC transporters vary within in the superfamily by the number of TMs and overall fold.[1] Common to all family members, transport occurs by ATP hydrolysis and results in large conformational changes within the transmembrane domain (TMD).[2, 3] The ABC transporters are known efflux pumps, using ATP hydrolysis to drive transport against a concentration gradient. Many cancer cells overexpress ABC transporters as a primary mechanism of therapeutic resistance. To better understand this role in resistance, Victor Ling's group extracted a large glycoprotein from the membrane of colchicine-resistant Chinese hamster ovarian cells.[4] This multidrug resistant protein was identified to be a 170 kDa transmembrane protein and was called P-glycoprotein (Pgp).[4] After its discovery, Pgp was identified to be expressed ubiquitously in the body, as well as many mammalian cell lines used in *in vitro* assays.[5]

This multidrug resistant transporter is encoded by the *ABCB1* (*MDR1*) gene located on chromosome 7q21.12.[6, 7] *ABCB1* is large, comprising of 28 exons and the cDNA spanning 4.5 kb.[8] There are ~50 SNPs found on *ABCB1*, with only a few leading to amino acid changes; such as A61G, T307C, G1199A, G2677T and G2955A.[8] Of these variants, no changes in drug disposition were found. The structure and overall function of Pgp is fairly robust against site-directed mutagenesis in both the TMD and NBD.[9, 10] However, mutations in the catalytic binding site of ATP will lead to loss of ATP hydrolysis and thus, transport.[11] Interestingly, loss of function

mutations have not been identified on *ABCB1*. One polymorphism that shows some evidence in altering drug disposition is the C3435T point mutation on exon 26, but the result is controversial.[7, 8] This polymorph has a "silent" mutation that does not alter the amino acid sequence and it is found in ~26% of the Caucasian population, ~17% in the Asian population and only ~2.3% in the African population. [2,37] More profoundly, this polymorph has been associated with epileptic patients that have a poor response to anti-epileptic drugs.[12] Other studies find no correlation with the C3435T polymorph and drug disposition.[3,8] This apparent polymorphic multidrug resistance is most likely the result of polymorphism of a neighboring gene, thus altering the expression of Pgp or vice versa.[8]

Induction of Pgp has been shown to be regulated by the orphan nuclear steroid and xenobiotic receptor (SXR) and the constitutive androstane receptor (CAR).[14–17] Nuclear receptors bind small molecules to elicit large changes in expression of their gene target; thus, the nuclear receptors play a large role in gene regulation.[18] Being an orphan nuclear receptor, the endogenous ligand that binds CAR and SXR are unknown.[19] SXR and CAR are both highly promiscuous and bind a wide range of xenobiotics.[19] Interestingly, many of these compounds have overlapping specificity with Pgp, such as paclitaxel, rifampicin and phenobarbital.[15] SXR and CAR have been shown to upregulate Pgp, as well as cytochrome P450 isoforms 3A (CYP3A) and 2C (CYP2C) expression.[15, 19, 20] It is not surprising that SXR and CAR are highly expressed in the small intestine and liver, as these organs are subjected to orally ingested xenobiotics first.[19] SXR expression has also been observed in the kidney and in blood-tissue barriers like the placenta-blood barrier.[15] Co-localization of SXR, CAR and Pgp suggest an important role in xenobiotic disposition and overall protection of cells.[19] By increasing transporter expression, the xenobiotic detected by SXR and CAR, assuming that it is a transporter substrate, is expelled from the cells by Pgp, freeing up the

SXR and CAR ligand binding sites. This positive feedback loop drives the clearance of xenobiotics from cells and protects tissues from xenobiotics.[20]

1.2.1.2 The function of P-glycoprotein at absorption and excretion related tissues

Pgp is expressed at low concentrations ubiquitously in the body; however, high expression levels can be found in absorption related tissues (i.e. small intestines) and excretion related tissues (i.e. liver and kidney), as well as blood-tissue barriers.[8, 21, 22] Based on tissue distribution, the function of Pgp is to limit the cellular uptake of xenobiotics that may compromise homeostasis.

Upon oral administration of xenobiotics, most absorption occurs within the small intestines. The primary function of the small intestine is to absorb nutrients through passive diffusion or influx transporters.[23, 24] The differentiated microvilli found on the apical surface of the small intestines increase the surface area, thus increasing the absorptive capabilities of the small intestines.[25] Microvilli have a high expression level of transporters, acting as the first line of defense against xenobiotics.[23] Pgp expression increases through the progression of the gastrointestinal (GI) tract, with the stomach having the lowest and the colon having the highest expression levels.[23, 25, 26] The role of Pgp at the apical surface of small intestinal enterocytes is to expel compounds back into the intestinal lumen, limiting the absorption of xenobiotics (Figure 1.1A). Thus, Pgp plays a major role in limiting the bioavailability of many therapeutics. This role was first identified in human intestinal epithelial cell lines like Caco-2, where inhibiting Pgp lead to the increased accumulation of vinblastine and many other Pgp substrates.[24]

Following small intestine absorption, xenobiotics enter the portal vein and are then filtered by the liver. Here, xenobiotics passively diffuse or become transported into hepatocytes and are metabolized prior to entering the systemic circulation.[24, 25] The liver has the highest expression of

both phase I or phase II metabolizing enzymes. Many metabolizing enzymes share substrate specificity with ABC efflux transporters, creating an effective mechanism of xenobiotic clearance at the liver.[14, 24, 27] ABC transporters are primarily expressed on the bile canaliculus, including Pgp (Figure 1.1B). Pgp expels xenobiotics or the metabolized product into the bile duct where it will return to the intestinal lumen for excretion. The role of Pgp at the bile canaliculus is essential in the removal of many xenobiotics. Pgp acts as a "safety net" ensuring any xenobiotics that escape efflux at the small intestine will be excreted into the bile.[25]

Pgp expression has also been identified in the kidney. The kidneys play a central role in the filtering of blood and removing metabolized or unchanged xenobiotics through the production of urine. Pgp is highly expressed on the apical surface of the proximal tubule, where reabsorption occurs (Figure 1.1C).[25, 28] Here, Pgp-mediated transport expels xenobiotics back into the nephron for excretion in urine.[25] Pgp expression was also identified at loop of Henle and in the collecting ducts of the kidney.[29] *In vivo* studies demonstrated that *Mdr1a/Mdr1b*(-/-) mice have decreased renal clearance of vinblastine compared to wildtype mice, demonstrating Pgp's role in xenobiotic clearance.[30] Furthermore, *Mdr1a/Mdr1b*(-/-) mice exposed to arsenic had a 2-fold increase in arsenic accumulation, producing severe lesions in the kidney when compared to wildtype mice.[31] Inability to transport xenobiotics into urine for elimination can lead to accumulation of harmful toxins in the body and severely compromise homeostasis.[30, 31]

Overall, the role of Pgp at tissues involved in xenobiotic absorption and excretion is to prevent harmful compounds from entering the circulation system or aid in the removal of these compounds from the blood. The expression of Pgp at the GI tracts acts as the first barrier to orally administered xenobiotics by preventing the passive diffusion through the small and large intestines. The liver hosts a high expression of metabolizing enzymes to metabolize xenobiotics

and ABC transporters to transport compounds into bile. The expression of Pgp at the kidneys is found primarily at the proximal tubule, where xenobiotics or metabolized product are transported into urine. Pgp plays an important role in the removal of these xenobiotics, thus limiting the compounds that reach blood-tissue barriers.

1.2.1.3 The function of P-glycoprotein at blood-tissue barriers

While Pgp acts as the first line of defense at tissues involved in xenobiotic distribution, compounds can circumvent active transport and enter the systemic circulation. Pgp is found expressed on the apical surface of many blood-tissue barriers to expel compounds back into lumen or blood.[25] The blood-brain barrier (BBB) is a unique barrier composed of continuous endothelial cells with both tight and adherens junction proteins to protect neurons in the central nervous system (CNS).[32, 33] Because many xenobiotics and endogenous compounds can passively diffuse through the BBB and enter the CNS, Pgp is expressed on the apical surface to efflux compounds back into blood (Figure 1.1.D).[34, 35] Pgp's role is critical in excluding substances from the brain and acts as the gatekeeper to the BBB. [34, 36] *In vivo Mdr1a/b*(-/-) knockout mice are 3-100 fold more sensitive to neurotoxicity associated with Pgp substrates when compared to wildtype mice.[37, 38] The Pgp substrates used in this study were ivermectin, vinblastine, indinavir and loperamide.[37, 38] The exclusion of many drugs by Pgp protects the CNS and maintains neuronal homeostasis, which is critical for survival.

The blood-placenta barrier (BPB) is a blood-tissue barrier that protects the fetus from exposure to xenobiotics.[25] This barrier is critical in maintaining homeostasis in the placenta for healthy development of the fetus. The apical membrane of the syncytiotrophoblast is exposed to maternal blood with the basolateral membrane in contact with the cytotrophoblast and fetal blood

vessels.[25] Pgp is expressed at high levels on the apical surface of the brush border of the syncytiotrophoblast layer (Figure 1.1E).[28, 39] *In vivo*, *Mdr1a/b(-/-)* mice were treated with avermectin and produced offspring with a cleft palate. *Mdr1a/b(-/+)* mice showed some sensitivity to avermectin, while wildtype mice offspring showed no abnormality.[38] Additionally, treatment of pregnant mice with a Pgp inhibitor resulted in equal accumulation of digoxin and saquinavir in fetuses, despite their genotypes.[40] Without Pgp expression at the BPB, the development of the fetus can become severely compromised, leading to birth defects and abnormalities.

Blood-tissue barriers are present in sex organs to protect maturing gametes from the exposure of xenobiotics and other endogenous compounds. The blood-testis barrier (BTB) functions to protect maturing sperm cells and is characterized with continuous endothelial cells joined by tight junctions, similar to the BBB.[25] Pgp is highly expressed on the apical surface of the capillary endothelial cells of the testis (Figure 1.1F).[37, 41] Pgp was also identified to be expressed in Leydig cells, testicular macrophages, peritubular cells, Sertoli cells and haploid spermatozoa.[42] *In vivo*, Mdr1a*/b*(-/-) knockout mice had an increase in the accumulated Pgp substrates (i.e. vinblastine and ivermectin) within testicular tissue when compared to wildtype.[37, 41] Additionally, treating wildtype mice with the Pgp inhibitor nelfinavir, a known Pgp substrate, increased levels within the testes ~4-fold.[41] Pgp expression at the BTB suggests that the primary role of the transporter at this tissue barrier is to protect maturing gametes and ultimately, the health of future offspring.

Pgp plays an important role in the exclusion and protection of sensitive tissues like the brain, fetus and testis. Pgp protects maturing gametes in the testis from toxic insults for healthy offspring, as well as protect the developing fetus at the BPB from abnormalities or birth defects from xenobiotic exposure. Additionally, Pgp expression at the BBB is critical in preventing

neurotoxicity from xenobiotics and endogenous substrates, which can become life threatening. The *in vivo* studies demonstrated the severity of eliminating Pgp at these tissues and shows the role of Pgp in protection by excluding xenobiotics. However, Pgp prevents many therapeutic compounds from reaching their target and make them ineffective. To better understand Pgp-mediated transport and improve drug design and development, the atomic structures of Pgp were sought out.

1.2.2 The structure of P-glycoprotein

1.2.2.1 Atomic resolution of the mouse and human P-glycoprotein structure

The atomic structure of Pgp is key in determining the mechanism of substrate recognition, ATP hydrolysis and drug transport, to overcome disposition of therapeutic compounds. Mouse Pgp (*m*Pgp, *Mdr1a*) has an 87% sequence identity to human Pgp (*h*Pgp, *ABCB1*), making it an excellent model for probing the structural mechanism of Pgp-mediated transport. The structure of *m*Pgp has been solved by many groups.[43–46]The first *mPgp* structure was determined in the absence of ligand and ATP at 3.8 Å by X-ray crystallography (Figure 1.2A, PDB ID 4M1M).[43, 46] Structural analysis of the nonglycosylated *m*Pgp revealed an inward facing conformation. The large binding cavity was exposed to the lipid membrane with the NBDs separated (Figure 1.2A).[44–46] The binding cavity of mPgp was identified to have a volume of 6,000 $Å^3$. [43, 47] This large cavity supports Pgp's ability to bind multiple substrates at once[43, 48] and transport large substrates like amyloid-β.[49] The *m*Pgp transporter was also solved in the presence of nontherapeutic cyclic peptides QZ59-RRR and QZ59-SSS at 4.4 and 4.35 Å, respectively. Both peptides were found in the binding cavity with QZ59-SSS occupying two sites.[43, 46] Interestingly, there are no polar residues lining this cavity, unlike the bacterial lipid flippase MsbA and Pgp's homolog MDR3 (*ABCB4*).[46] The

hydrophobic nature of the *m*Pgp and *h*Pgp cavity suggests that the endogenous role of Pgp is a xenobiotic pump, not as a lipid flippase.

Until recently, no protein structure of *h*Pgp has been solved. Youngjin Kim and Jue Chen solved the first *h*Pgp structure in D-dodecyl-ß-maltoside (DDM) micelle by cryo-EM at 3.4 Å resolution (Figure 1.2B, PDB ID 6C0V).[50] To achieve this structure, *h*Pgp was mutated in both nucleotide binding domains (E556Q/E1201Q) to make a catalytically inactive transporter and was fused with a GFP-tag for expression and purification.[50] The structure was then solved in the presence of saturating ATP (10 mM) and vinblastine (150 μM), a known Pgp substrate, with efforts to trap the protein in a ligand-bound conformation. While no electron density was found for vinblastine, both NBD domains contained one ATP molecule,[50] suggesting this conformation of *h*Pgp was following substrate release and before ATP hydrolysis. This structure is the first Pgp structure solved in the closed-conformation and correlates ATP binding to conformational changes.

1.2.2.2 Lipid-protein interactions are critical in mediating P-glycoprotein transport

Transmembrane proteins are in direct contact with the lipid membrane, forming unique lipid-protein interactions. An electron paramagnetic resonance spectroscopy study identified lipids that interact with integral membrane proteins to be motionally restricted.[51] These membrane proteins present multiple grooves and clefts on the surface, resulting in lipids that are distorted to fill the space around the protein.[51] In order to form a static, lipid shell around a protein, the lipid composition varies for each protein. This intimate lipid-protein interaction plays a role in protein structure and function.

The lipid composition surrounding Pgp is a significant component in substrate binding and transport. Previous studies demonstrated that lipid compositions in membranes extracted around Pgp are enriched with phosphatidylethanolamine and phosphatidylserine[52] and this composition can influence Pgp-mediated ATPase activity. Low cholesterol levels in a cell membrane have been shown to decrease Pgp-mediated ATPase activity.[53, 54] This decrease in activity is a result of poor packing of the TMD and instability in the NBDs due to the increased membrane fluidity. While the NBDs are located in the cytosol of a cell, ATPase activity is not present unless Pgp is interacting with lipids,[37] suggesting that the lipid environment is critical for substrate permeability, binding and transport.

1.2.3 The mechanism of P-glycoprotein-mediated transport

1.2.3.1 Substrate binding and transport models of P-glycoprotein

Pgp sits in the membrane with the extracellular portion of the transporter protruding ~25 Å from the membrane and the NBD ~75 Å (Fig. 1.2). The orientation of the transporter provides direct contact of the lipid membrane with the large, hydrophobic cavity, suggesting xenobiotics enter the cavity through the plasma membrane. However, the exact mechanism of substrate binding onto Pgp has yet to be resolved. Two models currently exist, the hydrophobic vacuum cleaner and the substrate flippase models (Figure 1.3).[5, 52] The hydrophobic vacuum cleaner model proposes that xenobiotics would become "sucked" into the binding pocket. Upon binding and conformational rearrangement, the transporter would expel xenobiotics back into lumen or blood (Figure 1.3A).

The substrate flippase model was proposed due to the sequence similarity to MDR3, which acts as a lipid flippase. In the unbound state, Pgp has both NBDs far apart with the TMD having

access to the inner leaflet, as seen in the X-ray crystal structure of *m*Pgp.[44–46] The substrate flippase model suggests that xenobiotics enter the binding cavity of Pgp through the inner leaflet of the membrane (Figure 1.3B).[5, 52] Conformational rearrangement of the transporter results in the cavity having access to the outer leaflet. Substrates diffuse back into the outer leaflet membrane, thus continuing a concentration gradient within the membrane, not across the membrane like the hydrophobic vacuum model.[5, 52] Considering the *h*Pgp structure in the closed conformation, there is strong evidence that the hydrophobic vacuum cleaner transport model is correct. The lateral opening formed by helices 4,6 and 10,12 is exposed to the inner membrane leaflet in the unbound state (i.e. *m*Pgp structure[45]) and remains in the inner leaflet upon conformational rearrangement to the closed-state (i.e. *h*Pgp structure[50]). Continuous closure to the outer leaflet from the open to closed-states of Pgp suggests that compounds are expelled back into lumen, not the membrane.

1.2.3.2 Alternating site model of P-glycoprotein-mediated ATP hydrolysis

The exact mechanism of ATP-coupling to transport is unknown; however, biochemical assays suggest that Pgp-mediated transport occurs through the alternating site model. In this model, binding of both ATP molecules is required for conformational rearrangement and only one ATP molecule is hydrolyzed to return Pgp to the unbound conformation.[55–57] For this mechanism to occur, one ATP binding site is "tight", while the other is "loose" and the affinity for these sites switch as the transport cycle continues. Orthovanadate (*Vi*) trapping is a technique used to trap the transporter in an $ADP\text{-}Vi\text{-}Mg^{2+}$ complex because V_i will replace inorganic phosphate (P_i) in a non-reversible mechanism.[5, 56, 57] Trapping one NBD in the $ADP\text{-}Vi\text{-}Mg^{2+}$ complex abolished the activity of the other NBD, suggesting both NBDs must be active to couple transport.[5, 56, 57] Therefore, only one NBD is catalytically active while the other NBD is in the transition state.

Considering the ATP-bound *h*Pgp structure, it appears that the binding of both ATP molecules is required to form the NBD sandwich. Unfortunately, the transition state of one NBD was unable to be resolved because the *h*Pgp structure was catalytically inactive.

The catalytically inactive *m*Pgp (E552Q/E1197Q) was solved in the presence of saturating ATP (18 mM) at 3.35 Å resolution by X-ray crystallography.[45] Here, only one ATP molecule was bound in NBD1 in the absence of a Pgp substrate and remained in the open conformation. This structure demonstrates that the affinity of ATP at each NBD is different; however, it is unsure why both NBDs did not occupy ATP molecules like the *h*Pgp structure at saturating concentrations. In this *m*Pgp structure, the loop that connects both halves of the transporter was shortened to reduce the protein dynamics and thus, improve the crystallization and structure resolution. This loop may play a role in communicating the affinity switch between the NBDs. This lack of communication is demonstrated in the reduction of the relative ATPase activity of the transporter upon shortening of the loop, resulting in only one NBD binding ATP.[45] However, by shortening this loop, the asymmetric ATP bound conformation was shown for *m*Pgp and supports the asymmetric catalytic cycle model. One ATP molecule binds to the NBD "tightly" (i.e. *m*Pgp structure[45]) a second ATP molecule binds in the other NBD "loosely" and shift the transporter into a closed conformation, (*h*Pgp structure[50]) Whether one hydrolysis or two hydrolysis events bring the transporter back to the open conformation is still unknown and requires additional research to resolve.

1.2.3.3 Bridging ligand binding to ATP hydrolysis: crosstalk between the transmembrane and nucleotide binding domains

Substrate binding onto Pgp induces large conformational rearrangements that lead to ATP hydrolysis and transport. In the absence of substrate and presence of ATP, Pgp displays a basal

rate of ATP hydrolysis.[48, 58] Upon substrate binding, basal ATP hydrolysis can be maintained, stimulated or inhibited[48, 58, 59] and the rate of ATP hydrolysis is not coupled to Pgp-mediated transport. The well-known transported Pgp substrate Hoechst 33342 maintains basal ATPase activity,[60] while the inhibitor tariquidar binds in the upper binding cavity and "locks" the transporter in the closed-conformation, thus stimulating ATP hydrolysis.[59, 61] Therefore, conformational changes in the TMD must translate to the NBD to alter the basal rate of ATP hydrolysis. This crosstalk between the domains has been hypothesized to occur in the intracellular loops (ICLs) of the transporter.[11]

The ICLs are the small loops connecting the TMs to the NBDs and are structural features common to most ABC transporters. The location of each ICL loop is labeled in Figure 1.4. ICL1 and ICL4 link the transmembrane helices to NBD1 and ICL 2 and 3 to NBD2. Specific helices on the ICLs interact directly with the NBDs and are called intracellular helices (IHs) 1-4. These IHs have direct contact with ATP in the closed conformation (Figure 1.4B). Thus IHs within the ICLs are believed to act as "ball-and-socket" joints on ABC transporters to relay conformational changes in the TMD to the NBD.[11] Introducing single cysteine mutants into IH1 and IH3 and crosslinking these residues resulted in a constitutively active transporter;[3] however, cross-linking IH2 with IH3 abolished ATPase activity.[11] Molecular dynamics simulations of *h*Pgp revealed rotation of IH2 may be a key feature for the binding of ATP.[62] Additionally, Tyr1087 on NBD2 creates a crucial interaction with IH3 upon conformational rearrangement and ATP binding[62] and reduced Pgp-mediated ATPase activity when mutated to Phe.[63] Interestingly, mutating residues that interact with ATP did not result in such large inhibition of Pgp-mediated ATPase activity.[63] Thus, the IH2/IH3/NBD2 interaction was found to be critical for ATP binding and ATPase activity[11, 63] and an ideal target for inhibiting Pgp-mediated transport.

1.2.4 Hurdles and strategies of P-glycoprotein mediated transport

1.2.4.1 Shared physiochemical properties of P-glycoprotein substrates

In 1997, Charles Lipinski identified five physiochemical properties that were shared among many successful drugs.[64] The Rule of 5 predicts the pharmacokinetic properties of a drug, not necessarily its effectiveness; however, compounds following this rule of thumb are more successful in clinical trials.[65, 66] The Rule of 5 states that an orally active therapeutic has no more than one violation of the following criteria: at most 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, a molecular mass less than 500 Daltons and a log*P* no greater than 5.[64, 65] Because Pgp binds a wide range of drugs, many studies have sought to determine the physiochemical properties that lead to Pgp recognition and transport. Structure activity relationship (SAR) studies have identified shared physiochemical properties among Pgp substrates: lipophilicity, aromatic rings, high degree of hydrogen bonding atoms and positively charged tertiary amine at physiological pH.[47, 67] Based on Lipinski's Rule of 5, it is not surprising that many drugs are Pgp substrates. Drugs that cannot reach their target due to Pgp-mediated efflux have reduced efficacy and cannot elicit their therapeutic effect. These Pgp substrates include, but are not limited to, HIV protease inhibitors, opioids, anti-epileptics, antibiotics, Ca^{2+} channel blockers and anticancer therapeutics.[49, 68, 69]

1.2.4.2 Future aspects of drug design and development

The function of Pgp prevents many successful therapeutic compounds from reaching their targets. This is especially prevalent at the BBB, where many neurotherapeutic compounds cannot penetrate the barrier due to Pgp-mediated transport, resulting in ineffective therapies.[68, 70–72] One method proposed to circumvent Pgp-mediated efflux is by inhibiting the transporter.[32, 73] While

this method seems ideal in promoting drug permeability, inhibiting Pgp may not be feasible.[32, 73] Global inhibition of the transporter may lead to drug accumulation and drug-drug interactions, resulting in many adverse side effects.[48, 74, 75] For example, the concentration required to inhibit Pgp at the BBB would be low in orally administered therapeutics due to the first pass effect and would require a much higher dose, which could induce adverse effects.[73] To reverse multidrug resistance in cancer, clinicians have co-administered Pgp inhibitors with anti-cancer drugs. Using inhibitors like verapamil has been unsuccessful in the clinic due to these associated toxicities of Pgp inhibition.[76, 77] Inhibiting Pgp at the binding cavity may not be a useful strategy in increasing therapeutic permeability; however, methods to avoid Pgp-mediated transport may be successful in mediating multidrug resistance.[34]

Due to the mechanism of Pgp-mediated transport, therapeutics that passively diffuse through the plasma membrane enter the binding cavity through the lateral opening in the TMD. Drug compounds can enter cells through different mechanisms that avoid passive diffusion and thus, Pgp-mediated transport. Large and/or charged compounds can penetrate cells by two methods: adsorptive-mediated transcytosis and receptor-mediated transcytosis. The rate of adsorptive-mediated and receptor-mediated transcytosis increases in the disease state, which is advantageous in the treatment of disorders and diseases.[32] Another method to administer drugs are by nanoparticles targeting specific cell types. Nanoparticles can be designed to be nontoxic, biodegradable, and stable with various therapeutic releasing technologies.[32] However, this method is fairly new with only a few nanoparticles currently in clinical trial for cancer treatment.

Methods to increase therapeutic permeability in a disease state are necessary to improve the treatment and quality of life of patients. The safety, risk and benefit must be evaluated to determine the most effective method to treat multidrug resistance and unwarranted drug

disposition. While utilizing transcytosis or nanoparticles to administer therapies seems promising, the transporter expression pattern for each disease, as well as patient complacency must be considered.[33] Understanding the structural basis of substrate recognition by Pgp is critical in teasing out the mechanism of transport and thus designing compounds that circumvent Pgp-mediated transport or target the IH3/IH2/NBD2 interface. This approach would avoid pitfalls of previously established Pgp inhibitors of the TMD and utilize small molecules instead of using biotechnologies or biotherapeutics.

1.2.5 Concluding Remarks

The multidrug resistant transporter Pgp is expressed ubiquitously through the body, with higher expression levels at clearance tissues and blood-tissue barriers. The role of Pgp is to protect sensitive tissues by limiting the entry of xenobiotics into the systemic circulation. While this is critical for homeostasis, Pgp-mediated transport limits the permeability of many therapeutics. Thus, Pgp has become a major target in the pharmaceutical industry in efforts to overcome Pgp-mediated multidrug resistance and drug disposition. SAR studies of Pgp identified substrates to be lipophilic and aromatic with many hydrogen bond donors and acceptors. It is not surprising that many drugs are Pgp substrates, since the characteristics of a Pgp substrate overlap with Lipinski's Rule of 5. Due to the multidrug resistance of Pgp in cancer and inability to target tissues like the CNS, efforts have been made to inhibit Pgp-mediated transport. Clinical trials of co-administering Pgp inhibitors with chemotherapeutics resulted in adverse side effects because Pgp inhibition occurred globally. Understanding the protein-ligand interactions that drive transport will provide a better understanding of the mechanism of Pgp-mediated transport and the development of safer, orally administered inhibitors that target Pgp or drugs that circumvent Pgp. To probe this

mechanism, the structural basis of Pgp-mediated transport of eletriptan (ETT) and sumatriptan (STT) were investigated due to their differences in their Pgp-mediated efflux ratios despite their structural similarity.

FIGURES

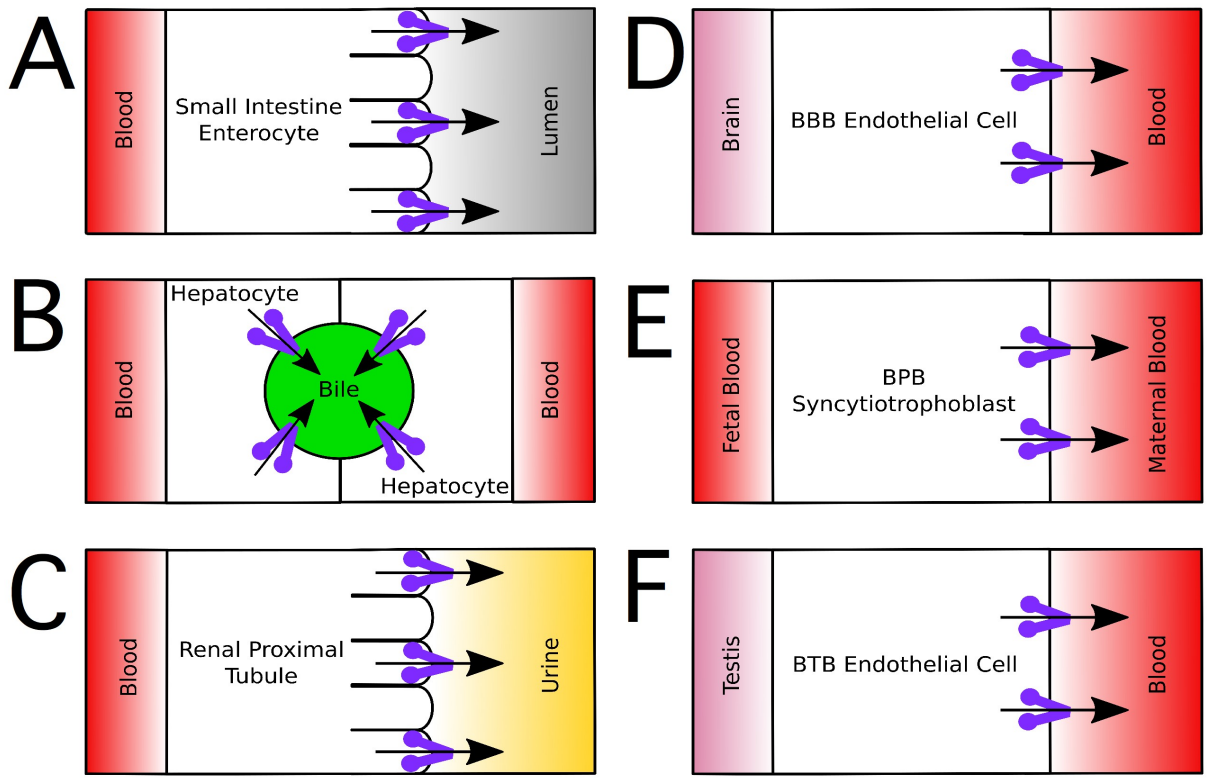

Figure 1.1 *The tissue distribution of P-glycoprotein (Pgp).* The tissue distribution of Pgp at the **(A)** small intestine enterocyte, **(B)** bile canaliculus, **(C)** renal proximal tubule, **(D)** BBB endothelial cell, **(E)** BPB syncytiotrophoblast and **(F)** BTB endothelial cell. Pgp is shown in purple and the arrows represent the direction of transport.

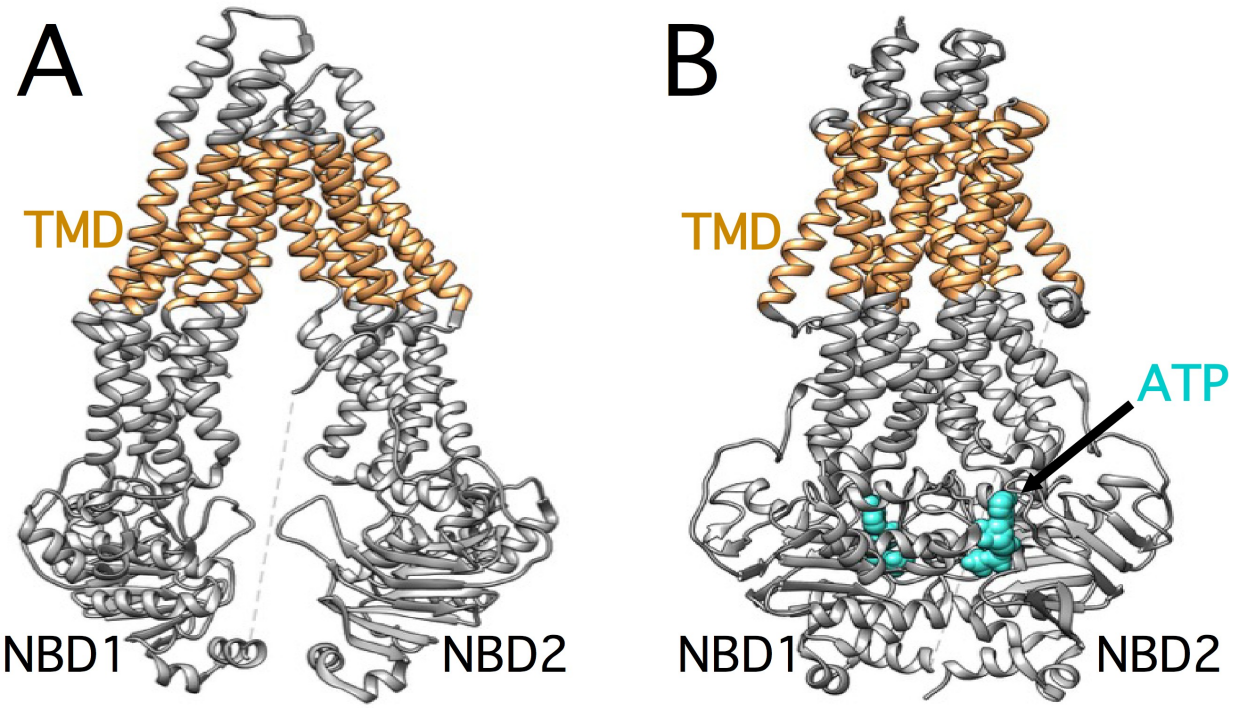

Figure 1.2 *The open and closed conformations of P-glycoprotein (Pgp)*. The **(A)** open conformation of *m*Pgp (PDB ID: 4M1M) and the **(B)** closed conformation of *h*Pgp with two ATP molecules bound (PDB ID: 6C0V) are shown as gray cartoons with the NBDs, TMD and ATP molecules labeled. The TMD spanning the membrane are colored tan. The ATP molecules bound into the NBDs are shown as Van der Waals spheres in turquoise.

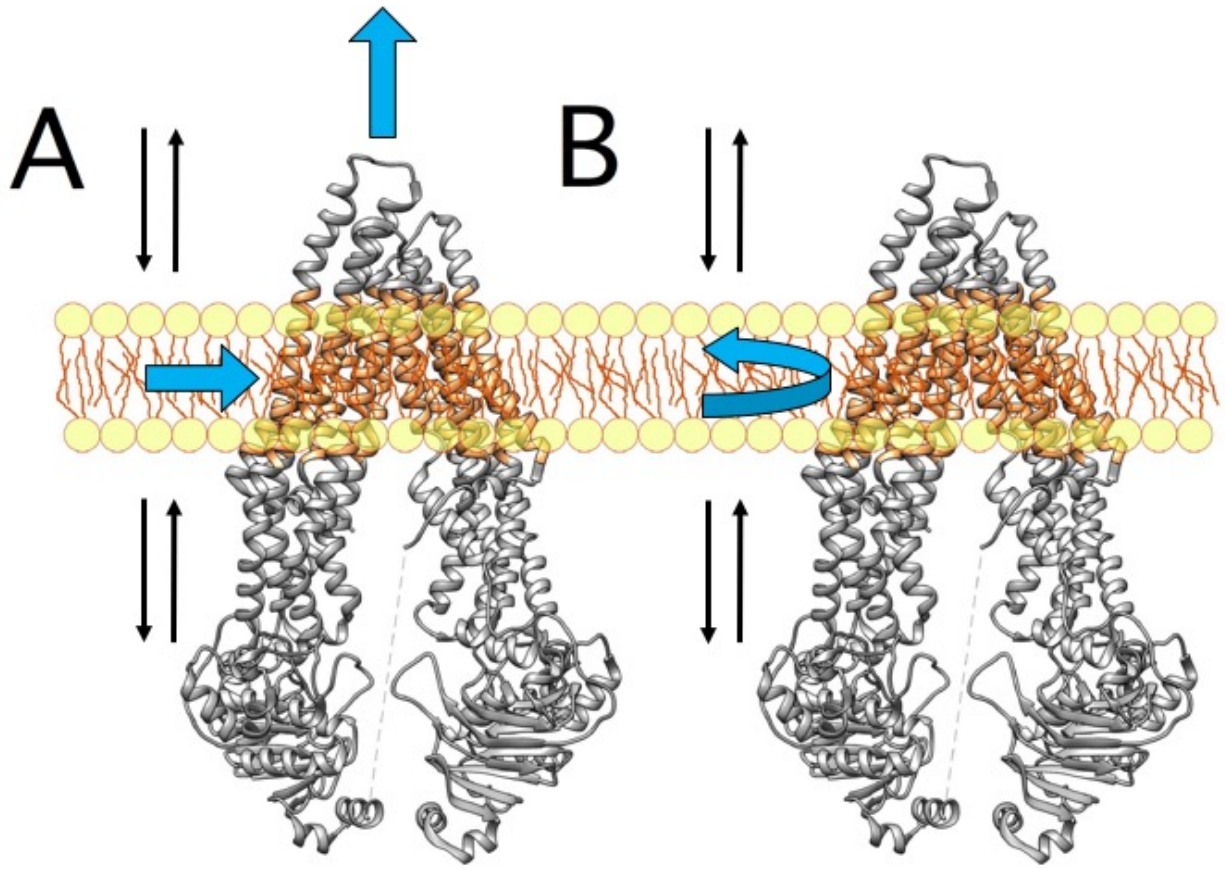

Figure 1.3 *The substrate binding and transport models of P-glycoprotein (Pgp).* The **(A)** hydrophobic vacuum cleaner and **(B)** substrate flippase models describing Pgp-mediated drug transport. The small black arrows show passive diffusion of drug molecules and the large blue arrows reflect Pgp-mediated drug transport.

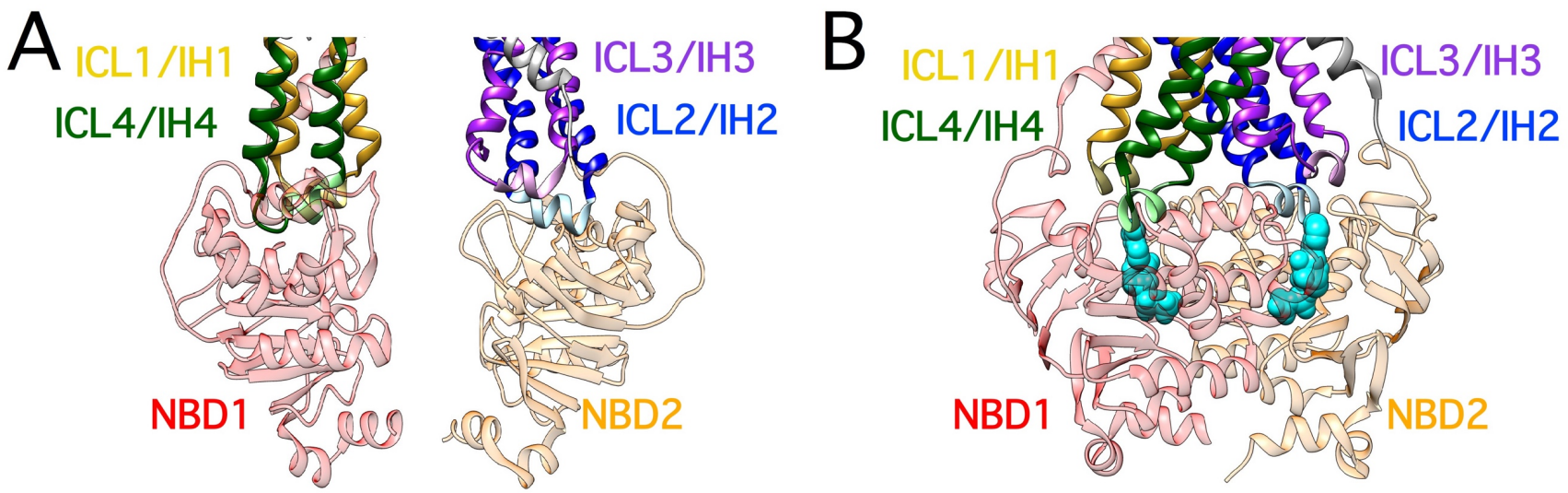

Figure 1.4 *The crosstalk region between the TMD and the NBD of P-glycoprotein (Pgp).* The intracellular loops (ICLs) of Pgp in the **(A)** open (PDB ID: 4M1M) and **(B)** closed conformation (PDB ID: 6C0V). The ICLs are color coded by name. The light-colored regions on the ICLs are the intracellular helices (IHs) that directly interact with the NBD and/or ATP. The ATP molecules bound into the NBDs are shown as Van der Waals spheres in turquoise.

CHAPTER 2

INSIGHTS INTO THE MOLECULAR MECHANISM OF TRIPTAN TRANSPORT BY P-GLYCOPROTEIN [1]

[1] Wilt, L.A., Nguyen, D. and A. G. Roberts (2017). *Journal of Pharm. Sci*. 106:1670-1679. Reprinted here with permission of the publisher.

2.1 ABSTRACT

The P-glycoprotein (Pgp) transporter reduces the penetration of a chemically diverse range of neurotherapeutics at the blood-brain barrier (BBB), but the molecular features of drugs and drug-Pgp interactions that drive transport remain to be clarified. In particular, the triptan neurotherapeutics, eletriptan (ETT) and sumatriptan (STT), were identified to have a greater than 10-fold difference in transport rates despite being from the same drug class. Consistent with these transport differences, ETT activated Pgp-mediated ATP hydrolysis ~2-fold, while STT slightly inhibited Pgp-mediated ATP hydrolysis by ~10%. The interactions between them were also non-competitive, suggesting that they occupy different binding sites on the transporter. Despite these differences, protein fluorescence spectroscopy revealed that the drugs have similar affinity to the transporter. NMR with Pgp and the drugs showed that they have distinct interactions with the transporter. Tertiary conformational changes probed by acrylamide quenching of Pgp tryptophan fluorescence with the drugs and a non-hydrolyzable ATP analog implied that the STT-bound Pgp must undergo larger conformational changes to hydrolyze ATP than ETT-bound Pgp. These results and previous transport studies were used to build a conformationally driven model for triptan transport with Pgp where STT presents a higher conformational barrier for ATP hydrolysis and transport than ETT.

2.2 INTRODUCTION

P-glycoprotein (Pgp) is a member of the ATP-binding cassette superfamily and acts as the gatekeeper of the blood-brain barrier (BBB) by limiting the penetration of a chemically and structurally diverse range of molecules.[68, 78] This function protects the brain from toxic insults, but also prevents the penetration of commercially available and potentially beneficial

neurotherapeutics,[79] which has made neurotherapeutic drug development a formidable challenge.[17,18] One strategy to overcome this barrier has been to modulate Pgp transport with inhibitors, but so far this has been unsuccessful in the clinic.[19,20] Therefore, there is a lot of interest in the pharmaceutical industry to identify molecular features of neurotherapeutics that drive Pgp transport and to determine molecular mechanisms of neurotherapeutic transport by Pgp.

Pgp is localized to the outer membrane of endothelial cells of the BBB.[14,21,22] The X-ray crystal structure of mouse Pgp revealed that the transporter is a 140 kDa asymmetric monomer comprised of twelve transmembrane helices and two nucleotide-binding domains (NBDs), which lie on the cytosolic side of the membrane and bind ATP.[12, 27] The promiscuity of Pgp to neurotherapeutics is the result of a 6,000 $Å^3$ cavity formed by the transmembrane helices, [23,24] which allows multiple drugs to bind to the transporter simultaneously.[28, 29] The hydrolysis of ATP at the NBDs drive directional transport as a result of large conformational changes that bring the NBDs within contact, as seen in functionally related bacterial transporters.[87, 88]

Triptan drugs represent a class of neurotherapeutics that abort migraines by acting on the 5-hydroxytryptamine 1B and 1D (5-$HT_{1B/1D}$) receptors that are distributed throughout the brain. Several triptan drugs are substrates for Pgp and show varying degrees of brain penetration and Pgp transport,[68, 69, 73, 92] but it is unclear what molecular features of triptans drive transport. At one extreme is the triptan drug sumatriptan (STT, Figure 2.1A). It was the first clinically available triptan drugh[93] and it is one of the weakest Pgp substrates. [e.g. 69] *In vitro* studies with the drug and different cell types overexpressing Pgp revealed a range of efflux ratios from 1.1 to 2.9.[68, 69, 73] Some studies have identified STT to be a good Pgp substrate, with efflux ratios higher than other triptan drug class members,[69] while other studies suggest STT is a non-transported substrate of Pgp.[68] This phenomenon has also been previously noted for Pgp and the cardiovascular drug

verapamil.[28, 48, 94] *In vivo*, STT was shown to be weakly transported by Pgp at the BBB in mice.[35] On the other extreme, eletriptan (ETT, Figure 2.1B) is a second-generation triptan drug[93] and a very good Pgp substrate. *In vitro* studies of ETT identified the drug to have Pgp efflux ratios ranging from 11[69] to 46.7[73]. *In vivo,* Pgp knockout mice showed a ~50 fold increase in ETT concentration in the brain compared to wildtype.[92]

The molecular mechanism of transport and the specific interactions of triptan drugs with Pgp are currently unknown. The large difference in log P values of 4.1 and 0.90 for ETT and STT, respectively, may provide an explanation for the large differences in their transport rates.[95] Therefore, we hypothesized that the higher transport rate of ETT by Pgp is the result of stronger interactions with Pgp because of its relatively high log P value when compared with STT. To test this hypothesis, the molecular interactions of ETT and STT with the transporter were examined. First, the effect of these drugs on Pgp-coupled ATP hydrolysis was determined. The interactions of the triptan drugs with Pgp were probed by protein fluorescence spectroscopy and the saturation transfer double difference (STDD) NMR technique. Finally, the effect of these drugs on Pgp conformation in the absence and presence of a non-hydrolyzable ATP analog was examined by acrylamide quenching of tryptophan fluorescence. These experimental results and previously published transport studies from the literature[68, 69, 73, 92] were used to build a conformationally gated model describing the molecular mechanism of triptan transport by Pgp.

2.3 MATERIALS AND METHDOS

2.3. 1. Materials

Cholesterol, Tris-HCl and disodium ATP (Na_2ATP) were purchased from Amresco (Solon, OH). *Escherichia coli* (*E. Coli*) total lipid extract powder was purchased from Avanti Polar Lipids Inc.

(Alabaster, AL). HEPES and acrylamide were purchased from Calbiochem (San Diego, CA). Deuterium oxide (D_2O) was purchased from Cambridge Isotope Laboratories Inc. (Tewksbury, MA). Deuterated dithiothreitol (d_{10}-DTT) was purchased from CDN Isotopes (Quebec, Canada). The detergent *n*-dodecyl-β-D-maltoside (DDM) was purchased from EMD Millipore Corporation (San Diego, CA). Sodium orthovanadate (Na_3VO_4) was purchased from Enzo Life Sciences (Farmingdale, NY). Sumatriptan was purchased from Fagron, Inc. (St. Paul, MN). Dithiothreitol (DTT) was purchased from Gold Biotechnology (Olivette, MO). Eletriptan was given as a gift from Pfizer Inc. (New York City, NY). Ammonium chloride (NH_4Cl) and deuterated DMSO (DMSO-d_6) was purchased from Sigma Aldrich (Milwaukee, WI). The 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) detergent was purchased from VWR International (Radnor, PA). All the remaining chemicals in this study were purchased from Thermo Fisher Scientific (Waltham, MA).

2.3.2 Purification and reconstitution of Pgp

The his-tagged wildtype mouse Pgp (*Abcb1a,* Mdr3) was purified from *Pichia pastoris* in two stages with nickel-nitrilotriacetic acid (Ni-NTA) (Thermo Fisher Scientific, Waltham, MA) and Whatman DE52 cellulose resin (Thermo Fisher Scientific, Waltham, MA) as previously described.[96, 97] Detergent solubilized Pgp was directionally integrated into liposomes with the NBDs on the outside using the procedure described in.[48, 94] These liposomes were composed of 80% w/v Avanti total *Escherichia (E.) coli* Total Lipid Extract (Avanti Polar Lipids, Alabaster, AL) and 20% w/v cholesterol with a lipid to protein ratio of 0.16 mg ml^{-1}.[48, 94] The Pgp orientation in the liposomes was determined by permeabilizing the liposomes to ATP with CHAPS detergent so that any NBDs located on the inside of the liposome would hydrolyze ATP. There was no

increase in the apparent ATPase activity as a result of NBDs located on the inside of the liposomes (data not shown) demonstrating that the NBDs are essentially on the external side of the liposome.[48] Proteoliposome aliquots were stored at -80°C in HEPES buffer (20 mM HEPES, 100 mM $NaCl_2$, 5 mM $MgCl_2$, 2 mM DTT pH 7.4). The concentration of protein was determined with the extinction coefficient of 1.28 mg mg^{-1} cm^{-1} and the DC Protein Assay Kit II (Bio-Rad, Hercules, CA).[48, 94]

2.3.3 ATPase activity measurements.

The ATPase activity of Pgp was measured using the Chifflet method on a FlexStation 3 spectrometer (Molecular Devices, Sunnyvale, CA), as previously described.[37,38,42] The concentration of free inorganic phosphate (*Pi*) after ATP hydrolysis was estimated by the absorbance signal at 850 nm from the formation of a *Pi*-molybdenum complex. The ATPase activity of ETT and STT was measured in the presence of 50 nM Pgp reconstituted in liposomes in Chifflet buffer (150 mM NH_4Cl, 5 mM $MgSO_4$, 0.02% w/v NaN_3, 50 mM Tris-HCl, pH 7.4).[48, 94, 98]

Simple enzyme kinetics is traditionally analyzed using linear transformations such as Lineweaver-Burk, Hans-Woolf and Eadie-Hofstee plots.[99, 100] Unfortunately, these transformations suffer from lack of variable independence across the axes and they have the tendency to bias the data points.[101–103] In contrast, non-linear regression is considerably more accurate than linear transformation, and is easily performed with modern computers.[101] Therefore, the Pgp-coupled ATP hydrolysis kinetics was fit by non-linear regression with Igor Pro 6.2 (Wavemetrics, Tigard, OR) and the following equations. Hyperbolic Pgp-coupled ATP hydrolysis kinetics of the drugs were fit to the modified Michaelis-Menten equation (Eq. 1):[99]

$$v = \frac{V_{\max} \cdot [L]}{K_m + [L]} + v_{basal} \quad (1)$$

where V_{max} is the maximum ATP hydrolysis, K_m is the Michaelis-Menten constant, [L] is the concentration of ligand and v_{basal} is the basal ATPase activity in the absence of drugs. For drug interactions that are competitive, the ATP hydrolysis curves were fit to Eq. 1 and an apparent Michaelis-Menten constant ($K_{m,app}$) shown in Eq. 2:[99]

$$K_{m,app} = K_m \cdot \left(1 + \frac{[I]}{K_I}\right) \quad (2)$$

where *[I]* is the inhibitor and K_I is the inhibitory constant. For non-competitive inhibition, the apparent V_{max} ($V_{max,app}$) can be calculated with the following equation (Eq.3):[99]

$$V_{\max,app} = \frac{V_{\max}}{1 + \frac{[I]}{K_I}} \quad (3)$$

This equation assumes that the enzyme velocity will approach 0 at saturating [I]. In cases where there is only partial inhibition of the enzyme at saturating [I], the $V_{max,app}$ will approach a new V_{max} ($V_{max,uninhibited}$). In this case, Eq. 3 can be modified to Eq. 4:

$$V_{\max,app} = \frac{V_{\max,inhibited}}{1 + \frac{[I]}{K_I}} + V_{\max,uninhibited}$$

(4)

where $V_{max,inhibited}$ represents the maximum inhibitable enzyme velocity at saturating [I].

2.3.4 Fluorescence Spectroscopy.

Quenching of protein fluorescence has been used to measure the dissociation constants (K_Ds) of a wide range of ligands to Pgp.[48, 94, 104, 105] This technique was used to measure the affinity

of ETT and STT to Pgp. An Olis DM 45 spectrofluorimeter (Olis Corp, Bogart, GA) was used to measure the drug-induced quenching of 1 μM Pgp reconstituted in liposomes in 100 mM potassium phosphate (KP_i) and 2 mM DTT (pH 7.4) to prevent cysteine disulfide linkage between Pgp monomers, as previously described.[48, 94] Pgp protein fluorescence emission was observed between 300 and 500 nm with a fluorescence emission maximum at ~330 nm, when the protein was excited at wavelengths between 260 and 295 nm.[106] The ETT and STT molecules also emit fluorescence in this region, which could potentially complicate Pgp protein fluorescence measurements. Fortunately, the drug fluorescence only partially overlaps with the Pgp protein fluorescence (Figure 2.S1), even at the highest concentrations used in the study. For drug concentrations less than 400 μM, the drug fluorescence contributions could be effectively subtracted from the overall fluorescence signal. The drug fluorescence contribution was further minimized for ETT and STT by exciting Pgp at 280 and 295 nm, respectively. The drug induced fluorescence quenching was corrected ($F_{corrected}$) for background fluorescence and inner filter effects with the following equation (Eq. 5):[94, 106]

$$F_{corrected} = (F - B)^{\frac{(\varepsilon_{ex} b_{ex} + \varepsilon_{em} b_{em})}{2}[Q]} \quad (5)$$

where F is the measured protein fluorescence at 330 nm, B is the background, $[Q]$ is the quenching ligand concentration, ε is the extinction coefficients for the excitation (ε_{ex}) and emission (ε_{em}) and b is the pathlength for the excitation (b_{ex}) and emission (b_{em}). The ε_{ex} at 280 nm and ε_{em} at 330 nm for ETT was 2.13 mM^{-1} cm^{-1} and 0.432 mM^{-1} cm^{-1}, respectively. For STT, the ε_{ex} at 295 nm was determined to be 0.808 mM^{-1} cm^{-1} and the ε_{em} at 330 nm was 0.344 mM^{-1} cm^{-1}.

Protein fluorescence quenching can occur by static or dynamic mechanisms. Static quenching occurs from ligand forming a complex with the protein and is directly correlated to the ligand's affinity.[106] In contrast, dynamic or collisional quenching is the result of random collisions

between a ligand and protein.[106] To differentiate between different quenching mechanisms, protein fluorescence quenching titrations were measured at different temperatures as described.[48, 94, 106] The corrected fluorescence ($F_{corrected}$) of a monophasic fluorescent quenching curve was fit to the following rearrangement of the Stern-Volmer equation (Eq. 6):[38,50]

$$F_{corrected} = \frac{F_{corrected,0}}{1 + K_{SV}[Q]} + F_{unquenched} \quad (6)$$

where $F_{corrected,0}$ *is fluorescence in the absence of a quenching ligand*, $F_{unquenched}$ is an offset related to the unquenched fluorescence, K_{SV} is the Stern-Volmer constant, and [Q] is the concentration of quenching ligand. Biphasic fluorescence quenching curves were fit to the following equation (Eq. 7):[106]

$$F_{corrected} = \frac{F_{L,0}}{1 + K_{SV,L}[Q]} + \frac{F_{H,0}}{1 + K_{SV,H}[Q]} + F_{unquenced} \quad (7)$$

where $F_{0,L}$ and $F_{0,H}$ are the fluorescence amplitudes at low and high quenching ligand concentrations, respectively. The $K_{SV,L}$ and $K_{SV,H}$ are the Stern-Volmer constants at low and high quenching ligand concentrations, respectively.

The collisional quenching of Pgp by acrylamide has been used to probe changes in exposure to solvent accessible tryptophan residues to determine conformational changes of Pgp upon ligand binding. [e.g. 48, 107, 108] Acrylamide is an uncharged polar compound that has minimal penetration into lipids and the hydrophobic core of proteins.[109] Fluorescence emission of 1 μM Pgp reconstituted in liposomes was monitored at 330 nm after excitation at 295 nm in the presence of saturating concentrations of ETT and STT in 100 mM KPi and 2 mM DTT (pH 7.4). The fluorescence emission was corrected for inner filter effects and background fluorescence with Eq. 5. To compare the results with the ATP hydrolysis experiments, the fluorescence quenching was monitored in the presence of 3.2 mM of the non-hydrolyzable ATP analog adenosine 5'-(β,γ-

imido)triophsophate (AMPPNP). *N*-acetyl-L-tryptophanamide (NATA) is an analog of solvent accessible tryptophans that was used to determine the maximum possible tryptophan quenching.[107] The Stern-Volmer plots were created by plotting the $F_{corrected,0}/F_{corrected}$ as a function of acrylamide concentrations. The K_{SV} value is a quantitative measure of exposure to solvent accessible tryptophans, which was estimated from the slopes of the Stern-Volmer plots as described previously. [e.g. 48, 106]

2.3.5 Saturation transfer double difference NMR technique

All NMR experiments were performed at 25°C on a Varian INOVA 600 MHz spectrometer with a 5mm z-gradient $^{1}H\{^{13}C/^{15}N\}$ cryoprobe and analyzed in iNMR software (http://ww.inmr.net) and Igor Pro 6.2, as described previously.[48, 94] The ^{1}H NMR peaks of 100 mM STT and ETT in DMSO-d_6 were assigned using standard ^{1}H 1D NMR techniques. The ^{1}H NMR peak assignments for STT and ETT are shown in Figure 2.1. This labeling scheme was used for the assigning ^{1}H NMR peaks.

The saturation transfer difference (STD) NMR technique is used to identify drug functional groups that interact with a protein receptor by selectively saturating the receptor and monitoring the saturation transfer from the receptor to the drug.[110–112] This observed degree of saturation transfer is directly correlated to the drug's interaction with the receptor.[110, 112] For membrane proteins reconstituted into a lipid system, the lipid-drug interactions can lead to non-specific interactions and interference in the ^{1}H STD NMR spectrum. The saturation transfer double difference (STDD) NMR technique was developed to extract specific NMR signals corresponding to drug-protein interactions from samples with lipid membranes.[48, 94, 111] The STDD NMR spectrum is created by subtracting the STD NMR spectrum of the drug in the presence of lipid

from the STD NMR spectrum of a drug in the presence of a protein embedded in the same lipid.[94, 111, 113, 114] This technique was used to probe the ETT and STT binding interactions with Pgp reconstituted in liposomes. NMR samples contained 1 μM reconstituted Pgp in 80% deuterated 100 mM KP_i, pD 7.4 and saturating concentrations of triptans, 250 μM. Control liposome samples were prepared identically to reconstituted Pgp samples. A train of 50 ms gaussian shaped pulses selectively saturated reconstituted Pgp for a total saturation of 2 s. The water suppression by gradient tailored excitation (WATERGATE) pulse sequence was added to suppress the water signal.[48] The STD NMR spectra were produced by phase cycling and alternating the saturation frequency between -1.5 and 40 ppm during 512 scans.[48, 94] To create the ^{1}H STDD NMR spectrum of the drugs with Pgp, the ^{1}H STD NMR spectrum of the control liposome sample with drug was subtracted from the ^{1}H STD NMR spectrum of the drug with reconstituted Pgp. This difference (ΔI) correlates to the molecular interactions between the drug and the transporter. This was quantitated using STDD amplification factors calculated from the following equation (Eq. 8):[48]

$$STDD\ Amplification\ Factor = \frac{[L]}{[P]}\frac{\Delta I}{I_0} \quad (8)$$

where *[P]* is the protein concentration, [*L*] is the concentration of ligand and I_0 is the amplitude of the ^{1}H NMR peaks in the spectrum without radio frequency saturation.

2.4 RESULTS

2.4.1 The effect of ETT and STT on Pgp ATP activation

Figure 2.2 shows the Pgp-coupled ATPase activity in the presence of ETT and STT. In the absence of drugs, the Pgp-coupled ATP hydrolysis activity was 560 ± 40 nmol min^{-1} mg^{-1} (Figure 2.2A), which is similar to previous estimates.[48, 94] The ATP hydrolysis activity was almost completely abolished in the presence of 200 μM of the Pgp inhibitor Na_3VO_4 (data not shown). In

Figure 2.2A (closed squares), the Pgp-coupled ATP hydrolysis activity was measured with a range of ETT concentrations. The ATP hydrolysis kinetics was monophasic and reached a maximum at ~1000 nmol min^{-1} mg^{-1}. Fitting the curve to the modified Michaelis-Menten equation (Eq. 1) generated a V_{MAX} and K_M value of 508 ± 27 nmol min^{-1} mg^{-1} and 14.8 ± 3.1 μM, respectively. Figure 2.2A (open circles) also shows the Pgp-coupled ATP hydrolysis kinetics in the presence of STT. In contrast to ETT, STT weakly inhibits Pgp-coupled ATP hydrolysis kinetics by < 10%. This effect is not inconsistent with STT being transported by Pgp. Several known Pgp transported substrates, colchicine, doxorubicin and chloramphenicol, have little impact on Pgp-coupled ATP hydrolysis kinetics[115–117] and the fluorescent probe substrate Hoechst 33342, in fact, inhibits ATP hydrolysis by Pgp, but is still transported.[60]

The fact that Pgp-mediated ATP hydrolysis kinetics in the presence of ETT and STT were distinct in Figure 2.2A allowed us to probe the competition of these drugs to Pgp. Classically, competitive inhibition will lead to an increase in K_m as a result of drug displacement from Pgp, while non-competitive inhibition will lead to a decrease in V_{max}, but will have little effect on the K_m.[99] Non-competitive inhibition also implies simultaneous binding of both drugs.[99] Figure 2.2B shows the effect of a range of STT concentrations on the activation of Pgp-coupled ATP hydrolysis by 125 μM ETT. Increasing STT concentrations only decreased the apparent ATP hydrolysis rate about 30%. The fact that the ATP hydrolysis is only partially inhibited implies that both drugs are bound simultaneously and that the interactions are non-competitive. Fitting the ATP hydrolysis kinetics in Figure 2.2B with Eqs. 1 and 4 gives a K_I, $V_{max,inhibited}$ and $V_{max,uninhibited}$ values of 30.2 ± 12.1 μM, 201 ± 21 nmol min^{-1} mg^{-1} and 187 ± 17 nmol min^{-1} mg^{-1}, respectively. To demonstrate that ETT and STT interactions were in fact non-competitive, an ETT titration of Pgp was performed in the presence of 500 μM STT in Figure 2.2C. Fitting the curve to Eq. 1 reveals that

the activated V_{max} decreases 30% to 341 ± 20 nmol min^{-1} mg^{-1} and the extracted K_m is 11.1 ± 2.8 μM. The decrease in V_{max} and essentially no change from the previous K_m value of 14.8 ± 3.1 μM (Figure 2.2A, closed squares) demonstrate that the interactions between ETT and STT are indeed non-competitive. The K_I for STT estimated by fitting Figure 2.2C with Eqs. 1 and 4 was 48.3 ± 18.0 μM, which is similar to the K_I obtained from the curve fit in Figure 2.2B.

2.4.2 Probing the interactions of ETT and STT with Pgp by intrinsic protein fluorescence

Equilibrium constants (i.e. K_m, K_I) derived from fitting Pgp-mediated ATP hydrolysis kinetics curves do not necessarily correlate directly to ETT and STT binding affinity, since the fitting equations assume specific enzyme mechanisms.[99] One might anticipate that there could be significant differences in the affinities of ETT and STT to Pgp because of large differences in their log P values[118] and their apparent transport rates. [e.g. 92] Ligand affinities to Pgp have been directly estimated with fluorescent or radiolabeled ligands,[85, 119, 120] fluorescence quenching of labeled Pgp[121, 122] and quenching of intrinsic Pgp tryptophan fluorescence.[48, 94, 104] Directly measuring drug binding with fluorescently labeled Pgp may be problematic due to interference from the label.[121, 122] Therefore, the affinities of ETT and STT were estimated by measuring their impact on Pgp protein fluorescence.[48, 94, 104]

Figure 2.3 shows the protein fluorescence and analysis of Pgp in the presence of ETT and STT. Figure 2.3A shows the Pgp fluorescence emission spectra in the presence of ETT excited at 280 nm with the background fluorescence subtracted. The protein fluorescence produces a fluorescence maximum at 330 nm (thin line) that decreases ~50% with saturating ETT (thick line). After adjustment for inner filter effects with Eq. 5, the amplitude at 330 nm was monitored as a function of ETT concentration (Figure 2.3B). Quenching of Pgp by ETT was biphasic with low

and high concentration phases. Fitting the quenching curve to Eq. 7 gave $K_{SV,L}$ and $K_{SV,H}$ of 7.15 ± 1.14 μM^{-1} and 8.69 x 10^{-3} ± 0.74 x 10^{-3} μM^{-1}, respectively. The titration was repeated at 37°C to determine the quenching mechanism for each phase as done previously.[48, 94] The Stern-Volmer constants, $K_{SV,L}$ and $K_{SV,H}$, increased and decreased, respectively (data not shown). Therefore, the low concentration-quenching phase associated with $K_{SV,L}$ quenches by a dynamic mechanism, and the high quenching concentration phase associated with $K_{SV,H}$ quenches by a static mechanism. Thus, the K_A for ETT was 8.69 x 10^{-3} ± 0.74 x 10^{-3} μM^{-1} and the corresponding dissociation constant (K_D) was 115 ± 10 μM. Figure 2.3C shows the effects of STT on Pgp fluorescence after excitation at 295 nm with the fluorescence background subtracted. Saturating STT quenched Pgp fluorescence about 20%. In Figure 2.3D, Pgp fluorescence emission corrected for inner filter effects was monitored at 330 nm as a function of STT concentration. The fluorescence quenching curve was monophasic and fitting the curve to Eq. 6 gave a K_{SV} value of 10.8 x 10^{-3} ± 2.9 x 10^{-3} μM^{-1}. The K_{SV} value increased when this titration was repeated at 10°C showing that the quenching was the result of a static mechanism (data not shown) and the K_D at 25°C was 92.6 ± 25 μM. The surprising similarities between the K_DS between ETT and STT reveal their binding is not driven by drug hydrophobicity. Instead, the binding may be driven by other factors such as the number and strength of hydrogen bonds as suggested by a SAR study.[123]

2.4.3 Functional groups of ETT and STT involved in molecular recognition by Pgp

Although ETT and STT affinities are similar, the effects of these drugs on Pgp-coupled ATP hydrolysis were distinct and the interactions between the two drugs on the transporter were non-competitive. Therefore, their interactions with the transporter were probed by the STDD NMR

technique. The degree of drug interaction with the transporter is reflected in the amplitude of the ^{1}H STDD NMR peaks and quantitatively by ^{1}H STDD NMR amplification factors.[94]

Figure 2.4 shows the STDD NMR spectra, STDD amplification factors and %ΔSTDD amplification factors of ETT and STT in the presence of Pgp. The figure follows the labeling scheme of the STT and ETT molecular structures in Figure 2.1. The ^{1}H STDD NMR spectrum of 125 μM ETT in the presence of 1 μM Pgp is illustrated in Figure 2.4A. ^{1}H STDD NMR peaks with significant saturation transfer are labeled. Protons emanating from the liposomes and exchange broadening prevented us from reliably measuring the amplitudes of several alkyl and an amine ^{1}H NMR peaks of ETT labeled 1, 10-11 and 21-25 (see Figure 2.1B). From the STDD amplitudes in Figure 2.4A, the STDD amplification factors are estimated in Figure 2.4B using Eq. 8. The average STDD amplification factor for ETT to Pgp was 15. The strongest STDD amplification factor of 20 was observed for aromatic phenyl protons labeled 15,19, while the weakest STDD amplification factor of 10 came from the methyl labeled 27 of the pyrrolidine functional group. To exemplify differences in the STDD amplification factors, they were normalized against protons of the pyrrolidine methyl in Figure 2.4C. The phenyl group (protons 15-19, black) of ETT had STDD amplification factors that were 72% larger than the pyrrolidine methyl. The indole ring (protons 2-7, gray) of ETT had amplification factors that were 34% larger than the pyrrolidine methyl. These results imply that the phenyl of ETT plays the largest role in molecular recognition of Pgp.

Figure 2.4D shows the ^{1}H STDD NMR spectrum of 250 μM STT in the presence of 1 μM Pgp. Several ^{1}H STDD NMR peaks of STT were identified and labeled in the figure. Like ETT, reliable ^{1}H STDD NMR measurements were not possible for several alkyl protons (I and J) and amine protons (A and G) due to overlap with proton NMR signals from the liposome and exchange broadening of amines from STT (see Figure 2.1A). The STDD amplification factors from Figure

2.4D were calculated using Eq. 8 and are shown in Figure 2.4E. The average STDD amplification factor for STT was 6, which is less than half of ETT. This difference does not necessarily correspond to weaker interaction of STT with the transporter than ETT, since fluorescence spectroscopy with these drugs and Pgp in Figure 2.3 has already demonstrated that they have similar K_D*s*. Instead, lower STDD amplification factors by STT can result from lower saturation transfer efficiency[110, 124] because of how STT is distinctly sequestered by the transporter. For STT, the strongest STDD amplification factor of 9 was observed with proton E of the indole ring, while the weakest STDD amplification factor of 3 was observed for the methyl protons labeled K. Interestingly, the analogous methyl protons on ETT also had the weakest STDD amplification factor. To contrast the differences in STDD amplification factors of STT, they were normalized against the methyl protons labeled K in Figure 2.4F. The indole ring had 130% higher STDD amplification factors (gray) than the methyl protons. The N-linked methyl near the sulfone labeled H had STDD amplification factors (black) that were 75% of the methyl protons. These results imply that the indole functional group of STT plays the largest role in molecular recognition of the transporter.

2.4.4 Drug-induced conformational changes of Pgp in the presence of ETT and STT

Both X-ray crystal structures of analogous bacterial transporters with non-hydrolyzable ATP analogs[87, 88] and studies with Pgp [e.g. 3, 125] have suggested that conformational changes are involved in Pgp-mediated ATP hydrolysis. The Pgp conformations were probed by estimating the relative exposure of solvent accessible tryptophan residues (e.g. residues 158, 799, 1104) through acrylamide quenching as we did previously.[48, 94] Figure 2.5 shows the Stern-Volmer plots in the presence of drugs and the non-hydrolyzable ATP analog AMPPNP, where the slope is equivalent

to the K_{SV} value. In all cases, the curves in the Stern-Volmer plots were linear with respect to acrylamide concentration, which shows that acrylamide does not directly bind to Pgp or affect Pgp's conformation.[106]

In the absence of drugs, the Stern-Volmer plot of Pgp in Figure 2.5A (closed squares) had a K_{SV} value of 1.52 ± 0.20 M^{-1} with 34% of the total tryptophan fluorescence quenched at 1 M acrylamide. This degree of quenching correlates well to the ratio of fully solvent accessible tryptophans (3) versus the total number of tryptophan residues (11) deduced from the X-ray crystal structure of Pgp.[10, 45] As a positive control, NATA was used to assess the maximum tryptophan quenching. NATA was quenched 93% at 1M acrylamide and had a K_{SV} value of 13.4 ± 0.30 M^{-1} (Figure 2.5A, dashed line).

The addition of the ATP analog AMPPNP (open circles) to Pgp reduced the K_{SV} value to 0.84 ± 0.20 M^{-1}, resulting in a ΔK_{SV} of 0.68 ± 0.20 M^{-1}. The change in tryptophan accessibility or the ΔK_{SV} value is consistent with nucleotide-induced conformational shift observed in previous crystallography studies. [e.g. 87, 88]

The effect of STT on the solvent accessibility of tryptophan residues from Pgp is shown in Figure 2.5B (closed squares). The STT-bound complex has a significantly higher K_{SV} value of 4.13 ± 0.20 M^{-1} and in the presence of AMPPNP (open circles), the K_{SV} value decreases to 2.69 ± 0.20 M^{-1}, resulting in a ΔK_{SV} value of 1.44 ± 0.20 M^{-1}. This ΔK_{SV} value is considerably larger than the ΔK_{SV} value in the absence of drugs, implying that the STT-bound complex undergoes a relatively large conformational change in the presence of the ATP analog. In the presence of ETT, the K_{SV} value increased modestly to 1.96 ± 0.07 M^{-1} (Figure 2.5C, closed squares). Furthermore, the K_{SV} value of the ETT-bound complex remains essentially unchanged by the addition of AMPPNP to 2.06 ± 0.05 M^{-1} (Figure 2.5C, open circles) with a ΔK_{SV} value of -0.10 ± 0.09 M^{-1}. These results

imply that the nucleotide-induced Pgp conformational transitions are more modest in the presence of ETT than STT.

2.5 DISCUSSION

A conformationally gated model of triptan transport is presented in Figure 2.6 to explain the molecular mechanism driving ETT and STT transport by Pgp. The model is based on our results, the conformational changes that Pgp undergoes in the presence of nucleotides and ligands[94, 126–128] and the observed differences in triptan transport. [e.g. 68, 69, 73, 92] By analogy with the bacterial transporters,[88] Pgp can assume a wide range of conformations. For simplicity, we define three conformations of Pgp: "open", "intermediate" and "closed", as we did previously.[48, 94] In the "open" conformation, the NBDs are separated with the binding cavity exposed to the cytosol. The NBDs are together in the "closed" conformation with the binding cavity exposed to the extracellular space. The "intermediate" conformation is between the "open" and "closed" Pgp conformations with equivalent exposure to the extracellular and cytosolic sides of the membrane.

Changes to tryptophan accessibility of Pgp by ETT, STT and the non-hydrolyzable analog AMPPNP were deduced by acrylamide fluorescence quenching in Figure 2.5. The K_{SV} and ΔK_{SV} values were used to correlate the exposure of solvent accessible tryptophan residues to Pgp conformational changes. An increase in the K_{SV} value implies a shift to a more "open" conformation, while a decrease in K_{SV} value suggests a shift to a "closed" conformation. Moreover, the ΔK_{SV} value infers the degree of this conformational change. However, tryptophan accessibility and the corresponding K_{SV} values cannot be used to assign specific ligand-bound conformations, since it does not correlate directly to a specific Pgp conformation. We know from X-ray crystallography, cryo-em, cross-linking and mutagenesis studies with Pgp and the analogous

bacterial transporters that the interaction of the NBDs is essential for ATP hydrolysis.[3, 129–133] Therefore, Pgp-mediated ATP hydrolysis rates, relative changes in tryptophan accessibility from Figure 2.5 and previous Pgp conformational assignments[48, 94] were used to direct conformational assignment of Pgp in the presence of nucleotides, ETT and STT. In general, when the ligand-induced activation of Pgp-mediated ATP hydrolysis is relatively high, we assume that Pgp is shifted toward the "closed" conformation. When there is ligand-induced inhibition or no inhibition in Pgp-mediated ATP hydrolysis, we assume that Pgp is shifted toward or in an "open" conformation.

We know from X-ray crystallography of Pgp in the absence of ligands that the NBDs are relatively far apart in the "open" conformation. [e.g. 43] Therefore, Pgp is depicted in the "open" conformation in the absence of ligands in Figure 2.6A. Figure 2.5A shows that tryptophan accessibility decreases in the presence of AMPPNP (ΔK_{SV} = 0.68 M^{-1}) suggesting a shift to the "closed" conformation. However, the baseline P-gp mediated ATP hydrolysis rate hovers around 600 nmol min^{-1} mg^{-1}. Therefore, we propose that Pgp in the presence of ATP, but in the absence of ligands, is in an "intermediate" conformation in Figure 2.6B as we hypothesized previously.[94]

Figure 2.5B shows that STT increases the K_{SV} value of Pgp by 2.61 M^{-1}, while STT decreases the Pgp-mediated ATP hydrolysis rate by ~10% in Figure 2.2. These results imply a ligand-induced separation of the NBDs and a shift to a more "open" conformation. This is consistent with X-ray crystallographic studies showing Pgp in an "open" conformation, when complexed with inhibitors.[43, 134] Therefore, we propose that the STT-bound Pgp structure is in an "open" conformation and bound near the extracellular side of the transporter in Figure 2.6C like inhibitors in the Pgp X-ray crystal structures.[43, 134] In the presence of AMPPNP, the K_{SV} value of the STT-bound Pgp decreases by 1.44 M^{-1}, implying a large conformational shift toward the

"closed" conformation. A conformational barrier to ATP hydrolysis and transport would explain the 10% reduction in Pgp-mediated ATP hydrolysis in the presence of STT (Figure 2.2A) and the significantly lower Pgp-mediated transport of STT versus ETT.[69] Since Pgp-mediated ATP hydrolysis in the presence of STT is similar to basal activity, the STT-nucleotide-bound complex is hypothesized to be in an "intermediate" conformation in Figure 2.6D. *In vitro* transport studies with STT with Pgp have demonstrated that Pgp-mediated STT transport is cell type dependent.[68, 69, 73, 92] Therefore, STT transport or lack thereof is represented as a dashed arrow in Figure 2.6.

ETT induces ~2-fold activation of Pgp-mediated ATP hydrolysis, which is similar to digoxin-induced activation but is half of the verapamil-induced activation of Pgp.[48] Digoxin was hypothesized to shift Pgp into an "intermediate" conformation, while verapamil shifted the transporter to the "closed" conformation. Therefore, we propose that Pgp in the presence of ETT is in an "intermediate" conformation in Figure 2.6E like the digoxin-bound complex.[48] The Stern-Volmer plot in Figure 2.5C shows that the K_{SV} value for acrylamide quenching of Pgp in the presence of ETT changes little with saturating AMPPNP, implying that the conformation is changed little by the non-hydrolyzable ATP analog. Therefore, we propose that Pgp in the presence of ETT and ATP is also in an "intermediate" conformation in Figure 2.6F. With the ETT-bound Pgp being in the same conformation in the presence and absence of ATP, the conformational barrier for ATP hydrolysis is minimized when compared to STT. This reduction in the conformational barrier correlates well to the 5-20 fold higher efflux ratios observed for ETT than STT.[68, 69] In addition, the interactions of STT and ETT on Pgp probed by Pgp-mediated ATP hydrolysis (Figure 2.2B and 4.2C) were non-competitive suggesting that they occupy distinct sites on the transporter. Since ETT induces significant activation of Pgp-mediated ATP hydrolysis, we place ETT far from the hypothetical STT binding site near the NBDs in Figure 2.6E.

Our results in this study imply that triptan transport is not merely driven by chemical properties of STT and ETT. Triptans have distinct interactions with the transporter and induce conformational changes that activate or hinder Pgp-mediated ATP hydrolysis.[68, 69, 73, 92] These studies are the first to provide direct insight into the molecular mechanism of triptan transport by Pgp. Although this study is only focused on two neurotherapeutics, our simplified transport model may apply to a range of neurotherapeutics.

2.6 FIGURES

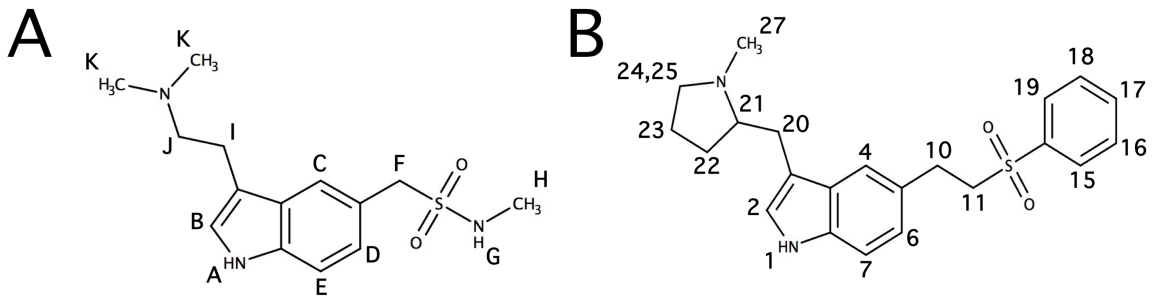

Figure 2.1. *The molecular structure of* ***(A)*** *STT and* ***(B)*** *ETT with proton assignments labeled.*

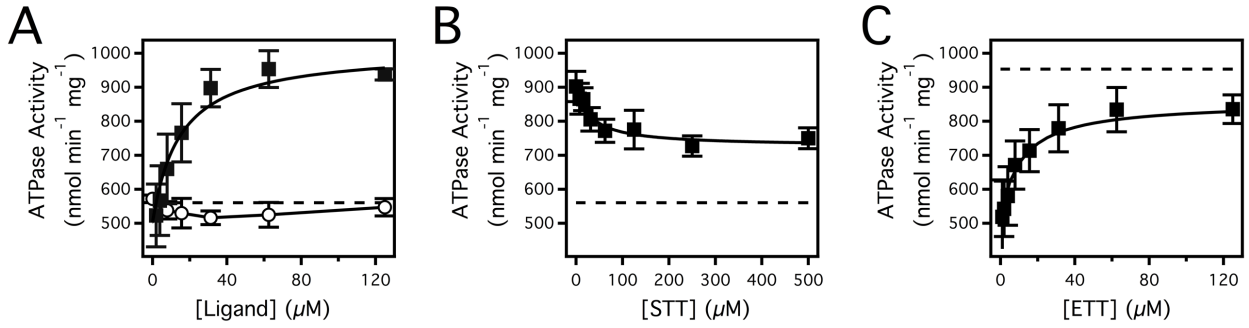

Figure 2.2. *The effect of ETT and STT on Pgp-mediated ATP hydrolysis.* **(A)** Pgp-coupled ATPase activity as a function of basal ATPase activity (dashed line), ETT (closed squares) and STT (open circles) concentrations. **(B)** Pgp-coupled ATPase activity in the presence of 125 μM ETT and a range of STT concentrations. For comparison, the level for basal ATPase activity is shown as a dotted line in the panel. **(C)** Pgp-coupled ATPase activity in the presence of 500 μM STT and a range of ETT concentrations. The dashed line in the panel represents the maximum Pgp-coupled ATP hydrolysis of ETT. Error bars represent the standard deviation and the points represent the average of at least three independent experiments.

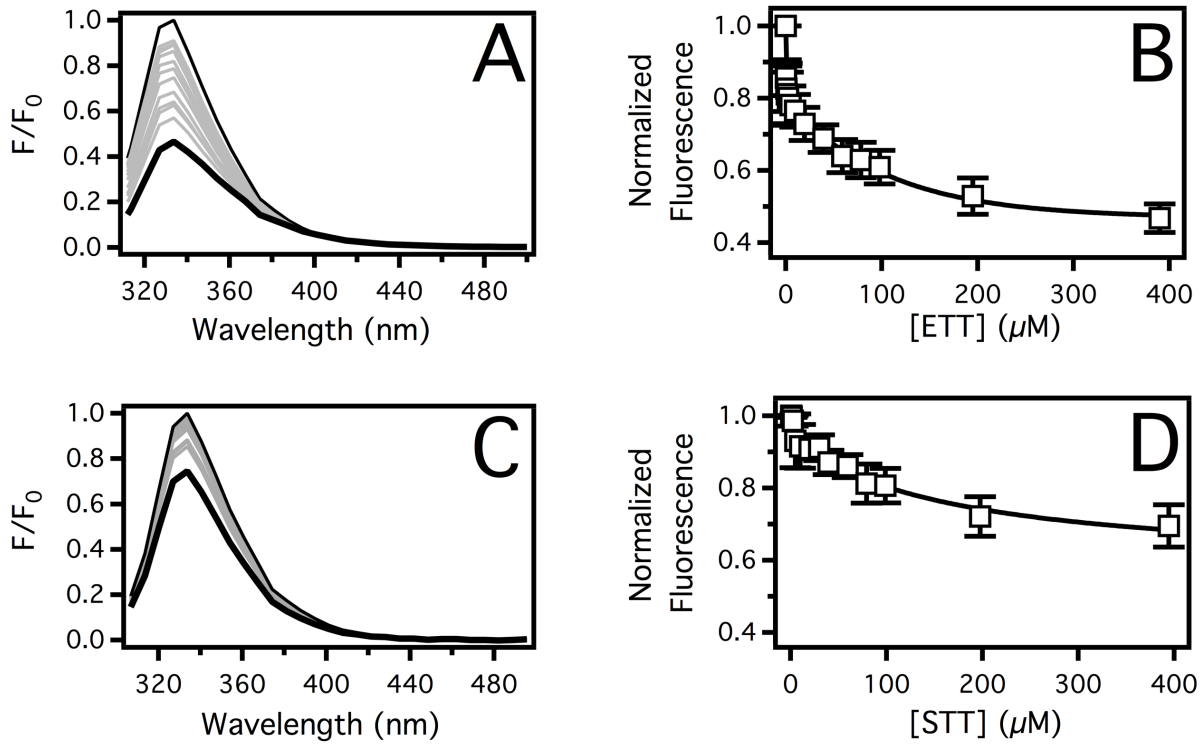

Figure 2.3. The affinity of *ETT and STT to Pgp deduced by protein fluorescence spectroscopy.* Pgp fluorescence emission spectra in the presence of a range of **(A)** ETT and **(C)** STT concentrations. The fluorescence emission spectrum at 0 μM and 400 μM triptan drugs are shown as thin and thick black lines, respectively. Spectra at intermediate concentrations are shown in gray. The corrected fluorescence emission amplitude at 330 nm was plotted as a function of **(B)** ETT and **(D)** STT concentration. The points and error bars represent the average and standard deviations, respectively, of at least three independent experiments.

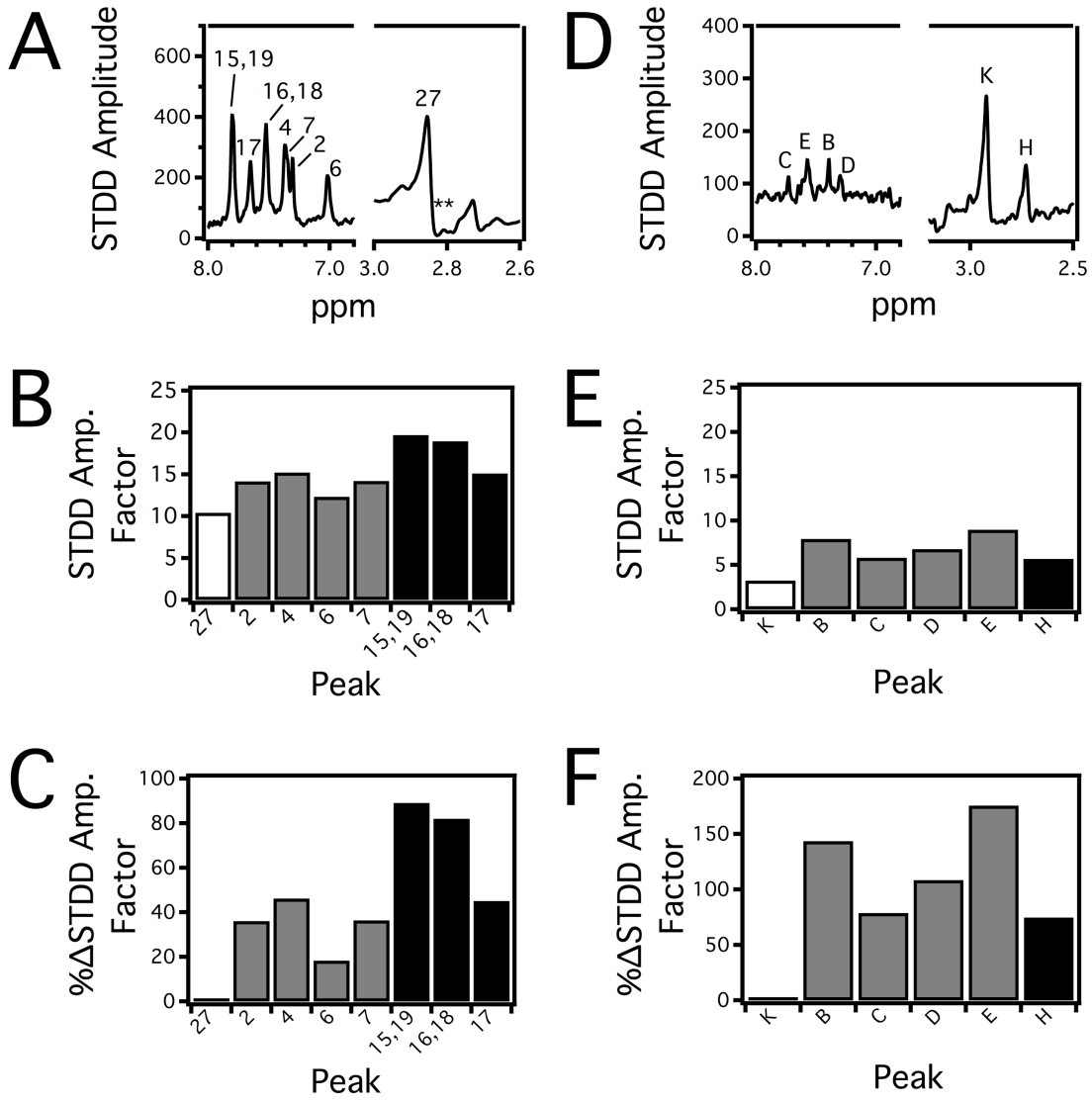

Figure 2.4. *ETT and STT interactions with Pgp investigated by the saturation transfer double difference (STDD) NMR technique.* The ^{1}H STDD NMR spectra with the peaks labeled of 250 μM **(A)** ETT and **(D)** STT in the presence of 1 μM Pgp. STDD amplification factors calculated for **(B)** ETT and **(E)** STT from the ^{1}H STDD NMR spectra. For clarity, STDD amplification factors from the indole protons are colored gray and protons from flanking functional groups are shown in white and black. To emphasize differences in the STDD amplification factors, they were normalized against the left most (weakest) STDD amplification factor as %ΔSTDD for **(C)** ETT and **(F)** STT.

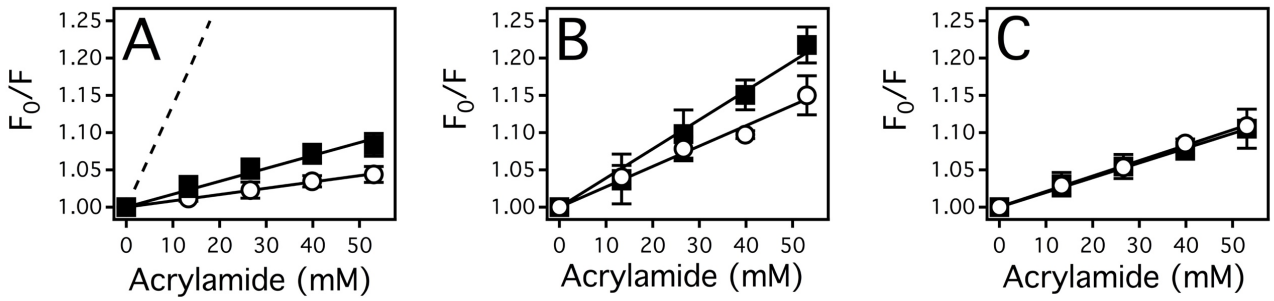

Figure 2.5. *ETT and STT induced conformational changes of Pgp in the absence and presence of an ATP analog AMPPNP inferred by fluorescence spectroscopy.* **(A)** The Stern-Volmer plots of NATA (dashed line), Pgp in the absence of drugs (closed squares) and in the presence of AMPPNP (open circles). **(B)** The Stern-Volmer plot of Pgp in the presence of STT (closed squares) and STT with AMPPNP (open circles). **(C)** The Stern-Volmer plot of Pgp in the presence of ETT (closed squares) and ETT with AMPPNP (open circles). The points represent the average and the error bars represent the standard deviation of at least three independent experiments.

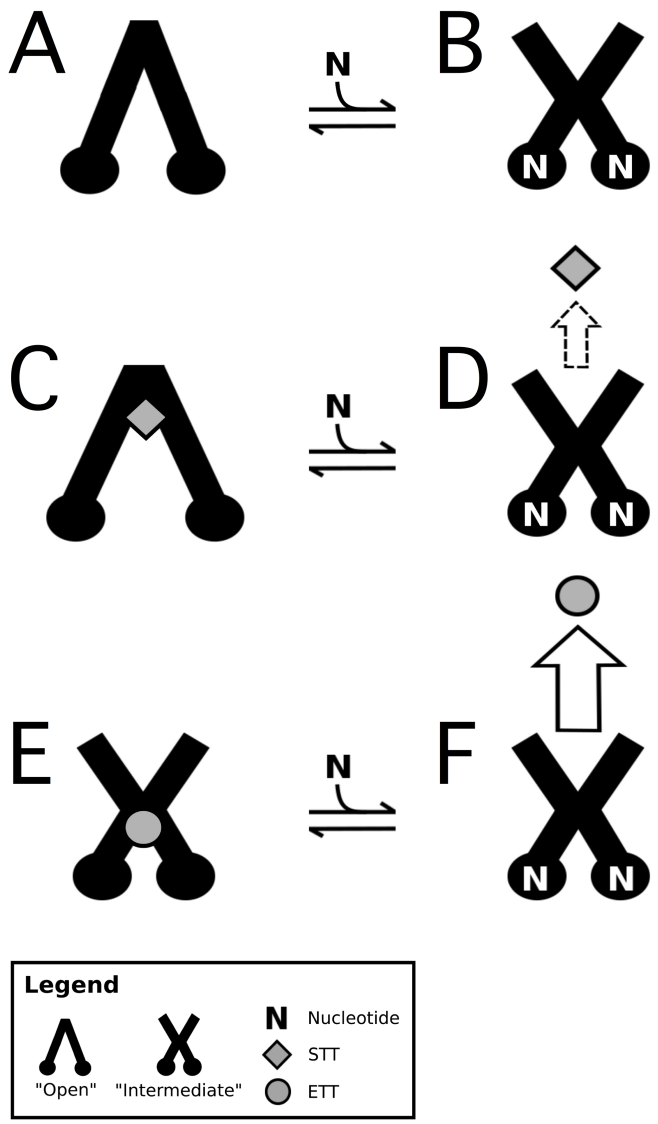

Figure 2.6. *Conformationally gated model of triptan transport by Pgp*. Pgp is depicted as a cartoon image of the "open" and "intermediate" states in the **(A,B)** absence of drug, the presence of **(C,D)** STT and **(E,F)** ETT. STT are shown as gray diamonds, ETT are gray circles and N is ATP. The horizontal arrows represent equilibrium between the nucleotide-bound conformational states of Pgp. The sizes of the vertical arrows represent the degree of transport. The dashed arrow reflects the fact that STT has been observed to be both transported[19] and not transported by Pgp[1] depending on the cell type.

2.7 SUPPLEMENTARY INFORMATION

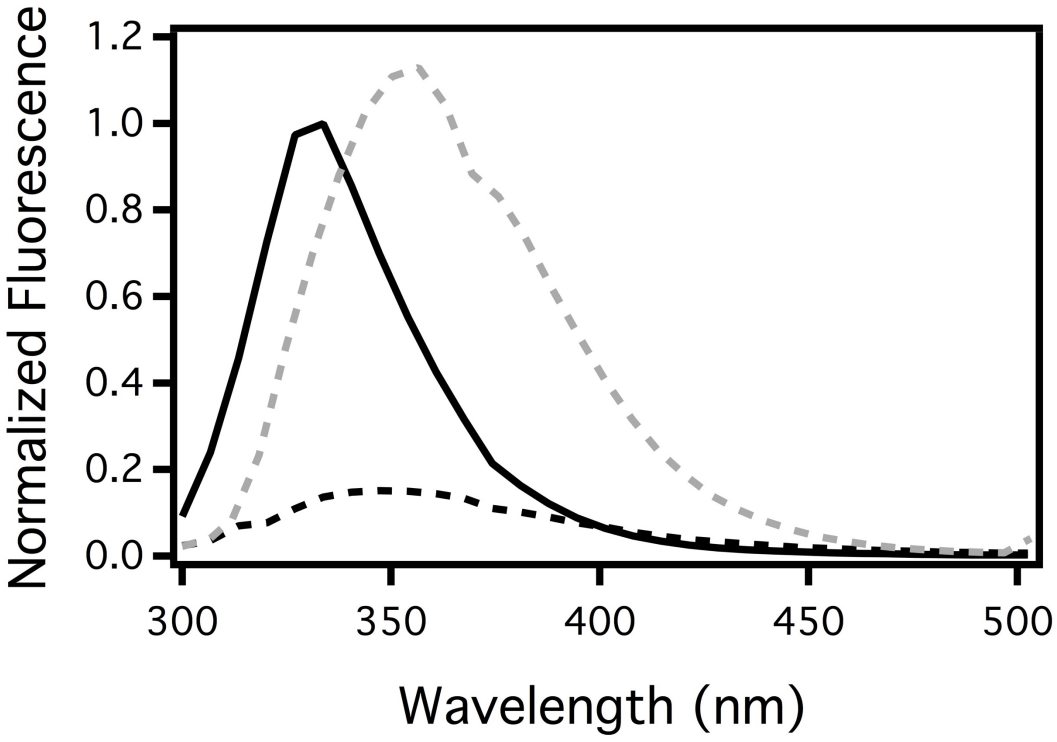

Figure 2.S1: *The maximum fluorescence emission overlap between ETT, STT and Pgp.* The figure shows the fluorescence emission spectra of 1 μM Pgp (solid, black line), 400 μM ETT (dashed, black line) and 400 μM STT (dashed, gray line) after excitation at 295 nm, 280 nm and 295 nm, respectively. The fluorescence emission spectra of ETT and STT in the figure are shown at the highest drug concentrations used in the study.

CHAPTER 3

HADDOCK-VINA: A HYBRID APPROACH FOR EXPERIMENTALLY-DRIVEN RECEPTOR-SMALL MOLECULE DOCKING [2]

[2] Wilt, L.A and A. G. Roberts. *Submitted* to *PROTEINS: Structure, Function, and Bioinformatics,* 03/15/18.

3.1 ABSTRACT

AutoDock Vina and High Ambiguity-Driven Biomolecular DOCKing (HADDOCK) protocols are powerful tools for molecular docking. AutoDock Vina is very efficient for blind docking and virtual drug screening of small molecules to receptors. Unfortunately, in many cases, there are discrepancies between docked and experimentally determined bound conformations. The docked conformations predicted by the data-driven HADDOCK software tend to be more refined because they are driven by experimentally derived distance restraints; however, small molecules can become trapped at sites far from their true binding sites. To take advantage of both docking procedures, a hybrid molecular docking wrapper called HADDOCK-Vina was developed that integrates a customized version of AutoDock Vina with HADDOCK. This wrapper also incorporates the PROPKA software to accurately protonate the receptor and other programs to simplify the process of docking. In addition, the wrapper uses a simplified experimental input to generate complex distance restraints. Initially, rigid molecular docking is performed with AutoDock Vina through the wrapper. Then the AutoDock Vina bound ligands are experimentally filtered using the HADDOCK scoring algorithm within the HADDOCK-Vina wrapper. These filtered ligand-receptor complexes are refined by molecular dynamics with and without explicit waters using the HADDOCK protocols. The outputted structures are clustered and analyzed. Ligands docked with HADDOCK-Vina in three cases reproduced the ligand positions in the corresponding ligand-bound protein X-ray crystal structures. The outputted ligand-bound structures from the wrapper were also considerably more accurate than docking the ligands with HADDOCK and AutoDock Vina by themselves.

3.2 INTRODUCTION

Determining how ligands interact with receptors can accelerate the design of novel drugs.[135, 136] With the modernization of high throughput screening and large compound libraries, molecular docking has been exploited to rapidly screen pharmacophores and lead compounds for a receptor.[135, 137] Selecting the appropriate molecular docking strategy that avoids common pitfalls is critical for accurate modeling of drug-bound complexes.[138]

Molecular docking is a computational approach that is used to predict the interactions and affinities of small molecules to receptors.[135, 139, 140] Currently, there are over 60 molecular docking programs that are available for academic or commercial research.[141] Molecular docking is typically performed by searching for likely binding sites and then evaluating them using a scoring function, which is often related to the binding free energy.[140, 142] The ligand-bound receptor structures can be clustered by root mean square deviation (RMSD) and their molecular interactions are analyzed.[135, 140]

One of the most popular molecular docking programs is AutoDock Vina.[143] AutoDock Vina has a much higher speed and accuracy than its predecessor AutoDock 4.0 and has comparable performance to other molecular docking packages such as Schrödinger GLIDE and DOCK.[143–145] The software also excels at "blind" molecular docking compared to other molecular docking packages.[146] Because the software is open source, several variants have been produced of the original program. [e.g. 147, 148] A limitation of the software is that it can only make a limited number of residues flexible and the molecular docking is not performed in solvent.[143] Another disadvantage is that distance restraints cannot be used to drive molecular docking simulations for this software.

There are molecular docking programs designed to use experimental data for docking such as Schrödinger GLIDE[149] and the High Ambiguity Driven biomolecular DOCKing (HADDOCK) protocol.[150–153] These software programs incorporate experimental information through distance restraints.[149, 150] HADDOCK is advantageous over Schrödinger GLIDE because it can do protein-protein docking simulations in addition to small molecule-receptor docking.[150, 152–157] The HADDOCK protocol also utilizes explicit Boolean operators, such as AND, OR and NOT, in the distance restraints within the Crystallography & NMR System (CNS).[151] Distance restraint Boolean operators allow the HADDOCK protocol and CNS to implement ambiguous distance restraints, which are useful for integrating experimental data into molecular docking simulations.[150] A disadvantage of the current implementation of the HADDOCK protocol is that most distance restraints must be inputted manually with the exception of site-directed mutagenesis in the web-based interface.[158] Therefore, inputting distance restraints can be tedious for complex molecular interactions.

Recently, a user-friendly web-based interface was developed for the HADDOCK software called the HADDOCK 2.2 webserver.[158] This webserver has four types of interfaces to allow researchers to dock proteins and small molecules: Easy, Prediction, Expert and Guru. The Guru interface of the HADDOCK 2.2 webserver provides access to over 500 parameters that are available within HADDOCK and CNS. An easy way to input a wide range of experimental data into the HADDOCK 2.2 webserver would be ideal. Currently, the webserver only provides a simplified input for residues that are involved in binding to the ligand and protein in the Easy and Prediction interfaces. As far as we are aware, all other types of experimental data must be manually inputted as distance restraints and requires minimum of Expert level access to the server. The webserver also automatically protonates residues through the Reduce program from

Molprobity.[159, 160] A limitation of the Reduce program is that the pH of the protonated system cannot be changed.[159]

The molecular docking with HADDOCK protocols proceeds in three successive stages. First, the stochastic rigid-body docking stage starts with the ligands in random orientations separated in space followed by steepest descent energy minimization (EM) combined with translation. In the next stage, both semi-flexible and fully flexible simulated annealing (SA) is performed for molecular docking in torsional angle space with the side chains and backbone both flexible. A final flexible molecular dynamics (MD) refinement is performed in Cartesian space with an 8-Å shell of explicit TIP3P model waters[161] (MD_{water}). The process of EM, SA and MD_{water} will be referred to as HADDOCK refinement in the rest of the manuscript. The final drug bound receptor models are clustered and scored on a per-cluster basis using a HADDOCK score (H_{score}), which is related to the intermolecular binding energy. One of the main disadvantages of the initial rigid docking stage is that the ligand starts in random positions around the receptor. Distance restraints between the small molecule and receptor can help guide it to the binding site, but ligands can become energetically "stuck" far from their actual drug binding sites or are unable to penetrate a drug binding site that is in a narrow crevasse. To overcome this issue, a buried binding site protocol for HADDOCK was developed.[157] However, at least for the ligands investigated in,[157] they had relatively high RMSD's around 4 Å with the corresponding ligand-bound X-ray crystal structure.

To overcome these limitations, a hybrid molecular docking wrapper called HADDOCK-Vina was developed. HADDOCK-Vina was designed with an experimental input tailored specifically to small molecule-receptor interactions that allows users to enter experimental data in a simplified format that outputs complex CNS-compatible distance restraint files. A sample

showing all the different types of experimental inputs is provided as Figure 3.S4 of the *Supplementary Information*. The correct protonation state of the receptor at a given pH is achieved through integration with the PROPKA[162, 163] software called within the HADDOCK-Vina wrapper. A customized version of AutoDock Vina was incorporated into the HADDOCK-Vina wrapper and used for the initial docking stage. In the manuscript, the effectiveness of ligand docking with the HADDOCK-Vina wrapper is demonstrated against three ligand-bound X-ray crystal protein structures and experimental data derived from the literature. The accuracy of our hybrid molecular docking approach is compared to molecular docking with HADDOCK and AutoDock Vina by themselves.

3.3 MATERIALS AND METHODS

3.3.1 Calculation of Root Mean Square Deviation (RMSD)

The RMSD between docked ligands were calculated using the definition and equation of RMSD by AutoDock Vina.[143] Essentially, the RMSD was measured between equivalent heavy (i.e. non-proton) atoms as an approximate overlap between ligand-bound conformations. This approach is advantageous because it can deal with chemically equivalent nuclei that have different atom names specified in the structure file. Another advantage of this approach is that it allows one to calculate the RMSD between molecules that are not necessarily identical. The algorithm also ignores the proton positions to deal with free rotation of proton atoms in functional groups such as methyls.

3.3.2 Calculation of Q-factor

A *Q*-factor (*Q*), which is analogous to an *R*-factor in X-ray crystallography, has been used to determine the relative agreement between calculated NMR NOE and relaxation rates to the corresponding experimental values.[37,81] The equation can be rearranged to allow determination of the correspondence between distances from the docked structures (r_{model}) to the inputted or calculated distances (r_{calc}) that have r^6 distance dependence:

$$Q = \sqrt{\frac{\sum_i \{r_{calc}^{-6}(i) - r_{model}^{-6}(i)\}^2}{\sum_i r_{calc}^{-6}(i)^2}} \quad (1)$$

An advantage of using this equation is that it can be used to calculate *Q*-factor from multiple experimental sources such as NMR NOE and relaxation rates, if there is an r^6 distance dependence with a given NMR measurement. However, Equation 1 does not consider the maximum and minimum values of r_{calc} as defined by r_{min} and r_{max}. Since *Q*-factor is analogous to *R*-factor in X-ray crystallography,[37,81] this uncertainty in the distance or NOE would be equivalent to uncertainty in X-ray crystallography, which is reflected in lower resolution electron density maps or high B-factors.[82] As long as the model fits within the low resolution electron density map, the calculated *R*-factor will be low.[82] Thus, a r_{model} that is within the r_{min} and r_{max} will have a Q-factor of 0, since the r_{model} is within the uncertainty of the distances. In this case, Equation 1 will be:

$$Q = \begin{cases} \sqrt{\dfrac{\sum_i \left(r_{\max}^{-6}(i) - r_{\text{model}}^{-6}(i) \right)^2}{\sum_i \left(r_{\max}^{-6}(i) \right)^2}} & r_{\text{model}} > r_{\max} \\ 0 & r_{\min} \le r_{\text{model}} \le r_{\max} \\ \sqrt{\dfrac{\sum_i \left(r_{\min}^{-6}(i) - r_{\text{model}}^{-6}(i) \right)^2}{\sum_i \left(r_{\min}^{-6}(i) \right)^2}} & r_{\text{model}} \le r_{\min} \end{cases} \quad (2)$$

In Equation 2, when the r_{model} is greater than r_{max}, the Q-factor will be calculated using the r_{max} in place of the r_{calc} in Equation 1. The Q-factor will be calculated using the r_{min} in place of r_{calc}, when the r_{model} is less than r_{min} in Equation 2.

3.3.3 Brief overview of the HADDOCK-Vina wrapper.

All simulations and calculations in the enclosed study were performed on a customized 48-CPU node Linux cluster with 2.2 GHz AMD Opteron 6174 processors and 64 Gbytes of RAM running openSUSE Leap 42.2. The software was also tested on an 8-CPU 2.8 GHz Intel Xeon processor and 12 Gbytes of RAM running CentOS 6.9, on the 32-CPU node NMRbox account running Ubuntu 16.04,[164] and on a 4 CPU VirtualBox virtual machine with 2 Gbytes of RAM with 100 Gbytes of hard drive space running Centos 7.4. The virtual machine configuration is our recommended minimum requirement to run HADDOCK-Vina. Figure 3.1 illustrates a flowchart of HADDOCK-Vina. This wrapper seamlessly integrates several software programs including the HADDOCK protocol with CNS,[150, 151] AutoDock Vina,[143] and PROPKA[162, 163] using scripts written in the Python programming language version 2.7. More than 20,000 lines of Python code were required to make the software programs compatible with each other. Additional

parameterization of the PARALLHDG force field[165] was also required to allow for protonation states of amino acid residues that are specified in the PROPKA program.[166, 167] In addition, an experimental input written in the Python programming language was designed to convert experimental data in a simplified format into complex distance restraints for CNS.

Briefly, a receptor file is acquired and processed in Figure 3.1A to remove unwanted protein subunits, ligands, ions, etc, which are specified by the user (Figure 3.S1). The receptor file is then protonated using the PROPKA software in Figure 3.1B.[162, 163] In Figure 3.1C, a ligand with appropriate parameters is acquired from the GlycoBioChem PRODRG2 Server.[168] The ligand is then docked to the receptor using a customized version of AutoDock Vina[143] in Figure 3.1D. The experimental data is entered into the experimental input and is used to generate a set of complex distance restraints within HADDOCK-Vina (Figure 3.1E). This experimental input uses a simplified language to greatly reduce the number of lines needed to produce a set of distance restraints (Figure 3.S4 in the *Supplementary Information*). These distance restraints are used to filter docked structures produced from AutoDock Vina (Figure 3.1F). By default, the docked structures with the top three H_{score}*'s* that follow specific H_{score} and energetic criteria are subjected to MD in CNS (Figure 3.1G). By default, HADDOCK-Vina recommended docked conformations from AutoDock Vina are energy minimized (EM) without randomizing their positions by HADDOCK in CNS. Then the ligand receptor complexes are subjected to three stages of SA in rigid and flexible stages. In the last stage, a MD simulation is performed in explicit waters (MD_{water}) on the docked structure. The final docked structures are clustered and analyzed by HADDOCK-Vina in Figure 3.1H. A detailed description of the default HADDOCK protocol parameters have been previously described.[150, 158]

3.3.4 Installing and Using HADDOCK-Vina

To use HADDOCK-Vina, appropriate software licenses for HADDOCK 2.2[150] and CNS[151] must be obtained. The compressed HADDOCK and CNS folders obtained from their respective parties are transferred into the HADDOCK-Vina wrapper. Compiled versions of AutoDock Vina,[143] MSMS,[169] PROPKA,[166, 167] Reduce,[159] and Volume Assessor[170] are provided within the wrapper. To prevent overloading, a variant of the job scheduling and dispatching software called Xeduler (http://xed.ch/project/xeduler/) is also provided within the HADDOCK-Vina wrapper. An installer is provided within the wrapper to automatically install all programs associated with HADDOCK-Vina, compile HADDOCK and CNS, and start the Xeduler variant.

Once HADDOCK-Vina is installed on the system and the paths are appropriately set, the user changes to the directory that they want to do the simulations in. The user executes the *HADDOCKVina* command, which spawns two directories: *Parameters* and *Scenarios*. The *Parameters* directory contains subdirectories that contain the parameters, Python scripts and configuration files that are needed to run the wrapper. A list of these directories and a description of their function are provided in Table 3.S1 of the *Supplementary Information*. The idea behind the *Scenarios* directory is to have a group of Python Scripts such as Step01, Step02, etc… to sequentially perform specific tasks and to simplify the process of performing the molecular docking simulation. In the Scenarios directory, there are three subdirectories: *Run_HaddockVina*, *Run_Haddock_Only* and *Run_Vina_Only*. The *Run_Haddock_Only* and *Run_Vina_Only* have Python scripts for running HADDOCK and AutoDock Vina by themselves. The names and descriptions of these Python scripts are provided in Tables 3.S2 and 3.S3 of the *Supplementary Information*, respectively. Simulations in these directories are intended for comparison with *HADDOCK-Vina* molecular docking simulations. In the *Run_HaddockVina* directory are the

Python scripts used to perform the hybrid molecular docking simulations. The names and descriptions of these Python scripts are provided in Table 3.S4 of the *Supplementary Information*.

3.3.5 Processing of the receptor files

The HADDOCK-Vina wrapper (Figure 3.1A) automatically processes the receptor protein file. This step is important to ensure correct parameterization for simulating in AutoDock Vina and CNS with the HADDOCK protocols. HADDOCK-Vina can accept files directly from the PDB database[171] or preprocessed PDB formatted files. The specified subunit, pH, metal and metal charge are specified in a *Receptor.txt* file (Figure 3.S1 of the *Supplementary Information*). All protons, ions and non-standard amino acids that are not parameterized by HADDOCK-Vina are removed from the receptor file. Currently, parameterization and detection of cysteine ligated hemes that are found in cytochromes P450 are provided in the current version of HADDOCK-Vina. In the future, there are plans to detect and parameterize a wide range of non-standard amino acids, covalently-linked amino acid labels, prosthetic groups and metals.

After the processing stage, the standard amino acids of the unprotonated receptor file are reprotonated at a specific pH (e.g. 7.4) using PROPKA 3.0 software[166, 167] and a PARSE force field called automatically within the HADDOCK-Vina wrapper (Figure 3.1B). PROPKA was chosen over other pKa prediction programs because of its relatively high accuracy.[162, 163] The PROPKA protonated amino acids and non-standard amino acids are combined into a single protonated receptor file by HADDOCK-Vina. This combined receptor file is converted into separate files that are compatible with the AutoDock Vina and the CNS programs. For use with AutoDock Vina, the receptor is converted into an appropriate PDBQT format with AutoDock Tools (ADT)[172] within the HADDOCK-Vina wrapper, but without affecting the PROPKA protonated state of the

receptor. AutoDock Vina does not use partial charges for its atoms, but treats atom charges as bond donors and acceptors, or essentially +1 or -1.[143] To approximate metal charge on the receptor file, additional dummy metal atoms were added to the AutoDock Vina receptor file in HADDOCK-Vina, as done previously.[35] To be compatible with the HADDOCK protocols and CNS, amino acids in different protonation states were given unique names. For example, a HIS protonated at the δN was named HSD. Also, specific proton names defined by PROPKA differed from the HADDOCK protocol. Therefore, they were changed throughout the receptor file with a Python function within the HADDOCK-Vina wrapper.

3.3.6 Processing of the ligand file

A ligand can be made with a molecular editor such as Avogadro,[174] be obtained from PubChem[95] or extracted from the PDB file using a Python utility script that we developed called *ExtractLigand*. This ligand file is submitted to the GlycoBioChem PRODRG2 Server in PDB, MDL Molfile or SYBYL Mol2 formats with full charges and energy minimization.[168] (Figure 3.1C) The DRGCNS.PAR, DRGCNS.TOP and DRGFIN.PDB files produced from the server are used with HADDOCK-Vina. A Python function within the HADDOCK-Vina wrapper removes an extraneous CNS command section in the DRGCNS.PAR file that interferes with electrostatic calculations by CNS. HADDOCK-Vina wrapper converts DRGFIN.PDB into a compatible format and defines the rotatable bonds of the ligand for rigid receptor molecular docking with AutoDock Vina by calling ADT.[172] The ligand PDB file extracted using the *ExtractLigand* Python script from the X-ray crystal structure was renamed TEMPLATE.PDB and was used to calculate the RMSD with the HADDOCK-Vina docked positions. These PDB files and the PRODRG parameter and topology files are specified in the *Ligand.txt* file in the ligand directory (Figure S2).

3.3.7 Rigid molecular docking of the ligand

In HADDOCK-Vina, flexible ligands are docked to a rigid protonated receptor using a customized version of AutoDock Vina (Figure 3.1D). The purpose of this customized version of AutoDock Vina is to populate the surfaces, grooves, and crevasses of the receptor with ligand binding sites. In the customized AutoDock Vina version, the number of outputted structures is equal to the number of binding modes. The default number of possible binding modes is 1000, which will coat the surface of most receptors with ligands. To increase ligand coverage, users can either increase the number of binding sites on the receptor or change the weights in the *vina.conf* file.

The rest of the parameters in the customized version of AutoDock Vina are the same as the original. In addition, the center and bounds of the search box can be set manually or automatically in the *Box.txt* file (Figure S3) based on the center and the dimensions of the receptor plus 10% to ensure complete coverage of the receptor. After docking, HADDOCK-Vina outputs a PyMOL script called *BindingAnalysis.pml* that can be used to visualize the coverage of the AutoDock Vina docked ligands to the receptor in PyMOL (Schrödinger, Cambridge, MA). The bound ligand structures are reprotonated using Babel 2.3[175] within the wrapper. HADDOCK-Vina wrapper also changes the atom names of the reprotonated file to match the original PRODRG files.[168]

3.3.8 Building distance restraints from the experimental input

Within the HADDOCK-Vina wrapper, there is an experimental input that allows one to enter a range of experimental data in a simplified language without the necessity of manually inputting the distance restraints between the small molecule and the receptor (Figure S4). This experimental input greatly simplifies the process of producing a distance restraint file for small

molecule-receptor interactions. The wrapper converts the experimental input that corresponds to experimental data such as site directed mutagenesis into CNS-compatible distance restraints. An example of the experimental input is shown as Figure 3.S4 in the *Supplementary Information*. This example has the 24 lines specifying experimental data that produces an CNS-compatible distance restraints file with over 8,000 lines. To deal with distances within ambiguous distance restraints, the distance restraint (d) is a "r^{-6}-summed" distance and is represented mathematically as:[15,40]

$$d = \left(\sum r_{ij}^{-6}\right)^{-\frac{1}{6}} \quad (3)$$

where r_{ij} are the individual intermolecular distances. Because of this summing, d is weighted toward the shortest distance. For example, if you have a thousand 5 Å intermolecular distances and one 2 Å intermolecular distance encompassing an ambiguous distance restraint, d will be 1.52 Å. This relationship should be kept in mind when setting the upper bound (r_{max}) of an ambiguous distance restraint. Previous studies set the r_{max}'s of the ambiguous distance restraint to 5 Å or 2 Å.[150, 176] In the HADDOCK-Vina wrapper, the default r_{max}'s depend on the type of experimental data defining the distance restraint.

The experimental input allows users to weigh different types of experimental data using two types of weights. There are weights that reduce the r_{max} of the distance restraint. This type of weight would be used when the distance is not explicitly given. For example, this is the case for entering site-directed mutagenesis data in the experimental input where only residues and weights are given. The other type of weight increases the number of distance restraints, which effectively increases the effective force constant of that distance restraint. For example, two distance restraints that have a force constant of 50 kcal mol^{-1} Å^{-2} will have a combined effective force constant of 100 kcal mol^{-1} Å^{-2}. This weight is used when the distances are explicitly provided in the experimental input, such as intramolecular NOEs.

There is a section labeled [MUT] in the experimental input that will take site-directed mutagenesis data as residue numbers and weights (Fig S4). This is analogous to the residue input of the Easy and Prediction interfaces of the HADDOCK 2.2 webserver.[158] Based on the fact that coulombic interaction energy between charges falls off at ~12 Å,[177] we considered this the maximum distance that a mutation can directly affect ligand binding. The number of distances associated with an ambiguous distance restraint between a residue and a ligand can range from the 100's to the low thousands. Because of the r^{-6} summation in Equation 3, the r_{eff} can easily be half this distance. Therefore, the default r_{max} of the distance restraint was set to 6 Å. Changing the weight applied to the [MUT] experimental restraints affect the r_{max} of the distance restraint. For example, a weight of 1 maintains the [MUT] default r_{max} of 6 Å and a weight of 2 changes that r_{max} of the distance restraint to 3 Å. This weighting is useful when one mutant has a larger effect on ligand binding affinity, while another mutant has a more modest effect.

There are sections for chemical shift and saturation transfer difference (STD) NMR data called [CS] and [STD], respectively. Only the ligand protons that exhibit saturation transfer or a chemical shift need to be defined in these sections. Non-exchangeable protons on the receptor are going to exhibit the most efficient saturation transfer[178] and the strongest effects on chemical shift, so these surface protons are determined with MSMS[169] program called within the HADDOCK-Vina wrapper. Experimental data inputted within these sections generate ambiguous distance restraint(s) between the ligand and all the non-exchangeable surface protons with the receptor. Example of experimental inputs in the [CS] and [STD] sections are provided in Figure S4. Determining the surface protons with MSMS[169] greatly reduces the number of distances associated with an ambiguous distance restraint and greatly improves the speed, but can still result in several thousand distances. Because of the weighting effect of the intermolecular distances on *d* by

Equation 3, the default r_{max} for this category of restraints was set to 3 Å. The r_{max} of these distance restraints can be reduced by increasing the weight.

There is a section called [HSQC] (Figure S4) in the experimental input that specifies shifts in receptor nuclei from heteronuclear single-quantum correlation (HSQC) or any other NMR experiment where receptor nuclei shift in the presence of small molecules. Experimental data within this section generates ambiguous distance restraint(s) between the shifted nuclei and the atoms of the ligand with default r_{max} of 6 Å using the same rational for this distance as the [MUT] section. Like the [MUT] section, this r_{max} can be reduced by dividing by the weight.

In drug development, functional groups of a drug or ligand that may be critical for molecular recognition by the receptor can be identified through structure activity relationships (SARs). [e.g. 179] This information can be inputted into the [SAR] section of the experimental input (Figure S4). From the experimental input, HADDOCK-Vina builds ambiguous distance restraints between specified atoms on the ligand and nuclei on the surface of the receptor determined using the MSMS[169] software that is called within the HADDOCK-Vina wrapper. Because a single ambiguous distance restraint produced in this section can be comprised of thousands of distances, the default r_{max} of the distance restraint was set to 2 Å. Increasing the weight will reduce the r_{max} of the distance restraint, like the [MUT], [CS], [STD] and [HSQC] sections of the experimental input.

The experimental input with HADDOCK-Vina also has sections for paramagnetic relaxation enhancement (PRE) NMR ([PRENMR]), hydrogen (H)-bonding or intermolecular NOEs ([HBOND]), intramolecular nuclear Overhauser effect (NOE) data ([NOE]) and exchange-transferred NOE ([EXNOE]) (Figure S4). The [PRENMR] section has an offset to deal with flexible paramagnetic labels if the restraint is put on the connected residue and not the label. This

would be useful in cases when there is no parameterization of the label for the docking simulation. The [HBOND] section can be used in cases of intermolecular NOEs or for known hydrogen bonds. Because distances specified in the [EXNOE] and [NOE] sections do not reflect specific intermolecular interactions, these restraints are put into a separate CNS-compatible distance restraint table called *unambig.tbl*. Because these restraints are in a separate file, these distance restraints are not used to calculate intermolecular distance restraint violation energy or E_{dist}.

Within the distance restraint, there are three distance parameters are specified in these sections: r_{min}, r and r_{max}. The r_{min} and the r_{max} reflect the upper and lower bounds of the distance restraint, respectively. These parameters are correlated to the r_{minus}, r and r_{plus} of the distance restraint file used by CNS program within the HADDOCK protocols.[151] The r's are equivalent between the experimental input and distance restraint file of CNS. The r_{minus} term in the CNS distance restraint file is the difference between r and r_{min} of the experimental input. The r_{plus} term in the CNS distance restraint file is the difference between r_{max} and r in the experimental input. The weights are also treated differently in this section than the previous sections since the distances are explicitly stated for these distance restraints. The weight reflects the copy number of the distance restraints, which affects the effective force constant of that restraint. For example, a weight of 2 will duplicate the distance restraint and a weight of 4 will make four copies of the distance restraint. If the force constant of a single distance restraint is 10 kcal mol^{-1} $Å^{-2}$, the effective force constant with a weight of 2 and 4 will be 20 kcal mol^{-1} $Å^{-2}$ and 40 kcal mol^{-1} $Å^{-2}$, respectively.

Finally, there is also a section with the experimental input for producing ambiguous distance restraints in the absence experimental data in a section labeled [SOI] for "sphere of influence" (Figure S2). The idea behind this section is to keep ligands constrained within a

specified distance of specific residues. It is analogous to the search box used in molecular docking programs such as AutoDock Vina[143]. In this section one specifies the residue numbers, the r_{max} of the distance restraint and a weight. The weights determine the number of copies of the distance restraint.

3.3.9 Experimentally filtering AutoDock Vina docked conformations

The distance restraints produced by the experimental input are used as an experimental filter for the docked conformations of the rigid docking stage by AutoDock Vina using the HADDOCK protocol (Figure 3.1F). In other words, AutoDock Vina docked conformations are selected based on their correspondence to the experimentally-derived distance restraints through a HADDOCK Score (H_{score}),[150] but are not actually restrained or constrained at this stage. Typically, the H_{score} reflects a combination of interaction weighted intermolecular energies such as Van der Waals (E_{vdw}), electrostatic (E_{elect}), and intermolecular distance restraint violation energy.[150] The H_{score} calculation utilizes the PARALLHDG force field parameters and OPLSX for non-bonded parameters, as previously described.[150, 165, 180, 181] By default, electrostatics are calculated with a constant dielectric and an epsilon scaling constant of 10, optimized for soluble proteins.[182] The ambiguous interaction restraint violation energy (E_{air}) term from HADDOCK was intended for ambiguous intermolecular distance restraints defined under the *ambig.tbl* of the HADDOCK protocols. In the HADDOCK-Vina wrapper, this energy term serves as intermolecular distance restraint violation energy, which encompasses both intermolecular ambiguous and unambiguous distance restraints. Therefore, to be more accurate with our usage, E_{air} term will be referred to E_{dist} in the manuscript. For the initial rigid AutoDock Vina stage, HADDOCK scoring is used to filter docked conformations that correlate well to experimental data. Thus, by default, the energetic

components, E_{vdw}, E_{elect}, and desolvation energies (E_{desolv}), are set to 0 and the H_{score} is equal to E_{dist} below:

$$H_{score}(AutoDockVina) = 0.0 \cdot E_{vdw} + 0.0 \cdot E_{elect} + 0.0 \cdot E_{desolv} + 1.0 \cdot E_{dist} \quad (4)$$

Using this Equation, docked conformations that correlate well to experimental data will have a $E_{dist} = 0$ and thus a $H_{score} = 0$ using Equation 4. After filtering, HADDOCK-Vina produces an MS Excel file in comma-separated values (*csv*) format called *Analysis.csv* with the interaction energies sorted by H_{score} and the E_{total} (E_{total}=$E_{VDW} + E_{elect} + E_{desolv}$). A PyMOL script called *Haddock.pml* is produced to allow visualization of the H_{score} ranking of the experimentally-filtered ligands. One can also determine the RMSD between a template ligand defined in *Ligand.txt* (Figure S2) and the AutoDock Vina docked ligand conformations.

3.3.10 HADDOCK Refinement

By default, the top three scored AutoDock Vina ligand-bound conformations using Equation 4 are analyzed by the wrapper. The wrapper then recommends ligand-bound conformations for further HADDOCK refinement (Figure 3.1F) using the following criteria. If any bound structures in the top three had a $H_{score} = 0$, they are selected over bound structures with an $H_{score} > 0$. For those structures with a H_{score}=0, those that have an E_{total}<0 indicating favorable binding are recommended over those with an E_{total}>0. Finally, if none of the structures has an H_{score}=0, then structures with an $E_{dist} < 1.37$ kcal mol^{-1}, which reflects less than a ten-fold shift in E_{dist} at standard temperature and pressure, and the E_{total}<0 are chosen. If the ligand-bound conformations fit none of these criteria, then no recommendations are given by the wrapper for HADDOCK refinement. In this case, one can adjust the experimentally-derived distance restraints,

adjust the *vina.conf* parameters or force the wrapper to use ligand bound structures for HADDOCK refinement by altering the *Run.txt* file used by the wrapper.

The ligand-bound conformations recommended by HADDOCK-Vina or manually specified by the user are subjected to HADDOCK refinement. The H_{score} is weighted during the EM, SA and MD_{water} stages of HADDOCK refinement. Appropriate weighting for the different stages of HADDOCK refinement has not been resolved for a given docking pair. Our intention was to not bias the molecular docking toward one of the energetic terms such as E_{vdw}, so we assumed that a ligand bound in the correct position on the receptor will have the same E_{total} regardless of the stage of HADDOCK refinement. Therefore, binding energetics (i.e. E_{vdw}, E_{elect} and E_{desolv}) were calculated for 20 ligand X-ray crystal structures in Table 3.S5 for the different stages of HADDOCK refinement. This was accomplished by running HADDOCK refinement for a single step at 0K to prevent the X-ray crystallographic ligands from being significantly perturbed by the simulation. The weights were normalized, so that the E_{total} of the *EM* and *SA* stages of HADDOCK refinement were the same as the MD_{water} stage. Our analysis in Table 3.S5 revealed that there was considerable variation in the E_{vdw}, E_{elect} and E_{desolv} for ligand binding between the ligand-bound structures. By normalizing the weights for energies of the MD_{water} stage to 1.0, the HADDOCK score weights were calculated for the EM and SA stages. In general, the E_{vdw} and E_{desolv} in the EM and SA stages were very similar to the MD_{water} stage. In contrast, the E_{elect} was an order of magnitude lower for the EM and SA stages (Table 3.S5). This is likely due to differences in the epsilon scaling constant between the stages. By default, the weight of the E_{dist} for all the stages was set to 10 to bias the molecular docking toward the experimental data. Based on Table 3.S5, the following weights for H_{score} were used for HADDOCK refinement:

$$H_{score}(EM) = 1.03 \cdot E_{vdw} + 12.47 \cdot E_{elect} + 1.40 \cdot E_{desolv} + 10.0 \cdot E_{dist} \quad (5)$$

$$H_{score}(SA) = 0.99 \cdot E_{vdw} + 12.94 \cdot E_{elect} + 1.04 \cdot E_{desolv} + 10.0 \cdot E_{dist} \quad (6)$$

$$H_{score}(MD_{water}) = 1.00 \cdot E_{vdw} + 1.00 \cdot E_{elect} + 1.00 \cdot E_{desolv} + 10.0 \cdot E_{dist} \quad (7)$$

3.3.11 Analysis of the HADDOCK refined docked ligand conformations

In HADDOCK-Vina, the binding energy, RMSD's, and clustering are determined for the docked structures (Figure 3.1H). The binding energies, which include E_{vdw}, E_{elect}, E_{desolv}, buried surface area (BSA), and H_{scores} are extracted from the HADDOCK directories and compiled into a csv formatted MS Excel file called *Analysis.csv*. A PyMOL script called *HADDOCK.pml* is also produced to allow visualization of ligand binding modes and H_{score} ranking during different stages of the simulation. An RMSD can be calculated for the ligand binding modes versus a template ligand such as an X-ray crystallographic ligand, as specified in the *Ligand.txt* file (Figure S2).

Receptor-bound ligand molecules were clustered using an RMSD-based algorithm[183] that is similar to the one implemented in AutoDock Vina.[143] The RMSD cutoff for clustering was set to 2 Å, which is typical for small molecule clustering. [e.g. 184] The bound ligand molecule with the lowest H_{score} as defined by Equation 7 defines the initial cluster. An RMSD is measured between this molecule and other bound ligand molecules from the HADDOCK-Vina simulation in order of H_{score}. The first bound molecule that is outside this RMSD cutoff defines the lowest H_{score} molecule for the next round of clustering. This algorithm is continued until all the molecules are in a cluster.

The fractional overlap ($f_{overlap}$) between a template ligand and the clusters is determined by calculating the van der Waals volume of the cluster ($V_{cluster}$), template ligand and their combined volumes ($V_{combined}$) using the Volume Assessor program[170] called within the HADDOCK-Vina wrapper and the following Equation:

$$f_{overlap} = 1 - \frac{V_{combined} - V_{cluster}}{V_{cluster}} \quad (8)$$

HADDOCK-Vina cluster analysis identifies the most populated cluster, the cluster that has the greatest fractional overlap to the original ligand-bound X-ray crystal structure, the lowest H_{score} docked ligand conformation within each cluster, the energetics and RMSDs between the docked ligand and the template ligand. This information is outputted to the tab-delimited *Clusters.txt* file that can be imported into the MS Excel program. These values can then be averaged in MS Excel as seen in Table 3.S6.

3.4 RESULTS AND DISCUSSION

3.4.1 Effectiveness of HADDOCK-Vina for small molecule docking

As a benchmark, molecular docking was performed on three ligand-bound X-ray crystal structures: 1) Human Abl tyrosine kinase in complex with imatinib (PDB ID: 2HYY[185]), 2) Human Immunodeficiency Virus type 1 (HIV-1) protease in complex with the inhibitor KNI-272 (PDB ID: 3FX5[186]) and 3) *E. coli* adenylate kinase (Adk) in complex with adenosine monophosphate (AMP) (PDB ID: 1ANK[187]). These X-ray crystal structures were chosen because experimental data associated with these structures can be used to drive the molecular docking. [e.g. 188] To determine the effect of experimentally-derived distance restraints, molecular docking was performed with distance restraints from a single experimental and combined experimental sources.

To minimize molecular docking bias for this benchmark, the ligands used for molecular docking in this benchmark were obtained from PubChem[95] rather than extracted directly from the X-ray crystal structure using the ExtractLigand Python script. The loops and missing atoms were added back into the receptor by running the homology modelling program MODELLER.[189, 190] Processing the receptor with MODELLER relaxes the side chains within the ligand binding site

but does not significantly perturb the C_α backbone. The side chain relaxation helps reduce the molecular docking bias from using a ligand-bound protein X-ray crystal structures with the residue side chains oriented to coordinate with the ligand. MODELLER processing does affect the overall orientation and position of the protein receptor. Therefore, to compare with the ligand-bound X-ray crystal structure, the MODELLER processed receptor was realigned with the original protein X-ray crystal protein structure. To accurately compare molecular docking of the MODELLER processed receptor with the ligand-bound X-ray crystal structure, C_α from residues within 4 Å of the ligand were identified. Using a Python utility script that we developed called *AlignBindingSite*, the corresponding C_α atoms of the MODELLER processed receptor were then aligned with the same atoms on the X-ray crystal structure. The RMSD after alignment was typically around 0.1 Å.

The quantitative results of molecular docking for the lowest H_{score} clusters are shown in Table 3.1. The left most column shows the protein and ligand names. These columns are followed by the experimental input that was used from a single experimental source or combined experimental sources such as site-directed mutagenesis, HSQC NMR and intermolecular NOEs. After this column is the lowest H_{score} that defines the initial ligand cluster (a.k.a. lowest H_{score} cluster). This column is followed by the average H_{score} of that cluster and the average binding interaction energetics: E_{vdw}, E_{elect}, E_{desolv}, E_{dist} and E_{total}. That last column contains the averaged RMSD values from the cluster. A complete table of all the cluster energetics is provided in Table 3.S6 in the *Supplementary Information*.

The molecular docking results are visualized in Figure 2. In the left most column (Figs. 2A, 2E and 2I) are the average RMSDs from the lowest H_{score} cluster derived from a single or combined experimental sources. The experimental input with the lowest RMSD is designated with

a pound symbol. The next column (Figs. 2b, 2f and 2j) shows the distribution of ligands within the lowest RMSD experimental input (i.e. #) and their fractional overlap with the ligand from the protein X-ray crystal structure. The cluster with the lowest H_{score} is labeled with an asterisk. The next column (Figs. 2c, 2g and 2k) shows the location of the cluster produced from the ligand with the lowest H_{score} with the ligand X-ray crystal structure shown in the background. The last column (Figs. 2d, 2h and 2l) shows the lowest RMSD ligand from the lowest H_{score} cluster overlaying the ligand X-ray crystal structure.

3.4.2 Case #1: Human Abl tyrosine kinase domain in complex with imatinib.

Human Abl tyrosine kinase is involved in cell signaling, differentiation and cell division.[185, 191] Mutated Abl tyrosine kinase was identified in chronic myelogenous leukemia (CML), forming the fusion protein BCR-ABL. Thus, the small molecule imatinib is a potent inhibitor of Abl tyrosine kinase activity. [e.g. 192] An X-ray crystal structure of Abl tyrosine kinase has been solved with the anti-cancer drug inhibitor imatinib (PDB ID: 2HYY[185]). Mutations of Abl tyrosine kinase and NMR experimental data were used as experimental restraints to dock imatinib onto Abl tyrosine kinase. The experimental input used to dock imatinib to the Abl tyrosine kinase domain is shown in Figure S5. Many cancer patients have been identified to have mutations in Abl tyrosine kinase rendering them resistant to imatinib. These individuals develop a C→T point mutation, resulting in a substitution of threonine at position 315 to isoleucine, serine or glycine. *In vitro*, this gatekeeper mutation leads to a more than a 20-fold increase in the IC_{50} of imatinib with mammalian baf3 cells.[188] This residue was added to the site-directed mutagenesis section (i.e. [MUT]) and given a weight of two in the experimental input due to the severity of this mutation (Figure S5). Five additional variants, L248R/V, G250E/R, Y253H/C/F, E255K and

E279K, were identified to increase the IC_{50} of imatinib in mammalian baf3 cells, but not as significantly as T315I/S/G.[188] Therefore, these residues were given a weight of 1 in the site-directed mutagenesis section of the experimental input (Figure S5). Binding of imatinib to Abl tyrosine kinase was found to induce relatively large shifts in the ^{1}H-^{15}N amide resonances of A385 and N316 in the ^{1}H-^{15}N transverse relaxation-optimized spectroscopy (TROSY) NMR spectrum compared to other residues on the receptor, suggesting protein-ligand interactions.[195] This information was put into the HSQC NMR section of the experimental input (Figure S5). For the initial docking stage, the repulsion weight for AutoDock Vina docking was reduced to 0.3 to increase the coverage of the kinase.

HADDOCK-Vina molecular docking of imatinib to Abl tyrosine kinase using only site-directed mutagenesis-derived distance restraints is shown in the top row of Table 3.1. Only the two top scored ligand-bound conformations out of 11 were recommended for HADDOCK refinement. They had H_{score}'s of 0 and E_{total}'s-1.89 and -6.79 kcal mol^{-1}. From HADDOCK refinement, two clusters were produced from the distance restraints produced from the site-directed mutagenesis data. The lowest H_{score} cluster had an average E_{total} -151±18.1 kcal mol^{-1} and the model distances were very favorable, with an E_{dist} of 0.00021±0.003 kcal mol^{-1} (Table 3.1). However, this cluster was not overlapping with the imatinib molecule of the X-ray crystal structure of Abl tyrosine kinase, with an average RMSD of 4.46±0.22 Å (Table 3.1).

The HSQC NMR data was used as an experimental input to dock imatinib onto Abl tyrosine kinase.[195] The top three AutoDock Vina ligand bound conformations were recommended by the wrapper for HADDOCK refinement because they all had an H_{score}=0 and E_{total}<0. After HADDOCK refinement, 4 clusters were produced 4 (Table 3.S6). The average H_{score} for the lowest H_{score} cluster was higher compared to the docking using only site-directed mutagenesis data (Table

3.1). The average E_{total} for the lowest H_{score} cluster was -153±28.6 kcal mol^{-1}, which was similar E_{total} of the lowest H_{score} cluster using site-directed mutagenesis data. Additionally, the cluster fit well within the experimental data, with an average E_{dist} of 0.04±0.20 kcal mol^{-1} suggesting a good correlation to the experimental data (Table 3.1). However, the cluster is still relatively far from the X-ray crystal structure ligand with an average RMSD of 4.22±0.43 Å (Table 3.1).

Lastly, the site-directed mutagenesis and HSQC NMR experimental data were combined to dock imatinib to the Abl tyrosine kinase. Only the top ligand bound conformation was selected by the wrapper for HADDOCK refinement because it was the only one with and H_{score} of 0 and $E_{total} < 0$ (i.e. -39.6 kcal mol^{-1}). This one molecule produced 4 clusters (Table 3.S6). The lowest H_{score} cluster was the most populated, with 264 docked imatinib molecules (Table 3.1). The average E_{total} for the cluster was -199±56.9 kcal mol^{-1}, which was more favorable than the docking with either site-directed mutagenesis or HSQC NMR derived distance restraints (Table 3.1). Combining the experimentally-derived distance restraints slightly increased the E_{dist} to 0.61±1.35 kcal mol^{-1}, but still reflects a good correlation to the experimental data. The average RMSD for this cluster was 2.62±0.54 Å, which was a significant improvement from using distance restraints derived from single experimental sources.

The effect of combining the experimentally-derived distance restraints is shown in Figure 2A. The decrease in the average RMSD using a combination of distance restraints shows that restraints derived from individual experimental sources cooperatively direct the imatinib molecule into the correct binding site of the Abl tyrosine kinase. The 4 clusters produced from the combination simulation is shown in Figure 2B. The most populated cluster was the lowest H_{score} cluster or cluster #1. The fractional overlap of this cluster with the X-ray crystal structure imatinib volume was 88%. This cluster encases the X-ray crystal structure of imatinib as shown in Figure

2C. There was significant ambiguity in the position of the piperazine ring among imatinib molecules in the cluster, implying dynamics for that functional group. The is consistent a pharmacophore study of tyrosine kinase inhibitors that showed that this functional group is not essential for potency.[192] The imatinib molecule in the cluster with the lowest RMSD of 1.76 Å with respect to the X-ray crystal structure ligand is shown in Figure 2D. The largest atomic deviations between the ligands occurs with the pyridine and the piperazine functional groups at the ends of the molecules.

3.4.3 Case #2: Human Immunodeficiency Virus type 1 (HIV-1) protease in complex with the inhibitor KNI-272.

HIV-1 protease is a homodimer aspartyl protease that is essential to the life-cycle of HIV.[196] This protein is a major drug target due to its essential role in HIV pathogenesis.[197] To aid in HIV drug development, X-ray crystallography and NMR spectroscopy was used to investigate the interactions of inhibitors with the protease.[186, 196, 198, 199] In one X-ray crystallographic study, HIV-1 protease was solved with the potent protease inhibitor KNI-272 (PDB ID: 3FX5[186]), which has also been investigated by NMR.[199] KNI-272 was docked to the HIV-1 protease X-ray crystal structure using the NMR data as experimental restraints, as described below. The experimental input used to dock KNI-272 onto HIV-1 protease is shown in Figure 3.S6 of the *Supplementary Information*.

KNI-272 was found to cause significant shifts in the ^{1}H-^{15}N HSQC NMR peaks corresponding to D25, D125, V82 and V182 of the HIV-1 protease,[199] which were used as experimental restraints in the HSQC NMR section of the experimental input (Figure S6). The KNI-272 molecule was also found to induce significant chemical shifts in ^{13}C NMR peaks

corresponding to several aspartic residues of the protease: D29, D30, D60 D129, D130 and D160.[199] Of these residues, D30 and D130 residues showed the largest ^{13}C shifts with the protease inhibitor. As a result, the weight for these residues was doubled in the experimental input (Figure S6), which decreases the r_{max} to 3 Å. A number of intermolecular NOEs were observed between protons of KNI-272 and amide protons of aspartic residues, D29, D30, D129 and D130.[199] These were inputted into the [HBOND] section of the experimental input (Figure S6). For the initial rigid docking of KNI-272 onto HIV-1 Protease, the coverage of ligand to the receptor by AutoDock Vina was improved by decreasing the repulsion weight to 0.3 and by increasing electrostatic weight (a.k.a. hydrogen weight) to 0.2 to increase ligand penetration and promote electrostatic interactions between the ligand and the receptor, respectively.

The influence of the experimentally-derived distance restraints from HSQC NMR data and intermolecular NOEs were examined. The first row of the HIV-1 protease section in Table 3.1 shows the results of docking KNI-272 with the protease using distance restraints derived from HSQC NMR data. To improve coverage of the receptor by KNI-272 molecules, the binding site number parameter was increased to 2,000. HADDOCK-Vina wrapper recommended the top two ligand bound conformations for HADDOCK refinement (H_{score}=0 and E_{total}<0). The lowest H_{score} cluster produced from HADDOCK refinement had an average E_{total} of -122.0±18.3 kcal mol^{-1} and E_{dist} of 0.56±0.57 kcal mol^{-1}, which indicates a good correlation to experimental data (Table 3.1). However, the positions of molecules within the cluster and the X-ray crystal structure ligand were good to fair with an average RMSD of 2.70±0.19 Å.

Results from using intermolecular NOE's derived distance restraints for molecular docking are shown in the second row of the HIV-1 protease section of Table 3.1. Because the unambiguous distance restraints from this docking simulation were between specific atoms of the receptor and

the ligand, they were expected to lead to a more accurate model of the KNI-272-HIV-1 protease complex. Only the top AutoDock Vina ligand bound conformation was recommended for HADDOCK refinement by the wrapper with an H_{score}=0 and an E_{total}<0. From these simulations, two clusters were produced from docking with the intermolecular NOEs. The average E_{dist} of the lowest H_{score} cluster was 0.01±0.03 kcal mol^{-1} and had an average Q-Factor of 0.044±0.035, which indicates excellent agreement of the molecules within the cluster and the experimental data. However, the average E_{total} of this cluster was slightly higher at -106±20.8 kcal mol^{-1} compared with the average E_{total} using only HSQC NMR data (Table 3.1). However, the average RMSD between molecules in the cluster and the X-ray crystal structure KNI-272 improved compared to the HSQC NMR only simulation, with a value of 2.36±0.17 Å.

In the last row of the HIV-1 protease section, the experimental sources were combined. From the AutoDock Vina initial docking, only the top ligand bound molecule was recommended by HADDOCK-Vina. The H_{score} using Equation 4 was positive for this molecule; however, the E_{dist}<1.37 kcal mol^{-1} and the E_{total} was -3.31 kcal mol^{-1}. HADDOCK refinement using a hundred copies of this molecule produced two clusters. The average Q-Factor of the lowest H_{score} cluster between the calculated and model distances was essentially zero (i.e. 0.0004±0.0005), showing that the model fit well within the defined distance restraints of the intermolecular NOEs (Table 3.1 1). The average E_{total} of the lowest H_{score} cluster was -121±18.4 kcal mol^{-1} (Table 3.1). The average E_{dist} was positive at 1.94±1.21 kcal mol^{-1}, suggesting some opposition between the HSQC- and intermolecular NOE-derived distance restraints. The average RMSD was even better for this cluster at 2.03±0.09 Å, when compared to clusters formed using distance restraints from individual experiments.

Figure 2E shows the average RMSD's between molecules within the lowest H_{score} cluster and the X-ray crystal structure KNI-272 using distance restraints from single and combined experimental sources. Molecular docking simulations combining the HSQC NMR data with the intermolecular NOE derived distance restraints inputted in the [HBOND] section of the experimental input had the lowest average RMSD to the X-ray crystal structure KNI-272. The clustering of the molecular docking simulation combining the experimental data is shown in Figure 2F. The $f_{overlap}$ of the cluster with the X-ray crystal structure KNI-272 was almost complete at 85% (Figure 2G and Table 3.S6). The RMSD KNI-272 molecule with the lowest RMSD of 1.82 Å within the cluster is shown in Figure 2H with the X-ray crystal structure KNI-272 and the surface of the protease is shown in the background. This relatively low RMSD shows that this docked ligand occupies the same space as the X-ray crystallographic ligand. However, there were some deviations of the functional groups and the backbone between the docked ligand and the X-ray crystallographic ligand.

3.4.4 Case #3: *E. coli* adenylate kinase (Adk) in complex with adenosine monophosphate (AMP).

Adenylate kinase (Adk) is an integral enzyme in cellular homeostasis by regulating the concentration of intracellular ATP, ADP and AMP.[200] The relatively small size of Adk and its expression in a variety of species has made it an ideal model for protein structure, function and dynamics.[201] An X-ray crystal structure of Adk from *E. Coli* was solved in the presence of two substrates: adenosine monophosphate (AMP) and a non-hydrolyzable ATP analog (PDB ID: 1ANK[187]). The HADDOCK-Vina approach was employed to dock AMP onto *E. coli* Adk using site-directed mutagenesis data from porcine Adk[202] and ^{1}H-^{15}N HSQC chemical shift

perturbations.[201] The experimental input combining all the experimental data used to dock AMP onto *E. Coli* Adk is shown in Figure 3.S7 in the *Supplementary Information*.

The steady state kinetics of mutants and wildtype porcine Adk was examined in the presence of AMP.[202] Of the mutants evaluated, the R44M, R97M and R149M mutants resulted in a ~30-fold or greater increase in the apparent K_m of AMP,[202] so these residues were used as restraints in the experimental input (Figure S7). Using a protein amino acid sequence alignment, the porcine residues correspond to R36, R88 and R167 on *E. coli* Adk. These residues were defined in the site-directed mutagenesis section of the experimental input and given a weight of 2 because of their large effect on the K_m of AMP (Figure S7). The weight of 2 effectively reduces the r_{max} of these site-directed mutagenesis related distance restraints to 3 Å. AMP also caused significant shifts in the ^{1}H-^{15}N HSQC NMR peaks of several residues of *E. coli* Adk.[201] The largest ^{1}H-^{15}N chemical shifts were observed in D61, L63 and A66 of *E. coli* Adk.[201] These three residues were entered into the HSQC NMR section of the experimental input (Figure S7). To improve ligand coverage during the initial rigid docking by AutoDock Vina, the repulsion weight was decreased to 0.3.

Results from using the site-directed mutagenesis data are shown in the first row of the Adk section of Table 3.1. Because the top three AutoDock Vina ligand conformations had a $H_{score} = 0$ (Equation 1) and an $E_{total}<0$, these AMP molecules were recommended by the wrapper for HADDOCK refinement. The molecular docking simulation produced 20 AMP clusters (Table 3.1 and Table 3.S6). The lowest H_{score} (Equation 7) AMP cluster had an average E_{total} of -324±39.0 kcal mol^{-1} and was populated with 63 molecules. This cluster also correlated well with experimental data, since E_{dist} was equal to 0. However, the cluster did not overlap well with the X-ray crystal structure AMP, with an average RMSD of 5.83±0.42 Å.

The next row in the Adk section of Table 3.1 1 shows the molecular docking simulation using HSQC NMR-derived distance restraints. None of the ligand bound conformations were recommended from the initial rigid AutoDock Vina docking by the wrapper because the H_{score}>0 and E_{dist}>1.37 kcal mol^{-1}. Changing AutoDock Vina parameters in the *vina.conf* file did not improve ligand coverage of Adk. However, the E_{total} for the top H_{score} ligand (Equation 4) was favorable with a value of -17.2 kcal mol^{-1}. Therefore, this ligand was manually selected for HADDOCK refinement within the *Run.txt* file that is used by the wrapper. HADDOCK refinement produced a single cluster with a hundred molecules. The lowest H_{score} cluster had an average E_{total} of -266±45.7 kcal mol^{-1} and an average E_{dist} of 0.33±0.37 kcal mol^{-1} indicating good correspondence to the experimental data. The average RMSD with respect to the X-ray crystal structure ligand was 1.53±0.17 Å.

A molecular docking simulation of AMP was performed with the distance restraints derived from both site-directed mutagenesis and the HSQC NMR data. The top two scored (Equation 4) ligand bound AutoDock Vina conformations were recommended for HADDOCK refinement (H_{score}=0, E_{total}<0). The lowest H_{score} cluster had an average E_{total} of -279±42.7 kcal mol^{-1}, indicating more favorable binding compared to the molecular docking simulation only using distance restraints derived from HSQC NMR data. The correlation to experimental data was similar to the molecular docking simulation using HSQC NMR data, with an average low E_{dist} of 0.31±0.38 kcal mol^{-1} and a correspondingly low average RMSD of 1.55±0.15.

Figure 2I shows the average RMSD from molecular docking as a function of the experimental restraint. The molecular simulation using HSQC NMR data produced a similar average RMSD as the simulation using all the experimental data. The molecular docking simulation combining experimental data produced a more favorable E_{total} than the molecular

docking simulation using HSQC NMR data (Table 3.1). The $f_{overlap}$ was also higher at 93% versus 89% using only the HSQC NMR data (Table 3.S6).

Docking AMP with the combined distance restraints produced two clusters that are shown in Figure 2J. The lowest H_{score} AMP cluster (H_{score}=-358 A.U. (Equation 7)) contained 99 AMP molecules while the second cluster produced from a H_{score} of -251 had only one AMP molecule. The lowest H_{score} cluster in Figure 2J is shown in panel K with the AMP-bound X-ray crystal structure of Adk. The cluster completely encloses the AMP molecule in the X-ray crystal structure, with a $f_{overlap}$ of 93%. From this cluster, the lowest RMSD AMP molecule is shown with the X-ray crystal structure AMP and a surface of Adk shown in the background. This molecule has an RMSD of 1.26 Å with the X-ray crystal structure AMP and shows small deviations in the sugar and nucleotide functional group.

3.4.5 Characteristics of HADDOCK-Vina molecular docking simulations that accurately reproduce X-ray crystal structures

In the previous section, molecular docking with HADDOCK-Vina was used to accurately reproduce ligand-bound X-ray crystal protein structures in three cases. The molecular docking was performed with distance restraints derived from a single experimental source or a combination of experimental sources. The most accurate models did not necessarily have the lowest in magnitude E_{total} as shown by the very low average E_{total} from several clusters for docking AMP to the Adk enzyme using only site-directed mutagenesis derived distance restraints. Although H_{score} is a useful tool to rank bound conformations within a molecular docking simulation, the H_{score} value for ligands between different simulations did not necessarily correspond to the most accurate ligand-bound receptor models. However, in the three molecular docking cases that we presented, the most

accurate models produced by HADDOCK-Vina were the lowest H_{scored} cluster (i.e. cluster #1). When there were multiple clusters present, the lowest H_{score} cluster had a significantly higher population than the rest of the clusters as we saw in all three cases. In contrast, when the ligand populations were distributed among several clusters, the average RMSD of those clusters were high, as we saw with the docking of AMP to the *E. coli* Adk with only site-directed mutagenesis derived distance restraints in Case #3. Finally, the ligand-bound receptor models that correlated best to the ligand-bound X-ray crystal structures had low E_{dist}, as seen in all three cases. This does not mean that the E_{dist} must equal 0, since experimentally-derived distance restraints from different experimental sources can sometimes oppose each other leading to an E_{dist} that is slightly positive.

3.4.6 The accuracy of our hybrid molecular docking with HADDOCK-Vina versus HADDOCK or AutoDock Vina.

The effectiveness of the HADDOCK-Vina approach was determined by comparing molecular docking with HADDOCK-Vina with HADDOCK and AutoDock Vina alone. Python scripts are provided with the HADDOCK-Vina wrapper to run HADDOCK and AutoDock Vina independently, within the *Run_Haddock_Only* (Table 3.S2) and *Run_Vina_Only* (Table 3.S3) directories of the wrapper, to allow users to make similar comparisons. All the HADDOCK-Vina parameters used in this comparison were default, recommended by the wrapper based on the H_{score}, E_{dist} and E_{total} criteria or used to increase ligand coverage of the receptor by AutoDock Vina. In kind, the parameters used to dock with HADDOCK protocols and AutoDock Vina program independently were mostly their default parameters. For molecular docking with HADDACK protocols alone, this includes the 1000 ligands energy minimized from random starting orientations in space followed by HADDOCK refinement of the 200 best scored ligand-bound structures. The

only non-default parameters used for docking with HADDOCK were the H_{score}, which was calculated using Equations 5, 6, and 7 for the EM, SA and MD_{water} stages, respectively, allowing direct comparison with the HADDOCK-Vina wrapper. For the docking simulation with AutoDock Vina, a search box that encompasses the entire protein receptor was used, like the one in the HADDOCK-Vina docking simulations. To compare with molecular docking of HADDOCK-Vina, Equation 7 instead of Equation 4 was used to calculate the H_{score} for the ligands docked with AutoDock Vina. Also, in this comparison, the receptors were not protonated using PROPKA,[162, 163] but were protonated by Reduce[159], which is used in the HADDOCK 2.2 webserver, and ADT[172] for the AutoDock Vina simulation. Lastly, the experimental inputs that were used to produce the lowest RMSD clusters in Figure 2 were used for the HADDOCK and AutoDock Vina simulations.

To compare HADDOCK-Vina to HADDOCK and AutoDock Vina, the lowest H_{score} cluster in the HADDOCK molecular docking simulation and the ligand-bound conformation with the lowest AutoDock score in AutoDock Vina simulations were evaluated. Table 3.2 shows the results of the molecular docking for the various software programs. The left three columns show the PDB ID, the protein and the ligand. The middle column shows the approach used for docking: HADDOCK-Vina, HADDOCK and AutoDock Vina. The next two columns are the lowest H_{score} and lowest E_{total}. The fractional overlap between the cluster with the lowest H_{score} ligand and the ligand from the X-ray crystal structure are shown in the next column for the HADDOCK-Vina and HADDOCK simulations. The final column shows the lowest RMSD between the best scored ligand and the X-ray crystal structure ligand from these molecular docking approaches.

The lowest H_{score} clusters using the HADDOCK-Vina wrapper from Figure 2C, 2G and 2K are shown in the first column of Figure 3. The top row of Figure 3 shows the docking of imatinib to the X-ray crystal structure of Abl tyrosine kinase I (PDB ID: 2HYY)[185] by HADDOCK-Vina,

HADDOCK and AutoDock Vina. The cluster shown in Figure 3A of the HADDOCK-Vina simulation shows essentially complete overlap with the imatinib molecule from the X-ray crystal structure. This is compared to the cluster with the lowest H_{score} ligand from the HADDOCK docking simulation, which does not overlap with the X-ray crystal structure ligand (Figure 3B). The best RMSD for ligands in this cluster was 13.7 Å compared with 1.76 Å for the best RMSD in the cluster from the HADDOCK-Vina molecular docking simulation. The lowest scored molecule had poor correlation to experimental data with a H_{score} of 281, but the E_{total} for this ligand was favorable with a value -237 kcal mol^{-1}. The RMSD of the best AutoDock scored ligand in Figure 3C was also large with a value of 5.32 Å. The H_{score} of the ligand (Equation 7) was smaller in magnitude compared to the HADDOCK simulation. Additionally, the E_{total} for the AutoDock Vina simulation was the largest, with 132 kcal mol^{-1}.

The docking of KNI-272 to the X-ray crystal structure of HIV-1 protease (PDB ID: 3FX5)[186] by the three software programs is shown in the middle row of Figure 3. The H_{score} and E_{total} for KNI-272 molecules docked with HADDOCK were larger in magnitude than the E_{total} from HADDOCK-Vina, 1,682 and 87.3 kcal mol^{-1}, respectively (Table 3.2). This is largely due to the orientation of the HADDOCK cluster, which only shares 46% $f_{overlap}$ with the X-ray crystal structure (Figure 3E). The lowest RMSD for ligands docked with HADDOCK were ~2-fold larger than the lowest RMSD ligand docked by HADDOCK-Vina. Docking with AutoDock Vina alone produced an even less accurate model of the ligand-bound receptor than the other two approaches. The lowest H_{score} and lowest E_{total} were unfavorable, with values of 4,451 A.U. and 4,507 kcal mol^{-1}, respectively. However, the AutoDock Vina score for this molecule was favorable, with a score of -8.0 kcal mol^{-1}. This surprisingly large disparity in energy values produced from the HADDOCK protocol and AutoDock Vina reflect differences in the scoring function used in

AutoDock Vina[143] and the method to estimate of E_{total} in HADDOCK.[150] Despite these relatively poor results, the AutoDock Vina docked KNI-272 molecule was still only 2.94 Å from the X-ray crystal structure KNI-272.

In the bottom row of Figure 3, AMP was docked into the X-ray crystal structure of adenylate kinase (PDB ID: 1ANK)[187] by these programs. The fractional overlap of the lowest H_{score} cluster with the X-ray crystal structure ligand was more than three times larger for HADDOCK-Vina than HADDOCK. This lack of overlap with the X-ray crystal structure ligand corresponds to the high RMSD for the best ligand in this cluster, 4.67 Å. Additionally, the H_{score} and E_{total} of the lowest H_{scored} molecule was ~100-fold higher in magnitude compared to the HADDOCK-Vina simulation. The best AutoDock Vina scored ligand had the largest RMSD compared to both HADDOCK-Vina and HADDOCK, with a value of 5.63 Å. The H_{score} and E_{total} for AMP docking to *E. coli* Adk was considerably larger in magnitude, at 489 A.U. and 471 kcal mol^{-1}, respectively, suggesting a poor correlation to the experimental data and docking of AMP to Adk.

Overall, the HADDOCK-Vina wrapper produced considerably more accurate ligand-bound receptor models than HADDOCK or AutoDock Vina by themselves. Fractional overlap between clusters produced by HADDOCK-Vina and the X-ray crystal structure were several times higher than the corresponding cluster in the HADDOCK simulation. In addition, HADDOCK-Vina had the lowest H_{score}*'s* and E_{total}*'s* with the same experimentally-derived distance restraints. There were also large differences in the H_{score}*'s* and E_{total}*'s* for the HADDOCK and AutoDock Vina molecular docking simulations, when compared to the differences in the HADDOCK-Vina molecular docking simulations. HADDOCK protocols have shown great utility in protein-protein docking using experimentally-derived distance restraints[150] and AutoDock Vina has proved itself with blind flexible docking to rigid receptors.[146] However, the HADDOCK-Vina wrapper's

strength over these previous approaches is its ability to dock small flexible molecules to flexible receptors using experimentally-derived distance restraints. Under these specific conditions, the ligand-bound models produced by HADDOCK-Vina were more accurate than ligand-bound models produced from the HADDOCK protocols or AutoDock Vina under similar conditions.

3.5 CONCLUSION

The HADDOCK-Vina wrapper uses a hybrid molecular docking approach for small molecule receptor docking driven by experimentally-derived distance restraints. Essentially, HADDOCK-Vina wrapper fused the rigid molecular docking capabilities of AutoDock Vina with the MD simulation capabilities of CNS through the HADDOCK protocol. This integration provides better control of the initial rigid molecular docking phase when compared with using the HADDOCK protocol by itself. This control allows ligands to penetrate narrow crevasses of the receptor and helps prevent small molecules from becoming energetically "stuck" far from their true binding sites. The HADDOCK-Vina wrapper also integrates features that are not typically found in molecular docking software packages. For example, the protonation state of receptor residues is automatically achieved through the PROPKA pKa prediction software[162, 163] that is called within the HADDOCK-Vina wrapper. There is also an experimental input that converts experimental data such as site-directed mutagenesis and intermolecular NOEs data into CNS-compatible distance restraints between small molecules and receptors. Molecular docking with three test ligands with experimentally-derived distances from a single source or combined sources versus their corresponding ligand-bound X-ray crystal structures provided insight into an appropriate strategy for producing accurate ligand-bound receptor models. The most accurate models came from simulations with the least number of generated clusters and the most populated

cluster had the molecule with the lowest H_{score} determined from Equation 7. We show the advantage of using the HADDOCK-Vina hybrid molecular docking approach versus doing small molecule docking with HADDOCK or AutoDock Vina. Overall, the HADDOCK-Vina wrapper provides a powerful and relatively easy method to use NMR and non-NMR data to produce experimentally correlated ligand-bound receptor models. The HADDOCK-Vina wrapper is provided free of charge to academic institutions by request to the corresponding author.

3.6 TABLES

Table 3.1 Cluster analysis of the lowest H_{score} cluster with HADDOCK-Vina

PDB ID	Protein	Ligand	Restraint Type[a]	Lowest H_{score} (A.U.)[c]	Average H_{score} ± S.D. (A.U.)[d,*]	Average E_{vdw} ± S.D. (kcal/mol)[*]	Average E_{elect} ± S.D. (kcal/mol)[*]	Average E_{desolv} ± S.D. (kcal/mol)[*]	Average E_{dist} ± S.D. (kcal/mol)[*]	Average Q-Factor ± S.D. (A.U.)[e,*]	Average E_{total} ± S.D. (kcal/mol)[f,*]	Average RMSD ± S.D. (Å)[g,*]
2HYY	Abelson tyrosine kinase 1	Imatinib	MUT	-194	-151 ± 18.1	-61.3 ± 4.45	-83.8 ± 16.7	-6.28 ± 5.04	2.1E-4 ± 0.003	--	-151 ± 18.1	4.46 ± 0.22
			HSQC	-245	-153 ± 28.6	-43.6 ± 3.55	-105 ± 27.8	-4.17 ± 5.81	0.04 ± 0.20	--	-153 ± 28.6	4.22 ± 0.43
			Comb	-359	-194 ± 59.5	-57.4 ± 6.89	-139 ± 59.2	-3.83 ± 5.60	0.61 ± 1.35	--	-199 ± 56.9	2.62 ± 0.54
3FX5	HIV-1 Protease	KNI-272	HSQC	-168	-116 ± 18.3	-78.3 ± 3.94	-44.5 ± 17.1	0.83 ± 3.49	0.56 ± 0.57	--	-122 ± 18.3	2.70 ± 0.19
			HBOND	-153	-106 ± 20.7	-61.2 ± 4.15	-46.0 ± 19.0	1.59 ± 3.84	0.01 ± 0.03	0.04 ± 0.04	-106 ± 20.8	2.36 ± 0.17
			Comb	-167	-101 ± 23.6	-69.7 ± 3.30	-52.5 ± 16.5	1.36 ± 4.07	1.94 ± 1.21	4.0E-4 ± 5.0E-4	-121 ± 18.4	2.03 ± 0.09
1ANK	Adenylate Kinase	AMP	MUT	-409	-324 ± 39.0	-22.7 ± 5.24	-307 ± 41.5	5.62 ± 4.59	0.00 ± 0.00	--	-324 ± 39.0	5.83 ± 0.42
			HSQC	-372	-263 ± 45.8	-34.1 ± 3.63	-234 ± 44.6	1.14 ± 5.40	0.33 ± 0.37	--	-266 ± 45.7	1.55 ± 0.15
			Comb	-358	-276 ± 44.3	-32.9 ± 3.45	-247 ± 43.4	0.87 ± 5.42	0.31 ± 0.38	--	-279 ± 42.7	1.53 ± 0.17

[*]Average ± standard deviation (S.D.) for each docked structure within the cluster. Clusters with a S.D. of 0 had only one docked structure populating the cluster

[a] Section of the experimental interface used

[b]The fractional overlap for each cluster using equation 8

[c]The lowest HADDOCK score (H_{score}) of the given cluster calculated with equation 7

[d]The average H_{scores} within each cluster

[e]The Q-Factor was calculated for each structure with intermolecular NOE experimental restraints. These values were then averaged for each cluster.

[f]The total intermolecular energy ($E_{total} = E_{vdw}+E_{elect}+E_{desolv}+E_{dist}$) was calculated for each structure and averaged for each cluster

[g]The RMSD was calculated for each structure. These RMSD values were then averaged for each cluste

Table 3.2 Comparing the lowest H_{scored} structure produced by HADDOCK-Vina to HADDOCK and Autodock-Vina.

PDB ID	Protein	Ligand	Software	Lowest H_{score} (A.U.)[a]	Lowest E_{total} (kcal/mol)[b]	Fractional Overlap[c]	RMSD[d]
2HYY	Abelson tyrosine kinase 1	imatinib	HADDOCK-Vina	-359	-359	0.88	1.76
			HADDOCK	281	-237	0.00	13.7
			Vina	148	132	--	5.32
3FX5	HIV-1 Protease	KNI-272	HADDOCK-Vina	-167	-169	0.85	1.82
			HADDOCK	1,682	87.3	0.46	3.29
			Vina	4,451	4,507	--	2.94
1ANK	Adenylate Kinase	AMP	HADDOCK-Vina	-358	-370	0.93	1.26
			HADDOCK	-39.7	-90.0	0.27	4.67
			Vina	489	471	--	5.63

[a]The structure with the lowest HADDOCK score (H_{score}) within the cluster calculated using equation 7.
[b]The total intermolecular energy ($E_{total} = E_{vdw}+E_{elect}+E_{desolv}+E_{dist}$) of the lowest H_{score} structure. Total energy is a sum of the van der Waals, electrostatics and desolvation energy terms from the docking simulation.
[c]The fractional overlap using equation 8.
[d]The lowest RMSD structure within the cluster for HADDOCK-Vina and HADDOCK. For Vina, the structure with the lowest RMSD from the simulation are shown. The RMSD was calculated using equations 1 and 2.

3.7 FIGURES

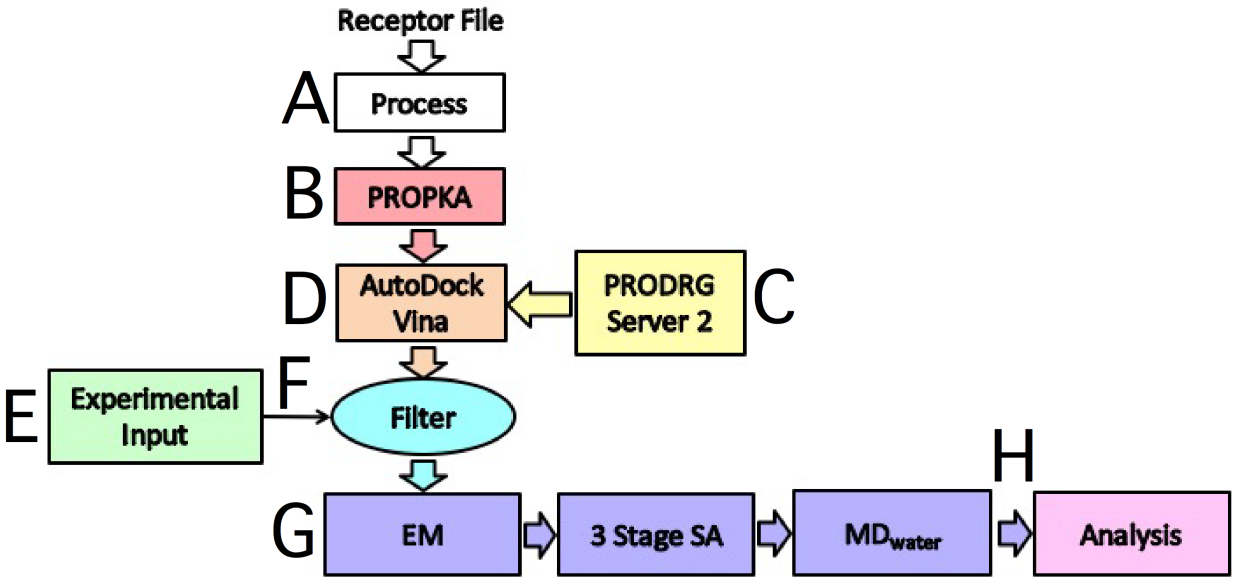

Figure 3.1 Flowchart of molecular docking using the HADDOCK-Vina wrapper. **(A)** The receptor file obtained from the protein data bank or other methods and processed for docking simulations (white). **(B)** The receptor is protonated by the PROPKA software at a specified pH (red). **(C)** A ligand file is obtained and processed through the PRODRG server 2 (yellow). **(D)** A customized version of the AutoDock Vina program was used to penetrate the receptor with ligand (orange). **(E)** An experimental input converts experimental data into ambiguous and unambiguous distance restraints for the HADDOCK-Vina protocol (green). **(F)** The experimentally-derived distance restraints are used to filter AutoDock Vina docked conformations (blue). **(G)** The best experimentally correlated drug-bound conformations, i.e. criteria 1 or 2, are refined further through energy minimization (EM), 3 stage simulated annealing (3 Stage SA), and molecular dynamics in explicit water (MD_{water}) (purple). **(H)** The outputted structures are clustered and analyzed (pink).

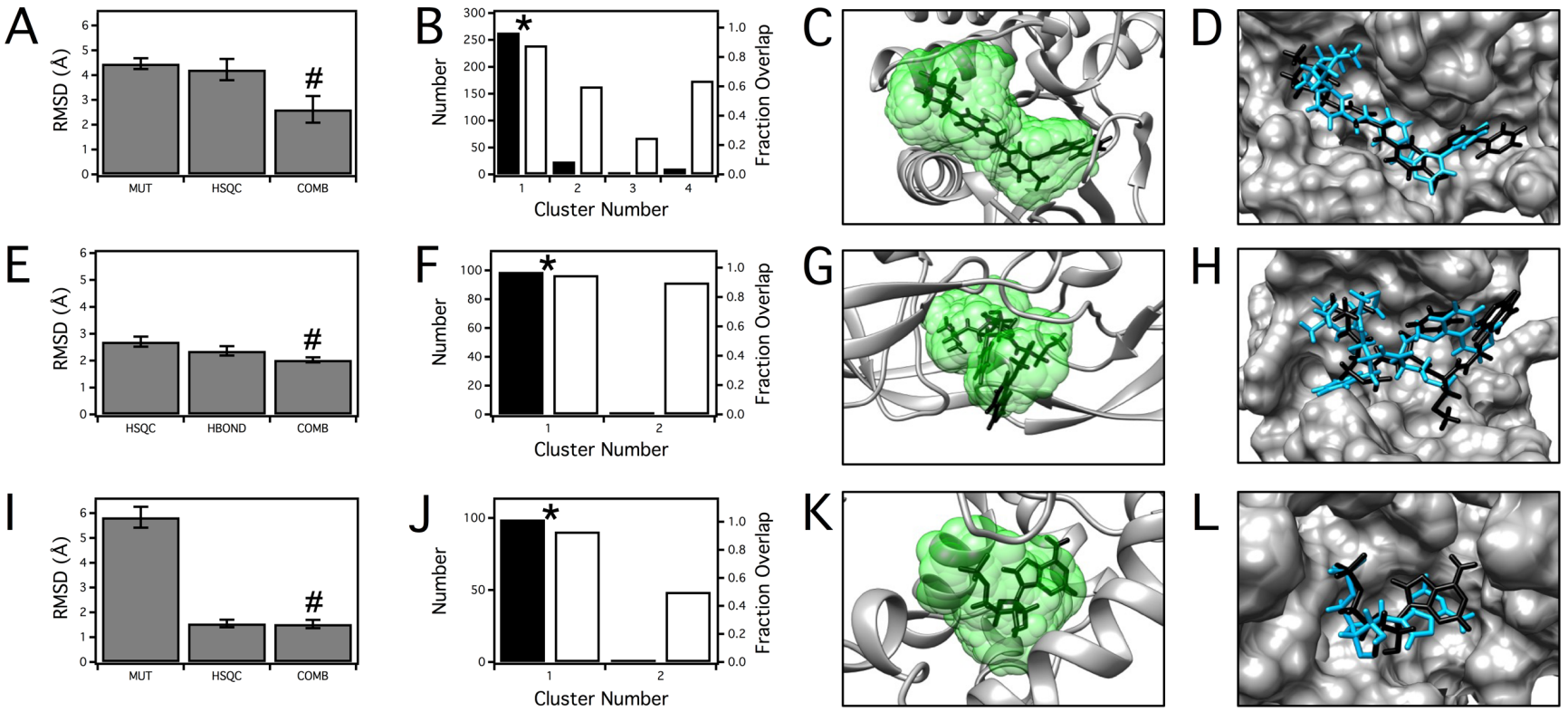

Figure 3.2 Experimentally-driven molecular docking by HADDOCK-Vina versus ligand-bound X-ray crystal structures. The docked conformations produced with the HADDOCK-Vina protocol with **(A-D)** imatinib and Abl tyrosine kinase, **(E-H)** KNI-272 and HIV-1 protease, and **(I-L)** AMP and Adk. **(A,E,I)** The first column of panels show the RMSD of the lowest H_{score} cluster to the original ligand-bound X-ray crystal structure for individual and combined (Comb.) experimental data restraints. The lowest RMSD cluster was used for further analysis and is designated with the pound symbol **(B,F,J)** The next column of panels show the population and $f_{overlap}$ of the cluster associated with the lowest ligand RMSD. The cluster with the lowest H_{score} are designated with an asterisk **(C,G,K)** The lowest H_{score}

clusters from panels B,F,J are transparent green with X-ray crystal structure ligand shown in black, and the protein is shown as a gray ribbon in the background. **(D,H,L)** Gray protein surface with the lowest RMSD ligand (blue) with respect to the X-ray crystallographic ligand (black).

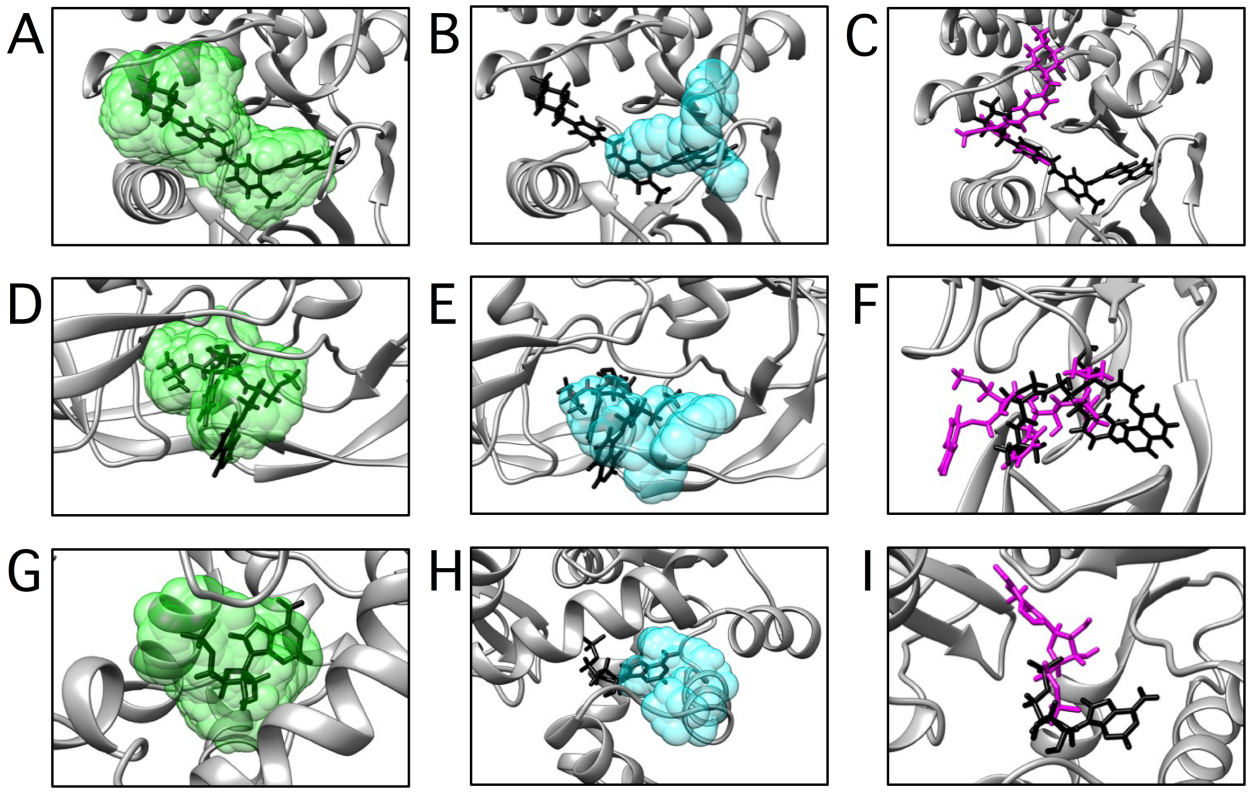

Figure 3.3 Comparison of the hybrid HADDOCK-Vina docking approach to the HADDOCK and AutoDock Vina molecular docking software. The docking output is shown for **(A-C)** imatinib and Abl tyrosine kinase, **(D-F)** KNI-272 and HIV-1 protease, and **(G-I)** AMP and adenylate kinase. **(A,D,G)** The first column shows the lowest H_{score} clusters from HADDOCK-Vina in transparent green. **(B,E,H)** The second column shows the lowest H_{score} cluster from HADDOCK in transparent cyan. **(C,F,I)** The third column shows the lowest scored structure by AutoDock Vina in magenta. In all panels, the original ligand bound X-ray crystal structure is shown in black and the protein is shown as a gray ribbon in the background.

3.8 SUPPLEMENTARY INFORMATION

Table 3.S1: *Subdirectories in the Parameters directory of HADDOCK-Vina.*

Directory	Description
00A-Receptor	The directory contains the receptor PDB formatted file and the receptor parameter file *Receptor.txt* (Figure 3.S1).
00B-Ligand	The directory contains the ligand PDB, parameter and topology files in addition to the ligand parameter file *Ligand.txt* (Figure S2).
00C-VinaConfiguration	The directory contains the AutoDock Vina configuration file *vina.conf* and the search box parameter file *Box.txt* (Figure S3).
00D-Restraints	The directory contains the *Restraints.txt* file (Figure S4) that is used to produce the CNS-compatible distance restraint files *ambig.tbl* and *unambig.tbl*.
00E-HADDOCKScore	The directory contains modified *run.cns.template* and *new.html.template* files for use with HADDOCK. There is also a *Scores.txt* file, which is used to weight the HADDOCK score (H_{score}) (Equation 4 in the manuscript) for ligand-bound structures produced from AutoDock Vina.
00F-HADDOCKRun	The directory contains the run.cns.template and new.html.template files. There are also three HADDOCK parameter files: *Run.txt*, *Setup.txt* and *Score.txt*. *Score.txt* sets weights for H_{score} for the HADDOCK refinement (Equations 5, 6 and 7 in the manuscript). *Run.txt* is for setting the molecular dynamics parameters. *Setup.txt* is for setting up the flexible regions of the receptor and setting up the force constants of the distance restraints.
00G-MDPARM	The directory contains modified CNS files and modified PARALLHDG force field parameter and topology files to be used with CNS.
00H-PyMOL	The directory contains PyMOL templates for use the Python scripts.
00I-VinaConfigurationOnly	The directory is similar to *00C-VinaConfiguration*, except that it is for doing simulations with AutoDock Vina by itself.
00J-HADDOCKRunOnly	The directory is similar to 00F-HADDOCKRun, except that it is for doing simulations with HADDOCK by itself.
00K-HADDOCKScore_VinaOnly	Same as 00E-HADDOCKScore directory, except that the *Score.txt* file uses weighting from Equation 7 in the manuscript by default.
VinaHaddockFunctions	This directory contains Python functions and classes used by the HADDOCK-Vina wrapper.

Table 3.S2: *Python scripts in the Run_Haddock_Only subdirectory for doing simulations with HADDOCK by itself.*

Python Script	Description
Step01-Process_Receptor.py	Python script for processing the ligand and receptor files for HADDOCK. This Python script protonates the receptor PDB file using the Reduce program. The program uses *Receptor.txt* (Figure 3.S1), *Ligand.txt* (Figure S2) and *Box.txt* (Figure S3).
Step02-BuildHaddockRestraints.py	Python script for translating the *Restraints.txt* file (Figure S4) into a distance restraint file that can be used with HADDOCK and CNS.
Step03-RunHaddock.py	Python script for automatically running HADDOCK protocols with CNS.
Step04-Analyze_Haddock.py	Python script for analyzing the energetics of the HADDOCK simulation.
Step05-Calculate_RMSD.py	Python script for calculating the root mean square deviation (RMSD) between the HADDOCK docked conformations and a template ligand specified in *Ligand.txt* (Figure S2).
Step06-Calculate_Qfactor.py	Python script for calculating the Q-factor as defined Equation 2 in the manuscript and the *Restraints.txt* file (*Figure S4*) for distances specified by NMR NOEs and relaxation, and the observed distances from docking in HADDOCK.
Step07-Analyze_Clusters.py	Python script for producing clusters of HADDOCK docked conformations based on the H_{score}.

Table 3.S3: *Python scripts in the Run_Vina_Only subdirectory for doing simulations with AutoDock Vina by itself.*

Python Script	Description
Step01-AutoDockVina.py	Python script to process the ligand and receptor files for AutoDock Vina. The Python script protonates the receptor PDB file using AutoDock Tools (ADT). The program uses *Receptor.txt* (Figure 3.S1), *Ligand.txt* (Figure S2) and *Box.txt* (Figure S3).
Step02-Process_Receptor.py	Python script for processing the ligand and receptor files for *HADDOCK* for use in scoring *AutoDock Vina* docked conformations. The program uses *Receptor.txt* (Figure 3.S1) and *Ligand.txt* (Figure S2).
Step03-BuildHaddockRestraints.py	Python script for translating the *Restraints.txt* file (Figure S4) into a distance restraint file that can be used with HADDOCK and CNS.
Step04-HADDOCKScoreVina.py	Python script for using HADDOCK to score the *AutoDock Vina* docked conformations.
Step05-AnalyzeVina.py	Python script for analyzing the energetics and scoring of the AutoDock Vina docked conformations by HADDOCK.
Step06-Calculate_RMSD.py	Python script for calculating RMSD between the AutoDock Vina docked conformations and a template ligand specified in *Ligand.txt*
Step07-Calculate-Qfactor.py	Python script for calculating the Q-factor as defined Equation 2 in the manuscript and the *Restraints.txt* file (Figure S4) for distances specified by NMR NOEs and relaxation, and the observed distances from the AutoDock Vina docked conformations.

Table 3.S4: *Python scripts in the Run_HaddockVina subdirectory for doing hybrid molecular docking simulations with HADDOCK-Vina.*

Python Script	Description
Step01-AutoDockVina.py	Python script to process the ligand and receptor files for AutoDock Vina and HADDOCK. The Python script protonates the receptor PDB file using the PROPKA program. The program uses *Receptor.txt* (Figure 3.S1), *Ligand.txt* (Figure S2) and *Box.txt* (Figure S3).
Step02-BuildHADDOCKRestraints.py	Python script for translating the *Restraints.txt* file (Figure S4) into a distance restraint file that can be used with HADDOCK and CNS.
Step03-HADDOCKScoreVina.py	Python script for using HADDOCK to score the *AutoDock Vina* docked conformations.
Step04-AnalyzeVina.py	Python script for analyzing the energetics and scoring of the AutoDock Vina docked conformations by HADDOCK.
Step05-Calculate_RMSD.py	Python script for calculating RMSD between the AutoDock Vina docked conformations and a template ligand specified in *Ligand.txt*
Step06-Calculate_QFactor.py	Python script for calculating the Q-factor as defined Equation 2 in the manuscript for distances specified by NMR NOEs and relaxation, and the observed distances from the AutoDock Vina docked conformations.
Step07-RunHaddock.py	Python script to select specific AutoDock Vina docked conformations and run HADDOCK protocols with CNS.
Step08-Analyze_Haddock.py	Python script for analyzing the energetics of the HADDOCK simulation.
Step09-Calculate_RMSD.py	Python script for calculating the root mean square deviation (RMSD) between the HADDOCK docked conformations and a template ligand specified in *Ligand.txt* (Figure S2).
Step10-Calculate-QFactor.py	Python script for calculating the Q-factor as defined Equation 2 in the manuscript and the Restraints.txt file (*Figure S4*) for distances specified by NMR NOEs and relaxation, and the observed distances from docking in HADDOCK.
Step11-Analyze_Clusters.py	Python script for producing clusters of HADDOCK docked conformations based on the H_{score}.

Table 3.S5 CNS calculated binding energies for ligands of 20 X-ray crystal structures

			EM			SA			MD_{water}		
PDB ID	**Protein**	**Ligand**	**E_{vdw} (kcal/mol)**	**E_{elec} (kcal/mol)**	**E_{desolv} (kcal/mol)**	**E_{vdw} (kcal/mol)**	**E_{elec} (kcal/mol)**	**E_{desolv} (kcal/mol)**	**E_{vdw} (kcal/mol)**	**E_{elec} (kcal/mol)**	**E_{desolv} (kcal/mol)**
1ANK	Adenylate Kinase	AMP-PNP	59.67	-20.71	12.85	-48.95	-21.39	12.45	-44.43	-245.5	10.66
1GKY	Guanylate Kinase	GMP	-30.55	-11.31	6.518	-35.41	-10.67	2.673	-26.82	-155.8	2.202
1H4J	Methanol Dehydrogenase	PQQ	-41.78	-3.211	1.389	-47.73	-1.005	2.767	-31.88	-160.1	3.536
1HKB	Hexokinase	G6P	12.68	-9.877	9.398	-7.734	-10.96	8.231	-11.73	-151.5	17.05
1IKW	HIV Reverse Transcriptase	Efavirenz	-29.13	-2.089	-4.056	-34.70	-1.881	-4.215	-39.46	-25.95	-15.66
1JEF	Lysozyme	NAG3	-33.30	-1.681	0.2337	-34.66	-1.585	-0.0722	-34.54	-21.61	0.5090
1Q0Y	Mouse Fab 9B1	Morphine	-28.94	-10.51	-5.823	-18.46	-7.375	-6.599	-17.24	-60.44	-3.605
1RL9	Arginine Kinase	ADP	-52.12	-22.80	9.918	-48.00	-24.33	9.414	-44.77	-285.4	11.75
2H42	Phospho-diesterase-5	Sidenafil	-55.84	-1.136	-3.422	-58.37	-1.312	-3.519	-58.28	-25.22	1.667
2Q72	LeuT	L-leucine	-45.92	-4.869	0.2737	-43.23	-3.247	-1.824	-49.88	-23.86	2.313
2P0W	HAT1	Acetyl-CoA	-63.31	-13.36	3.945	-72.54	-13.50	4.892	-70.30	-165.4	3.581
3B60	MsbA	AMPPNP	10.41	-5.710	7.889	-19.81	-5.799	8.143	-25.23	-110.7	6.803
3Q70	Aspartic Protease	Ritonavir	-71.17	-3.399	1.638	-59.34	-2.384	3.377	-66.95	-20.27	-2.148
4FH2	Beta Lactamase	Sublactam	-0.1374	-12.12	0.4278	-18.58	-7.981	1.588	-10.62	-115.1	6.264
4JK4	Bovine Serum Albumin	Diiodo-salicylic Acid	-19.59	-8.313	-2.093	-23.21	-8.960	1.979	-15.36	-129.3	-5.232
4Q93	Tyrosyl tRNA Synthetase	Resver-atrol	-25.91	0.9841	-0.6178	-32.67	1.086	0.5744	-33.02	9.212	-0.2322
4TS1	Tyrosyl tRNA Synthetase	L-Tyrosine	-3.498	-3.064	-0.0127	-28.30	-3.888	1.024	-24.89	-68.66	-5.938
4X2L	BACE-1	Comp. #6	-12.66	-4.587	-1.027	-20.32	-3.287	-2.260	-22.97	-20.94	2.529
4ZRW	Mincle	Trehalose	-20.39	1.589	7.178	-22.57	1.301	5.322	-20.83	2.934	3.260
5KMH	CavAB	Br-verapamil	-8.444	-0.6058	-0.9907	-7.631	-0.6921	-1.545	-8.137	-7.873	-1.465
Average ± S.D.			**-20.87 ± 29.58**	**-6.839 ± 6.746**	**2.181 ± 5.142**	**-34.11 ± 17.69**	**-6.393 ± 7.013**	**2.120 ± 4.913**	**-32.87 ± 18.25**	**-89.07 ± 84.97**	**1.892 ± 7.032**
Calculated Weights			**1.03**	**12.47**	**1.40**	**0.99**	**12.94**	**1.04**	**1.0**	**1.0**	**1.0**

Table 3.S5: *Average Weights for Energy Terms used in MD Simulations.* The weights applied to the energy terms used for the MD simulations for HADDOCK-Vina were calculated from 20 co-

crystal structures from the PDB. References for the structures are shown in the table. Structures currently in press are designated with an asterisk. The energy terms associated for each MD stage were calculated: energy minimization (it0), simulated annealing (it1) and molecular dynamics with explicit water (it2). The energy terms calculated were van der Waals (E_{vdw}), electrostatics (E_{elec}) and desolvation (E_{desolv}). Energy weights are represented as an average for all energy term

Table 3.S6 Cluster analysis of all clusters produced from HADDOCK-Vina.

X-ray Structure	Restraint Type[a]	Cluster Number	Number in Cluster	$f_{overlap}$[b]	Lowest H_{score} (A.U.)[c]	Average H_{score} ± S.D.[d,*]	Average E_{vdw} ± S.D.* (kcal/mol)	Average E_{elec} ± S.D.* (kcal/mol)	Average E_{desolv} ± S.D.* (kcal/mol)	Average E_{dist} ± S.D.* (kcal/mol)	Average Q-Factor ± S.D. (A.U.)[e,*]	Average E_{total} ± S.D.[f,*] (kcal/mol)	Average RMSD ± S.D.[g,*] (Å)
Abelson Tyrosine Kinase I and Imatinib (2HYY)	MUT	1	143	0.53	-194	-151 ± 18.1	-61.3 ± 4.45	-83.8 ± 16.7	-6.28 ± 5.04	2.1E-4 ± 0.003	--	-151 ± 18.1	4.46 ± 0.22
		2	56	0.45	-184	-155 ± 12.4	-57.1 ± 3.07	-89.4 ± 11.4	-8.33 ± 5.31	0.00 ± 0.00	--	-155 ± 12.4	4.70 ± 0.24
	HSQC	1	52	0.52	-245	-153 ± 28.6	-43.6 ± 3.55	-105 ± 27.8	-4.17 ± 5.81	0.04 ± 0.20	--	-153 ± 28.6	4.22 ± 0.43
		2	45	0.52	-206	-160 ± 23.8	-46.4 ± 3.43	-111 ± 20.9	-2.54 ± 5.96	0.00 ± 0.00	--	-160 ± 23.8	3.93 ± 0.36
		3	1	0.31	-185	-185 ± 0.00	-44.1 ± 0.00	-129 ± 0.00	-12.5 ± 0.00	0.00 ± 0.00	--	-185 ± 0.00	4.22 ± 0.00
		4	2	0.46	-167	-134 ± 45.6	-50.7 ± 6.05	-83.9 ± 39.3	0.11 ± 0.24	0.00 ± 0.00	--	-134 ± 45.6	3.81 ± 0.54
	Comb	1	264	0.88	-359	-194 ± 59.5	-57.4 ± 6.89	-139 ± 59.2	-3.83 ± 5.60	0.61 ± 1.35	--	-199 ± 56.9	2.62 ± 0.54
		2	24	0.60	-203	-165 ± 23.8	-47.4 ± 3.69	-133 ± 19.0	-4.64 ± 4.94	1.99 ± 1.66	--	-183 ± 18.1	3.61 ± 0.21
		3	1	0.25	-202	-202 ± 0.00	-54.8 ± 0.00	-148 ± 0.00	0.72 ± 0.00	0.00 ± 0.00	--	-202 ± 0.00	4.88 ± 0.00
		4	11	0.64	-156	-138 ± 26.1	-61.8 ± 5.99	-74.7 ± 25.0	-5.14 ± 3.80	0.35 ± 0.90	--	-141 ± 22.6	2.77 ± 0.67
HIV-1 Protease and KNI-272 (3FX5)	HSQC	1	100	0.72	-168	-116 ± 18.3	-78.3 ± 3.94	-44.5 ± 17.1	0.83 ± 3.49	0.56 ± 0.57	--	-122 ± 18.3	2.70 ± 0.19
		2	100	0.89	-153	-119 ± 14.0	-75.8 ± 4.56	-54.6 ± 11.3	0.28 ± 4.21	1.06 ± 0.76	--	-130 ± 13.3	1.96 ± 0.08
	HBOND	1	97	0.85	-153	-106 ± 20.7	-61.2 ± 4.15	-46.0 ± 19.0	1.59 ± 3.84	0.01 ± 0.03	0.04 ± 0.04	-106 ± 20.8	2.36 ± 0.17
		2	3	0.67	-153	-95.7 ± 25.4	-61.0 ± 5.56	-37.7 ± 19.6	2.74 ± 0.74	0.03 ± 0.06	0.01 ± 0.01	-96.0 ± 25.1	2.17 ± 0.08
	Comb	1	99	0.85	-167	-101 ± 23.6	-69.7 ± 3.30	-52.5 ± 16.5	1.36 ± 4.07	1.94 ± 1.21	4.05E-4 ± 5.0E-4	-121 ± 18.4	2.03 ± 0.09
		2	1	0.51	-73	-73.0 ± 0.00	-69.3 ± 0.00	-11.8 ± 0.00	4.53 ± 0.00	0.35 ± 0.00	0.00 ± 0.00	-76.6 ± 0.00	2.15 ± 0.00
	MUT	1	63	0.26	-409	-324 ± 39.0	-22.7 ± 5.24	-307 ± 41.5	5.62 ± 4.59	0.00 ± 0.00	--	-324 ± 39.0	5.83 ± 0.42
		2	23	0.34	-391	-249 ± 38.3	-26.0± 2.97	-264 ± 39.4	5.06 ± 4.42	0.00 ± 0.00	--	-285 ± 38.3	4.76 ± 0.34
		3	89	0.75	-385	-282 ± 42.3	-25.4 ± 3.68	-260 ± 43.4	3.27 ± 5.19	0.00 ± 0.00	--	-282 ± 42.3	3.39 ± 0.51
		4	24	0.31	-380	-291 ± 37.1	-26.5 ± 4.51	-267 ± 41.1	3.03 ± 5.20	0.02 ± 0.08	--	-290 ± 36.9	6.09 ± 0.59
		5	69	0.23	-376	-292 ± 38.0	-24.9 ± 4.21	-271 ± 39.6	3.63 ± 5.94	0.01 ± 0.07	--	-292 ± 38.0	5.34 ± 0.44

Adenylate Kinase and AMP (1ANK)		6	1	0.15	-331	-331 ± 0.00	-20.8 ± 0.00	-312 ± 0.00	2.64 ± 0.00	0.00 ± 0.00	--	-331 ± 0.00	4.88 ± 0.00
		7	1	0.03	-329	-329 ± 0.00	-29.7 ± 0.00	-308 ± 0.00	8.91 ± 0.00	0.00 ± 0.00	--	-329 ± 0.00	7.31 ± 0.00
		8	4	0.29	-323	-302 ± 29.5	-23.2 ± 1.85	-286 ± 30.0	7.74 ± 5.01	0.00 ± 0.00	--	-302 ± 29.5	5.11 ± 0.31
		9	6	0.77	-313	-231 ± 51.0	-31.4 ± 3.26	-202 ± 46.9	2.59 ± 3.59	0.00 ± 0.00	--	-231 ± 51.0	1.95 ± 0.25
		10	6	0.17	-308	-271 ± 26.2	-29.2 ± 4.89	-249 ± 25.8	5.21 ± 3.60	0.20 ± 0.45	--	-273 ± 25.4	7.46 ± 0.33
		11	2	0.2	-305	-301 ± 5.71	-241.3 ± 2.79	-284 ± 16.4	4.26 ± 7.91	0.00 ± 0.00	--	-301 ± 5.71	4.88 ± 0.37
		12	2	0.4	-287	-270 ± 24.1	-23.2 ± 0.57	-253 ± 27.9	5.71 ± 3.25	0.00 ± 0.00	--	-270 ± 24.1	4.13 ± 0.88
		13	1	0.11	-282	-282 ± 0.00	-26.0 ± 0.00	-261 ± 0.00	4.55 ± 0.00	0.00 ± 0.00	--	-282 ± 0.00	4.50 ± 0.00
		14	1	0.28	-275	-275 ± 0.00	-24.1 ± 0.00	-156 ± 0.00	-5.65 ± 0.00	0.00 ± 0.00	--	-275 ± 0.00	4.50 ± 0.00
		15	1	0.47	-271	-271 ± 0.00	-38.7 ± 0.00	-233 ± 0.00	0.92 ± 0.00	0.00 ± 0.00	--	-271 ± 0.00	1.99 ± 0.00
		16	1	0.14	-256	-256 ± 0.00	-25.7 ± 0.00	-238 ± 0.00	8.13 ± 0.00	0.01 ± 0.01	--	-256 ± 0.00	8.01 ± 0.00
		17	2	0.32	-247	-197 ± 70.4	-25.1 ± 1.40	-175 ± 71.1	2.84 ± 2.10	0.00 ± 0.00	--	-197 ± 70.4	4.87 ± 0.45
		18	1	0.11	-238	-238 ± 0.00	-26.4 ± 0.00	-220 ± 0.00	7.95 ± 0.00	0.00 ± 0.00	--	-238 ± 0.00	5.29 ± 0.00
		19	1	0.46	-160	-160 ± 0.00	-29.0 ± 0.00	-133 ± 0.00	2.36 ± 0.00	0.00 ± 0.00	--	-160 ± 0.00	3.22 ± 0.00
		20	2	0.44	-113	-113 ± 0.81	-27.2 ± 2.79	-84.9 ± 8.27	-0.80 ± 4.67	0.00 ± 0.00	--	-113 ± 0.81	3.58 ± 0.04
	HSQC	1	100	0.89	-372	-263 ± 45.8	-34.1 ± 3.63	-234 ± 44.6	1.14 ± 5.40	0.33 ± 0.37	--	-266 ± 45.7	1.53 ± 0.17
	ComB	1	99	0.93	-358	-276 ± 44.3	-32.9 ± 3.45	-247 ± 43.4	0.87 ± 5.42	0.31 ± 0.38	--	-279 ± 42.7	1.55 ± 0.15
		2	1	0.5	-251	-251 ± 0.00	-38.2 ± 0.00	-218 ± 0.00	-1.18 ± 0.00	0.61 ± 0.00	--	-255 ± 0.00	1.76 ± 0.00

[*]Average ± standard deviation (S.D.) for each docked structure within the cluster. Clusters with a S.D. of 0 had only one docked structure populating the cluster and has no average.
[a] Section of the experimental interface used.
[b]The fractional overlap for each cluster using Equation 8.
[c]The lowest HADDOCK score (H_{score}) of the given cluster calculated with Equation 7.
[d]The average H_{scores} within each cluster.
[e]The Q-Factor was calculated for each structure with intermolecular NOE experimental restraints. These values were then averaged for each cluster.
[f]The total intermolecular energy ($E_{total} = E_{vdw}+E_{elect}+E_{desolv}+E_{dist}$) was calculated for each structure and averaged for each cluster.
[g]The RMSD was calculated for each structure. These RMSD values were then averaged for each cluster.

Table 3.S6: *Cluster analysis of all clusters produced from HADDOCK-Vina.* All the clusters produced from molecular docking simulations using experimentally derived distance restraints with HADDOCK-Vina were analyzed. In the analysis, the cluster number and the number of

structures in each cluster are reported. The $f_{overlap}$ is also provided for each cluster. The $H_{score,}$ E_{vdw}, E_{elect}, E_{desolv}, E_{dist}, E_{total} and RMSD was averaged for each cluster and provided with a standard deviation (S.D.). In the case where the S.D. is 0.00, there was only one structure populating the cluster.

```
; Receptor information

Receptor 2hyy-random.pdb ; put your receptor PDB file here

; 2hyy-random.pdb is default PDB file
; 2hyy-random.pdb was processed by Modeller,
; where the side chains have been relaxed without ligand present

; Chain to use for molecular docking

Chain First ; put your chain here

; First, Last, or another chain designation such as A

; Simulation pH

pH  7.4

; Metal Parameters

Metal XX ; Metal type

; Label Metal XX or none for no metal
; Other possibilities are CA, FE, MG, MN, CL, BR and ZN
; However, there must be specific parameterization for it within
; HADDOCK.

Metal Charge +0 ; Charge of the metal
```

Figure 3.S1. *Example of the Receptor.txt file*. The comments are shown in black and parameters red by the HADDOCK-Vina wrapper are shown in **red**.

; Ligand Information

Ligand DRGFIN.PDB ; Ligand PDB file

Ligand Parameter DRGCNS.PAR ; Ligand Parameter file

Ligand Topology DRGCNS.TOP ; Ligand Topology file

Ligand Template TEMPLATE.PDB ; Ligand Template file

Figure 3.S2. *Example of the Ligand.txt file*. The comments are shown in black and parameters red by the HADDOCK-Vina wrapper are shown in **red**.

```
; Search Box Parameters

AutoSet Box True

; By setting to true, a box is automatically built
; around the receptor and the parameters below are ignored

Center X 20.0
Center Y 20.0
Center Z 20.0

Size X 60
Size Y 60
Size Z 60
```

Figure 3.S3. *Example of the AutoDock Vina search box parameter file Box.txt.* The comments are shown in black and parameters red by the HADDOCK-Vina wrapper are shown in **red**.

[LIGANDS] ; The section defines the ligand

; Residue Name | Residue Number

STI -1

; If the residue is less than the number of residues in the
; receptor PDB file. The ligand residue number is the number
; of receptor residues + 1.

; Description: Ligand named STI and -1 indicates that the ligand
; number will be equal to number of receptor residues + 1.

[MUT] ; This section defines the residues that interact with the
; receptor

; Residue number(s) | Weight

248 253 256 | 1

; Description: Ambiguous distance restraints with an upper
; bound of 6 Å and a weight of 1 between the ligand and residues ; 248, 253 and 256

[SOI] ; This section defines the residues that will serve
; as non-experimental constraints for the ligand.

; Residue numbers | Constraint Distance (r) | Weight (Integer)

248 253 256 | 8 | 3

; Description: Ambiguous distance restraint with an upper bound ; of 3 Å and a weight of 3 between the ligand and residues 248, ; 253 and 256

[STD] ; This section defines the ligand nuclei that demonstrated
; protein-ligand interactions by STD NMR

; Proton(s) | Weight

H6S H6T H20 | 1

; Description: Ambiguous distance restraint with an upper bound ; of 3 Å and weight of 1 between protons H6S, H6T and H20 on
; the ligand.

[CS] ; This section defines the ligand nuclei that exhibited a
; chemical shift (CS) by NMR

; Proton(s) | Weight

H6S H6T H20 | 1

; Description: Ambiguous distance restraint with an upper bound
; of 3 Å and a weight of 1 between protons H6S, H6T and H20 on
; the ligand.

[PRENMR] ; This section defines the ligand nuclei that exhibited
; changed in relaxation rate in the presence of a
; paramagnetic label by NMR

; Ligand | Receptor | Distance

; Proton(s) | Residue Number , Atom Name | rlower r rupper | Offset | Weight

H6S H6T H20 | 315 HG21 HG22 HG23 | 0.00 4.00 6.00 | 0.00 | 1

; Description: Ambiguous distance restraint with a weight of 1 and no offset between protons H6S, H6T and H20 on imatinib and the paramagnetic center of the methyls on T315 HG21 HG22 HG23

[HBOND] ; This section defines the ligand nuclei and residue
; nuclei between non-paramagnetic nuclei like
; intermolecular NOEs and H-bonds.

; Ligand | Receptor | Distance
; Proton(s) | Residue Number , Atom Name | rlower r rupper | Offset | Weight

H6S H6T H20 | 315 HG21 HG22 HG23 | 0.00 4.00 6.00 | 0.00 | 1

; Description: Ambiguous distance restraint with a weight of 1 and no offset between protons H6S, H6T and H20 of the ligand and the paramagnetic center of the methyls on T315 HG21 HG22 HG23.

[EXNOE] ; This section defines ligand nuclei that display
; exchange transferred NOEs

; Ligand #1 | Ligand #2 | Distance
; Proton(s) | Proton(s) | rlower r rupper | Weight

H11 | H6S H6T H20 | 0.00 3.00 5.00 | 1

; Description: Ambiguous distant restraint with a weight of 1 between H11 and H6S H6T H20 on imatinib

[NOE] ; This section defines receptor intranuclear NOEs

; Receptor | Receptor | Distance
; Residue #, Atom Name | Residue #, Atom Name | rlower r rupper | Weight

429 HD21 HD22 HD23 | 471 HD21 HD22 HD23 | 0.00 3.00 4.00 | 1

; Description: Ambiguous distant restraint with a weight of 1 between the methyls on L429 and L471

[HSQC] ; This section defines HSQC shifted peaks of receptor
; nuclei by NMR

; Residue Number | Atom Number | Weight

248 CD1 HD11 HD12 HD13 | 1

; Description: Ambiguous distance restraint with an upper bound ; of 6 Å and weight of 1 between L248 nuclei and imatinib

[SAR] ; This section defines functional groups of the ligand that

```
    ; are known to interact with the receptor

; Ligand Atom Numbers | Weight

61-end, 38-50, begin-3 | 1

; Description: Ambiguous distance restraint with an upper bound of 2 Å and
; weight of 1 between atoms 61 – end of imatinib, atoms 38-50 and the beginning
; of ligand to atom 3 with the receptor

[END] ; End of the experimental restraint file
```

Figure 3.S4. *Example experimental input and comments for HADDOCK-Vina wrapper showing possible experimental restraints that can be used.* The comments are shown in black and simplified language used to produce the CNS-compatible distance restraints are shown in **red**. These sections are described in detail at http://128.192.148.219/COPClasses/index.php/HADDOCK-Vina.

```
[LIGANDS]
LIG -1

[MUT]
; Reference = Azam et al. Table 3
248 | 1 ; Mutations L248R/V
250 | 1 ; Mutations G250E/R
253 | 1 ; Mutations Y253H/C/F
255 | 1 ; Mutation E255K
279 | 1 ; Mutation E279K
315 | 2 ; Gatekeeper mutations T315I/S/G

[HSQC]
; Reference = Skora et al. Figure 2
385 HN | 1 ; Large shift on A385
316 HN | 1 ; Large shift on N316

[END]
```

Figure 3.S5. *Experimental input used to drive the molecular docking of imatinib onto Abl tyrosine kinase.* The comments are shown in black and the simplified language used to produce the CNS-compatible distance restraints are shown in **red**. These sections are described in detail at http://128.192.148.219/COPClasses/index.php/HADDOCK-Vina.

```
[LIGANDS]
LIG -1

[HBOND]

; Reference = Wang et al. 1996.
; Intermolecular NOEs

HAX HB5 HB6 HB7 HB8 HAY HB9 HCA HAZ | 29 HN | 2.00 5.00 7.00 | 0.00 | 1
; t-butyl of the ligand and the amide of residue #29
HAX HB5 HB6 HB7 HB8 HAY HB9 HCA HAZ | 30 HN | 2.00 5.00 7.00 | 0.00 | 1
; t-butyl of the ligand and the amide of residue #30
HCD HCE HBK | 129 HN | 2.00 5.00 7.00 | 0.00 | 1
; thiol methyl of the ligand and the amide of residue #129
; *Note: amino acid number change from Modeller
HCD HCE HBK | 130 HN | 2.00 5.00 7.00 | 0.00 | 1
; thiol methyl of the ligand and the amide of residue #130
; *Note: amino acid number change from Modeller
HAJ | 29 HN | 2.00 5.00 7.00 | 0.00 | 1
; HAJ proton (proton 22 in the paper) of the ligand and amide of residue #29
HAL | 129 HN | 2.00 5.00 7.00 | 0.00 | 1
; HAL proton (proton 14 in the paper) of the ligand and amide of residue #129
HBJ HCC | 129 HN | 2.00 5.00 7.00 | 0.00 | 1
; HBJ and HCC protons (ambiguous proton 10 in paper) and amide of residue #129

[HSQC]

; Reference = Wang et al. 1996. Figures 2 and 5

25 HN | 1 ; D25 HSQC shift
125 HN | 1 ; D125 HSQC shift
82 HN | 1 ; V82 HSQC shift
182 HN | 1 ; V182 HSQC shift
30 CG | 2  ; Large D30 HSQC shift
130 CG | 2 ; Large D130 HSQC shift
129 CG | 1 ; D129 HSQC shift
29 CG | 1  ; D29 HSQC shift

[END]
```

Figure 3.S6. *Experimental input used to drive the molecular docking of KNI-272 onto HIV-1 Protease.* The comments are shown in black and the simplified language used to produce the CNS-compatible distance restraints are shown in **red**. These sections are described in detail at http://128.192.148.219/COPClasses/index.php/HADDOCK-Vina.

```
[LIGANDS]
LIG -1

[MUT]

; Reference = Tsai and Yan. 1991.

36 | 2  ;  44R in Tsai and Yan 1991
88 | 2  ; 97R in Tsai and Yan 1991
167 | 2  ; 149R in Tsai and Yan 1991

[HSQC]

; Reference = Aden et al. 2007. Figure 3A

61 HN | 1 ; Large HSQC shift on D61
63 HN | 1 ; Large HSQC shift on L63
66 HN | 1 ; Large HSQC shift on I66

[END]
```

Figure 3.S7. *Experimental input used to drive the molecular docking of AMP onto E. Coli Adk.*

The comments are shown in black and the simplified language used to produce the CNS-compatible distance restraints are shown in **red**. These sections are described in detail at http://128.192.148.219/COPClasses/index.php/HADDOCK-Vina

CHAPTER 4

A NMR-DERIVED MODEL OF THE ACETYLSALICYLIC ACID HIGH AFFINITY BINDING SITE ON BOVINE SERUM ALBUMIN USING THE HYBRID MOLECULAR DOCKING WRAPPER HADDOCK-VINA [3]

[3] Wilt, L.A and A. G. Roberts. To be submitted to *PROTEINS: Structure, Function, and Bioinformatics.*

4.1 ABSTRACT

Deciphering the interactions of drugs with their corresponding receptors can greatly accelerate their development. Two powerful structural biology techniques that have been successfully used to investigate these interactions are X-ray crystallography and protein NMR. Using these approaches has been extremely informative. Unfortunately, building drug-bound receptor complexes *de novo* for drug discovery and development is often an arduous and time-consuming process. To overcome this barrier, we developed a wrapper called HADDOCK-Vina in Chapter 3 that can utilize experimental data to drive molecular docking to receptors solved by X-ray crystallography and NMR. The HADDOCK-Vina wrapper is a hybrid molecular docking approach that fuses the rigid molecular docking of AutoDock Vina with the molecular dynamic capabilities of HADDOCK to accurately reproduce small molecule-bound X-ray crystal receptor structures with sparse experimental data. To demonstrate the ability of the HADDOCK-Vina wrapper to be used for building drug-bound receptor complexes, an aspirin (ASA) site on bovine serum albumin (BSA) was determined using NMR relaxation, ligand competition and information from an X-ray crystal structure of a related ligand. The experimentally-derived ASA-BSA model revealed that the ASA binding site overlaps with a related non-steroidal anti-inflammatory drug (NSAID) located within subdomain IB of BSA. This experimentally-derived position is consistent with previous investigations of ASA with BSA.

4.2 INTRODUCTION

There are more than 20,000 ligand-bound X-ray crystal structures currently in the protein database (PDB).[171] In many cases, these structures have provided valuable insight that could be exploited to accelerate drug design and development.[135,185] Unfortunately, successful co-

crystallization of protein receptors with drugs still remains a major experimental hurdle.[203, 204] The quality of the electron density surrounding the ligand is critical for determining an accurate ligand-protein complex;[203, 204] however, the electron density of the ligand can be ambiguous due to the presence of several ligand-bound orientations, partial occupancy in the binding pocket[205, 206] or no electron density despite high ligand concentrations.[173] These limitations are particularly true for ligands that have weak binding affinities. As a result, a major fraction of drug-bound X-ray crystals in the PDB database are with high affinity inhibitors,[171] when many drugs are actually low affinity.

NMR is particularly suited for analyzing the interactions of low affinity ligands to receptors.[207] There are several ligand-based NMR techniques that are amenable to probing the interactions of weakly bound ligands to receptors.[207] The saturation transfer difference (STD) NMR technique can identify functional groups of a weakly bound ligand that are involved in the molecular recognition of the receptor.[14,15] The exchange transfer nuclear Overhauser effect (etNOE) experiments can reveal the bound ligand conformation.[208] The position of a weakly bound ligand can be ascertained using paramagnetic relaxation enhancement (PRE)[209–211] or pseudo contact shifts (PCS) NMR.[18,20] Ideally, this experimental information can be combined with already solved X-ray crystal receptor structures to accurately reproduce the drug-bound X-ray protein crystal structures. For this purpose, the hybrid molecular docking approach HADDOCK-Vina was developed by our laboratory.

The HADDOCK-Vina wrapper is a hybrid molecular docking approach which combines the High Ambiguity Driven biomolecular DOCKing (HADDOCK) protocols[150] with AutoDock Vina[143] and PROPKA[162, 163] to accurately model ligand-protein complexes.[213] The wrapper also provides an easy experimental input to produce experimentally derived distance restraints for the molecular docking simulations. In the previous chapter, the HADDOCK-Vina wrapper accurately

reproduced drug-bound X-ray crystal structures using sparse NMR and non-NMR data.[213] Therefore, this approach was employed to determine the high affinity site of ASA on BSA using NMR distance restraints.

The BSA-ASA complex has not been determined by X-ray crystallography; however, a ligand-bound X-ray crystal structure of BSA with a closely related drug 3,5-diiodosalicylic acid (DIS) is available (PDB ID 4JK4[214]). From the DIS-BSA X-ray crystal structure, DIS occupies four binding sites on BSA in subdomains IB, IIA and IIIA. Assuming ASA overlaps with one of the DIS binding sites on BSA, this DIS-bound BSA X-ray crystal structure will provide a qualitative measure of accuracy. ASA is known to occupy a relatively high affinity and low affinity binding site on BSA.[215] Fluorescence spectroscopy was used to determine the affinities of the "low" and "high" affinity binding sites of BSA. ASA is relatively weakly bound,[216–218] so it is amenable to a number of ligand-based NMR approaches.[219] Since BSA has a single free cysteine (C34) that is readily accessible to labeling,[220] PRE NMR was used to identify the ASA high affinity site. To isolate the "high" affinity binding sites, the NMR experiments were performed at stoichiometric concentrations of BSA and ASA. PRE NMR revealed the relative location ASA with respect to a flexible paramagnetic label. PRE NMR measurements of ASA in the presence of DIS also limited possible locations of the ASA binding sites. To ascertain the effect of specific restraints on the final ASA-docked BSA model, separate HADDOCK-Vina simulations were performed with each set of experimental restraints and with all the restraints. The effects of restraints on the molecular docking and the final ASA-bound BSA model are discussed.

4.3 MATERIALS AND METHODS

4.3.1 Materials.

Dibasic potassium phosphate was purchased from Acros Organic (Geel, Belgium). Zinc chloride, anhydrous was purchased from Alfa Aesar (Haverhill, MA). Manganese sulfate monohydrate was purchased from Amresco (Solon, OH). Deuterium oxide was purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA). Monobasic potassium phosphate and Amicon 30 MWCO spin concentrators were purchased from Fisher Scientific (Waltham, MA). All the remaining chemicals that were used in this study were purchased from Sigma Aldrich (Milwaukee, WI). The high purity BSA lyophilized powder (Sigma, ≥99.9% purity, product no. A-7638) was used for this study because it has been useful in several structural studies of the protein. [e.g. 221–223] Confirming the BSA purity, only a single band was detected by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (data not shown). For NMR and fluorescence experiments, this BSA was suspended in 100 mM potassium phosphate (KPi) pH 7.4.

4.3.2 Labeling BSA

Thiol-reactive EDTA label (N-[S-(2-pyridyldithio)cysteinyl]ethylenediamine-N,N,N',N'-tetraacetic acid monamine) was incubated with either paramagnetic Mn^{2+} or diamagnetic Zn^{2+} metal ion at a 1:1 ratio in unbuffered water for an hour at room temperature.[224] The resulting Mn^{2+} and Zn^{2+} chelated EDTA labels will be referred as Mn^{2+}-EDTA and Zn^{2+}-EDTA, respectively, in the text. To saturate the known Ca^{2+} binding sites on BSA,[214, 225] 2-5 mM of BSA was dissolved in 200mM $CaCl_2$ solution and 100 mM KPi pH 7.4 15 minutes prior to labeling. To ensure labeling of the single free C34 residue on BSA, Mn^{2+}-EDTA and Zn^{2+}-EDTA labels were added at stoichiometric concentrations and incubated at room temperature for two hours in two separate

reactions. Labeled BSA samples were extensively dialyzed against 100 mM KPi pH 7.4 to remove excess label and metals. The degree of labeling was verified by a fluorescent thiol detection assay (Cayman Chemical, Ann Arbor, MI) and found that ≥ 98% of BSA was labeled by this approach (data not shown).

4.3.3 NMR spectroscopy

NMR experiments of ASA and BSA were performed on a Unity INOVA 600 MHz spectrometer at 25°C with a 5 mm z-gradient $^{1}H\{^{13}C/^{15}N\}$ cryoprobe (Agilent Technologies, Santa Clara, CA). All NMR samples contained 100 mM KPi, pD 7.4 with 80% D_2O. ^{1}H NMR peaks of ASA were assigned according to the Spectral Database of Organic Compounds (SDBS). Both longitudinal T_1 and spin-spin T_2 relaxation experiments were performed on the NMR samples containing 500 μM ASA and 500 μM Mn^{2+}-EDTA or Zn^{2+} EDTA labeled BSA. Figure 4.1 shows diagrams of the pulse programs. The T_1 relaxation pulse sequence in Figure 4.1A is comprised of a composite 180^{o} pulse followed by an inversion recovery delay and 3-9-19 WATER suppression by GrAdient Tailored Excitation (WATERGATE) to suppress background water signals.[226–228] A 30 ms $T_1\sigma$ spin lock filter[229] was added to reduce interference from background BSA protein signals. T_2 relaxation pulse program shown in Figure 4.4.1BB used the standard Carr-Purcell-Meiboom-Gill (CPMG) spin echo sequence followed by 3-9-19 WATERGATE for water suppression.[226–228] Details of the pulse sequence are in the Figure legend.

4.3.4 Processing the NMR Spectra

The NMR spectra were processed and analyzed using the NMR processing software, iNMR (http://www.inmr.net), and Igor Pro 6.2 (Wavemetrics, Tigard, OR), as previously described.[209,

[230] The NMR spectra were converted into ASCII text format in iNMR and imported into Igor Pro software package. In Igor Pro, residual background protein signals were removed using Igor Pro and scripts written in the Python programming language.[209] Proton peaks arising from BSA were separated from the ASA proton peaks using the multiple peak-fitting package 2.0 in Igor Pro 6.2.[209] For processing the T_1 relaxation NMR spectra, the fitted peaks were used with singular value decomposition (SVD) analysis to estimate relaxation-induced changes in the ASA NMR spectra.[209] The resulting exponential decay curves were fit to:

$$M_z(t) = M_z(1 - 2 \cdot fe^{-t \cdot R_1} \quad (1)$$

where *Mz* is the *z* component of nuclear spin magnetization, $M_z(t)$is the M_z at thermal equilibrium, R_1 is the longitudinal relaxation rate, *t* is inversion recovery time and *f* is to adjust for small errors in the 180° pulse. The paramagnetic relaxation enhancement (PRE) of the EDTA labels on ASA due to spin-lattice relaxation (R_{P1}) was estimated by taking the difference in the relaxation rates of BSA with the paramagnetic Mn^{2+}-EDTA ($R_{1,\ para}$) and diamagnetic Zn^{2+}-EDTA labels ($R_{1,dia}$), which are calculated from the inverse of the spin-lattice relaxation decay constant:

$$R_{P1} = R_{1,para} - R_{1,dia} = \frac{1}{T_{1,para}} - \frac{1}{T_{1,dia}} \quad (2)$$

where $T_{1,para}$ and $T_{1,dia}$ are the spin lattice decay constants of ASA with BSA labeled with Mn^{2+}-EDTA and Zn^{2+}-EDTA, respectively. The error was estimated from the mean and standard deviation of several R_{P1} measurements.

Because of protein interference on the baseline, the two point approximation method was used to estimate the PRE of the EDTA chelated metal labels on ASA due to spin-spin relaxation (R_{P2}).[231] The intensities of the ^{1}H NMR peaks of ASA in the T_2 relaxation spectra in the presence Mn^{2+}-EDTA (I_{para}) and Zn^{2+}-EDTA labeled BSA (I_{dia}) were taken at two times, t_a and t_b, after the

initial decay of the transverse magnetization. R_{P2} will have the following mathematical relationship:[231]

$$R_{P2} = R_{2,para} - R_{2,dia} = \frac{1}{t_a - t_b} \ln \frac{I_{dia}(t_b) \cdot I_{para}(t_a)}{I_{dia}(t_a) \cdot I_{para}(t_b)} \tag{3}$$

where $R_{2,para}$ and $R_{2,dia}$ are the spin-spin relaxation rates of ASA in the presence of Mn^{2+}-EDTA and Zn^{2+}-EDTA labeled BSA, respectively. As noted earlier,[231, 232] the effects of ^3J coupling during the decay of the transverse magnetization are cancelled out by using identical times, t_a and t_b, and intensity ratios in Equation 3. Errors in R_{P2} or $\sigma(R_{P2})$ can be estimated through error propagation with the following Equation:[231]

$$\sigma\left(R_{P2}\right) = \frac{1}{t_b - t_a} \sqrt{\left\{\frac{\sigma_{dia}}{I_{dia}(t_a)}\right\}^2 + \left\{\frac{\sigma_{dia}}{I_{dia}(t_b)}\right\}^2 + \left\{\frac{\sigma_{para}}{I_{para}(t_a)}\right\}^2 + \left\{\frac{\sigma_{para}}{I_{para}(t_b)}\right\}^2} \tag{4}$$

where σ_{para} and σ_{dia} are the standard deviations in the noise of the ^{1}H NMR spectra of ASA in the presence Mn^{2+}-EDTA and Zn^{2+}-EDTA labeled BSA.

4.3.5 Calculating distances from PRE

PRE through T_1 and T_2 relaxation between a paramagnetic center and a nucleus can be used to estimate their distances to ~35 Å[232] . The paramagnetic relaxation rates, R_{P1} and R_{P2}, the EDTA label and ASA can be described with modified versions of the Solomon Bloembergen Equations: e.g. 209–211, 233

$$R_{P1} = \alpha \frac{2}{5} \left(\frac{\mu_0}{4\pi}\right)^2 \gamma^2 g_e^2 \mu_B^2 S(S+1) J_{SB}(\varpi_N) \tag{5}$$

$$R_{P2} = \alpha \frac{1}{15} \left(\frac{\mu_0}{4\pi}\right)^2 \gamma^2 g_e^2 \mu_B^2 S(S+1)(4J_{SB}(0) + 3J_{SB}(\omega_N)) \tag{6}$$

where μ_0 is the magnetic permeability of free space, γ_N is the nuclear gyromagnetic ratio, g_e is the electronic g-factor, μ_B is the Bohr magneton, and $\varpi_N/2\pi$ is the Larmor frequency of a proton. The fraction of bound ligand (α) can be approximated with ([E]/(K_D+[L])), assuming a single ligand binding model, where K_D is the dissociation constant determined by ASA induced quenching of BSA fluorescence, *[L]* is the ASA concentration, and *[E]* is the BSA concentration. Fractional binding of a ligand onto a receptor can affect the observed relaxation rate and should be considered when calculating the R_{P1} and R_{P2}. Details of implementing the dissociation constant into distance calculations involving mobile ligands and receptors can be found in [209, 211, 234]. The generalized spectral density function $J_{SB}(\varpi)$ can be approximated with:[210]

$$J_{SB}(\omega_N) = r_{app}^{-6} \frac{\tau_C}{1+\omega_N^2 \tau_C^2} \tag{7}$$

where r_{app} is the apparent time-averaged electron nuclear distance and τ_C is the correlation time for the nuclear-electron interaction vector. When the paramagnetic label is rigid, the $\tau_{C,rigid}$ is defined by $(\tau_R^{-1} + \tau_S^{-1})^{-1}$, τ_R is the rotational correlation of the BSA and τ_S is the spin relaxation time of the Mn^{2+}-EDTA label. A τ_S value of 1.3 ns was used in our distance measurements from values calculated previously of Mn^{2+}-EDTA labeled protein.[235] For a paramagnetic label that is mobile, the $\tau_{C,flexible}$ is equal to $(\tau_R^{-1} + \tau_S^{-1} + \tau_i^{-1})^{-1}$, where τ_i is the internal motion of the label on the order of low ns, which can significantly affect the distance calculation.[210] The τ_i has significant effects on the R_{P1}, but not the R_{P2}. Therefore, the R_{P1} Equation was modified with an S^2 order parameter:[210]

$$R_{P1} = S^2 \cdot R_{P1}(\tau_{C,rigid}) + (1-S^2) \cdot R_{P1}(\tau_{C,flexible}) \tag{8}$$

Curie-spin relaxation, which results from dipole-dipole interaction and time averaging of the electronic magnetization, can affect the R_{P2} under certain conditions.[210] This relaxation

mechanism will not likely be significant for the Mn^{2+}-EDTA labeled BSA in this work because Mn^{2+} is an isotropic metal and BSA is a medium sized macromolecule, as previously described.[210]

4.3.6 Parameterizing the Mn^{2+}-EDTA label for Molecular Dynamics

The Mn^{2+}-EDTA was attached to a cysteine residue (Mn^{2+}-EDTA-Cys) in Avogadro 1.1.1 using the Ghemical force field.[174] The Mn^{2+}-EDTA-Cys was parameterized for the GROMOS 53A6 force field, which is optimized for simulations in explicit water,[236] in GROningen MAchine for Chemical Simulations (GROMACS) 4.5.7.[237] The Mn^{2+} metal ion and the surrounding nitrogen and oxygen atoms of EDTA were given partial Mulliken charges of 0.762, -0.068 and -0.282, respectively, for a Mn Schiff base complex in the ground state.[238] The C6 and C12 Lennard-Jones (LJ) parameters for the Mn^{2+} were 3.63119×10^{-5} kJ mol^{-1} nm^6 and 1.21636×10^{-8} kJ mol^{-1} nm^{12}, respectively, and were calculated using a distance between atoms at their lowest potential energy of 0.2635 ± 0.0072 nm and a well depth of $0.02710042 \pm 0.01063420$ kJ mol^{-1}, averaged from several water models.[239] The remaining atoms of Mn^{2+}-EDTA-Cys were given force field parameters derived from cysteine or functional groups such as carboxylic acids found in the GROMOS 53A6 force field.[236] The Mn^{2+}-EDTA-Cys label was integrated into the GROMOS 53A6 force field with the CED three letter amino residue designation.

4.3.7 Estimating the S^2 order parameter and simulating Mn^{2+}-EDTA label dynamics

The S^2 order parameter and internal correlation time (τ_i) for the Mn^{2+}-EDTA label were estimated from a 100 ns molecular dynamics (MD) simulation of BSA labeled with Mn^{2+}-EDTA-Cys.[236, 237] The Mn^{2+}-EDTA-Cys label replaced cysteine 34 in the X-ray crystal structure of bovine serum albumin (BSA) (PDB ID: 4JK4[214]) using a script written in the Python programming

language and the Biopython module.[240] To reduce the computational expense, the 100 ns MD simulations was performed on domain IA of BSA (i.e. first 111 residues) that has the Mn^{2+}-EDTA label. This was simulated in a box having 50 Å sides and containing ~11,000 explicit simple point charge (SPC) model waters and 12 sodium atoms to balance out the charges.[161] Particle mesh Ewald method[241] was used for electrostatics of the protein domain and periodic boundary conditions were provided in all dimensions. The domain was first energy minimized by steepest descent method[242] with a tolerance of 10 kJ mol^{-1} nm^{-1} and maximum step size of 1 nm. Then a short 250ps MD simulation at 300K was performed with the C_α backbone positionally restrained and Berendsen temperature coupling with a time constant of 0.1 ps[243] to allow the amino acid side chains and Mn^{2+}-EDTA-Cys to equilibrate and to minimize van der Waals overlap. This was followed by a 100 ns MD simulation performed at 300K with the C_α backbone positionally restrained in a GROMOS 53A6 force field with Berendsen temperature coupling (time constant=0.1 ps).[243] After the simulation, the internal autocorrelation function ($C_i(t)$) was calculated using a second order Legendre polynomial of the vector between the Mn^{2+} and the amide proton of the cysteine group of the Mn^{2+}-EDTA-Cys label using the g_rotacf program within GROMACS.[236, 237] $C_i(t)$ for the MD simulation is shown as a function of simulation time in Figure 2. The $C_i(t)$ stabilizes after 10 ns reaching a value of approximately 0.9. Assuming that the internal motions are not coupled with the overall tumbling rate, $C_i(t)$ for the internal motion of the Mn^{2+}-amide proton interaction vector can be approximated by:[232]

$$C_i(t) = S^2 + (1 - S^2)e^{-\frac{t}{\tau_i}} \tag{9}$$

where t is the MD simulation time. Fitting the $C_i(t)$ curve shown in Figure 4.2 gives values for the S^2 order parameter and τ_i of 0.91 and 6.64 ns, respectively. This S^2 order parameter is within range

of values calculated for amino acid side chains,[244] although the fact that it is close to 1 indicates that it is fairly rigid.

4.3.8 Parameterizing Mn^{2+}-EDTA label for HADDOCK-Vina

HADDOCK-Vina lacks specific parameterization of a Mn^{2+}-EDTA label.[213] Therefore, the Mn^{2+}-EDTA-Cys was parameterized for the PARALLHDG force field. Force field parameters such as charges and bond energies for the EDTA functional group and Mn^{2+} of the label were extracted from parameters of the deoxythymine labeled with Mn^{2+} chelated by EDTA (Mn^{2+}-EDTA-dT).[232] Parameters for the cysteine functional group of Mn^{2+}-EDTA-Cys used parameters from the cysteine residue of the PARALLHDG force field.[165] Like GROMACS, this Mn^{2+}-EDTA-Cys label was added to the PARALLHDG force field and given the CED three letter amino residue designation. The Mn^{2+}-EDTA-Cys label in 100 ns MD simulation of the BSA domain IA was in many different orientations. To produce a Mn^{2+}-EDTA labeled BSA for HADDOCK-Vina, Python scripts utilizing the Biopython module were used to manipulate the label and the BSA X-ray crystal structure (PDB ID: 4JK4[214]). The Mn^{2+}-EDTA-Cys labels were first extracted from the GROMACS MD snapshots and the amide nitrogen, carbonyl carbon, $C_{\checkmark}$ and C_{β} of the cysteine residue were aligned, which made them resemble a flower bouquet. Then the label ensemble was protonated using Babel 2.31[175] and the atoms were renamed to make them compatible with the modified PARALLHDG force field. Lastly, the C34 of the BSA X-ray crystal structure was replaced with a label with the least amount of overlap from the ensemble. Using this single conformation does not prevent the label from sampling many different orientations during HADDOCK refinement. In fact, the label assumes a relatively similar distribution of label

conformations after 100 HADDOCK-Vina runs as the label assumed during the MD simulation (data not shown).

4.3.9 Fluorescence spectroscopy

Protein fluorescence spectroscopy is a powerful tool to investigate protein conformation and investigate ligand-protein interactions.[106] Monitoring ligand-induced quenching of intrinsic tryptophan fluorescence is a popular method to probe ligand interactions with proteins by fluorescence.[106] The fluorescence technique has already been used in many studies to determine the dissociation constants (K_DS) of ligands with BSA. [e.g. 218, 222, 245–247] The technique was used here to measure the binding affinity of ASA to BSA. These experiments were performed with 1 μM BSA and a range of ASA concentrations from 0 to 1 mM in 100 mM KPi pH 7.4 on an Olis DM 45 spectrophotometer (Olis Corp, Bogart, GA). BSA was excited at 295 nm and the fluorescence was monitored between 300-500 nm. The fluorescence quenching was corrected for inner filter effects and changes in volume using the following Equation:[58, 106]

$$F_{corrected} = (F - B)10^{\frac{(\varepsilon_{ex} b_{ex} + \varepsilon_{em} b_{em})[Q]}{2}} \quad (10)$$

where F is the observed fluorescence, B is the background fluorescence, $[Q]$ is the concentration of the quenching ligand, ε is the extinction coefficient for the excitation (ε_{ex}) and emission (ε_{em}) wavelengths and b is the pathlength. For ASA, the extinction coefficients (ε) at 295 and 330 nm were 2.24 $M^{-1}cm^{-1}$ and 0.18 $M^{-1}cm^{-1}$, respectively.

For fluorescence quenching that exhibits monophasic dependence with respect to ligand concentration, the quenching curve can be fit to:[58, 106]

$$F_{corrected} = \frac{F_{corrected,0}}{1 + K_{SV}[Q]} + F_{unquenched} \quad (11)$$

where K_{SV} is the Stern-Volmer constant, *[Q]* is the concentration of the quenching ligand, $F_{corrected,0}$ is the corrected fluorescence in the absence of quenching ligand, and $F_{unquenched}$ is an offset related to the unquenched fluorescence. In cases where the quenching has biphasic dependence, the quenching curve can be fit to the following Equation:[58, 106]

$$F_{corrected} = \frac{F_{L,0}}{1 + K_{SV,L}[Q]} + \frac{F_{H,0}}{1 + K_{SV,H}[Q]} + F_{unquenched} \quad (12)$$

where $F_{L,0}$ and $F_{H,0}$ are the fluorescence amplitudes at low and high ligand concentrations, respectively. The K_L and K_H are the Stern-Volmer constants at low and high quenching ligand concentrations, respectively. Intrinsic tryptophan fluorescence quenching can occur by static and dynamic or collisional quenching mechanism.[106] Static quenching is correlated to ligand binding with the protein and can be used to calculate the binding affinity.[106] The K_{SV} value, in this case, is equivalent to an association constant (K_A) or $1/K_D$. In contrast, dynamic quenching is associated with random collisions between the ligand and the protein.[106] These quenching mechanisms can be differentiated by examining the temperature dependence of quenching.[48, 106]

4.3.10 Estimated τ_R by dynamic light scattering (DLS)

DLS spectroscopy is a useful technique to gauge the size of proteins.[248, 249] The size of BSA was measured in a Malvern Zetasizer Nano ZS (Malvern Instruments Ltd., UK) in silanized 105.251-QS ultra-micro fluorescence quartz cuvettes (Hellma USA, Plainview, NY). DLS samples contained 1 μM BSA in 100 mM KPi (pH 7.4) at 25°C as described for the NMR experiments. The DLS chromatograms were analyzed with Zetasizer 7.03 software using standard refractive indexes for a protein in aqueous solution. Analysis of the DLS chromatograms of BSA revealed a single peak with a hydrodynamic radius (r_H) of 32 ± 1 Å (data not shown). The r_H was very close

to the average ~35 Å radius that we measured of the X-ray crystal structure of BSA (PDB ID: 4JK4[214]) and that was determined in previous DLS studies of BSA.[250] Using the experimentally determined r_H, a $\boldsymbol{\tau}_R$ was calculated using the Einstein-Stokes Equation:[251]

$$\tau_R = \frac{4\pi\eta r_H^3}{3K_B T} \tag{13}$$

where K_B is the Boltzmann constant, T is the temperature (i.e. 25°C) and η is the viscosity of water (i.e. 8.94 x 10^{-4} Pa sec). The τ_R was determined from Equation 13 was 29.8 ns, which is within values determined previously for BSA. [e.g. 252]

4.3.11 Calculation of Root Mean Square Deviation (RMSD)

The root mean square deviation (RMSD) using the definition of RMSD was calculated between ASA and 3,5-diiodosalicylic acid (DIS) located in subdomain IB (DIS3) from the DIS-bound BSA X-ray crystal structure (PDB ID: 4JK4[214]) and was used to determine clustering of ASA molecules bound to BSA. The RMSD definition used for AutoDock Vina[143] was used for the HADDOCK-Vina simulations.[213] This approach is advantageous because it can calculate the RMSD between molecules that are not identical, as the case with ASA and DIS.[213]

4.3.12 Calculation of Q-Factor

The Q-Factor is analogous to the R-Factor used in X-ray crystallography and is used to calculate the agreement of calculated NOEs or relaxation rates with respect to the experimental values.[232, 253] The Q-Factor used in HADDOCK-Vina will be used in our docking simulations when comparing the docked ASA molecules to the experimental distances used as restraints.[213]

Briefly, if the Q-Factor is equal to 0, the docked conformation of ASA falls within the minimum and maximum experimental distances used as restraints.

4.3.13 Clustering of ASA molecules

The ASA molecules were clustered using the algorithm described in[213]. Briefly, the resulting MD protein structures are aligned. The ASA molecules are clustered using an RMSD-based algorithm[183] that is similar to the one implemented in AutoDock Vina.[143] The RMSD cutoff was set to 2Å for clustering bound ASA molecules which is typical for small molecule clustering. [e.g. 184] The ASA with the lowest H_{score} defines the initial cluster.[213] An RMSD is measured between this ASA and other BSA-bound ASA molecules from the HADDOCK-Vina simulation according to the H_{score}. The first ASA molecule that is outside this RMSD cutoff defines the ASA molecule that is used for the next cluster and so on.

4.4 RESULTS AND DISCUSSION

4.4.1 K_D of ASA to BSA deduced by fluorescence spectroscopy.

The K_D must be experimentally determined in order to estimate the fraction of ligand bound or α in Eqs. 5 and 6. This parameter can be determined using intrinsic tryptophan fluorescence.[106] Figure 4.3 shows the fluorescence spectra and analysis with BSA and ASA. The fluorescence emission spectra of BSA with a range of ASA concentrations after excitation at 295 nm are shown in Figure 4.3A. These spectra were corrected for background fluorescence emanating from ASA by least squares fitting equivalent ASA fluorescence emission spectra in the absence of BSA. [e.g. 254] Increasing concentrations of ASA induces a 30% decrease in intrinsic tryptophan fluorescence of BSA. The large tryptophan quenching effect has been attributed to direct or indirect interaction

of ASA with W134 and/or W213 of BSA, which lie in the IB and IIA subdomains, respectively.[217] The amplitude at 330 nm in Figure 4.3A was plotted as a function of ASA concentration in Figure 4.3B. The fluorescence quenching curve was biphasic, so it was fit to Equation 12. The $K_{SV,L}$ and $K_{SV,H}$ values determined from fitting with Equation 12 were 0.391 ± 0.021 μM^{-1} and 0.00862 ±0.00210 μM^{-1}. These values decreased, when the titration was repeated at higher temperature (data not shown), indicating that ASA induced quenching occurs through a static mechanism and that these K_{SV} values are association constants (K_A's). The K_D's for ASA determined from the K_A's were 2.56 ± 0.014 μM and 116 ± 28.3 μM for each ASA binding site. These results are consistent with the observation that there is a high and low affinity ASA binding sites on BSA.[215, 217] The average affinity is in the range of ASA affinities determined by double reciprocal plot analysis and Scatchard Plot analysis of UV-visible and fluorescence titration curves.[215–217, 255]

4.4.2 PRE NMR experiments with ASA and BSA

PRE NMR of paramagnetically spin-labeled protein and protein with paramagnetic metal centers has been used successfully to create accurate protein and ligand-bound protein structures.[230, 256–258] The distances between paramagnetic centers and nuclei can be measured up to ~ 30 Å.[258] For ligands that are in fast exchange with bulk solvent and protein, distances as short as 10 Å can be estimated from PRE NMR.[230] For example, a distance of 2.6 Å was calculated from PRE between cytochrome P450 heme and bound water versus 2.2 Å for the water bound X-ray crystal structure.[259, 260] In this study, the distances will be calculated using PRE between mobile Mn^{2+}-ETDA label on BSA and an ASA molecule. The effect of 3,5-diiodosalicylic acid (DIS), whose four binding sites are known from an X-ray crystallographic study,[214] on the PRE on ASA was also examined.

One potential complicating factor to our PRE NMR measurements is the fact that ASA has a high and a low affinity binding site on BSA. [e.g. 217] Luckily, PRE by the Mn^{2+}-EDTA label on ASA will only report the binding site that is closest to the label because of its r^6 distance dependence. [e.g. 210] Furthermore, the NMR experiments were performed with high concentration BSA and a stoichiometric amount of ASA so that the high affinity site would be preferentially populated to isolate the PRE effects on ASA occupying the high affinity binding site from the low affinity binding site. Using a Mn^{2+}-EDTA label as a paramagnetic center is advantageous in that distances between the Mn^{2+} and ^1H can be detected at ~35 Å.[232] The ASA binding sites have been identified to lie in subdomain IB and IIA,[217] both within the ~35Å range. Therefore, we are confident that the high affinity is within range of detection with the Mn^{2+}-EDTA label.

Another potential complicating factor to our PRE experiments was that serum albumins are known to have esterase activity that could potentially convert ASA into salicylic acid.[261, 262] A previous NMR study with ASA and BSA already demonstrated that BSA does not hydrolyze ASA under their conditions.[217] Since the NMR experiments in this study were performed at high BSA concentrations, ASA conversion to salicylic acid in the presence of high BSA concentrations (i.e. 500 μM) was monitored by NMR. No significant changes were observed in ASA ^1H peak amplitudes or ^1H chemical shift after 24 hours of incubation at room temperature (data not shown). This confirms that BSA does not hydrolyze ASA.

Figure 4.4 shows the PRE NMR experiments with 500 μM ASA and 500 μM BSA. Although the amplitude of the methyl ^1H NMR peak labeled 5 of ASA was larger than the aromatic peaks, there was too much interference from BSA ^1H NMR signals to reliably measure relaxation (data not shown). Therefore, R_{P1} and R_{P2} were estimated using the aromatic ^1H NMR peaks of ASA labeled 1-4 (Figure 4.4A). Figure 4.4B shows a waterfall plot of aromatic region of the ^1H

ASA T_1 relaxation NMR spectra in the presence of Mn^{2+}-EDTA labeled BSA. The peaks of the ^1H NMR spectra were negative at short inversion recovery times and become positive at longer inversion recovery times. After fitting the ^1H NMR overlapping peaks, the relaxation-induced changes in the ASA ^1H NMR spectra were determined by SVD analysis. Figure 4.4C shows SVD analysis of ^1H ASA NMR peak labeled 1. Fitting the relaxation curve in Figure 4.4C with Equation 1 produces an $R_{1,para}$ of 1.26 ± 0.02 sec^{-1}. Figure 4.4D shows the two point R_{P2} approximation method applied to 500 μM ASA and 500 μM Mn^{2+}-EDTA labeled BSA. In the panel, the ^1H ASA NMR spectra were taken with transverse magnetization decay times of 8 ms and 32 ms. The ^1H NMR peak amplitudes at 32 ms were about half than at 8 ms. From those amplitudes, R_{P2} values were calculated using Equation 3 and are shown in Table 4.1. The effect of competing DIS with ASA on the R_{P1} of ASA protons are shown Figure 4.4e. In the absence DIS, the R_{P1} values ranged from about 0.1 to 0.4 sec^{-1}. To saturate the four DIS binding sites on 500 μM of BSA, 2 mM of DIS was added to the NMR sample. No ^1H DIS NMR signals were observed in the PRE NMR spectra under these conditions, which suggests that DIS is in slow exchange with BSA (data not shown). In the presence of DIS, the R_{P1} was reduced to ~0 sec^{-1}, which indicates that ASA was displaced far away from the label at C34 and it shows that the high affinity ASA binding site overlaps with one of the known DIS binding sites on BSA.

Analysis of the data in Figure 4.4 allowed us to estimate the distances between ASA and the Mn^{2+}-EDTA label using Eqs. 5, 6 and 8 (Table 4.1). The numbers of left most column of the Table are the proton numbers of the molecular structure shown in Figure 4.4A. The T_1 and T_2 relaxation results are shown on the top and bottom of the Table, respectively. Under the T_1 relaxation results section, the R_{P1} values that range from 0.09 to 0.17 sec^{-1}. The $r_{1,min}$, $r_{1,avg}$ and $r_{1,max}$ was calculated with Eqs. 5 and 8 using the minimum, average and maximum R_{P1} values,

respectively. The calculated distances range from about 22 to 31 Å. The T_2 relaxation results section shows the R_{P2} calculated from the amplitudes of the NMR spectra. The R_{P2} values are significantly higher with relaxation values ranging from 12 to 22 sec^{-1}. The distances calculated using Equation 6 ranged from 17 to 21 Å. The distances calculated from the T_1 and T_2 relaxation are in the same general range. The differences may be the result of internal motion effects on the R_{P1} and/or limitations in the two point approximation method for estimating R_{P2}.[210]

4.4.3 Molecular docking of ASA to Mn^{2+}-EDTA labeled-BSA using HADDOCK-Vina

Without molecular docking with HADDOCK-Vina, our experimental data already reveals the approximate location of the high affinity ASA binding site. The quenching of tryptophan fluorescence (Figure 4.3) implies that ASA binds near to W134 of subdomain IB or to W213 of subdomain IIA.[214] The T_1 and T_2 relaxation experiments suggest that ASA binds within ~20-30 Å of the Mn^{2+}-EDTA label (Table 4.1). Displacement of ASA by DIS, revealed by decreases in R_{P1} (Figure 4.4D), suggest that binding occurs at one of the DIS binding sites. The DIS sites (DIS1 and DIS4) located near W213 in subdomain IIA exceed 30 Å and would not detect ASA ^1H relaxation from the paramagnetic label at C34. The only DIS binding site that lies within the experimentally-derived T_1 and T_2 distances (Table 4.1) is in subdomain IB (i.e. residues 112-205 of BSA) or DIS3.[214] Therefore, the DIS3 binding site was used as a restraint in HADDOCK-Vina. The DIS3 binding site is defined by eighteen residues, that lie within 4 Å of the DIS molecule:[214] 114, 115, 116, 140, 141, 142, 159, 160, 161, 180, 181, 182, 184, 185, 186, 187, 188 and 189. To restrain the ASA molecule, the DIS3 residue numbers were inputted into the experimental input of HADDOCK-Vina in [MUT] section as 18 separate restraints with a weight of 1. The

experimental input used for the docking of ASA onto BSA is shown in *Supplementary Information* (Figure 4.S1).

The T_1 and T_2 relaxation-derived ASA to Mn^{2+}-EDTA labeled BSA distances reflect a distance weighted toward the shortest ASA to Mn^{2+} distance because of the r^6 distance dependence.[210] In addition, a free ASA molecule that is in fast exchange between the bulk solvent and BSA will assume many bound orientations with respect to BSA, so the r_{app} in *Equation 7* reflects the weighted distance average of a ligand bound in many different orientations. Distances are inputted into the [PRENMR] section of the experimental input of HADDOCK-Vina as distance (r), lower bound distance (r_{lower}) and upper bound distance (r_{upper}), which are all used to define the distance restraint within the HADDOCK protocols and CNS. The r_{lower} and the r of the distance restraint was set to r_{min} and r_{max}, respectively, while the r_{upper} was set to r_{max} plus the long axis of ASA (i.e. 8.6 Å) to account for free rotation of ASA (Figure 4.S1).

The receptor used for the HADDOCK-Vina simulations is the DIS-bound BSA crystal structure (PDB ID: 4JK4).[214] Since the sidechains are oriented for DIS binding, the sidechains are essentially randomized for ASA binding and side chain movement occurs during HADDOCK refinement. To dock ASA onto BSA, extraneous metals and ligands were removed from the receptor, as described previously.[213] The ASA molecule was extracted from PubChem and processed on the PRODRG server.[168, 213] To effectively coat BSA surface, 2,000 ASA molecules were docked with a weight repulsion of 0.3. From these ligands, ASA molecules that had a H_{score} of 0 and a favorable E_{total} were selected for HADDOCK refinement, as previously described.[213] Each selected AutoDock Vina conformation was subjected to 100 simulations of energy minimization, simulated annealing and molecular dynamics with explicit solvent in HADDOCK refinement.[213] The bound ASA molecules were clustered according to their H_{score} and a 2 Å RMSD

cutoff. The number of clusters with significant populations of molecules will decrease as the RMSD between the ligand-bound model and the ligand-bound X-ray crystal structure decreases, as observed previously.[213] Thus, we anticipate that the number of populated clusters will decrease as we approach an accurate solution. In this study, there is no ASA-bound BSA X-ray crystal structure; therefore, the ASA bound conformations will be compared with the structurally-related DIS molecule bound at the DIS3 binding site determined from the X-ray crystal structure (PDB ID 4JK4[214]). Of course, getting a low RMSD between an ASA bound on BSA and the DIS bound on the X-ray crystal structure of BSA (PDB ID 4JK4[214]) does not necessarily indicate a good ASA-bound BSA model. Most importantly, a good ASA-bound model should have an E_{dist} and Q-Factor that approaches 0, suggesting the experimental data fit well within the model. In addition to a reduced number of populated molecular clusters, the cluster formed from the bound ASA molecule with the lowest H_{score} should have a significantly higher population than the remaining clusters.[213]

ASA was docked to Mn^{2+}-EDTA labeled BSA using restraints derived from the NMR relaxation experiments (Figure 4) and HADDOCK-Vina. To determine the effect distance restraints on cluster distribution, HADDOCK-Vina molecular docking simulations were performed with the distance restraints of each NMR experiment separately and combined in Figure 4.5. Table 4.2 shows the average H_{scores}, interaction energies and RMSDs for the lowest H_{score} cluster in the ASA molecular docking simulation with HADDOCK-Vina. A complete set of statistics is provided in Table 4.S1 of the Supplementary Information. In Table 4.2, the left column is the restraint type. For T_1 and T_2 relaxation, the distance restraints are found in Table 4.1. The DIS3 competition restraint type are distance restraints derived from residues within the DIS3 binding site, and all is a combination of all the distance restraints. The lowest H_{score} value that

defines the first cluster is in the next column. The next columns have the average H_{scores} and interaction energies for that cluster. To correlate the ASA modeled site to the DIS3 site of X-ray crystal structure of BSA,[214] the average RMSD between ASA in the cluster and DIS was calculated in the second to last column using Eqs. 14 and 15. In the last column, a Q-factor was calculated for the ligand in the cluster with the lowest H_{score} from the simulations using the T_1 and T_2 relaxation derived distance restraints using Equation 15.

Molecular docking of ASA with BSA was first examined using only distance restraints derived from T_1 relaxation data. The resulting distribution of the 58 clusters from molecular docking with HADDOCK-Vina is shown in Figure 4.5B and the BSA X-ray crystal structure with the location of the ASA cluster with the lowest H_{score} (red) and ASA clusters with E_{dist} = 0, indicating good correlation with experimental data (blue) is shown in Figure 4.5C. The lowest H_{score} cluster was formed from 41 ASA molecules and was ~28 Å from the Mn^{2+}-EDTA label, which fits within the T_1 distances (Table 4.1). The lowest H_{score} of the ASA molecule occupying the cluster was -169 A.U. (Table 4.2). This molecule also had Q-factor of 0.00 and was positioned about ~11 Å from the DIS molecule occupying the DIS3 binding site of BSA (Table 4.2). The cluster associated with this molecule had an average E_{total} of -95.7 ± 36.1 kcal mol^{-1} and an average E_{dist} of 0.00 ± 0.04, indicating that the location of the docked ASA molecules corresponded well within the experimental data (Table 4.2). Unfortunately, 48 other clusters also have average E_{dist} of 0 including clusters (Table 4.S2). While the lowest H_{score} structure is the most populated, cluster #27 had 38 ASA molecules, as well as #21 and #34 clusters are also highly populated. Therefore, additional experimentally derived distance restraints will be needed to reduce the ambiguity of ASA molecules on BSA.

Figs 5d and 5e show the distribution of 62 clusters and the position of ASA clusters on the BSA X-ray crystal structure, respectively, using distance restraints derived only from T_2 relaxation measurements. The lowest H_{score} from the initial cluster was -194, which was comprised of 22 ASA molecules. This cluster was ~21 Å from the Mn^{2+}-EDTA label, within range of the experimental data (Table 4.1). The lowest H_{score} structure within this cluster had both a E_{dist} and the Q-factor were 0 showing that its position fits well with respect to the experimental data. The molecule lies about ~17 Å away from the bound DIS molecule in the DIS3 binding site (Figure 4.5C). One of the issues with the molecular docking using distances derived from a single experimental source is that the remaining 55 ASA clusters have an E_{dist} of 0. This implies that the other bound ASA molecules are equally likely to represent the ASA-bound complex, since they fit well within the experimental data. In addition, one of the clusters has a much higher populations of ASA molecules, cluster #16, which has 60 ASA molecules. Molecular docking driven by distances restraints derived from either the T_1 and T_2 relaxation experiments from a single label is clearly not sufficient to confidently identify a single ASA binding site on BSA.

Molecular docking of ASA with BSA using only DIS3 binding site residues as restraints are shown in Figures 5F and 5G. The first lowest H_{score} cluster (Figure 4.5F) was found within the IB subdomain of BSA (cf. Figure 4.5G and 5A), where we anticipate the ASA binding site lies. The lowest H_{score}, average H_{scores} and average E_{totals} of the ASA molecules with the cluster was -131 A.U., -80.1 ± 28.0 and -80.2 ± 27.8 kcal mol^{-1}, respectively (Table 4.2). This ASA cluster essentially overlapped with the DIS3 binding site, with an average RMSD of 2.42 ± 0.54 Å (Table 4.1). The population of the second cluster (Figure 4.5F) was about half of this cluster. The lowest H_{score} of cluster #2 had a very similar H_{score} to the lowest H_{score} cluster, with a value of -130. This cluster had an E_{dist} of 0 kcal mol^{-1} versus the E_{dist} of the lowest H_{score} cluster of 0.01 ± 0.10 kcal

mol^{-1} (Table 4.S1), which shows that cluster #2 correlates better to the experimentally derived distance restraints. In addition, the lowest H_{score} ligand in cluster #2 is considerably closer to the DIS3 binding site with an RMSD around 1.79 Å (Table 4.S2). Additionally, clusters #4 and #6 also had E_{dist} of 0 (Figure 4.5F). The DIS competition data by itself is clearly not sufficient to resolve the ASA binding site on BSA.

Figs. 5h and 5i show the molecular docking with HADDOCK-Vina with all the experimentally-derived distances. The molecular docking of ASA with these distance restraints resulted in 6 clusters (Figure 4.5H). All the clusters are located within domain I of BSA. The initial cluster formed from the bound ASA molecule with the lowest H_{score} of -167 A.U. was populated with ~89% of the ASA molecules from the HADDOCK-Vina runs (Figure 4.5H). This cluster had an average E_{total} and E_{dist} of -49.3 ± 28.7 and 1.70 ± 1.55 kcal mol^{-1}, respectively (Table 4.S1). The average E_{dist} is the highest value compared to the lowest H_{score} clusters derived distance restraints from a single experimental source because the combined distance restraints will oppose each other, as seen previously.[213] However, the ASA molecule with the lowest H_{score} that defines the cluster had a Q-factor and E_{dist} of 0 and 0.14 kcal mol^{-1}, respectively, showing that it fits well with all the experimental data. The molecule also had the E_{total} of -168 kcal mol^{-1}. The RMSD of this bound ASA molecule to the DIS molecule occupying the DIS3 binding site of the BSA X-ray crystal structure was 1.57 Å, showing that they have overlapping binding sites.

4.4.4 ASA-bound BSA model versus DIS-bound BSA X-ray crystal structure

The molecular docking simulations of ASA to BSA using all the experimental restraints were analyzed further in Figure 4.6. Figure 4.6A shows the most populated cluster from Figure 4.5H (red) with respect to the Mn^{2+}-EDTA label and DIS (black). The cluster is located in the 1B

subdomain and completely envelopes the DIS molecule occupying the DIS3 binding site of BSA. The RMSD of the lowest H_{score} ASA occupying the cluster and DIS was 1.57 Å. The cluster associated with this ASA molecule completely envelopes the DIS molecule from the X-ray crystal structure with an average RMSD of 2.68 ± 0.41 Å (Table 4.S1). The cluster position is consistent with the competition observed between DIS and ASA with BSA in Figure 4e. The next panel (Figure 4.6B) shows the DIS molecule within the DIS3 binding site and interacting residues of BSA. The carboxylic acid functional group of DIS forms hydrogen bonds with R185, while the iodine functional groups interact with the backbone of R185 in addition to Y137, Y160 and M184 as noted in the original publication.[214] The ASA molecule with the lowest H_{score} from the cluster in panel a is shown in Figure 4.6C with surrounding residues. The ASA molecule was 26 Å from the Mn^{2+}-EDTA label, corroborating the distances determined by PRE NMR (Table 4.1). The R185 was displaced ~3 Å from its original position in the X-ray crystal structure. The residues R185 and K114 both form hydrogen bonds with the carboxylic acid functional group of ASA. There are also hydrophobic interactions between the phenyl functional group of ASA and L115. The two tyrosine residues that interacted with DIS are shown in panel C for comparison. While both Y137 and Y160 are too far away to interact with ASA, in the absence of DIS3, the two residues overlap to satisfy pi-pi stacking interactions between them.

4.5 CONCLUSION

In the previous publication, the HADDOCK-Vina protocol integrated several software programs to experimentally drive docking. With the software, we essentially reproduced the ligand positions from the X-ray crystal protein structures. Using the molecular docking strategy developed in the previous manuscript, HADDOCK-Vina was used to dock a weakly interacting

ligand in fast exchange to a macromolecule under conditions that would be difficult to investigate by X-ray crystallography. In HADDOCK-Vina, information from PRE NMR relaxation experiments with paramagnetically labeled BSA and competition with DIS were used to drive docking of ASA to BSA. Using the PRE NMR relaxation derived distance restraints alone was not sufficient to unambiguously identify the high affinity ASA binding site. In these cases, the cluster number and distribution showed a lot of variation between molecular docking simulations. When additional restraints were added, the ASA molecules coalesced toward the cluster derived from the lowest H_{score} ASA molecule. Our experimentally-driven docking revealed that the ASA binding site was posited on the subdomain IB of BSA. The approach applied to the BSA and ASA model system can be translated for docking other weakly interacting ligands such as substrates and drugs. Therefore, HADDOCK-Vina protocol can serve as a powerful tool for receptor-small molecule docking with experimental data.

4.6 TABLES

Table 4.1 Paramagnetic relaxation rates and the calculated distances

T_1 relaxation results

proton	$R_{1,para}$ (sec^{-1})[a]	$R_{1,dia}$ (sec^{-1})[b]	R_{P1} (sec^{-1})[c]	$r_{1,min}$ (Å) *	$r_{1,avg}$ (Å) *	$r_{1,max}$ (Å) *
H-4	1.19 ± 0.01	1.34 ± 0.06	0.09 ± 0.03	26.92	28.38	30.68
H-2	0.967 ± 0.06	1.43 ± 0.05	0.34 ± 0.07	22.09	22.79	23.68
H-3	1.04 ± 0.04	1.45 ± 0.02	0.27 ± 0.04	23.04	23.61	24.29
H-1	1.26 ± 0.02	1.62 ± 0.02	0.17 ± 0.01	25.10	25.44	25.80

T_2 relaxation results

proton	$I_{para}(t_a)$,$I_{para}(t_b)$[d]	$I_{dia}(t_a)$,$I_{dia}(t_b)$[e]	$R_{2,P}$ (sec^{-1})[f]	$r_{2,min}$ (Å) *	$r_{2,avg}$ (Å) *	$r_{2,max}$ (Å) *
H-4	0.92, 3.55	1.25, 3.61	11.8 ± 4.31	18.73	19.73	21.27
H-2	1.33, 3.39	3.63, 5.42	22.3 ± 1.31	17.58	17.75	17.93
H-3	0.74, 3.03	2.65, 6.46	21.7 ± 3.96	17.34	17.83	18.44
H-1	2.40, 6.25	5.68, 9.46	18.6 ± 1.32	18.09	18.30	18.53

[a,b] The T_1 relaxation rates for the paramagnetic ($R_{1,para}$) and diamagnetic ($R_{1,dia}$) were calculated by exponential fitting for protons 1-4.
[c] The relaxation rates (R_{P1}) were calculated by taking the difference of the $R_{1,Para}$ and $R_{1,Dia}$ rates for each proton.
[d] The intensities of the ^{1}H NMR peaks of ASA in the paramagnetic samples at two time points.
[e] The intensities of the ^{1}H NMR peaks of ASA in the diamagnetic samples at two time points.
[f] The T_2 relaxation rates (R_{P2}) were calculated by the two time-point measurement method for each proton.
* For both T_1 and T_2, the average distance (r_{avg}) was calculated using the modified Solomon-Bloembergen equation. The minimum distance (r_{min}) and the maximum distance (r_{max}) were calculated from the maximum and minimum R_p values.

Table 4.2 Cluster analysis of ASA-BSA complex with HADDOCK-Vina.

Restraint Type	H_{score} (A.U.)[a]	Average H_{score} ± S.D. (A.U.)[b,*]	Average E_{vdw} ± S.D. (kcal/mol)[*]	Average E_{elect} ± S.D. (kcal/mol)[*]	Average E_{desolv} ± S.D. (kcal/mol)[*]	Average E_{dist} ± S.D. (kcal/mol)[*]	Average E_{total} ± S.D. (kcal/mol)[c,*]	Average RMSD ± S.D. (Å)[d,*]	Q-Factor[e]
T_1 PRE NMR	-169	-95.6 ± 36.3	-13.6 ± 2.56	-92.1 ± 37.7	10.0 ± 9.64	0.00 ± 0.04	-95.7 ± 36.1	10.2 ± 0.55	0.00
T_2 PRE NMR	-194	-119 ± 34.5	-11.3 ± 4.31	-119 ± 30.2	12.0 ± 9.30	0.00 ± 0.00	-119 ± 34.5	20.5 ± 1.05	0.00
DIS Binding Site	-131	-80.1 ± 28.0	-20.6 ± 2.80	-67.5 ± 24.4	7.91 ± 11.0	0.01 ± 0.10	-80.2 ± 27.8	2.42 ± 0.54	--
All	-167	-32.3 ± 34.0	-15.9 ± 2.24	-40.8 ± 26.3	7.41 ± 9.52	1.70 ± 1.55	-49.3 ± 28.7	2.68 ± 0.41	0.00

[*]Average ± standard deviation (S.D.) for each docked structure within the cluster. Clusters with a S.D. of 0 had only one docked structure populating the cluster.
[a]The lowest HADDOCK score (H_{score}) structure of the lowest H_{scored} cluster
[b]The average H_{scored} structures within the lowest H_{scored} cluster
[c]The total intermolecular energy ($E_{total} = E_{vdw}+E_{elect}+E_{desolv}+E_{dist}$) was calculated for each structure and averaged for the lowest H_{scored} cluster
[d]The RMSD was calculated for each structure and averaged for the lowest H_{scored} cluster
[e]The Q-Factor of the lowest H_{score} ligand in the lowest H_{scored} cluster

4.7 FIGURES

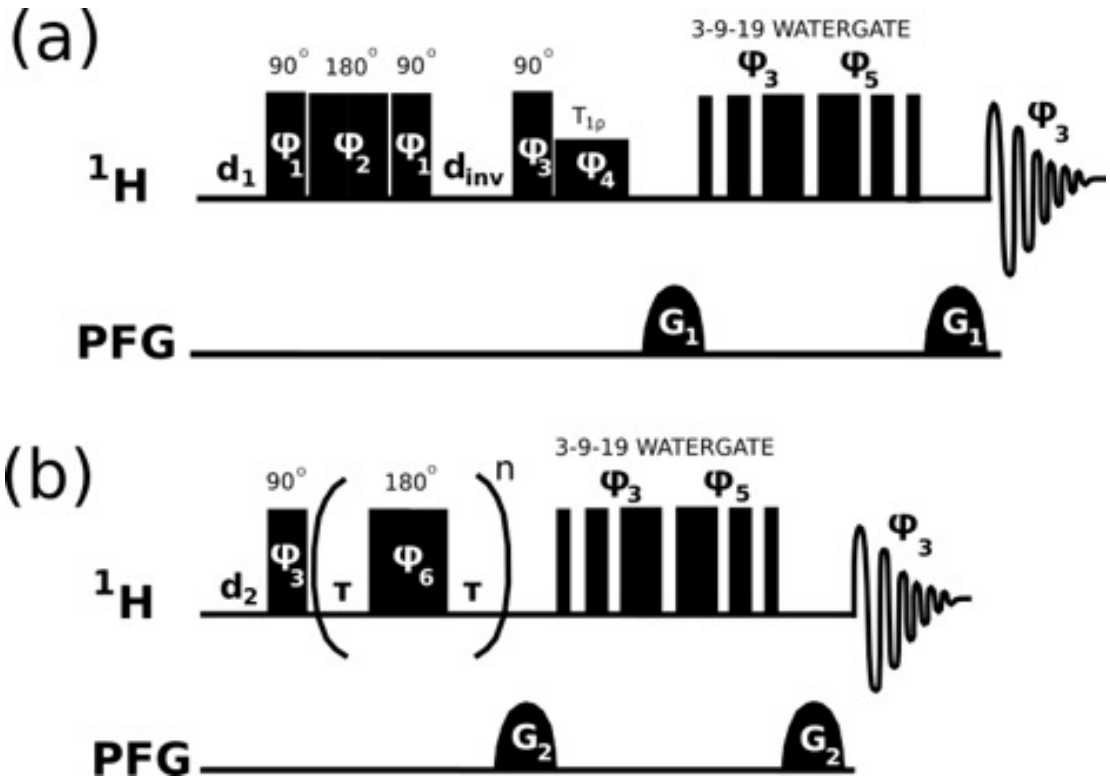

Figure 4.1. Pulse sequences for measuring PRE. **(A)** R_{P1} and **(B)** R_{P2} showing the ^1H and pulse field gradient (PFG) channels. The bars reflect square pulses, while the parabolas are gradient pulses. The 90°, 180° and $T_{1\rho}$ spin lock pulses are labeled. The 6 pulses titled with 3-9-19 WATERGATE were used for water suppression. The relaxation delays, d_1 and d_2, were 15 and 4 s, respectively. The d_{inv} is the inversion recovery time. The τ in **(B)** was set to 1 ms. Phase cycling: ϕ_1=x; ϕ_2=y; ϕ_3=(x,x,y,y,-x,-x,-y,-y,y,y,-x,-x,-y,-y,x,x,-x,-x,-y,-y,x,x,y,y,-y,-y,x,x,y,y,-x,-x); ϕ_4=(-y,y); ϕ_5=(-x,-x,-y,-y,x,x,y,y,-y,-y,x,x,y,y,-x,-x,x,x,y,y,-x,-x,-y,-y,y,y,-x,-x,-y,-y,x,x); ϕ_6=(y,y,-x,-x,y,y,-x,-x,-x,-x,y,y,-x,-x,y,y). The field gradients, G_1 and G_2, were optimized to suppress water. The variable n is the number of spin echo repetitions, which was calculated by dividing the transverse magnetization decay time by 2* τ.

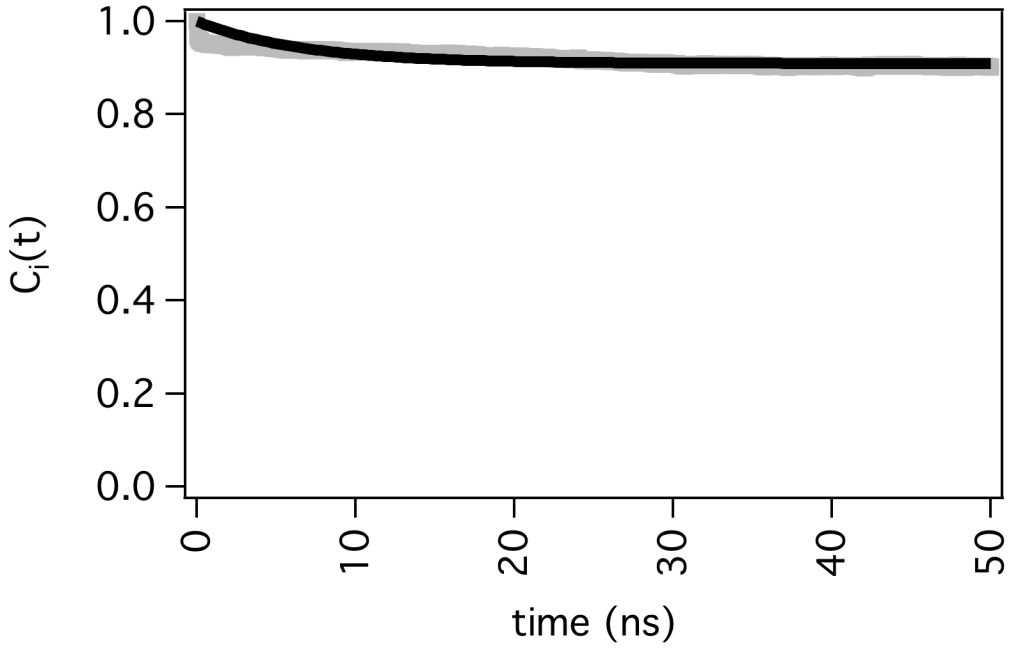

Figure 4.2. Internal autocorrelation function ($C_i(t)$) of the vector between the amide proton and the Mn^{2+} nucleus for the Mn^{2+}-EDTA labeled C34 on domain IA of BSA as a function of MD simulation time. The $C_i(t)$ values are shown as solid gray squares and the fit to Equation 9 is shown as a black line.

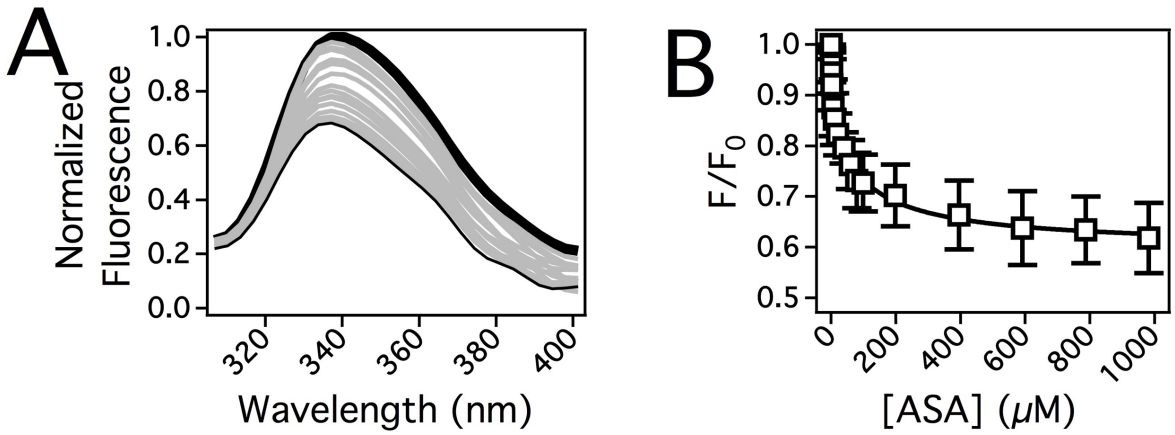

Figure 4.3. Intrinsic tryptophan fluorescence quenching of BSA in the presence of ASA. **(A)** The normalized BSA tryptophan fluorescence spectra (310-400 nm) with increasing concentrations of ASA. The BSA tryptophan fluorescence spectrum at 0 and 3.0 mM ASA are represented as thick and thin black lines, respectively. The intermediate concentrations are shown as gray lines. **(B)** The amplitude at 333 nm corrected for inner filter effects, background and volume is plotted as a function of ASA concentrations and normalized. The fit to Equation 12 is shown as a black line. The data points represent an average and the bars represent the standard deviation of at least three experiments.

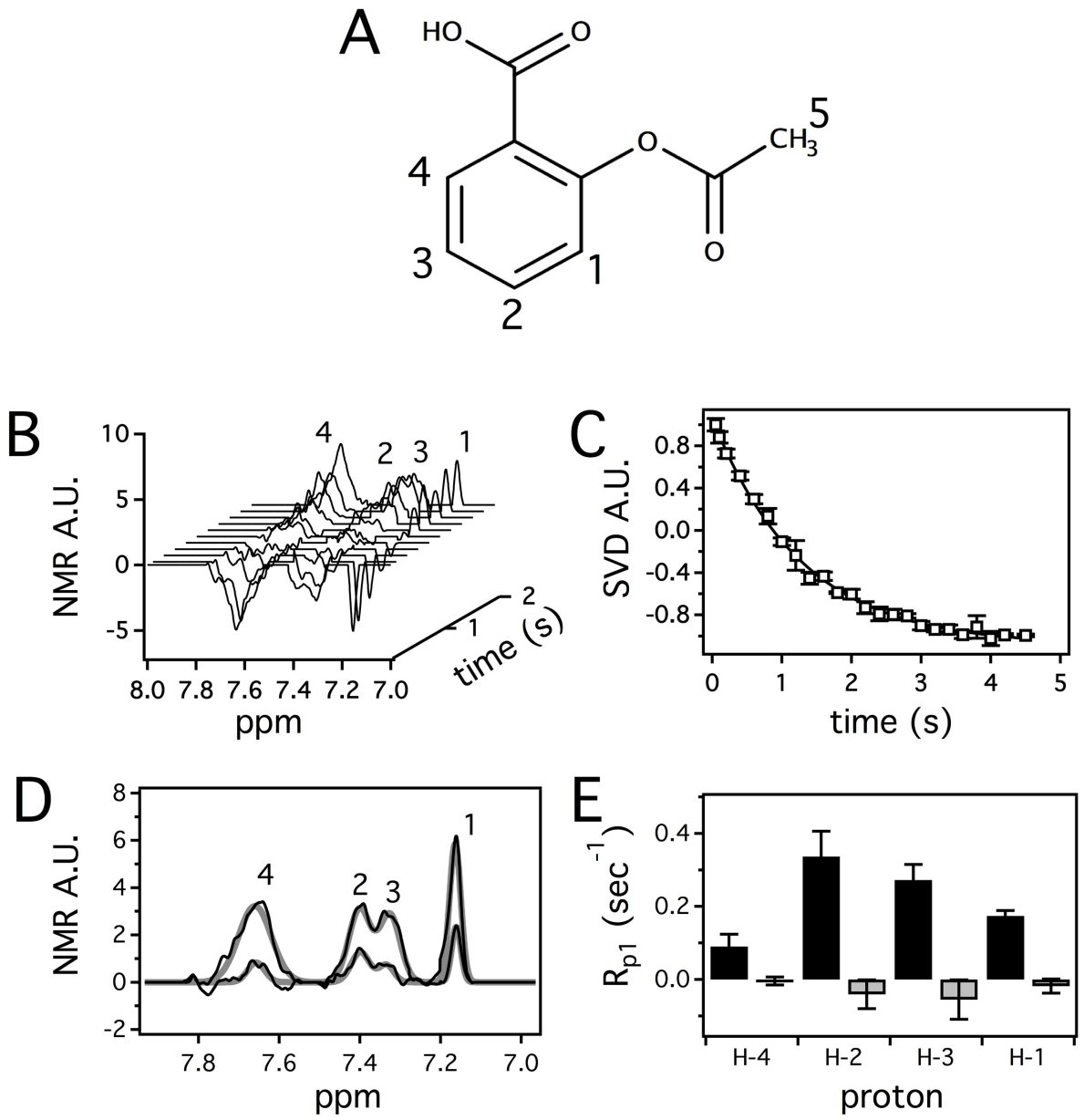

Figure 4.4. Relaxation of ^{1}H NMR peaks from 500 μM ASA in the presence of 500 μM Mn^{2+}-EDTA labeled BSA. **(A)** A waterfall plot of ^{1}H ASA NMR spectra at various inversion recovery times. **(B)** SVD analysis of proton labeled 1 as a function of the inversion recovery time. **(C)** The ^{1}H ASA NMR spectra at 8 ms and 32 ms T_2 relaxation times (black) and the least squares fits (gray). **(D)** The R_{P1} of aromatic protons of ASA in the absence (black bars) and the presence of DIS (gray bars).

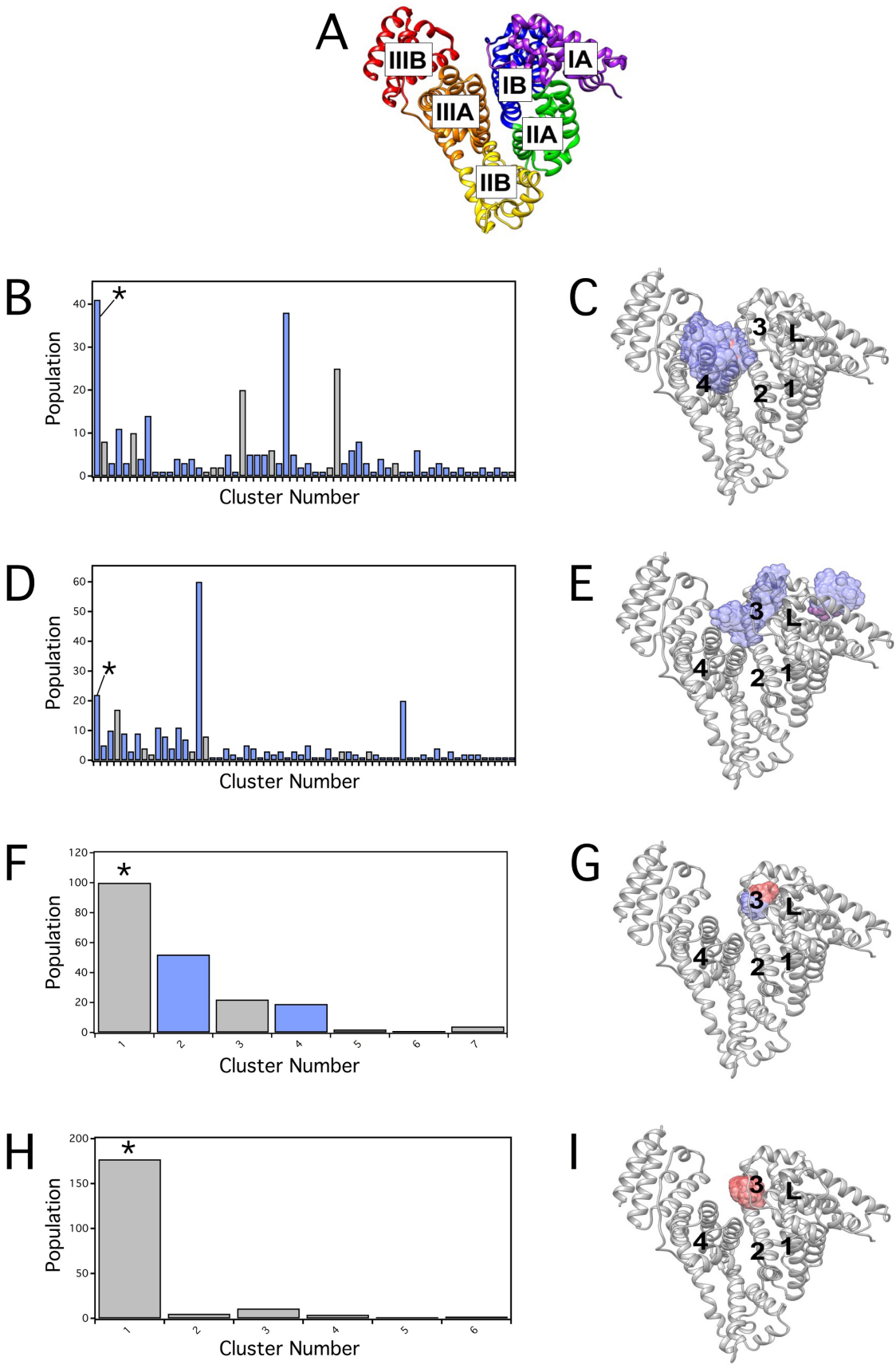
A
IIIB
IA
IB
IIIA
IIA
IIB
B
Population
Cluster Number
C
3
L
4
2
1
D
Population
Cluster Number
E
3
L
4
2
1
F
Population
Cluster Number
G
3
L
4
2
1
H
Population
Cluster Number
I
3
L
4
2
1

Figure 4.5. Experimentally-driven molecular docking of ASA to BSA with the HADDOCK-Vina protocol. **(A)** The structure of BSA with the subdomains labeled. Molecular docking of ASA using distance restraints derived from **(B,C)** the T_1 relaxation (Table 4.1), **(D,E)** the T_2 relaxation (Table 4.1), **(F,G)** the DIS competition (Figure 4D), and **(H,I)** all the NMR experiments. The clustering distribution from the molecular docking simulations is shown **B,D,F,H** with the highest populated cluster denoted with an asterisk. Clusters with a $E_{dist} = 0$ are designated with a pound sign. The clusters are organized, so that ASA molecules with the lowest H_{scores} in each cluster are ordered first. Clusters with the lowest H_{score} (red) and $E_{dist} = 0$ (blue) are rendered as van der Waals volume in **C,E,G,I**. The overlaid numbers, 1-4, correspond to the DIS binding sites, DIS1-DIS4, that were noted in the DIS-bound BSA X-ray crystal structure[214] . The "L" represents the location of the Mn^{2+}-EDTA label.

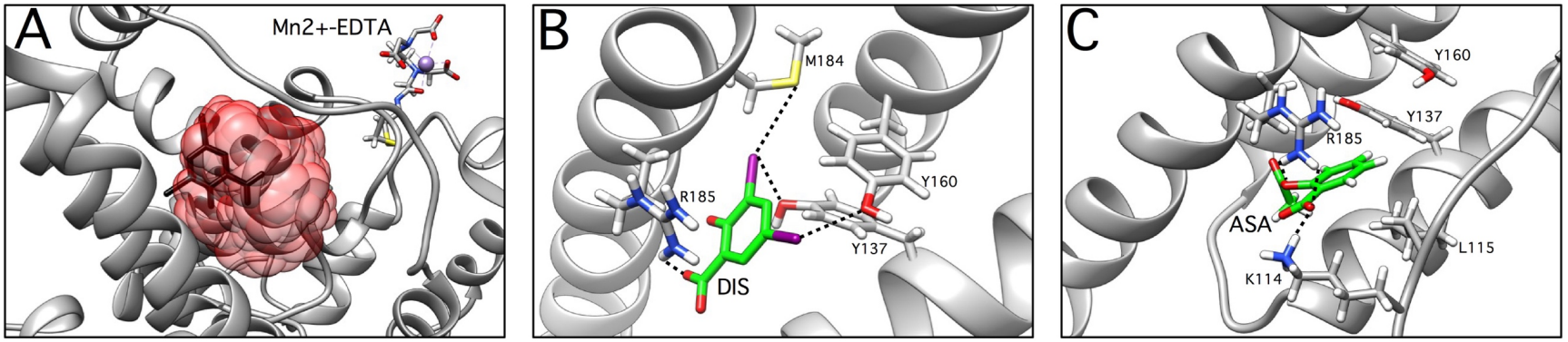

Figure 4.6. The high affinity ASA binding site of BSA deduced by HADDOCK-Vina. **(A)** The most populated cluster of ASA using all experimental data derived distance restraints (transparent red) with respect to the Mn^{2+}-EDTA label. The DIS molecule from the X-ray crystal structure is shown within the ASA cluster (black). **(B)** The bound DIS molecule in the DIS3 binding site within subdomain IB of the BSA X-ray crystal structure[214] . **(C)** The ASA molecule with the lowest H_{score} and E_{total} bound near to the DIS3 binding site of BSA. The residues that interact with DIS and ASA in **B,C** are shown as sticks and labeled. In **B,C**, the carbons of DIS and ASA are colored green, while the carbons of interacting residues are colored gray. The dashed lines in **B,C** denote H-bonds and interactions between the ligands and BSA.

4.8 SUPPLEMENTARY INFORMATION

Table 4.S1 *Cluster analysis of all clusters produced from HADDOCK-Vina*

Restraint Type	Cluster Number	Number in Cluster	$f_{overlap}$[a]	Lowest H_{score}[b]	Average H_{score} ± S.D.[c,*]	Average E_{vdw} ± S.D.* (kcal/mol)	Average E_{elect} ± S.D.* (kcal/mol)	Average E_{desolv} ±S.D.* (kcal/mol)	Average E_{dist} ± S.D.* (kcal/mol)	Average E_{total} ± S.D.[d,*] (kcal/mol)	Average RMSD ± S.D.[c,*] (Å)
T1	1	41	0	-169	-95.6 ± 36.3	-13.6 ± 2.56	-92.1 ± 37.7	10.0 ± 9.64	0.00 ± 0.04	-95.7 ± 36.1	10.2 ± 0.55
	2	8	0	-158	-83.4 ± 45.2	-9.05 ± 2.75	-90.0 ± 46.0	14.7 ± 14.3	0.09 ± 0.24	-84.3 ± 46.3	9.19 ± 0.73
	3	3	0	-151	-124 ± 28.5	-12.4 ± 3.09	-116 ± 27.2	4.07 ± 12.2	0.00 ± 0.00	-124 ± 28.5	10.3 ± 0.35
	4	11	0	-136	-76.6 ± 36.6	-4.61 ± 3.00	-84.3 ± 36.9	12.3 ± 8.67	0.00 ± 0.00	-76.6 ± 36.6	27.2 ± 0.66
	5	3	0	-132	-105 ± 26.9	-18.5 ± 2.64	-94.4 ± 18.5	7.65 ± 11.6	0.00 ± 0.00	-105 ± 26.9	10.9 ± 0.18
	6	10	0	-131	-52.4 ± 44.6	-16.9 ± 1.53	-40.7 ± 44.0	4.74 ± 5.39	0.05 ± 0.13	-52.9 ± 44.3	9.97 ± 0.39
	7	4	0	-129	-82.3 ± 44.5	-10.3 ± 3.10	-76.9 ± 46.0	4.93 ± 3.95	0.00 ± 0.00	-82.3 ± 44.5	8.59 ± 0.84
	8	14	0	-127	-67.5 ± 25.3	-18.0 ± 2.01	-60.3 ± 24.2	10.8 ± 16.3	0.00 ± 0.00	-67.5 ± 25.3	11.2 ± 0.45
	9	1	0	-125	-125 ± 0.00	-10.7 ± 0.00	-130 ± 0.00	16.2 ± 0.00	0.00 ± 0.00	-125 ± 0.00	8.72 ± 0.00
	10	1	0	-112	-112 ± 0.00	-12.5 ± 0.00	-118 ± 0.00	19.0 ± 0.00	0.00 ± 0.00	-112 ± 0.00	10.1 ± 0.00
	11	1	0	-105	-105 ± 0.00	-5.12 ± 0.00	-106 ± 0.00	5.81 ± 0.00	0.00 ± 0.00	-105 ± 0.00	9.34 ± 0.00
	12	4	0	-105	-69.1 ± 32.4	-19.7 ± 1.22	-49.8 ± 34.9	0.41 ± 7.24	0.00 ± 0.00	-69.1 ± 32.4	11.0 ± 0.44
	13	3	0	-104	-61.5 ± 36.9	-8.40 ± 3.91	-59.0 ± 21.6	5.91 ± 12.3	0.00 ± 0.00	-61.5 ± 36.9	27.6 ± 0.74
	14	4	0	-100	-83.2 ± 14.0	-20.5 ± 2.58	-76.4 ± 20.2	13.7 ± 7.19	0.01 ± 0.02	-83.2 ± 14.0	11.5 ± 0.52
	15	2	0	-88.5	-34.6 ± 76.3	-3.82 ± 0.66	-45.8 ± 71.3	15.1 ± 4.29	0.00 ± 0.00	-34.6 ± 76.3	25.3 ± 0.63
	16	1	0	-76.3	-76.3 ± 0.00	-10.4 ± 0.00	-84.7 ± 0.00	18.4 ± 0.00	0.04 ± 0.00	-76.7 ± 0.00	9.55 ± 0.00
	17	2	0	-69.8	-46.4 ± 33.1	-14.9 ± 4.67	-46.5 ± 33.9	13.4 ± 6.05	0.16 ± 0.22	-48.0 ± 35.2	9.86 ± 0.41
	18	2	0	-64.8	-53.1 ± 16.6	-19.9 ± 3.10	-40.1 ± 23.3	6.03 ± 1.85	0.13 ± 0.18	-54.4 ± 18.3	10.7 ± 0.31
	19	5	0	-63.3	-25.1 ± 34.5	-5.00 ± 2.28	-35.1 ± 25.5	15.0 ± 11.9	0.00 ± 0.00	-25.1 ± 34.5	26.4 ± 0.98
	20	1	0	-56.9	-56.9 ± 0.00	-16.4 ± 0.00	-51.5 ± 0.00	10.9 ± 0.00	0.00 ± 0.00	-56.9 ± 0.00	11.0 ± 0.00
	21	20	0	-56.4	-0.42 ± 18.6	-3.99 ± 1.79	-5.25 ± 16.0	8.70 ± 7.33	0.01 ± 0.03	-0.54 ± 18.5	25.4 ± 0.74
	22	5	0	-54.4	-27.6 ± 23.3	-4.09 ± 0.70	-33.4 ± 25.4	9.93 ± 6.60	0.00 ± 0.01	-27.6 ± 23.2	25.6 ± 0.56
	23	5	0	-51.4	-3.21 ± 30.2	-2.27 ± 0.71	-13.3 ± 25.3	12.4 ± 14.1	0.00 ± 0.00	-3.21 ± 30.2	26.2 ± 0.72
	24	5	0	-36.4	-1.90 ± 23.2	-4.44 ± 2.29	-6.29 ± 14.6	8.84 ± 10.5	0.00 ± 0.00	-1.90 ± 23.2	27.0 ± 0.43
	25	6	0	-32.5	-3.31 ± 21.4	-4.21 ± 0.73	-9.73 ± 16.7	10.6 ± 9.62	0.01 ± 0.01	-3.36 ± 21.4	26.3 ± 0.72
	26	3	0	-32.0	-12.5 ± 19.9	-1.60 ± 0.56	-14.0 ± 173	3.03 ± 30.2	0.00 ± 0.01	-12.6 ± 19.9	27.9 ± 1.56
	27	38	0	-28.9	-1.84 ± 8.94	-5.26 ± 1.95	-3.04 ± 3.01	6.46 ± 8.30	0.00 ± 0.00	-1.84 ± 8.94	32.3 ± 0.84
	28	5	0	-24.2	-9.56 ± 13.4	-7.42 ± 3.07	-14.5 ± 15.5	12.4 ± 7.31	0.00 ± 0.00	-9.56 ± 13.4	25.8 ± 0.79
	29	2	0	-20.3	-10.3 ± 14.2	-6.16 ± 0.75	-29.4 ± 2.42	-1.18 ± 15.9	0.00 ± 0.00	-10.3 ± 14.2	26.3 ± 0.33
	30	3	0	-18.9	-9.01 ± 13.9	-7.20 ± 1.32	-6.47 ± 3.44	4.66 ± 15.4	0.00 ± 0.00	-9.01 ± 13.9	33.1 ± 0.63
	31	1	0	-17.3	-17.3 ± 0.00	-7.69 ± 0.00	-8.83 ± 0.00	-0.74 ± 0.00	0.00 ± 0.00	-17.3 ± 0.00	31.2 ± 0.00
	32	1	0	-12.8	-12.8 ± 0.00	-5.63 ± 0.00	-4.02 ± 0.00	-3.18 ± 0.00	0.00 ± 0.00	-12.8 ± 0.00	31.4 ± 0.00
	33	2	0	-12.3	-0.93 ± 16.0	-5.33 ± 0.92	-11.6 ± 13.3	14.4 ± 1.48	0.15 ± 0.21	-2.44 ± 13.9	25.9 ± 0.02
	34	25	0	-11.9	4.20 ± 13.7	-4.86 ± 2.2.0	0.25 ± 2.14	8.70 ± 12.8	0.01 ± 0.04	4.10 ± 13.8	32.1 ± 0.80
	35	3	0	-11.9	-3.40 ± 7.88	-6.24 ± 0.31	0.42 ± 2.37	2.42 ± 5.89	0.00 ± 0.00	-3.40 ± 7.88	32.5 ± 0.19
	36	6	0	-9.10	2.90 ± 9.04	-0.88 ± 0.42	-1.81 ± 1.32	5.60 ± 9.76	0.00 ± 0.00	2.90 ± 9.04	33.5 ± 0.73
	37	8	0	-8.15	13.2 ± 13.6	-1.54 ± 0.89	1.74 ± 4.97	13.0 ± 10.4	0.00 ± 0.00	13.2 ± 13.6	27.9 ± 0.90
	38	3	0	-7.89	3.92 ± 11.0	-1.14 ± 0.69	-7.79 ± 12.8	12.8 ± 1.40	0.00 ± 0.00	3.92 ± 11.0	27.5 ± 0.42
	39	1	0	-7.34	-7.34 ± 0.00	-0.82 ± 0.00	-14.3 ± 0.00	7.79 ± 0.00	0.00 ± 0.00	-7.34 ± 0.00	29.5 ± 0.00
	40	4	0	-7.04	3.85 ± 13.2	-6.65 ± 2.22	-0.29 ± 1.16	10.8 ± 11.0	0.00 ± 0.00	3.85 ± 13.2	32.5 ± 0.62
	41	2	0	-5.47	-1.09 ± 6.19	-2.11 ± 0.98	2.25 ± 2.53	-1.23 ± 4.64	0.00 ± 0.00	-1.09 ± 6.19	34.4 ± 0.54
	42	3	0	-4.59	0.89 ± 5.21	-1.61 ± 0.74	-0.97 ± 1.41	2.73 ± 6.81	0.07 ± 0.13	0.14 ± 6.42	30.8 ± 0.43

	43	1	0	-4.58	-4.58 ± 0.00	-6.07 ± 0.00	7.42 ± 0.00	-5.93 ± 0.00	0.00 ± 0.00	-4.58 ± 0.00	32.6 ± 0.00
	44	1	0	-3.83	-3.83 ± 0.00	-4.14 ± 0.00	-0.05 ± 0.00	0.35 ± 0.00	0.00 ± 0.00	-3.83 ± 0.00	34.1 ± 0.00
	45	6	0	-3.27	7.66 ± 10.5	-3.45 ± 1.35	6.40 ± 4.06	4.71 ± 6.87	0.00 ± 0.00	7.66 ± 10.5	27.4 ± 0.75
	46	1	0	-0.89	-0.89 ± 0.00	-0.06 ± 0.00	0.34 ± 0.00	-1.17 ± 0.00	0.00 ± 0.00	-0.89 ± 0.00	31.6 ± 0.00
	47	2	0	-0.67	9.61 ± 14.5	-0.73 ± 0.51	-0.59 ± 0.37	10.9 ± 14.4	0.00 ± 0.00	9.61 ± 14.5	33.6 ± 0.24
	48	3	0	0.69	11.7 ± 9.74	-2.83 ± 0.40	-0.22 ± 1.49	14.8 ± 8.46	0.00 ± 0.00	11.7 ± 9.74	32.4 ± 0.29
	49	2	0	0.91	9.87 ± 12.7	-0.59 ± 0.09	-23.8 ± 7.51	34.3 ± 5.24	0.00 ± 0.00	9.87 ± 12.7	27.0 ± 0.16
	50	1	0	0.95	0.95 ± 0.00	-0.49 ± 0.00	-1.27 ± 0.00	2.71 ± 0.00	0.00 ± 0.00	0.95 ± 0.00	33.1 ± 0.00
	51	2	0	1.89	3.45 ± 2.20	-2.43 ± 2.68	2.33 ± 0.69	3.55 ± 0.21	0.00 ± 0.00	3.45 ± 2.20	27.7 ± 0.54
	52	1	0	2.58	2.58 ± 0.00	-0.14 ± 0.00	0.28 ± 0.00	2.44 ± 0.00	0.00 ± 0.00	2.58 ± 0.00	27.5 ± 0.00
	53	1	0	3.64	3.64 ± 0.00	-2.06 ± 0.00	-2.88 ± 0.00	8.58 ± 0.00	0.00 ± 0.00	3.64 ± 0.00	26.9 ± 0.00
	54	2	0	3.87	7.08 ± 4.54	-5.74 ± 3.05	9.61 ± 0.30	3.22 ± 1.78	0.00 ± 0.00	7.08 ± 4.54	33.8 ± 0.76
	55	1	0	6.16	6.16 ± 0.00	-1.84 ± 0.00	2.27 ± 0.00	5.73 ± 0.00	0.00 ± 0.00	6.16 ± 0.00	25.8 ± 0.00
	56	2	0	6.40	11.3 ± 6.92	-1.15 ± 0.01	0.56 ± 0.13	11.9 ± 7.06	0.00 ± 0.00	11.3 ± 6.92	32.7 ± 0.42
	57	1	0	8.59	8.59 ± 0.00	-3.43 ± 0.00	13.7 ± 0.00	-1.67 ± 0.00	0.00 ± 0.00	8.59 ± 0.00	33.9 ± 0.00
	58	1	0	10.7	9.77 ± 0.00	-4.73 ± 0.00	18.3 ± 0.00	-3.80 ± 0.00	0.00 ± 0.00	9.77 ± 0.00	34.2 ± 0.00
	1	22	0	-194	-119 ± 34.5	-11.3 ± 4.31	-119 ± 30.2	12.0 ± 9.30	0.00 ± 0.00	-119 ± 34.5	20.5 ± 1.05
	2	5	0	-181	-122 ± 36.0	-8.18 ± 2.76	-129 ± 40.4	15.3 ± 5.69	0.00 ± 0.00	-122 ± 36..0	9.50 ± 1.05
	3	10	0	-168	-117 ± 36.4	-9.68 ± 4.42	-115 ± 36.2	7.58 ± 6.94	0.00 ± 0.01	-117 ± 36.4	21.4 ± 0.85
	4	17	0.01	-159	-102 ± 38.1	-10.6 ± 2.67	-106 ± 36.3	14.6 ± 7.02	0.01 ± 0.06	-102 ± 38.1	7.23 ± 0.85
	5	9	0	-158	-113 ± 18.6	-12.1 ± 2.55	-112 ± 13.8	11.0 ± 7.19	0.00 ± 0.00	-113 ± 18.6	8.37 ± 1.19
	6	3	0	-141	-85.8 ± 52.1	-12.7 ± 0.94	94.8 ± 54.7	21.7 ± 16.9	0.00 ± 0.00	-85.8 ± 52.1	7.00 ± 0.40
	7	9	0.05	-141	-75.3 ± 37.6	-8.41 ± 2.01	-80.5 ± 38.3	13.6 ± 16.3	0.00 ± 0.00	-75.3 ± 37.6	6.79 ± 1.29
	8	4	0	-135	-61.4 ± 53.9	-9.02 ± 2.79	-57.2 ± 48.5	4.80 ± 6.18	0.01 ± 0.01	-61.4 ± 53.9	7.65 ± 0.23
	9	2	0	-134	-118 ± 23.5	-7.31 ± 3.75	-114 ± 30.7	3.23 ± 3.86	0.03 ± 0.04	-118 ± 23.1	7.56 ± 1.147
	10	11	0	-132	-86.2 ± 31.7	-5.96 ± 2.29	-92.2 ± 30.2	12.0 ± 8.18	0.00 ± 0.00	-86.2 ± 31.7	22.2 ± 0.62
	11	8	0	-131	-65.0 ± 38.0	-3.52 ± 1.40	-74.2 ± 35.2	12.8 ± 5.44	0.00 ± 0.00	-65.0 ± 38.0	23.6 ± 0.68
	12	4	0	-131	-111 ± 26.0	-12.0 ± 3.80	-112 ± 18.0	12.8 ± 6.06	0.00 ± 0.00	-111 ± 26.0	19.9 ± 0.83
	13	11	0	-129	-77.0 ± 50.9	-14.0 ± 4.16	-72.7 ± 49.8	9.67 ± 6.85	0.00 ± 0.00	-77.0 ± 50.9	8.36 ± 0.80
	14	7	0	-123	-75.2 ± 37.6	-6.98 ± 2.88	-74.4 ± 42.9	6.16 ± 7.04	0.00 ± 0.00	-75.2 ± 37.6	23.0 ± 0.66
	15	3	0	-122	-93.1 ± 29.4	-10.2 ± 1.78	-95.8 ± 32.6	12.5 ± 3.41	0.04 ± 0.05	-93.5 ± 29.0	9.95 ± 0.41
	16	60	0.87	-117	-75.3 ± 28.5	-21.5 ± 1.86	-63.1 ± 28.3	9.30 ± 9.80	0.00 ± 0.00	-75.3 ± 28.5	2.96 ± 0.69
T2	17	8	0	-114	-52.9 ± 33.5	-13.3 ± 2.34	-51.7 ± 31.6	11.9 ± 9.47	0.02 ± 0.05	-53.1 ± 33.7	8.44 ± 1.14
	18	1	0	-112	-112 ± 0.00	-10.0 ± 0.00	-111 ± 0.00	8.84 ± 0.00	0.00 ± 0.00	-112 ± 0.00	21.3 ± 0.00
	19	1	0	-112	-112 ± 0.00	-7.86 ± 0.00	-120 0.00	16.4 ± 0.00	0.00 ± 0.00	-112 ± 0.00	19.3 ± 0.00
	20	4	0.01	-107	-58.6 ± 56.1	-11.2 ± 1.98	-51.4 ± 59.0	4.04 ± 4.75	0.00 ± 0.00	-58.6 ± 56.1	6.74 ± 0.29
	21	2	0	-104	-72.9 ± 43.6	-6.21 ± 0.34	-75.3 ± 32.6	8.58 ± 10.6	0.00 ± 0.00	-72.9 ± 43.6	7.68 ± 0.80
	22	1	0	-104	-104 ± 0.00	-12.0 ± 0.00	-92.3 ± 0.00	0.71 ± 0.00	0.00 ± 0.00	-104 ± 0.00	7.29 ± 0.00
	23	5	0	-102	-75.2 ± 27.6	-5.10 ± 2.60	-82.8 ± 29.5	12.7 ± 4.54	0.00 ± 0.00	-75.2 ± 27.6	23.1 ± 0.88
	24	4	0	-102	-94.0 ± 5.46	-7.26 ± 1.12	-96.8 ± 1.07	10.0 ± 6.29	0.00 ± 0.00	-94.0 ± 5.46	11.7 ± 0.91
	25	1	0	-96.3	-96.3 ± 0.00	-14.0 ± 0.00	-85.0 ± 0.00	2.76 ± 0.00	0.00 ± 0.00	-96.3 ± 0.00	7.71 ± 0.00
	26	3	0.78	-94.9	-64.3 ± 29.0	-19.7 ± 3.75	-61.0 ± 35.1	16.4 ± 4.30	0.00 ± 0.00	-64.3 ± 29.0	2.00 ± 0.27
	27	2	0	-94.9	-50.6 ± 62.6	-9.21 ± 2.73	-56.3 ± 68.2	14.9 ± 2.82	0.00 ± 0.00	-50.6 ± 62.6	10.8 ± 0.66
	28	3	0	-91.8	-68.5 ± 34.6	-15.8 ± 1.59	-61.6 ± 26.9	8.89 ± 6.21	0.00 ± 0.00	-68.5 ± 34.6	10.4 ± 0.76
	29	1	0	-90.9	-90.9 ± 0.00	-4.90 ± 0.00	-93.5 ± 0.00	7.46 ± 0.00	0.00 ± 0.00	-90.9 ± 0.00	22.6 ± 0.00
	30	3	0	-88.7	-75.1 ± 14.6	-7.49 ± 0.49	-71.5 ± 16.0	3.91 ± 2.74	0.00 ± 0.00	-75.1 ± 14.6	21.7 ± 0.65
	31	2	0.12	-88.0	-72.0 ± 22.7	-14.1 ± 1.90	-64.3 ± 11.5	6.46 ± 9.36	0.00 ± 0.00	-72.0 ± 22.7	4.07 ± 0.04
	32	5	0.11	-86.8	-40.1 ± 35.0	-16.3 ± 1.65	-32.3 ± 42.6	8.46 ± 8.03	0.00 ± 0.00	-40.1 ± 35.0	6.24 ± 0.49
	33	1	0	-81.3	-81.3 ± 0.00	-8.66 ± 0.00	-88.3 ± 0.00	15.6 ± 0.00	0.00 ± 0.00	-81.3 ± 0.00	11.9 ± 0.00

	34	1	0	-80.8	-80.8 ± 0.00	-6.22 ± 0.00	-80.4 ± 0.00	5.80 ± 0.00	0.00 ± 0.00	-80.8 ± 0.00	26.0 ± 0.00
	35	4	0	-79.8	-43.6 ± 37.6	-6.80 ± 2.85	-50.0 ± 35.7	13.2 ± 3.49	0.00 ± 0.00	-43.6 ± 37.6	25.9 ± 0.83
	36	1	0	-75.7	-75.7 ± 0.00	-9.89 ± 0.00	-57.6 ± 0.00	-8.27 ± 0.00	0.00 ± 0.00	-75.7 ± 0.00	6.02 ± 0.00
	37	3	0	-72.7	-45.5 ± 24.6	-12.1 ± 3.82	-39.9 ± 14.2	6.26 ± 10.4	0.03 ± 0.05	-45.8 ± 38.0	14.1 ± 1.22
	38	3	0	-72.4	-41.6 ± 38.0	-3.97 ± 1.30	-45.5 ± 29.6	7.95 ± 12.4	0.00 ± 0.00	-41.6 ± 38.0	24.1 ± 0.11
	39	2	0	-67.0	-32.0 ± 49.5	-8.19 ± 0.97	-44.6 ± 46.7	20.8 ± 1.80	0.00 ± 0.00	-32.0 ± 49.5	13.7 ± 0.17
	40	1	0.14	-66.3	-66.3 ± 0.00	-20.5 ± 0.00	-55.0 ± 0.00	9.19 ± 0.00	0.00 ± 0.00	-66.3 ± 0.00	5.37 ± 0.00
	41	3	0.85	-66.2	-61.2 ± 7.30	-18.9 ± 1.04	-47.4 ± 7.64	4.91 ± 9.90	0.02 ± 0.04	-61.4 ± 0.00	1.47 ± 0.08
	42	2	0	-62.8	-33.6 ± 41.3	-5.09 ± 0.92	-38.3 ± 54.5	9.83 ± 14.1	0.00 ± 0.00	-33.6 ± 41.3	22.9 ± 0.07
	43	1	0	-62.3	-62.3 ± 0.00	-5.57 ± 0.00	-61.4 ± 0.00	4.66 ± 0.00	0.00 ± 0.00	-62.3 ± 0.00	22.8 ± 0.00
	44	1	0	-61.1	-61.1 ± 0.00	-10.4 ± 0.00	-68.8 ± 0.00	18.0 ± 0.00	0.00 ± 0.00	-61.1 ± 0.00	12.3 ± 0.00
	45	1	0.42	-60.5	-60.5 ± 0.00	-23.7 ± 0.00	-42.7 ± 0.00	5.95 ± 0.00	0.00 ± 0.00	-60.5 ± 0.00	2.65 ± 0.00
	46	20	0.34	-60.2	-21.5 ± 15.8	-19.3 ± 2.78	-8.60 ± 14.0	6.39 ± 8.84	0.00 ± 0.00	-21.5 ± 15.8	5.47 ± 0.79
	47	1	0.03	-54.2	-54.2 ± 0.00	-9.58 ± 0.00	-56.6 ± 0.00	12.0 ± 0.00	0.00 ± 0.00	-54.2 ± 0.00	5.06 ± 0.00
	48	1	0.13	-46.7	-46.7 ± 0.00	-20.7 ± 0.00	-19.6 ± 0.00	-6.39 ± 0.00	0.00 ± 0.00	-46.7 ± 0.00	3.99 ± 0.00
	49	2	0	-41.2	-36.9 ± 6.00	-5.20 ± 1.15	-36.6 ± 15.0	4.90 ± 7.81	0.00 ± 0.00	-36.9 ± 6.00	22.7 ± 1.05
	50	1	0	-37.0	-37.0 ± 0.00	-6.73 ± 0.00	-23.7 ± 0.00	-6.59 ± 0.00	0.00 ± 0.00	-37.0 ± 0.00	13.4 ± 0.00
	51	4	0	-32.5	0.46 ± 26.2	-3.20 ± 1.31	-8.55 ± 12.0	12.2 ± 21.2	0.00 ± 0.00	0.46 ± 26.2	25.0 ± 0.52
	52	1	0	-28.4	-28.4 ± 0.00	-4.71 ± 0.00	-13.7 ± 0.00	-9.98 ± 0.00	0.00 ± 0.00	-28.4 ± 0.00	16.7 ± 0.00
	53	3	0	-27.3	-8.93 ± 22.7	-8.25 ± 3.09	-12.0 ± 19.7	11.3 ± 4.34	0.00 ± 0.00	-8.93 ± 22.7	26.6 ± 0.82
	54	1	0	-25.4	-25.4 ± 0.00	-10.4 ± 0.00	-18.3 ± 0.00	3.27 ± 0.00	0.00 ± 0.00	-25.4 ± 0.00	9.93 ± 0.00
	55	2	0	-25.3	-9.36 ± 22.5	-4.17 ± 3.57	-8.20 ± 16.6	3.01 ± 2.30	0.00 ± 0.00	-9.36 ± 22.5	27.3 ± 0.24
	56	2	0.07	-23.8	-17.8 ± 8.43	-17.1 ± 0.04	-4.07 ± 1.86	3.92 ± 6.60	0.00 ± 0.00	-17.8 ± 8.43	7.23 ± 0.17
	57	2	0	-22.4	-13.5 ± 12.6	-2.73 ± 1.97	-14.0 ± 11.8	3.19 ± 2.79	0.00 ± 0.00	-13.5 ± 12.6	24.0 ± 0.38
	58	1	0	-21.6	-21.6 ± 0.00	-3.40 ± 0.00	-26.6 ± 0.00	8.40 ± 0.00	0.00 ± 0.00	-21.6 ± 0.00	25.5 ± 0.00
	59	1	0	-13.4	-13.4 ± 0.00	-16.8 ± 0.00	2.46 ± 0.00	0.92 ± 0.00	0.00 ± 0.00	-13.4 ± 0.00	8.96 ± 0.00
	60	1	0	-7.79	-7.79 ± 0.00	-4.47 ± 0.00	-13.0 ± 0.00	9.63 ± 0.00	0.00 ± 0.00	-7.79 ± 0.00	26.2 ± 0.00
	61	1	0	14.3	14.3 ± 0.00	-9.46 ± 0.00	18.7 ± 0.00	51.2 ± 0.00	0.00 ± 0.00	14.3 ± 0.00	10.2 ± 0.00
	62	1	0	27.0	27.0 ± 0.00	-1.69 ± 0.00	10.1 ± 0.00	18.6 ± 0.00	0.00 ± 0.00	27.0 ± 0.00	27.1 ± 0.00
DIS Binding Site	1	100	0.94	-131	-80.1 ± 28.0	-20.6 ± 2.80	-67.5 ± 24.4	7.91 ± 11.0	0.01 ± 0.10	-80.2 ± 27.8	2.42 ± 0.54
	2	52	0.99	-130	-74.2 ± 23.7	-18.4 ± 2.92	-65.2 ± 24.3	9.42 ± 8.36	0.00 ± 0.00	-74.2 ± 23.7	1.72 ± 0.28
	3	22	0.7	-124	-80.4 ± 35.8	-17.8 ± 2.82	-69.5 ± 34.8	4.14 ± 7.13	0.28 ± 0.74	-83.1 ± 32.8	2.94 ± 0.62
	4	19	0.87	-97.2	-58.9 ± 23.4	-21.5 ± 2.28	-48.1 ± 18.4	10.8 ± 10.7	0.00 ± 0.00	-58.9 ± 23.4	2.32 ± 0.45
	5	2	0.32	-53.8	-39.6 ± 20.1	-17.7 ± 0.22	-41.1 ± 19.7	4.30 ± 12.2	1.50 ± 1.20	-54.5 ± 32.1	2.80 ± 0.19
	6	1	0.47	-48.4	-48.4 ± 0.00	-22.0 ± 0.00	-24.9 ± 0.00	-1.48 ± 0.00	0.00 ± 0.00	-48.4 ± 0.00	1.88 ± 0.00
	7	4	0.42	-46.8	-41.0 ± 8.08	-16.9 ± 0.40	-40.1 ± 6.42	10.8 ± 5.79	0.51 ± 0.48	-46.1 ± 9.92	3.35 ± 0.66
All	1	177	0.65	-167	-32.3 ± 34.0	-15.9 ± 2.24	-40.8 ± 26.3	7.41 ± 9.52	1.70 ± 1.55	-49.3 ± 28.7	2.68 ± 0.41
	2	5	0.33	-81.7	-48.4 ± 27.8	-16.5 ± 1.19	-60.1 ± 29.1	10.8 ± 5.94	1.75 ± 0.83	-65.9 ± 26.4	2.79 ± 0.24
	3	11	0.16	-62.7	51.8 ± 85.0	-14.2 ± 3.14	-1.92 ± 28.3	5.13 ± 8.36	6.28 ± 7.18	-11.0 ± 31.0	4.33 ± 0.43
	4	4	0.32	-47.0	-14.9 ± 26.5	-13.8 ± 2.84	-35.5 ± 18.7	7.25 ± 5.48	2.71 ± 1.30	-42.0 ± 15.6	2.96 ± 0.12
	5	1	0.5	-36.6	-36.6 ± 0.00	-15.8 ± 0.00	-26.1 ± 0.00	4.54 ± 0.00	0.07 ± 0.00	-37.3 ± 0.00	1.64 ± 0.00
	6	2	0	286	303 ± 23.6	-12.8 ± 1.68	17.8 ± 30.1	10.9 ± 0.74	28.7 ± 0.56	-16.0 ± 29.1	5.15 ± 0.33

[*] Average ± standard deviation (S.D.) for each docked structure within the cluster. Clusters with a S.D. of 0 had only one docked structure populating the cluster.

[a] The fractional overlap for each cluster.

[b] The lowest HADDOCK score (H_{score}) found in the given cluster.

[c] The H_{score} was calculated for each structure and averaged.

[d] The total intermolecular energy ($E_{total} = E_{vdw}+E_{elec}+E_{desolv}+E_{dist}$) was calculated for each structure and averaged.

[e] The RMSD was calculated for each structure and averaged.

```
[LIGANDS]
LIG -1

[MUT]

; Residues within 4 angstroms of the DIS 603 binding site

114 | 1
115 | 1
116 | 1
140 | 1
141 | 1
142 | 1
144 | 1
145 | 1
159 | 1
160 | 1
161 | 1
163 | 1
164 | 1
180 | 1
181 | 1
182 | 1
184 | 1
185 | 1
186 | 1
187 | 1
188 | 1
189 | 1
191 | 1
192 | 1

[PRENMR]

; Ligand    | Receptor                | Distance

; Proton(s) | Residue Number , Atom Name | Min. Avg. Max. | Offset | Weight

HAE | 34 MN | 26.92 30.68 39.28 | 0 | 1 # T1
HAG | 34 MN | 22.09 23.68 32.28 | 0 | 1 # T1
HAF | 34 MN | 23.04 24.29 32.89 | 0 | 1 # T1
HAH | 34 MN | 25.10 25.80 34.40 | 0 | 1 # T1
HAE | 34 MN | 18.73 21.27 29.87 | 0 | 1 # T2
HAG | 34 MN | 17.58 17.93 26.53 | 0 | 1 # T2
HAF | 34 MN | 17.34 18.44 27.04 | 0 | 1 # T2
HAH | 34 MN | 18.09 18.53 27.13 | 0 | 1  # T2

[end]
```

Figure 4.S1. *Experimental input for the combination experimental input ASA-BSA HADDOCK-Vina simulation.* [MUT] experimental input containing the 18 residues in the DIS3 binding site with a weight of 1. The [PRENMR] experimental input containing both the T_1 and T_2 distances. The ligand protons are listed, the Mn on residue 34, the r_{min}, r_{max} and the r_{max} + 8 Å. There are no offsets and each restraint has a weight of 1. The comments are shown in black and the simplified language used to produce the CNS-compatible distance restraints are shown in **red**.

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 SUMMARY

5.1.1 The molecular mechanism of ETT and STT transport by P-glycoprotein.

In this dissertation, the structural basis of eletriptan (ETT) and sumatriptan (STT) transport by the multidrug resistant transporter P-glycoprotein (Pgp) was investigated. The triptan class of drugs is used to treat migraines by acting on the Serotonin 1B/1D ($5HT_{1B/1D}$) receptors in the CNS.[89–91] Despite their structural similarities (Figure 2.1), ETT was shown to be actively removed from the CNS by Pgp,[69, 73] while STT showed little to no Pgp-mediated transport in cell-based assays.[68, 69, 73, 92] These studies suggest that the chemical features found solely on ETT lead to an increase in the rate of Pgp-mediated transport. Due to their large differences in efflux ratios, we hypothesized that the observed increase in ETT-induced transport was due to the log P. ETT is more lipophilic, with a log P of 4.1, while STT had a log P of 0.90.[95] To test this hypothesis, the Pgp-mediated ATPase activity was probed in the presence of ETT and STT (Figure 2.2). ETT was shown to stimulate ATPase activity and saturate the transporter at 125 μM, with a K_m of 14.8 ± 3.1 μM. In contrast, STT maintained basal ATPase activity and could not be fit to the Michaelis-Menten equation. However, this difference in ATPase activity allowed us to determine if the drugs share a similar binding site on Pgp. ETT and STT displayed non-competitive inhibition, revealing the two compounds occupy distinct sites within the TMD of Pgp, despite their structural similarities (Figure 2.2).

The drugs occupying two distinct sites on the transporter and the variations in their logP suggest that ETT and STT would have differences in their binding affinity on Pgp. However, using intrinsic protein fluorescence quenching, ETT and STT were found to have similar binding affinities on reconstituted Pgp, 115 ± 10 μM and 92.6 ± 25 μM, respectively (Figure 2.3). These data suggest that the driving force for ETT and STT binding was not lipophilicity but another feature determined by the SAR studies.[123] To identify the functional groups on ETT that stimulates Pgp-mediated ATPase activity and transport, saturation transfer double difference (STDD) NMR was employed on reconstituted Pgp.[113, 114] The benzene ring on ETT was found to have the highest interaction with the transporter, a structural feature found solely on ETT (Figure 2.4), suggesting the aromaticity of Pgp substrates identified by the SAR study may be the driving force of ETT binding and Pgp-mediated efflux.

Upon substrate binding, Pgp undergoes large conformational changes to render ATP binding; thus, shifting the transporter into the closed conformation and promoting substrate release. In this study, our goal was to determine if the ETT-bound conformation was more favorable for ATP binding than the STT-bound by determining the change in tryptophan accessibility (ΔK_{SV}) between the substrate-bound and closed-conformations. In the absence of drug, the transporter undergoes basal ATPase activity. This transition from the open to closed conformation resulted in a ΔK_{SV} of 0.68 M^{-1} (Figure 2.5A). Upon the addition of STT, the STT-bound conformation undertakes large conformational changes to reach the closed-state, producing a ΔK_{SV} of 1.44 ± 0.20 M^{-1} (Figure 2.5B). In contrast, the ATP-induced closed conformation of the ETT-bound complex required little to no change, with a ΔK_{SV} of -0.10 ± 0.09 M^{-1} (Figure 2.5C). These data suggest that ETT binding induces a favorable conformation for ATP binding, while the STT-bound conformation undergoes large conformational rearrangement.

Using these findings, previous transport studies[68, 69, 73, 92] and X-ray crystal structures of Pgp,[43, 45, 46] a simplified transport model was developed (Figure 2.6). In the absence of drugs, Pgp adopts an open conformation, as observed in the *m*Pgp X-ray crystal structures (Figure 2.6A). Upon the addition of ATP, Pgp undergoes basal ATPase activity; thus, Pgp was depicted to be in an intermediate conformation (Figure 2.6B). STT binding increases tryptophan accessibility and maintains basal ATPase activity. Therefore, STT-binding may induce a more "open" conformation, as observed in X-ray crystal structures of Pgp in complex with inhibitors (Figure 2.6C).[46, 134] Upon ATP binding, STT maintains basal ATPase activity and adopts an intermediate conformation, similar to unbound Pgp (Figure 2.6D). ETT induces ~2-fold activation in Pgp-mediated ATPase activity and the tryptophan accessibility of the ETT-bound conformation is similar to the closed-conformation of Pgp. Therefore, ETT binding shifts Pgp in an intermediate conformation (Figure 2.6E). This conformation is favorable for ATP binding, leading to an increase in Pgp-mediated ATPase activity (Figure 2.6F). We believe this favorable ETT-bound conformation is distinct from STT binding because of the benzene ring found on ETT. STT has no additional aromaticity to the structure and may bind in a site that induces a conformational change that is unfavorable for ATP binding. Therefore, STT transport would be minimal, as observed in previous transport studies.[68, 69, 73, 92]

5.1.2 The data-driven molecular docking wrapper: HADDOCK-Vina.

ETT and STT were identified to occupy distinct sites on the transporter and induce conformational changes that correlate to the differences in their observed transport rates.[68, 69, 73, 92] Identifying the ETT and STT binding sites would provide better understanding on the structural basis of these differences. To date, solving substrate-bound structures of Pgp has been unsuccessful

by X-ray crystallography and CryoEM.[45, 46, 50, 88] Due to this roadblock, using molecular docking to identify the ETT and STT binding sites onto the transporter would provide key structural information on Pgp-mediated triptan transport. Using experimental data to drive the molecular docking would be advantageous in achieving an accurate triptan-bound Pgp model. Therefore, data-driven molecular docking approaches were explored to model the ETT and STT binding sites.

HADDOCK is a data-driven software that was developed to dock and model protein-protein interactions.[150, 182, 263] Recently, HADDOCK protocols were adapted for ligand-protein docking;[150, 182, 263] however, the experimental data interface for ligand-protein docking is limited. Users must input specific ligand atoms that interact with the receptor to proceed with the molecular docking.[263] This is problematic when the ligand atoms interacting with the transporter are unknown, in the case of site-directed mutagenesis or HSQC NMR data. Furthermore, small molecules can become stuck on the surface when using HADDOCK for ligand-protein docking. To overcome these limitations, HADDOCK was combined with AutoDock Vina[143] to make the wrapper HADDOCK-Vina. A customizable version of AutoDock Vina was used for initial docking to effectively coat the surface of the receptor, including narrow crevasses and deep ligand binding pockets. Additionally, HADDOCK-Vina as well as incorporates PROPKA[166] to accurately set the protonation states of polar side chains and has a simplified experimental interface to input ligand-protein interactions.

The effectiveness of HADDOCK-Vina was demonstrated by docking three previously determined X-ray crystal structures: imatinib onto Abl tyrosine kinase (PDB ID: 2HYY[185]), KNI-272 onto HIV-1 protease (PDB ID: 3FX5[186]) and AMP onto *E. Coli* Adk (PDB ID: 1ANK[187]). These three ligand-bound X-ray crystal structures were selected for benchmarking because they have experimental data that can be used to drive the molecular docking. To use the receptors for

the simulations, missing atoms and loops were modelled back onto the receptors using MODELLER. This also relaxed the sidechains and eliminated the ligand-bound orientation of these residues, which reduced any bias in our docking simulation. Furthermore, ligands used for docking were taken from PubChem,[95] eliminating any conformation or coordinate bias of the ligand. From our molecular docking simulations, combining experimental data for all three cases led to a low RMSD cluster (Figure 3.2). Additionally, the most populated cluster from the combination simulations were the lowest H_{score} cluster, i.e. cluster #1. This population distribution was large, with the lowest H_{score} cluster having a much larger population than the subsequent clusters. The lowest H_{score} for all three cases had favorable E_{total}'s, E_{dist}'s and Q-Factors, which suggest excellent correlation to the experimental data used for the docking simulation (Table 3.1). These lowest H_{score} clusters encased the ligand X-ray crystal structure and had $f_{overlap}$'s >85% (Figure 3.2). Also, the average RMSD with the lowest H_{score} cluster for all three cases were ~2 Å. The lowest RMSD structures within these clusters were <2 Å and showed good overlap to the original structure (Figure 3.2).

The accuracy of HADDOCK-Vina was compared to HADDOCK and AutoDock Vina with minimal modifications to default parameters, allowing direct comparison to HADDOCK-Vina. In all three cases, the HADDOCK-VINA simulation was more accurate than HADDOCK and AutoDock Vina (Table 3.2). HADDOCK-Vina produced clusters that overlapped with the X-ray ligand structure and produced RMSDs <2 Å. Unlike HADDOCK, which produced <46% overlap and RMSD values 3.29 to 13.7 Å. The lowest AutoDock Vina structures produced highly unfavorable H_{score}'s and E_{total}'s, with RMSD values ranging from 2.94 to 5.63 Å. These studies demonstrated that the molecular docking wrapper HADDOCK-Vina can produce more accurate

and energetically favorable models using a combination of experimental data when compared to HADDOCK and AutoDock Vina.

5.1.3 Docking the high affinity acetylsalicylic acid binding site on bovine serum albumin using HADDOCK-Vina.

Paramagnetic relaxation enhancement (PRE) NMR is a technique used to determine the binding sites of small molecules on receptors. The changes in the relaxation rate of ligand nuclei are detected in the presence of paramagnetic centers in a distance-dependent manner. This method will be employed to accurately determine the binding sites of ETT and STT on Pgp and the PRE distances will be used to guide the molecular docking by HADDOCK-Vina. However, the parameterization and methodology of using a paramagnetic label covalently attached to a receptor in HADDOCK-Vina has not been established.

To demonstrate this method, acetylsalicylic acid (ASA) was modeled on bovine serum albumin (BSA) using sparse PRE NMR data to drive the molecular docking by HADDOCK-Vina. ASA occupies two sites on BSA, a high and weak affinity site, 2.56 ± 0.014 μM and 116 ± 28.3 μM, respectively (Figure 4.3). To date, there are no ASA-BSA X-ray crystal structures available. Fortunately, the structurally similar 3,5-diiodosalicylic acid (DIS)-BSA structure has been determined.[214] To identify the high affinity site, this site was selectively populated by collecting NMR spectra at a 1:1 stoichiometry. The paramagnetic label, Mn^{2+}-EDTA, was covalently attached to the single free cysteine found on BSA, Cys34. The T_1 and T_2 relaxation rates of the ASA nuclei were measured in the presence and absence of the Mn^{2+}-EDTA label (Figure 4.4). Both measurements identified the ASA molecule to occupy a site ~23 Å from the label (Table 4.1). To determine if the high affinity site overlaps with a DIS binding site, ASA T_1 enhanced relaxation

rates were monitored in the presence of excess DIS. The ASA T_1 relaxation rates were eliminated in the presence of DIS, suggesting competitive displacement at a shared binding site (Figure 4.4). Examining the Dis-BSA crystal structure (PDB ID:4JK4[214]) revealed one DIS binding site that is within the experimentally determined T_1 and T_2 distances from the paramagnetic label, DIS3. Therefore, the residues surrounding the DIS3 site were used as an experimental restraint for docking, as well as the T_1 and T_2 distances.

The individual experimental restraints (T_1, T_2 and DIS3 binding site) and a combination of all restraints were used to dock ASA onto BSA. From our previous HADDOCK-Vina study in Chapter 3, the most accurate docking simulation was from the lowest H_{score} cluster. Additionally, this cluster had a higher population compared to the other clusters produced. For simulations using T_1 experimental restraints, the most populated cluster was the lowest H_{score} cluster (Figure 4.5B). However, the lowest H_{score} cluster was not the most populated for the simulations using solely T_2 experimental restraints (Figure 4.5D). For both the T_1 and T_2 simulations, multiple clusters produced from the simulation that had a E_{dist} of 0. These clusters were within the experimental data but mainly lie on the surface of the receptor (Figure 4.5C,E). Additionally, 58 and 62 clusters were produced for the T_1 and T_2 simulations, respectively (Table 4.S1). Therefore, it is unclear which cluster produced the most accurate ASA-docked structure for these simulations.

Using the residues surrounding the DIS3 binding site greatly reduced the number of clusters produced, with the lowest H_{score} cluster being the most populated; however, the cluster #2 population was half of cluster #1 and had a E_{dist} of 0 (Figure 4.5F). Furthermore, all the clusters produced from this experimental restraint were within subdomain IB and overlapped with the DIS3 molecule, making it unclear which cluster is the most accurate (Figure 4.5G). To clarify this ambiguity, all three experimental restraints were combined. Here, the lowest H_{score} cluster was the

most populated (177 ASA molecules) and this population was much larger than subsequent clusters (Figure 4.5H). Similar to the DIS residue simulation, the clusters from this simulation were found in subdomain IB (Figure 4.5I). While the average E_{dist} for the lowest H_{score} cluster was the largest of the simulations, the lowest H_{score} structure within this cluster had a E_{dist} of 0.14, indicating excellent correlation to the combination of experimental data. This structure had a H_{score} of -167 A.U. and an E_{total} of -168 kcal mol^{-1}. The RMSD of this docked-ASA molecule compared to DIS3 was 1.57 Å. This ASA molecule was docked into subdomain IB, with hydrogen bond interactions between Arg185 and Lys114. (Figure 4.6C) The Arg185 residue that interacts with DIS3 rotates ~3 Å to achieve hydrogen bonding with ASA (Figure 4.6B,C). Additionally, Tyr137 and Tyr160 that hydrogen bond with DIS3 rotate to satisfy pi-pi stacking in the absence of DIS3. (Figure 4.6B,C) In this study, using sparse PRE NMR data to dock a small-molecule onto a receptor with a covalently attached paramagnetic label proved to be feasible and accurate by HADDOCK-Vina.

5.1.4 The ETT and STT binding sites using sparse experimental data to drive molecular docking by HADDOCK-Vina.

HADDOCK-Vina was shown to be an accurate method in docking small molecules onto receptors using sparse experimental data. While PRE NMR experiments were not implemented for ETT and STT, preliminary docking simulations of the triptan drugs with Pgp were performed. The large intracellular loop connecting the two halves of the transporter and missing atoms at the N- and C-terminal were added back onto *m*Pgp using MODELLER.[189, 264] The experimental inputs used to dock ETT and STT are shown in Figures 5.1 and 5.2, respectively. Briefly, 126 residues that line the large binding cavity were used as experimental restraints. These residues were inputted

into the [SOI] experimental input and given a 20 Å distance range due to the large size of the binding cavity. Additionally, the STDD of the nuclei observed for both ETT and STT were inputted in the [STD] experimental input. Finally, the molecular features that lead to Pgp recognition from the SAR studies[123] were used as experimental restraints. The aromatic functional groups, tertiary amines and the hydrogen bond donor/acceptors were entered in the [SAR] section of the experimental input. For the docking of ETT and STT, default HADDOCK-Vina parameters were used except the weight repulsion, which was changed to 0.3 to increase ligand coverage on the transporter, as done previously in Chapters 3 and 4.

Of the 1000 molecules used to dock ETT onto Pgp, the top two AutoDock Vina conformations were selected for HADDOCK refinement because they had a $H_{score} = 0$. From these two ligands, 14 clusters were produced and the lowest H_{score} cluster was the most populated (Figure 5.3A). From Chapter 3 and 4, we know that the lowest H_{score} cluster will be the most populated, with the remaining clusters having a much lower population resulting in a large distribution of clustering. Here, cluster #2 has a population almost equal of cluster #1, providing some ambiguity within the docking (Figure 5.3A). Additional experimental data would be needed to increase the population distribution; however, cluster #1 is more favorable than cluster #2, with an average E_{total} of -124 ± 21.3 kcal mol^{-1} (Table 5.1). The lowest H_{score} structure within the docking simulation is in cluster #1, -130 A.U., and is shown in panel C. This ETT molecule interacts with transmembrane (TM) helices 2, 3, and 10, 11, which form ICLs 1 and 4, respectively. ETT-induced conformational changes on these TM helices would translate to the IH1/IH4/NBD1 interface, thereby stimulating Pgp-mediated ATPase activity as seen in crosslinking studies of the NBDs.[3, 11] This favorable binding and induced conformational changes provides a rationale for the observed increase in ETT transport by Pgp. In this site, the benzene on ETT has direct interaction

with hydrophobic residues within the cavity, i.e. L875 and A879. The protonated Lys883 on TM 10 is ~1.7 Å from the sulfonyl groups on ETT and would promote hydrogen bonding interactions. The Asp184 residue on TM 2 is ~5 Å from the nitrogen on the indole ring of ETT, forming electrostatic interactions. From this preliminary docking study, ETT binding is driven by hydrogen bonding, hydrophobic interactions and electrostatic interactions.

For the docking of STT, the default top three AutoDock Vina of the 1000 initial structures were selected for HADDOCK refinement. From this simulation, 15 clusters were produced (Figure 5.3D). The lowest H_{score} cluster was also the most populated with 97 STT molecules; however, the population distribution reveals some ambiguity with the docking. The population of cluster #5 was almost as populated as cluster #1 with 76 molecules. Additional experimental data may relieve this ambiguity; however, by examining the average E_{total}, the binding of STT within cluster #1 was more favorable than cluster #5, with an average E_{total} of -118 ± 18.4 kcal mol^{-1} (Table 5.1). Viewing the lowest H_{score} cluster, i.e. cluster #1, STT was bound at the upper region on the binding cavity (Figure 5.3E). From this cluster, the lowest H_{score} STT molecule (-134 A.U.) had the indole ring interacting with TM 12 (Figure 5.3F). The S989 residue on TM12 drives this protein-ligand binding through electrostatic interactions with the amine found on the indole ring, ~4 Å. The Q343 residue on TM6 is ~1.8 Å from the sulfonyl group on STT and would form hydrogen bond interactions. Interestingly, both TM6 and TM12 do not form the ICLs found on Pgp, as described in Chapter 1; however, TM6 and TM12 form the interface for substrate binding,[265, 266, 267] Additionally, previous cross-linking studies of TM6 to TM12 showed that TM6 rotates upon ATP binding to form new residue contacts with TM12.[268] If STT binding in this interface locks TM6, the transporter would require large conformational changes to rotate upon ATP binding and promote changes in the IH1/IH4/NBD1 interface. Inability to produce conformational changes

would lead to a large energetic barrier to reach the closed conformation for transport, as seen in our acrylamide quenching data.

5.2 FUTURE DIRECTIONS

While these preliminary docking simulations of ETT and STT onto Pgp provide insight into the structural basis of Pgp-mediated transport, the population of the clusters adds uncertainty to the precision of the model. The addition of unambiguous experimental restraints would more accurately determine the binding sites of ETT and STT on Pgp. Therefore, PRE NMR will be used to determine the binding sites of ETT and STT onto Pgp. These distances will be used as experimental restraints for docking, as done with the ASA-BSA model in Chapter 4. To do this, single cysteine mutants would be engineered onto Pgp to allow paramagnetic label attachment.

Pgp has been previously demonstrated to be robust against site-directed mutagenesis with little to no changes in ATPase activity or overall structure.[9, 43, 265, 269] Therefore, cysless Pgp would be engineered with all cysteines mutated to alanine, as done previously.[270] Single cysteine residues would be reintroduced to the cysless transporter, allowing paramagnetic label attachment at these positions. To improve the expression screening and stability of the transporter, cysless Pgp variants will be engineered with a C-terminal green fluorescent protein (GFP) by a linker containing a TEV protease cleavage site. Addition of a GFP tag has been previously demonstrated to provide increased expression levels and allow for rapid detection of successful plasmid integration into the *P. pastoris* choromosome.[271]

To accurately identify the ETT and STT binding sites by PRE NMR, at least three mutants would be required to triangulate the binding sites of the drugs on the transporter. The paramagnetic label is effective in measuring relaxation rates at a maximum distance of ~35 Å.[232] Therefore,

selection of the mutants needed to measure the relaxation rates of ETT and STT is critical for accurate detection. Using the preliminary docking model may provide valuable insight in selecting mutants for PRE studies. Residues within 35 Å of the preliminary model will be selected for cysteine mutation; however, residues that lie in the binding cavity will be avoided as to not disturb the triptan binding site or lipid interactions.

Despite the robust nature of Pgp, the overall fold and activity of these mutants will be compared to wildtype Pgp by circular dichroism and Pgp-mediated ATPase activity of the probe substrate verapamil. Once these mutants are confirmed to be fully functional, the Mn^{2+}-EDTA and Zn^{2+}-EDTA labels will be incubated with the transporter at a 1:1 ratio, as done previously in Chapter 4. After extensive dialysis, the T_1 and T_2 relaxation rates will be measured for all three mutants for ETT and STT binding. These distances would be inputted into the [PRE] experimental interface for HADDOCK-Vina. The T_1 and T_2 distances would be combined with the previous experimental restraints, [SOI], [STD] and [SAR]. We hypothesize that the ETT and STT binding sites determined by the HADDOCK-Vina simulation using these combined restraints would be similar to the preliminary docking study, with STT binding in the upper cavity of Pgp and ETT interacting with TM helices that contribute to the IH1/IH4/NBD1 interface.

To validate the ETT and STT binding sites determined by HADDOCK-Vina, mutants will be introduced to disrupt ETT and STT binding interactions. Again, the overall structure and Pgp-mediated ATPase activity will be evaluated for the Pgp mutants and compared to wildtype. As long as the mutants are stable and active, the effect of the mutants will be evaluated with ETT and STT. Since STT maintains basal ATPase activity, measuring the effect of the mutation in the STT binding site by Pgp-mediated ATPase activity would yield little to no changes. Therefore, the changes in binding affinity of ETT and STT will be probed by protein fluorescence quenching of

reconstituted Pgp mutants. We anticipate that mutations introduced to disrupt ETT and STT binding would reduce or eliminate the binding affinity of these drugs, thus corroborating our ETT- and STT-bound Pgp models determined by HADDOCK-Vina.

5,3 RESEARCH OUTLOOK

Using sparse experimental data to guide the molecular docking of ETT and STT by HADDOCK-Vina has provided a rationale and the structural basis for the large differences in the observed Pgp-mediated transport of the triptan class of drugs. This method presents an advantageous approach to understand protein-ligand interactions driving transport and can improve the drug development process. Pgp is a major hurdle in the pharmaceutical industry by preventing many therapeutic compounds from reaching target sites, profoundly in multidrug resistant tumors overexpressing Pgp. Efforts to relieve Pgp-mediated transport have been to inhibit or circumvent the transporter and to date, these approaches have been unsuccessful in the clinic. By using biophysical and biochemical data as experimental restraints, Pgp substrates and inhibitors can be docked accurately onto the transporter by HADDOCK-Vina; thus, eliminating the requirement of a substrate-bound structure and greatly improving the speed and accuracy of identifying key protein-ligand interactions driving Pgp-mediated transport. While Pgp inhibition has been unsuccessful, this methodology can be used to design selective inhibitors of the IH2/IH3/NBD2 interface and rapidly screen compounds for effectiveness. By understanding the protein-ligand interactions that lead to Pgp recognition by HADDOCK-Vina, rational drug modification of Pgp substrates can be employed to circumvent Pgp-mediated transport. We believe this approach will aid in the development of novel drugs that circumvent or inhibit Pgp-mediated transport and mitigate the multidrug resistance and drug disposition of P-glycoprotein.

5.4 TABLES

Table 5.1 *Cluster analysis of all clusters produced by HADDOCK-Vina for ETT and STT docking onto P-glycoprotein.*

Triptan	Cluster Number	Number in Cluster	H_{score} (A.U.)[a]	Average H_{score} ± S.D. (A.U.)[b,*]	Average E_{vdw} ± S.D. (kcal/mol)[*]	Average E_{elect} ± S.D. (kcal/mol)[*]	Average E_{desolv} ± S.D. (kcal/mol)[*]	Average E_{dist} ± S.D. (kcal/mol)[*]	Average E_{total} ± S.D. (kcal/mol)[c,*]
ETT	1	84	-130	-79.4 ± 24.3	-38.1 ± 6.18	-53.1 ± 19.3	-32.5 ± 12.0	4.42 ± 1.88	-124 ± 21.3
	2	65	-125	-70.1 ± 26.1	-43.8 ± 4.01	-37.0 ± 13.0	-31.3 ± 11.8	4.21 ± 2.00	-112 ± 18.1
	3	23	-114	-56.6 ± 21.0	-47.4 ± 3.93	-8.62 ± 5.24	-38.4 ± 11.4	3.77 ± 1.83	-94.4 ± 13.1
	4	2	-112	-82.6 ± 9.11	-32.3 ± 2.23	-43.7 ± 5.80	-29.5 ± 12.2	2.28 ± 0.66	-105 ± 15.8
	5	7	-110	-43.5 ± 32.1	-36.0 ± 1.86	-9.38 ± 7.23	-41.7 ± 16.7	4.36 ± 2.75	-87.1 ± 22.2
	6	6	-110	-22.3 ± 69.0	-31.0 ± 5.58	-49.9 ± 11.4	-31.4 ± 6.77	9.00 ± 6.84	-112 ± 8.21
	7	1	-108	-74.3 ± 0.00	-42.3 ± 0.00	-40.2 ± 0.00	-41.1 ± 0.00	5.22 ± 0.00	-127 ± 0.00
	8	1	-106	-70.5 ± 0.00	-33.7 ± 0.00	-23.7 ± 0.00	-46.5 ± 0.00	3.34 ± 0.00	-104 ± 0.00
	9	1	-104	-68.3 ± 0.00	-42.4 ± 0.00	-68.3 ± 0.00	-18.7 ± 0.00	6.10 ± 0.00	-129 ± 0.00
	10	3	-101	-32.5 ± 37.8	-37.7 ± 2.12	-10.6 ± 4.83	-23.0 ± 3.44	3.88 ± 3.78	-71.4 ± 6.28
	11	1	-98.8	-55.6 ± 0.00	-26.7 ± 0.00	-27.9 ± 0.00	-48.5 ± 0.00	4.75 ± 0.00	-103 ± 0.00
	12	2	-97.1	-22.4 ± 28.5	-31.3 ± 3.78	3.80 ± 1.27	-36.1 ± 2.79	4.13 ± 3.08	-63.6 ± 2.25
	13	3	-96.8	-27.9 ± 10.1	-37.5 ± 2.84	-4.80 ± 9.14	-33.8 ± 4.42	4.82 ± 1.87	-76.1 ± 8.84
	14	1	-96.1	73.8 ± 0.00	-27.3 ± 0.00	-4.49 ± 0.00	-14.9 ± 0.00	12.0 ± 0.00	-46.6 ± 0.00

STT	1	97	-134	-82.3 ± 23.1	-33.3 ± 3.11	-51.0 ± 13.3	-33.4 ± 12.6	3.54 ± 1.82	-118 ± 18.4
	2	46	-130	-89.8 ± 23.7	-31.9 ± 2.94	-46.8 ± 10.9	-33.6 ± 12.4	2.25 ± 1.46	-112 ± 17.1
	3	2	-107	-93.4 ± 18.8	-27.0 ± 1.49	-54.6 ± 12.7	-32.5 ± 2.34	2.07 ± 0.53	-114 ± 13.5
	4	21	-104	-70.5 ± 24.8	-29.6 ± 3.78	-33.5 ± 13.7	-36.6 ± 11.4	2.92 ± 1.81	-99.7 ± 18.6
	5	76	-101	-60.4 ± 22.8	-27.5 ± 4.03	-23.9 ± 16.8	-39.0 ± 12.2	3.00 ± 1.64	-90.5 ± 18.7
	6	2	-100	34.1 ± 189	-17.5 ± 2.51	-41.6 ± 37.3	-42.1 ± 5.71	13.5 ± 15.5	-101 ± 34.1
	7	27	-96.9	-59.2 ± 20.8	-27.9 ± 3.56	-44.9 ± 13.0	-28.9 ± 11.8	4.26 ± 1.66	-102 ± 18.0
	8	15	-96.7	-55.3 ± 21.7	-29.4 ± 4.18	-36.3 ± 16.4	-37.0 ± 9.46	4.74 ± 1.95	-103 ± 16.0
	9	1	-84.2	-84.2 ± 0.00	-25.5 ± 0.00	-67.5 ± 0.00	-42.3 ± 0.00	5.10 ± 0.00	-135 ± 0.00
	10	7	-82.6	-63.2 ± 20.6	-26.74 ± 3.53	-42.9 ± 5.10	-28.8 ± 9.47	3.53 ± 2.00	-98.5 ± 13.0
	11	1	-70.2	-70.2 ± 0.00	-30.5 ± 0.00	-29.4 ± 0.00	-27.1 ± 0.00	1.69 ± 0.00	-87.1 ± 0.00
	12	2	-69.5	-56.5 ± 18.5	-28.4 ± 0.54	-58.3 ± 21.9	-22.8 ± 0.32	5.30 ± 0.33	-109 ± 21.7
	13	1	-58.2	-58.2 ± 0.00	-28.1 ± 0.00	-33.1 ± 0.00	-50.0 ± 0.00	5.27 ± 0.00	-111 ± 0.00
	14	1	-56.1	-56.1 ± 0.00	-37.0 ± 0.00	-8.82 ± 0.00	-30.6 ± 0.00	2.03 ± 0.00	-76.4 ± 0.00
	15	1	-40.4	-40.4 ± 0.00	-33.8 ± 0.00	-40.6 ± 0.00	-29.8 ± 0.00	6.37 ± 0.00	-104 ± 0.00

[*] Average ± standard deviation (S.D.) for each docked structure within the cluster. Clusters with a S.D. of 0 had only one docked structure populating the cluster.

[a] The lowest HADDOCK score (H_{score}) found in the given cluster.

[b] The H_{score} was calculated for each structure and averaged.

[c] The total intermolecular energy ($E_{total} = E_{vdw}+E_{elect}+E_{desolv}+E_{dist}$) was calculated for each structure and averaged.

5.5 FIGURES

[LIGANDS]
DRG -1

[SOI]
; Residue numbers | Restraint Distance (r) | Weight (Integer)

180 | 20 | 1
184 | 20 | 1
188 | 20 | 1
137 | 20 | 1
176 | 20 | 1
875 | 20 | 1
133 | 20 | 1
883 | 20 | 1
927 | 20 | 1
872 | 20 | 1
179 | 20 | 1
136 | 20 | 1
132 | 20 | 1
879 | 20 | 1
876 | 20 | 1
938 | 20 | 1
144 | 20 | 1
886 | 20 | 1
177 | 20 | 1
942 | 20 | 1
343 | 20 | 1
346 | 20 | 1
868 | 20 | 1
60 | 20 | 1
864 | 20 | 1
977 | 20 | 1
981 | 20 | 1
861 | 20 | 1
195 | 20 | 1
336 | 20 | 1
64 | 20 | 1
979 | 20 | 1
174 | 20 | 1
350 | 20 | 1
916 | 20 | 1
169 | 20 | 1
170 | 20 | 1
894 | 20 | 1
168 | 20 | 1
173 | 20 | 1
172 | 20 | 1
930 | 20 | 1
934 | 20 | 1
949 | 20 | 1
945 | 20 | 1
946 | 20 | 1

941 | 20 | 1
114 | 20 | 1
129 | 20 | 1
165 | 20 | 1
166 | 20 | 1
890 | 20 | 1
191 | 20 | 1
347 | 20 | 1
952 | 20 | 1
68 | 20 | 1
332 | 20 | 1
339 | 20 | 1
728 | 20 | 1
306 | 20 | 1
71 | 20 | 1
990 | 20 | 1
295 | 20 | 1
986 | 20 | 1
724 | 20 | 1
292 | 20 | 1
774 | 20 | 1
287 | 20 | 1
822 | 20 | 1
221 | 20 | 1
810 | 20 | 1
773 | 20 | 1
778 | 20 | 1
781 | 20 | 1
281 | 20 | 1
816 | 20 | 1
710 | 20 | 1
721 | 20 | 1
302 | 20 | 1
303 | 20 | 1
762 | 20 | 1
819 | 20 | 1
239 | 20 | 1
291 | 20 | 1
827 | 20 | 1
769 | 20 | 1
298 | 20 | 1
232 | 20 | 1
288 | 20 | 1
225 | 20 | 1
785 | 20 | 1
243 | 20 | 1
992 | 20 | 1
812 | 20 | 1
294 | 20 | 1
989 | 20 | 1
815 | 20 | 1
996 | 20 | 1
993 | 20 | 1
988 | 20 | 1
236 | 20 | 1
985 | 20 | 1

```
982 | 20 | 1
335 | 20 | 1
974 | 20 | 1
975 | 20 | 1
978 | 20 | 1
300 | 20 | 1
67  | 20 | 1
72  | 20 | 1
75  | 20 | 1
113 | 20 | 1
121 | 20 | 1
326 | 20 | 1
327 | 20 | 1
328 | 20 | 1
329 | 20 | 1
333 | 20 | 1
725 | 20 | 1
729 | 20 | 1
732 | 20 | 1
946 | 20 | 1
971 | 20 | 1
953 | 20 | 1
110 | 20 | 1
340 | 20 | 1

[STD]
; Proton(s) | Weight

HAA HAE | 1 ; Protons 15,19
HAB HAD | 1 ; Protons 16,18
HAC | 1 ; Proton 17
HAL | 1 ; Proton 4
HAN | 1 ; Proton 7
HAR | 1 ; Proton 2
HAM | 1 ; Proton 6


[SAR]
; Ligand Atom Numbers | Weight

21-34,44-end| 1 ; aromaticity
42,43 | 1 ; hydrogen bond acceptors
5 | 1 ; tertiary amine

[end]
```

Figure 5.1 *Experimental input used to drive the molecular docking of eletriptan (ETT) onto P-glycoprotein (Pgp).* The comments are shown in black and the simplified language used to produce the CNS-compatible distance restraints are shown in **red**.

[LIGANDS]
DRG -1

[SOI]
; Residue numbers | Restraint Distance (r) | Weight (Integer)

180 | 20 | 1
184 | 20 | 1
188 | 20 | 1
137 | 20 | 1
176 | 20 | 1
875 | 20 | 1
133 | 20 | 1
883 | 20 | 1
927 | 20 | 1
872 | 20 | 1
179 | 20 | 1
136 | 20 | 1
132 | 20 | 1
879 | 20 | 1
876 | 20 | 1
938 | 20 | 1
144 | 20 | 1
886 | 20 | 1
177 | 20 | 1
942 | 20 | 1
343 | 20 | 1
346 | 20 | 1
868 | 20 | 1
60 | 20 | 1
864 | 20 | 1
977 | 20 | 1
981 | 20 | 1
861 | 20 | 1
195 | 20 | 1
336 | 20 | 1
64 | 20 | 1
979 | 20 | 1
174 | 20 | 1
350 | 20 | 1
916 | 20 | 1
169 | 20 | 1
170 | 20 | 1
894 | 20 | 1
168 | 20 | 1
173 | 20 | 1
172 | 20 | 1
930 | 20 | 1
934 | 20 | 1
949 | 20 | 1
945 | 20 | 1
946 | 20 | 1
941 | 20 | 1
114 | 20 | 1
129 | 20 | 1

165 | 20 | 1
166 | 20 | 1
890 | 20 | 1
191 | 20 | 1
347 | 20 | 1
952 | 20 | 1
68 | 20 | 1
332 | 20 | 1
339 | 20 | 1
728 | 20 | 1
306 | 20 | 1
71 | 20 | 1
990 | 20 | 1
295 | 20 | 1
986 | 20 | 1
724 | 20 | 1
292 | 20 | 1
774 | 20 | 1
287 | 20 | 1
822 | 20 | 1
221 | 20 | 1
810 | 20 | 1
773 | 20 | 1
778 | 20 | 1
781 | 20 | 1
281 | 20 | 1
816 | 20 | 1
710 | 20 | 1
721 | 20 | 1
302 | 20 | 1
303 | 20 | 1
762 | 20 | 1
819 | 20 | 1
239 | 20 | 1
291 | 20 | 1
827 | 20 | 1
769 | 20 | 1
298 | 20 | 1
232 | 20 | 1
288 | 20 | 1
225 | 20 | 1
785 | 20 | 1
243 | 20 | 1
992 | 20 | 1
812 | 20 | 1
294 | 20 | 1
989 | 20 | 1
815 | 20 | 1
996 | 20 | 1
993 | 20 | 1
988 | 20 | 1
236 | 20 | 1
985 | 20 | 1
982 | 20 | 1
335 | 20 | 1
974 | 20 | 1

```
975 | 20 | 1
978 | 20 | 1
300 | 20 | 1
67  | 20 | 1
72  | 20 | 1
75  | 20 | 1
113 | 20 | 1
121 | 20 | 1
326 | 20 | 1
327 | 20 | 1
328 | 20 | 1
329 | 20 | 1
333 | 20 | 1
725 | 20 | 1
729 | 20 | 1
732 | 20 | 1
946 | 20 | 1
971 | 20 | 1
953 | 20 | 1
110 | 20 | 1
340 | 20 | 1

[STD]
; Proton(s) | Weight

HAD | 1 ; Proton 4
HAF | 1 ; Proton 7
HAG | 1 ; Proton 2
HAJ | 1 ; Proton 6


[SAR]
; Ligand Atom Numbers | Weight

21-30 | 1 ; aromaticity
35,36 | 1 ; hydrogen bond acceptors
5 | 1 ; tertiary amine

[end]
```

Figure 5.2 *Experimental input used to drive the molecular docking of sumatriptan (STT) onto P-glycoprotein (Pgp).* The comments are shown in black and the simplified language used to produce the CNS-compatible distance restraints are shown in **red**.

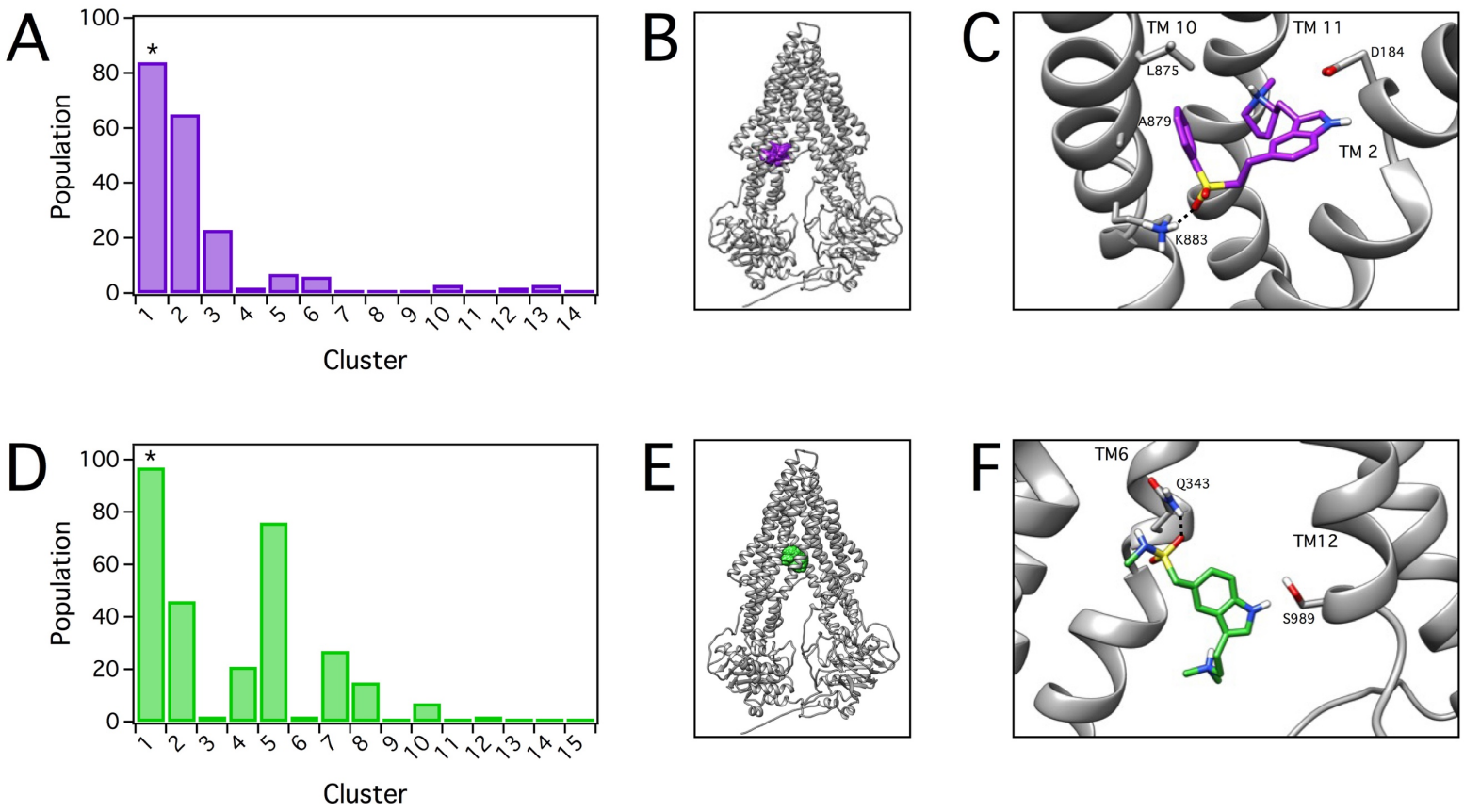

Figure 5.3 *The ETT and STT binding site of P-glycoprotein (Pgp) deduced by HADDOCK-Vina.* The molecular docking of **(A-C)** ETT and **(D-F)** of STT. The clustering distribution from the molecular docking simulations is shown **A,D** and the lowest H_{score} cluster is denoted with an asterisk. The lowest H_{score} cluster is shown on the structure of Pgp rendered as a gray cartoon in **B,E**. Within these clusters, the lowest H_{score} structure is shown in **C,F**. The transmembrane helices are shown in gray and labeled. The residues interacting with the ligand are labelled. Dashed lines denote hydrogen bond interactions

REFERENCES

1. Schinkel AH, Jonker JW. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. Adv Drug Deliv Rev 2003;55:3–29.
2. Loo TW, Bartlett MC, Clarke DM. Human P-glycoprotein is active when the two halves are clamped together in the closed conformation. Biochem Biophys Res Commun 2010;395:436–440.
3. Loo TW, Bartlett MC, Detty MR, Clarke DM. The ATPase activity of the P-glycoprotein drug pump is highly activated when the N-terminal and central regions of the nucleotide-binding domains are linked closely together. J Biol Chem 2012;287:26806–26816.
4. Juliano RL, Ling V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta BBA - Biomembr 1976;455(1):152–162.
5. Sharom FJ. The P-glycoprotein multidrug transporter. Essays Biochem 2011;50:161–178.
6. Callen DF, Baker E, Simmers RN, Seshadri R, Roninson IB. Localization of the human multiple drug resistance gene,MDR1, to 7q21.1. Hum Genet 1987;77(2):142–144.
7. Feng X, Xie H-G. Applying Pharmacogenomics in Therapeutics. : CRC Press; 2016.
8. Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM. P-glycoprotein: from genomics to mechanism. Oncogene 2003;22(47):7468–7485.
9. Loo TW, Clarke DM. Membrane topology of a cysteine-less mutant of human p-glycoprotein. J Biol Chem 1995;270:843–848.
10. Swartz DJ, Weber J, Urbatsch IL. P-glycoprotein is fully active after multiple tryptophan substitutions. Biochim Biophys Acta BBA-Biomembr 2013;1828:1159–1168.

11. Loo TW, Clarke DM. Locking intracellular helices 2 and 3 together inactivates human P-glycoprotein. J Biol Chem 2014;289:229–236.
12. Kwan P, Baum L, Wong V, Ng PW, Lui CH, Sin NC, Hui ACF, Yu E, Wong LKS. Association between ABCB1 C3435T polymorphism and drug-resistant epilepsy in Han Chinese. Epilepsy Behav 2007;11(1):112–117.
13. Bournissen FG, Moretti ME, Juurlink DN, Koren G, Walker M, Finkelstein Y. Polymorphism of the MDR1/ABCB1 C3435T drug-transporter and resistance to anticonvulsant drugs: A meta-analysis. Epilepsia 2009;50(4):898–903.
14. Benet LZ, Cummins CL. The drug efflux–metabolism alliance: biochemical aspects. Adv Drug Deliv Rev 2001;50:S3–S11.
15. Synold TW, Dussault I, Forman BM. The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. Nat Med 2001;7:584–590.
16. Slosky LM, Thompson BJ, Sanchez-Covarrubias L, Zhang Y, Laracuente M-L, Vanderah TW, Ronaldson PT, Davis TP. Acetaminophen Modulates P-Glycoprotein Functional Expression at the Blood-Brain Barrier by a Constitutive Androstane Receptor–Dependent Mechanism. Mol Pharmacol 2013;84(5):774–786.
17. Wu J, Lin N, Li F, Zhang G, He S, Zhu Y, Ou R, Li N, Liu S, Feng L, Liu L, Liu Z, Lu L. Induction of P-glycoprotein expression and activity by Aconitum alkaloids: Implication for clinical drug–drug interactions. Sci Rep 2016;6.
18. Evans RM, Mangelsdorf DJ. Nuclear Receptors, RXR, and the Big Bang. Cell 2014;157(1):255–266.
19. Mukherjee S, Mani S. Orphan Nuclear Receptors as Targets for Drug Development. Pharm Res 2010;27(8):1439–1468.

20. Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ. Nuclear Receptors and Lipid Physiology: Opening the X-Files. Science 2001;294(5548):1866–1870.
21. Sharom FJ. Multidrug resistance protein: P-glycoprotein. Drug Transp Mol Charact Role Drug Dispos 2007:223–262.
22. Sharom FJ. ABC multidrug transporters: structure, function and role in chemoresistance. 2008.
23. Takano M, Yumoto R, Murakami T. Expression and function of efflux drug transporters in the intestine. Pharmacol Ther 2006;109:137–161.
24. Chan LM, Lowes S, Hirst BH. The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability. Eur J Pharm Sci 2004;21:25–51.
25. Leslie EM, Deeley RG, Cole SP. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. Toxicol Appl Pharmacol 2005;204:216–237.
26. Zhang Y, Benet LZ. The gut as a barrier to drug absorption. Clin Pharmacokinet 2001;40:159–168.
27. Suzuki H, Sugiyama Y. Role of metabolic enzymes and efflux transporters in the absorption of drugs from the small intestine. Eur J Pharm Sci 2000;12:3–12.
28. Ledwitch KV, Roberts AG. Cardiovascular ion channel inhibitor drug-drug interactions with P-glycoprotein. AAPS J 2017;19(2):409–420.
29. Sylvain E, Rajaraman S, Megyesi J, Bello-Reuss EN. Expression of MDR1 (multidrug resistance) gene and its protein in normal human kidney. Nephron Basel 1997;77(3):284–289.
30. Schinkel AH, Wagenaar E, Deemter L van, Mol CA, Borst P. Absence of the mdr1a P-Glycoprotein in mice affects tissue distribution and pharmacokinetics of dexamethasone, digoxin, and cyclosporin A. J Clin Invest 1995;96(4):1698–1705.

31. Liu J, Liu Y, Powell DA, Waalkes MP, Klaassen CD. Multidrug-resistance mdr1a/1b double knockout mice are more sensitive than wild type mice to acute arsenic toxicity, with higher arsenic accumulation in tissues. Toxicology 2002;170(1):55–62.
32. Chen Y, Liu L. Modern methods for delivery of drugs across the blood–brain barrier. Adv Drug Deliv Rev 2012;64:640–665.
33. Abbott NJ. Anatomy and physiology of the blood–brain barriers. In: Drug Delivery to the Brain. Springer; 2014. p 3–21.
34. Tachikawa M, Uchida Y, Ohtsuki S, Terasaki T. Recent progress in blood–brain barrier and blood–csf barrier transport research: pharmaceutical relevance for drug delivery to the brain. In: Drug Delivery to the Brain. Springer; 2014. p 23–62.
35. Zlokovic BV. The blood-brain barrier in health and chronic neurodegenerative disorders. Neuron 2008;57:178–201.
36. Agarwal S, Uchida Y, Mittapalli RK, Sane R, Terasaki T, Elmquist WF. Quantitative proteomics of transporter expression in brain capillary endothelial cells isolated from P-glycoprotein (P-gp), breast cancer resistance protein (Bcrp), and P-gp/Bcrp knockout mice. Drug Metab Dispos 2012;40:1164–1169.
37. Schinkel AH, Smit JJ, Tellingen O van, Beijnen JH, Wagenaar E, Deemter L van, Mol CA, Valk MA van der, Robanus-Maandag EC, Riele HP te, et al. Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood-brain barrier and to increased sensitivity to drugs. Cell 1994;77(4):491–502.
38. Lankas GR, Cartwright ME, Umbenhauer D. P-glycoprotein deficiency in a subpopulation of CF-1 mice enhances avermectin-induced neurotoxicity. Toxicol Appl Pharmacol 1997;143(2):357–365.

39. Atkinson DE, Greenwood SL, Sibley CP, Glazier JD, Fairbairn LJ. Role of MDR1 and MRP1 in trophoblast cells, elucidated using retroviral gene transfer. Am J Physiol-Cell Physiol 2003;285(3):C584–C591.
40. Smit JW, Huisman MT, Tellingen O van, Wiltshire HR, Schinkel AH. Absence or pharmacological blocking of placental P-glycoprotein profoundly increases fetal drug exposure. J Clin Invest 1999;104(10):1441–1447.
41. Choo EF, Leake B, Wandel C, Imamura H, Wood AJ, Wilkinson GR, Kim RB. Pharmacological inhibition of P-glycoprotein transport enhances the distribution of HIV-1 protease inhibitors into brain and testes. Drug Metab Dispos Biol Fate Chem 2000;28(6):655–660.
42. Melaine N, Liénard M-O, Dorval I, Le Goascogne C, Lejeune H, Jégou B. Multidrug Resistance Genes and P-Glycoprotein in the Testis of the Rat, Mouse, Guinea Pig, and Human. Biol Reprod 2002;67(6):1699–1707.
43. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 2009;323(5922):1718–1722.
44. Ward AB, Szewczyk P, Grimard V, Lee C-W, Martinez L, Doshi R, Caya A, Villaluz M, Pardon E, Cregger C. Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. Proc Natl Acad Sci 2013;110:13386–13391.
45. Esser L, Zhou F, Pluchino KM, Shiloach J, Ma J, Tang W-K, Gutierrez C, Zhang A, Shukla S, Madigan JP. Structures of the Multidrug Transporter P-glycoprotein Reveal Asymmetric ATP Binding and the Mechanism of Polyspecificity. J Biol Chem 2016:jbc. M116. 755884.

46. Li J, Jaimes KF, Aller SG. Refined structures of mouse P-glycoprotein. Protein Sci 2014;23:34–46.
47. Eckford PD, Sharom FJ. ABC efflux pump-based resistance to chemotherapy drugs. Chem Rev 2009;109:2989–3011.
48. Ledwitch KV, Barnes RW, Roberts AG. Unravelling the complex drug-drug interactions of the cardiovascular drugs, verapamil and digoxin, with P-glycoprotein. Biosci Rep 2016;36(2).
49. Cirrito JR, Deane R, Fagan AM, Spinner ML, Parsadanian M, Finn MB, Jiang H, Prior JL, Sagare A, Bales KR. P-glycoprotein deficiency at the blood-brain barrier increases amyloid-β deposition in an Alzheimer disease mouse model. J Clin Invest 2005;115:3285.
50. Kim Y, Chen J. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. Science 2018;359(6378):915–919.
51. Fyfe PK, McAuley KE, Roszak AW, Isaacs NW, Cogdell RJ, Jones MR. Probing the interface between membrane proteins and membrane lipids by X-ray crystallography. Trends Biochem Sci 2001;26:106–112.
52. Sharom FJ. Complex interplay between the P-glycoprotein multidrug efflux pump and the membrane: Its role in modulating protein function. Front Oncol 2014;4(41):1–19.
53. Kimura Y, Kioka N, Kato H, Matsuo M, Ueda K. Modulation of drug-stimulated ATPase activity of human MDR1/P-glycoprotein by cholesterol. Biochem J 2007;401(2):597–605.
54. Kodan A, Shibata H, Matsumoto T, Terakado K, Sakiyama K, Matsuo M, Ueda K, Kato H. Improved expression and purification of human multidrug resistance protein MDR1 from baculovirus-infected insect cells. Protein Expr Purif 2009;66(1):7–14.
55. Kerr KM, Sauna ZE, Ambudkar SV. Correlation between steady-state ATP hydrolysis and vanadate-induced ADP trapping in Human P-glycoprotein. Evidence for ADP release as the

rate-limiting step in the catalytic cycle and its modulation by substrates. J Biol Chem 2001;276(12):8657–8664.

56. Sharom FJ. Shedding light on drug transport: structure and function of the P-glycoprotein multidrug transporter (ABCB1). Biochem Cell Biol 2006;84:979–992.
57. Siarheyeva A, Liu R, Sharom FJ. Characterization of an asymmetric occluded state of P-glycoprotein with two bound nucleotides: implications for catalysis. J Biol Chem 2010;285:7575–7586.
58. Wilt LA, Nguyen D, Roberts AG. Insights into the molecular mechanism of triptan transport by P-glycoprotein. J Pharm Sci 2017;106:1670–1679.
59. Loo TW, Clarke DM. Tariquidar inhibits P-glycoprotein drug efflux but activates ATPase activity by blocking transition to an open conformation. Biochem Pharmacol 2014;92(4):558–566.
60. Clay AT, Sharom FJ. Lipid Bilayer Properties Control Membrane Partitioning, Binding, and Transport of P-Glycoprotein Substrates. Biochemistry (Mosc) 2013;52(2):343–354.
61. Loo TW, Clarke DM. Mapping the Binding Site of the Inhibitor Tariquidar That Stabilizes the First Transmembrane Domain of P-glycoprotein. J Biol Chem 2015;290:29389–29401.
62. Pajeva IK, Hanl M, Wiese M. Protein Contacts and Ligand Binding in the Inward-Facing Model of Human P-Glycoprotein. ChemMedChem 2013;8(5):748–762.
63. Loo TW, Clarke DM. The transmission interfaces contribute asymmetrically to the assembly and activity of human P-glycoprotein. J Biol Chem 2015;290:16954–16963.
64. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published

in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. Adv Drug Deliv Rev 2001;46(1):3–26.

65. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 2004;1:337–341.
66. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov 2007;6(11):881–890.
67. Seelig A. A general pattern for substrate recognition by P-glycoprotein. Eur J Biochem 1998;251:252–261.
68. Mahar Doan KM, Humphreys JE, Webster LO, Wring SA, Shampine LJ, Serabjit-Singh CJ, Adkison KK, Polli JW. Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. J Pharmacol Exp Ther 2002;303:1029–1037.
69. Evans DC, O'Connor D, Lake BG, Evers R, Allen C, Hargreaves R. Eletriptan metabolism by human hepatic CYP450 enzymes and transport by human P-glycoprotein. Drug Metab Dispos 2003;31:861–869.
70. Fromm MF. P-glycoprotein: a defense mechanism limiting oral bioavailability and CNS accumulation of drugs. Int J Clin Pharmacol Ther 2000;38(2):69–74.
71. Pardridge WM. CNS drug design based on principles of blood-brain barrier transport. J Neurochem 1998;70(5):1781–1792.
72. Doran A, Obach RS, Smith BJ, Hosea NA, Becker S, Callegari E, Chen C, Chen X, Choo E, Cianfrogna J, Cox LM, Gibbs JP, Gibbs MA, Hatch H, Hop CE, Kasman IN, Laperle J, Liu J, Liu X, Logman M, Maclin D, Nedza FM, Nelson F, Olson E, Rahematpura S, Raunig D, Rogers S, Schmidt K, Spracklin DK, Szewc M, Troutman M, Tseng E, Tu M, Van Deusen

JW, Venkatakrishnan K, Walens G, Wang EQ, Wong D, Yasgar AS, Zhang C. The impact of P-glycoprotein on the disposition of drugs targeted for indications of the central nervous system: evaluation using the MDR1A/1B knockout mouse model. Drug Metab Dispos Biol Fate Chem 2005;33:165–174.

73. Kikuchi R, Morais SM de, Kalvass JC. In vitro P-glycoprotein efflux ratio can predict the in vivo brain penetration regardless of biopharmaceutics drug disposition classification system class. Drug Metab Dispos Biol Fate Chem 2013;41:2012–2017.
74. Fenner KS, Troutman MD, Kempshall S, Cook JA, Ware JA, Smith DA, Lee CA. Drug-drug interactions mediated through P-glycoprotein: clinical relevance and in vitro-in vivo correlation using digoxin as a probe drug. Clin Pharmacol Ther 2009;85:173–181.
75. Finch A, Pillans P. P-glycoprotein and its role in drug-drug interactions. Aust Prescr 2014;37(4):137–139.
76. Callaghan R, Luk F, Bebawy M. Inhibition of the Multidrug Resistance P-Glycoprotein: Time for a Change of Strategy? Drug Metab Dispos 2014;42(4):623–631.
77. Wu C-P, Calcagno AM, Ambudkar SV. Reversal of ABC drug transporter-mediated multidrug resistance in cancer cells: Evaluation of current strategies. Curr Mol Pharmacol 2008;1(2):93–105.
78. Ramakrishnan P. The role of p-glycoprotein in the blood-brain barrier. Einstein Q J Biol Med 2003;19:160–165.
79. Jose M, Thomas SV. Role of multidrug transporters in neurotherapeutics. Ann Indian Acad Neurol 2009;12(2):89–98.
80. Kaitin KI, Milne CP. A dearth of new meds. Sci Am 2011;305(2):16.

81. Cutler NR, Sramek JJ, Murphy MF, Riordan H, Bieck P, Carta A. The Impending Crisis in CNS Drug Development. In: Critical Pathways to Success in CNS Drug Development. Wiley-Blackwell; 2010. p 1–13.
82. Binkhathlan Z, Lavasanifar A. P-glycoprotein Inhibition as a Therapeutic Approach for Overcoming Multidrug Resistance in Cancer: Current Status and Future Perspectives. Curr Cancer Drug Targets 2013;13(3):326–346.
83. Bendayan R, Lee G, Bendayan M. Functional expression and localization of P-glycoprotein at the blood brain barrier. Microsc Res Tech 2002;57(5):365–380.
84. Lee G, Bendayan R. Functional expression and localization of P-glycoprotein in the central nervous system: relevance to the pathogenesis and treatment of neurological disorders. Pharm Res 2004;21(8):1313–1330.
85. Martin C, Berridge G, Higgins CF, Mistry P, Charlton P, Callaghan R. Communication between Multiple Drug Binding Sites on P-glycoprotein. Mol Pharmacol 2000;58(3):624–632.
86. Safa AR. Identification and characterization of the binding sites of P-glycoprotein for multidrug resistance-related drugs and modulators. Curr Med Chem Anticancer Agents 2004;4(1):1–17.
87. Dawson RJP, Hollenstein K, Locher KP. Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. Mol Microbiol 2007;65(2):250–257.
88. Ward A, Reyes CL, Yu J, Roth CB, Chang G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci 2007;104(48):19005–19010.
89. Barnes NM, Sharp T. A review of central 5-HT receptors and their function. Neuropharmacology 1999;38:1083–1152.

90. Napier C, Stewart M, Melrose H, Hopkins B, McHarg A, Wallis R. Characterisation of the 5-HT receptor binding profile of eletriptan and kinetics of [3H] eletriptan binding at human 5-HT 1B and 5-HT 1D receptors. Eur J Pharmacol 1999;368:259–268.
91. Tepper SJ, Rapoport AM, Sheftell FD. Mechanisms of action of the 5-HT1B/1D receptor agonists. Arch Neurol 2002;59:1084–1088.
92. Kalvass JC, Maurer TS, Pollack GM. Use of plasma and brain unbound fractions to assess the extent of brain distribution of 34 drugs: comparison of unbound concentration ratios to in vivo p-glycoprotein efflux ratios. Drug Metab Dispos Biol Fate Chem 2007;35:660–666.
93. The Discovery and Development of the Triptans, a Major Therapeutic Breakthrough - Humphrey - 2008 - Headache: The Journal of Head and Face Pain - Wiley Online Library. .
94. Ledwitch KV, Gibbs ME, Barnes RW, Roberts AG. Cooperativity between verapamil and ATP bound to the efflux transporter P-glycoprotein. Biochem Pharmacol 2016.
95. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA. PubChem substance and compound databases. Nucleic Acids Res 2015:gkv951.
96. Lerner-Marmarosh N, Gimi K, Urbatsch IL, Gros P, Senior AE. Large scale purification of detergent-soluble P-glycoprotein from Pichia pastoris cells and characterization of nucleotide binding properties of wild-type, Walker A, and Walker B mutant proteins. J Biol Chem 1999;274(49):34711–34718.
97. Bai J, Swartz DJ, Protasevich II, Brouillette CG, Harrell PM, Hildebrandt E, Gasser B, Mattanovich D, Ward A, Chang G, Urbatsch IL. A Gene Optimization Strategy that Enhances Production of Fully Functional P-Glycoprotein in Pichia pastoris. PLOS ONE 2011;6(8):e22577.

98. Chifflet S, Torriglia A, Chiesa R, Tolosa S. A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. Anal Biochem 1988;168(1):1–4.

99. Segel IH. Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems. New York: John Wiley & Sons, Inc.; 1975.

100. Cook PF, Cleland WW. Enzyme Kinetics and Mechanism. London ; New York: Garland Science; 2007.

101. Martin RB. Disadvantages of Double Reciprocal Plots. J Chem Educ 1997;74:1238.

102. Leatherbarrow RJ. Using linear and non-linear regression to fit biochemical data. Trends Biochem Sci 1990;15(12):455–458.

103. Ranaldi F, Vanni P, Giachetti E. What students must know about the determination of enzyme kinetic parameters. Biochem Educ 1999;27(2):87–91.

104. Liu R, Siemiarczuk A, Sharom FJ. Intrinsic fluorescence of the P-glycoprotein multidrug transporter: sensitivity of tryptophan residues to binding of drugs and nucleotides. Biochemistry (Mosc) 2000;39:14927–14938.

105. Qu Q, Russell PL, Sharom FJ. Stoichiometry and affinity of nucleotide binding to P-glycoprotein during the catalytic cycle. Biochemistry (Mosc) 2003;42(4):1170–1177.

106. Lakowicz JR. Principles of fluorescence spectroscopy. New York: Kluwer Academic/Plenum; 1999.

107. Russell PL, Sharom FJ. Conformational and functional characterization of trapped complexes of the P-glycoprotein multidrug transporter. Biochem J 2006;399:315–323.

108. Sonveaux N, Vigano C, Shapiro AB, Ling V, Ruysschaert JM. Ligand-mediated tertiary structure changes of reconstituted P-glycoprotein. A tryptophan fluorescence quenching analysis. J Biol Chem 1999;274(25):17649–17654.

109. Eckford PD, Sharom FJ. Functional characterization of Escherichia coli MsbA: interaction with nucleotides and substrates. J Biol Chem 2008;283:12840–12850.

110. Mayer M, Meyer B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc 2001;123:6108–6117.

111. Venkitakrishnan RP, Benard O, Max M, Markley JL, Assadi-Porter FM. Use of NMR saturation transfer difference spectroscopy to study ligand binding to membrane proteins. Membr Protein Struct Dyn Methods Protoc 2012:47–63.

112. Viegas A, Manso J, Nobrega FL, Cabrita EJ. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. J Chem Educ 2011;88(7):990–994.

113. Claasen B, Axmann M, Meinecke R, Meyer B. Direct observation of ligand binding to membrane proteins in living cells by a saturation transfer double difference (STDD) NMR spectroscopy method shows a significantly higher affinity of integrin αIIbβ3 in native platelets than in liposomes. J Am Chem Soc 2005;127:916–919.

114. Haselhorst T, Munster-Kuhnel AK, Oschlies M, Tiralongo J, Gerardy-Schahn R, Itzstein M von. Direct detection of ligand binding to Sepharose-immobilised protein using saturation transfer double difference (STDD) NMR spectroscopy. Biochem Biophys Res Commun 2007;359:866–870.

115. Borgnia MJ, Eytan GD, Assaraf YG. Competition of hydrophobic peptides, cytotoxic drugs, and chemosensitizers on a common P-glycoprotein pharmacophore as revealed by its ATPase activity. J Biol Chem 1996;271(6):3163–3171.

116. Litman T, Zeuthen T, Skovsgaard T, Stein WD. Competitive, non-competitive and cooperative interactions between substrates of P-glycoprotein as measured by its ATPase activity. Biochim Biophys Acta BBA - Mol Basis Dis 1997;1361(2):169–176.

117. Richter O von, Glavinas H, Krajcsi P, Liehner S, Siewert B, Zech K. A novel screening strategy to identify ABCB1 substrates and inhibitors. Naunyn Schmiedebergs Arch Pharmacol 2009;379:11–26.

118. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006;34:D668–D672.

119. Döppenschmitt S, Spahn-Langguth H, Regårdh CG, Langguth P. Radioligand-binding assay employing P-glycoprotein-overexpressing cells: testing drug affinities to the secretory intestinal multidrug transporter. Pharm Res 1998;15:1001–1006.

120. Busche R, Tummler B, Cano-Gauci DF, Riordan JR. Equilibrium, kinetic and photoaffinity labeling studies of daunomycin binding to P-glycoprotein-containing membranes of multidrug-resistant Chinese hamster ovary cells. Eur J Biochem 1989;183(1):189–197.

121. Romsicki Y, Sharom FJ. The membrane lipid environment modulates drug interactions with the P-glycoprotein multidrug transporter. Biochemistry (Mosc) 1999;38:6887–6896.

122. Liu R, Sharom FJ. Site-directed fluorescence labeling of P-glycoprotein on cysteine residues in the nucleotide binding domains. Biochemistry (Mosc) 1996;35:11865–11873.

123. Seelig A, Landwojtowicz E. Structure-activity relationship of P-glycoprotein substrates and modifiers. Eur J Pharm Sci Off J Eur Fed Pharm Sci 2000;12(1):31–40.

124. Angulo J, Enríquez-Navas PM, Nieto PM. Ligand–Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. Chem – Eur J 2010;16(26):7803–7812.

125. Scian M, Acchione M, Li M, Atkins WM. Reaction Dynamics of ATP Hydrolysis Catalyzed by P-Glycoprotein. Biochemistry (Mosc) 2014;53(6):991–1000.

126. Loo TW, Bartlett MC, Clarke DM. Drug binding in human P-glycoprotein causes conformational changes in both nucleotide-binding domains. J Biol Chem 2003;278:1575–1578.

127. Verhalen B, Ernst S, Börsch M, Wilkens S. Dynamic Ligand-induced Conformational Rearrangements in P-glycoprotein as Probed by Fluorescence Resonance Energy Transfer Spectroscopy. J Biol Chem 2012;287(2):1112–1127.

128. O'Mara ML, Mark AE. Structural Characterization of Two Metastable ATP-Bound States of P-Glycoprotein. PLOS ONE 2014;9(3):e91916.

129. Frank GA, Shukla S, Rao P, Borgnia MJ, Bartesaghi A, Merk A, Mobin A, Esser L, Earl LA, Gottesman MM, Xia D, Ambudkar SV, Subramaniam S. Cryo-EM analysis of the conformational landscape of human P-glycoprotein (ABCB1) during its catalytic cycle. Mol Pharmacol 2016;90:35–41.

130. Beaudet L, Gros P. Functional dissection of P-glycoprotein nucleotide-binding domains in chimeric and mutant proteins. Modulation of drug resistance profiles. J Biol Chem 1995;270(29):17159–17170.

131. Hrycyna CA, Ramachandra M, Germann UA, Cheng PW, Pastan I, Gottesman MM. Both ATP sites of human P-glycoprotein are essential but not symmetric. Biochemistry (Mosc) 1999;38(42):13887–13899.

132. Lawson J, O'Mara ML, Kerr ID. Structure-based interpretation of the mutagenesis database for the nucleotide binding domains of P-glycoprotein. Biochim Biophys Acta BBA - Bioenerg 2008;1778:376–391.

133. Urbatsch IL, Sankaran B, Bhagat S, Senior AE. Both P-glycoprotein Nucleotide-binding Sites Are Catalytically Active. J Biol Chem 1995;270:26956–26961.

134. Nicklisch SC, Rees SD, McGrath AP, Gökirmak T, Bonito LT, Vermeer LM, Cregger C, Loewen G, Sandin S, Chang G. Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and cocrystal structure. Sci Adv 2016;2:e1600001.

135. Ferreira LG, Santos RN dos, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules 2015;20:13384–13421.

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

137. Bohacek RS, McMartin C, Guida WC. The art and practice of structure-based drug design: A molecular modeling perspective. Med Res Rev 1996;16:3–50.

138. Scior T, Bender A, Tresadern G, Medina-Franco JL, Martínez-Mayorga K, Langer T, Cuanalo-Contreras K, Agrafiotis DK. Recognizing pitfalls in virtual screening: A critical review. J Chem Inf Model 2012;52(4):867–881.

139. Manju K, Gunjan P, Anju G. Review on introduction to molecular docking software technique in medicinal chemistry. Int J Drug Res Technol 2017;2:8.

140. Morris GM, Lim-Wilby M. Molecular docking. Mol Model Proteins 2008:365–382.

141. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev 2017;9(2):91–102.

142. Liu J, Wang R. Classification of current scoring functions. J Chem Inf Model 2015;55:475–482.

143. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31(2):455–461.

144. Castro-Alvarez A, Costa AM, Vilarrasa J. The performance of several docking programs at reproducing protein-macrolide-like crystal structures. Molecules 2017;22.

145. Durrant JD, Friedman AJ, Rogers KE, McCammon JA. Comparing neural-network scoring functions and the state of the art: applications to common library screening. J Chem Inf Model 2013;53:1726–1735.

146. Azam SS, Abbasi SW. Molecular docking studies for the identification of novel melatoninergic inhibitors for acetylserotonin-O-methyltransferase using different docking routines. Theor Biol Med Model 2013;10:63.

147. Nivedha AK, Thieker DF, Makeneni S, Hu H, Woods RJ. Vina-carb: improving glycosidic angles during carbohydrate docking. J Chem Theory Comput 2016;12:892–901.

148. Quiroga R, Villarreal MA. Vinardo: a scoring function based on autodock vina improves scoring, docking, and virtual screening. PLOS ONE 2016;11(5):e0155183.

149. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–1749.

150. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–1737.

151. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.

152. Shah DM, Ab E, Diercks T, Hass MA, Nuland NA van, Siegal G. Rapid protein–ligand costructures from sparse NOE data. J Med Chem 2012;55:10786–10790.

153. Proudfoot A, Bussiere DE, Lingel A. High-confidence protein–ligand complex modeling by nmr-guided docking enables early hit optimization. J Am Chem Soc 2017;139:17824–17833.

154. Sandal M, Duy TP, Cona M, Zung H, Carloni P, Musiani F, Giorgetti A. GOMoDo: a GPCRs online modeling and docking webserver. PloS One 2013;8:e74092.

155. Qin H, Shi J, Noberini R, Pasquale EB, Song J. Crystal structure and NMR binding reveal that two small molecule antagonists target the high affinity ephrin-binding channel of the EphA4 receptor. J Biol Chem 2008;283:29473–29484.

156. Stark JL, Powers R. Application of NMR and molecular docking in structure-based drug discovery. Top Curr Chem 2012;326:1–34.

157. Kurkcuoglu Z, Koukos PI, Citro N, Trellet ME, Rodrigues J, Moreira IS, Roel-Touris J, Melquiond AS, Geng C, Schaarschmidt J. Performance of HADDOCK and a simple contact-based protein–ligand binding affinity predictor in the D3R Grand Challenge 2. J Comput Aided Mol Des 2018;32:175–185.

158. Zundert GCP van, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, Dijk M van, Vries SJ de, Bonvin AMJJ. The HADDOCK2.2 web server:

user-friendly integrative modeling of biomolecular complexes. J Mol Biol 2016;428(4):720–725.

159. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton. J Mol Biol 1999;285(4):1735–1747.

160. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 2010;66:12–21.

161. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem B 2001;105:24a-24a.

162. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins 2005;61:704–721.

163. Davies MN, Toseland CP, Moss DS, Flower DR. Benchmarking pK(a) prediction. BMC Biochem 2006;7:18.

164. Maciejewski MW, Schuyler AD, Gryk MR, Moraru, Romero PR, Ulrich EL, Eghbalnia HR, Livny M, Delaglio F, Hoch JC. Nmrbox: a resource for biomolecular nmr computation. Biophys J 2017;112:1529–1534.

165. Linge JP, Nilges M. Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. J Biomol NMR 1999;13:51–59.

166. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. Propka3: consistent treatment of internal and surface residues in empirical pka predictions. J Chem Theory Comput 2011;7(2):525–537.

167. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pka values. J Chem Theory Comput 2011;7(7):2284–2295.

168. Schuttelkopf AW, Aalten DM van. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr Biol Crystallogr 2004;60:1355–1363.

169. Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 1996;38:305–320.

170. Voss NR, Gerstein M. 3V: cavity, channel and cleft volume calculator and extractor. Nucleic Acids Res 2010;38:W555-62.

171. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.

172. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–2791.

173. Roberts AG, Katayama J, Kaspera R, Ledwitch KV, Le Trong I, Stenkamp RE, Thompson JA, Totah RA. The role of cytochrome P450 BM3 phenylalanine-87 and threonine-268 in binding organic hydroperoxides. Biochim Biophys Acta BBA-Gen Subj 2016;1860:669–677.

174. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminformatics 2012;4:1.

175. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open babel: an open chemical toolbox. J Cheminform 2011;3:33.

176. Clore GM, Schwieters CD. Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H/15N chemical shift mapping and backbone 15N-1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. J Am Chem Soc 2003;125:2902–2912.

177. Loewenthal R, Sancho J, Reinikainen T, Fersht AR. Long-range surface charge-charge interactions in proteins. Comparison of experimental results with calculations from a theoretical method. J Mol Biol 1993;232:574–583.

178. Jayalakshmi V, Krishna NR. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. J Magn Reson 2002;155:106–118.

179. Meylan WM, Howard PH. A review of quantitative structure-activity relationship methods for the prediction of atmospheric oxidation of organic chemicals. Env Toxicol Chem 2003;22:1724–1732.

180. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins Struct Funct Bioinforma 2003;50:496–506.

181. Dijk M van, Dijk AD van, Hsu V, Boelens R, Bonvin AM. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. Nucleic Acids Res 2006;34:3317–3325.

182. De Vries SJ, Dijk M van, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 2010;5:883–897.

183. Daura X, Gademann K, Jaun B, Seebach D, Gunsteren WF van, Mark AE. Peptide folding: when simulation meets experiment. Angew Chem-Int Ed 1999;38:236–240.

184. Hazan C, Boudsocq F, Gervais V, Saurel O, Ciais M, Cazaux C, Czaplicki J, Milon A. Structural insights on the pamoic acid and the 8 kDa domain of DNA polymerase beta complex: towards the design of higher-affinity inhibitors. BMC Struct Biol 2008;8:22.

185. Cowan-Jacob SW, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, Rheinberger P, Centeleghe M, Fabbro D, Manley PW. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. Acta Crystallogr D Biol Crystallogr 2007;63(Pt 1):80–93.

186. Adachi M, Ohhara T, Kurihara K, Tamada T, Honjo E, Okazaki N, Arai S, Shoyama Y, Kimura K, Matsumura H, Sugiyama S, Adachi H, Takano K, Mori Y, Hidaka K, Kimura T, Hayashi Y, Kiso Y, Kuroki R. Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. Proc Natl Acad Sci U A 2009;106:4641–4646.

187. Berry MB, Meador B, Bilderback T, Liang P, Glaser M, Phillips GN. The closed conformation of a highly flexible protein: the structure of E. coli adenylate kinase with bound AMP and AMPPNP. Proteins Struct Funct Bioinforma 1994;19:183–198.

188. Azam M, Latek RR, Daley GQ. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. Cell 2003;112:831–843.

189. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Methods Enzym 2003;374:461–491.

190. Roberts AG, Yang J, Halpert JR, Nelson SD, Thummel KT, Atkins WM. The structural basis for homotropic and heterotropic cooperativity of midazolam metabolism by human cytochrome P450 3A4. Biochemistry (Mosc) 2011;50:10804–10818.

191. Hantschel O. Structure, regulation, signaling, and targeting of abl kinases in cancer. Genes Cancer 2012;3:436–446.

192. Cui J, Fu R, Zhou L-H, Chen S-P, Li G-W, Qian S-X, Liu S. BCR-ABL tyrosine kinase inhibitor pharmacophore model derived from a series of phenylaminopyrimidine-based (PAP) derivatives. Bioorg Med Chem Lett 2013;23(8):2442–2450.

193. Mian AA, Schüll M, Zhao Z, Oancea C, Hundertmark A, Beissert T, Ottmann OG, Ruthardt M. The gatekeeper mutation T315I confers resistance against small molecules by increasing or restoring the ABL-kinase activity accompanied by aberrant transphosphorylation of endogenous BCR, even in loss-of-function mutants of BCR/ABL. Leukemia 2009;23:1614–1621.

194. Bose P, Park H, Al-Khafaji J, Grant S. Strategies to circumvent the T315I gatekeeper mutation in the Bcr-Abl tyrosine kinase. Leuk Res Rep 2013;2:18–20.

195. Skora L, Mestan J, Fabbro D, Jahnke W, Grzesiek S. NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. Proc Natl Acad Sci 2013;110:E4437–E4445.

196. Weber IT, Agniswamy J. HIV-1 protease: structural perspectives on drug resistance. Viruses 2009;1:1110–1136.

197. Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. HIV AIDS Auckl 2015;7:95–104.

198. Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PY, Jadhav PK, et al. Flexibility and function in HIV-1 protease. Nat Struct Biol 1995;2:274–280.

199. Wang YX, Freedberg DI, Yamazaki T, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y, Torchia DA. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. Biochemistry (Mosc) 1996;35:9945–9950.

200. Dzeja P, Terzic A. Adenylate kinase and amp signaling networks: metabolic monitoring, signal communication and body energy sensing. Int J Mol Sci 2009;10:1729–1772.

201. Aden J, Wolf-Watz M. NMR identification of transient complexes critical to adenylate kinase catalysis. J Am Chem Soc 2007;129:14003–14012.

202. Tsai MD, Yan H. Mechanism of adenylate kinase: site-directed mutagenesis versus X-ray and NMR. Biochemistry (Mosc) 1991;30:6806–6818.

203. Davis AM, St-Gallay SA, Kleywegt GJ. Limitations and lessons in the use of X-ray structural information in drug design. Drug Discov Today 2008;13:831–841.

204. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. Angew Chem Int Ed 2003;42:2718–2736.

205. Chilingaryan Z, Yin Z, Oakley AJ. Fragment-based screening by protein crystallography: successes and pitfalls. Int J Mol Sci 2012;13:12857–12879.

206. Pearce NM, Bradley AR, Krojer T, Marsden BD, Deane CM, Delft F von. Partial-occupancy binders identified by the pan-dataset density analysis method offer new chemical opportunities and reveal cryptic binding sites. Struct Dyn 2017;4:032104.

207. Cala O, Guilliere F, Krimm I. NMR-based analysis of protein-ligand interactions. Anal Bioanal Chem 2014;406:943–956.

208. Post CB. Exchange-transferred NOE spectroscopy and bound ligand structure determination. Curr Opin Struct Biol 2003;13:581–588.

209. Roberts AG, Sjogren SE, Fomina N, Vu KT, Almutairi A, Halpert JR. NMR-derived models of amidopyrine and its metabolites in complexes with rabbit cytochrome P450 2B4 reveal a structural mechanism of sequential N-dealkylation. Biochemistry (Mosc) 2011;50:2123–2134.

210. Clore GM, Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 2009;109:4108–4139.

211. Dwek RA. Nuclear magnetic resonance (N.M.R.) in biochemistry : applications to enzyme systems. Oxford Eng.: Clarendon Press; 1973.

212. Pintacuda G, John M, Su XC, Otting G. NMR structure determination of protein-ligand complexes by lanthanide labeling. Acc Chem Res 2007;40:206–212.

213. Wilt LA, Roberts AG. HADDOCK-Vina: a hybrid approach for experimentally-driven receptor-small molecule docking. Proteins Struct Funct Bioinforma 2018; Submitted.

214. Sekula B, Zielinski K, Bujacz A. Crystallographic studies of the complexes of bovine and equine serum albumin with 3, 5-diiodosalicylic acid. Int J Biol Macromol 2013;60:316–324.

215. Su YY, Jirgensons B. Optical activity studies of drug--protein complexes. The interaction of acetylsalicylic acid with human serum albumin and myeloma immunoglobulin. Biochem Pharmacol 1978;27:1043–1047.

216. Bojko B, Sułkowska A, Maciążek M, Rownicka J, Njau F, Sułkowski WW. Changes of serum albumin affinity for aspirin induced by fatty acid. Int J Biol Macromol 2008;42:314–323.

217. Sulkowska A, Bojko B, Rownicka J, Sulkowski WW. Competition of cytarabine and aspirin in binding to serum albumin in multidrug therapy. Biopolymers 2006;81:464–472.

218. Ni Y, Zhu R, Kokot S. Competitive binding of small molecules with biopolymers: a fluorescence spectroscopy and chemometrics study of the interaction of aspirin and ibuprofen with BSA. Analyst 2011;136:4794–4801.

219. Fielding L, Rutherford S, Fletcher D. Determination of protein-ligand binding affinity by NMR: observations from serum albumin model systems. Magn Reson Chem 2005;43:463–470.

220. Simons SS. Selective covalent labeling of cysteines in bovine serum albumin and in hepatoma tissue culture cell glucocorticoid receptors by dexamethasone 21-mesylate. J Biol Chem 1987;262:9669–9675.

221. Liu D, Chu X, Lagi M, Zhang Y, Fratini E, Baglioni P, Alatas A, Said A, Alp E, Chen S-H. Studies of phononlike low-energy excitations of protein molecules by inelastic X-ray scattering. Phys Rev Lett 2008;101:135501.

222. Togashi DM, Ryder AG. Time-resolved fluorescence studies on bovine serum albumin denaturation process. J Fluoresc 2006;16:153–160.

223. Zhang F, Skoda MW, Jacobs RM, Martin RA, Martin CM, Schreiber F. Protein interactions studied by SAXS: effect of ionic strength and protein concentration for BSA in aqueous solutions. J Phys Chem B 2007;111:251–259.

224. Ebright YW, Chen Y, Pendergrast PS, Ebright RH. Incorporation of an EDTA-metal complex at a rationally selected site within a protein: application to EDTA-iron DNA affinity cleaving with catabolite gene activator protein (CAP) and Cro. Biochemistry (Mosc) 1992;31:10664–10670.

225. Majorek KA, Porebski PJ, Dayal A, Zimmerman MD, Jablonska K, Stewart AJ, Chruszcz M, Minor W. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. Mol Immunol 2012;52:174–182.

226. Liu ML, Mao XA, Ye CH, Huang H, Nicholson JK, Lindon JC. Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. J Magn Reson 1998;132:125–129.

227. Piotto M, Saudek V, Sklenář V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR 1992;2:661–665.

228. Kowalewski J, Mäler L. Nuclear spin relaxation in liquids : theory, experiments, and applications. New York: Taylor & Francis; 2006.

229. Scherf T, Anglister J. A T1 rho-filtered two-dimensional transferred NOE spectrum for studying antibody interactions with peptide antigens. Biophys J 1993;64:754.

230. Gay SC, Roberts AG, Maekawa K, Talakad JC, Hong W-X, Zhang Q, Stout CD, Halpert JR. Structures of cytochrome P450 2B4 complexed with the antiplatelet drugs ticlopidine and clopidogrel. Biochemistry (Mosc) 2010;49(40):8709–8720.

231. Iwahara J, Tang C, Clore GM. Practical aspects of 1 H transverse paramagnetic relaxation enhancement measurements on macromolecules. J Magn Reson 2007;184:185–195.

232. Iwahara J, Schwieters CD, Clore GM. Ensemble approach for nmr structure refinement against 1h paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J Am Chem Soc 2004;126(18):5879–5896.

233. Solomon I. Relaxation processes in a system of two spins. Phys Rev 1955;99:559.

234. Cameron MD, Wen B, Roberts AG, Atkins WM, Patricia Campbell A, Nelson SD. Cooperative binding of acetaminophen and caffeine within the P450 3A4 active site. Chem Res Toxicol 2007;20:1434–1441.

235. Pintacuda G, Moshref A, Leonchiks A, Sharipo A, Otting G. Site-specific labelling with a metal chelator for protein-structure refinement. J Biomol NMR 2004;29:351–361.

236. Oostenbrink C, Soares TA, Van der Vegt NF, Van Gunsteren WF. Validation of the 53A6 GROMOS force field. Eur Biophys J 2005;34:273–284.

237. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–1718.

238. Abdallah HH, Gadzhiev OB, Adnan R. Comparative study of theoretical partial charges of Zn and Mn-Schiff base complexes. In: ; 2009; . MDPI. p g016: 1-g016: 10.

239. Li P, Roberts BP, Chakravorty DK, Merz Jr KM. Rational design of particle mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit solvent. J Chem Theory Comput 2013;9:2733–2748.

240. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25:1422–1423.

241. Darden T, York D, Pedersen L. Particle mesh ewald: an n· log (n) method for ewald sums in large systems. J Chem Phys 1993;98:10089–10092.

242. Petrova SS, Solov'ev AD. The origin of the method of steepest descent. Hist Math 1997;24:361–375.

243. Berendsen HJ, Postma J van, Gunsteren WF van, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.

244. Li DW, Bruschweiler R. A dictionary for protein side-chain entropies from NMR order parameters. J Am Chem Soc 2009;131:7226–7227.

245. Abuin E, Lissi E. Quenching of BSA intrinsic fluorescence by alkylpyridinium cations: its relationship to surfactant-protein association. J Photochem Photobiol Chem 2003;155:157–162.

246. Liu J, Tian J, Zhang J, Hu Z, Chen X. Interaction of magnolol with bovine serum albumin: a fluorescence-quenching study. Anal Bioanal Chem 2003;376:864–867.

247. Soares S, Mateus N, De Freitas V. Interaction of different polyphenols with bovine serum albumin (BSA) and human salivary α-amylase (HSA) by fluorescence quenching. J Agric Food Chem 2007;55:6726–6735.

248. Schurr JM. Dynamic light scattering of biopolymers and biocolloids. CRC Crit Rev Biochem 1977;4:371–431.

249. Zakharov P, Scheffold F. Advances in dynamic light scattering techniques. Light Scatt Rev 4 Single Light Scatt Radioact Transf 2009:433–467.

250. Yunfang Li, Guangzhong Yang, Zhinan Mei. Spectrscopic and dynamic light scattering studies of the interaction between pterodontic acid and bovine serum albumin. Acta Pharm Sin B 2012;2:53–59.

251. Edward JT. Molecular Volumes and Stokes-Einstein Equation. J Chem Educ 1970;47:261-.

252. James TL, Matson GB, Kuntz ID. Protein rotational correlation times determined in aqueous-solution by c-13 rotating frame spin-lattice relaxation in presence of an off-resonance radiofrequency field. J Am Chem Soc 1978;100:3590–3594.

253. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 1998;120:6836–6837.

254. Miles CI, Schenk GH. Fluorescence of acetylsalicylic acid in solution and its measurement in presence of salicylic acid. Anal Chem 1970;42:656–659.

255. Nafisi S, Sadeghi GB, PanahYab A. Interaction of aspirin and vitamin C with bovine serum albumin. J Photochem Photobiol B 2011;105:198–202.

256. Jacob J, Baker B, Bryant RG, Cafiso DS. Distance estimates from paramagnetic enhancements of nuclear relaxation in linear and flexible model peptides. Biophys J 1999;77:1086–1092.

257. Constantine KL. Evaluation of site-directed spin labeling for characterizing protein-ligand complexes using simulated restraints. Biophys J 2001;81:1275–1284.

258. Liang B, Bushweller JH, Tamm LK. Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. J Am Chem Soc 2006;128:4389–4397.

259. Philson SB, Debrunner PG, Schmidt PG, Gunsalus IC. The effect of cytochrome P-450_{cam} on the NMR relaxation rate of water protons. J Biol Chem 1979;254:10173–10179.

260. Lee YT, Wilson RF, Rupniewski I, Goodin DB. P450cam visits an open conformation in the absence of substrate. Biochemistry (Mosc) 2010;49:3412–3419.

261. Yang F, Bian C, Zhu L, Zhao G, Huang Z, Huang M. Effect of human serum albumin on drug metabolism: structural evidence of esterase activity of human serum albumin. J Struct Biol 2007;157:348–355.

262. Córdova J, Ryan JD, Boonyaratanakornkit BB, Clark DS. Esterase activity of bovine serum albumin up to 160 C: a new benchmark for biocatalysis. Enzyme Microb Technol 2008;42:278–283.

263. Zundert GC van, Bonvin AM. Modeling protein–protein complexes using the HADDOCK webserver "modeling protein complexes with HADDOCK." Protein Struct Predict 2014:163–179.

264. Webb B, Sali A. Protein structure modeling with modeller. Methods Mol Biol 2017;1654:39–54.

265. Loo TW, Clarke DM. Determining the dimensions of the drug-binding domain of human P-glycoprotein using thiol cross-linking compounds as molecular rulers. J Biol Chem 2001;276:36877–36880.

266. Loo TW, Clarke DM. Location of the rhodamine-binding site in the human multidrug resistance P-glycoprotein. J Biol Chem 2002;277:44332–44338.

267. Loo TW, Bartlett MC, Clarke DM. Substrate-induced conformational changes in the transmembrane segments of human P-glycoprotein Direct evidence for the substrate-induced fit mechanism for drug binding. J Biol Chem 2003;278:13603–13606.

268. Loo TW, Clarke DM. Cross-linking of human multidrug resistance P-glycoprotein by the substrate, tris-(2-maleimidoethyl)amine, is altered by ATP hydrolysis. Evidence for rotation of a transmembrane helix. J Biol Chem 2001;276(34):31800–31805.

269. Loo TW, Clarke DM. Mutational analysis of the predicted first transmembrane segment of each homologous half of human P-glycoprotein suggests that they are symmetrically arranged in the membrane. J Biol Chem 1996;271(26):15414–15419.

270. Tombline G, Urbatsch IL, Virk N, Muharemagic A, White LB, Senior AE. Expression, purification, and characterization of cysteine-free mouse P-glycoprotein. Arch Biochem Biophys 2006;445(1):124–128.

271. Subedi GP, Johnson RW, Moniz HA, Moremen KW, Barb A. High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. J Vis Exp JoVE 2015(106).

ABBREVIATIONS

ABC: ATP-binding cassette

Adk: adenylate kinase

Amp: amplitude

AMP: adenosine monophosphate

AMPPNP: adenosine 5'-(β,γ-imido)triophsophate

ASA: acetylsalicylic acid

ATP: adenosine triphosphate

BBB: blood-brain barrier

BSA: bovine serum albumin

CHAPS: 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate

CNS: central nervous system

CNS: Crystallography & NMR System

CS: chemical shift

DDM: *n*-dodecyl-β-*D*-maltoside

DEAE: diethylaminoethyl cellulose

$\boldsymbol{E_{air}}$**:** ambiguous interaction restraint violation energy

EM: energy minimization

$\boldsymbol{E_{desolv}}$**:** desolvation energies

$\boldsymbol{E_{dist}}$**:** ambiguous interaction restraint violation energy

$\boldsymbol{E_{elect}}$**:** electrostatic energies

***E_{total}*:** total energy

ETT: eletriptan

***E_{vdw}*:** Van der Waals energies

EXNOE: exchange-transferred NOE

GI: gastrointestinal

HADDOCK: High Ambiguity-Driven Biomolecular DOCKing

HIV-1: Human Immunodeficiency Virus type 1

***H_{score}*:** HADDOCK score

HSQC: heteronuclear single quantum correlation

ICLs: intracellular loops

IHs: intracellular helices

***K_D*:** Dissociation constant

K_{Pi}: potassium phosphate

***K_{SV}*:** Stern-Volmer quenching constant

MD_{water}: molecular dynamics with explicit TIP3P model waters

MUT: site-directed mutagenesis

n: number of binding sites

NATA: N-acetyl-L-tryptophanamide

NBDs: nucleotide-binding domain

Ni-NTA: nickel-nitrilotriacetic acid

NMR: nuclear magnetic resonance

NOE: intramolecular nuclear Overhauser effect

Pgp: P-glycoprotein

***Pi*:** inorganic phosphate

PRE: paramagnetic relaxation enhancement

RMSD: root mean square deviation

SA: simulated annealing;

SAR: structure-activity relationship

SOI: sphere of influence

STD: saturation transfer difference

STDD: saturation transfer double difference

STT: sumatriptan

SXR: steroid and xenobiotic receptor

***T_1*:** longitudinal relaxation

***T_2*:** transverse relaxation

TM: transmembrane helix

TMD: transmembrane domain

TROSY: ^{1}H-^{15}N transverse relaxation-optimized spectroscopy

***V_i*:** orthovanadate

WATERGATE: water suppression by gradient tailored excitation.