UiT

THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Health Sciences

Department of Medical Biology

Medical Pharmacology and Toxicology Research Group

Computational study of SERT using new X-ray crystal structures and initial experimental verification of potential GABAB receptor compounds

Gustav Godtliebsen

Master's thesis in Biomedicine (MBI-3911) – May 2018

<u>Acknowledgements</u>

Abstract

Depression is currently one of the leading causes worldwide of suicide and disability, the most common treatment is antidepressants. Most antidepressants work by increasing the monoamine levels in the central nervous system by inhibition of the reuptake of monoamines into the presynaptic neuron, and thereby ensure accumulation of the monoamines in the synaptic cleft. The main target for most antidepressants is the serotonin transporter, which is responsible for transporting serotonin, a monoamine, from the synaptic cleft back into the presynaptic neuron. The antidepressant (S)-citalopram is an antidepressant targeting the serotonin transporter that is well tolerated among patient populations. It is also the inhibitor that was rendered with best resolution in the recent crystal structures of the serotonin transporter.

A computational study of the serotonin transporter with a focus on key amino acids for its function in both the central and allosteric sites using (S)-citalopram and its substrate serotonin as ligands in four MD simulations was performed. This shed light on some of the internal molecular mechanisms of the serotonin transporter, especially the interactions at the binding sites. Each simulation identified key amino acids between each respective ligand and binding site. For the central site simulation there was identified one key amino acid, TYR95, that both ligands had as their primary interaction and point of contact. There was also seen a disparity between (S)-citalopram and serotonin in terms of interaction types, with their preferences being hydrophobic and H-bonds respectively. For the allosteric site simulations, the interaction type trends were the same as for the central site. In the allosteric simulations the amino acids PHE335 and GLU494 were identified as the strongest interacting partners for (S)-citalopram and serotonin respectively. The interactions between the ligands and the allosteric site were also not as strong overall as the interactions of the central site. A docking study was also performed with verified inhibitors of the serotonin transporter that sought to investigate binding site interactions between the docked inhibitors and the protein. An overview of amino acid interactions was created, which allowed for the identification of amino acids that interacted with the vast majority of ligands docked into the respective binding sites. The data from the docking was also used to attempt to elucidate a connection between *in silico* and experimental results.

In order for the study to involve experimental work an initial functional cAMP assay screening was incorporated into the project from an ongoing study in the research group. The assay tested 10 compounds that had been identified through a virtual screening approach as being potential

GABA$_B$ receptor antagonists. None of the 10 compounds showed antagonistic effects on the GABA$_B$ receptor, but there were indications that three of them may be agonists. Further assays are required to confirm the three compounds status as agonists. The experimental work provided experience working in a laboratory environment and carrying out every step of a functional cAMP assay.

__Abbreviations__

2D   – Two Dimensional

3D   – Three Dimensional

5G solution   – 5mM D-glucose solution prepared in deionized water

7TM   – Heptahelical Transmembrane Domain

ADMET   – Absorption, distribution, metabolism, excretion, toxicity

CADDD   – Computer-Aided Drug Discovery and Development

cAMP   – cyclic adenosine monophosphate

Cas   – Ca^{2+}-sensing

CNS   – Central Nervous System

CRD   – Cysteine Rich Domain

DA   – Dopamine

DMSO   – Dimethyl sulfoxide

EL   – Extracellular loops

GABA   – γ-aminobutyric acid

GABA$_B$ – γ-aminobutyric acid B

GPCR   – G-protein-coupled receptor

GRAFS   – Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, and Secretin

H-bonds   – Hydrogen Bonds

HBSS   – HEPES-bufered hank's balanced salt solution

HEPES   – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

IL   – Intracellular loops

K   – Kelvin (Temperature)

kDa   – Kilodaltons

K_i	–	Inhibitor Constant
MD	–	Molecular Dynamics
MAOIs	–	Monoamine Oxidase Inhibitors
mGlu	–	Metabotropic glutamate
MM	–	Molecular Mechanics
ml	–	Milliliter
mM	–	Millimolar
µl	–	Microliter
µM	–	Micromolar
NE	–	Norepinephrine
nM	–	Nanomolar
NMR	–	Nuclear Magnetic Resonance
NPT	–	Constant-temperature, constant-pressure ensemble
ns	–	Nanosecond
OPM	–	Orientations of Proteins in Membranes
PDB	–	Protein Data Bank
PNS	–	Periphery Nervous System
ps	–	Picosecond
QM	–	Quantum Mechanics
QM/MM	–	Molecular Mechanics/Quantum Mechanics
RLU	–	Relative Light Unit
RMSD	-	Root Mean Square Deviation
RMSF	–	Root Mean Square Fluctuation
SERT	–	Serotonin Transporter

SEM   – Standard Error of Mean

SID   – Simulation Interactions Diagram

SLC6   – Solute Carrier 6

SSRI   – Selective Serotonin Reuptake Inhibitor

TCAs   – Tricyclic Antidepressants

TM   – Transmembrane α-helix

UniProt   – Universal Protein Resource

VFT   – Venus Flytrap

VMAT1   – Vesicular Monoamine Transporters 1

VMAT2   – Vesicular Monoamine Transporters 2

Å   – Ångström

Table of Contents

1. Introduction

1.1 The Nervous System

The nervous system plays a critical role in the human body and has a vital role in the body's ability to maintain homeostasis. Communication of both the internal and external environment, as well as the coordination of corresponding activity between the widely disperse cells, tissues and organs that make up the human body is the primary function of the nervous system. A few examples of the many functions that the nervous system is responsible for managing within the body is the activation of muscle contraction, integrating respiratory system activity with the levels of blood oxygen, carbon dioxide and pH levels, as well as the modulation of the motility and secretions of the digestive system. Being one of two major control systems within the body, the other being the endocrine system with its relatively slow and long-lasting signals, the nervous system functions through the use of rapid electrical signals that travel through the body along the cellular membranes of nerve cells (1).

The nervous system itself is divided into two parts, the Central Nervous System (CNS) and the Peripheral Nervous System (PNS). The CNS is made up of the brain and spinal cord, while the PNS is composed of the nerves connecting the brain and spinal cord to the tissues of the body. Both the CNS and the PNS are made up of the same basic unit, nerve cells, which are also called neurons (1). The structure of the neuron cells is itself highly polarized with distinct subcellular domains which help to promote a variety of functions. From a morphological standpoint, the typical neuron has three clearly definable regions; the cell body, the dendrites emanating from the cell body and the axon. The cell body itself, which is also known as the perikaryon, is where the nucleus and major cytoplasmic organelles are contained. The dendrites emanating from the cell body vary in number and the area they stretch across as well as their size and shape varies with the specific type of neuron. The dendrites act as signal receivers and accept signals from the axon terminals of other neuron cells. The axon, the last defined region of a neuron cell, conducts electrochemical and action potentials that the cell has received. It will in most cases stretch significantly further from the cell body than the emanating dendrites, it also emits a varying number of branches known as collateral axons. The axon and collateral axons have what is known as axon terminals at their end-points, which are used to send signals transmitted through the axon to the dendrites emanating from other neurons (2). The nervous system transmits information through these rapid electrical signals, which are generated by the neurons and used to communicate with either the same cell or neighboring cells. For most

neurons the response to receiving an electrical signal is the release of chemical messengers called neurotransmitters that serve as a means of communication with other cells (1, 2).

There are numerous different neurotransmitters that act in the CNS and PNS, the exact number of these is not known, but it is known to more than 100. Although there is a significant degree of diversity among them it is still possible to classify them into two broad categories. The first of these categories is small-molecule neurotransmitters, which in general mediate rapid synaptic actions. The second category is neuropeptides, that most commonly modulate functions that are slower and ongoing. Despite the neurotransmitters being possible to categorize into two categories they all undergo a similar cycle of use (Figure 1). The first stage of this cycle is the synthesis of the neurotransmitter and its subsequent packaging into vesicles, all of which occurs in the presynaptic cell, the cell the signal originates from. The actual synthesis and packaging of the neurotransmitter varies somewhat between the small-molecule neurotransmitters, which are synthesized in the axon terminals, and the neuropeptides, which are synthesized in the cell body alongside enzymes that are needed to modify them before release. The neuropeptides are transported from the cell body through the axon before they are enzymatically modified in the axon terminals and subsequently released. The second stage involves the release of the neurotransmitter from the presynaptic cell and its subsequent binding to receptors on one or more postsynaptic cell, the postsynaptic cells are the receivers of the signal. The third and final stage is the rapid removal of the neurotransmitter and/or its degradation. This stage also has some variations between small-molecule neurotransmitters and neuropeptides. Small-molecule neurotransmitters either undergo reuptake directly or they are enzymatically degraded and their metabolites undergo reuptake. After this reuptake, they are ready to undergo another round of synthesis (only the metabolites), packaging, release and removal. Neuropeptides do not undergo reuptake, they instead diffuse away from the axon and are degraded by proteolytic enzymes (3).
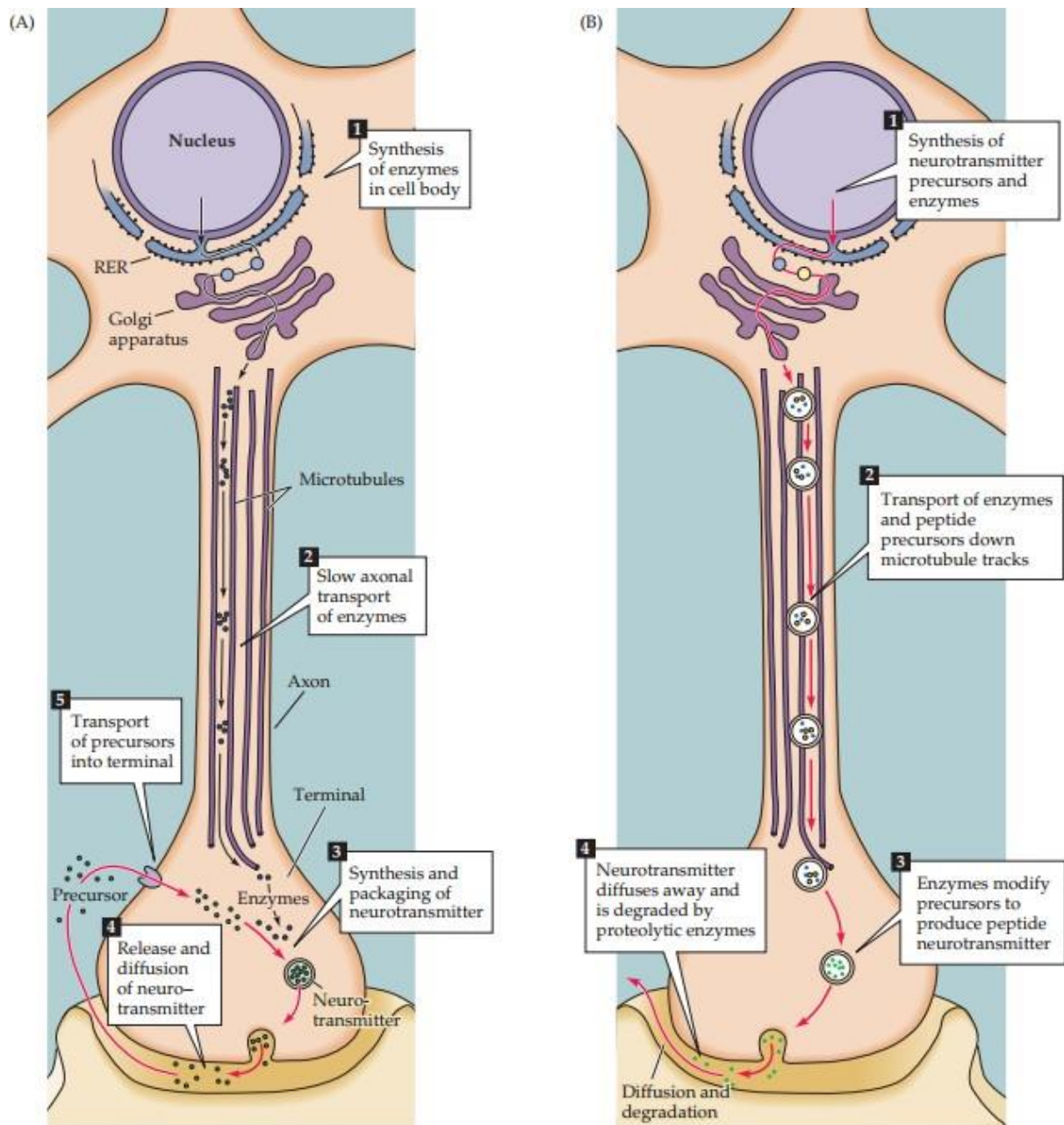
Figure 1: The figure displays the process of synthesis and release, followed by reuptake or degradation for respectively neurotransmitters (left) and neuropeptides (right). There are also monoamine transporters that transport neurotransmitters back into the presynaptic neuron, these transporters are part of step 5 in the (left) figure and are not visualized in the figure. The figures has been modified from (3).

These neurotransmitters serve a crucial role in the communication of the nervous system and abnormalities in their functionality are significant contributing factors to a wide range of both neurological and psychiatric disorders. Many therapeutic strategies aiming at treating such conditions therefore target various stages of a neurotransmitters functional life (3).

1.2 Neurological and Psychiatric Disorders

Abnormalities related to neurotransmitters, such as abnormal regulation of their release or irregular extracellular concentration levels of the neurotransmitters, has a wide variety of impact upon an individual's health. These abnormalities are the basis of hypotheses regarding the neuronal foundation of both behavioral and cognitive disorders. They are also considered symptomatic in the diagnosis of neuropsychiatric and neurodegenerative disorders. The abnormalities themselves are indications for that the ability of neurons to process information is being disrupted (4). This can cause, as evidenced by the many disorders such disruption is associated with, substantial health issues that are difficult to treat and the treatments often have numerous side-effects (5).

1.2.1 Depression

A common but serious disorder that is known to be related to abnormalities in neurotransmitter function is depression (6), also known as major depressive disorder and clinical depression. The symptoms of depression are wide and can vary significantly between patients, the degree of severity these symptoms exhibit also varies significantly. Core symptoms of a major depressive episode are having a «depressed mood» and the "loss of interest and pleasure in nearly all activities", symptoms such as fatigue, anxiety and sleep disturbance are also quite common (7). The details are to this day not known regarding the underlying biological cause of depression. The hypothesis that has been dominant is that depression develops because of a low level of the monoamines, serotonin, dopamine (DA) and norepinephrine (NE), throughout certain areas of the CNS, and particularly in the limbic systems. The basis for this was that the first drugs with a therapeutic effect in depression acted by increasing the levels of these monoamines. It is not certain that the low levels of monoamines is the cause of depression, it is a possibility that it is a symptom of depression rather than a cause (8).

Depression is sometimes characterized into two subtypes based on clinical symptoms displayed and rated on a scale based on the severity of the symptoms. The subtypes are melancholic depression and atypical depression, which historically were recommended different pharmacological treatments (7). Melancholic depression had tricyclic antidepressants (TCAs) and electroconvulsive therapy as the recommended treatment (9), while atypical depression had classical monoamine oxidase inhibitors (MAOIs) as the recommended treatment as patients with atypical depression seemed to respond better to MAOIs than tricyclic

antidepressants (10). The current generation of antidepressants have less of a distinction between them in regard to the treatment effectivity of depression subtypes, and there are indications for that depression is more complex than just melancholic and atypical subtypes (7).

1.2.2 Antidepressants

All of this is indicative of the complexity and individual variation inherent in the disorder and by extension the complexity and variation of the abnormalities in neurotransmitter function. Despite this complexity and variation between patients there are numerous treatment options available for depression. These treatment options range from psychological therapy, electroconvulsive therapy and pharmacotherapy to a combination of those (11). The drug treatments affect the primary irregularities that are seen to play a major role in causing the symptoms of depression. These medications do not work for all patients with depression, and patients must often try numerous drugs before they find one that treats their symptoms. These medications also frequently cause a wide array of side effects (12).

The current generation of antidepressants, such as the selective serotonin reuptake inhibitors (SSRIs), have a greater degree of selectivity than the previous generation, they also offer an improved level of safety of use and tolerability. The various TCAs and MAOIs among the first-generation of antidepressants often possessed unwanted side-effects and also toxic effects in cases of overdoses, factors that limited their application in treatment to various degrees. Between the current generation antidepressants and the first-generation there is a common feature among all the antidepressants that remain on the market, they all increase the monoamine levels though by a variety of therapeutic mechanisms. Specifically they increase the levels DA, NE and/or serotonin (12). The improved selectivity of the newer generation allows the antidepressants to more narrowly affect the nervous system and thereby avoid some of the side effects caused by the compound interacting with unintended targets (13).

1.3 Targets for development of antidepressant drugs

The CNS and PNS both have a significant variety in types of neurons that they are made up of, with there being about 100 distinct types of neurons possessing a wide variety of functions and morphologies. These distinct neuron types have all inherited the same gene complement, with their differences stemming from the restricted set of genes that they actively express. The restricted expression ensures only certain molecules are produced, be they enzymes, secretory products such as neurotransmitters, membrane constituents or structural proteins, while others are absent from that specific type of neuron. What makes one cell type unique compared to another is the specific set of molecules that it produces, the specific molecules enabling the cell to carry out its role (14).

Of particular interest in the study of treatments for neurological and psychiatric disorders are the various membrane constituents, such as receptors and ion channels, and the various secretory products of neurons. These are the means by which a neuron communicates and disruption at this level of communications are theorized to be the cause or significant contributing factors to a large variety of disorders that afflict human beings of all ages (15). In modern drug development the receptors spread throughout the CNS and PNS are seen as critical targets in the treatment of many disorders, with some receptors standing out as of particular interest for certain diseases. Both the serotonin transporter (SERT) and the γ-aminobutyric acid B (GABA$_B$) receptor are among those standing out as of interest, with both proteins playing critical roles in the function of the nervous system (16, 17).

1.3.1 Monoamine Transporters

Located in the plasma membranes of monoaminergic neurons the monoamine transporters serve a crucial role in the regulation of a wide variety of bodily functions (18). They are highly significant transmembrane proteins, including among their numbers SERT, the dopamine transporter and the norepinephrine transporter (19). All of these proteins use ion gradients (Na$^+$, Cl$^-$) as their energy source for the transportation of their respective monoamines into or out of neurons. These transports play a critical regulatory role with their major function being the termination of monoamine transmission, which is accomplished by the removal of substrates from the synaptic cleft by inward transport into the presynaptic neuron (18).

There is also the vesicular monoamine transporters 1 and 2 (VMAT1 and VMAT2) located in the membrane of intracellular synaptic vesicles. These transporters make use of a proton gradient as their source of energy for sequestering cytosolic monoamines into vesicles which are subsequently released by exocytosis into the synaptic cleft (18). The VMAT transporters have been indicated by numerous studies to play a crucial role in the storage of, sorting and release of neurotransmitters. They have also been implicated in the fine-tuning of the informational output of both endocrine and neuronal sources (20-24).

These transporters share an overall function in that they regulate the tempo-spatial components of monoamine transmission, and are in fact so crucial that the loss of even a single type of these transporters can cause severe diseases or even death (18). One example showcasing this in humans for the dopamine transporter would be that only two loss-of-function mutations, specifically L368Q and P395L, results in infantile parkinsonism-dystonia in humans (25). The monoamine transporters play a vital role in the human body, but they are also therefore a vulnerability.

1.3.1.1 Solute Carrier Family 6

The human genome contains 20 members of the solute carrier 6 (SLC6) family, these 20 members are comprised of transporters of amino acids, energy metabolites, osmolytes and neurotransmitters. The members of the SLC6 family play crucial roles in neurotransmission as well as homeostasis at both the cellular and whole body level (26). The transporters in the family are secondary transporters that function by coupling the potential energy that is stored in preexisting ion gradients to enable them to transfer molecules across the cell membrane against the concentration gradient of the molecules. These transporters use a substrate translocation model known as the alternate access model, which entails the use of a central binding site that is connected to the intracellular and extracellular environments of the cell through the use of permeation pathways. As from the name, alternate access model, this type of transporter has its binding site only accessible to one permeation pathway at a time, which is accomplished by the transporter shifting conformations alternatively between the inward- and outward-facing conformations of the transporter (27, 28).

The family, based on sequence similarity, can be subdivided into four branches (Figure 2); the monoamine transporter branch, γ-aminobutyric acid (GABA) transporter branch and the amino acid transporter (I) and (II) branches (26).
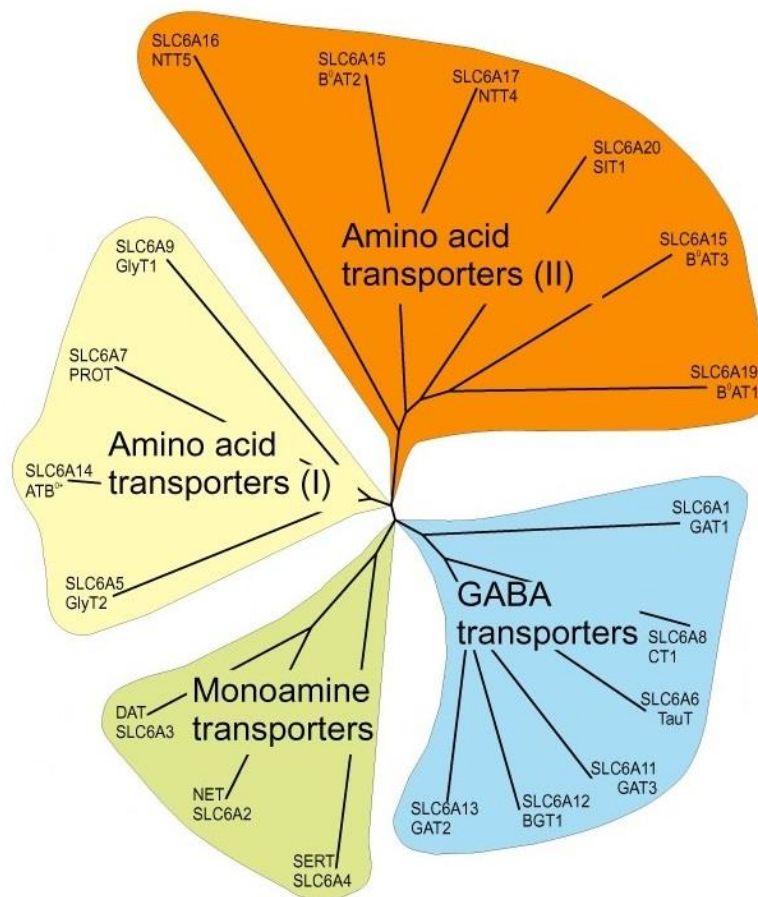
Figure 2: The figure displays a visualization of the sequence similarities of the human SLC6 family transporters. T-coffee (29) was used to align all peptide sequences of the human SLC6 members and the similarities were then visualized using the TreeView (30) application. The figure is modified from (26).

Each of the different branches in the SLC6 family has a substrate preference that is different from the others. For the monoamine transporters branch the preferred substrates are serotonin, NE or DA. Each of these neurotransmitters have regulatory roles in the CNS and affect the activity of numerous pathways. Their regulation has a particularly prominent effect on mood, appetite, anxiety, attention, addiction, aggression, depression etc. (31, 32). The GABA transporter branch has transporters for GABA, creatine, betaine and taurine. GABA is of particular interest among these, being the major inhibitory neurotransmitter that affects the brain. Creatine has high-energy phosphate bonds that make it excellent at replenishing ATP, it is used as a storage compound for this purpose by primarily, but not exclusively, muscle tissue and the brain (33). Betaine and taurine both play a role in the regulation of the maintenance of cell volume and fluid balance through osmosis, both compounds fall within the group of compounds known as osmolytes. Then there are the amino acid transporter branches I and II.

The amino acid transporter branch I is made up of transporters for proline, glycine and also the general amino acid transporter ATB(0,+) (34, 35). The amino acid transporter branch II has amino acid transporters that are involved in the transport of amino acids in epithelial tissue and the brain (36). The majority of transporters in this branch are involved in amino acid homeostasis by virtue of their ability to accept a variety of neutral amino acids (26).

1.3.1.2 The Serotonin Transporter

One member of the SLC6 family is SERT, a monoamine transporter that plays a key role in the regulation of the serotonergic system. Specifically, it plays a role in terminating serotonergic neurotransmission by transporting serotonin into the presynaptic neuron from the synaptic cleft (Figure 3). Once in the presynaptic neuron the serotonin is either recycled into vesicles for storage or the cell converts it into an inactive metabolite (28). As it plays an important role in the serotonergic system, which is implicated in many physiological processes and behaviors as well as psychiatric disorders. It is considered an important target in the development of pharmaceuticals and is the target of numerous antidepressants (37-39).
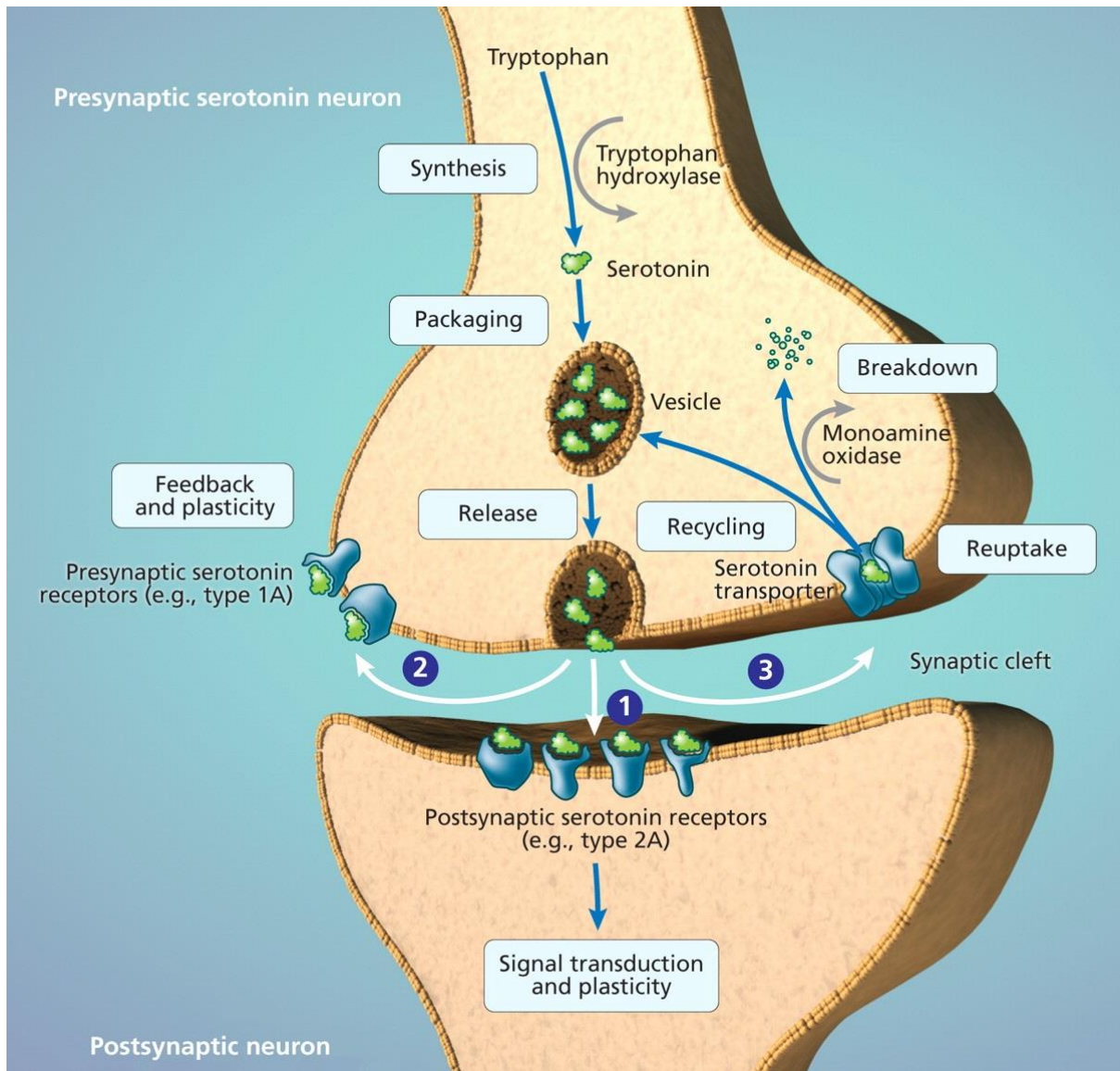
Figure 3: The figure displays the serotonin synapse. In the presynaptic neuron serotonin is synthesized by an enzyme named tryptophan hydroxylase from tryptophan. Once serotonin completes the subsequent packaging step, where it is packaged into vesicles, it is released into the synaptic cleft. The release into the synaptic cleft will only occur when the neuron has undergone sufficient stimulation. The serotonin that is released into the synaptic cleft from the presynaptic neuron has several actions it can undergo. 1) The serotonin binds to the appropriate receptors on the postsynaptic neuron, resulting in the signal that stimulated the presynaptic neuron being further transduced. 2) The serotonin binds to receptors on the presynaptic neuron, i.e. the neuron which released the serotonin, enabling feedback which in turn plays a role in regulating the neurons plasticity. 3) The serotonin undergoes reuptake by the presynaptic neuron through SERT, subsequently either being recycled for release in the future or broken down by the enzyme monoamine oxidase (40). The image is modified from an image by Lianne Friesen and Nicholas Woolridge.

As with other members of the SLC6 family it is also a secondary transporter that uses a changing conformation, switching between inward- and outward-facing conformations, to

transport its substrate across the cell membrane against the concentration gradient. With these conformation shifts the binding site of SERT is only available from one side of the cell membrane at a time (27, 28). SERT gains the energy required for the transport of its substrate, serotonin, from an electrochemical gradient made up of sodium and chloride ions (41-44). In order for SERT to undergo the necessary changes to its conformation that are required for translocation all of its solutes are thought to be required to be bound in a 1:1:1 stoichiometry. The solutes required to bind for translocation are specifically serotonin, Na^+ and Cl^-, in addition there is an efflux of a single K^+ ion that occurs separate from SERT itself (45, 46).

For SERT the ions Na^+ and Cl^- also play a key role in the ability of inhibitors to bind to the protein. All known SERT inhibitors, with ibogaine being the only known exception, are stimulated to bind to SERT by Na^+. Ibogaine is thought to be an exception to this because it is believed to stabilize an inward-facing conformation of SERT, the affinity of this inhibitor for SERT increases with the removal of Na^+ (47-50). There are also several SERT inhibitors that are stimulated by Cl^-, the most notable among these are citalopram, fluoxetine, imipramine and sertraline (48, 50).

The structure of SERT has been the subject of study for many years (51, 52), but it is only in recent years that it was successfully crystalized (53). Having a crystal structure of SERT allows for the *in silico* study of the protein with a degree of accuracy that could not be achieved using only the amino acid sequence or a homology model (51, 54). The crystal structure that was rendered though X-ray crystallography is in the outward-open conformation (Figure 4), having the antidepressant drug (S)-citalopram bound to its orthosteric binding site, which is also often referred to as the central binding site. The central binding site of SERT is located halfway through the membrane in a cavity that is formed by residues from the transmembrane (TM) α-helices 1, 3, 6, 8 and 10. Another rendering of the SERT crystal structure found that (S)-citalopram was also bound to an allosteric site located in the extracellular vestibule that was approximately 13Å from the central site (53). The allosteric site of SERT has been shown to affect the inhibitory potential of (S)-citalopram, lowering the inhibitory potential if the allosteric site is disabled through mutation (55). SERT has a total of 12 TM helices that span the membrane, TMs 1 and 6 have short non- helical conformation regions in the area they skirt the central site for ligand binding. These short non- helical regions are contributors of residues that bind inhibitors and also take part in the coordination of NA^+ and Cl^- ions (53).
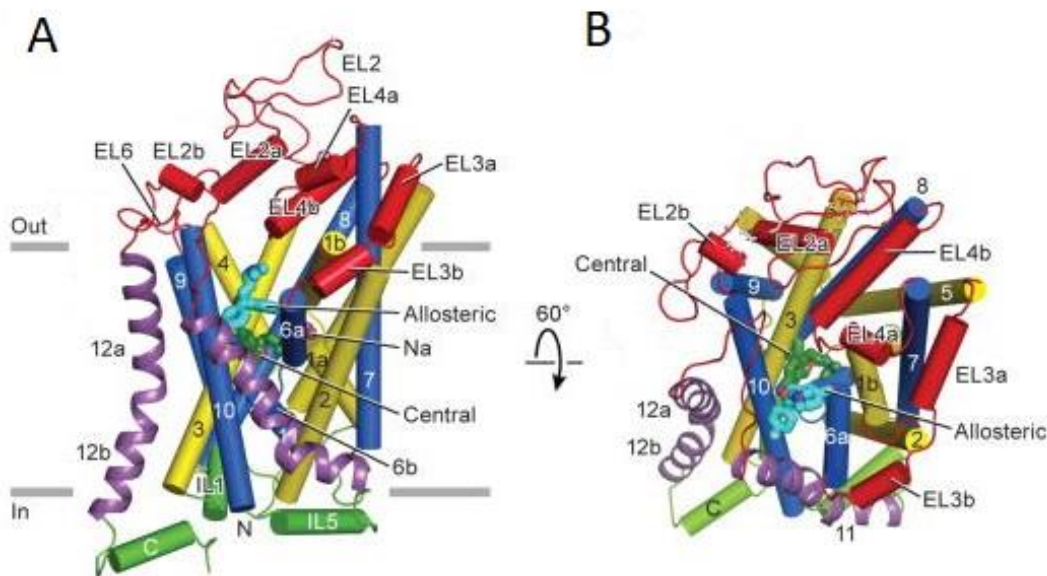
Figure 4: The above figure displays the structure of SERT. A. The structure of SERT parallel to the membrane. B. The structure of SERT from the extracellular side of the membrane. (S)-citalopram in the central and allosteric sites are shown as sticks, they are dark green and cyan in color respectively. The salmon colored spheres are Cl$^-$ ions (53). The numbered columns are the TM helices, EL refers to extracellular loops and IL refers to intracellular loops. The figure is modified from (53).

SERT is the target of the antidepressant SSRI citalopram, which has been indicated as the best tolerated SSRI among patients (13). (S)-citalopram is the co-crystalized ligand for both the crystal structure with the best resolution (PDB ID 5I71, 3.15 Å) and also the crystal structure where the central and allosteric site was found to bind a ligand (PDB ID 5I73, 3,24 Å).

1.3.2 G-Protein-Coupled Receptors

A large group of proteins that is considered to contain numerous therapeutic targets of great potential for a wide spectrum of diseases are the G-protein-coupled receptors (GPCRs). This group of receptors is responsible for mediating the majority of physiological responses to neurotransmitters, hormones and environmental stimulants. The environmental stimulants that GPCRs are responsible for mediating are taste, olfaction and vision (56). There is also a significant variety and wide range to the ligands which may interact with GPCRs, included among these ligands are photons (light), ions, amino acids and large proteins (57). This large group of proteins play crucial rolls throughout the human body and defects or irregularities in many of the groups member proteins may lead to severe diseases (58).

14

When it comes to families of membrane proteins the GPCR family is the largest, with all members, at the most basic level, sharing several structural characteristics. The GPCRs all share the structural feature of having seven α-helical segments that span the cellular membrane, these segments are separated by loop regions that alternate between being intracellular and extracellular (56). This type of structural feature is called a heptahelical transmembrane domain, 7TM for short (59).

The GPCRs in vertebrates can be divided into five families using the GRAFS classification system (60). The term GRAFS comes from the names of the families; Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, and Secretin. Of these families three of them, rhodopsin (Family A), secretin (Family B) and glutamate (Family C), correspond to an earlier GPCR classification system which split the GPCRs into a clan system going from A-F (61, 62). The adhesion and Frizzled/Taste2 families were not part of the GPCR clan system and were introduced with the GRAFS classification system (60). While the GPCR families do share similarities, there are significant differences between even individual GPCRs. Their signal-transduction activities are enabled through unique combinations of a multitude of G-protein subtypes, G-protein-independent signaling pathways and the numerous complex processes with which they are regulated (56). Together this makes for a complex web of potential avenues for drugs to target in order to treat various diseases that occur as a consequence of these signal-transduction pathways being disrupted (58, 63).

1.3.2.1 Family C receptors

Among the GPCRs the Family C receptors, also known as the glutamate family of receptors, share the 7TM domain with the other GPCRs, but are distinguished by two structural features that are unique to the Family C receptors among the GPCRs. The first of these features is a large extracellular domain featuring a Venus flytrap (VFT) module as well as a cysteine rich domain (CRD). The VFT module contains the orthosteric sites of Family C receptors and the extracellular domain itself is situated away from the center of the 7TM helices. While the 7TM domain is a common feature among all GPCRs and is a conserved feature, it is noteworthy that the Family C receptors only possess allosteric sites in the 7TM domain. In the Family C receptors the majority of receptors contain a CRD, with only the $GABA_B$ receptor not containing one. The second of the features is that these GPCRs form constitutive dimers with modes of activation that are unique when compared with the other families of GPCRs. This

takes the form of homodimers for the metabotropic glutamate (mGlu) and Ca^{2+}-sensing (Cas) receptors, while for the GABA$_B$ receptor and T1Rs it takes the form of heterodimers. This dimerization leads to an activation process that is particularly complex with allosteric interactions occurring between different dimer domains (57).

In addition to the mGlu, Cas and GABA$_B$ receptors the Family C GPCRs are made up of pheromone receptors, amino acid and sweet taste receptors, odor receptors in fish and a few orphan receptors (59). The mGlu, Cas and GABA$_B$ receptors play critical roles in numerous disorders that may affect the CNS and calcium homeostasis, they are therefore considered therapeutic targets of great significance (64, 65). The taste receptors in the Family C receives a great deal of attention from the food industry, with many food companies being invested in taste additives which target these receptors (65).

1.3.2.2 GABA$_B$ Receptor

A member of the Family C of GPCRs that is of particular note in the study of numerous diseases and development of new pharmaceuticals is the GABA$_B$ receptor (17, 66, 67). It is one of three native receptors for GABA, the other two native receptors for GABA are the ionotropic GABA$_A$ and GABA$_C$ receptors. As GABA is the primary inhibitory neurotransmitter of the body there are many physiological functions that are regulated through GABA-mediated neurotransmission. Found in both pre- and postsynaptic inhibitory and excitatory synapses the GABA$_B$ receptor is spread throughout the CNS and the also the autonomic division of the PNS (68, 69).
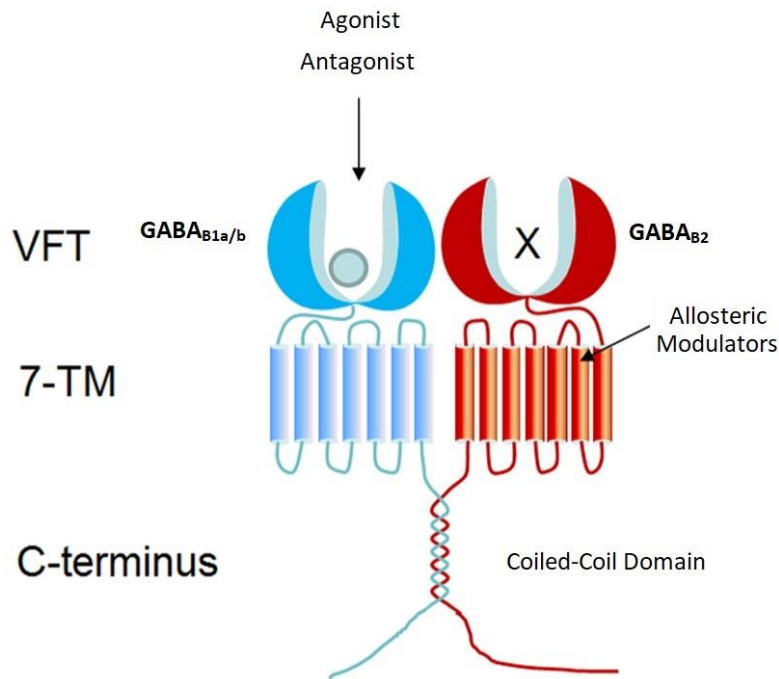
Figure 5: The figure displays the structural organization of the GABA$_B$ receptor. The subunits that compose the heterodimer of the GABA$_B$ receptor, GABA$_{B1a/b}$ and GABA$_{B2}$, can be seen in blue and red respectively. There are two variants of the GABA$_{B1a/b}$ subunit termed GABA$_{B1a}$ and GABA$_{B1b}$. Binding of agonists and antagonists binding occur in the N-terminal VFT domain of the GABA$_{B1a/b}$ subunit. The figure is modified from (70).

The GABA$_B$ receptor is as mentioned previously a heterodimer, which is made up of two subunits called GABA$_{B1a/b}$ and GABA$_{B2}$ (Figure 5). Both these subunits are made up of three domains that are distinct from one another, the N-terminal extracellular VFT domain, the 7TM domain that all GPCRs possess and the intracellular C-terminal tail. There are two binding sites for ligands that have been identified in the GABA$_B$ receptor. The orthosteric binding site which is recognized by GABA, other agonists and antagonists is located in the VFT domain of the GABA$_{B1a/b}$ subunit, while an allosteric binding site is located in the 7TM of the GABA$_{B2}$ subunit (Figure 5). The allosteric site is capable of binding both positive and negative allosteric modulators (68).

1.3.2.3 $GABA_B$ Receptor Signaling

As the $GABA_B$ receptor is one of three receptors for GABA, the primary inhibitory neurotransmitter of the body, the signaling pathways that involve the $GABA_B$ receptor are complex and interconnected. The receptor is located both pre- and post-synaptically and is responsible for the regulation of synaptic transmission, which is accomplished through either inhibiting neurotransmitter release or the dampening of postsynaptic excitability (71). It is associated with a subset of G-proteins which themselves regulate specific ion channels or trigger or inhibit cyclic adenosine monophosphate (cAMP) cascades (72). Both presynaptic and postsynaptic activation leads to the G-protein mediated inhibition of adenylyl cyclase and thereby also a decrease in cAMP for the cell. Voltage gated calcium channels are also inhibited by both presynaptic and postsynaptic activation, with the presynaptic inhibition of the voltage gated calcium channels leading to a reduced release rate of GABA, serotonin, NE and DA (73).

The complexity and multitude of the signaling pathways associated with the $GABA_B$ receptor is the reason why it is difficult to create a drug which only affects the $GABA_B$ receptor in one very specific manner with no side effects. The idea of creating a biased signaling drug, essentially a drug which would prefer one specific pathway available to a receptor, is of particular interest within drug discovery and development focused on proteins such as GPCRs with their multiple potential pathways (74). Biased signaling is therefore a focus of much research, also with the $GABA_B$ receptor, where it is hoped that biased signaling may provide the path to the development of an antidepressant targeting the $GABA_B$ receptor with less side effects than is possible with today's drugs (75, 76).

1.4 In silico methods in Drug Discovery and Development

For new compounds to come to the market as antidepressants or other forms for medications it has to go through the process of drug discovery and development. This process has changed and evolved significantly over time, new technology allowing for new possibilities. From laboratory testing of compounds to determine their effects in the early stages of drug development to the current focused multi-step methodology that aims to bring highly selective and specific compounds to the market at a lower cost (77, 78).

In the past compounds were discovered through trial and error in the laboratory and this process did indeed discover numerous compounds with effects worthy of clinical use, there was

however often numerous problematic side effects associated with these compounds. Over time related compounds were discovered that sometimes had less side effects, but these discoveries could take many years to arrive and many more to enter clinical settings (79). In the modern age with very steep costs of bringing a new compound to the market there have been developed numerous methods of evaluating a compounds efficiency as a pharmaceutical. What makes these methods different from those of the past is that many of them can be carried out *in silico*, removing the need to buy large libraries of compounds and testing them individually. Compounds can be evaluated *in silico* based on their ADMET (absorption, distribution, metabolism, excretion, toxicity) qualities, screening out any compounds that had qualities that would make them unfit to be used in marketable pharmaceuticals. The compounds can also be screened against both known ligands of the target protein to select ones with similar qualities and against the actual protein, docking the compound as a ligand into the protein to calculate how it would affect the protein. These methods, known as ligand and structure based virtual screening, are two methods in the wide library of methodologies that have become available with the advent of Computer-Aided Drug Discovery and Development (CADDD) (80).

Modern methods of CADDD are targeted searches for compounds that affect a target protein in a desired fashion, preferably with a great deal of specificity as to avoid unwanted side effects. One of the areas where this has been of particular interest is in the development of pharmaceuticals to treat behavioral and cognitive disorders, as well as neuropsychiatric and neurodegenerative disorders (16, 17).

1.5 Molecular Modeling

In order to understand the intricacies of biology it is critical to study the macromolecular structures that are responsible for the molecular interactions which are the basis of biological function. The first structure determinations occurred in the 1950s and since then, both for proteins and nucleic acids, there has been a continuous advance in the understanding of how macromolecular structures are built and function (81). Alongside these advances in structural biology there have been advances in computational technology which have been harnessed as the technology developed in order to aid in the understanding of the complexities of large macromolecules. The computational advances allowed for the methodology known as molecular modelling to be developed, a method which allows for analysis and visualization of the structure of biological macromolecules in a three dimensional system (82). A good example

showing how far the technology has come since then can be seen by comparing the capabilities of molecular dynamics (MD) simulations then and now. An MD simulation is basically the simulation of molecules and how they interact with each other on a nanosecond (ns) to millisecond timescale, there is a more in-depth explanation of the method further below. The first MD simulations were run in the late 70s (81, 83, 84), and contained only a few hundred atoms, with current day computational power the MD simulations are regularly run containing systems of 50,000-100,000 atoms with some simulations reaching up to around 500,000 atoms with the appropriate computer facilities. The length of the simulations has also increased greatly, going from being in the low picoseconds to reaching multiple milliseconds if necessary (81).

The advances in computation has made molecular modeling a valuable tool in the study of large macromolecules. The basis of the methodology is the creation of the computational models of the three-dimensional molecular structures of the macromolecules. The aim of the molecular models varies from project to project but a common goal is being able to relate biological activity to the structure of the molecule (81, 82). As molecules in nature always seek the lowest possible energy state this is something which must be reflected in any molecular model. In practice this is done by computing the theoretical potential energy of a molecule as a function of the positions of the constituent atoms. Molecular modeling has two main methods for computing this potential energy, the first of which is the Quantum Mechanics (QM) method and the second of which is the Molecular Mechanics (MM) method (83). Both methods have their own advantages and disadvantages, with the QM method being a more accurate and computationally intensive methods suited for systems of up to around 100 atoms and the MM method sacrificing some accuracy for speed, allowing it to handle significantly larger systems (85).

1.5.1 Molecular Mechanics

As mentioned previously molecular modeling has two methods for computing potential energy when doing molecular modeling, these being QM and MM. As QM takes into account the sub-atomic particles when performing energy calculations, it is much computationally intensive than MM, which treats each atom as a single unit with its own force field which approximates the effects of sub-atomic particles through classical mechanics. There also exists hybrid methods, called QM/MM, which use QM methods in areas of interest and MM for the rest of

model system. This method has the advantage of providing greater accuracy in the area of interest as well as retaining some of the speed provided by MM (85). While this method is efficient it is also in some ways more difficult, the borders for QM calculations must be set and changes to these borders affect both the results and the time required to compute the results (86). Both QM and QM/MM have their own merits when used appropriately, the same can be said for MM. If a study does not require the level of accuracy QM provides then using only MM is less computationally intensive and therefore beneficial.

With MM treating each individual atom as a single unit the force fields governing them become very important to the accuracy of the computer model. Each individual unit is governed by numerous energy terms, these terms describe the deformation of bond lengths, torsion angels and the bond angles between the various individual units. The computer model also requires energy terms that apply to non-bonded atom pairs, describing the attractive and repulsive forces of van der Waals forces and a term that considers the coulombic interactions between charges (82). Even with the numerous energy terms having to be taken into consideration for each atom MM is still, computationally, significantly faster with the atom being treated as a single unit rather than a collection of sub-atomic units (85).

1.5.2 Force Fields

In computational models using MM their collection of energy terms, mentioned above, are what make up the forcefield that is used to calculate the interactions between the atoms making up a molecule and between separate molecules. These energy terms have reference values for what an "unstrained" bond length, angle, torsion or nonbonded interaction is and this is used to calculate the total energy of a molecule by looking at deviations from these reference values. The term force field, when referring to MM, is a collection of empirically derived values of "unstrained" force constants (87). The basic formula used to describe force fields is possible to write in the following manner:

$$E_{tot} = E_{bonded} + E_{nonbonded}$$

E_{tot} is the potential energy, while E_{bonded} and $E_{nonbonded}$ describes the covalent and noncovalent boning energy terms. This formula can be further subdivided, giving a clearer understanding of the terms that make up E_{bonded} and $E_{nonbonded}$ (87). A further subdivided formula can be written as:

$$E_{tot} = (E_{bond} + E_{angle} + E_{dihedral}) + (E_{elec} + E_{vdw})$$

where E_{bond} describes bond stretching, E_{angle} describes angle bending and $E_{dihedral}$ describes the torsional energy terms. For the nonbonded energy terms there is E_{elec} which describes the electrostatic forces and E_{vdw} which describes the van der Waals forces (87).

All the energy terms that are used to build the force fields used in MM are empirically derived, meaning there are not any energy functions or parameters that are "correct", but rather close approximations of the measurable phenomena of the natural world. What this translates to in computational models is slight variations in projects using MM with different force field models (88).

1.5.3 Structure Determination

In order to employ the use of computational methods in the study of a protein it is necessary to have the structure of that protein rendered in a virtual environment. There are several ways for a structure to be rendered virtually, with some methods being more accurate in the structure determination than others (89). Of methods that seek to determine the structures of proteins there are several, with the most common and powerful methods being X-ray crystallography, followed by Nuclear Magnetic Resonance (NMR) spectroscopy. An X-ray crystallography structure determination has no limitations on the size of the protein that is being determined, it is also the technique which provides the most precise atomic detail. The downsides to X-ray crystallography is that it requires a protein that can be crystalized, information regarding a proteins dynamics may also be limited if it is determined by X-ray crystallography. NMR spectroscopy does not require a protein to be crystalized, but rather works on macromolecules in solution. This makes NMR spectroscopy a technique suitable for use in cases where X-ray crystallography is not possible, however NMR spectroscopy does have a limitation on the size of the molecule it can determine (molecular weights below 50 kDa) and also has a lower resolution than that provided by X-ray crystallography (90). The two techniques can be used in conjunction and deliver complementary information, especially useful in the structure determination of small proteins (91). Of the entries in the PDB database 125271 of them were determined by X-ray crystallography, while 12206 were determined by NMR spectroscopy as of May 2018 (92). NMR spectroscopy is however the dominating technique for structure determination of molecules with a size of less than 10 kDa molecular weight (90).

There is also a third technique for structure determination, cryo-electro microscopy, which has recently achieved resolutions which allow for structure determination. This method involves the three-dimensional reconstruction of electron microscopy images, the methodology is based on the methods of X-ray crystallography and in earlier stages it required ordered samples, such as 2D crystals or helical arrays (93). The improvement in resolution in recent years have taken image resolutions from sub-10 Å resolution being significant 10 years ago to current times where sub-5 Å resolution is considered a common accomplishment. This method works best with biological molecules, especially those that are larger than 500 kDa, and may therefore be difficult to resolve with X-ray crystallography or NMR spectroscopy. Although the resolutions achieved using this method have improved they are still subpar when used alone compared to that achieved by X-ray crystallography. The use of both methods in conjunction can however be highly effective and achieve a potentially better result than using only one of the methods on its own (94). In the PDB database there are 2110 entries as of May 2018 that have been attributed to determination by electro microscopy, this is the smallest category be a large margin (92), but with further improved resolution there is promising potential that a larger portion of structures may in the future be determined through electro microscopy (94).

If a project aims to study a protein which has yet to have its structure determined there is also a computational method that can be used to build a model of the target protein. This method is called homology modeling, and involves the construction of a computational model based on an already resolved protein that is homologous to the target protein and the amino acid sequence of the target protein. This creates a model of the target protein that should to a degree have the same or similar structure to the target protein, how accurate the model is depends on the sequence similarity with the homologous protein that was used as a scaffold for the model (95). The determination of structure through experimental methods is still a complex process, especially for membrane bound proteins. This makes homology modeling a key *in silico* tool for obtaining structural information on unresolved proteins.

When a computational model has been acquired, whether it has been resolved from the target protein or if there has been created a homology model, the structure needs to assume a low-energy conformation. The need for a low-energy conformation in computational models comes from molecules spontaneously taking on low-energy conformations in nature, as would be their natural state. It is in this state the proteins are assumed to interact with and bind to their ligands in nature, making it crucial to minimize the energy of a resolved structure to create as similar an environment as possible between *in silico* and nature. For those working *in silico* there is a

computer-based method called Energy Minimization which seeks optimize the computational model towards the low-energy state the molecule is thought to assume in a natural environment (28).

1.5.5 Molecular Dynamics

One of the most important tools for the computational study of the physical basis and function of biomolecules is MD simulations (96). Biological functions are based on the interactions of molecules and many molecular interactions only occur as a consequence of macromolecular structures. This makes the study of these macromolecular biomolecule structures paramount to truly be able to decipher biological functions. Advancements in technology and improved computing power has allowed MD simulations to reach simulation times that are close to relevant within biological timeframes. This has allowed the technique to be used more effectively to understand the relationship between structure and function that biomolecules such as transporter proteins and membrane receptors possess. It also allows for the study of how these large macromolecular structures interact with smaller molecules that affect their function (81).

An MD simulation is an *in silico* simulation of a molecular system that is designed to reproduce how molecules interact as closely as possible. The molecules of the system are defined by the force fields of the MM model that the system is using. The movement of molecules within the simulation is calculated at regular intervals called frames (81). Newton's second law of motion is used to calculate new velocities and positions for atoms in each new frame, once the calculations are complete the atoms are move to their new position generating a new conformation which is subsequently added to the MD simulations trajectory (28). For the MD simulation Newton's second law of motion is solved as:

$$F_i\,(t) = m_i\,a_i\,(t)$$

Here F_i is the force that is applied to atom i at time t, m_i describes the mass of the atom i and a_i describes the acceleration of atom i at time t. With each new frame being a new t, the simulation continues for the preset number of frames with a set in-system time-interval between these frames that is part of the movement calculations applied at each frame (28, 87).

1.5.6 Docking and Scoring

For the prediction of the binding orientations and affinities of ligands in the binding sites of their macromolecular targets docking is a widely used method. Most modern docking programs make use of a semi-flexible docking approach as it is not computationally feasible to have both the ligand and receptor molecules fully flexible. In the semi-flexible approach, the smaller ligands are flexible while the macromolecule it is to be docked into is rigid (97). There is also an approach that introduces protein flexibility into the docking by performing refinement of protein side chains, and in some occasions also the protein backbone, that is in the vicinity of the docked ligands, a method called induced-fit docking (98). The increased flexibility introduced into the binding site by induced-fit docking (IFD) protocols allows for the ligands to be placed in conformations not allowed by a rigid protein and allows for finding a position in the binding site where the ligand is placed more optimally (98).

When a ligand has been docked into a protein there is a need to score the docking, a number that describes the interaction between the ligand and the protein. A scoring function can this way be used to rank a large number of orientations for a ligand in a binding site to find the ligands optimal placement, it can also be used for the prediction of an absolute binding affinity between a ligand and a protein (99). The Gibbs-Helmholtz equation is used for the calculation of the binding free energy (ΔG) of a docking pose. In the Gibbs-Helmholtz equation:

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

ΔH is the enthalpy, T is the temperature in Kelvin, ΔS is the entropy, R is the gas constant and K_i is the binding constant (87).

2. Aim of the Study

One aim of the current study is to gain a greater understanding and proficiency in the use of both computational (*in silico*) methods for studying biologic molecules as well as some laboratory (*in vitro*) methods for testing biologic molecules. The study is essentially separated into two separate but related parts. The first part is the computational study of SERT, which will itself be split into two separate parts. The second part is the laboratory work which will be focusing on identification of putative new $GABA_B$ receptor compounds as a part of a larger study in the Molecular Pharmacology and Toxicology research group. Compounds predicted as putative $GABA_B$ receptor compounds in a previous virtual screening approach (Evenseth et al. 2018, unpublished work) will be tested by *in vitro* functional studies.

The computational part of the study will be performed first, and as stated previously will be divided into two parts. Each of these parts will employ different methods in order to broaden the scope of methods used and allow for a wider understanding of computational methods to gain greater proficiency in.

The first of the parts is MD simulations, which will be performed using the recently (53) crystalized structures of the SERT protein. There will in total be performed four MD simulations, each at a length of 200 ns. The ligands to be used during the simulations are (S)-citalopram, a known inhibitor of SERT, and serotonin, which is the endogenous substrate of SERT. As (S)-citalopram is the native ligand of both crystal structures that were used for the project, as well as being a well-tolerated SSRI (13), it was chosen to be used as an inhibitor for the computational part of the project. The MD simulations will focus on the central site of the 5I71 crystal structure (53) and the allosteric site of the 5I73 crystal structure (53). For both the central and allosteric site (S)-Citalopram and serotonin will have their own simulations without the presence of another ligand in the model system.

The second part of the computational study is the docking of SERT inhibitors. These inhibitors fall into two groups themselves. The known inhibitor of SERT, which is (S)-citalopram, and inhibitors found in an earlier virtual ligand screening approach of SERT using a homology model. The inhibitors from the previous homology model study have been experimentally verified as functional inhibitors of SERT using [^{3}H]-citalopram as a radioligand (100). The X-ray structure of SERT indicates that (S)-citalopram is capable of binding to both the central site and the allosteric site of SERT. It is therefore not clear if the identified inhibitors bind to the central, allosteric or both sites on SERT.

The specific goals of the computational parts of the study are as following:

- Gain a greater understanding of and higher proficiency in the use of computational methods used in the study and analysis of biological molecules.
- Elucidate the molecular mechanism of action of (S)-citalopram and serotonin by studying the interactions with surrounding amino acids, ions and water molecules in both the central and allosteric site over the course of a 200 ns MD simulation.
- Compare the results of the induced fit docking of inhibitors from the previous experimental verification with the results of the MD simulation with a focus on ligand-protein interactions in both binding sites.
- Attempt to elucidate a correlation between docking score in the crystal structure with the previous results of the experimental verification of the inhibitors (100).

The laboratory part of the study will be performed after the computational part. This part will focus on the acquisition and practice of laboratory techniques during *in vitro* binding studies using a functional cAMP assay in order to verify predicted hits from a virtual screening approach.

The goals of the laboratory parts of the study are as following:

- Gain a greater understanding of and higher proficiency in the use of laboratory techniques.
- Experimental verification of predicted hits from a previous *in silico* virtual screening approach in order to identify new orthosteric $GABA_B$ receptor compounds.

3. Methods

3.1 Computational Work

3.1.1 Software and Databases

3.1.1.1 Software

Schrödinger Software Release 2017-2

The 2017-2 version of the Schrodinger software package is a collection of tools designed for molecular modeling, drug design and the study of molecular structures.

The small molecule discovery suite put together by Schrodinger is a powerful tool in the study of complex biological proteins and their interactions. The suite possesses a wide array of tools that allow many different approaches in the study of the target of interest. Among the suites capabilities are the visualization of the molecules ranging from single atoms all the way to complex proteins, the ability to analyze the structure of a target and its binding modes, numerous utilities for ligand structures, the ability to simulate the targets motion in an *in silico* environment and tools for analyzing the results of the targets simulated motion. The tools that were used during this study include the unified interface Maestro (101), the Protein Preparation Wizard (102, 103), the protein structure prediction tool Prime (104), the ligand preparation module LigPrep (105), the Residue Scanning module (106), the Desmond molecular dynamics system module (107), the Simulation Interactions Diagram (107), the Event Analysis module (107) and the Induced Fit Docking protocol (108, 109).

3.1.1.2 Databases

The Protein Data Bank

The Protein Data Bank (PDB) (92) archive (https://www.rcsb.org/) is the universal repository of data on the 3D structure of large biological molecules, this includes proteins and nucleic acids. There are several methodologies used in acquiring the structures that are submitted to the PDB archive, these are X-ray crystallography, NMR spectroscopy and cryoelectron microscopy. Most scientific journals requiring that any solved structures be submitted to the PDB archive, making it an extensive and key resource in the field of structural biology.

The Orientations of Proteins in Membranes Database

The Orientations of Proteins in Membranes (OPM) (110) Database (http://opm.phar.umich.edu/) provides spatial orientations for membrane proteins from the PDB database with respect to the hydrocarbon core of the lipid bilayer. The spatial orientation provided for the available proteins has the protein positioned in the lipid bilayer, by having the transfer energy from water to the membrane minimized it is possible to adjust the thickness of the lipid bilayer. The OPM database contains every unique experimental structure of transmembrane proteins (110).

The Universal Protein Resource

The Universal Protein Resource (UniProt) (111) (www.uniprot.org) contains a comprehensive collection of the worlds protein sequences and annotation data. Many of these entries are obtained from genome sequencing projects. The database contains extensive data on the structure and function of registered proteins. The information the database contains is acquired from research literature, with most journals requiring that any relevant data be submitted to the database. The UniProt database is a collaborative work between the Protein Information Resource, European Bioinformatics Institute, and the Swiss Institute of Bioinformatics (111).

ChEMBL

The ChEMBL database (https://www.ebi.ac.uk/chembl) is a large-scale open database of bioactivity. The majority of the databases bioactivity data content originates from medicinal chemistry literature. This data has a myriad of applications, among them allowing researchers to identify compounds known to interact with potential therapeutic targets, investigating compounds with similar features and the identification of potential off-target effects for specific chemotypes (112).

3.1.2 Computational Methodology

For the computational part of the study the work was carried out, generally, in accordance with the flowchart below (Figure 6). Exceptions to the methodology presented in the flowchart are mostly in regards to some steps having to be repeated for verification, this mostly relates to the docking steps.
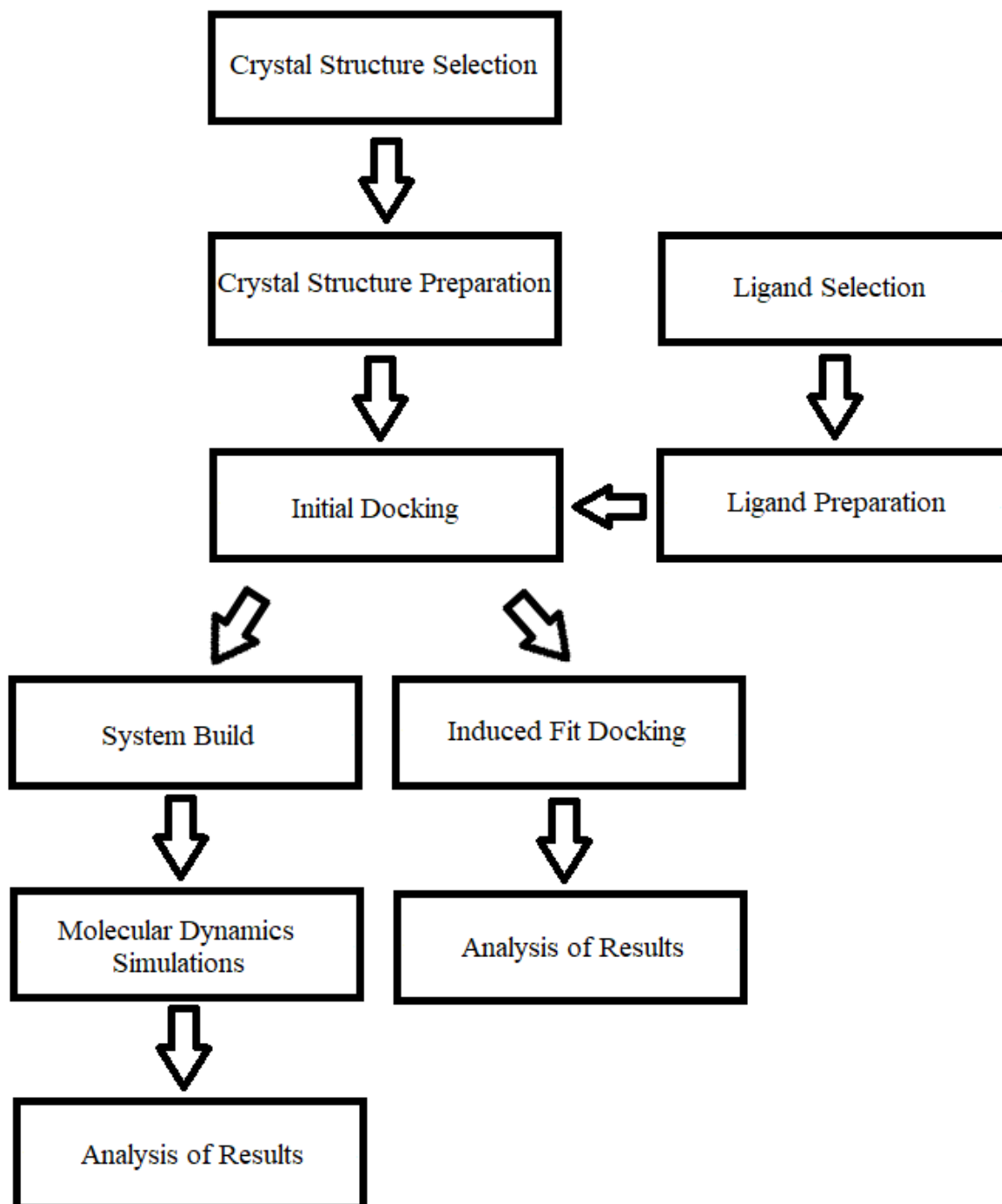

Figure 6: The figure displays a flowchart of the computational methodology that was employed in the completion of this project.

3.1.2.1 Crystal Structure Selection

For the computational components of the study it was necessary to have the structure of SERT as an *in silico* model. Previous studies (54, 113, 114) have used homology models of SERT for this, but recently a number of models were rendered through X-ray crystallography (53). This allowed for the current study to select a rendered crystal structure to be used for MD simulations and docking studies. The structures that were selected were used for both the MD simulations as well as for the docking study, the primary criteria for their selection was therefore their resolution and that they had the inhibitor (S)-citalopram bound as their ligand. Of the two structures that were selected one had (S)-citalopram bound to only its central site and a resolution of 3.15 Å (PDB ID 5I71) while the second selected structure had (S)-citalopram bound to both the central and allosteric sites and had a resolution of 3.24 Å (PDB ID 5I73). The use of the same ligand, (S)-citalopram, ensured a greater degree of consistency between the structural conformation of the structures, a feature intended to give less deviation in the results of the docking study.

3.1.2.2 Crystal Structure Preparation

The preparation of the rendered crystal structures for the use in computational studies started with the use of the Residue Scanner module (106) to reverse mutations made to the amino acid sequence of SERT to enable the rendering of the structure through X-ray crystallography. The mutations introduced to amino acid 110, 291 and 439 to increase thermostability were reversed. In addition the surface exposed cysteines, amino acids 554 and 580, were mutated to alanine in the crystal structure. These mutations were also reversed back to cysteine during the crystal structure preparation. With the mutations reversed the structures were processed with the Protein Preparation Wizard (102). The first step is the preprocessing of the structure, which was carried out with the default settings apart the generation of het states using Epik (115) had the pH set to 7.0 +/- 0.2. In addition to the default settings, Prime (116) was also used to fill in any missing side chains or missing loops and the termini were also capped. Once preprocessing was completed the preparation of the protein structure moved on to the review and modify stage. Here the workspace was analyzed and any additions to the crystal structure for the sake of the crystallization process were removed. All water molecules in the system were also removed. With the review and modification stage completed the final stage of the protein preparation was the refine stage. The first step here was the H-bond assignment, which was

carried out using the default settings. The second step was the restrained minimization, which was also carried out with the default settings and the default OPLS3 force field.

3.1.2.3 Ligand Selection

For the computational part of the study ligands had to be selected for both the MD simulations as well as for the docking study. For the MD simulations, as the focus here was studying the interactions of the substrate and an inhibitor with the amino acids of the central and allosteric binding sites, serotonin and (S)-citalopram were selected. Serotonin is the substrate of SERT (44) and therefore seeing how it interacts with the amino acids of SERT while it is in its outward-open conformation is of interest. (S)-citalopram is an SSRI, an inhibitor of SERT (117) and seeing what amino acids it interacts with to achieve inhibition and comparing this with what serotonin interacts with is the focus of the MD portion of the computational study. (S)-citalopram was also the ligand of the selected crystal structures and the structures are therefore not expected to undergo any significant structural changes over the course of the MD simulations. It is however interesting to identify the structural consequences upon the SERT structure when it binds a substrate or an inhibitor.

For the docking study ligands were selected from a previous study (100), which had created a ranked list of compounds that had been initially run through virtual screening using a homology model of SERT before being experimentally tested. The ligands were arranged into clusters based on structure and the compounds with the best experimental values in a competitive assay against [3H]-citalopram were selected from each of the clusters for use in the docking study. When selecting ligands the experimental values that were considered were the percent of inhibition a compound achieved, with higher being better, and also a low K_i. For most clusters the highest ranked compound was selected, any clusters that did not have compounds with a K_i of less than 1000 nM were not included in the docking study. A total of 15 ligands were selected from 13 clusters for the docking study, the docking study also included (S)-citalopram as a known inhibitor of SERT.

3.1.2.4 Ligand Preparation

The ligands that were selected for both the MD simulations and the docking study were run through Schrodingers LigPrep module (105). During the ligand preparation using the LigPrep module the default settings were used apart from setting the target pH of possible states to be generated to 7.0 +/- 0.2.

3.1.2.5 Induced Fit Docking

In this project Schrödinger Induced Fit Docking protocol (109) was used to dock the selected ligand, serotonin, into the prepared crystal structures of SERT for both the central and allosteric binding sites. Induced Fit Docking was also used in the study of the experimentally verified compounds that had been selected for the Docking part of the project.

For the Induced Fit Docking that was carried out on the prepared ligands the default OPLS3 force field was used with the protocol set to extended sampling, which would generate up to 80 poses for each ligand. The ligand binding pocket in the receptor was defined by selecting the ligand, (S)-citalopram, that was in its default position in the 5I71 structure for the central site, and 5I73 structure for the allosteric site. All other settings were set to default for Induced Fit Docking jobs.

3.1.2.6 Molecular Dynamics Simulation

Four 200 ns MD simulations of SERT were carried out using Desmond (107). The MD simulations were performed with the PDB ID 5I71 crystal structure for the central site simulations and the PDB ID 5I73 crystal structure for the allosteric site simulations. Of the four simulations two were of the central site and two were of the allosteric site, each site had one simulation with serotonin and another with (S)-citalopram.

For the (S)-citalopram simulations the placement of the ligand was left as it was in the crystal structure. The ligand was removed for ligand preparation and was later merged back into the crystal structure in the same position as before the ligand preparation, this was done for both the (S)-citalopram simulations in the central site and the allosteric site. For the simulations using serotonin the ligand was first prepared, before being placed into each respective binding pocket. For the central binding pocket it was placed by Induced Fit docking, but for the

allosteric site this method failed and it was instead merged into the allosteric site in the same position as the best scoring serotonin docking pose that was generated for the allosteric binding pocket using Induced Fit docking. The reason the SP docking failed is not clear, but the same result was achieved by merging the entries. All structures were minimized once more after having their ligands reinserted.

With the protein-ligand complex ready the next step was the creation of the model system, this was done in Schrodinger's System Builder module. The OPLS3 force field was used for the amino acid interactions and the water model simulation used the TIP3P method. As SERT is a membrane protein an important part of the model system setup is the addition of the membrane. For the addition of the membrane the structure was prealigined on a template that was acquired from the OPM database, the membrane itself was setup with the POPE (310 K) membrane model. With the membrane setup settings completed the boundary conditions of the model system were set. Here a 10 Å × 10 Å × 10 Å water box with orthorhombic dimensions was used, as is the default setting in the system builder. The final stage of the model system construction was the balancing of the net charge with Cl^- ions to neutralize the system, followed with the addition of a salt concentration of 0.15 M. For the added salt concentration Na^+ was used as the positive ion and Cl^- was used as the negative ion.

Before initiating the MD simulations each of the model systems were run through Schrodingers relaxation protocol for membrane-bound proteins using the command line. The script for this was run before the MD simulation, which is why the default option to relax the model system before beginning the simulation was deselected.

For the MD simulation the system was subjected to 310 K for 200 ns, 200 000 ps, using the default NPT ensemble class. The recording interval was set to 50 ps. The pressure was left at the default 1.0132 bar and the option to relax the model system before the simulation was deselected as mentioned above.

3.1.2.7 Analysis of MD trajectories

The MD simulations were analyzed with the Simulation Interactions Diagram (SID) module (107) and also the Event Analysis module (107). The data provided by the SID module was included in the results, among this was the Protein-Ligand RMSD, the Protein RMSF and Protein-Ligand contact analysis. The Event Analysis module was used to look for differences in the behavior of amino acids in the central site by measuring the varying distances over the course of the simulations between opposing amino acids known to be important (TYR95, ASP98, ILE172, ASN177, PHE341, SER438) for the function of SERT (118). The measurements data was then plotted using the data presentation options of the Event Analysis module.

3.1.2.8 Analysis of docking results for experimentally verified SERT antagonists

The docking portion of the study gave docking scores for every compound for each binding site. This data could then be compared between binding sites, with the experimental Ki values of the compounds, and with the degree of inhibition that the compounds exhibited. These data points were charted to see if there was any correlation between them. The amino acids each ligand interacted with was also recorded for comparison with the MD simulation data.

3.2 Experimental testing of potential GABA$_B$ receptor antagonists

3.2.1 Materials

Materials purchased from DiscoveRx:

- cAMP Hunter™ CHO-K1 GABBR1+GABBR2 Gi Cell Line (Cat.# 95-0165C2)
    - The CHO-K1 Wild Type (WT) cell line was provided by the Tumor biology research group of UiT, The Arctic University of Norway.
- AssayComplete™ Revive CHO-K1 Medium (Cat.# 92-0016RM2S)
- AssayComplete™ CHO-K1 Cell Culture Kit 35 (Cat.# 92-0018G2R2)
- AssayComplete™ Cell Detachment Reagent (Cat.# 92-0009)
- AssayComplete™ Freezing Reagent F2 (cat.# 92-5102FR)
- HitHunter® cAMP Assay for Small Molecules (cat.# 90-0075SM2)
- White clear bottom, tissue culture treated 384-well (cat.# 92-0013)

Chemicals purchased from Sigma-Aldrich:

- CaCl2 (Cat.# C7902)
- HEPES (Cat.# H3375)
- MgCl2·6H2O (Cat.# M9272)
- KCl (Cat.# 746436)
- NaCl (Cat.# 746398)
- D-(+)-Glucose (Cat.# G7021)
- NaOH (Cat.# 30620)
- GABA (Cat.# A5835)
- DMSO (Cat.# 472301)
- Water soluble forskolin NKH477 (Cat.# N3290).

The test compounds used were purchased from Molport.

3.2.2 Method

3.2.2.1 Cell culture and harvesting

A cell culture was started with a CHO-K1 cell line stably co-expressing the human $GABA_{B(1b)}$ and $GABA_{B2}$ receptor subunits (DiscoveRx, Cat.# 95-0165C2) using the cell culture reagents and guideline from DiscoverX.

A vial of cells ($GABA_B$ or CHO-K1 WT) that had been cryopreserved in the gas phase of a liquid nitrogen tank was quickly thawed using a 37°C water bath. The cells were thawed under sterile conditions until only a small ice crystal remained, the majority of the cell pellet was thawed at this point. The thawed cells were then immediately transferred from the cryopreservation vial over into a sterile 15 ml tube containing 7 ml of pre-warmed culture medium. For the WT cells there was no antibiotics added to the culture medium. The next step was the centrifugation of the cells at 300 G for 4 minutes. The supernatant was then aspirated and the cell pellet immediately resuspended in 10 ml of pre-warmed culture medium for the $GABA_B$ cells and 15 ml pre-warmed culture medium for the WT cells (no antibiotics for WT). The following step was the transfer of 5 ml of ($GABA_B$) cell suspension into two T25 flasks, one intended as a backup. The T25 flasks were then marked and grown at 37°C, 5% CO_2. For the WT cells the 15 ml of cell suspension was transferred to a T75 flask, marked and grown at 37°C, 5% CO_2. The cells reached >70% confluency after about 24 hours for the WT cells (T75) and between 36-48 hours for the $GABA_B$ cells (2x T25). The WT cells could be harvested for assay after 24 hours, the $GABA_B$ cells had to first be grown in two T25 flasks, then transferred to two T75 flasks (one for each T25) and grown again.

With the ($GABA_B$) cells at >70% confluency the medium in the culture flasks was aspirated and the cells were quickly washed with 37°C Ca^{+2}/Mg^{+2} free Phosphate-buffered saline (PBS) 3 times. Once the final wash of 37°C PBS was aspirated, 0.8 ml of AssayComplete™ Cell Detachment Reagent was added to the T25 flask and it was then incubated at 37°C for 2 min, at which point the cells are detached. The cells were then collected using 20 ml of 37°C PBS into a 50 ml conical tube, which was subsequently centrifuged for 4 minutes at 300 G. Once the centrifugation was complete the supernatant was aspirated and the cell pellet was re-suspended using 15 ml of culture medium. The cell suspension was then transferred over to a T75 flask for continued growth. The second T25 flask underwent the same procedure.

The ($GABA_B$) cells were harvested after 2 days of growth for the cAMP functional assay, when the cells reached a confluency of about 75%. The WT cells were harvested from their T75 flask

after 1 day of growth for cAMP assay, at which point they had reached about 70% confluency. On the day of the assay culture medium was first aspirated from the T75 flask and the cells were quickly washed with 37°C PBS 3 times in quick succession. Then, 1.6 ml of AssayComplete™ Cell Detachment Reagent was quickly added to the T75 flask and incubated at 37°C for 2 min. The cells were then collected using 40 ml of 37°C PBS into a 50 ml conical tube, before being centrifuged for 4 minutes at 300 G. Once the centrifugation was complete the supernatant was aspirated and the cell pellet was re-suspended using 20 ml assay buffer HBSS (HEPES-buffered hank's balanced salt solution: 2 mM $CaCl_2$, 5 mM D-Glucose, 10 mM HEPES, 3 mM KCl, 1 mM $MgCl_2$, 136 mM NaCl, pH adjusted to 7.4 at 18°C using 1 M NaOH). The cell suspension was centrifuged again for 4 minutes at 300 G, with the supernatant subsequently being aspirated and the cell pellet re-suspended using 3 ml of the assay buffer.

For determining the cell concentration, 40 µl of fresh cell suspension was mixed with 40 µl of 0.4% trypan triple blue solution. 10 µl of this mixture was then loaded into two Countess® chamber slides and counted using a Countess® automated cell counter. The cells were then quickly diluted to a concentration of 0.5 million cells/ml and incubated at the assay volume of 30 µl/0.2 ml tube in an assay plate at 18°C for 1 hour before the test compounds were added (Figure 7).
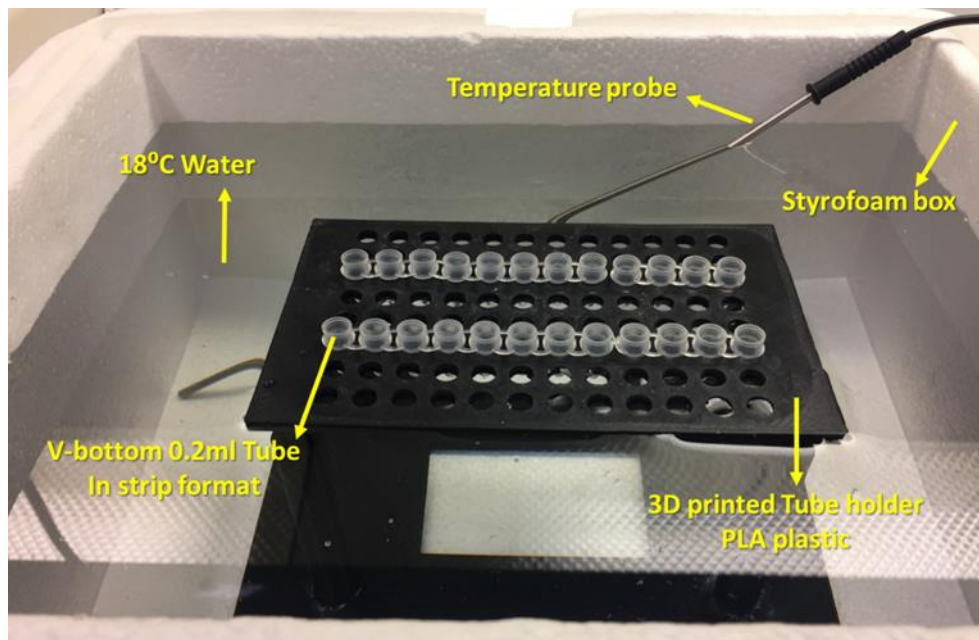

Figure 7: The cAMP assay plate and assay environment. Image by Imin Wushur.

3.2.2.2 Test compound preparation and testing on cells

The test compounds dissolved in 100% DMSO with a stock concentration of 10 mM transferred to aliquots and stored at -20°C freezer. The solubility of the test compounds was tested in assay buffer at a concentration of 60 µM. Five of the test compounds were found to have good and stable solubility, the other five test compounds were not soluble at 60 µM or 30 µM concentration, but were soluble in 5 mM D-glucose solution at a concentration of 30 µM. The final concentration of the test compounds when used in the assay is therefore 10 µM and 5 µM.

On the day of the assay, the frozen test compounds were retrieved from the freezer and placed at room temperature to thaw. A mixture of forskolin, GABA (for the $GABA_B$ expressing cells) and test compounds was prepared in assay buffer at the desired concentrations. When the cells had finished their 1 hour of pre-incubation at 18°C the compound mixture was added, 15 µl of it, to each of the 0.2 ml tubes (Figure 7). For each of the components of the mixture the final assay concentration was: 1µM GABA (for $GABA_B$ expressing cells), 30 µM Forskolin and 10 µM or 5 µM test compound (depending on solubility of compound). For the controls only 1 µM GABA (for $GABA_B$ expressing cells) and 30 µM Forskolin mixture was used. The reaction was allowed 25 minutes before the cAMP reagents were added.

3.2.2.3 Measurement of cAMP levels

Once the reaction had been active for 25 minutes at 18°C the cAMP levels were measured through the use of the HitHunter® cAMP Assay for Small Molecules (DiscoveRx) by following the instructions provided by the manufacturer. The cAMP antibody reagent and the cAMP working detection solution were both added to all the wells in rapid succession, which was followed by an incubation period of 1 hour in the dark at 18°C. After the 1 hour incubation period the cAMP solution A was finally added to all the wells and the wells were left for a 12-14 hour incubation at room temperature in the dark.

After the 12-14 hour incubation the assay mixture was transferred in the dark to a white clear bottom tissue culture treated 384-well plate. For each well 50 µl was transferred, each compound was given triplicate wells. The cAMP signal was then immediately read on ClarioStar® plate reader (BMG LABTECH) for the luminescence readout. The results of the assay were then analyzed using the GraphPad Prism software (GraphPad Software Inc., San Diego, CA).

4. Results

4.1 Molecular Dynamics

The four MD simulations (Figure 8) that were run for this study provided a very significant amount of data. The choice was made for the data to be analyzed from two points, these were the structural dynamics of the protein over the course of the simulation and the interactions between the protein and the ligand over the course of the simulation. As the MD simulations were rather short, only 200 ns, no substantial changes were expected in the proteins overall structure, with the focus instead being on stability of the protein model and the movement of amino acids throughout the simulation. For the protein-ligand interactions the focus was on the interactions between the ligand and the various amino acids in the binding pockets. Whether these interactions were consistent with what had been found to be key amino acids in the central site or not, and also investigating which amino acids in the allosteric site were favored by the ligands.

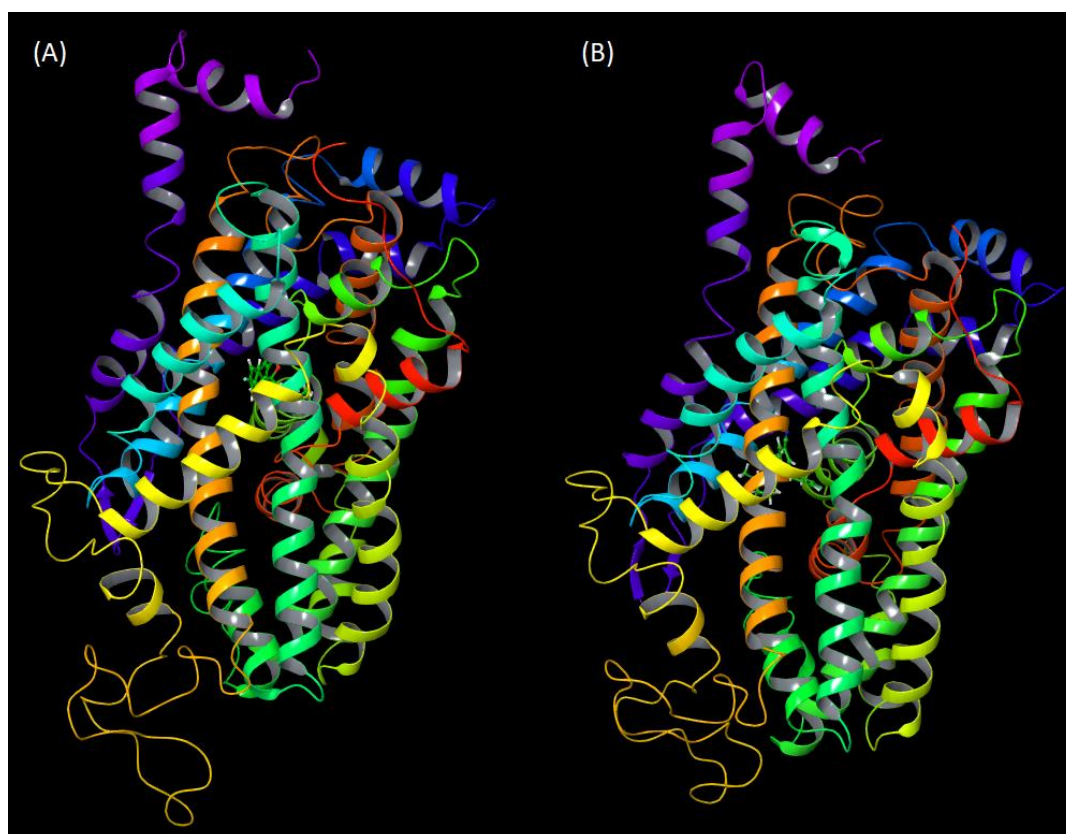Figure 8: The above figure shows the ribbon diagrams of the SERT crystal structures used for the MD simulations. (A) is the PDB ID 5I71 crystal structure with (S)-citalopram in the central site, was used for all MD simulations with a ligand in the central site. (B) is the PDB ID 5I73 crystal structure with (S)-citalopram in the allosteric site, was used for all MD simulations with a ligand in the allosteric site.

4.1.1 Protein Structure Analysis

4.1.1.1 Protein-Ligand RMSD

In order to measure how significant a change there has been in the placement of a selection of atoms over the course of an MD simulation compared with a reference frame the Root Mean Square Deviation (RMSD) is used. The RMSD is calculated by measuring the average change in displacement for the atom selection in a particular frame with respect to a reference frame. To find the RMSD of a MD simulation timeframe this must be done for every frame in the trajectory. The Schrodinger Simulation Interaction Diagram module by default generates a Protein RMSD and a Ligand RMSD when analyzing and MD simulation. The Protein RMSD allows for a deeper understanding of a proteins structural conformation throughout the MD simulation and can also give indications if the simulation is equilibrated or not. The Ligand RMSD shows an indication of the degree of stability the ligand possesses with respect to the protein and its binding pocket. If the Ligand RMSD has significantly higher values than that of the Protein RMSD then this is an indication that the ligand has moved away from its initial binding site (107).
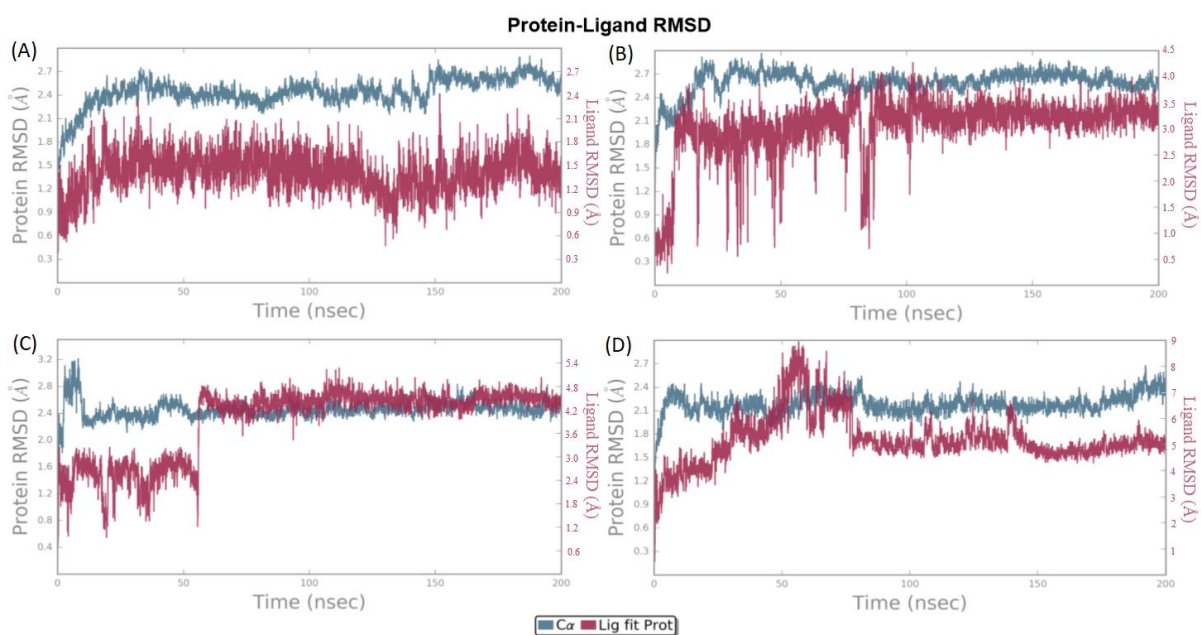

Figure 9: The above figure displays the plots showing the RMSD evolution of both SERT (left Y-axis, Cα) and the designated ligand for each respective simulation (right Y-axis, Lig fit Prot) compared with their reference position at the initiation of each of the respective simulations, the distance values are in angstrom (Å). The X-axis for each of the plots shows the timeframe of the simulation in ns. The figure is organized by binding pocket, the two top most plots are for the central site and are for (A) (S)-citalopram and (B) serotonin. The two bottommost plots are for the allosteric site and are for (C) (S)-citalopram and (D) serotonin.

For the Protein RMSD all the protein frames are aligned on the reference frame backbone, with the RMSD then being calculated based on the atom selection. The Ligand RMSD, Lig fit Prot, displays the RMSD of the ligand at the point where the protein-ligand complex is initially aligned on the protein backbone of the reference complex and then the RMSD of the ligand heavy atoms is measured (107).

As can be seen from Figure 9 the Protein RMSD and Ligand RMSD was calculated for each of the four simulations that were performed for the project. The (A) section of the figure gives an overview for (S)-citalopram in the central site, (S)-citalopram is the native ligand of the crystal structure and both the protein and the ligand remained quite stable for the duration of the simulation. For the (B) section of the simulation, which had serotonin in the central site we see that the protein is quite stable throughout the simulation. The ligand also remains mostly stable with a few spikes in movement. Comparing (S)-citalopram and serotonin in the central site shows that while both were stable for the duration of the simulation it is clear that serotonin moved substantially more from its reference frame in the beginning of the simulation than (S)-citalopram did. For the allosteric simulation with (S)-citalopram (C), which is again the native ligand of the crystal structure, the protein remained stable throughout the simulation. The ligand however underwent numerous small spikes throughout the first 60 ns before increasing significantly and subsequently becoming stable for the remainder of the simulation. The allosteric simulation with serotonin (D) saw the protein being stable for the duration of the simulation, while the ligand underwent numerous spikes throughout the simulation before reaching apparent stability at 150 ns. The (D) simulation was the one that saw the most activity with the ligand, where it moved around significantly throughout the simulation. The ligand did not however leave the allosteric binding pocket during the simulation, despite the relatively large shift in Ligand RMSD.

4.1.1.2 Protein RMSF

The movement of amino acids along the protein chain can be characterized by the Root Mean Square Fluctuation (RMSF).
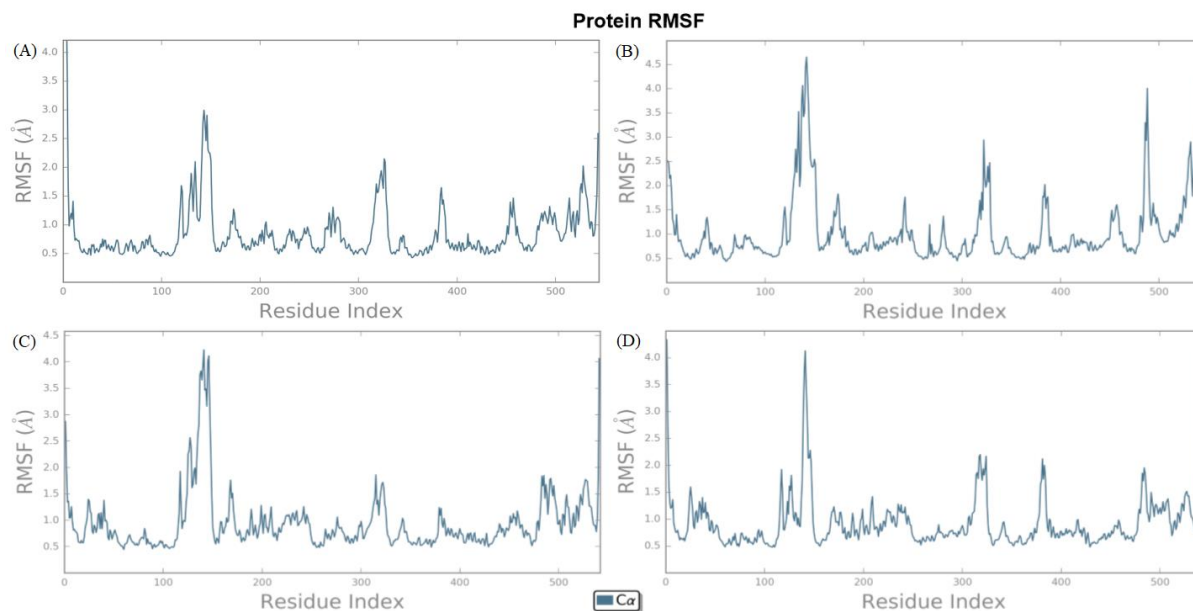

Figure 10: The above figure displays the plots showing the Protein RMSF evolution of all four MD simulations. The X-axis for each of the plots display the residue index, i.e. each point along the X-axis represents an amino acid belonging to SERT ordered after residue number. The Y-axis for each of the plots displays degree of fluctuation by each amino acid in angstrom, functionally characterizing local changes along the protein chain. The figure is organized by binding pocket, the two top most plots are for the simulations with the ligand in the central site and are for (A) (S)-citalopram and (B) serotonin. The two bottommost plots are for the simulations with the ligand in the allosteric site and are for (C) (S)-citalopram and (D) serotonin.

The plots in Figure 10 show the amount of fluctuation individual amino acids underwent over the course of the simulation. The peaks in the plot are indicative of areas that underwent the most fluctuations, the degree of the fluctuation can be seen on the Y-axis of Figure 10. The amino acids that make up the secondary structure elements of the protein, such as α helices and β strands, normally have a higher degree of rigidity than parts of the protein that are not part of a secondary structure. The secondary structure parts of the protein are therefore expected to fluctuate less than the proteins loop regions (107).

When examining the plots in Figure 10 with the amino acids known to be in the central and allosteric binding sites (Figure 11, Table 2, Table 3) in mind it also becomes apparent that for the most part these amino acids have fluctuations that fall on the lower end of the spectrum.

The amino acids that interact with the ligand but still undergo significant fluctuations are the ones in the range above 480 residues, although there are dips down into lower fluctuation degrees in some of the simulations. For the late five hundred amino acids it seems the fluctuations are especially extreme in the MD simulation with serotonin in the central site (Figure 10, (B)). For the other simulations, and the simulations in general most amino acid fluctuations are fairly similar between simulations with only a few variations. There also seems to be some variations in fluctuation from the difference in crystal structure, although this may also potentially be caused by these models having no ligand in the central site.

4.1.2 Protein-Ligand Interactions Analysis

4.1.2.1 Protein-Ligand Contacts – Binding Type

The interactions between the protein and the ligand can be monitored throughout the entirety of the MD simulation. These interactions can then be categorized as seen in the figure below (Figure 11). The stacked bar charts are normalized over the course of the trajectory. An example of this in practice would be that a value of 0.5 means that the specific interaction was maintained for 50% of the simulation time. It is possible for the values to be greater than 1.0 by virtue of some protein residues making multiple contacts of the same subtype with the ligand (107).

Here we look at what types of interactions the ligand of each simulation had with the amino acids of the binding site and what percentage of the simulation these interactions were maintained for.
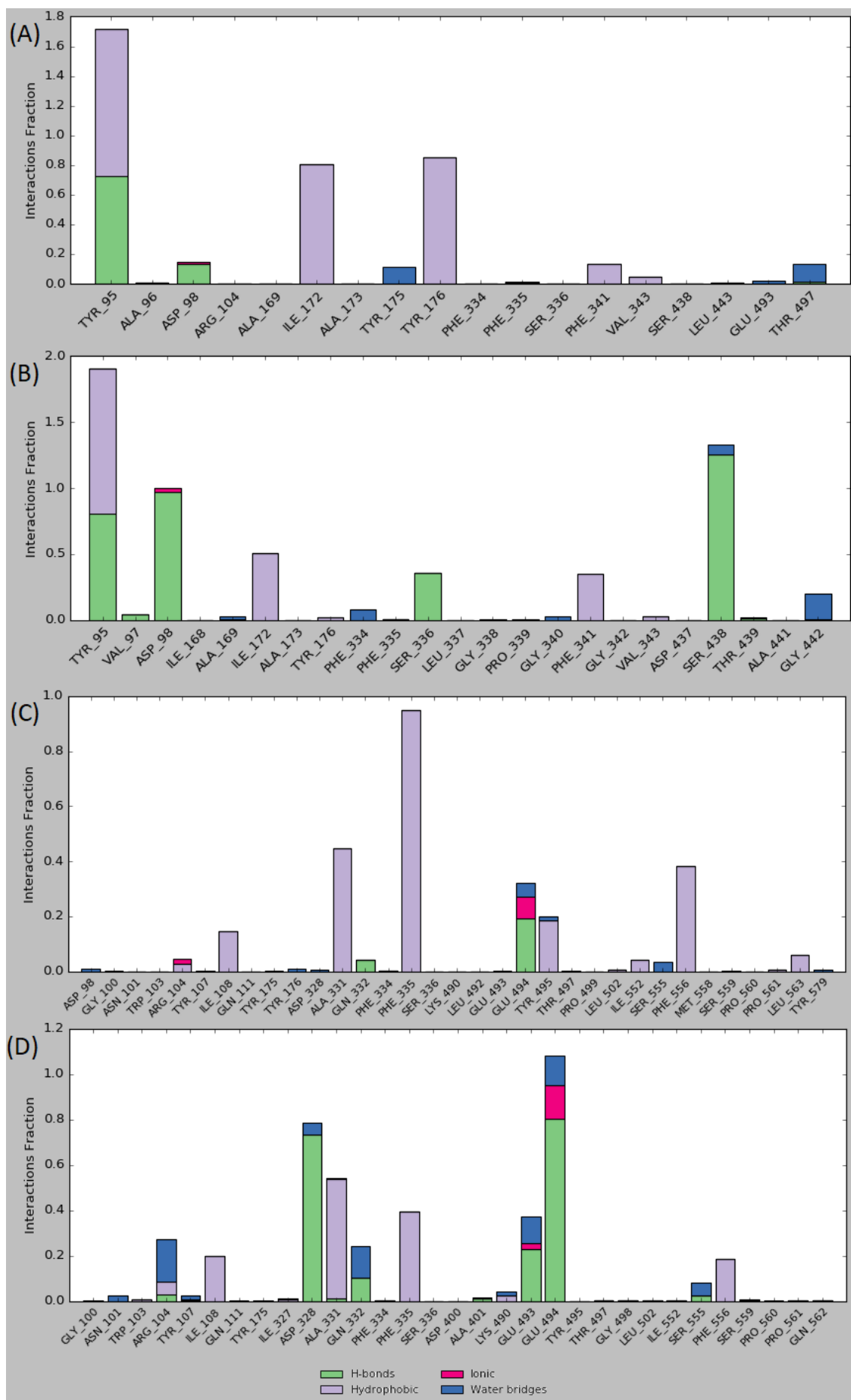
Figure 11: The figure displays the amino acids that the ligands interacted with over the course of each of the four the simulations. The figure is organized by binding pocket, the two top most bar charts are for the central site and are for (A) (S)-citalopram and (B) serotonin. The two bottommost bar charts are for the allosteric site and are for (C) (S)-citalopram and (D) serotonin. Each of the ligands was placed in the respective binding site of SERT for that simulation from the onset of the simulation, each entire system having already been run though a relaxation protocol script for membrane bound proteins before initiating their MD simulation. The Y-axis of each bar chart displays a numerical value equal to the percentage of the time that residue interacted with the ligand over the course of the simulation, values above 1.0 (100%) are possible if the amino acid had multiple contacts with the ligand. The X-axis of each bar chart displays the names of the interacting amino acids.

As can be seen from Figure 11 there were numerous connections between the amino acids of the binding pockets and the ligand. It is also quite apparent that the ligands, both (S)-citalopram and serotonin, interacted with significantly more amino acids when in the allosteric binding pocket than when in the central binding site. To be precise (S)-citalopram interacted with 18 amino acids in the central binding site and 33 amino acids in the allosteric binding site. Which is slightly lower than for serotonin in the central site, with serotonin interacting with 23 amino acids in the central binding pocket. For the allosteric binding pocket serotonin interacted with fewer amino acids, interacting with 31 over the course of the simulation.

The amino acids that each ligand had a high degree of interaction with over the course of the simulations were for the most part within 3Å of the ligand at the end of the simulation, at which point the systems for each MD simulations was quite stable with only small overall fluctuations. For the central site simulation with (S)-citalopram the amino acids that showed the highest degree of interaction were TYR95, ILE172 and TYR176 (Figure 11 (A)). TYR95 had the highest interaction value, at about 1.7, which means the interaction was maintained for the majority of the simulation and also had multiple interactions simultaneously with the ligand. As can be seen from the above figure, the interactions with TYR95 were of both a hydrophobic nature and in the form of H-bonds. ILE172 and TYR176 had values of about 0.8 and 0.9 respectively and were solely hydrophobic in nature. As can be seen from Figure 12 (A) all three of these amino acids were within 3Å of (S)-citalopram at the end of the simulation.

The central site simulation with serotonin showed that the amino acids with the highest degree of interaction were TYR95, ASP98 and SER438 (Figure 11 (B)). These amino acids had interaction values of about 1 for ASP98 and SER438, with TYR95 having the highest interaction value at nearly 2.0. As can be seen the figure above, the interactions with the TYR95 amino acid were of both a hydrophobic nature and also in the form of H-bonds. The interactions

of both ASP98 and SER438 were for the most part H-bonds with the ligand, but a small portion of ASP98's interactions were ionic and a slightly larger portion of SER438's interactions were in the form of water bridges. Serotonin also had ILE172, SER336 and PHE341 which maintained a moderate degree of interaction at the 0.5-0.3 level. These interactions were of a hydrophobic nature for ILE172 and PHE341, while SER336 had H-bond interactions with the ligand. As can be seen from Figure 12 (B) both the high and moderate degree interaction value amino acids were within 3Å of serotonin at the end of the simulation.

Figure 12: The above figure displays the Schrodinger Maestro workspace containing each respective MD simulation ligand and any amino acids that where within 3Å of the ligand in the central site at the end of the 200 ns simulation. (A) is for (S)-citalopram and (B) is for serotonin. Color coding of atoms: red; oxygen, blue; nitrogen, green; ligand carbon atoms, gray; amino acid carbon atoms, white; hydrogen atoms.

The allosteric simulations showed the ligands interacting with a larger quantity of amino acids, but there were no interaction values that exceeded 1.2, unlike the central site which had an interaction value for TYR95 exceeding 1.6 with both (S)-citalopram and serotonin. For the simulation with (S)-citalopram in the allosteric site the only amino acid that showed a high degree of interaction comparable to that seen in the central site was PHE335 (Figure 11 (C)) which had an interaction score of 1.0. The interaction between PHE335 and (S)-citalopram in the allosteric site was of a hydrophobic nature, unlike the very minor interactions seen between this amino acid and the ligands in the central site, which took the form of water bridges. There were also three interactions for (S)-citalopram that would fall into the previously labeled moderate interaction category, these are ALA331, PHE556 and GLU494 in order of the degree of interaction. Both ALA331 and PHE556 had solely hydrophobic interactions with the ligand, their interaction values being approximately 0.45 and 0.39 respectively. GLU494 has an interaction value of about 0.3, making it on the low side for a moderate interaction. The interactions GLU494 had with (S)-citalopram were made up of H-bonds, water bridges and ionic interactions, making it the most varied interaction for (S)-citalopram in both the central and allosteric sites. Amino acids that had a high to moderate degree of interaction with (S)-citalopram in the allosteric site can all be seen to be within 3Å of the ligand in the system workspace at the end of the simulation (Figure 13 (C)).

The simulation of serotonin in the allosteric site showed two amino acids that had a high degree of interaction, these were GLU494 and ASP328 in order of degree of interaction. GLU494 had an interaction value of about 1.1, with the interaction types being split between H-bonds, water bridges and ionic interactions, the H-bonds being the most common by a significant margin. ASP328 had an interaction value slightly above 0.8, as with GLU494 this interaction was also mostly H-bonds, with a small portion being made up of water bridges. There are a further three amino acids that can be labeled as moderate interactions, these are ALA331, PHE335 and GLU493. ALA331 had an interaction value slightly above 0.5 with the majority of the interactions being of the hydrophobic variety and a very small portion being H-bonds with the ligand. The second of the moderate interactions is PHE335, with an interaction value of about 0.4, with the entirety of its interactions being hydrophobic. The last of the moderate interactions is GLU493 with an interaction value of about 0.35, which similar to GLU494, has its interactions split between H-bonds, water bridges and ionic interactions, with their relative quantity in that respective order. As can be seen from Figure 13 (D) the high interaction amino

acids GLU494 and ASP328 are within a 3Å range of the ligand at the end of the simulation, so are the moderate interaction amino acids ALA331 and PHE335. The moderate interaction amino acid GLU493 is however not within 3Å of the ligand at the end of the simulation.
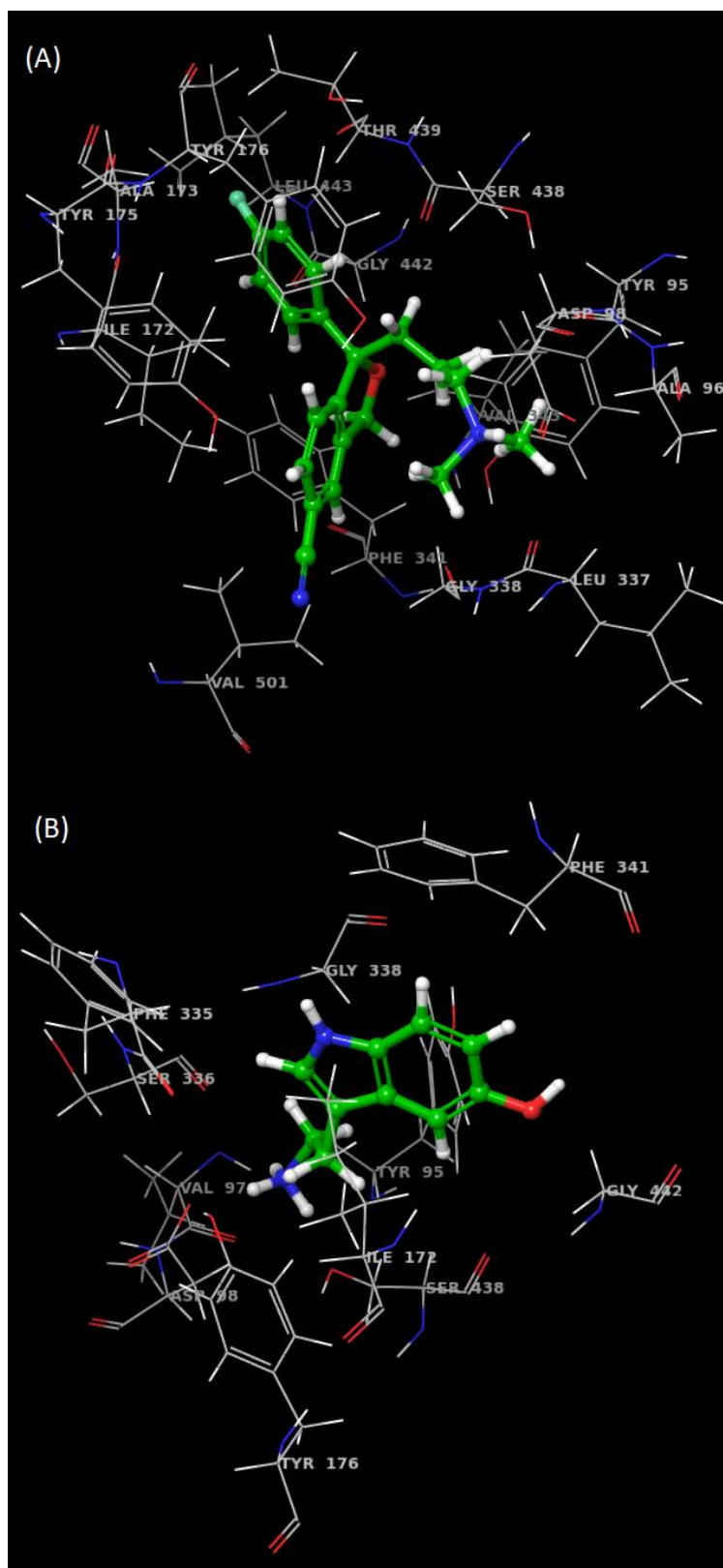

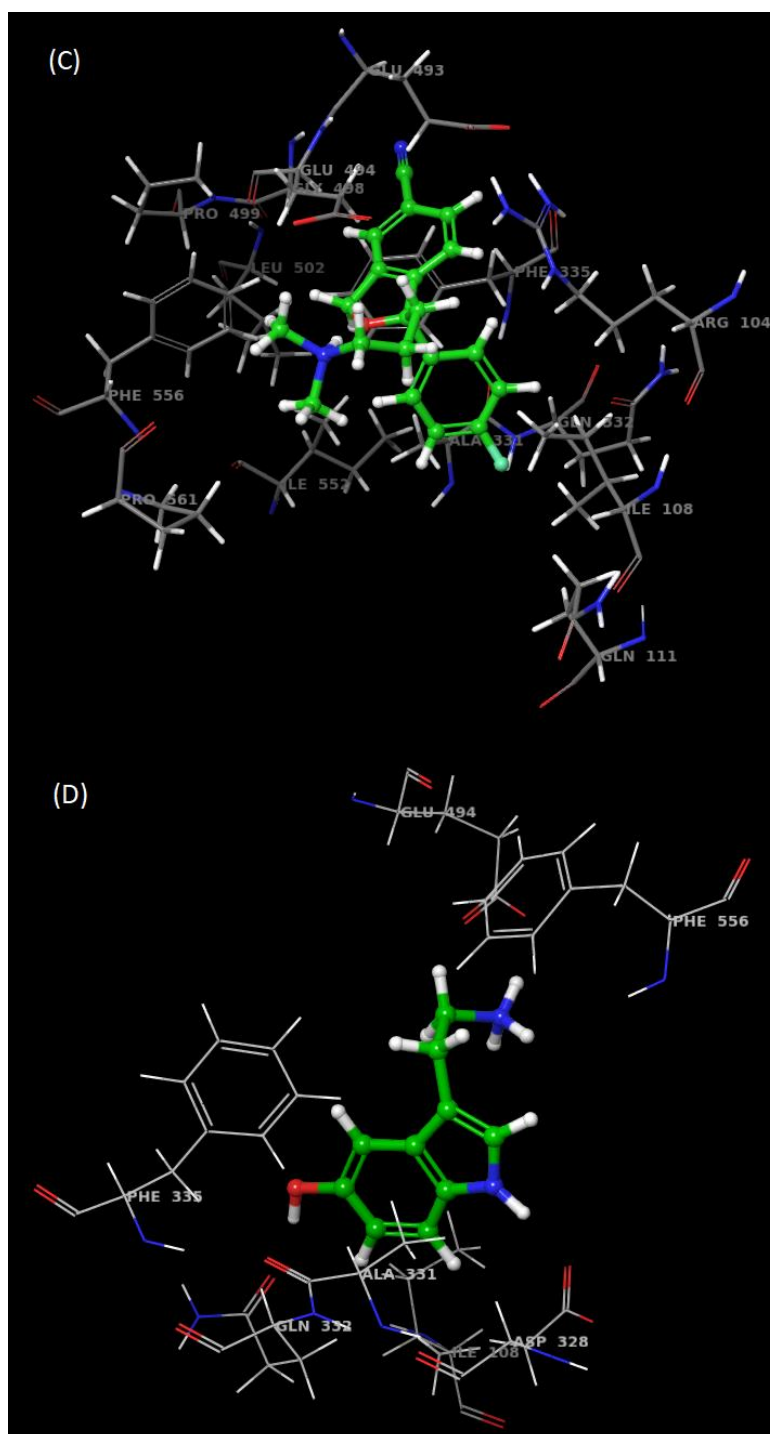Figure 13: The above figure displays the Schrodinger Maestro workspace containing each respective MD simulation ligand and any amino acids that where within 3Å of the ligand in the allosteric site at the end of the 200 ns simulation. (C) is for (S)-citalopram and (D) is for serotonin. Color coding of atoms: red; oxygen, blue; nitrogen, green; ligand carbon atoms, gray; amino acid carbon atoms, white; hydrogen atoms.

4.1.2.2 Protein-Ligand Contact – Frequency and Quantity

The figures below give a timeline representation of the interactions and contacts (H-bonds, Hydrophobic, Ionic, Water Bridges) that were presented in the 4.1.2.1 Protein-Ligand Contact – Binding Type section of the results.
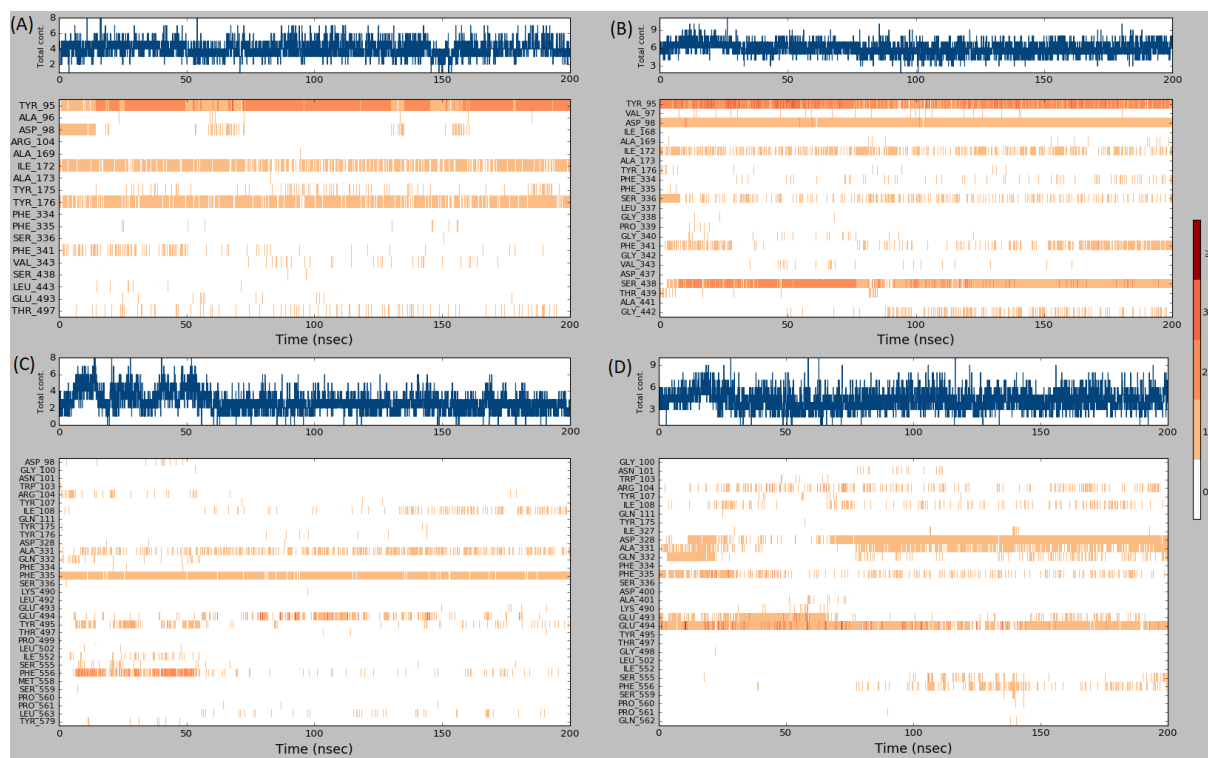

Figure 14: The above figure shows the quantity of contacts and frequency of contacts between the ligand and the protein over the course of the simulation in each of the four MD simulations that were run. The top row shows the MD simulations of the central site, with (A) (S)-citalopram being on the left and (B) serotonin being on the right. The bottom row shows the MD simulations of the allosteric site, with (C) (S)-citalopram being on the left and (D) serotonin being on the right. Each part of the figure is itself divided in two between number of contacts over time, top, and frequency of individual amino acid contacts over time, bottom. The top part of the figure displays the total number of contacts (Y-axis) between the ligand and the protein over the course of the simulation (X-axis). The bottom part of the figure displays the contacts between individual amino acids belonging to the protein and the ligand (Y-axis) over the course of the simulation (X-axis). The color of the lines indicating a contact between the amino acid and the ligand is color coded using the scale (# of contacts) on the right side of the figure. The type of contact the amino acids had with the ligand can be seen in the corresponding Protein-Ligand Binding Type figure (Figure 11).

The knowledge of the frequency of contacts between an amino acid and the ligand as well as the number of contacts between the ligand and amino acid at each registered point in time is invaluable for gaining greater insight into the specific interactions of the ligand within the binding site. The total number of interactions with the ligand also gives insight into the number

of amino acids that are holding the ligand in place and preventing it from drifting out of the binding site. From Figure 11 it was clear that TYR95 was a very important amino acid for the ligands in the central site, but from Figure 14 it became clear that not only was the interaction maintained with few gaps for the entirety of each central site simulation, but maintained with a significant degree of multiple simultaneous interactions. Interactions that were steadily maintained suggest that the interactions were quite stable, possibly holding the ligand in place. For (S)-citalopram in the central site in addition to TYR95 there were also ILE172 and TYR176, which maintained steady contact with the ligand for the majority of the simulation. While serotonin in its central site simulation had very steady contacts throughout the simulation with both ASP98 and SER438, in addition to TYR95 which was shared for both the central site simulations.

For the allosteric site the only amino acid that (S)-citalopram maintained a steady near uninterrupted connection to was PHE335, with ALA331 having a fairly steady rate of interactions over the course of the simulations as well. PHE556 and GLU494 had steady interactions with (S)-citalopram for portions of the simulation, with PHE556 being in the first 55 ns and GLU494 being between roughly 70 ns and 150 ns. PHE335 and ALA331 had the highest and second highest interaction scores respectively in Figure 11, which reflects on their steady interaction rates. With serotonin in the allosteric site there was a steady near uninterrupted interaction with GLU494, ASP328 also had a steady near uninterrupted interaction with serotonin from roughly the 75 ns mark. ALA331 had a steady interaction with serotonin from about the 85 ns point in the simulation. PHE335 which had a very strong and steady interaction with (S)-citalopram in the allosteric site had a fairly steady but sporadic interaction with serotonin, the connection being stronger in the early stages of the simulation.

Figure 14 also shows the number of interactions that each ligand maintained over the course of the simulations, giving further insight into the interactions of the binding site. By comparing the central site (Figure 14 (A) & (B)) total contacts over time with those of the allosteric site (Figure 14 (C) & (D)) it appears that the central site on average maintains about one more contact on average throughout the simulation with the ligand than the allosteric site does. When comparing the ligands with one another it appears that (S)-citalopram (Figure 14 (A) & (C)) maintains on average less contacts with its binding sites than serotonin (Figure 14 (B) & (D)) does.

4.1.2.3 Protein-Ligand Contacts – Contact Strength

Below figures display contacts that had maintained a contact strength of a minimum of 30% for the duration of the simulation. 30% was the standard setting used by the Simulation Interaction Diagram, and indicates the number of interactions that were maintained between contacts and for what degree of the simulation they were maintained.
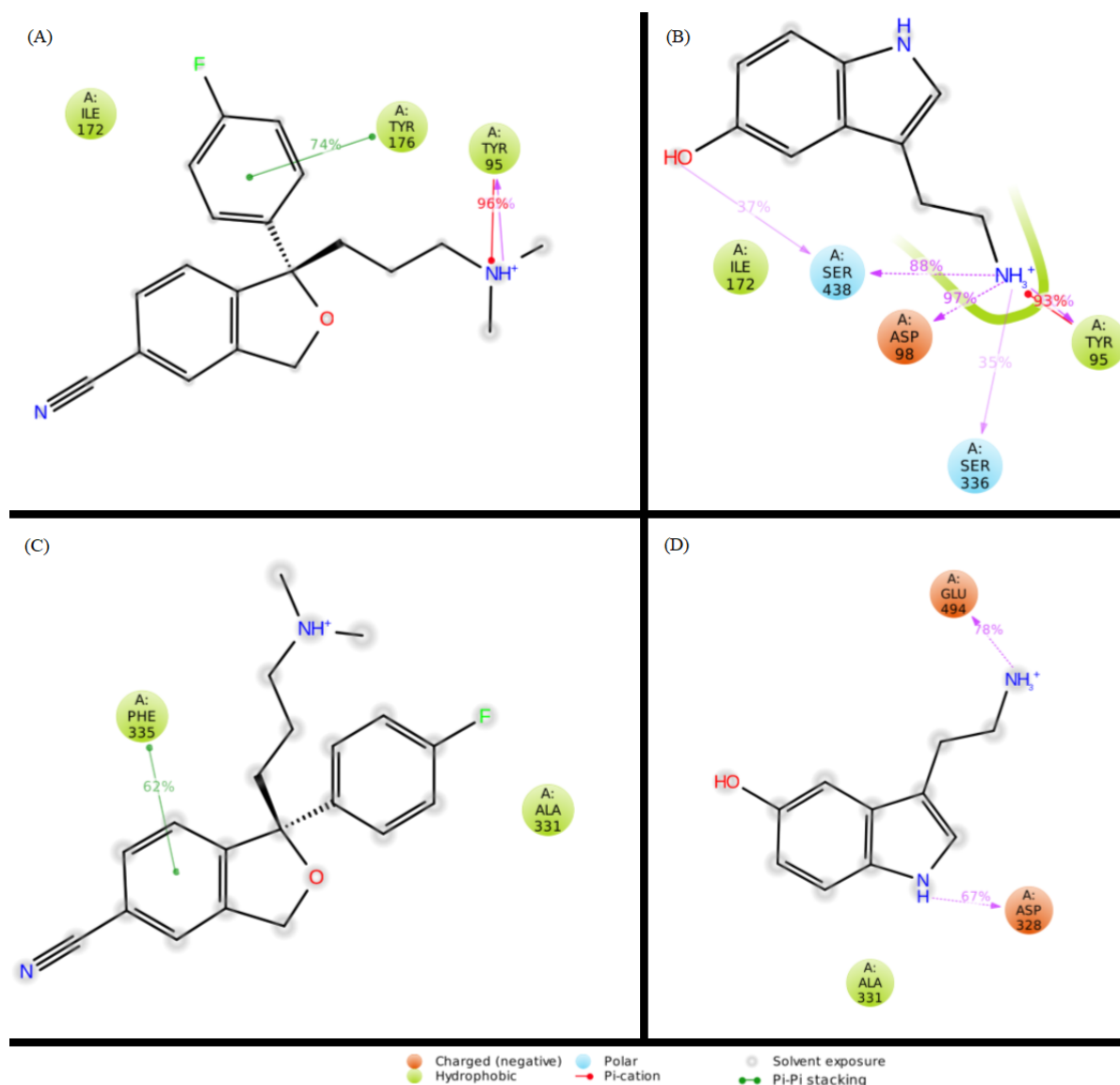
Figure 15: The figure displays the contacts between amino acids and the ligand for each of the MD simulations. For a contact to be displayed it was required to have a contact strength above the standard threshold of 30%. The top part of the figure displays the results from the central binding pocket, with (A) being (S)-citalopram and (B) being serotonin. The bottom part of the figure displays the results from the allosteric binding pocket, (C) is (S)-citalopram and (D) is serotonin.

As can be seen from Figure 15 each of the ligands were under the effects of solvent exposure throughout the simulation, a contributing factor to the water bridge interactions that were observed to have occurred between the ligands and some amino acids in Figure 11. With regards to the amino acids that each ligand is shown to have a contact strength of above 30% within each respective binding site it is clear that the amino acids that were prominent in Figure 11 and Figure 14 are also represented in Figure 15. For the (S)-citalopram in the central site we see TYR95, ILE172 and TYR176. Of particular note here is the interactions between TYR95 and the NH^+ of (S)-citalopram, showing which part of the antagonist is interacting most strongly with the central binding site. For serotonin in the central site we see a larger number of contacts that are above the 30% contact strength threshold (Figure 15). First among these are TYR95, ASP98 and SER438, which Figure 11 and Figure 14 clearly show have strong and steady connections with serotonin. TYR95, ASP98 and SER438 all interact with NH_3^+, indicating that the interactions of NH_3^+ with the central site are of significant importance. SER438 also has an interaction with the OH group of serotonin. NH_3^+ is also seen to interact with SER336, although Figure 14 shows that this interaction is steady, but with regular gaps. ILE172 was present above the 30% contact strength threshold for both central site simulations, with Figure 11 showing that the interactions it had with the respective ligands was of a hydrophobic nature.

The allosteric site simulation with (S)-citalopram had only PHE335 and ALA331 meeting the 30% contact strength threshold (Figure 15). Figure 14 did also show quite clearly that there were very few amino acids in the allosteric site that maintained steady interactions with (S)-citalopram for longer stretches of the simulation. The allosteric site serotonin simulation saw three amino acids meet the 30% contact strength threshold. First among them being GLU494 which interacted with NH_3^+ on serotonin, a contact that was heavily favored for serotonin in the central site. Second is ASP328, which interacted with NH on serotonin, a contact that was not seen in the central site above the 30% contact strength threshold. The final amino acid for serotonin in the allosteric site is ALA331, which was present for both allosteric site ligands above the 30% contact strength threshold. As was seen in Figure 11 the interactions from ALA331 were of a hydrophobic nature, although for serotonin there was also a small portion of the interactions that took the form of H-bonds.

4.1.2.4 Central Site – amino acid movements

In order to study the behavior of the amino acids of the central binding site measurements of the distance between selected amino acids were made. The amino acids that were selected are known to be crucial to the function of SERT (TYR95 (TM1), ASP98 (TM1), ILE172 (TM3), ASN177 (TM3), PHE341 (TM6), SER438 (TM8)) and are all part of the central binding site. The measurements that were made were between amino acids that were opposite to one another with the ligand between them, each pair had the measurements made between different transmembrane helices. The atoms that were selected to be used for measurements were selected oxygen or nitrogen side chain atoms facing the center of the binding site at the start of the simulation. In this fashion three measurements were made that together give some insight into the movements of these key amino acids relative to one another over the course of the simulations (Figure 16-18).

Figure 16: The figure displays distance measurements made for all four MD simulations, measuring the distance between the amino acids TYR95 and ASN177 over the course of each respective simulation. The TYR95 amino acid is located in TM1 and ASN177 is located in TM3. The top row shows the MD simulations of the central site, with (A) (S)-citalopram being on the left and (B) serotonin being on the right. The bottom row shows the MD simulations of the allosteric site, with (C) (S)-citalopram being on the left and (D) serotonin being on the right.

From Figure 16 it can be seen that there were some differences between the simulations when looking at the distance between the TYR95 and ASN177 amino acids over time. Of particular note among these measurements are those for (S)-citalopram in the central site (Figure 16 (A)) and serotonin in the allosteric site with an empty central site (Figure 16 (D)). The simulation with (S)-citalopram in the central site appeared to be the one which had the greatest degree of instability in terms of distance between the measured amino acids. The simulation with serotonin in the allosteric site is however notable because the slowly rising distance between the measured amino acids for the first ~75 ns of the simulation, the somewhat stability displayed during the next 75 ns, then a slow slight decrease in average distance.
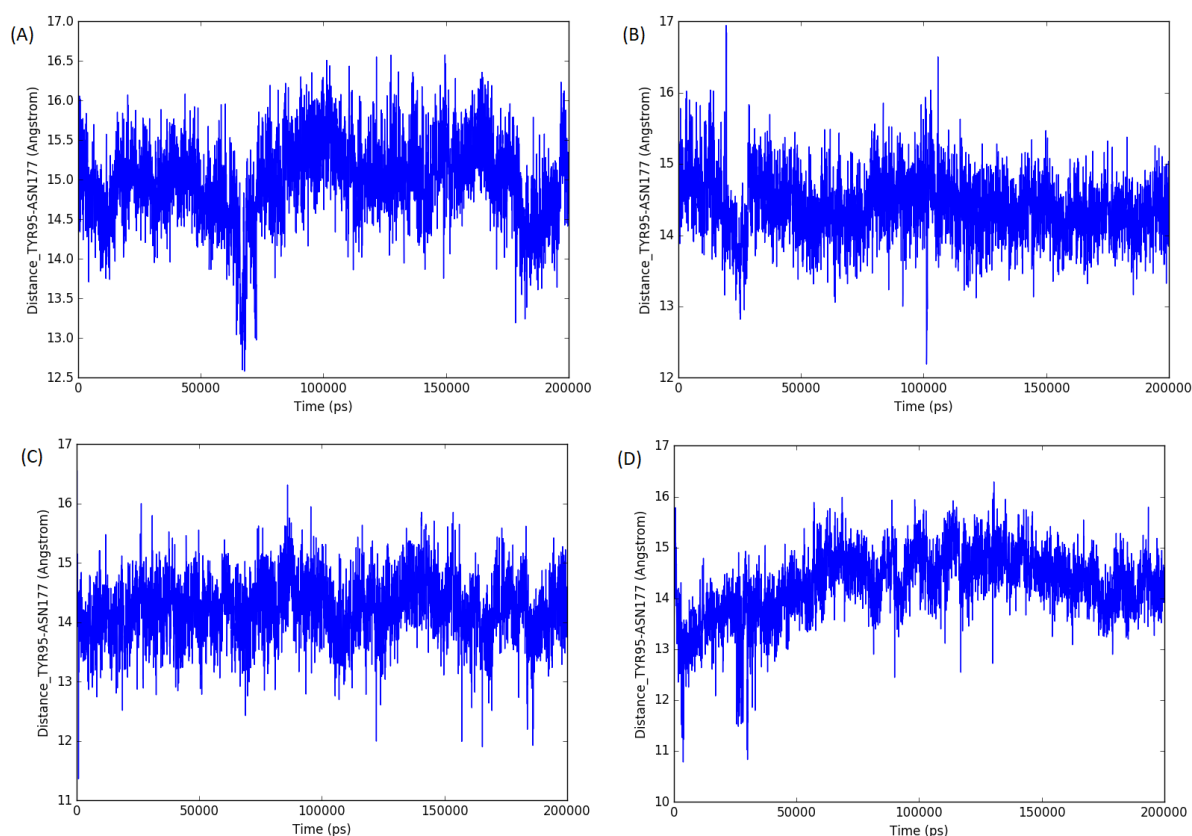

Figure 17: The figure displays distance measurements made for all four MD simulations, measuring the distance between the amino acids ASP98 and PHE341 over the course of each respective simulation. The ASP98 amino acid is located in TM1 and PHE341 is located in TM6. The top row shows the MD simulations of the central site, with (A) (S)-citalopram being on the left and (B) serotonin being on the right. The bottom row shows the MD simulations of the allosteric site, with (C) (S)-citalopram being on the left and (D) serotonin being on the right.

From Figure 17 it can be seen that the measured distances between ASP98 and PHE341 contained significantly more fluctuations then the measured distances between TYR95 and

ASN177. Which in Figure 17 actually makes (A) unique in that it does not contain numerous maintained spikes in fluctuation. Rather the amino acids seem to be continuously fluctuating more than for the other simulations, the only area that can be described as a maintained spike in activity for (A) is the area between 80 ns and 160 ns, which appears to be a small lengthy spike in distance between the two amino acids. In addition to this the approximate average Å distance that is maintained is about 1.5 Å higher than for the other simulations, lying at 15.5 Å – 17.0 Å. The simulation with serotonin in the central site (B) may be the simulation with the most stable fluctuations in distance, even considering the two small and two tiny maintained spikes in activity. When not spiking in distance the distance between the amino acids fluctuates at a very stable 13-14 Å, which is quite stable when compared with the other sections of Figure 17.

For the two simulations with no ligand in the central site and (S)-citalopram (C) and serotonin (D) in the allosteric site the most notable feature of their distance measurements is the very numerous maintained spikes in distance. Both these simulations started with a substantially lower distance between the amino acids than for the central site simulation, but they quickly increase this distance until they are both at about the average Å distance seen in (B).
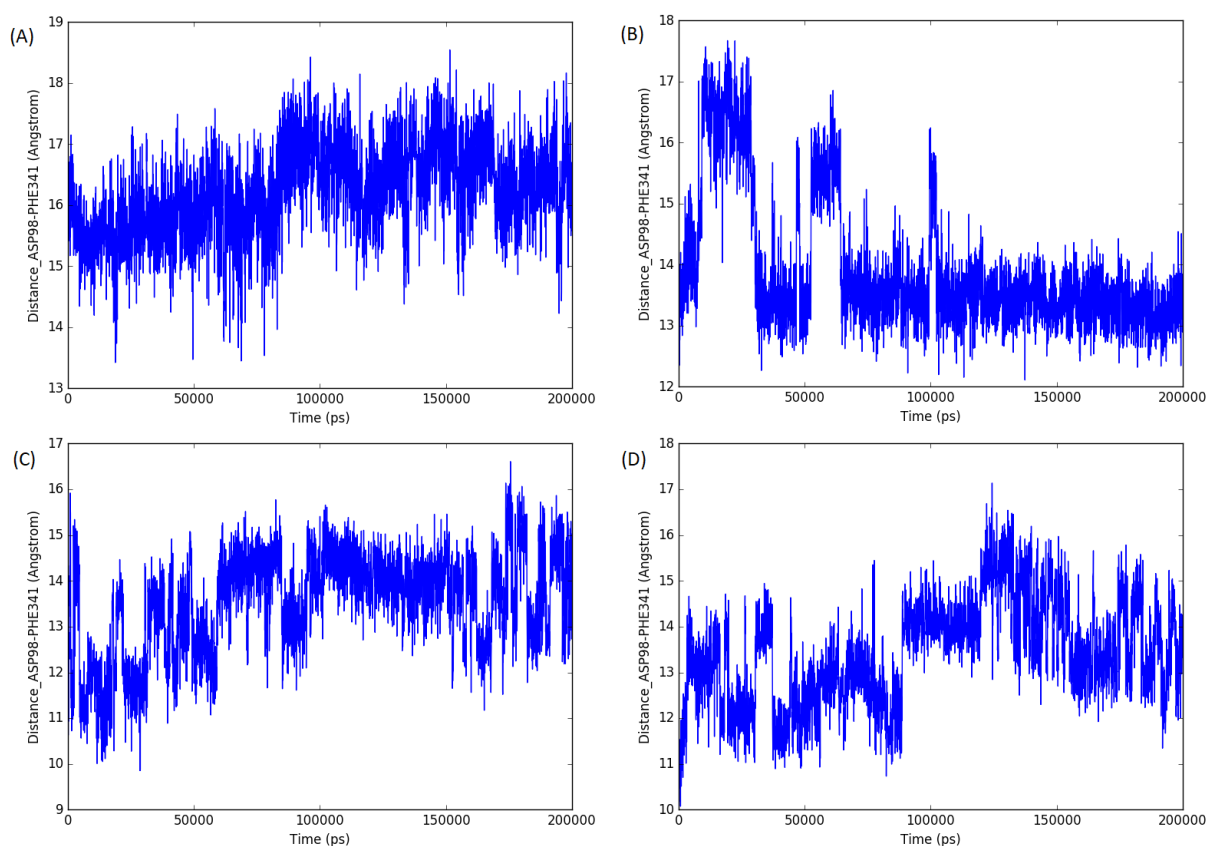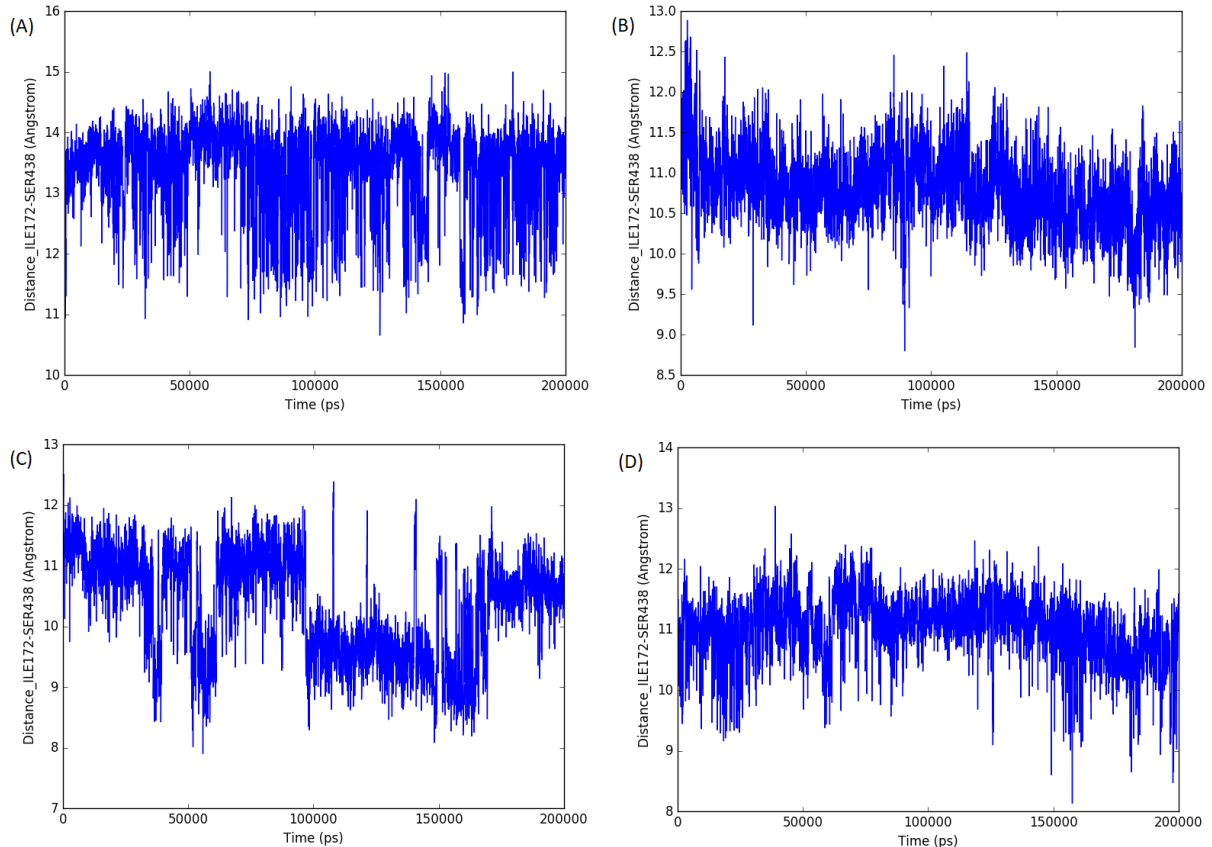
Figure 18: The figure displays distance measurements made for all four MD simulations, measuring the distance between the amino acids ILE172 and SER438 over the course of each respective simulation. The ILE172 amino acid is located in TM3 and SER438 is located in TM8. The top row shows the MD simulations of the central site, with (A) (S)-citalopram being on the left and (B) serotonin being on the right. The bottom row shows the MD simulations of the allosteric site, with (C) (S)-citalopram being on the left and (D) serotonin being on the right.

From Figure 18 the distance between amino acids ILE172 and SER438 were measured over the course of each simulation. These measurements are notably between TM3 and TM8, while the measurements in Figure 16 and Figure 16 are both from TM1 towards respectively TM3 and TM6. The central site simulation with (S)-citalopram is notable for the continuous large fluctuations that occurred for the majority of the simulation, fluctuating between 11-15 Å. This is in contrast to the simulation with (S)-citalopram in the allosteric site, where the measured distance between the amino acids is at its highest 12.5 Å and it goes as low as 8 Å. The allosteric site simulation had smaller fluctuations, although it did have some maintained spikes in measured distance. Both simulations with serotonin are comparatively more stable, with no large maintained spikes in distance and the fluctuations in measured distance between the amino acids staying in the range of about 2 Å between 10-12 Å. There does appear to be a slight trend towards the distance shortening from the 150 ns point, but an extended simulation would have been required to ascertain this.
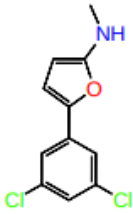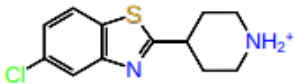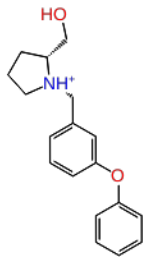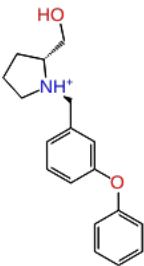
4.2 Induced Fit Docking of Experimentally verified Inhibitors
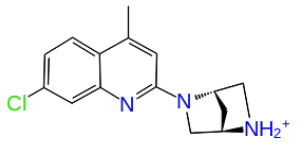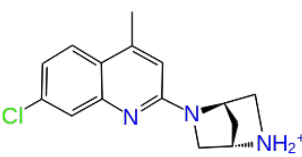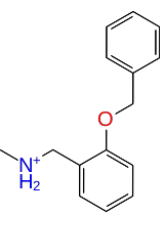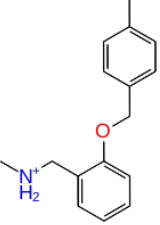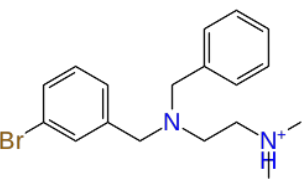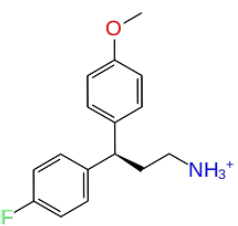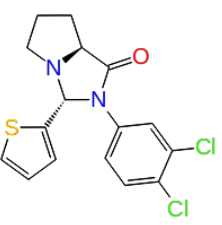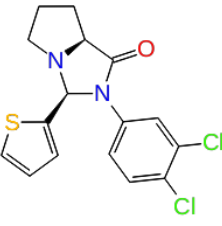
Inhibitors were selected from a previous study in our research group (100), all selected inhibitors (apart from citalopram) had been experimentally verified in that paper. Inhibitors had also been run through a virtual screening using a SERT outward-facing homology model.
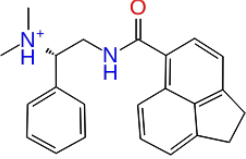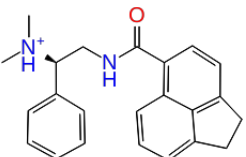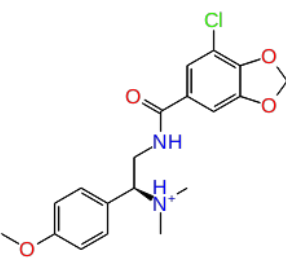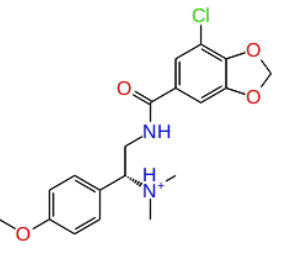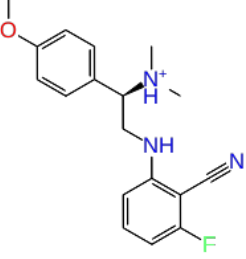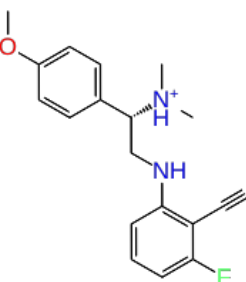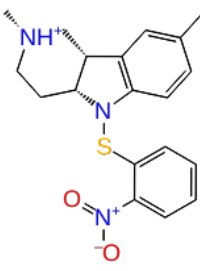
4.2.1 Docking Score
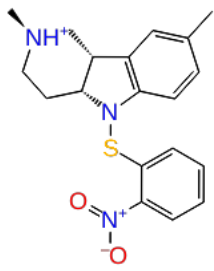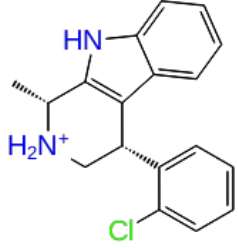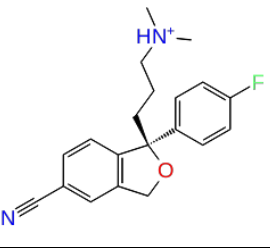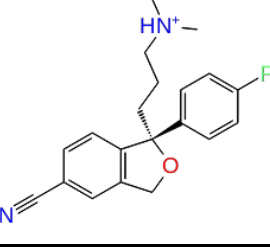
The most promising inhibitory compounds were selected from among structurally based clusters in the paper, if a cluster had no compounds of sufficient quality it would not be taken along for docking studies. For an overview of structural clusters see Appendix Table 1.

Table 1: The table shows the results of the docking of selected inhibitory compounds previously tested in an earlier study in our research group (100). The ID refers to the compounds ID number, which is used to identify a compounds name in the Appendix (Appendix Table 1). The Structure PDB ID is the crystal structure the inhibitory compound was docked into, the Site refers to which binding site within the crystal structure it was docked into, either Central or Allosteric. The Docking Score is the score the ligand got when docked into the docking site, scores below -6.0 are acceptable, lower is better. Ranking refers to the compounds Docking score relative to the other docked compounds.

Structure	ID	Structure PDB ID	Site	Docking Score	Ranking
	1	5I71	Central	-8.56	18/32
	1	5I73	Allosteric	-8.02	22/32
	2	5I71	Central	-8.99	15/32
	2	5I73	Allosteric	-7.11	30/32
	3	5I71	Central	-9.89	9/32
	3	5I73	Allosteric	-7.63	27/32
	4	5I71	Central	-11.05	2/32
	4	5I73	Allosteric	-9.28	13/32

Structure	#	PDB	Site	Energy	Rank
(7-chloro-4-methylquinolin-2-yl azetidine amine structure)	5	5I71	Central	-9.44	12/32
(7-chloro-4-methylquinolin-2-yl azetidine amine structure)	5	5I73	Allosteric	-7.52	28/32
(2-(benzyloxy)benzyl methylamine structure)	6	5I71	Central	-7.95	24/32
(2-(benzyloxy)benzyl methylamine structure)	6	5I73	Allosteric	-6.19	32/32
(2-((4-methylbenzyl)oxy)benzyl methylamine structure)	7	5I71	Central	-8.34	19/32
(2-((4-methylbenzyl)oxy)benzyl methylamine structure)	7	5I73	Allosteric	-6.86	31/32
(3-bromobenzyl benzyl ethylenediamine structure)	8	5I71	Central	-10.24	5/32
(3-bromobenzyl benzyl ethylenediamine structure)	8	5I73	Allosteric	-7.33	29/32
(4-methoxyphenyl 4-fluorophenyl ethanamine structure)	9	5I71	Central	-9.85	10/32
(4-methoxyphenyl 4-fluorophenyl ethanamine structure)	9	5I73	Allosteric	-7.99	23/32
(pyrrolidine-fused imidazolone with thiophene and dichlorophenyl structure)	10	5I71	Central	-10.13	7/32
(pyrrolidine-fused imidazolone with thiophene and dichlorophenyl structure)	10	5I73	Allosteric	-8.72	17/32

	11	5I71	Central	-12.10	1/32
	11	5I73	Allosteric	-8.77	16/32
	12	5I71	Central	-9.58	11/32
	12	5I73	Allosteric	-7.79	25/32
	13	5I71	Central	-10.13	7/32
	13	5I73	Allosteric	-7.65	26/32
	14	5I71	Central	-10.23	6/32

	14	5I73	Allosteric	-8.32	20/32
	15	5I71	Central	-10.93	3/32
	15	5I73	Allosteric	-9.01	14/32
	16	5I71	Central	-10.58	4/32
	16	5I73	Allosteric	-8.16	21/32

The induced fit docking of the experimentally verified inhibitors alongside citalopram provided a large quantity of data relating both which amino acids the compounds interacted with in their best scoring poses (Appendix Figure 1-8) and also how the compounds scored between the binding sites comparatively. The compounds were ranked according to docking score, as can be seen from Table 1.

4.2.2 Docking Score – Statistical Analysis

Here we look at what the data gained from the induced fit docking part of the study can yield when analyzed statistically. The sample small size is limiting, with only 15 experimentally tested inhibitors being used, the results are also potentially significantly influenced by outliers.
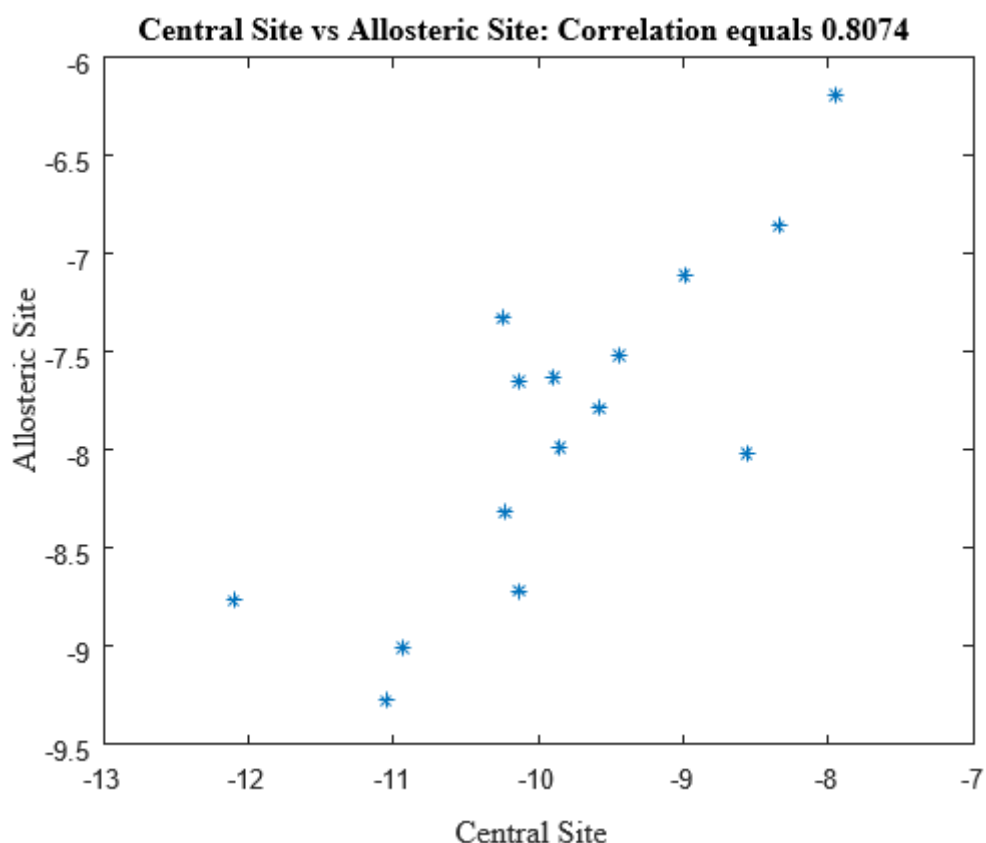
Figure 19: The figure displays the docking scores for each of the experimentally verified compounds, the score for the central site is along the X-axis and the score for the allosteric site is along the Y-axis. There was found to be a correlation equal to 0.8074 between the docking scores of the central site and the allosteric site.

When comparing the docking scores of each compound between the central site and the allosteric site (Figure 19) it became clear that there was a high positive correlation, specifically a value equal to 0.8074, between the scores of the respective sites for each compound. From the plot, it is seen that a compound will have a higher docking score in the central site than in the allosteric site. The average docking score in the central site is -9.827 and -7.879 for the allosteric site, which gives a difference of -1.948 between the averages.
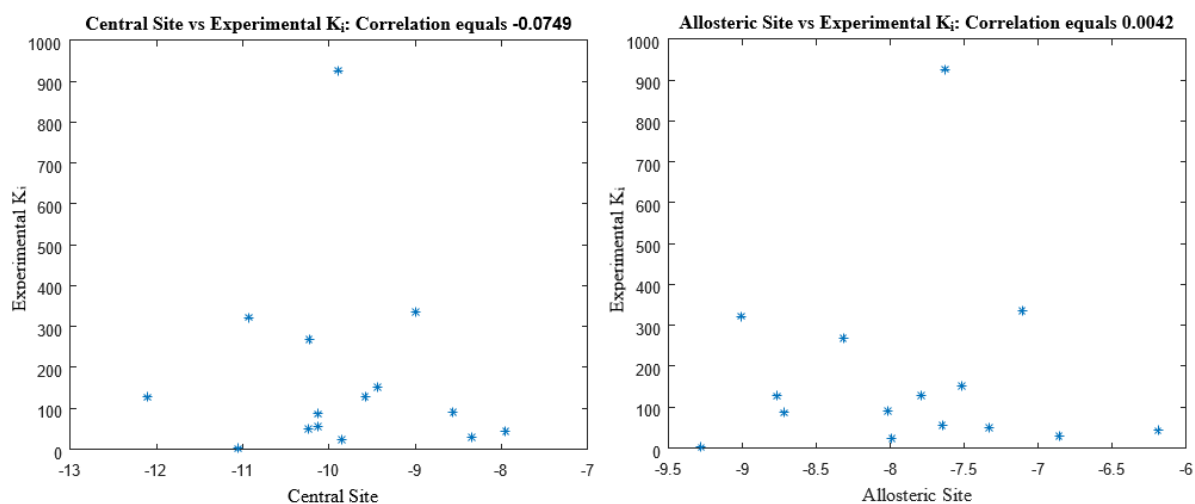
Figure 20: The figure above displays the docking scores for all the experimentally verified compounds in the central site (left) and allosteric site (right) against the experimental K_i (Appendix Table 1) that was found for the compounds in the experimental study (100). The docking score is along the X-axis and the experimental K_i is along the Y-axis. There was found a correlation equal to -0.0749 for the central site and of size 0.0042 for the allosteric site.

This data in Figure 20 indicates that there is neither a linear relationship between the docking scores of the central site and the experimental K_i nor between the allosteric site and the experimental K_i.
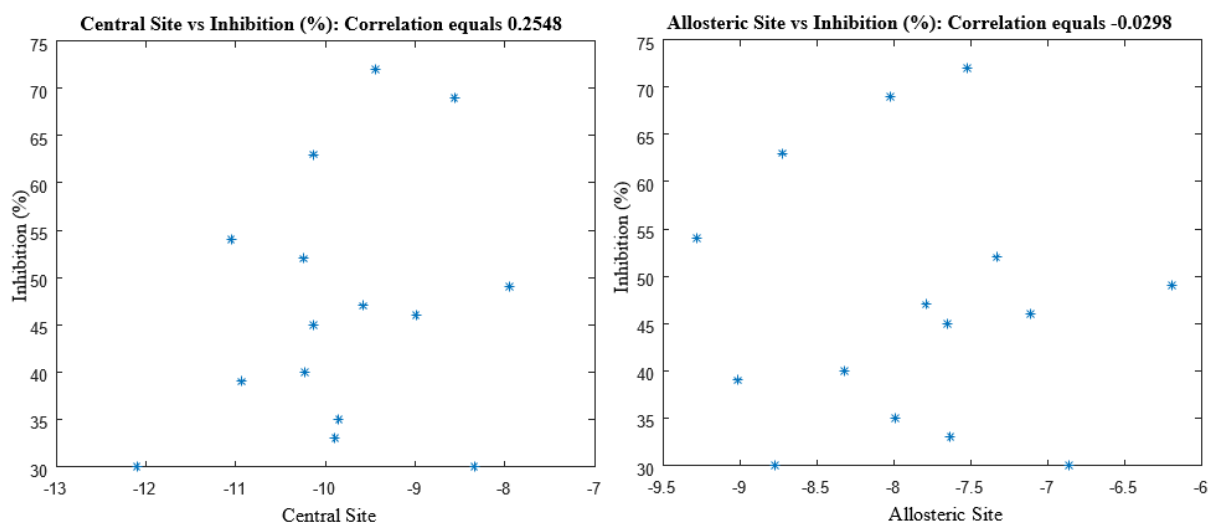
Figure 21: The figure above displays the docking scores for all the experimentally verified compounds in the central site (left) and allosteric site (right) against the inhibition in percentage (Appendix Table 1) that was achieved for the compounds in the experimental study (100) with the Ki listed in (Appendix Table 1). The docking scores is along the X-axis and the inhibition percentage the compound achieved is along the Y-axis. There was found to be a correlation equal to 0.2548 for the central site and a correlation of -0.0298 for the allosteric site.

For the central site (Figure 21, right) there was found to be a small degree of correlation between the docking score of the compound and its inhibition percentage. However, given the small sample size the degree of correlation that was found may in fact not be significant. For the allosteric site there was found to be no correlation between the docking score and the inhibition percentage.

4.2.3 Amino Acid interactions of docked compounds

The compounds that were docked during the docking study also provide data through the location of the ligand in their best scoring docking pose. Specifically, by looking at which amino acids, and ions, the ligand may or may not have interacted with while in each respective binding site. For the central site there are numerous amino acids that are known to be crucial to the function of SERT, but for the recently discovered allosteric site the amino acids that are interacted with may give insight into the specific interactions that would occur in the allosteric site of SERT *in vitro*.

Table 2: The table shows the interactions between each respective ligand and the amino acids of the central site. The interactions were obtained from the Ligand Interaction Diagram using the default settings (See Appendix Figure 1-4). The numbers in the first row designate each compounds ID, the leftmost column designates the interacting molecule. An X indicates if there was an interaction listed in the Ligand Interaction Diagram between the respective ligand and the molecule listed in the leftmost column.

Interactions	Ligand															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
TYR95	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ALA96	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
VAL97			X		X								X	X		X
ASP98		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
GLY100			X													
ARG104			X										X			
ILE168					X					X			X	X		
ALA169	X	X	X	X	X	X	X	X	X	X	X	X			X	X
ILE172	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ALA173		X	X	X	X	X	X	X	X	X	X	X	X			X
TYR175			X							X		X		X	X	X
TYR176	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ASN177		X	X	X			X				X					
MET180				X							X					
PHE334					X	X		X	X		X	X		X		
PHE335	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SER336	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LEU337	X	X	X		X			X		X	X	X	X	X	X	X
GLY338	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
PRO339								X			X			X		
PHE341	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
VAL343						X		X			X				X	
ASN368			X											X		
GLY435				X												
SER438	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
THR439	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X

GLY442	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
LEU443		X	X	X	X	X	X	X	X	X	X	X				X
VAL446											X					
GLU493			X								X		X			
THR497			X						X		X		X		X	X
GLY498			X								X		X		X	X
VAL501	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X
NA707		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

As can be seen from Table 2 there are numerous amino acids and one ion that the ligands tend to interact with in the central site. There are however some of these interactions which occur on a consistent basis with all or nearly all the ligands. The molecules that are seen to interact with the ligand in each or all but one case are; TYR95, ALA96, ASP98, ILE172, TYR176, PHE335, SER336, GLY338, PHE341, SER438, THR439, GLY442, VAL501 and NA707. When it comes to amino acids that are known to be crucial to the function of SERT these are TYR95, ASP98, ILE172, ASN177, PHE341 and SER438 (118). Among the molecules that interact with the ligand in all but one case five out of the six amino acids that are known to be crucial for the function of SERT are included, with the final amino acid ASN177 having interactions with only a few of the ligands that were docked.

Table 3: The table shows the interactions between each respective ligand and the amino acids of the allosteric site. The interactions were obtained from the Ligand Interaction Diagram using the default settings (See Appendix Figure 5-8). The numbers in the first row designate each compounds ID, the leftmost column designates the interacting molecule. An X indicates if there was an interaction listed in the Ligand Interaction Diagram between the respective ligand and the molecule listed in the leftmost column.

Interactions	Ligand															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
TYR95			X							X					X	
ALA96										X					X	
ASP98			X	X						X		X			X	
GLY100				X							X			X		X
ASN101				X												
ARG104	X	X	X	X	X	X	X	X	X		X	X	X	X		X
ILE108						X	X	X			X					X
ALA169										X					X	
ILE172			X		X					X		X		X	X	
ALA173										X						
TYR175	X		X		X			X	X			X		X	X	
TYR176			X							X		X			X	
ILE327						X	X				X		X			X
ASP328						X	X				X					X
ALA331	X	X		X	X	X	X	X	X		X	X	X	X		X
GLN332		X		X		X	X				X		X	X		X
PHE335	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SER336				X						X	X			X	X	
LEU337										X					X	
GLY338			X							X					X	
PHE341			X							X					X	
SER438			X							X					X	
THR439										X					X	
GLY442										X					X	
LEU443										X						
LYS490								X								

GLU493	X	X	X	X	X	X		X	X		X	X	X	X	X	X
GLU494	X	X	X	X	X	X	X	X	X		X	X	X	X		X
TYR495		X			X											
THR497		X			X			X	X			X		X		
GLY498	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
PRO499	X	X	X	X	X	X	X	X	X		X	X	X	X		X
VAL501		X								X		X				
LEU502	X	X	X	X	X	X	X	X	X		X	X	X	X		X
ILE552	X	X		X	X	X	X	X	X		X	X	X	X		X
ILE553							X	X	X			X				
SER555											X		X			
PHE556	X	X		X	X	X	X	X	X		X	X	X	X		X
SER559													X			
PRO561		X			X	X		X	X		X	X	X			X
NA707				X											X	

As can be seen from Table 3 there are numerous amino acids and one ion that the ligands tend to interact with when docked into the allosteric site. As with the central site there are some of these amino acids that that tend to interact with the ligands more consistently than others. The specific amino acids that tend to interact with the ligands are also, for the most part, different from the central site. Something in part caused by the approximate distance of 13 Å between the central and allosteric binding sites. Using the same parameters as with the central site for selecting molecules that interact with the ligand in all or all but one case gives us only the amino acids PHE335 and GLY498 that all ligands interacted with when docked into the allosteric site. If we however adjust the parameters for selecting high interaction rate amino acids to discount for ligands 3, 10 and 15 we find that the following amino acids now meet the updated requirements; ARG104, ALA331, PHE335, GLU493, GLU494, GLY498, PRO499, LEU502, ILE552 and PHE556. Among these there are no amino acids that are known to be crucial to the function of SERT, although the same study did find that mutations to PHE335 did give a significant reduction to the transport activity of SERT (118).

Ligands 3, 10 and 15 are, as can be seen from Table 3, in many ways outliers that interact with amino acids that other ligands docked in the allosteric site did not interacts with. The most clear

and notable examples is TYR95, which is an amino acid that all ligands in the central site interacted with (Table 2). The optimal placement by the IFD protocol of each of these ligands in the workspace is responsible for this, having placed them in the area between the central and allosteric sites, partially inside the central site.

4.3 Experimental testing of potential GABA$_B$ receptor antagonists

For the experimental portion of the project, ten compounds were selected from an ongoing study (Evenseth et al. 2018, unpublished work) and their activities were tested towards the GABA$_B$ receptor using a cAMP functional assay. This experimental screening was first tested on a CHO-K1 Wild Type cell line to determine if the compounds would interact with any receptors other than the GABA$_B$ receptor. The initial screening then tested the compounds against a CHO-K1 cell line stably expressing the GABA$_B$ receptor in order to investigate if the compounds had any direct effect on the GABA$_B$ receptor. The compounds being tested experimentally had been selected based on their scoring in an *in silico* virtual screening approach for antagonists of the GABA$_B$ receptor. The results included in this project are for only a single run of the assay, not a triplicate run. Numerous delays and malfunctioning equipment pushed the schedule for performing the assay back so far that it was no longer feasible to perform a triplicate of the assay within the projects timeframe. Due to this, the results of the assay are not totally reliable as the lack of triplicate testing leaves significant room for error at multiple stages.

The need for two solvents, and therefore two controls, for the test compounds came from the low solubility of five of the compounds. Specifically compound 2, 3, 6, 8 and 9 were not soluble in the standard assay buffer HBSS and required a different more soluble solvent, a D-glucose solution, referred to as 5G. The soluble compounds were 1, 4, 5, 7 and 11, all of which were readily and stably dissolved in the assay buffer and did not form any crystals.
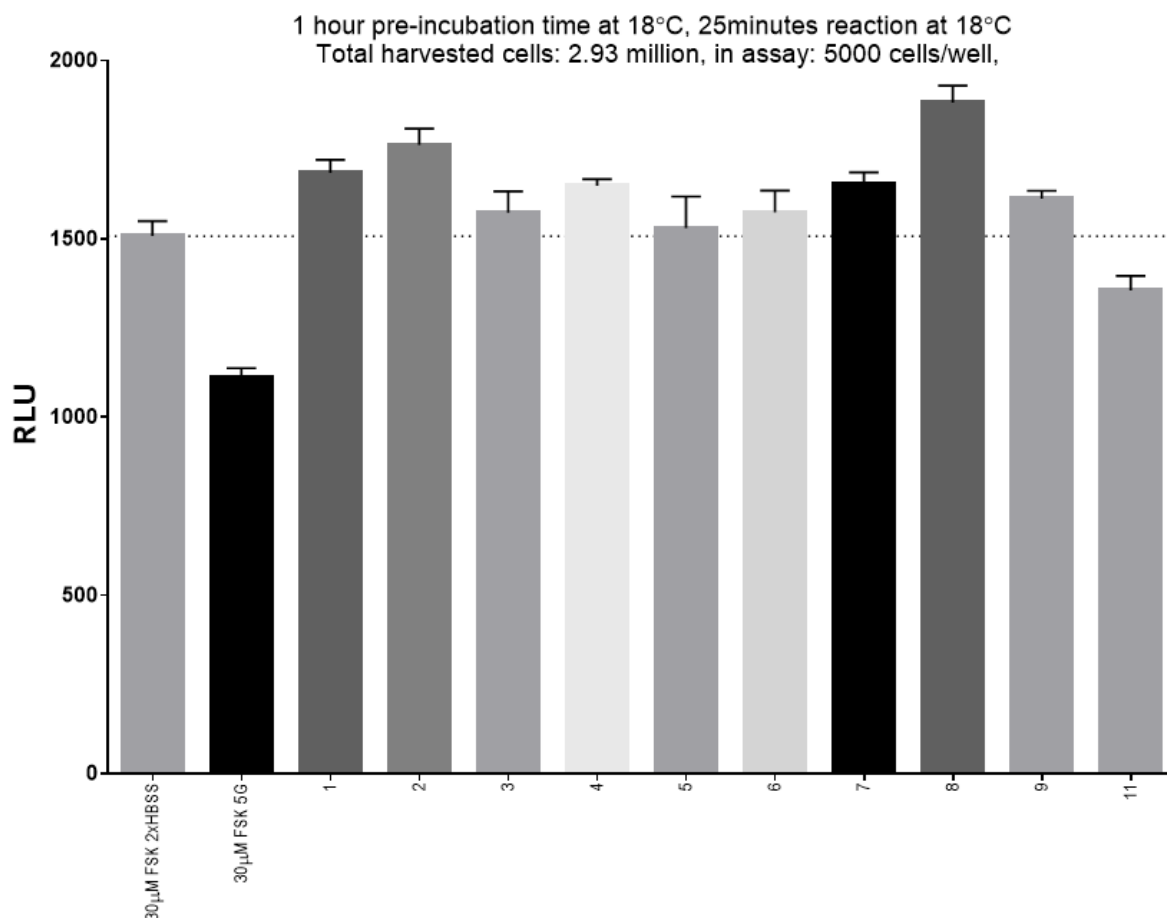
Figure 22: The above bar chart shows the results of the functional cAMP assay to test the ten selected compounds against a CHO-K1 Wild Type cell line in order to determine if the compounds had any effect on the receptors of the native Wild Type cells. The X-axis shows the contents of the tested wells, with the two columns on the left being the controls and the ten on the right being the compounds that were being tested. The Y-axis shows the relative light unit (RLU) of each well that has a linear relationship to the cAMP expression level. Each well was done in triplicate during the assay and the above results shows the average, with the standard deviation displayed on top of each column. The dotted line was set to the average of the control (stimulation of 30 µM forskolin alone).

Test compounds' activity against CHO-K1 WT is shown on Figure 22. The two leftmost columns are the controls, one for each solvent of the respective compounds. As can be seen from Figure 22 the 30 µM FSK 5G control was clearly anomalous, as it was supposed to be of equal level with the 30 µM FSK 2xHBSS control. It was therefore decided to use the 30 µM FSK 2xHBSS as a baseline to see if any of the compounds displayed irregular activity. As the bar chart in Figure 22 is on an individual assay, the results are less certain with more room for the effects of experimental errors. However, compared to the forskolin control and their

standard deviation, three compounds (compound 1, 2 and 8) showed a higher cAMP level than the rest of the compounds. This might indicated that compound 1, 2 and 8 interacted with an endogenous receptor or another cell component in the WT CHO-K1 cells.
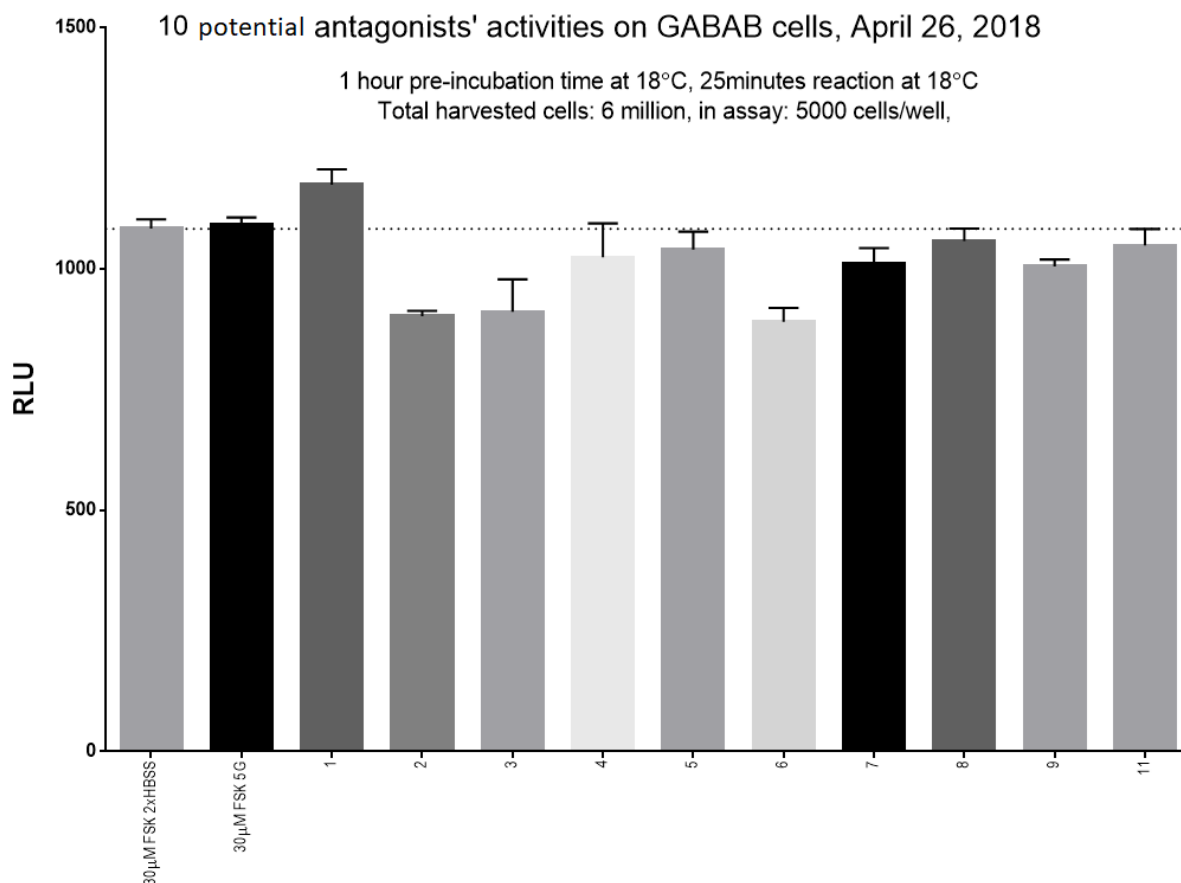
Figure 23: The above bar chart shows the results of the cAMP assay to test the ten selected compounds against a CHO-K1 cell line expressing the $GABA_B$ receptor in order to determine if the compounds had an inhibitory effect on the $GABA_B$ receptor. The X-axis shows each the contents of the tested wells, with the two columns on the left being the controls and the ten on the right being the compounds that were being tested. The Y-axis shows the relative light unit (RLU) of each well that has a linear relationship to the cAMP expression level. Each well was done in triplicate during the assay and the above results shows the average, with the standard deviation displayed on top of each column. The dotted line was set to the average level of the controls (stimulation of 1 µM GABA + 30 µM forskolin).

The second assay performed was to test if the chosen compounds showed any reaction with the $GABA_B$ receptor transfected cell line. A reaction in this cell line would indicate that the compound interacted with $GABA_B$ receptor in an agonistic or antagonistic manner, if there was no similar reaction from the compound in the wild type cell line. As can be seen from Figure 23 the only compound that has a higher RLU was compound 1, but compound 1 also showed a similar degree of interaction during the wild type assay. This raises the possibility of

compound 1 reacting with the cell through a different pathway than the GABA$_B$ receptor. The other compounds showed RLUs close to that of the controls, indicating there was potentially no inhibitory reaction. The compounds with lower RLUs than controls may indicate agonist like activity against the GABA$_B$ receptor. Compounds 2, 3 and 6 showed the strongest agonist-like activity, at least three repeated individual assays are needed to confirm that the observed activity is not caused by experimental error.

Table 4: The below table shows an overview of the data from Figure 22 and Figure 23. The leftmost column shows the compound IDs and name of controls for each respective row. The solubility column shows if the compound had a high or low solubility. Compounds with high solubility were dissolved in the assay buffer (HBSS), compounds with low solubility were dissolved in the 5G buffer. The following columns displays the percentage value for each of the bars in the bar charts of Figure 22 and Figure 23 in comparison to the non-anomalous controls. For the WT assay all compounds are compared with the assay buffer control, including the 5G buffer control which was anomalous. For the GABA$_B$ assay the high solubility compounds are compared with the assay buffer and the low solubility compounds are compared with the 5G buffer. For both of these assays each rows standard deviation is also listed in percentage.

ID	Solubility	Score vs Control (%) WT	Standard Deviation (%)	Score vs Control (%) (GABA$_B$)	Standard Deviation (%)
1	High	111.73	± 2.14	92.17	± 2.69
2	Low	116.90	± 2.56	120.95	± 1.25
3	Low	104.31	± 3.75	119.76	± 7.39
4	High	109.28	± 1.13	105.76	± 6.83
5	High	101.39	± 5.81	104.13	± 3.53
6	Low	104,37	± 3.84	122.54	± 3.21
7	High	109.48	± 2.12	107.02	± 3.12
8	Low	124.72	± 2.55	103.12	± 2.41
9	Low	106.89	± 1.36	108.56	± 1.50
11	High	89.86	± 2.98	103.24	± 3.24
30 µM FSK 2xHBSS (Control - High)		100	± 2.77	100	± 1.80
30 µM FSK 5G (Control - Low)		73.56	± 2.43	100	± 1.44

The bar charts of Figure 22 and Figure 23 do not provide exact values for their columns, but do provide a good comparative overview between the controls and different compounds. For a more exact overview there is Table 4, which has all the values listed as a percentage value of that of the controls. As the low solubility buffer (5G) was anomalous in the WT assay it was not used for comparative purposes. The anomalous nature of the WT 5G buffer control is clear from its very low comparative value with that of the other control and all the compounds. Comparing it with the overview observable in Figure 22 and Figure 23 reinforces this.

5. Discussion

5.1 Molecular Dynamics

The four MD simulations that were performed for this project were 200 ns in length. This makes the MD simulations unsuitable for the study of changes to the overall structure of the protein, but allows the project to instead focus more closely on the interactions between the ligand and the protein over the course of the simulation. It also permits closer study of the behavior of amino acids in the central that are known to be crucial for the function of SERT (118). The analysis of the projects MD simulation results shall therefore be focused on the interactions of the respective ligands in both binding sites, and also on how these ligands affect the above mentioned crucial amino acids located in the central site.

5.1.1 Structural Stability

While the simulations that were performed for the project are too short to study potential changes to overall protein structure it is still important to ascertain that the protein structures were stable throughout their simulations. The Protein RMSD observable in Figure 9 is indicative of the protein structures being stable in each simulation after a short adjustment period at the start of the simulation. Each simulation shows that the movements of the protein structures relative to the reference frame backbone are highly stable, with no fluctuations exceeding 0.5 Å after the initial 50 ns of the simulation. The Ligand RMSD on the other hand is more variable between simulations, but all the ligands do stabilize before the 100 ns point of the simulation with no ligand having left its binding site. The Protein RMSF observable in Figure 10 shows that there are similar movements between amino acids in each protein structure over the course of the simulations, although there are some amino acid movement degrees that differ slightly from one another along the lines of which binding site was used. The difference is either caused by the use of different crystal structures between the central site simulations and the allosteric site simulations, or potentially the placement of the ligand may be affecting the behavior of the amino acids in close vicinity.

5.1.2 Ligand Interactions

The interactions between the ligands, (S)-citalopram and serotonin, and the binding sites for each simulation provided a lot of data regarding what amino acids interact with each ligand, the type of interaction and also to what degree they interact. The analysis of the simulation trajectory data allows insight into the complex interactions that occur within the central and allosteric binding sites of SERT with both the native substrate present and a well-tolerated SSRI (13).

5.1.2.1 Central Site

There have been previous computational studies (54) that studied SERT through the use of MD simulations, but these simulations were for only the central binding site using a homology model. With the use of a crystal structure there is a higher degree of certainty that the protein structure in the model system is reflective of what the actual structure of SERT is, especially in areas greatly affected by structural shifts such as binding sites.

As was described in the results section there were a total of six amino acids that passed the 30% contact strength threshold (Figure 15) within the central site. Two of these, TYR95 and ILE172, were shared for both ligands as high strength contacts. TYR176 was the third high contact strength amino acid for (S)-citalopram, but for serotonin it was a minor hydrophobic influence (Figure 11 (B)). Something which suggests that TYR176 perhaps plays a minor role within SERT for the transport of its substrate serotonin, but plays a more important role for the binding of an inhibitor. Further computational studies using an inhibitor other than (S)-citalopram would be able to investigate this further. The other three high contact strength amino acids that interacted with serotonin were ASP98, SER336 and SER438. Of these ASP98 was also shared with (S)-citalopram, although it had an interaction value of about 0.15 compared to serotonins value of 1.0, given its proximity to TYR95 within the central binding site any ligand that interacts with TYR95 would probably have some interactions with ASP98 as well. For the last two amino acids, SER336 and SER438, they were listed in Figure 11 (A), indicating that the ligand had at least one interaction with each of them over the course of the simulation, but the interaction value is very low. From Figure 14 one can see that (S)-citalopram had one and two points of interaction respectively for SER336 and SER438 over the course of the entire simulation. Such low interaction values with (S)-citalopram are indicative of the amino acids being substantially more important for serotonin than (S)-citalopram.

When comparing the amino acids that the ligands are interacting with in the central site with the six previously mentioned amino acids known to be crucial for SERT activity (TYR95, ASP98, ILE172, ASN177, PHE341, SER438), it can be seen (Figure 11 (A) and (B)) that only ASN177 does not interact with the ligands in the central site. With the exception of ILE172 it is also quite clear that serotonin, the substrate of SERT, interacts more strongly with all of these known crucial amino acids than (S)-citalopram.

When looking at the nature of the interactions that are occurring throughout the central site (Figure 11 (A) and (B)) there is clear divide between (S)-citalopram which favors hydrophobic interactions and serotonin which has H-bonds as its favored interaction. Both ligands however have a similar ratio of hydrophobic to H-bond interactions with TYR95. Ionic interactions are the least common, with only a small portion of ASP98' interactions being of an ionic nature for both ligands. The final interaction type is of particular interest as it deals with the dynamics of an MD simulation, water bridge interactions. As a water bridge interaction is a hydrogen-bonded protein-ligand interaction mediated by a water molecule it is the movement of water molecules in the binding site that allows them to form between the ligand and various amino acids. Water bridges are less common interactions than H-bonds and hydrophobic interactions (Figure 11), but still make up a substantial number of protein-ligand interactions. A portion of the interactions with serotonin of the known crucial amino acid SER438 are water bridges. Other amino acids in the central site that formed water bridges with the ligand had this as their sole form for interaction with each respective ligand. For (S)-citalopram there were two amino acids, TYR175 and THR497, both of which had interaction values of about 0.1. Serotonin also had another two amino acids with exclusively water bridges with the ligand, these being PHE334 and GLY442. PHE334 had an interaction value of about 0.1, while GLY442 had an interaction value of about 0.2, making GLY442 the amino acid with the largest amount of water bridge interactions in the central site.

5.1.2.2 Allosteric Site

Several in vitro experimental studies have indicated that inhibitors including citalopram could allosterically modify SERT activity (119), but the allosteric binding site was not firmly identified due to lack of detailed structural information. As the allosteric site of SERT was identified with the rendering of the crystal structure (53), the previous study of SERT using a homology model did not investigate it (54). With the allosteric site having not previously been investigated, any insights gained from the data regarding the interactions of the ligands in the allosteric site are potential avenues of further investigation. The 30% contact strength threshold is a good starting point for insight into the allosteric site, and the allosteric simulations provided a total of four amino acids that passed this threshold (Figure 15 (C) and (D)). Among these only ALA331 was included above the threshold for both the ligands. (S)-citalopram had PHE335 in addition to ALA331, while serotonin had ASP328 and GLU494. For both ligands ALA331 was a source of hydrophobic interactions with an interaction value of about 0.5 for both (Figure 11 (C) and (D)), although serotonin also did form some H-bonds with the amino acid. Both ASP328 and GLU494 had primarily H-bond interactions with serotonin, but they both also had a small portion of water bridges and GLU494 even had some ionic interactions. PHE335 was also a fairly prominent interaction for serotonin (Figure 11 (D)), even if it did not meet the 30% contact strength threshold. For both ligands it provided exclusively hydrophobic interactions, and while it was not a crucial amino acid, the mutational mapping (118) study did find that the mutation of PHE335 did have significant effects on the activity of SERT. With the mutational mapping study measuring SERT activity using a [^{3}H]-serotonin uptake assay as a percent of activity of their WT cells it shows clearly that the various mutations of PHE335 would give a significant reduction of SERT transport activity. The mutation of PHE335 also caused changes in (S)-citalopram potency in some cases (118), indicating that it does indeed play an important role for SERT during *in vitro* testing. A new mutational mapping study, this time focusing on the allosteric site, could have a look at some of the amino acids with higher interaction values (Figure 11 (C) and (D)). As the mutation of PHE335 did affect the overall activity of SERT it is possible that some of the other prominent amino acids in the allosteric site also play key roles in its activation process. Further computational studies could also yield greater insights, especially if longer MD simulations were performed in order to study potential structural shifts that could occur both for (S)-citalopram and for serotonin. If the simulation was substantially longer there would perhaps be a chance to see the ligand migrate from the allosteric site into the central site, such a migration would of great interest to study. Such a

migration is hypothesized to be possible (120), there are also indications for the presence of a ligand occupying the allosteric site being capable of blocking egress from the central site (53).

The interactions occurring within the allosteric site were once more mostly hydrophobic for (S)-citalopram and mostly H-bonds for serotonin, indicating that each respective ligand may have a greater propensity for certain types of interactions. (S)-citalopram has five amino acids (Figure 11 (C)) with interaction values above 0.1 that were exclusively or near exclusively hydrophobic, with the only amino acid with a noteworthy interaction value and no hydrophobic interactions being GLU494. Serotonin has four amino acids (Figure 11 (D)) with interaction values above 0.1 that are exclusively or near exclusively hydrophobic, but serotonin has an additional five amino acids that it interacts with. Only one of these possessing a small portion of hydrophobic interaction with the ligand, and two of the others are ASP328 and GLU494, having the highest interaction values for serotonin and both being within the 30% contact strength threshold. Another notable difference in interactions between (S)-citalopram and serotonin is the relative amounts of water bridge interactions, with serotonin having both more amino acids with water bridge interactions, but also higher interaction values for those amino acids (Figure 11 (D)). A contributing factor to this disparity could be the fact that serotonin has substantially more contacts throughout the allosteric site for the duration of its simulation when compared with (S)-citalopram (Figure 14 (C) and (D)). If a future computational study were to be carried out with a focus on protein-ligand interactions in the allosteric site the use of multiple allosteric inhibitors would be of interest to study, if there is a general propensity for hydrophobic interactions among inhibitors or if this is a feature unique to (S)-citalopram.

5.1.3 Central Site Amino Acid Movement

The measurements of distance between amino acids in the central site gave some insight into the fluctuations of the central site as a consequence of ligand placement and if the ligand was the substrate or an inhibitor. The first thing to note here is that there clearly was a difference between the four simulations in regards to the fluctuations between these amino acids (Figure 16-18), with some amino acid measurement pairs having more fluctuations than others. This means that the central binding site is affected by presence of a ligand, or lack thereof. If there was also performed a simulation for both crystal structures with no ligand in either site this would have been valuable in order to compare if the central site was affected by a ligand being in the allosteric site or simply from lack of a ligand in the central site. From the plots that were

made of the measurement data it is clearly possible to see that the movements are not identical between the simulations, but seeing if there is any significance hidden in the fluctuations would require a statistical model. The lack of a simulation without the presence of ligands in any of the binding sites as a control makes it so no certain conclusions can be drawn from these measurements other than that there is variation between each of the simulations. This is in itself an indication that the presence of the ligands may have an effect on the fluctuations of the central binding site.

5.2 Induced Fit Docking of Experimentally verified Inhibitors

The experimentally verified inhibitors that were selected for the docking portion of the study were all initially identified through a virtual screening approach (100). This virtual screening used the ligand-based methods, 2D fingerprinting and a 3D pharmacophore follow-up, in combination with a structural screening using a homology model of SERT (54) based on the LeuT crystal structure. The combined databases used for the virtual screening approach contained approximately 2.5 million compounds, the approach reduced this to 182 compounds that could then be selected to go on to the biological evaluation where they could be verified. The compounds made it through the virtual screening by having characteristics of known inhibitors for the ligand-based portion and for having good docking scores in the central site of the homology model for the structure-based portion. This in essence means that the compounds were in part selected for their high docking scores in the central site. It is therefore not very surprising that there was found to be a high positive correlation between a higher central site docking score and a lower allosteric site docking score (Figure 19). The correlation was equal to 0.8074, with a 1 meaning that the relationship would have been completely linear. An even more interesting observation is that the two docking scores are not very different. In fact, the average difference is only 1.948 lower for the allosteric site. Several related factors may contribute to this relationship between the docking scores of the allosteric and central sites. The first of these is that the pharmacophores that were developed for the ligand-based portion of the virtual screening were based on known substrates and inhibitors, some of which may have affected the allosteric site. The attributes of the compounds that affected the allosteric site would therefore be among the selecting factors during the ligand-based approach. The next factor relates to the structure of SERT and the proximities of the central and allosteric sites to one another. As these binding sites are quite close to one another, approximately 13 Å apart (53), it has been hypothesized that moving through the allosteric site is necessary to reach the central site (120). For a compound to be able to reach the central site it may therefore be necessary for it to have an affinity with the allosteric site above a certain threshold. Whichever may be the case there is however a strong correlation within this small sample group between high docking scores in the central site and lower docking scores in the allosteric site.

Which brings us the experimental data of the compounds that were selected for the docking study. It should be stated first that given the complexity of the activation of a transporter protein and the many factors of molecular interactions, it was not expected to find a correlation between the docking scores of compounds and their experimental K_i. The mutation of a single amino

acid such as PHE335 can alter the K_i of (S)-citalopram by several fold, as an example of just one such factor. A docking score of a compound while placed in the optimal position found by the IFD protocol is therefore not going to find any linear correlation between a K_i and docking score (Figure 20). The primary reason to perform the analysis between these two values was to show that there would be found no correlation between them. Which brings us the next data that was investigated for correlation, the relationship between docking score by site and (%) inhibition of SERT. In modern drug development it would be an immense boon if there existed a direct relationship between a compounds docking score in an orthosteric site and its (%) inhibition of that protein. Any compound with a high docking score would have a high (%) inhibition if there was a correlation of 1 between these two values. There is however no such direct correlation, and this we can see in Figure 21. There is however a slight tendency for compounds with higher docking scores in the central site to have a higher (%) inhibition. While the small sample size is a limiting factor for giving this statistical significance, it is a weak indication that compounds with high docking scores may also provide a high (%) inhibition. The issue is complicated by the fact that it is unknown whether these compounds bind to the central or allosteric binding site to enact their inhibitory effect upon SERT. The additional uncertainty this adds to the statistical relevance of even the small correlation that was found is therefore questionable. Nonetheless, there was found to be a small correlation between the central site docking scores of the experimentally verified compounds and their (%) inhibition for SERT at their optimal K_i.

5.2.1 Central Site

The interactions between the docked ligands and the central site were studied through the use of the Ligand Interactions Diagram (101), a part of the unified Maestro interface. The default settings were used to study which amino acids (S)-citalopram and each of the experimentally verified compounds interacted with in the central binding site (Appendix Figure 1-4). This data was then compiled into Table 2, where we have an overview of how many of the inhibitors interacted with each of the listed amino acids. As was mentioned in the results all but one of the crucial amino acids identified by a previous study (118) were found to interact with all or all but one of the inhibitors when docked into the central site. But there were also additional amino acids that were interacted with, amino acids that only saw small degrees of interactions with the ligands over the course of the MD simulations. These were ALA96, PHE335,

GLY338, THR439 and VAL501, which varied between small interactions with both central site ligands or no interactions with one of the ligands during the central site MD simulations. In the docking of SERT inhibitors they were however featured in every docking in the case of PHE335 and GLY338, and in all but one docking for the other three. This is indicative on these amino acids some effect on the optimal position of the ligand within the binding site. The mutational mapping study did investigate both PHE335 and THR439 for how they affected SERT activity, both of which significantly reduced SERT activity when mutated, with the degree of reduction varying between what amino acid they were replaced with. ALA96, GLY338 and VAL501 should be considered potential candidates for a mutational study to see how their mutation would affect SERT activity, alongside high interaction value central site amino acids from the MD simulation that have not been previously investigated. There was one additional molecule that was seen in all but one of the interaction overviews, the ion NA707. The importance of the Na^+ ions for the initiation of the translocation process for SERT has not been strictly determined, but there are indications that they do play an important role in this (121). To investigate how the Na^+ ion NA707 affected the docking scores of the compounds the dockings could be repeated without the presence of the two Na^+ ions and with the ligands fixed in their previous optimal positions.

5.2.2 Allosteric site

The interactions of the docked ligands in the allosteric site (Appendix, Figure 5-8) were studied in the same manner as with the central site and the interactions were compiled into Table 3. This gives an overview of which interactions occur between the ligands and the binding site. For the allosteric binding site however, there were only two amino acids (PHE335 and GLY498) that interacted with every ligand. The reason for the relatively few amino acids that interacted with every ligand, or all but one, is the outliers mentioned in the results. A visual inspection of the positioning of ligand 3, 10 and 15 showed that they had managed to find an optimal docking position partially inside the central site. The proximity of the allosteric and central sites does make this possible, as the binding site for the docking is defined as an area within a three-dimensional box centered on the native allosteric site ligand (S)-citalopram's position. By including amino acids that interacted with all or all but one ligands and removing the outliers, the amino acids that meet the updated criteria are now ten. Among these ten amino acids there were six (ARG104, ALA331, PHE335, GLU493, GLU494 and PHE556) that saw

small to large interactions with the ligands of the allosteric site simulations. The remaining four amino acids (GLY498, PRO499, LEU502 and ILE552) all saw negligible interactions throughout the simulations. The amino acids with negligible interactions may possibly not interact well in the allosteric site with the ligands that were used for the MD simulations. If additional simulations were carried out using the compounds of the docking study it would be seen if these ligands show a greater degree of interaction with some of the amino acids that had less interactions during the two allosteric site simulations. The disparity between (S)-citalopram and serotonin in terms of amino acids interacted with indicates that a wider set of ligands for MD simulations would perhaps have uncovered additional amino acids that showed high interaction values.

If a mutational study of the allosteric site was to be carried out, as mentioned in the MD simulations discussion, it would perhaps be of interest to include the amino acids found to have interactions with the majority of ligands docked as candidates. Although the above mentioned further MD simulations could first be carried out to attempt to uncover signs further indicating the importance of individual amino acids as way to narrow the number of amino acids that would be included in such a study. The additional MD simulations would also serve to reinforce the importance of any amino acid that is frequently seen to interact significantly over the course of the respective MD simulations. Using the mutational mapping study (118) as a backdrop it would also be interesting how the mutations they applied would affect an *in silico* model system of SERT. The mutation of the crucial amino acids would perhaps lead to the complete disruption of the sustained bindings that were observed during this project.

5.3 Experimental testing of potential GABA$_B$ receptor antagonists

The potential GABA$_B$ receptor antagonists that were experimentally tested using a functional cAMP assay were chosen from a list of compounds that had been identified through a combined ligand and structure based virtual screening approach (Evenseth et al. 2018, unpublished work). The ligand based section was based on the structure of 55 known GABA$_B$ receptor ligands, 42 agonists and 13 antagonists, that were clustered to separate agonists from antagonists, giving a total of 6 clusters after manual modification. Pharmacophores were created based on these clusters, but only one pharmacophore per cluster was validated as of sufficient quality. The validated pharmacophore models were used to screen a compound database of five million compounds, the hits from this screening were then docked into the VFT domain of the GABA$_B$ receptor. Their MM-GBSA scores were calculated and a selection of the compounds was brought forwards to the *in vitro* testing stage. The first initial screening to be performed on a selection of these compounds was carried out as part of the present project.

As this was an initial assay screening that was not performed in a triplicate of assays, the current results might be affected by experimental errors. However, the assay results may still serve as an indication of the test compounds' properties. With the study of these compounds still being ongoing and the work related to them unpublished, the compounds themselves are not the focus of this portion of the project, rather the focus is on the methods used to test the compounds and also on the experimental errors that may affect the results of the assay.

Different factors during the functional assay may contribute to the deviation of assay results. The following analysis will therefore focus on the major factors. The first part of the assay procedure where error can be introduced is contamination during the cell culture stage. This could be bacterial or fungal, especially the WT cell line that used culture medium without adding antibiotics. The GABA$_B$ cell lines medium does contain antibiotics ensures that the cells continue to actively express the GABA$_B$ receptor, it also protects the cells to a certain degree from bacterial contamination. Both cell lines are however highly vulnerable to fungal contamination and proper sterilization and care during the cell growth stages are therefore crucial.

Other issues can be using PBS or culture medium that is too old. Both can have adverse effects on the cells and thereby potentially affect the cell behavior during the assay. During the cell growth and harvesting stages, cell quality can be affected by leaving cells in AssayComplete™ Cell Detachment Reagent for longer than 2 minutes. Although the exact mechanism is

unknown, the unstable results and reduced GABA$_B$ receptor activity were observed from an earlier experiment. Since ethanol can inhibit forskolin-activated adenylate cyclase (122), ethanol contamination during the cell harvesting stage can therefore create a false agonist like activity in the assay results or reduced forskolin stimulation. Cell aggregation in the cell suspension is the major problem causing incorrect cell counting. After centrifugation, when the cell pellet is not resuspended properly, the counted cell number could be lower than what it should be. Even worse is, the aggregated cells could resuspend into the cell suspension during the dilution step, and therefore result in the calculated cell concentration being lower than the actual assay cell concentration. If the assay is to be repeated multiple times, then the cell concentration deviation between assay repetitions could cause deviation in the final assay result. This would affect the repeatability of the assay.

The second part of the assay procedure where error can be introduced has to do with the mixing of buffers, stock solutions, the addition of compounds, essentially anything requiring measurements and/or pipetting. Human error and errors introduced by assay equipment, such as a not properly calibrated pipette are also important factors. The volume being pipetted by a pipette is very accurate if it has been properly calibrated, but it is not 100% precise. Following procedure can improve the pipetting accuracy: precalibrate the pipette, avoiding bubbles, wetting the inside wall of pipette before pipetting larger volumes, change the tip during serial dilution of GABA and also checking for liquid stuck on the outside of the pipette. When making stock solutions and serial dilution it is also better to pipette larger volumes. If a mixture requires a 9:1 ratio, then mixing 90µl+10µl provides a larger room for error than if 900µl+100µl was pipetted.

In the creation of stock solutions and assay buffers there is also a need to weigh the various compounds that are to be dissolved. A mistake here, for instance having some of the compound end up on the scale instead of in the measurement container, would introduce incorrect solute concentration. With a test compound this can cause a wrong concentration being used in every assay. If however the mistake is with one of the components of the assay buffer, such as calcium concentration which is known to affect GABA$_B$ receptor activity, this will then affect the measured test compound activity also (123, 124). Therefore, prioritizing caution and patience in these steps can introduce lower errors. Ensure the scales are clean, be careful not to spill any of the compounds/assay buffer components outside their measurement containers. When dealing with powders be aware of static. If the compound is somehow contaminated it is better to throw it away and try again.

For the assay that was performed with the 10mM stock compounds solution, there were several time points when error may have been introduced when pipetting. These were when mixing the previously frozen stock solutions of test compounds in preparation for the assay, the volume to be pipetted was only 6 µl, which went into 994 µl of the appropriate buffer. Even a small deviation in volume for the test compound would have affected the final concentration of the compound used in the assay quite significantly. There was also room for error when creating the forskolin stock in a 1:1 mix, it is better to use the same stock forskolin solution for each test compound. Different concentrations of forskolin can result in different maximum cAMP formation in the assay. This can be easily observed from the results of the forskolin control (Figure 22). The difference of forskolin in the controls in Figure 22 could be caused by pipetting error of forskolin stock into the 5G control. A visual inspection of the leftovers in the compound preparation tube confirmed this on the day of the assay. For the $GABA_B$ cells there was also the additional issue of adding GABA into forskolin and test compound mixture. The additional pipetting leaves room for more potential errors. Then there is the cell numbers, which are affected by pipetting even after cell counts have been correctly performed. Immediately after the cells have been diluted to the appropriate concentration of 0.5 million cells/ml, cells were incubated for 1-hour incubation (Figure 7) at 18°C. At this stage there is also the ever-present risk of pipetting the wrong volume. Since cells rapidly sink towards the bottom of the tray they are being pipetted from, it is therefore necessary to perform a quick premix (such as shaking the tray or pipetting up and down two times) of the cell suspension before transferring diluted cells into the incubation tubes. If the cells are allowed to sink to the bottom of the tray without being mixed again before the next round of pipetting, then there would be a higher cell concentration at the bottom of the solution in the tray than near the top.

The main issue with pipetting at the final stages of the assay is that the pipetting must happen quickly because the reaction begins immediately upon the liquids mixing, a difficult task complicated further by the entire procedure being performed in the dark because of the reagents being light sensitive. The issue of quickly pipetting can be solved in part through the use of Eppendorf Multipette M4, a positive displacement high precision pipette. This pipette is especially suitable to viscous or volatile liquids, such as the antibody solution in our assay. In addition, it is important to carefully pre-plan where to pipette each of the reagents, in a manner where the tip of the pipette would not be contaminated by droplets from the pipetting of previous reagents or cell-compound mixture in the incubation tube. During the assay this was solved with pipetting the compounds against the wall of the wells at different heights and sides,

starting at a low angle and working upwards. Fast and accurate pipetting in the dark raised an additional challenge during the assay. With proper planning it is possible to determine how many wells can be completed with a Multipette M4 pipette for each of the reagents. The darkness can make it difficult to determine if there is any liquid inside the pipette tip, a potential solution here is to carefully check the reagent inside the tips, as much as the low light conditions allow.

Light exposure is the last factor that could introduce experimental error. As the reagents used to measure the cAMP levels are light sensitive, exposure to light will severely affect the results of the assay. Which is why after adding the final reagent, Solution A, it is a good precaution to cover the entire assay setup (Figure 7) in aluminum foil to prevent light exposure during the 12-14 hour long incubation. The next stage where there is a danger of light exposure is when the transfer from the assay setup to the clear bottom 384-well assay plate is occurring. Light exposure at this stage can come from many sources, the best option to avoid it is to minimize the opportunity for unexpected light to occur. During transport to the ClarioStar® plate reader for the luminescence readout the plate should be covered in aluminum foil. Before the assay plate is put into the plate reader, turn off the light in the room where the plate reader is located.

During the performance of the assays for the project, there were two particularly notable things that were done to minimize the chance of assay failure. The first is to perform the assay at 18°C to slow the reactions occurring within the incubation tubes. This slight reduction in temperature prolongs the reaction time, and therefore gives more time to add both test compounds and reagents without adversely affecting the results of the assay. Submerging incubation tubes in pre-cooled water baths ensures that each assay is carried out at the same temperature. An assay carried out at room temperature during winter may be carried out at 19°C, while a follow-up assay that is carried out during the summer might have a temperature of 25°C. Earlier experiments showed that forskolin stimulated cAMP formation in our GABA$_B$ cells reached maximum level at a lower speed at lower temperature (data not shown). The significant degree of temperature variation would affect the cell reactivity and could lead to the assay results showing significant divergence from one another. Keeping a constant temperature for assays is a good step to minimize the variation between assays.

The second notable thing is related to ensuring sufficient cells for performing the assay. The initial cell culture harvests that were carried out tended to have a somewhat low cell counts considering the quantity of cells that should have been in a T75 flask at ~70% confluency. The

difference in total cell count can be seen quite clearly on Figure 22 and Figure 23. The WT assay which was performed first had roughly 2.93 million cells after the harvesting was complete, while the GABA$_B$ cells had 6 million cells. This was not the first cell harvesting, as there had been done several times prior to the assay as practice and the results from those trial runs were fairly consistent with that of the WT assay. There are numerous other factors that may have contributed to this sizable difference and the benefit of washing four times has been proven to give better assay results, perhaps due to reducing the nonspecific binding of GABA to leftover culture medium components.

The purpose of this part of the project was to gain a greater understanding of the laboratory work, experimental techniques and problem solving in a laboratory environment. The second purpose was the results of the assay, gain an indication of the test compounds status as a GABA$_B$ receptor antagonist. The results of the assay did give an indication that several of the compounds may act as GABA$_B$ receptor agonists, but further multiple repeats of the initial screening assay as well as the effect of selected potential hit compounds on GABA dose-response are required to confirm and characterize any hit compounds. The ones that are confirmed to affect the GABA$_B$ receptor would then have to be tested at multiple concentrations in order to establish dose-response curves.

6. Conclusions

The recently rendered X-ray crystal structure of SERT was used for a computational study investigating the central and allosteric binding sites of SERT. Emphasis was placed on which amino acids most heavily interacted with the ligands and differences between the results of the MD simulation and docking study of experimentally verified SERT inhibitors was explored. Each simulation identified key amino acids between each respective ligand and binding site. For the central site simulation it was seen that both (S)-citalopram and serotonin had very strong interactions with TYR95, and serotonin having more numerous strong interactions than (S)-citalopram. For the allosteric site simulations they identified PHE335 and GLU494 as the strongest interacting partners for (S)-citalopram and serotonin respectively. The interactions between the ligands and the allosteric site were also not as strong overall as the interactions of the central site. From the combined data of the MD simulations and the docking study numerous amino acids were identified as potential targets of a mutation mapping study looking at effects on SERT activity. Most of the amino acids that were of interest for such as study were located in the allosteric site (ARG104, ASP328, ALA331, GLU493, GlU494, GLY498, PRO499, LEU502 ILE552 and PHE556), as the central site has been partially investigated though such methods in the past. The MD simulation also found that (S)-citalopram and serotonin each had a favored interaction type regardless of binding site, hydrophobic interactions for (S)-citalopram and H-bonds for serotonin. Serotonin was also observed to maintain more interactions on average throughout the simulation than (S)-citalopram did, regardless of binding site. The experimental portion of the study successfully carried out an initial screening of ten compounds identified as potential GABA$_B$ receptor antagonists. None of the tested compounds showed any significant indications of antagonistic effect on the GABA$_B$ receptor, but three of the compounds did show indications of being GABA$_B$ receptor agonists. The compounds identified as potentially being agonists require further assays to confirm that their perceived activity was not caused by experimental error. The experimental portion of the study also provided experience working in a laboratory environment and proficiency in performing the methodologies required for a functional cAMP assay.

7. References

1.   Eric P. Widmaier HR, Kevin T. Strang. Vander's Human Physiology. Boston: McGraw-Hill Higher Education; 2014.
2.   Hof PR, Trapp BD, de Vellis J, Claudio L, Colman DR. CHAPTER 1 - Cellular Components of Nervous Tissue A2 - BYRNE, JOHN H. In: Roberts JL, editor. From Molecules to Networks. Burlington: Academic Press; 2004. p. 1-12.
3.   Neuroscience. Third Edition ed. Purves D AG, Fitzpatrick D, Hall WC, LaMantia AS, McNamara JO, Williams SM, editor. Sunderland (MA): Sinauer Associates; 2001.
4.   Sarter M, Bruno JP, Parikh V. Abnormal neurotransmitter release underlying behavioral and cognitive disorders: toward concepts of dynamic and function-specific dysregulation. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology. 2007;32(7):1452-61.
5.   Greenshaw AJ. Neurotransmitter interactions in psychotropic drug action: beyond dopamine and serotonin. Journal of Psychiatry and Neuroscience. 2003;28(4):247-50.
6.   Kalia M. Neurobiological basis of depression: an update. Metabolism: clinical and experimental. 2005;54(5 Suppl 1):24-7.
7.   Kennedy SH. Core symptoms of major depressive disorder: relevance to diagnosis and treatment. Dialogues in Clinical Neuroscience. 2008;10(3):271-7.
8.   Hirschfeld RM. History and evolution of the monoamine hypothesis of depression. The Journal of clinical psychiatry. 2000;61 Suppl 6:4-6.
9.   Parker G. Differential effectiveness of newer and older antidepressants appears mediated by an age effect on the phenotypic expression of depression. Acta psychiatrica Scandinavica. 2002;106(3):168-70.
10.   Stewart JW, Thase ME. Treating DSM-IV depression with atypical features. The Journal of clinical psychiatry. 2007;68(4):e10.
11.   Ravindran AV, da Silva TL. Complementary and alternative therapies as add-on to pharmacotherapy for mood and anxiety disorders: a systematic review. Journal of affective disorders. 2013;150(3):707-19.
12.   Penn E, Tracy DK. The drugs don't work? Antidepressants and the current and future pharmacological management of depression. Therapeutic Advances in Psychopharmacology. 2012;2(5):179-88.
13.   Ferguson JM. SSRI Antidepressant Medications: Adverse Effects and Tolerability. Primary Care Companion to The Journal of Clinical Psychiatry. 2001;3(1):22-7.
14.   Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ. Principles of Neural Science. 2012. New York: McGraw-Hill. 5th.
15.   McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A. Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. The American journal of psychiatry. 2017;174(7):676-85.
16.   White KJ, Walline CC, Barker EL. Serotonin transporters: Implications for antidepressant drug development. The AAPS Journal. 2005;7(2):E421-33.
17.   Cryan JF, Kaupmann K. Don't worry 'B' happy!: a role for GABA(B) receptors in anxiety and depression. Trends in pharmacological sciences. 2005;26(1):36-43.
18.   Lin Z, Canales JJ, Björgvinsson T, Thomsen MM, Qu H, Liu Q-R, et al. Monoamine Transporters: Vulnerable and Vital Doorkeepers. Progress in Molecular Biology and Translational Science. 2011;98:1-46.
19.   Torres GE, Gainetdinov RR, Caron MG. Plasma membrane monoamine transporters: structure, regulation and function. Nature Reviews Neuroscience. 2003;4:13.

20.   Eiden LE, Schafer MK, Weihe E, Schutz B. The vesicular amine transporter family (SLC18): amine/proton antiporters required for vesicular accumulation and regulated exocytotic secretion of monoamines and acetylcholine. Pflugers Archiv : European journal of physiology. 2004;447(5):636-40.

21.   Schuldiner S. A molecular glimpse of vesicular monoamine transporters. Journal of neurochemistry. 1994;62(6):2067-78.

22.   Gasnier B. The loading of neurotransmitters into synaptic vesicles. Biochimie. 2000;82(4):327-37.

23.   Henry JP, Botton D, Sagne C, Isambert MF, Desnos C, Blanchard V, et al. Biochemistry and molecular biology of the vesicular monoamine transporter from chromaffin granules. The Journal of experimental biology. 1994;196:251-62.

24.   Wimalasena K. Vesicular Monoamine Transporters: Structure-Function, Pharmacology, and Medicinal Chemistry. Medicinal research reviews. 2011;31(4):483-519.

25.   Kurian MA, Zhen J, Cheng SY, Li Y, Mordekar SR, Jardine P, et al. Homozygous loss-of-function mutations in the gene encoding the dopamine transporter are associated with infantile parkinsonism-dystonia. The Journal of clinical investigation. 2009;119(6):1595-603.

26.   Bröer S, Gether U. The solute carrier 6 family of transporters. British Journal of Pharmacology. 2012;167(2):256-78.

27.   Tanford C. Mechanism of free energy coupling in active transport. Annual review of biochemistry. 1983;52:379-409.

28.   Gabrielsen M. Structure, function and inhibition of the serotonin transporter studied by molecular docking, -dynamics and virtual screening. Tromsø: University of Tromsø UIT; 2011.

29.   Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology. 2000;302(1):205-17.

30.   Page RD. TreeView: an application to display phylogenetic trees on personal computers. Computer applications in the biosciences : CABIOS. 1996;12(4):357-8.

31.   Hahn MK, Blakely RD. The functional impact of SLC6 transporter genetic variation. Annual review of pharmacology and toxicology. 2007;47:401-41.

32.   Ramamoorthy S, Shippenberg TS, Jayanthi LD. Regulation of monoamine transporters: Role of transporter phosphorylation. Pharmacology & therapeutics. 2011;129(2):220-38.

33.   Wallimann T, Tokarska-Schlattner M, Schlattner U. The creatine kinase system and pleiotropic effects of creatine. Amino acids. 2011;40(5):1271-96.

34.   Javitt DC. Glycine transport inhibitors for the treatment of schizophrenia: symptom and disease modification. Current opinion in drug discovery & development. 2009;12(4):468-78.

35.   Mager S, Sloan J. Possible role of amino acids, peptides, and sugar transporter in protein removal and innate lung defense. European journal of pharmacology. 2003;479(1-3):263-7.

36.   Broer S. Apical transporters for neutral amino acids: physiology and pathophysiology. Physiology (Bethesda, Md). 2008;23:95-103.

37.   Russo S, Kema IP, Bosker F, Haavik J, Korf J. Tryptophan as an evolutionarily conserved signal to brain serotonin: molecular evidence and psychiatric implications. The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry. 2009;10(4):258-68.

38.   Murphy DL, Lerner A, Rudnick G, Lesch KP. Serotonin transporter: gene, genetic disorders, and pharmacogenetics. Molecular interventions. 2004;4(2):109-23.

39.   Zhou Z, Zhen J, Karpowich NK, Law CJ, Reith MEA, Wang DN. Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures. Nature structural & molecular biology. 2009;16(6):652-7.

40.   aan het Rot M, Mathew SJ, Charney DS. Neurobiological mechanisms in major depressive disorder. Canadian Medical Association Journal. 2009;180:305-13.

41.   Rudnick G. Active transport of 5-hydroxytryptamine by plasma membrane vesicles isolated from human blood platelets. The Journal of biological chemistry. 1977;252(7):2170-4.

42.   Gu HH, Wall S, Rudnick G. Ion coupling stoichiometry for the norepinephrine transporter in membrane vesicles from stably transfected cells. The Journal of biological chemistry. 1996;271(12):6911-6.

43.   Rudnick G. Bioenergetics of neurotransmitter transport. Journal of bioenergetics and biomembranes. 1998;30(2):173-85.

44.   Rudnick G. Serotonin transporters--structure and function. The Journal of membrane biology. 2006;213(2):101-10.

45.   Nelson PJ, Rudnick G. Coupling between platelet 5-hydroxytryptamine and potassium transport. The Journal of biological chemistry. 1979;254(20):10084-9.

46.   Zhang YW, Rudnick G. The cytoplasmic substrate permeation pathway of serotonin transporter. The Journal of biological chemistry. 2006;281(47):36213-20.

47.   Cool DR, Leibach FH, Ganapathy V. Modulation of serotonin uptake kinetics by ions and ion gradients in human placental brush-border membrane vesicles. Biochemistry. 1990;29(7):1818-22.

48.   Humphreys CJ, Wall SC, Rudnick G. Ligand Binding to the Serotonin Transporter: Equilibria, Kinetics, and Ion Dependence. Biochemistry. 1994;33(31):9118-25.

49.   Jacobs MT, Zhang YW, Campbell SD, Rudnick G. Ibogaine, a noncompetitive inhibitor of serotonin transport, acts by stabilizing the cytoplasm-facing state of the transporter. The Journal of biological chemistry. 2007;282(40):29441-7.

50.   Tavoulari S, Forrest LR, Rudnick G. Fluoxetine (Prozac) Binding to Serotonin Transporter Is Modulated by Chloride and Conformational Changes. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009;29(30):9635-43.

51.   Lesch KP, Gross J, Franzek E, Wolozin BL, Riederer P, Murphy DL. Primary structure of the serotonin transporter in unipolar depression and bipolar disorder. Biological psychiatry. 1995;37(4):215-23.

52.   Rudnick G. Structure/function relationships in serotonin transporter: new insights from the structure of a bacterial transporter. Handbook of experimental pharmacology. 2006(175):59-73.

53.   Coleman JA, Green EM, Gouaux E. X-ray structures and mechanism of the human serotonin transporter. Nature. 2016;532(7599):334-9.

54.   Gabrielsen M, Ravna AW, Kristiansen K, Sylte I. Substrate binding and translocation of the serotonin transporter studied by docking and molecular dynamics simulations. Journal of molecular modeling. 2012;18(3):1073-85.

55.   Matthaus F, Haddjeri N, Sanchez C, Marti Y, Bahri S, Rovera R, et al. The allosteric citalopram binding site differentially interferes with neuronal firing rate and SERT trafficking in serotonergic neurons. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2016;26(11):1806-17.

56.   Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature. 2009;459(7245):356-63.

57.   Chun L, Zhang W, Liu J. Structure and ligand recognition of class C GPCRs. Acta Pharmacologica Sinica. 2012;33(3):312-23.

58.   Schoneberg T, Schulz A, Biebermann H, Hermsdorf T, Rompler H, Sangkuhl K. Mutant G-protein-coupled receptors as a cause of human diseases. Pharmacology & therapeutics. 2004;104(3):173-206.

59.   Pin JP, Galvez T, Prezeau L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. Pharmacology & therapeutics. 2003;98(3):325-54.

60.   Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Molecular pharmacology. 2003;63(6):1256-72.

61.   Attwood TK, Findlay JB. Fingerprinting G-protein-coupled receptors. Protein engineering. 1994;7(2):195-203.

62.   Kolakowski LF, Jr. GCRDb: a G-protein-coupled receptor database. Receptors & channels. 1994;2(1):1-7.

63.   Williams C, Hill SJ. GPCR signaling: understanding the pathway to successful drug discovery. Methods in molecular biology (Clifton, NJ). 2009;552:39-50.

64.   Rondard P, Goudet C, Kniazeff J, Pin JP, Prezeau L. The complexity of their activation mechanism opens new possibilities for the modulation of mGlu and GABAB class C G protein-coupled receptors. Neuropharmacology. 2011;60(1):82-92.

65.   Urwyler S. Allosteric modulation of family C G-protein-coupled receptors: from molecular insights to therapeutic perspectives. Pharmacological Reviews. 2011;63(1):59-126.

66.   Heaney CF, Kinney JW. Role of GABA(B) receptors in learning and memory and neurological disorders. Neuroscience and biobehavioral reviews. 2016;63:1-28.

67.   Froestl W, Gallagher M, Jenkins H, Madrid A, Melcher T, Teichman S, et al. SGS742: the first GABA(B) receptor antagonist in clinical trials. Biochemical pharmacology. 2004;68(8):1479-87.

68.   Freyd T, Warszycki D, Mordalski S, Bojarski AJ, Sylte I, Gabrielsen M. Ligand-guided homology modelling of the GABA(B2) subunit of the GABA(B) receptor. PLoS ONE. 2017;12(3).

69.   Hyland NP, Cryan JF. A Gut Feeling about GABA: Focus on GABA(B) Receptors. Frontiers in Pharmacology. 2010;1: 124.

70.   Xu C, Zhang W, Rondard P, Pin JP, Liu J. Complex GABA(B) receptor complexes: how to generate multiple functionally distinct units from a single receptor. Frontiers in Pharmacology. 2014;5: 12.

71.   Gassmann M, Bettler B. Regulation of neuronal GABA(B) receptor functions by subunit composition. Nature reviews Neuroscience. 2012;13(6):380-94.

72.   Padgett CL, Slesinger PA. GABAB receptor coupling to G-proteins and ion channels. Advances in pharmacology (San Diego, Calif). 2010;58:123-47.

73.   Conn PJ, Lindsley CW, Meiler J, Niswender CM. Opportunities and challenges in the discovery of allosteric modulators of GPCRs for treating CNS disorders. Nature reviews Drug discovery. 2014;13(9):692-708.

74.   Onfroy L, Galandrin S, Pontier SM, Seguelas MH, N'Guyen D, Sénard JM, et al. G protein stoichiometry dictates biased agonism through distinct receptor-G protein partitioning. Scientific Reports. 2017;7.

75.   Sturchler E, Li X, de Lourdes Ladino M, Kaczanowska K, Cameron M, Griffin PR, et al. GABA(B) receptor allosteric modulators exhibit pathway-dependent and species-selective activity. Pharmacology Research & Perspectives. 2017;5(2).

76.   Ghose S, Winter MK, McCarson KE, Tamminga CA, Enna SJ. The GABA(B) receptor as a target for antidepressant drug action. British Journal of Pharmacology. 2011;162(1):1-17.

77.   Waldman SA, Terzic A. Systems-based discovery advances drug development. Clinical pharmacology and therapeutics. 2013;93(4):285-7.

78.   DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. Journal of health economics. 2003;22(2):151-85.

79.   Heath G, Colburn WA. An evolution of drug development and clinical pharmacology during the 20th century. Journal of clinical pharmacology. 2000;40(9):918-29.

80.   Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational Methods in Drug Discovery. Pharmacological Reviews. 2014;66(1):334-95.

81.   Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular dynamics simulations: advances and applications. Advances and Applications in Bioinformatics and Chemistry : AABC. 2015;8:37-47.

82.   Forster MJ. Molecular modelling in structural biology. Micron. 2002;33(4):365-84.

83.   McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977;267(5612):585-90.

84.   Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. Journal of molecular biology. 1976;103(2):227-49.

85.   van der Kamp MW, Mulholland AJ. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. Biochemistry. 2013;52(16):2708-28.

86.   Senn HM, Thiel W. QM/MM Methods for Biomolecular Systems. Angewandte Chemie International Edition. 2009;48(7):1198-229.

87.   Höltje HD, Sippl W, Rognan D, Folkers G. Molecular Modeling: Basic Principles and Applications: Wiley; 2008.

88.   Cherrak Z, Lagant P, Benharrats N, Semmoud A, Hamdache F, Vergoten G. Density functional theory and empirical derived force fields for the delocalized polaron form of polyaniline. Application to properties determination of an isolated oligomer using molecular dynamics simulations. Spectrochimica acta Part A, Molecular and biomolecular spectroscopy. 2005;61(7):1419-29.

89.   Miguet L, Zhang Z, Barbier M, Grigorov MG. Comparison of a homology model and the crystallographic structure of human 11beta-hydroxysteroid dehydrogenase type 1 (11betaHSD1) in a structure-based identification of inhibitors. Journal of computer-aided molecular design. 2006;20(2):67-81.

90.   Krishnan VV, Rupp B. Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy. eLS. 2012.

91.   Yee AA, Savchenko A, Ignachenko A, Lukin J, Xu X, Skarina T, et al. NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. Journal of the American Chemical Society. 2005;127(47):16512-7.

92.   Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic acids research. 2000;28(1):235-42.

93.   Carroni M, Saibil HR. Cryo electron microscopy to determine the structure of macromolecular complexes. Methods (San Diego, Calif). 2016;95:78-85.

94.   Venien-Bryan C, Li Z, Vuillard L, Boutin JA. Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discoveryThis paper is an extension of Boutin et al. [(2016), Med. Sci. (Paris), 32, 758-767]. Acta Crystallographica Section F. 2017;73(4):174-83.

95.   Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. Indian Journal of Pharmaceutical Sciences. 2012;74(1):1-17.

96.   Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature structural biology. 2002;9(9):646-52.

97.   Leach AR. Molecular Modelling. Principles and applications. 2nd ed. Edinburgh Gate: Pearson Education Limited; 2001.

98.   Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. Journal of medicinal chemistry. 2006;49(2):534-53.

99.   Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Physical chemistry chemical physics : PCCP. 2010;12(40):12899-908.

100.   Gabrielsen M, Kurczab R, Siwek A, Wolak M, Ravna AW, Kristiansen K, et al. Identification of Novel Serotonin Transporter Compounds by Virtual Screening and Experimental Verification. Journal of Chemical Information and Modeling. 2014;54(3):933-43.

101.   Schrödinger Release 2017-2. Maestro. Schrödinger, LLC: New York, NY; 2017.

102.   Schrödinger Release 2017-2: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2016; Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY. 2017.

103.   Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. Journal of computer-aided molecular design. 2013;27(3):221-34.

104.   Jacobson MP, Friesner RA, Xiang Z, Honig B. On the role of the crystal environment in determining protein side-chain conformations. Journal of molecular biology. 2002;320(3):597-608.

105.   Schrödinger Release 2017-2: LigPrep, Schrödinger, LLC, New York, NY. 2017.

106.   Schrödinger Release 2017-2: BioLuminate, Schrödinger, LLC, New York, NY. 2017.

107.   Schrödinger Release 2017-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2017. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY. 2017.

108.   Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of medicinal chemistry. 2004;47(7):1739-49.

109.   Schrödinger Release 2017-2: Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY. 2017.

110.   Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: orientations of proteins in membranes database. Bioinformatics (Oxford, England). 2006;22(5):623-5.

111.   UniProt: the universal protein knowledgebase. Nucleic acids research. 2017;45(D1):D158-D69.

112.   Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic acids research. 2017;45(Database issue):D945-54.

113.   Bøgwald I. Binding mode of novel multimodal serotonin transporter compounds in 5-hydroxytryptaminereceptors. Tromsø: The Arctic University of Norway; 2016.

114.   Gabrielsen M, Kurczab R, Ravna AW, Kufareva I, Abagyan R, Chilmonczyk Z, et al. Molecular mechanism of serotonin transporter inhibition elucidated by a new flexible docking protocol. European journal of medicinal chemistry. 2012;47(1):24-37.

115.   Schrödinger Release 2017-2: Epik, Schrödinger, LLC, New York, NY. 2017.

116.   Schrödinger Release 2017-2: Prime, Schrödinger, LLC, New York, NY. 2017.

117.   Ravna AW, Sylte I, Dahl SG. Molecular mechanism of citalopram and cocaine interactions with neurotransmitter transporters. The Journal of pharmacology and experimental therapeutics. 2003;307(1):34-41.

118.   Andersen J, Olsen L, Hansen KB, Taboureau O, Jørgensen FS, Jørgensen AM, et al. Mutational Mapping and Modeling of the Binding Site for (S)-Citalopram in the Human Serotonin Transporter. The Journal of biological chemistry. 2010;285(3):2051-63.

119.   Chen F, Larsen MB, Neubauer HA, Sanchez C, Plenge P, Wiborg O. Characterization of an allosteric citalopram-binding site at the serotonin transporter. Journal of neurochemistry. 2005;92(1):21-8.

120.   Jaronczyk M, Chilmonczyk Z, Mazurek AP, Nowak G, Ravna AW, Kristiansen K, et al. The molecular interactions of buspirone analogues with the serotonin transporter. Bioorganic & medicinal chemistry. 2008;16(20):9283-94.

121.   Felts B, Pramod AB, Sandtner W, Burbach N, Bulling S, Sitte HH, et al. The Two Na(+) Sites in the Human Serotonin Transporter Play Distinct Roles in the Ion Coupling and Electrogenicity of Transport. The Journal of biological chemistry. 2014;289(3):1825-40.

122.   Huang RD, Smith MF, Zahler WL. Inhibition of forskolin-activated adenylate cyclase by ethanol and other solvents. Journal of cyclic nucleotide research. 1982;8(6):385-94.

123.   Wise A, Green A, Main MJ, Wilson R, Fraser N, Marshall FH. Calcium sensing properties of the GABA(B) receptor. Neuropharmacology. 1999;38(11):1647-56.

124.   Chalifoux JR, Carter AG. GABA(B) receptor modulation of voltage-sensitive calcium channels in spines and dendrites. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2011;31(11):4221-32.
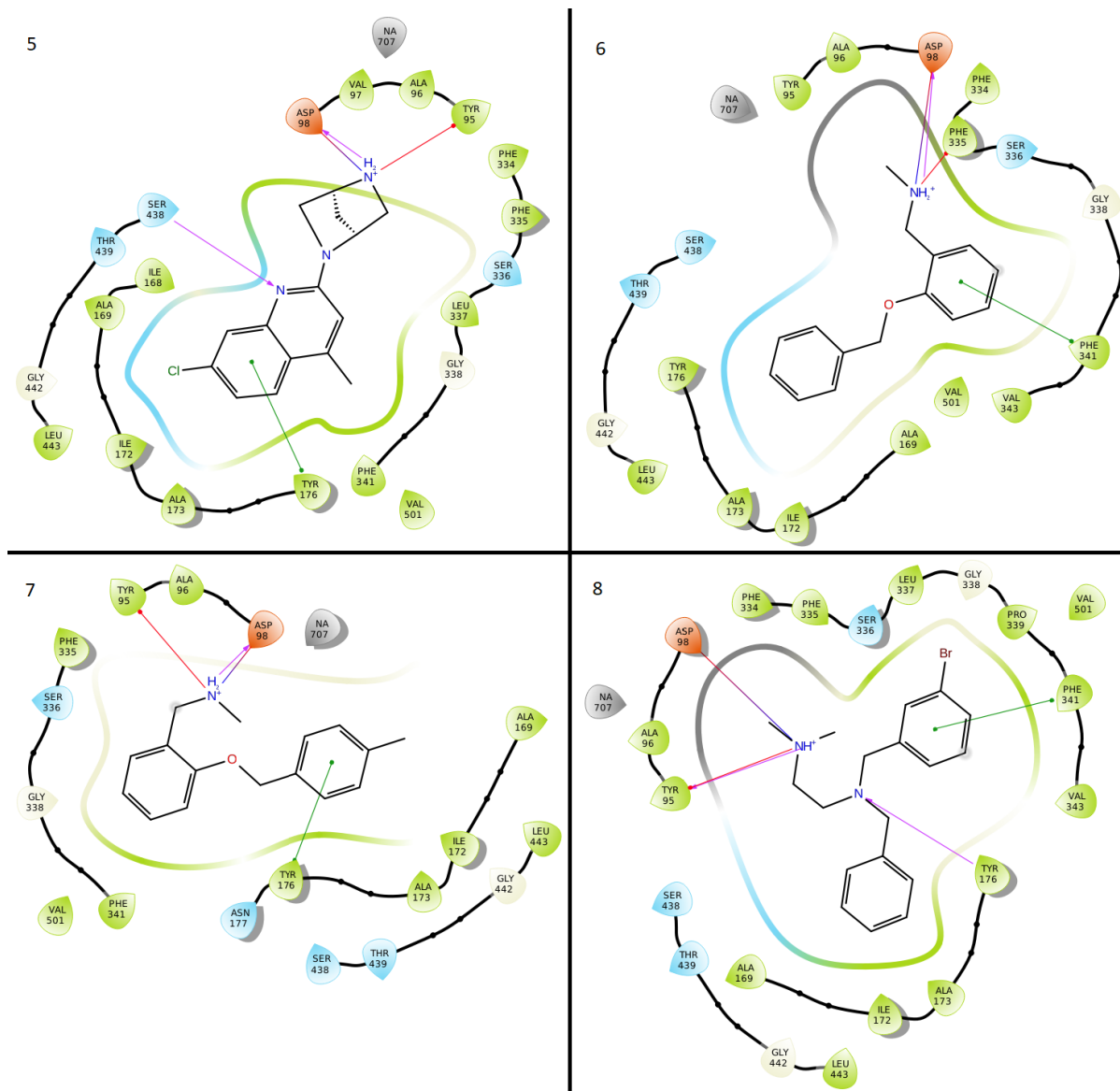
8. Appendix

8.1 Induced Fit Docking

Appendix Table 1: The table below show additional information regarding the experimentally verified (100) compounds that were selected for the *in silico* docking study. The ID refers to the compounds ID number, which is used to keep the compounds organized and sorted. The Compound Name refers to the name of the compounds in their respective vendor database, Vendor refers to the company that the compound was acquired from. Cluster refers to what cluster the compound was placed in during the virtual screening process before it was experimentally tested. Path refers to the method by which the compound was discovered to possibly be an inhibitory compound of SERT before the experimental verification. K_i [nM] ± SEM is the necessary K_i in nM for the compound to exhibit an optimal inhibitory effect, the standard error of mean is also listed. The % is the percent of inhibition the compound was found to provide during in vitro testing with the optimal concentration of the inhibitory compound in a competitive binding assay with [^{3}H]-citalopram.

| ID	Compound Name	Vendor	Cluster	Path	Ki [nM] ± SEM	%
1	ASN13153175	AS	C01	H2L	90.5 ± 8.7	69
2	T5544147	EN	C02	VS	336 ± 41.5	46
3	5417988	CB	C03	H2L	926 ± 57	33
4	T6125232	EN	C04	H2L	1.5 ± 0.3	54
5	46524932	CB	C05	VS	152 ± 12.7	72
6	9029307	CB	C06	VS	42.5 ± 1.3	49
7	9066608	CB	C06	VS	28.4 ± 0.3	30
8	5458751	CB	C07	VS	50 ± 1.7	52
9	BAS12193322	AS	C08	H2L	22.3 ± 2.8	35
10	C466-0145	CD	C09	VS	86.1 ± 8.1	63
11	T6406455	EN	C10	H2L	127 ± 4.5	30
12	T5777260	EN	C10	H2L	129 ± 9.8	47
13	T6574432	EN	C11	VS	56 ± 8.2	45
14	T0502-9459	EN	C12	VS	268 ± 16	40
15	EN300-08612	EN	C13	VS	322 ± 29	39
16	Citalopram		N/A	N/A		
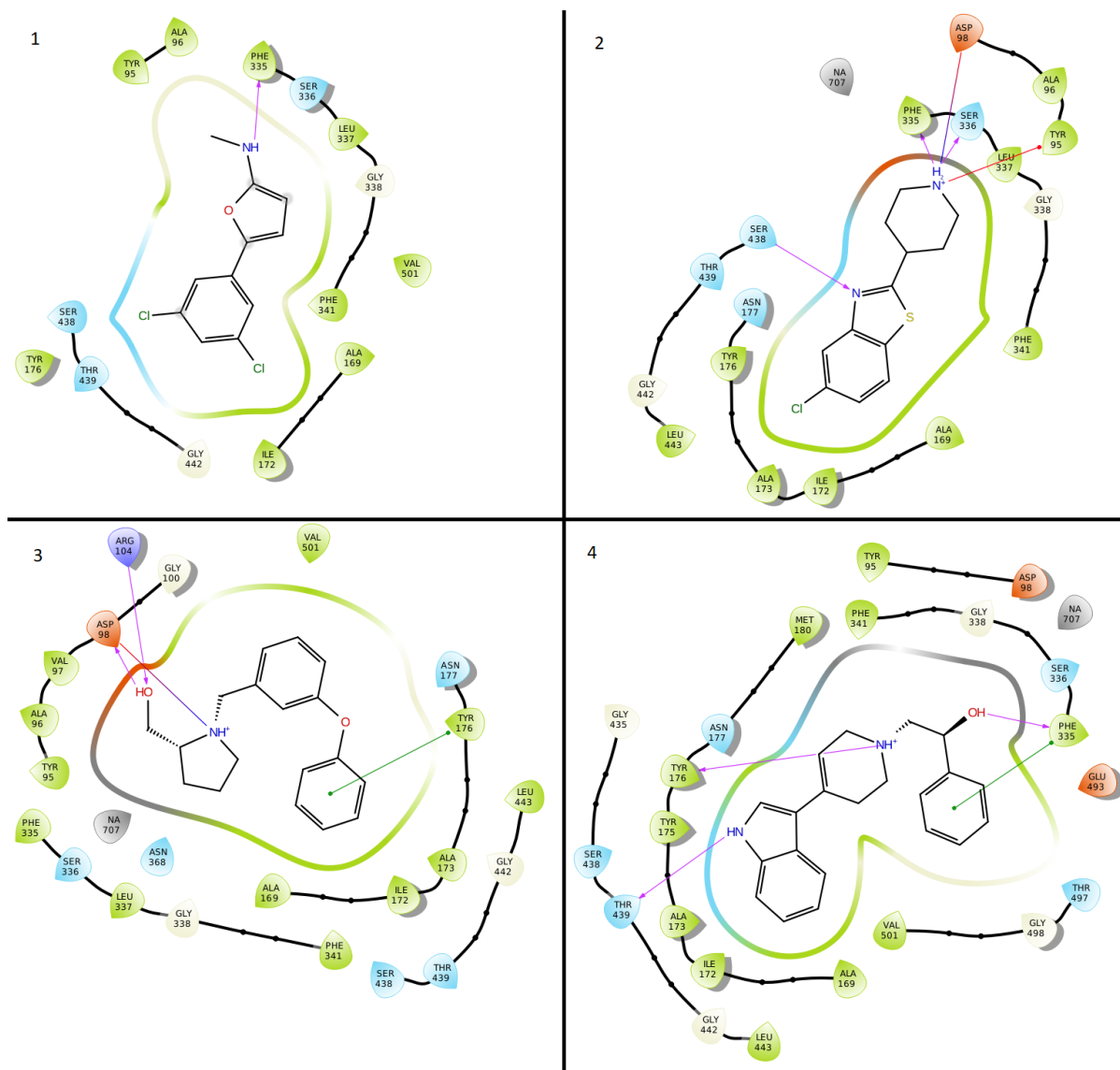
Appendix Figure 1: The figure displays the interactions the docked ligands had with the amino acids of the **central** site. The number in the left corner of each interaction diagram (1-4) is the ID of the ligand (See Appendix Table 1).
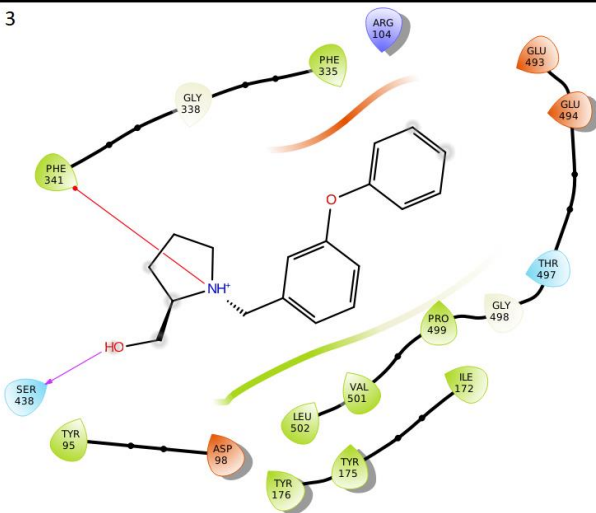
Appendix Figure 2: The figure displays the interactions the docked ligands had with the amino acids of the **central** site. The number in the left corner of each interaction diagram (5-8) is the ID of the ligand (See Appendix Table 1).
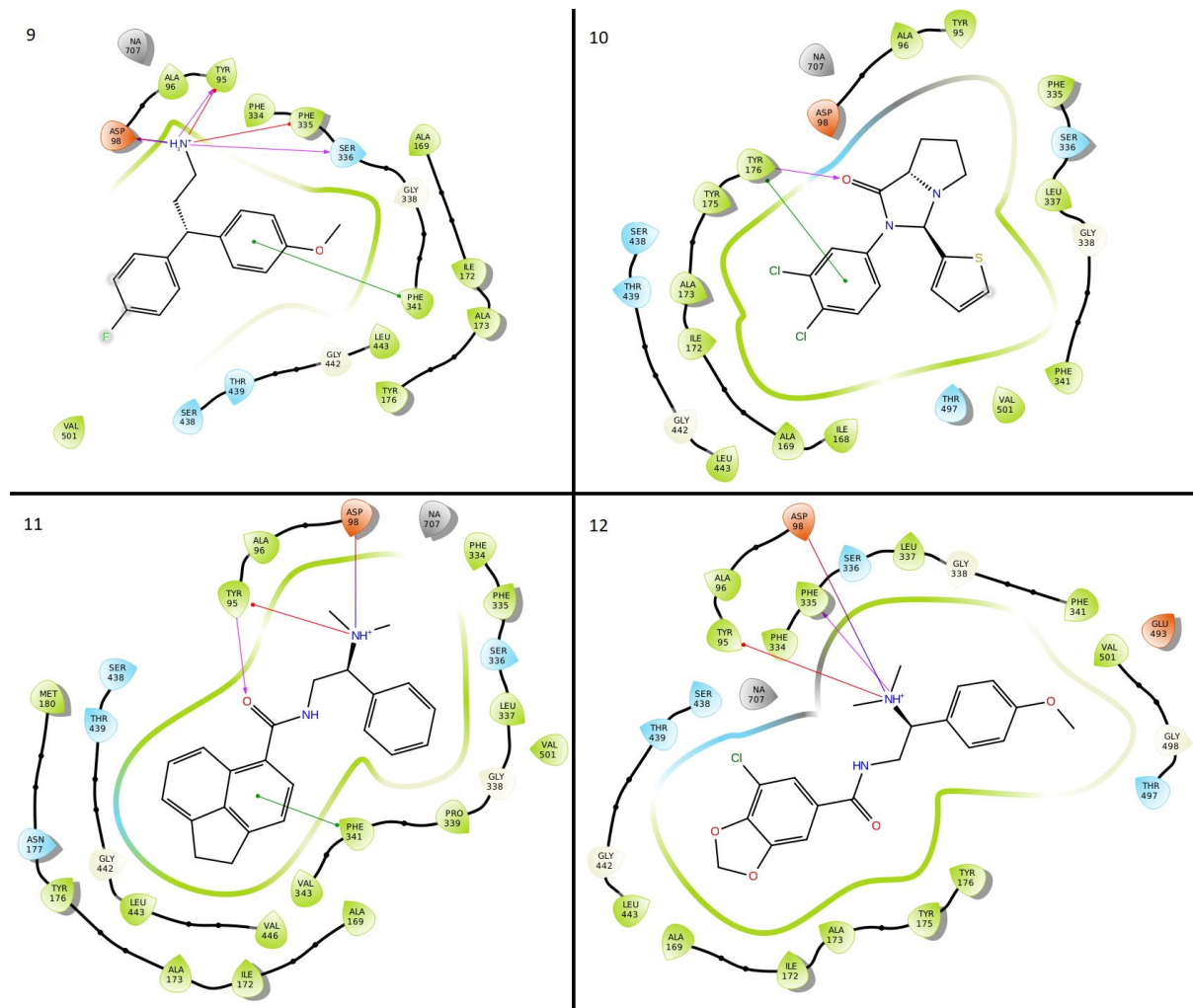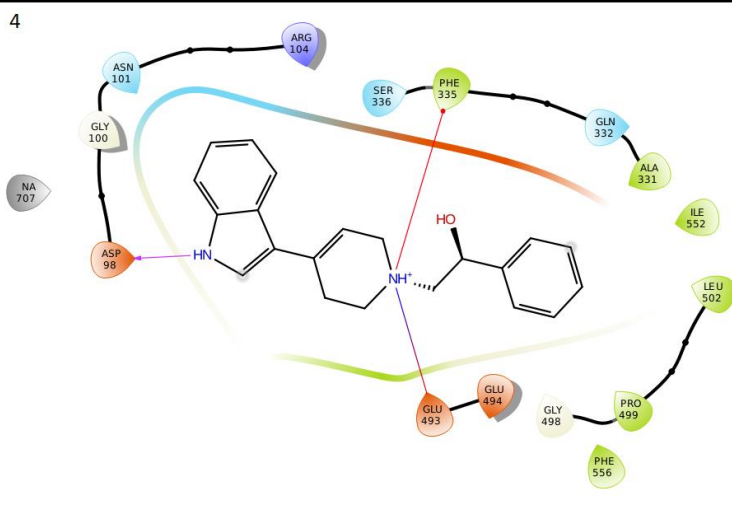
Appendix Figure 3: The figure displays the interactions the docked ligands had with the amino acids of the **central** site. The number in the left corner of each interaction diagram (9-12) is the ID of the ligand (See Appendix Table 1).
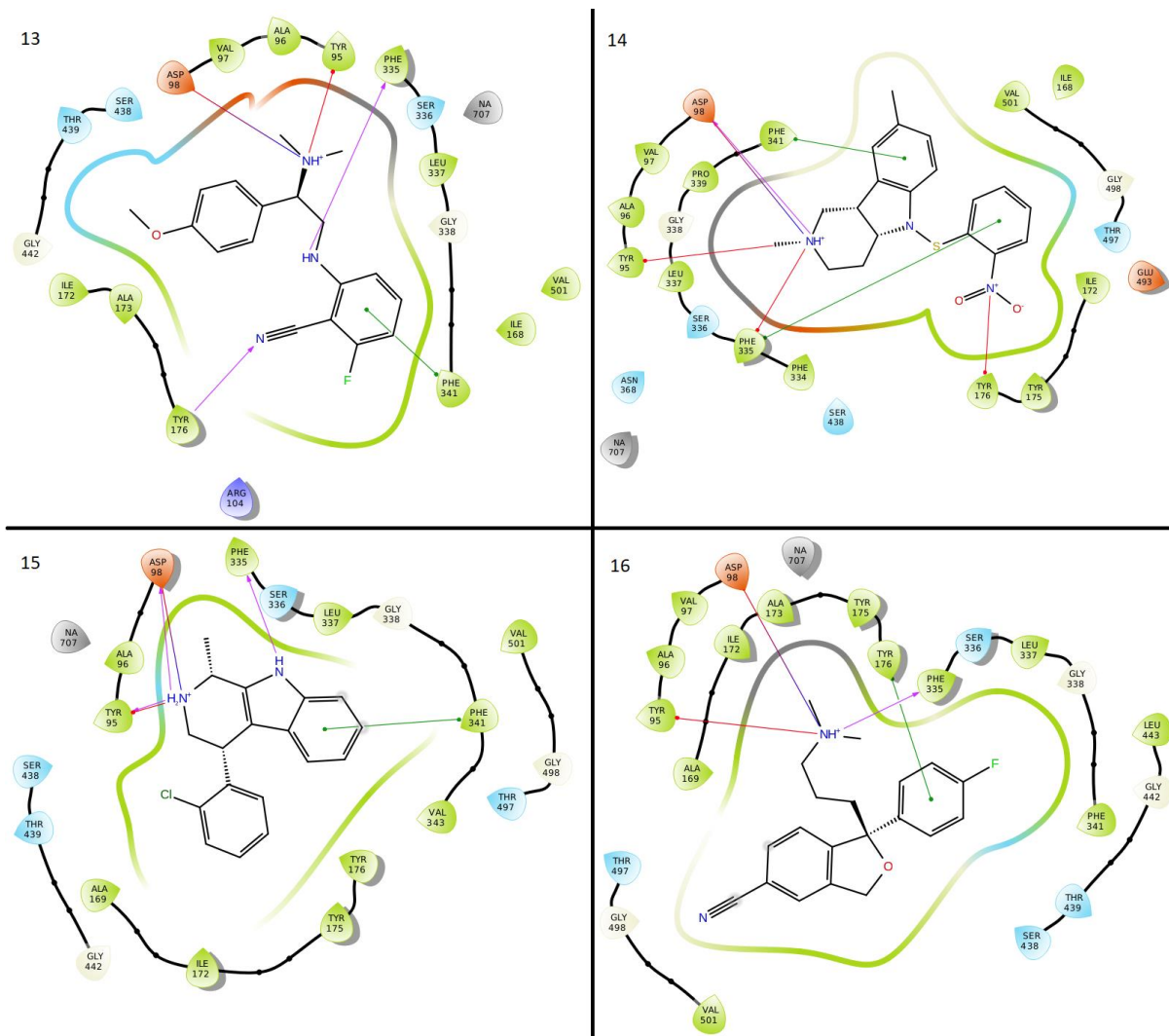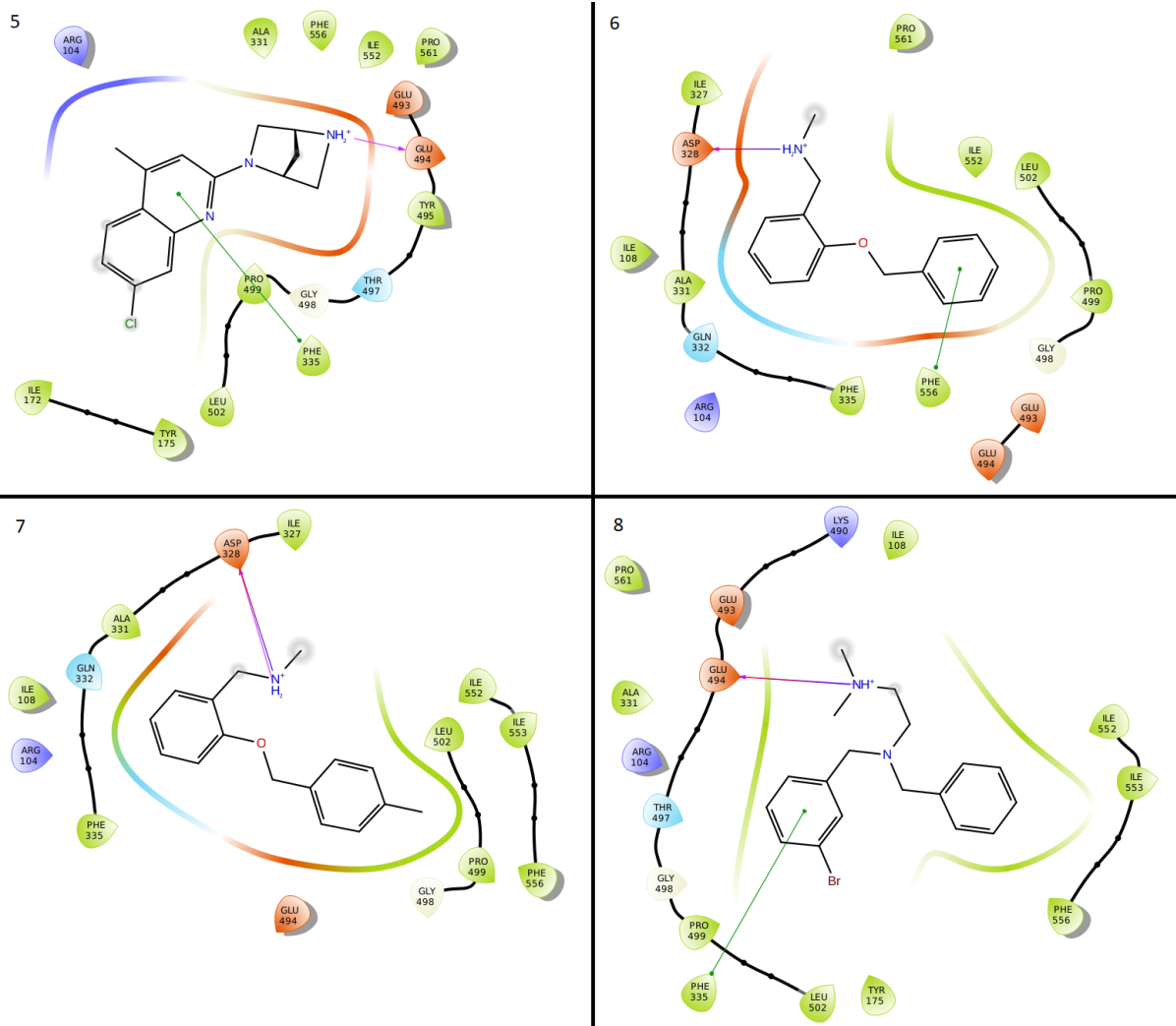
Appendix Figure 4: The figure displays the interactions the docked ligands had with the amino acids of the **central** site. The number in the left corner of each interaction diagram (13-16) is the ID of the ligand (See Appendix Table 1).
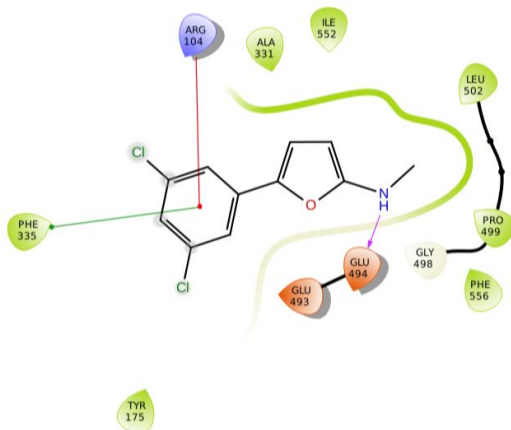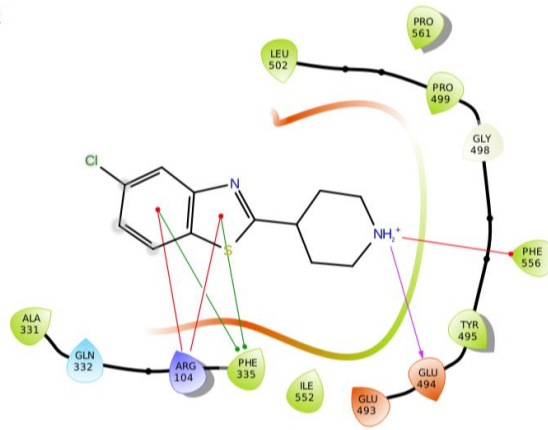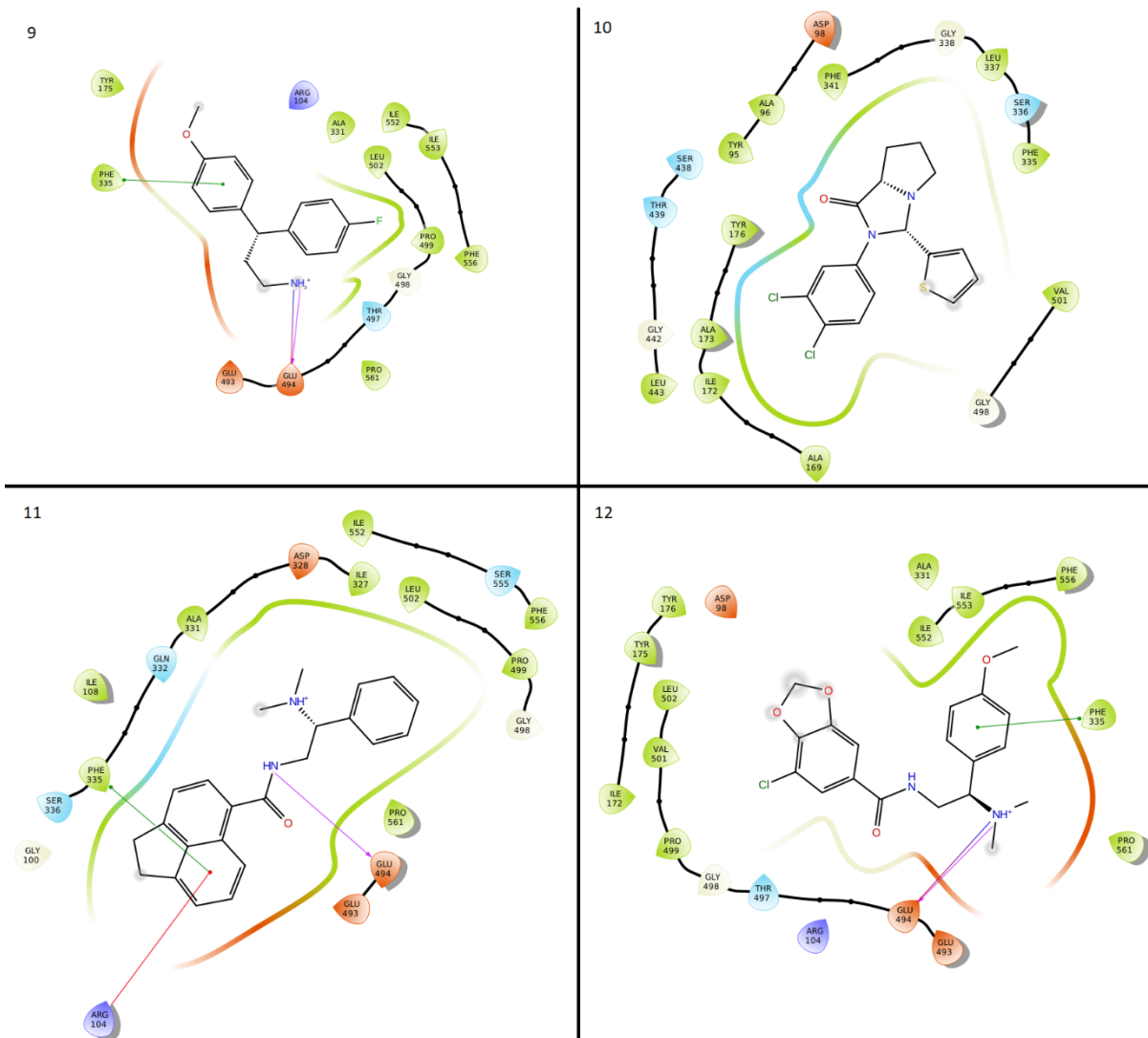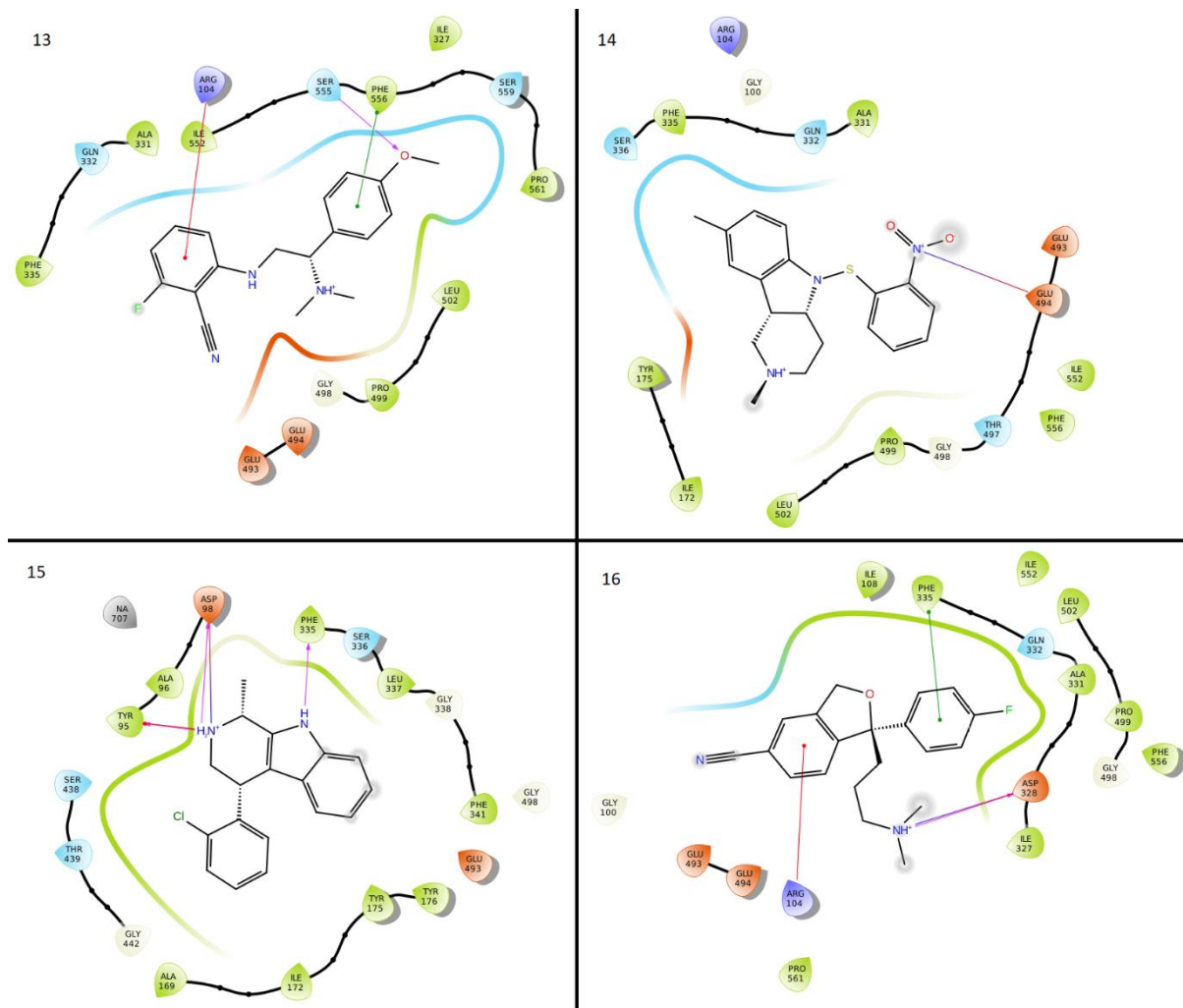
Appendix Figure 5: The figure displays the interactions the docked ligands had with the amino acids of the **allosteric** site. The number in the left corner of each interaction diagram (1-4) is the ID of the ligand (See Appendix Table 1).

Appendix Figure 6: The figure displays the interactions the docked ligands had with the amino acids of the **allosteric** site. The number in the left corner of each interaction diagram (5-8) is the ID of the ligand (See Appendix Table 1).

Appendix Figure 7: The figure displays the interactions the docked ligands had with the amino acids of the **allosteric** site. The number in the left corner of each interaction diagram (9-10) is the ID of the ligand (See Appendix Table 1).

Appendix Figure 8: The figure displays the interactions the docked ligands had with the amino acids of the **allosteric** site. The number in the left corner of each interaction diagram (13-16) is the ID of the ligand (See Appendix Table 1).