

IN SILICO SCREENING FOR GAT-3 INHIBITORS

Reem Alem Gebregazabhier

Master thesis in pharmacy (FAR-3911)

May 2017



ACKNOWLEDGEMENT

The master thesis was written at the Medical Pharmacology and Toxicology research group, Department of Medical Biology, Faculty of Health Science in collaboration with the Department of Pharmacy (IFA) at Uit The Arctic University of Norway from August 2016 to May 2017.

First, I would like to express my sincere gratitude to my supervisors Kurt Kristiansen and Ingebrigt Sylte for their guidance, support, and for always opening their doors whenever I needed help. This master thesis would not have been completed with out their help.

I also grateful to Linn Evenseth for sharing her knowledge and being available at anytime to help out with the methods in the laboratory.

To my mum and dad, I can't thank you enough for the endless love, guidance and support that you have given me throughout my life. I feel so lucky to have such wonderful parents. I would like also to thank my wounderful sibilings, Saba, Henock and Robel, I couldn't wish for greater sibilings. Love you all!

Finally, A big thank to you all my friends for all the support and encouraging word throughout the master period.

Reem Alem Gebregazabhier

Tromsø, May 2017

ABSTRACT

γ -aminobutyric acid (GABA) is the main inhibitory neurotransmitter of the central nervous system (CNS). GABA exerts its function by binding to three different receptor subtypes, the GABA_A, GABA_B and GABA_C receptor. The GABA level in different brain regions are regulated by four GABA transporters (GATs); GAT-1, GAT-2, GAT-3 and BGT-1. GAT-3 is located in glial cells that is controlling GABA function in the synapses.

A study has shown that Alzheimer's disease (AD) patients have an elevated GABA levels in the cerebrospinal fluid, while a transgenic mouse model of AD showed an unusual high GABA content in dentate gyrus (DG) and enhanced inhibition. The high GABA content in DG is a result of transport by the GABA transporter, GAT-3, and it is suggested that GAT-3 inhibitors may be a novel therapy. AD is the most common form of dementia, and is a worldwide disease with increasing incidence with age. There is no treatment that can cure the AD today and GAT-3 inhibitors may represent a new direction in the search for new therapeutic strategies.

The three dimensional (3D) structure of GAT-3 is unsolved. Therefore, X-ray structures of the drosophila dopamine transporter (dDAT) and the human serotonin transporter (hSERT) were used to construct homology models of GAT-3. The homology models were evaluated by docking a set of known inhibitors, substrates and decoys, and the best performing models were used in combined ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS) in order to identify potential GAT-3 inhibitors compounds from the ENAMINE database.

Four homology models were selected based on their ability to separate binders from non-binders by BEDROC calculation. 40 hit compounds from ENAMINE were selected with good docking score that may be potential GAT-3 inhibitors drug candidates. These hit compounds need evaluation by experimental testing.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	II
ABSTRACT	IV
1 INTRODUCTION.....	1
1.1 Alzheimer’s disease	1
1.1.1 Pathogenesis of AD.....	1
1.1.2 Treatment of AD	2
1.1.3 Drugs crossing the blood brain barrier	2
1.2 GABA neurotransmission.....	3
1.2.1 GABA receptors.....	4
1.3 The solute carrier family 6	5
1.3.1 GABA transporters	6
1.3.2 GAT-3 transporter	7
1.4 Molecular modeling.....	8
1.4.1 Homology modeling.....	9
1.4.1.1 Template identification	10
1.4.1.2 Target-template Alignment.....	11
1.4.1.3 Building the model.....	11
1.4.1.4 Model Refinements.....	11
1.4.1.5 Model validation of stereochemical quality.....	11
1.4.2 Virtual ligand screening.....	12
1.4.2.1 Ligand-based virtual screening.....	13
1.4.2.2 Structure-based virtual screening	14
2 AIM	16
3 METHODS	17
3.1 Software and databases	17
3.1.1 Software.....	17
3.1.2 Databases	18
3.2 Workflow of the study.....	19
3.3 Selection of active compound of GAT-3 and decoys.....	21

3.4	Ligand-based virtual screening.....	21
3.4.1	Clustering of active ligands.....	21
3.4.2	Calculation of model fingerprints.....	21
3.4.3	Determination of screening cut-off values.....	22
3.4.4	Screening of ENAMINE database	22
3.5	Homology modeling	22
3.5.1	Template identification	22
3.5.2	Target-template Alignment.....	23
3.5.3	Building the model	24
3.5.4	Evaluation of models.....	24
3.6	Structure-based virtual screening.....	25
3.6.1	Active ligands and decoys preparation	25
3.6.2	Semi-flexible docking.....	25
3.6.3	Induced fit docking.....	25
3.6.4	Evaluation of homology models with BEDROC.....	26
3.6.5	Docking the compounds obtained by the ligand based approach.....	26
4	RESULTS.....	28
4.1	Ligand-based virtual screening.....	28
4.1.1	Clustering of active ligands.....	28
4.1.2	Determination of screening cut-off values.....	29
4.1.3	Screening of ENAMINE database	30
4.2	Homology modeling	32
4.3	Structure-based virtual screening.....	37
4.3.1	Semi-flexible docking.....	37
4.3.2	Induced fit docking scores	38
4.3.3	Semi-flexible docking of active ligands into receptor obtained by IFD.....	40
4.3.4	Evaluation of the models, BEDROC scores.....	41
4.3.5	Docking the compounds obtained by the ligand based approach.....	44
4.3.5.1	4XPA-based model-I.....	45
4.3.5.2	4XP4-based model-II	47
4.3.5.3	5I73 based model-I, allosteric binding site.....	49
4.3.5.4	5I73 based model- II, orthosteric binding site	51

5	DISCUSSION.....	55
5.1	Ligand-based virtual screening.....	56
5.1.1	Clustering of active ligands.....	56
5.1.2	Use of known active ligands in searching for GAT-3 inhibitors.....	56
5.2	Homology modeling	57
5.3	Structure-based virtual screening.....	58
5.3.1	Evaluation of the selected final models, BEDROC score	58
5.3.2	Docking the compounds obtained by the ligand based approach.....	60
6	CONCLUSION.....	63
7	FUTURE DIRECTIONS	64
8	REFERENCE.....	65
9	APPENDIX.....	70

INDEX OF FIGURES, TABLES AND APPENDIX

FIGURES

Figure 1: 2D structure of GABA	4
Figure 2: Overview of SLC6 transporters.....	5
Figure 3: The GAT-3 amino acid sequence	8
Figure 4: The main steps in homology modeling.....	10
Figure 5: Overview of the two main approaches of virtual ligand screening	12
Figure 6: Workflow of the study	20
Figure 7: Homology model of 5I73-based model with orthosteric binding site and allosteric binding site defined by the co-crystallized ligand of the template.....	33
Figure 8: Multiple sequence alignment of GAT-3, GAT-1, GAT-2 and BGT-1 and dDAT (4XP4).....	34
Figure 9: Multiple sequence alignment of GAT-3, GAT-1, GAT-2 and BGT-1 and hSERT (5I73).....	35
Figure 10: Enrichment plot of 4XPA-based model-I (left) and 4XP4-based model-I (right)..	43
Figure 11: Enrichment plot of 5I73 allosteric-based model-I (left) and 5I73 orthosteric-based model-II (right).....	43
Figure 12: The binding mode of ENAMINE Z1428205595 (blue carbons) and the active ligand Compound 16 (Dark green carbons) in the binding site of 4XPA-based model-I.	53
Figure 13: The binding mode of ENAMINE Z31336192 (blue carbons) and the active ligand (R)-5d (dark green carbons) in the binding site of 4XP4-based model-II.....	53
Figure 14: The binding mode of ENAMINE Z31371528 (blue carbons) and the active ligand NNC05-0341 (dark green) in the binding site of 5I73 allosteric-based model-I.	54
Figure 15: The binding mode of ENAMINE Z103085506 (blue carbons) and the active ligand Compound 18 (green carbons) in the binding site of 5I73 orthosteric-based model-II..	54

TABLES

Table 1: An overview of Lipinski's "rule of five" and Veber for other therapeutic indications and CNS drugs.	3
Table 2: Overview of the four GATs (12).....	7
Table 3: X-ray crystal structures that were utilized as the templates in homology modeling of GAT-3.	23
Table 4: Five clusters of active ligands.....	28
Table 5: Determination of screening cut-off values for the 2D fingerprints.....	30
Table 6: Number of hits from the ligand-based screening.....	31
Table 7: Homology model evaluation.....	36
Table 8: Complexes selected for IFD.....	38
Table 9: Docking score and IFD score after IFD of the selected complexes.....	39
Table 10: The docking evaluation of the models from Table 9.	41
Table 11: BEDROC score of models after IFD..	42
Table 12: Clustering of ENAMINE compounds in 4XPA-based model-I.....	45
Table 13: ENAMINE compounds in 4XP4-based model-II.	47
Table 14: ENAMINE compounds in 5I73 allosteric-based model-I.....	49
Table 15: ENAMINE compounds in 5I73 orthosteric-based model:	51

APPENDIX

Appendix 1 : Overview of the 72 known active ligands

ABBREVIATIONS

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
AD	Alzheimer's disease
ADMET	Absorption-distribution-metabolism-excretion-toxicity
APP	Amyloid precursor protein
Aβ	Beta-amyloid
BBB	Blood brain barrier
BEDROC	Boltzmann-Enhanced Discrimination of Receiver-Operation Characteristics
BGT-1	Betaine transporter
BLAST	Basic Local Alignment Search Tool
C-terminal	Carbone terminal
Ca²⁺	Calcium ion
Cl⁻	Chloride ion
ClogP	Calculated log P
CNS	Central nervous system
DAT	Dopamine transport
dDAT	Drosophila dopamine transport
DG	Dentate gyrus
DUD.E	Databased of Useful Decoys: Enhanced
E_{angle}	Angle binding energy
E_{bonded}	Bonded energy
E_{dihedral}	Torsional energy
E_{elec}	Electrostatic energy
E_{non-bonded}	Non-bonded energy
E_{vdw}	Van der Waals energy
GABA	γ -aminobutyric acid
GAD	Glutamic acid decarboxylase
GAT-1	GABA transporter 1
GAT-2	GABA transporter 2

GAT-3	GABA transporter 3
HBA	Hydrogen bonding acceptors
HBD	Hydrogen bond donors
hSERT	Human serotonin transport
HTVS	High throughput virtual screening
IC₅₀	Half maximal inhibitory concentration
ICM	The Internal Coordinate mechanics
IFD	Induced fit docking
K⁺	Potassium
K_i	Binding affinity constant
LBVS	ligand-based virtual screening
LeuT	Leucine transporter
MM	Molecular mechanics
Mw	Molecular weight
N-terminal	Nitrogen terminal
NMDA	N-methyl-D-aspartate
NMR	Nuclear magnetic resonance
NRI	Norepinephrine reuptake inhibitor
PDB	Protein Data Bank
PDB ID	Protein Data Bank Identification
PIC₅₀	Logarithmic half maximal inhibitory concentration
PSA	Polar surface area
QM	Quantum mechanics
QSAR	Quantitative structure-activity relationship
R	Indicated the configuration of a stereocenter in a molecule
S	Indicated the configuration of a stereocenter in a molecule
SAVES	Structural Analysis and Verification Server
SBVS	Structure-based virtual screening
SERT	Serotonin transport
SLC6	Solute carrier 6
SMILES	Simplified molecular-input line-entry system
SP	Standard precision
SSRI	Selective serotonin reuptake inhibitor

TM	Transmembrane
UniProtKB	Universal Protein Resource Knowledgebase
vdW	Van der Waals
XP	Extra precision
Å	Ångström

1 INTRODUCTION

1.1 Alzheimer's disease

Alzheimer's disease (AD) is a neurodegenerative disease known as the most common form of dementia constituting about 60% of all types of dementias (1). According to the world Alzheimer Report 2016 (Alzheimer's International), 47 million people worldwide are affected by dementia, which is considered to rise to 131 million by 2050 (2). AD is named after Dr. Alois Alzheimer who described this disease for the first time in 1906 (3).

AD is an irreversible disease that particularly affects most elderly and develops gradually until symptoms become perceptible. The first sign of AD is memory problems. AD is categorised in three main stages: mild (early-stage), moderate (middle-stage) and severe (late-stage). In the early-stage the patient can function independently but has problems with remembering names of new people, getting lost and having personality changes. The middle-stage is the stage when a patient has trouble with elementary actions such as language and recognizing familiar persons and the memory gets worse. In the late final stage of the disease, patients can have no communication and need daily care (4).

1.1.1 Pathogenesis of AD

The cause of AD remains unknown, but we know that there are two pathological hallmarks for AD. The first hallmark is amyloid plaques, protein containing β -amyloid ($A\beta$) from an amyloid precursor protein (APP). The second hallmark is neurofibrillary tangles that accumulate in the brain in Alzheimer's disease patient. Neurofibrillary tangles are also proteins that are part of microtubule-associated protein (tau). These proteins cause damage to the activity between nerve cells in the brain and death of nerve cells in the hippocampus that have functions such as memory and other symptoms such as speech problems, mood swings and confusion (4-6).

The risk of AD increases with age, from about 5% at 65 years to 90% or more at 95 years. Other risk factors that affect this disease are genetic mutation, health, environmental, lifestyle and many others factors (5).

1.1.2 Treatment of AD

At present, there is no treatments for AD, but there are some drugs that can treat symptoms of the disease. Drugs such as cholinesterase inhibitors and NMDA receptor antagonists are approved for treatment of AD symptoms (7). γ -secretase inhibitors that reduce A β production were considered promising, but clinical studies showed limited efficacy, and they even worsened the AD symptoms in some patients (8). New drugs are therefore urgently needed. New research has proposed GAT-3 inhibitors as new drug candidates for treatment of AD.

Recent research shows that AD patients have a very high level of γ -aminobutyric acid (GABA) in the cerebrospinal fluid and that the high GABA content is caused by the GABA transport, GAT-3 (8). Compounds inhibiting the GAT-3 transporter are therefore suggested to be promising drug candidates in AD treatment.

1.1.3 Drugs crossing the blood brain barrier

Drugs for AD need to cross the blood brain barrier (BBB), which is a layer of endothelial cells within the brain capillaries and has a protective function. Development of drug that should reach the brain is a challenging task, and today there are few drugs that are able to cross the BBB. Today 98% of all small molecular drugs and 100% of large drug molecules do not cross the BBB (9).

There are two putative mechanisms for drugs to pass the BBB: active transport or passive transport. Active transport is an energy-requiring process that moves molecules against a concentration gradient from low to high concentration and is able to transport polar molecules through the BBB. Passive transport is a transport that moves molecules through the cell membrane down the concentration gradient and is able to transport lipophilic drugs through the BBB. For passive transport of lipophilic drugs across the BBB the drugs need to have favourable pharmacokinetic properties for transport connected to the molecular weight (MW), hydrophobicity/calculated oil/water distribution coefficient (ClogP), number of hydrogen bond donors (HBD), and hydrogen bonding acceptors (HBA), polar surface area (PSA) and molecular flexibility (number of rotatable bonds) (Table 1). These properties are described by the Lipinski's "rule of five", or the Veber rules/suggestions that gives drugs drug-like

properties to avoid absorption-distribution-metabolism-excretion-toxicity (ADMET) issues. Studies have shown that a successful central nervous system (CNS) drugs have lower ClogP value and lower MW, HBD, HBA, PSA and rotatable bonds than other therapeutic indications (10). These properties need to be taken into account early in the phase of drug development (Table 1).

Table 1: An overview of Lipinski's "rule of five" and Veber for other therapeutic indications and CNS drugs.

Properties	Physicochemical properties	Other therapeutic	CNS drug
Lipinski's "rule of five"	MW	< 500	< 450
	ClogP	< 5	< 5*
	HBD	< 5	< 3
	HBA	< 10	< 7
Veber	PSA	< 140 Å ²	< 60-70 Å ²
	Rotatable bonds	< 10	< 8

MW: Molecular weight, ClogP: calculated oil/water distribution coefficient (hydrophobicity), HBD: Number of hydrogen bond donor, HBA: Number of hydrogen bond acceptor, PSA: Polar surface area, rotatable bonds: number of molecular flexibility. * clogp as low as possible

1.2 GABA neurotransmission

GABA is the main inhibitory transmitter in the brain. GABA is synthesized from the amino acid glutamate by the enzyme glutamic acid decarboxylase (GAD) in the presynaptic neuron. GABA is stored into synaptic vesicles by a vesicular transporter in presynaptic neuron. The neurotransmitter is released from synaptic vesicles into the synaptic cleft and binds to receptors in the postsynaptic neuron, which triggers a response. The remaining GABA molecules in the synaptic cleft must be cleared and this occurs by reuptake into presynaptic neuron and glia cells by GABA transporters (11). The GABA transporter belongs to the solute carrier family 6 (SLC6) family of transporter proteins(12).

GABA is the main inhibitory neurotransmitter in the brain. GABA act as a neurotransmitter at approximately 40% of all synapses in the brain (13), and in many brain regions GABA is found in a concentration 1000 times higher than that of the monoamine neurotransmitters. GABA is involved in several neurological and psychiatric disorders in the CNS such as mental retardation, development malfunctions, epilepsy, sleep disorders, anxiety, depression, drug dependence, sensorimotor processing and motor coordination disorders (11). GABA is also found in other tissues of the body such as the liver, spinal cord, eyes, bladder and parathyroid (12)

GABA has a major function in CNS and abnormal GABA levels can result in different brain diseases. Protein implicated in GABA neurotransmission can therefore be valuable drug candidates (14). The development of GABA transport inhibitors as drugs is limited, and Tiagabine (GAT-1 inhibitor) is the only drug available on the market inhibiting GABA transport. Tiagabine is approved for the treatment of epilepsy (15).

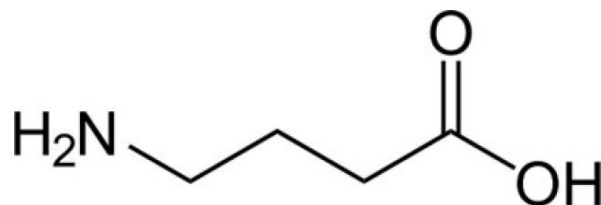


Figure 1: 2D structure of GABA

1.2.1 GABA receptors

There are three main types of GABA receptors: GABA-A, GABA-B and GABA-C receptors. The GABA-A receptor is a ligand-gated chloride-ion channel receptor located on the postsynaptic membrane. GABA-A receptors are pentamers consisting of 5 subunits (α - β - α - β - γ) with a GABA binding site between the α and β subunits giving a total of two GABA binding sites on each GABA-A receptor complex. Activation of GABA-A receptor causes membrane hyperpolarisation by opening of chloride (Cl^-) channels between the subunits. By Cl^- ions influx into postsynaptic cell, the membrane potential increases, which cause the cell to become less excitable. Many CNS drugs have GABA-A receptors as their target, including barbiturates and benzodiazepines.

The GABA-B receptor is a G-protein-coupled receptors of family C located both on presynaptic and postsynaptic neurons. Stimulation of GABA-B receptors reduces cAMP that decreases the Ca^{2+} influx by voltage-gated calcium channels and increase the K^{+} efflux by inwardly rectifying potassium channels, which results in reducing postsynaptic excitability. GABA-C is a ligand-gated chloride-ion channels receptor containing ρ subunits ($\rho 1$ - $\rho 3$) (11, 16).

1.3 The solute carrier family 6

Among solute carrier (SLC) families, SLC6 is the family containing most members of the human genome (20 transporters) (17). SLC6 is also recognized as the neurotransmitter sodium symporter family (NSS) or $\text{Na}^{+}/\text{Cl}^{-}$ neurotransmitter transporter family (18, 19), which includes four groups of transporters: GABA, monoamine, amino acid (I) and amino acid (II) (Figure 2). The SLC6 transporters have important roles in removal of neurotransmitters from the synaptic cleft by transporting them into the presynaptic neurons and glial cells, but they also have essential role in other tissues, such as liver, kidney, pancreas, intestine etc. (20, 21).

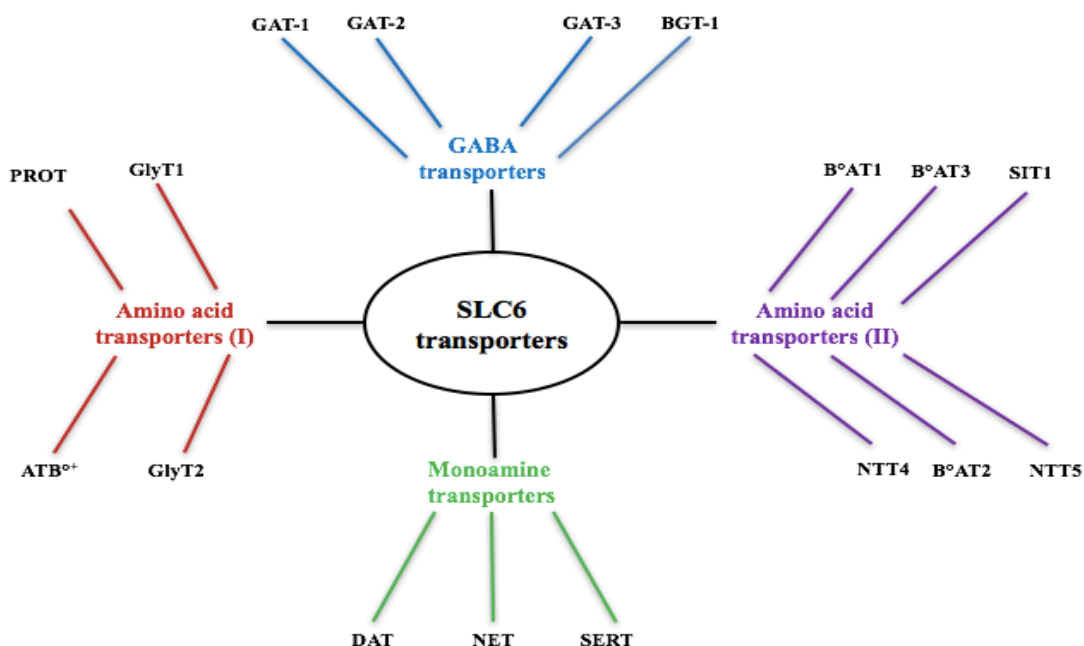


Figure 2: Overview of SLC6 transporters (12).

In 2005 the first X-ray crystal structure of a prokaryotic SLC6 member was published. The structure was of the *Aquifex aerolicus* leucine transporter (LeuT), at resolution of 1.65 Å (21). *Aquifex aerolicus* is a thermophilic bacterium. More recently, X-ray crystal structures of the dopamine (DAT) and serotonin transporter (SERT) have been determined. These X-ray crystal structures can be used as templates for homology modeling of GAT-3. Today, around 177 eukaryotic and 167 prokaryotic transporters of the SLC6 family have been classified (22).

Almost all of the SLC6 family members use the electrochemical gradient of Na⁺ and Cl⁻ for co-transport of GABA into the cell. The serotonin transporter (SERT) is the only family member that in addition to influx of Na⁺ and Cl⁻ transports K⁺ out of the cell (23). Although these transports: GABA, monoamine, amino acid (I), amino acid (II) belong to the same family, the stoichiometry between substrate, Na⁺ and Cl⁻ varies between family members (12, 21).

1.3.1 GABA transporters

The subfamily of GABA transporters consists of four transporters (GATs): GAT-1, GAT-2, GAT-3 and betaine (BGT-1) that all are located in the brain (24). Substrates for all the GABA transporters containing a C-terminal carboxyl group (COOH). Substrated for three of the transporters (GAT-1, GAT-2 and GAT-3) have an amino group (NH₂) N-terminally, while the BGT-1 in addition has the amino group methylated. Most of GABA transporters cotransport two Na⁺ ions, one substrate molecule and one Cl⁻ ion, except for BGT1, which cotransport one substrate molecule, three Na⁺-ions and one Cl⁻ion (Table 1) (12). GAT-1, GAT-2, GAT-3 and BGT-1 consist of 599, 602, 631 and 614 amino acids, respectively, with the differences in the length of the loops. These transporters are expressed in CNS, but they are also expressed in other tissues (21) (Table 2).

Table 2: Overview of the four GATs (12).

Endogenous substrate	Transporter	Stoichiometry (Substrate/ Na ⁺ /Cl ⁻)	Tissue distribution
GABA	SLC6A1/GAT-1	1:2:1	Brain, bladder, liver, parathyroid
GABA	SLC6A13/GAT-2	1:2:1	Brain, kidney, liver, eye
GABA	SLC6A11/GAT-3	1:2:1	Brain, eye, spinal cord
GABA, Betaine	SLC6A12/BGT-1	1:3:1	Brain, kidney, liver

1.3.2 GAT-3 transporter

The GAT-3 transporter regulates the extracellular GABA level in CNS. The GAT-3 levels are high in the CNS and is primarily expressed in glial cells, and also expressed in glomerular layer of the olfactory bulb, the inner nucleus of the retina, the thalamic paraventricular nucleus, and the globus pallidus (25). The GAT-3 transporter is located in glial cells and the glial cell uptake of GABA has an important role for controlling the accessibility of GABA neurotransmitter in the synapses, which has potential clinical relevance in epilepsy and stroke (14, 26). The physiological roles and therapeutic potential of the GAT-3 transporter can be further investigated by developing GAT-3 inhibitors (14). As previously stated, recent research indicate that the inhibitors of the GAT-3 transportes also may have a therapeutical potential in AD (8).

The detailed structure of the GAT3 transporters has not been resolved by X-ray crystallography. As shown in Figure 3, the GAT-3 contains 12 transmembrane α -helices (12TMs) spanning the membrane connected with intracellular and extracellular loops, while the N- and C-terminals are located intracellularly. The extracellular loop between TM3 and TM4 contains N-linked glycosylation sites (21).

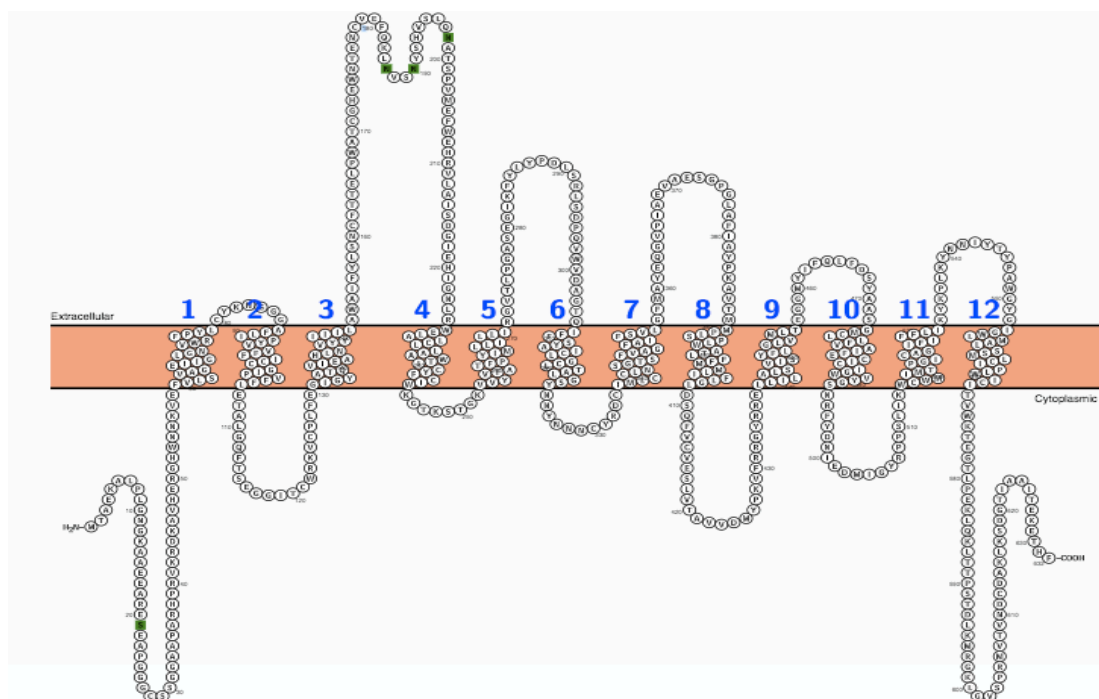


Figure 3: **The GAT-3 amino acid sequence.** Snake-like plot showing the membrane topology of human GAT-3 amino acid sequence. The figure has been generated by using the tool <http://wlab.ethz.ch/protter/start/> Uniprot kode: p48066 for the transporter.

1.4 Molecular modeling

Molecular modelling studies molecular structures and properties by using techniques such as computational chemistry, molecular dynamics simulations and structure and sequences analysis methods. Molecular modelling is often used for the discovery of new compounds with desired activity, and in drug development. The two main computational approaches for molecular modelling are molecular mechanics (MM) and quantum mechanics (QM).

However, today also a combination of these two methods may be used. Very often choosing MM or QM is dependent on the size of the molecular system to be calculated (27, 28).

MM is used for large molecules and is suitable for doing energy minimization, identifying stable conformations, energy calculations for specific conformations, generating different conformations and studying molecular motion. MM calculates the total potential energy (E_{tot}) from the sum of all bonded (E_{bonded}) and non-bonded ($E_{\text{non-bonded}}$) molecular interactions. E_{bonded} is the sum of the bond stretching energy (E_{bond}), angle binding energy (E_{angle}) and torsional energy (E_{dihedral}). $E_{\text{non-bonded}}$ is the sum of Van der Waals energy (E_{vdw}) and

electrostatic energy (E_{elec}) (27-29). The total energy form is described by the following equation (equation 1):

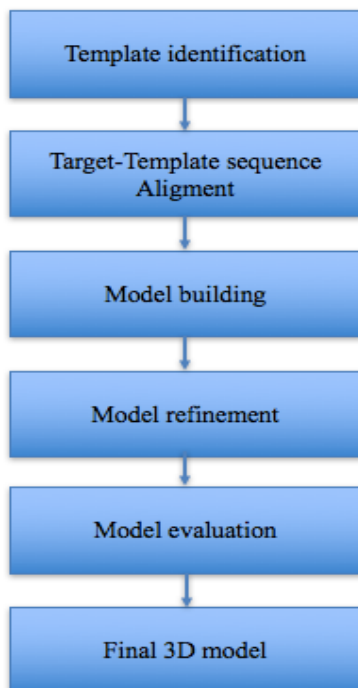
$$E_{tot} = E_{bonded} + E_{non-bonded}; \text{ or}$$
$$E_{tot} = (E_{bond} + E_{angle} + E_{dihedral}) + (E_{vdw} + E_{elec})$$

QM takes into account the electronic nature of each atom, and is used to model small molecules. The QM method is more accurate but also more time consuming than MM and is well qualified for calculating heat of formation for specific conformations, partial atomic charges calculated from molecular orbital coefficients, electrostatic potential, molecular orbital energies and coefficients (27).

1.4.1 Homology modeling

Proteins are large molecules with several functions in our body such as transporters, enzymes, antibodies, and structural components. The shape of the protein decides the function of the protein, and therefore it is important to have a knowledge and understanding of the three dimensional (3D) protein structure. The human genome consists of about 8000 membrane proteins and to understand their 3D structure and drugability is very important for drug discovery and development (30). But a problem with membrane proteins for 3D structure determination is their hydrophobic surface, which leads to poor amount of protein expressing in the cellular membrane. Proteins are flexible and unstable which create difficulties for keeping the protein functional after solubilisation and purification, and for keeping the membrane protein in one stable conformational state long enough to be able to crystallize the protein (31). NMR spectroscopy and X-ray crystallography are the two main techniques for experimental determination, and examines small macromolecules in solution. X-ray crystallography gives atomic resolution rigid structures and about 90% of the known 3D protein structures have been solved by X-ray crystallography. Today NMR and X-ray crystallography have solved total 11808 and 116306 structures of proteins, nucleic acids, protein/nucleic acid complexes and others (<http://www.rcsb.org/pdb/home/home.do>). Both those techniques are laborious, time-consuming and have some common problems particularly with membrane proteins (32).

Homology modeling can be utilized to construct 3D models of the proteins of unknown structure if there are 3D structures of related proteins available. The protein of interest is the “target”, while the protein homologue with known 3D structure is the template. Homology models of proteins are useful for understanding their structures and functions (32), and are predicted structures of the protein of interest. The assumption for this technique is that 3D structures of protein homologues will have a similar overall 3D fold. Homology modelling contains several steps for constructing the 3D structure of a protein based on a template (28).



.Figure 4: The main steps in homology modeling.

1.4.1.1 Template identification

Template selection is the first step of homology modeling. In this step it is important to find suitable template with similar 3D-structure as the target. Suitable templates for modelling the target can be search for by using tools such as FAST and Basic Local Alignment Search Tool (BLAST). The template that are selected should have as high amino acid similarity as possible to the target sequence and also high crystallographic resolution (33).

1.4.1.2 Target-template Alignment

Target-template alignment is the second step of homology modeling. It may be necessary to manually adjust the sequence alignment between target and template in order to avoid gaps in secondary structures in the alignment. If you have more than one possible template, the 3D structures of the templates may be superimposed and the alignment may be adjusted based on the structural superimposing. It may also be important to include many homologous sequences in the alignment (multiple sequence alignment) in order to more easily identify structurally homologous regions. A correct sequence alignment is very important for the model, otherwise the result of homology modelling will be inaccurate (33).

1.4.1.3 Building the model

The 3D structure of a target is built on the background of the target-template sequence alignments. Model building includes three stages: (1) generation of amino acid backbone of structurally conserved regions, (2) generation of the non-conserved loop regions, and (3) optimization of side chains (33). Modelling of loops is the most challenging in spite of high sequence similarity and maestro has two methods to overcome the loop region: knowledge-based and energy-based. Knowledge based is an approach that searches in Protein Data Bank (PDB) for matching residue that can be placed in the loop regions. Energy based is an approach that minimizes the energy function to obtain the best loop conformation by using Monte Carlo or molecular dynamics (33).

1.4.1.4 Model Refinements

Model refinement is used to optimize and correct the structure of 3D models. In this step, energy minimization, Monte Carlo Simulation and/or molecular dynamics calculations can be used to help the model to optimize the interactions between amino acids (33).

1.4.1.5 Model validation of stereochemical quality

Model validation is the last step of homology modeling. This step is to check the quality of the model by using the structural Analysis and verification server (SAVES; <http://nihserver.mbi.ucla.edu/SAVES/>). Molecular docking can also be used to evaluate if the model is able to distinguish between the known binders and compounds that do not bind (non-binders).

1.4.2 Virtual ligand screening

Virtual ligand screening (VLS) is an *in silico* technique of drug discovery used for screening of huge compounds libraries. The main goal with virtual ligand screening is to identify new lead molecules by searching database containing million of molecules (e.g ENAMINE).

There are two main approaches of virtual ligand screening: ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS). The LBVS approach is using information about known compounds for the target and, is often used when detailed structural information about the target is lacking. The SBVS approach is used when the 3D structure of target is known, and the structure is used for docking the compound library and calculate the score of each compound (34). Both VLS techniques can be combined as long as the structure of ligands and target are known.

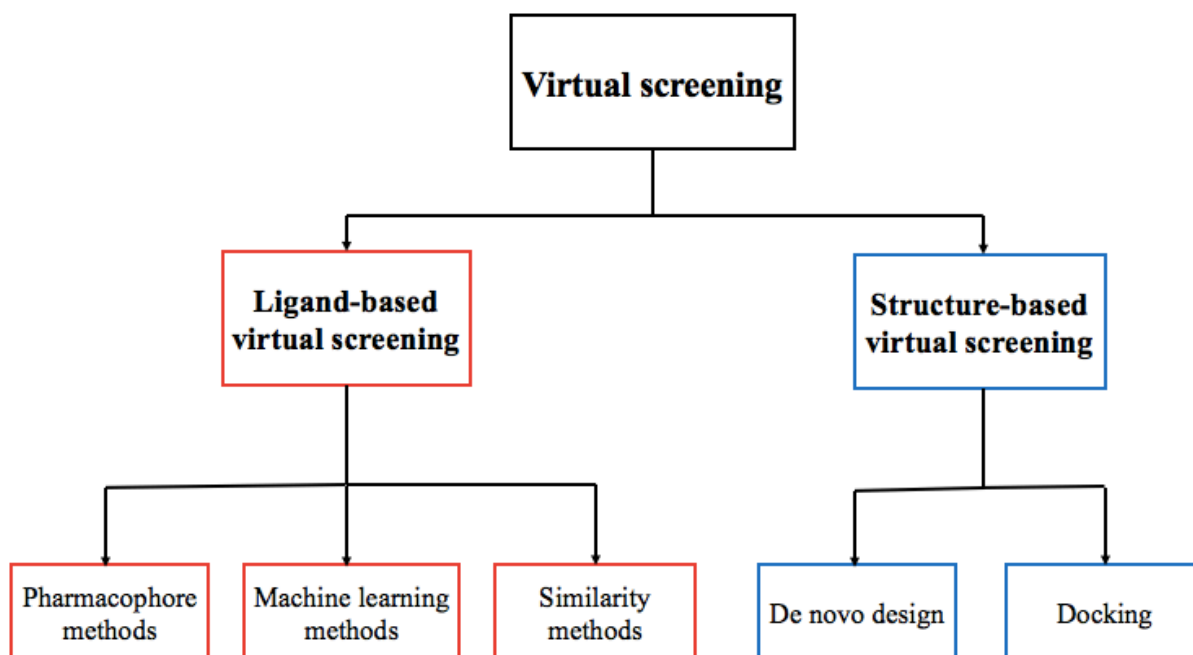


Figure 5: **Overview of the two main approaches of virtual ligand screening:** ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS).

1.4.2.1 Ligand-based virtual screening

Ligand-based virtual screening (LSVS) is based on structural similarity between compounds binding to the target. The method is searching for new molecules intended to bind to the target by using information about the properties of a known active ligands. It is assumed that compounds with similar structure are likely to have similar activity and are affected in the same manner at the target (35).

LSVS include pharmacophore methods, machine learning methods (including QSAR) and similarity methods (2D-fingerprints) (Figure 5). Pharmacophore methods use the structural knowledge about active ligands and identify what they have in common for optimal target interactions. Quantitative structure-activity relationship (QSAR) methods calculate the relationship between chemical structure and pharmacological activity of known compounds and use that to find new compounds in a database. The similarity methods such as 2D fingerprints are based on calculating the structural similarity between an active reference compound and compounds in a database. The 2D fingerprints calculations are performed by screening a database of compounds against active reference compounds. There are several metrics that can be a measure of the degree of similarity between two molecules, and Tanimoto similarity is the most used metric (36, 37). The general form of Tanimoto similarity metric can be written as (equation 2):

$$\text{Tanimoto similarity} = \frac{c}{(a + b - c)}$$

Where a is number of fragments in compound 1, b is number of fragments in compound 2 and c is the number of fragments in compound 1 and compound 2 (36).

1.4.2.2 Structure-based virtual screening

Structure-based virtual screening (SBVS) is a method that uses the knowledge about the 3D structure of the target to identify compounds that can be developed into drugs. SBVS includes, both de novo design and docking. De novo design is based on building the ligand molecule step-by-step into the binding pocket. Docking is a technique where the goal is to find the correct fit between a ligand and a target and it is the most used computational tool for structure-based drug design in research-based pharmaceutical industry (38).

In order to perform docking, the program requires atomic resolution structures of a target and ligand(s) and also an idea of where the binding site is. It is up to the docking programs to solve where to exactly fit the ligand(s), the conformation of the ligand and target and to evaluate the interacting energies and relate that to activity. It is possible to generate different complexes between the ligand and the binding site. A combination of a search algorithm that intend to suggest several possible ligand poses, and a scoring function aiming to identify the true (native) binding pose is used by the program. Most docking programs today treat the ligand as flexible and the target rigid (semi-flexible docking), but there are other approaches like induced fit docking (IFD) that also include flexibility into the target. The scoring calculates the strength of the interaction between a ligand and the target. The molecular docking software docks the ligands into the binding site and then calculate the free energy of binding between ligand and target. After running the calculation, the software lists up the compounds ranked based on their score from the highest to lowest score. The aim of the scoring function is to identify the most reliable binding pose and to distinguish the active ligands from the inactive and decoys. In molecular docking the scoring functions have three essential applications, first to identify the binding site and binding mode of the ligand in the target, second predict the binding affinity between ligand and target, and third is searching for potential drug hit/lead compounds that bind to the target by using virtual screening (39). The calculated free energy of binding is defined by the Gibbs-Helmholtz equation (equation 3):

$$\Delta G = \Delta H - T\Delta S$$

ΔG is the free energy of the binding, ΔH is the enthalpy, T is the temperature in Kelvin and ΔS is the entropy. Entropy is the degree of disorder in a system, when a system becomes more

disordered, the entropy increases, while the enthalpy is order. The relation between the ΔG and the affinity between a ligand and a target (K_i) is described by (equation 4):

$$\Delta G = -RT\ln K_i$$

R is the gas constant, T is the temperature in Kelvin and K_i is the inhibition constant.

There are four types of scoring functions in molecular docking: Force-field scoring functions, empirical scoring functions, knowledge-based scoring functions and consensus score functions. The force field functions are based on molecular mechanical energy of receptor-ligand interaction and internal ligand energy and is based on non-bonding interactions such as vdW interactions and electrostatic interactions and bonded interactions such as stretching/bending/torsional forces. Empirical is based on interactions such as vdW, hydrogen bond (H-bond), hydrophobicity, electrostatics, desolvation, entropy, etc. Knowledge-based scoring is statistical potential of ligand-target complexes from structural information of experimentally determined structures. Consensus score apply a combination of the three mentioned types of score functions to adjust their inaccurate score functions (39).

2 AIM

The GAT-3 transporter is responsible for the elevated levels of GABA in the dentate gyrus (DG) of transgenic AD mouse. The main aim of this project was to use a combination of LBVS and SBVS to predict GAT-3 inhibitors that may be drug candidates in AD.

Constructed homology models help in the understanding of the binding site of the target and docking of known compounds will help to understand the molecular interactions at the binding site.

The specific aims of the study were to:

1. Use different fingerprint types to identify putative GAT-3 inhibitors from the ENAMINE database by using the knowledge of known active ligands.
2. Construct homology models by using drosophila DAT (dDAT) and human SERT (hSERT) X-ray structures as templates.
3. Perform molecular docking of known active ligands and decoys.
4. Evaluate the homology models by using BEDROC to select good performing models of the orthosteric and the allosteric binding sites.
5. Use SBVS to dock all compounds from LBVS into selected models and identify putative GAT-inhibitors.

3 METHODS

3.1 Software and databases

3.1.1 Software

Schrödinger Release 2016-3

The Schrödinger software is a computational technology software package, used for molecular modelling in pharmaceutical, chemical, biotechnology and materials sciences research, and is also the leading software for drug design and development (<https://www.schrodinger.com/maestro>). In this study the Schrödinger software was used for homology modelling, sketching of active ligands, docking (Glide), induced fit docking (IFD), preparation of ligands before docking (Ligprep), preparation of proteins before docking (protein preparation wizard), and for virtual screening (virtual screening workflow).

Schrodinger Canvas

Canvas is a cheminformatics package of Schrödinger used for structural and data analysis. Canvas can be used for ligand fingerprinting and similarity searching, substructure searching, selection of compounds by diversity, structural clustering of compounds, and building regression and classification models (40). The software was used for 2D fingerprinting and clustering of 72 known active ligands and 100 best scoring hits from the ligand based screening in the ENAMINE database. Similarity matrixes were also generated using this application.

Molsoft Internal Coordinates Mechanics Software (Version

The Internal Coordinate mechanics (ICM) is a software with different application for molecular modelling and drug design (<http://www.molsoft.com/technology.html>). In this project the ICM software was used to convert decoys, from SMILES code to 2D structures.

3.1.2 Databases

Protein Data Bank

The Protein Data Bank (PDB) is a worldwide and freely available database that contains 3D structures of macromolecules obtained by X-ray crystallography, NMR spectroscopy, or electron microscopy. The 3D structures in the PDB are large biological molecules of proteins and nucleic acid from a variety of organisms. The PDB has many users from biologists, chemists, scientists, students, educators, media writers, illustrators, textbook authors, and to the general public (41). In this study, the X-ray crystal structures of the templates used for homology modeling were downloaded from the PDB database (<http://www.rcsb.org/pdb/home/home.do>).

The Universal Protein Resource Knowledgebase

The Universal Protein Resource Knowledgebase (UniProtKB) is a database that provides information about protein sequences and their functions. UniProtKB consisting of two different section: Manually annotated (UniProtKB/Swiss-Prot) and computationally analyzed (UniProtKB/TrEMBL) sequenced. UniProtKB/Swiss-Prot was used to download the amino acid sequences of GAT-1, GAT-2, GAT-3 and BGT-1 prior to building homology models of GAT-3 and construction of a multiple sequence alignment (<http://www.uniprot.org/>).

Databased of Useful Decoys: Enhanced

Databased of Useful Decoys: Enhanced (DUD.E) is a database with useful decoys. Decoys are molecules that have similar physiochemical properties as active compounds, but with different chemical structures so that they are assumed to be non-binders. In this study the DUD.E database was used to generate decoys for each of the 72 active ligands in a ratio 50:1 (50 decoys for each active compound) (<http://dude.docking.org/generate>).

ENAMINE database

The ENAMINE database is collection of known active compounds for biological targets, and is often utilized in the drug discovery process. The Enamine subset of advanced Collection with 294 995 active compounds was used in this study (<http://www.enamine.net/>).

3.2 Workflow of the study

An approach combining LBVS and SBVS were performed in order to predict new putative GAT-3 inhibitors. The 72 known active ligands were clustered into five structural groups, and four types of fingerprint were calculated and the cut-off value was determined. The cut-off value is the minimum value of that fingerprint among the known binders. The calculated fingerprints were used to search the ENAMINE database for similar structures. This resulted in 2150 compounds identified by using ligand fingerprinting. Homology models of human GAT-3 were built based on different templates, and 72 active ligands and decoys were docked into the homology models to select the most appropriate models for the SBVS approach. Boltzmann-Enhanced Discrimination of Receiver-Operation Characteristics (BEDROC) statistical analyses and the number of known actives docket were used to select among the different models. The selected models were improved by IFD and active ligands and decoys were re-docked into the new ligand-optimized conformations of the different GAT-3 models, and BEDROC calculations were repeated. Based on the docking, four models were selected for docking of the ENAMINE compounds. The 2150 compounds from the ligand based search in ENAMINE were prepared for docking by Ligprep (increased to 5876 compounds) and docked by virtual screening. The 100 compounds with the best docking score in each model were clustered into ten structural clusters and studied more in details. A workflow of the study is presented in Figure 6.

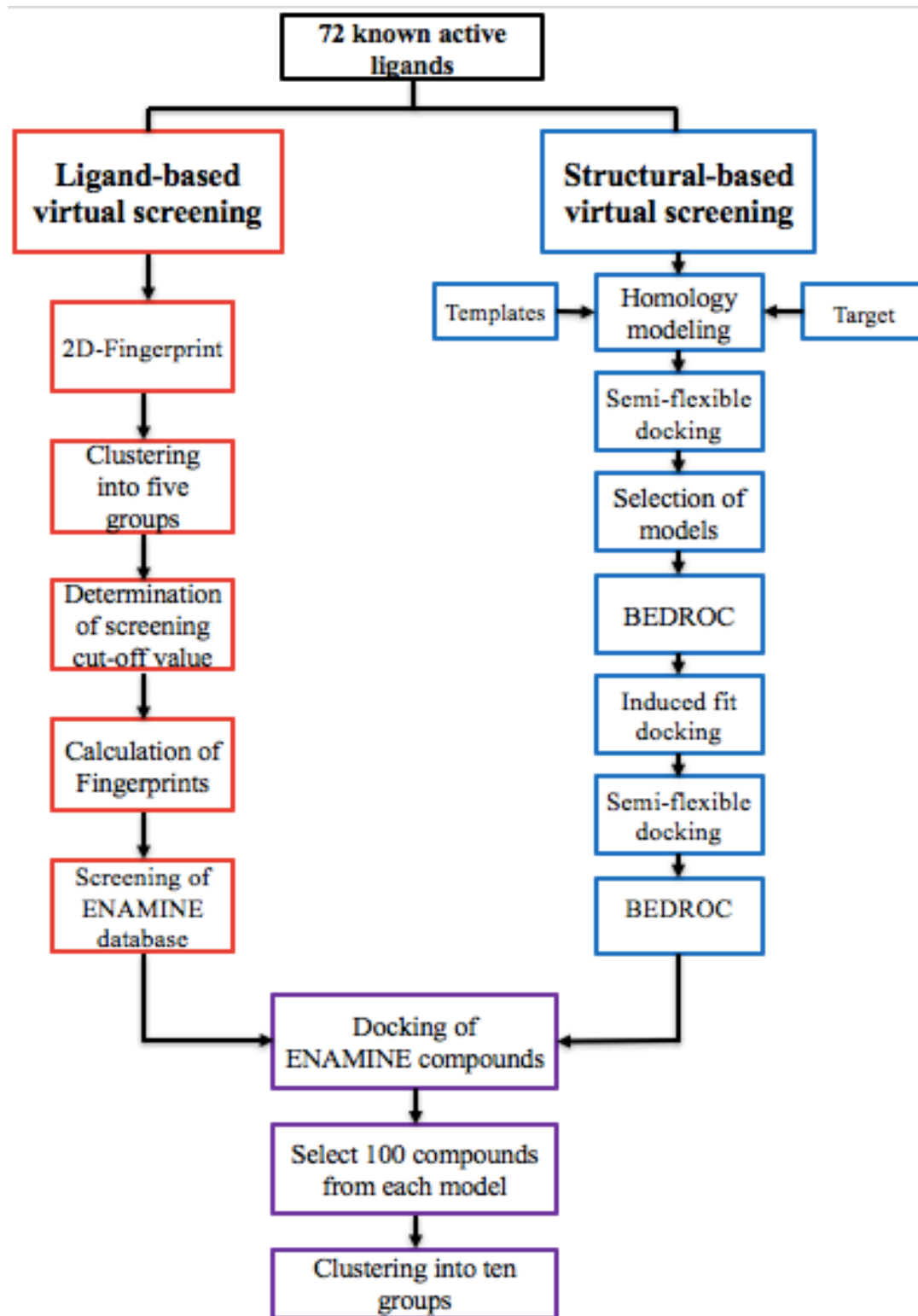


Figure 6: Workflow of the study

3.3 Selection of active compound of GAT-3 and decoys

A set of known binders of GAT-3 was obtained from the literature. A total of 72 active ligands were selected based on appropriate inhibition of GABA uptake (Appendix 1). As the structure of the compounds could not be found in any database, they were manually sketched in Maestro using 2D sketch (version 2016-3).

A total of 3849 decoys were generated by using the 72 active ligands as references and retrieved from DUD.E (<http://dude.docking.org/generate>). The decoys were downloaded as SMILES and converted into 2D structures by the ICM software and then imported to Maestro.

3.4 Ligand-based virtual screening

3.4.1 Clustering of active ligands

All the 72 active ligands were imported into Canvas to perform structural similarity clustering. In this study the Hierarchical clustering method with the radial fingerprint type was used to cluster the active ligands. Hierarchical clustering is a method that cluster up to 5000 compounds based on their structural similarity and present them in a dendrogram. The dendrogram has three methods for clustering the compounds. The compounds can be clustered by the number of clusters, by the Kelly criterion or by the merging distances. The 72 active compounds were clustered into five groups based on Kelley criterion with some manual modifications. The Kelly criterion is a measure that finds the most adequate specific clustering numbers for the input compounds. The similarity matrix based on the chosen fingerprint was performed to calculate and appraise the selected radial fingerprint.

3.4.2 Calculation of model fingerprints

Linear, Radial, Dendritic and MOLPRINT2D fingerprints were generated with the purpose of selecting the most appropriate for this particular set of ligands (40). The active ligands and decoys for each cluster were imported to Canvas and the four fingerprint types were calculated by using binary fingerprints from of known structures of all five clustered. The modal fingerprints were used to calculate average fingerprints of active ligands, which collect information from several query molecules into a single fingerprint (40). Active ligands in each cluster were used to create modal fingerprints for all five clusters.

3.4.3 Determination of screening cut-off values

Modal fingerprints for each of the five clusters of known compounds were calculated to screen the generated decoys. Canvas has 24 different indices to calculate similarity between the modal fingerprints and the structures in the active ligands and decoys, but in this study only Tanimoto similarity metrics was used. The structures of active binders and decoys for each cluster were sorted by their similarity to the modal fingerprints. Those with the highest similar value (max binder) to the modal fingerprints were sorted at the top and those with the lower value (min binder) were sorted below. The cut-off values (minimum binder values) are the minimum similarity to the modal fingerprints of the active ligands.

3.4.4 Screening of ENAMINE database

The 294 995 compounds of the ENAMINE database advanced collection were downloaded, and the modal fingerprints calculated for the active ligands were used to screen the compounds. The cut-off value (min binder) from active ligands for each fingerprint types and cluster were used to select ENAMINE structures that were above or equal to the cut-off values. All structures that had a value above or equal to the cut-off were exported and collected. The procedure comprised duplicate compounds that were removed.

3.5 Homology modeling

The 3D structure of GAT-3 has not been resolved. Homology modelling was used to construct 3D models of human GAT-3 based on the resolved X-ray crystal structures of drosophila DAT (dDAT) and human SERT (hSERT) for later use in the discovery of putative new GAT-3 inhibitors.

3.5.1 Template identification

Available X-ray crystal structures of dDAT and hSERT were downloaded from the PDB database (<http://www.rcsb.org/pdb/home/home.do>). Today there are several X-ray crystal structures of dDAT, but only the four dDAT (4XP4, 4XP9, 4XPA, 4XNU and 4XNX) with highest resolution were selected. We were mainly interested in inhibitor induced template for

our screening. However, the dDAT complex with a substrate having highest resolution was included.

Only two of the available hSERT structures (PDB ID: 5I6X and 5I73) were selected for this study. 5I6X was the one with the highest resolution, while the 5I73 crystal structure was selected as a template because it has ligands co-crystallized both in the orthosteric and allosteric binding site.

Table 3: X-ray crystal structures that were utilized as the templates in homology modeling of GAT-3.

PDB	Receptor	Ligand	Resolution (Å)	Binding site occupied by ligand
4XP4	dDAT	Cocain (inhibitor)	2.8	Orthosteric
4XP9	dDAT	D-amphetamine (Substrate)	2.8	Orthosteric
4XPA	dDAT	3,4-dichlorophenethylamine (Substrate)	2.95	Orthosteric
4XNU	dDAT	Nisoxetine (NRI*)	2.98	Orthosteric
4XNX	dDAT	Reboxetine (NRI*)	3.0	Orthosteric
5I6X	hSERT	Paroxetine (SSRI*)	3.14	Orthosteric
5I73	hSERT	Citalopram (SSRI*)	3.24	Orthosteric & allosteric

NRI*: Norepinephrine reuptake inhibitor, SSRI*: Selective serotonin reuptake inhibitor

3.5.2 Target-template Alignment

The complete amino acid sequence of GAT1, GAT2, GAT3 and BGT-1 were obtained from the UniProt database (www.uniprot.org) with accession numbers: P30531, Q9NSD5, P48066 and P48065, respectively. GAT-3 with P48066 Uniprot accession number was used as the query protein (target). The amino acid sequence of the four GATs was aligned with the template sequences by using the multiple sequence alignment in Maestro (version 2016-3).

3.5.3 Building the model

GAT-3 homology models were built based on the obtained alignment of the GATs and crystal structure (PDB ids) 4XP4, 4XP9, 4XPA, 4XNU, 4XNX, 5I6X and 5I73. The models were built by using the knowledge-based method of the Maestro software (version 2016-3), that use segments from known PDB structures for closing the alignment gaps. The ions present in the different template were included in the models. 5I73 was the only template with both an allosteric and an orthosteric binding site.

3.5.4 Evaluation of models

The constructed homology models were evaluated using the SAVES metaserver. The programs PROCHECK, ERRAT and Verify_3D were used to validate the models. PROCHECK is a program that evaluates the stereochemical quality of the protein structures and the residue geometry. The result of PROCHECK analysis may be represented on a graphical Ramachandran plot that shows the backbone dihedral angles phi (Ψ) and psi (Φ) of the structure and amino acids in allowed and disallowed regions. ERRAT calculates the non-bonded interactions between different atom types. Verify_3D evaluates the compatibility of the 3D structure with own amino acid sequence (1D) and comparing the result of the model with a good quality model (SAVES; <http://nihserver.mbi.ucla.edu/SAVES/>).

3.6 Structure-based virtual screening

3.6.1 Active ligands and decoys preparation

The active ligands and the generated decoys were prepared by using LigPrep before the docking procedure in Maestro. Despite chemically correct input structures, the geometry of the compounds may not be correct, and in addition to protonation states need attention. LipPrep was used to adjust the geometry by minimize the energy with correct chiralities of ligands and decoys (42). In this study, ligands and decoys were ionized at pH of 7.2 +/- 0.2 to conform the physiological conditions, and at most 1 stereoisomer per ligand.

3.6.2 Semi-flexible docking

Before docking calculations, the Receptor Grid Generation in maestro was used to define binding sites in the homology models. The Receptor Grid Generation has several options and the van der Waals radius scaling factor was set to 1.0 and partial charge cutoff 0.25 was used to avoid close contacts of atoms. The co-crystallized ligand of the template was used to specify the binding site in the homology models generated based on the template. In this study, the grid box was selected around the ligand in each model with an inner box of 10 Å and an outer box of 25 Å. The calculated receptor grids (grid maps) for each target were used to dock the prepared 72 active binders and 3849 decoys.

3.6.3 Induced fit docking

In a standard docking calculation the ligands are flexible and the target homology model rigid (semi-flexible docking). Protein structures have side-chain or backbone motion and are very flexible in nature, but with rigid binding sites in the docking process, this flexibility is not taken into account. Induced fit docking (IFD) treats the ligand as flexible and also include flexibility of the amino acids in the binding pocket of the target. A strategy is therefore to do an initial semi-flexible docking followed by IFD that optimize the receptor conformation in the presence of a ligand. Before starting the IFD process, the protein structures were

prepared using the Protein Preparation Wizard in Maestro. Hydrogen bonds were optimized, and the proteins were minimized in terms of energy.

The binding site for each model was defined by selecting all the residues within 4 Å sphere radius around the ligand, using the ligand co-crystallized with the template as a reference. For each cluster the ligand with highest binding affinity from ligand binding studies, and the ligand which obtained the highest docking score from the initial semi-flexible glide docking were selected for the IFD, giving altogether 13 complexes for IFD. After IFD, the binding site of the 13 complexes was defined by the active ligand that was used in IFD to re-dock the active ligands and decoys into receptor obtained by IFD.

3.6.4 Evaluation of homology models with BEDROC

BEDROC was used to evaluate the homology models ability to distinguish between active compounds and decoys on a graphical illustration. The BEDROC calculation was first performed with a semi-flexible docking protocol, but the BEDROC score was so low that models had to be improved with IFD protocol in maestro (version 2016-3) before selection of models for use in the virtual screening. BEDROC calculations were also performed after re-docking into the conformations from IFD (Figure 10 and 11).

3.6.5 Docking the compounds obtained by the ligand based approach

The models with the best BEDROC score were selected for virtual screening of the compounds obtained by the ligand-based approach. The selected models that included the orthosteric binding site were generated from the following template; 4XPA, 4XP4, and 5I73, while the selected model of the allosteric binding site that was generated from the 5I73 template. LigPrep was used to prepare the 2151 compounds (obtained by ligand-based screening) from ENAMINE in the same manner as active ligands and decoys. At most ten stereoisomers were generated giving a total of 5876 ligands for docking. The virtual screening workflow has three different docking stages: high throughput virtual screening (HTVS), standard precision (SP) and extra precision (XP). HTVS is the fastest screening method and can be used for a very large numbers of compounds. SP docking has more conformational

arrangements than HTVS docking and is used for screening of a large number of compounds with unknown quality. XP is the slowest, but a powerful method with more strict procedures and many scoring functions (43).

The 5876 compounds were used as a source of ligands for docking into the orthosteric binding site of 4XPA, 4XP4 and 5I73 based models, and the allosteric binding site of the 5I73 based model. For each model grids for the receptor had to be specified to be able to dock the 5876 compounds. All the three docking stages were used in this study. The first stage was HTVS docking, the second was SP docking and the last was XP docking. In each docking stages 100% of best compounds were kept after docking.

After virtual screening, the 100 compounds with highest docking score in each model were clustered into ten clusters by calculation their radial fingerprints. The compound with the best docking score in each model was also selected for IFD to optimize the interactions. The chosen compounds were prepared by using the Protein Preparation Wizard before IFD was performed.

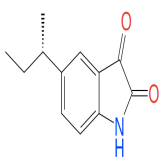
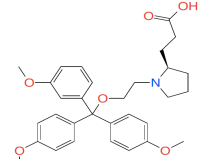
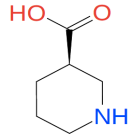
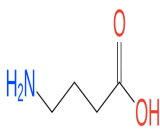
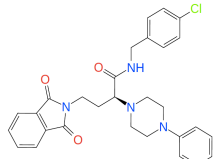
4 RESULTS

4.1 Ligand-based virtual screening

4.1.1 Clustering of active ligands

Hierarchical clustering with 2D fingerprints were used to cluster 72 active GAT-3 binders retrieved from the scientific literature (Appendix 1) and 3849 property matched compounds that were used as decoys. The 72 active compounds used as reference ligands consisted both of substrates and inhibitors. In addition, it is also reasonable to believe that some of them bind solely to the orthosteric site and others to the allosteric site, while some may bind to both sites. However, most probably most of them are orthosteric binders. Four types of fingerprints (Linear, Radial, Dendritic, MOLPRINT2D) were calculated for the active ligands and decoys. However, the Radial fingerprints and application of the Kelly criterion were used to cluster the 72 active ligands and the 3849 decoys into five clusters: cluster 1, cluster 2, cluster 3, cluster 4 and cluster 5 (Table 4). Most of the active ligands were in cluster 2 and only three compounds were in cluster 5. These three active ligands were outliers with different structures that did not fit with any of the other clusters and therefore they were clustered together manually.

Table 4: **Five clusters of active ligands.** The 72 active ligands and their 3849 property matched decoys clustered into five groups by radial fingerprints. One representative active ligand from each cluster is shown.

	Reference active ligands				
	Cluster 1	Cluster 2	Cluster3	Cluster 4	Cluster 5
Structures					
# active ligands	21	32	9	7	3
# decoys	1100	1699	550	350	150

In addition the Tanimoto similarity metrics of the 72 active ligands were calculated to examine the similarity between the five clusters. The Tanimoto similarity metrics values range between 0 and 1, where 0 means most dissimilar compounds, while 1 means most similar compounds, but it does not mean that the compounds are identical. Some of the active ligands have R and S enantiomers that are classified as similar to each other, but otherwise none of the active binders were identical to each other.

4.1.2 Determination of screening cut-off values

The modal values of the four fingerprint types were calculated for each cluster of compounds and are shown in Table 5. These modal values (screening cut-off values) were used to perform the 2D-fingerprint screening in the ENAMINE database. Each cluster and fingerprint type had a cut-off value (min binder). Table 5 shows that only three decoys in cluster 2 had values (Dendritic fingerprint) above the min value for active (three false positives).

Table 5: **Determination of screening cut-off values for the 2D fingerprints:** min - the minimum fingerprint value within the particular cluster, max - the maximum fingerprint value within the particular cluster and numbers of decoys (# decoys) that are above the cut-off value for each fingerprint type.

Reference ligand group	Value	Fingerprint type			
		Linear	Radial	Dendritic	MOLPRINT2D
Cluster 1	Max active	0.193	0.108	0.233	0.198
	Min active	0.095	0.064	0.157	0.136
	Max decoys	0.041	0.035	0.073	0.056
	# decoys	0	0	0	0
Cluster 2	Max active	0.284	0.098	0.271	0.235
	Min active	0.046	0.040	0.071	0.096
	Max decoys	0.041	0.028	0.091	0.092
	# decoys	0	0	3	0
Cluster 3	Max active	0.598	0.310	0.493	0.310
	Min active	0.171	0.172	0.155	0.207
	Max decoys	0.135	0.135	0.117	0.029
	# decoys	0	0	0	0
Cluster 4	Max active	0.374	0.261	0.400	0.273
	Min active	0.144	0.148	0.148	0.182
	Max decoys	0.071	0.071	0.068	0.048
	# decoys	0	0	0	0
Cluster 5	Max active	0.366	0.480	0.492	0.522
	Min active	0.350	0.293	0.258	0.348
	Max decoys	0.062	0.081	0.108	0.100
	# decoys	0	0	0	0

4.1.3 Screening of ENAMINE database

The modal fingerprints values (Table 5) were used for screening of the ENAMINE database, selecting compounds with fingerprint values above the min value of the active compounds. A subset of the ENAMINE database consisting of 294 995 drug like compounds were screened. Cluster 2 was the cluster with the highest number of identified compounds in the ENAMINE database. 1840 compounds had Dendritic fingerprint above the min value of cluster 2, while no compounds of the screened dataset had fingerprint values above the min value of cluster 5. Dendritic fingerprints gave the highest number of hits compared to the others. After removing of duplicates, a total 2151 compounds were identified by the ligand-based approach (Table 6).

This corresponds to 0.73% of the screened ENAMINE dataset of compounds. These compounds were then used for the structure-based virtual screening.

Table 6: **Number of hits from the ligand-based screening.** The cut-off values for each fingerprint type (min active in Table 5) were used to identify compounds with higher or equal value in the dataset. The sum of hit compounds was reduced because the duplicates were removed.

Reference ligand group	Fingerprint type	Fingerprint threshold	Number of compounds
Cluster 1	Linear	0.095	9
	Radial	0.064	2
	Dendritic	0.157	14
	Molprint2D	0.136	10
Cluster 2	Linear	0.046	354
	Radial	0.040	0
	Dendritic	0.071	1840
	Molprint2D	0.096	154
Cluster 3	Linear	0.171	74
	Radial	0.172	0
	Dendritic	0.155	173
	Molprint2D	0.207	0
Cluster 4	Linear	0.144	11
	Radial	0.148	0
	Dendritic	0.148	21
	Molprint2D	0.182	0
Cluster 5	Linear	0.350	0
	Radial	0.293	0
	Dendritic	0.258	0
	Molprint2D	0.348	0
SUM (non-redundant)			2662 (2151)

4.2 Homology modeling

Homology modeling was used to build seven 3D models of the target GAT-3 using the resolved crystal structures of dDAT (PDB ID: 4XP4, 4XP9, 4XPA, 4XNU, 4XNX) and hSERT (PDB ID: 5I6X, 5I73) as templates. In addition to the target (GAT-3) and the template, the sequences of GAT-1, GAT-2 and BGT-1 were also included in the multiple sequence alignments used for building the models (Figure 7 and 8). Multiple sequence alignments were performed, and the amino acid sequence identity between the template and the different targets were calculated. The sequence identities with GAT-3 were 40% for dDAT (4XP4, 4XP9, 4XPA, 4XNU, 4XNX) and 38% for hSERT (5I6X and 5I73). All the build models had characteristic the SLC6 transporters including the 12 TMs with extracellular and intracellular loops. The extracellular loop between TM3 and TM4 and the intracellular N- and C-terminals had N-linked glycosylation sites.

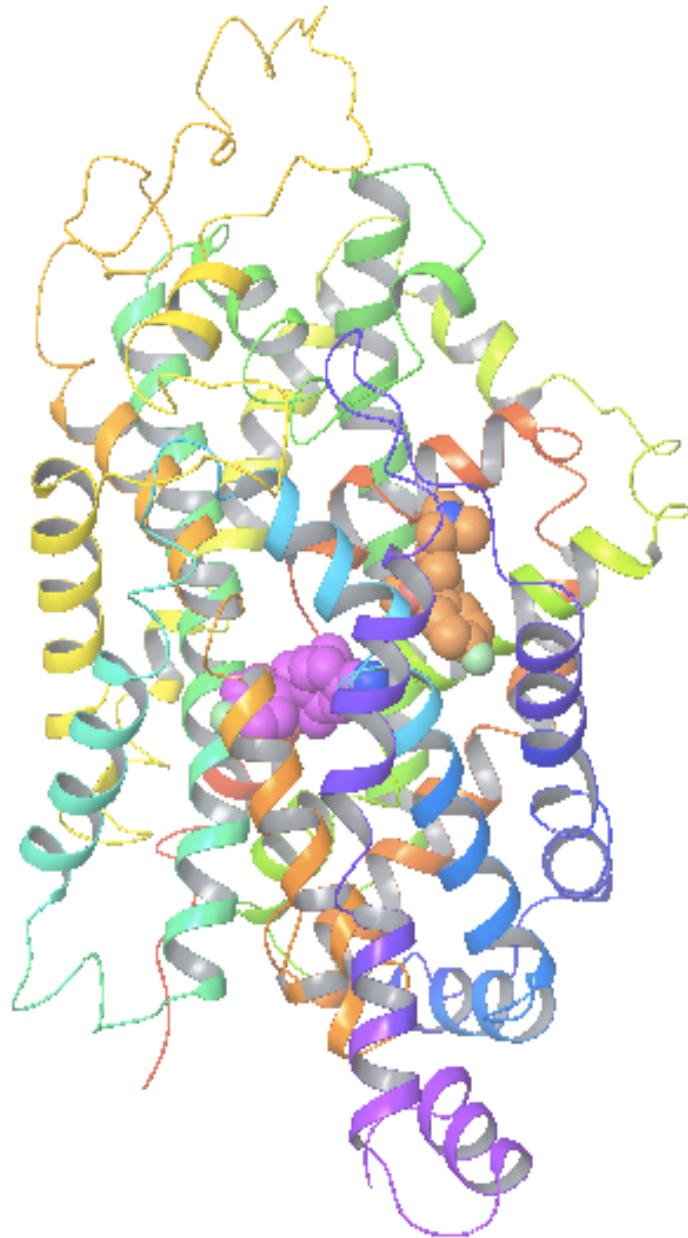


Figure 7: **Homology model of 5I73-based model with orthosteric binding site (purple) and allosteric binding site (orange) defined by the co-crystallized ligand of the template.** The extracellular side is up in the figure.

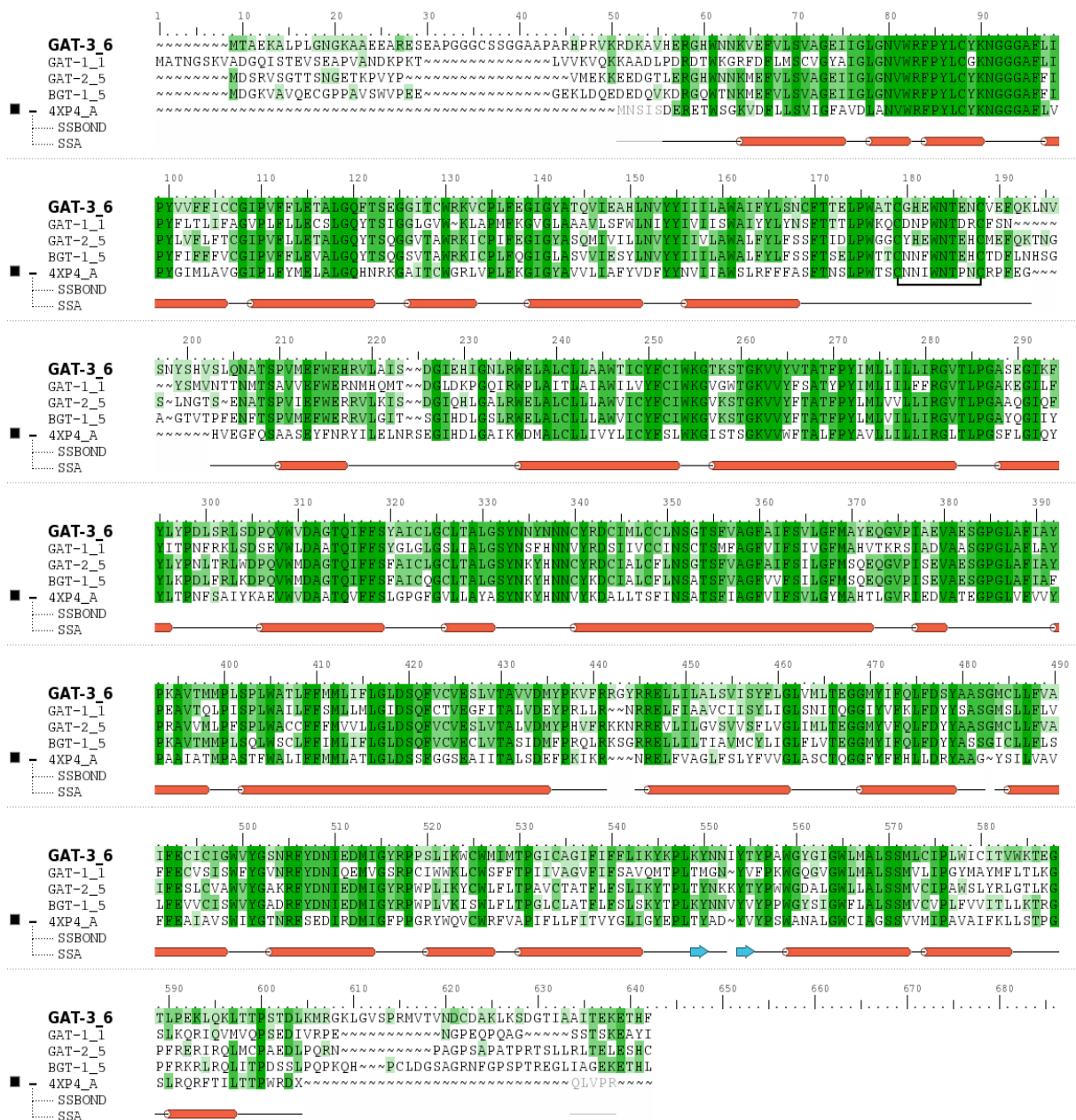


Figure 8: Multiple sequence alignment of GAT-3, GAT-1, GAT-2 and BGT-1 and dDAT (4XP4). The dark green color indicates identical residues. The red cylinders are a-helices, the blue arrows are b-sheets and the black box is a disulfide bond in the structure of the template.

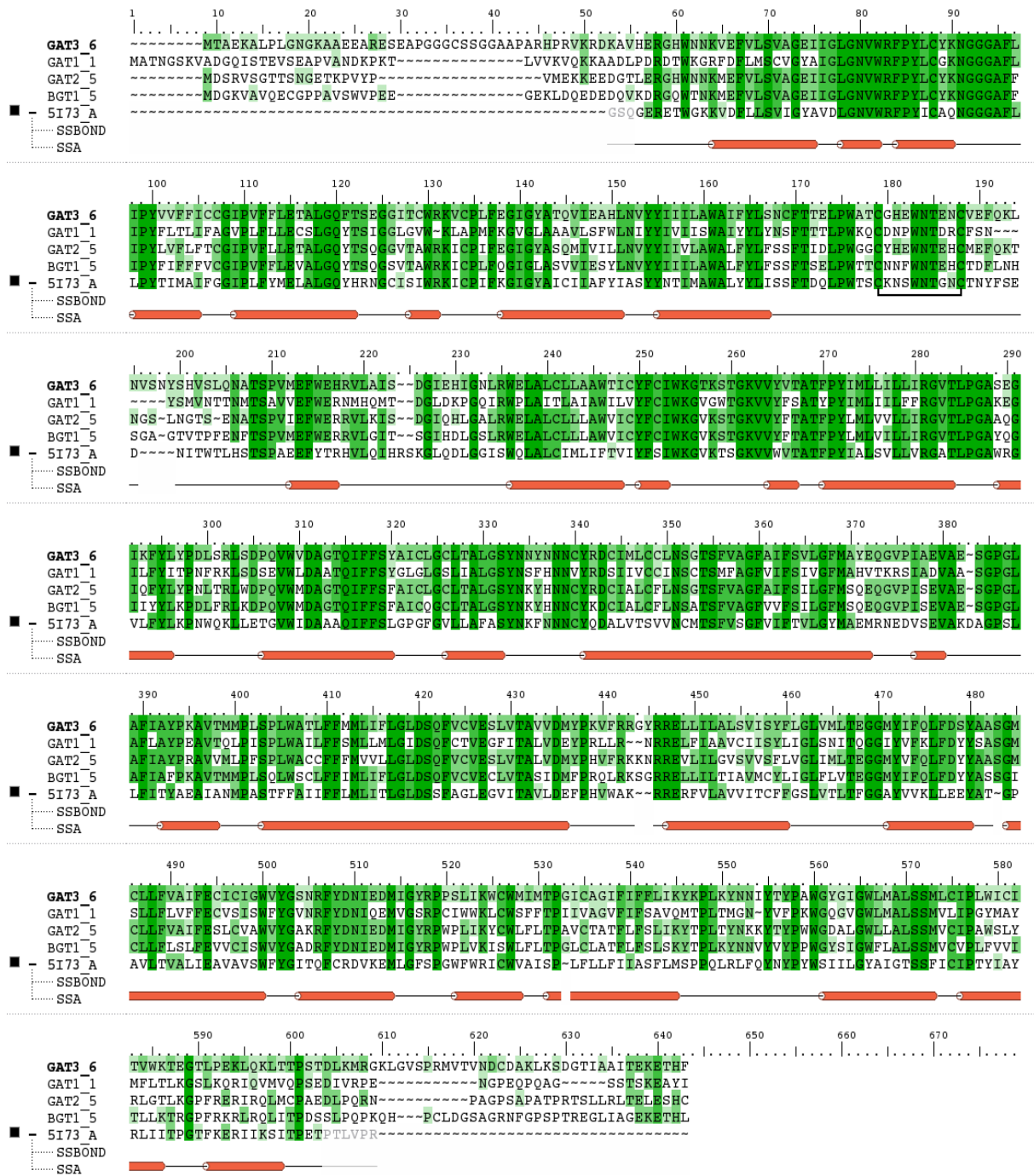


Figure 9: Multiple sequence alignment of GAT-3, GAT-1, GAT-2 and BGT-1 and hSERT (5I73). The dark green color indicates identical residues, the red cylinders are α -helices, and the black box is a disulphide bond in the structure of the template.

The stereochemistry and geometry of the homology models were evaluated by PROCHECK, ERRAT and Verify (Table 7). The PROCHECK results were presented by Ramachandran plots that shows the amount (%) of residues in most favoured, additionally allowed, generously allowed and disallowed regions. According to PROCHECK, a model with over 90% of amino acid in the most favoured regions is a model with good quality. 4XP4, 4XP9, 4XPA, 4XNX based models had more than 90% of the amino acid in the most favoured regions and 4XNU, 5I6X, 5I73 had less than 90% in most favoured regions, but the 5I73-based model was close to the acceptable value. Based on these results, we can conclude that 4XP4, 4XP9, 4XPA, 4XNX, 5I73- based models were satisfactory quality models.

The ERRAT program was also used for quality check. A model with an ERRAT value higher than 50 is consider to be a high quality model (44). The higher the score of ERRA is, the better is the quality of the model and all the models had an ERRAT score above 70.2, which are better than the acceptable range. The table also shows the Verify_3D value of all models. Except for 5I73-based models, at least 80% of the amino acids had an average 3D-1D score > 0.2.

Table 7: **Homology model evaluation.** The table shows a summary of the Ramachandran plot statistics and the ERRAT, and Verify_3D values in percent for all the homology models that were build in this study.

Homology Models	PROCHECK				ERRAT	Verify_3D
	Most favored regions	Additionally allowed regions	Generously allowed regions	Disallowed regions		
4XP4-based model	92%	6.1%,	1.5%,	0.4%	76.1	87.4%
4XP9-based model	92.3%	6.7%	0.4%	0.6%	85.4	85.0%
4XPA-based model	91.8%	6.5%	1.1%	0.6%	85.7	88.1%
4XNU-based model	88.6%	10.1%	0.8%	0.4%	82.0	85.2%
4XNX-based model	92.7%	6.5%	0.2%	0.6%	-	88.5%
5I6X-based model	88.8%	8.6%	1.5%	1.1%	75.0	83.03%
5I73-based model	89.5%	9.1%	0.6%	0.8%	76.0	79.0%

- The program ERRAT score was not obtained.

4.3 Structure-based virtual screening

4.3.1 Semi-flexible docking

The seven homology models were evaluated by docking the library 72 active binders and the 3849 property matched decoys. The active ligands were clustered into five structural cluster during the LBVS, and the same ligand clustering was used in the docking evaluation of the models in order to select models for SBVS. The binding site of the models was identified by using the ligand in the co-crystallized complex with the template to define the binding pocket. The initial docking of the library gave bad BEDROC values, and 12 complexes were therefore selected for IFD (Table 8). These complexes were four 4XPA-based complexes, three 4XP4-based complexes, one 4XP9- based complex, and one 5I73-based with ligands in the orthosteric sites, and three 5I73-based complexes with ligands in the allosteric sites.

From each cluster, the complex with highest docking score and the complex with the ligand with highest experimental detected activity (Appendix 1) were selected. The selected 4XPA-based model, 4XP4-based model, the orthosteric 5I73-based model, the allosteric 5I73-based model and 4XP9-based model were able to dock respectively 36, 71, 65, 72 and 46 of the active ligands. In cluster 2, the complex with the ligand NNC05-0341 in the allosteric site was included for IFD since the complex had the best docking score in the allosteric site of that cluster and the model was able to dock all the 72 active ligands in the allosteric site. In cluster 4, GABA was the ligand with the highest experimental activity and had the highest docking score in 5I73 orthosteric model. The allosteric site of this model was therefore also further tested with GABA. The docking indicated that GABA has higher score in the orthosteric binding site than in the allosteric, but quite low docking score in both. The complexes selected from cluster 1 had the highest ligand docking score of all.

Table 8: **Complexes selected for IFD.** From each cluster, the complex with highest docking score and the complex with the ligand with highest experimental detected activity were selected. However, only one complex with cluster 5 compounds in the orthosteric site was selected. Uptake inhibition corresponds 50% inhibition of substrate transport.

Ligand cluster	Model	Active ligand	Docking score	Uptake inhibition (μM)	Reference
Cluster 1	4XPA-based model	Compound 16	-8.97	15	(14)
	4XPA-based model	Compound 20	-8.45	6	(14)
Cluster 2	4XP4-based model	(R)-5d	-7.99	11.2	(45)
	4XP4-based model	DDPM-1457	-5.81	5.87*	(46)
	5I73 allosteric-based model-I	NNC05-0341	-7.56	2.8**	(47)
Cluster 3	4XPA-based model	(R)-nipecotic-acid	-6.44	14	(48)
	4XPA-based model	(R)-Isoserine	-6.41	4.3	(48)
Cluster 4	4XP9-based model	Beta_alanine	-5.73	12	(48)
	5I73 orthosteric-based model	GABA	-4.70	1.4	(14)
	5I73 allosteric-based model	GABA	-4.02	1.4	(14)
Cluster 5	4XP4-based model	Clomipramine	-6.84	40.2	(49)
	5I73 allosteric – based model	Compound18	-5.42	8.5	(50)

* pIC₅₀: Logarithmic half maximal inhibitory concentration.

** K_i: Binding affinity constant

4.3.2 Induced fit docking scores

The IFD was performed for complexes selected based docking score (the complex with highest score in each cluster) from the initial semi-flexible docking and the complex with the most active ligand (best experimental value). The IFD optimize the binding site and ligand interactions of the selected complex, and generates additional conformations of the binding site. The IFD improved the docking score of most of the selected ligands. The active ligands in cluster 2 had the highest docking score and the selected cluster 4 ligands had the lowest docking score after IFD. The docking score of Beta-alanine was lower after performing IFD than before, while the docking score of GABA in 5I73-allosteric did not improve much

during IFD. The active ligand Compound 18 in cluster 5 did not fit into the orthosteric binding site in the initial semi-flexible docking, but after IFD the ligand was re-docked and obtained the highest docking score of all the ligands in that clusters (Table 9). That complex was therefore further evaluated, and the model conformation was used in re-docking of the active and decoys.

Table 9: **Docking score and IFD score after IFD of the selected complexes.** IFD generates new models that were used for docking. For five clusters with each one active ligand with high binding affinity from ligand binding studies, and one binder which obtained high docking score from the initial semi-flexible glide docking with the selected models. Uptake inhibition corresponds 50% inhibition of substrate transport.

Ligand cluster	Model	Active ligand	Docking score	IFD score	Uptake inhibition (μM)
Cluster 1	4XPA-based model-I	Compound 16	-10.38	-998.82	15
	4XPA-based model-II	Compound 20	-10.36	-997.67	6
Cluster 2	4XP4-based model-I	(R)-5d	-12.06	-1004.75	11.2
	4XP4-based model-II	DDPM-1457	-12.08	-10005.16	5.87*
	5I73 allosteric-based model-I	NNC05-0341	-9.48	-879.23	2.8**
Cluster 3	4XPA-based model-III	(R)-nipecotic-acid	-8.15	-994.32	14
	4XPA-based model-IV	(R)-Isoserine	-6.18	-988.90	4.3
Cluster 4	4XP9-based model-I	Beta-alanine	-4.64	-990.92	12
	5I73 orthosteric-based model-I	GABA	-6.01	-860.27	1.4
	5I73 allosteric-based model-II	GABA	-4.68	-859.01	1.4
Cluster 5	4XP4-based model-III	Clomipramine	8.71	-996.94	40.2
	5I73 allosteric – based model-III	Compound18	-8.56	-943.32	8.5
	5I73 orthosteric–based model-II	Compound18	-12.36	-944.34	8.5

*pIC₅₀: Logarithmic half maximal inhibitory concentration.

** K_i: Binding affinity constant

4.3.3 Semi-flexible docking of active ligands into receptor obtained by IFD

After IFD, the selected models were evaluated based on docking of all the 72 active ligands and 3849 property matched decoys by using the ligand in the IFD complexes in each cluster to define the binding pocket of the model (Table 10). Good performing models should select the known binders in front of the decoys. The selected models based on cluster 2 and cluster 5 ligands were able to dock 71-72 active ligands, while the 4XPA-based model-III in cluster 3 was able to dock only 29 active ligands (Table 10). Cluster 2 had ligands with highest docking score and cluster 4 had active ligands with the lowest docking score. Figure 12-15 shows active ligands with the highest docking score for 4XPA-based model, 4XP4-based model, 5I73 allosteric-based model and 5I73 orthosteric-based model. The BEDROC scores for all models in each cluster were calculated and the models with the best BEDROC score were selected for docking the ENAMINE compounds.

Table 10: **The docking evaluation of the models from Table 9.** The table shows the docking score of the compounds from table 8 and 9 with their experimental uptake inhibition, and the number of active docked to the particular GAT-3 conformation and the experimental uptake inhibition.

Cluster	Model	Active ligand	Docking score	#Ligands docked	Uptake inhibition (μM)
Cluster 1	4XPA-based model-I	Compound 16	-10.43	38	15
	4XPA-based model-II	Compound 20	-9.74	38	6
Cluster 2	4XP4-based model-I	(R)-5d	-13.15	72	11.2
	4XP4-based model-II	DDPM-1457	-10.97	72	5.87*
	5I73 allosteric-based model-I	NNC05-0341	-9.44	71	2.8**
Cluster 3	4XPA-based model-III	(R)-nipecotic-acid	-7.83	29	14
	4XPA-based model-IV	(R)-Isoserine	-5.59	48	4.3
Cluster 4	4XP9- based model-I	Beta-alanine	-4.52	48	12
	5I73 orthosteric-based model-I	GABA	-6.11	45	1.4
	5I73 allosteric-based model-II	GABA	-4.34	67	1.4
Cluster 5	4XP4-based model-III	Clomipramine	-8.75	71	40.2
	5I73 orthosteric-based model-II	Compound18	-11.52	72	8.5
	5I73 allosteric – based model-III	Compound18	-8.72	72	8.5

* pIC50: Logarithmic half maximal inhibitory concentration.

** K_i: Binding affinity constant

4.3.4 Evaluation of the models, BEDROC scores

BEDROC were performed to evaluate the ability of the models to separate between active ligands and the decoys. The BEDROC was first calculated for the initial semi-flexible docking (results not shown) and then for the finally selected models (Table 11). The BEDROC values are between 1 and 0, where 1 indicates best performance to separate between the active ligands and decoys and 0 indicates that the model do not separate decoys from active ligands. The BEDROC value was calculated with $\alpha=20$, that weights the first ~ 8% of the screen results. The 5I73 orthosteric-based model was the best model with BEDROC value of 0.995, which mean that the model had better distinguishing properties

between active and decoys than the other models. The 4XPA-based model-IV had the lowest BEDROC score. Based on this evaluation, models were selected for the structure-based screening of the 2151 hits from the ligand-based approach that were prepared by Ligprep, and most ten stereoisomers of each were included giving a total of 5876 compounds for screening.

Table 11: **BEDROC score of models after IFD**. Four were finally selected for the structure based virtual screening for the 5876 compounds obtained by ligand-based screening.

Cluster	Model	BEDROC score ($\alpha=20$)
Cluster 1	4XPA-based model-I	0.261
	4XPA-based model-II	0.216
Cluster 2	4XP4-based model-I	0.368
	4XP4-based model-II	0.369
Cluster 3	5I73 allosteric-based model-I	0.116
	4XPA-based model-III	0.175
	4XPA-based model-IV	0.047
Cluster 4	4XP9- based model-I	0.084
	5I73 orthosteric-based model-I	0.095
	5I73 allosteric-based model-II	0.054
Cluster 5	4XP4-based model-III	0.147
	5I73 orthosteric-based model-II	0.995
	5I73 allosteric – based model-III	0.049

Four final models were selected based on best BEDROC score and docking score to perform virtual screening of the 5876 ENAMINE compounds. These four models were selected: the 4XPA-based model-I, 4XP4-based model-I, 5I73 allosteric-based model-I and 5I73 orthosteric-based model-II. The 5I73 allosteric-based model-I from cluster 2 docking was selected as a model to represent the allosteric binding site although the BEDROC score was low, which may reflect that most of the known binders presumably are orthosteric binders.

The BEDROC scores of the four final models were presented as enrichment plots (Figure 10 and 11). The Enrichment plot shows the sensitivity rate in Y-axis, specificity in X-axis, a black diagonal line and blue area. The sensitivity rate indicate the active ligands as true

binding ligands (true positive), the specificity indicate the decoys as true binding compounds (false positive), the black diagonal line indicate random results and the blue area indicate compounds that were docked into the model.

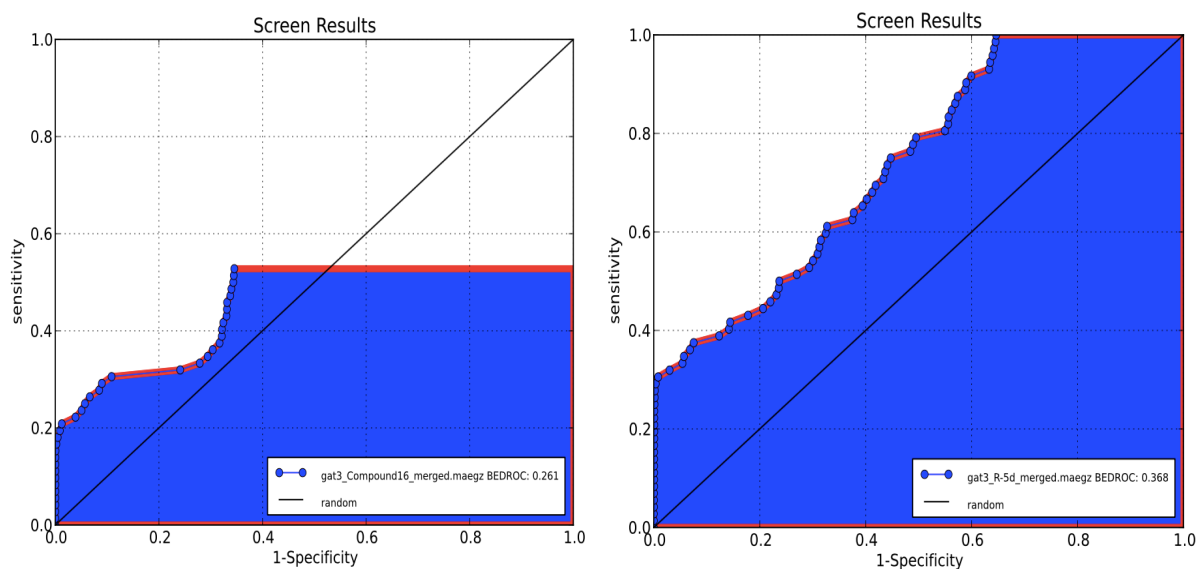


Figure 10: **Enrichment plot of 4XPA-based model-I (left) and 4XP4-based model-I (right)**. Both are dDAT-based models but 4XP4-based model have higher BEDROC score, which mean that the model have better distinguishing properties between active and decoys than 4XPA-based model.

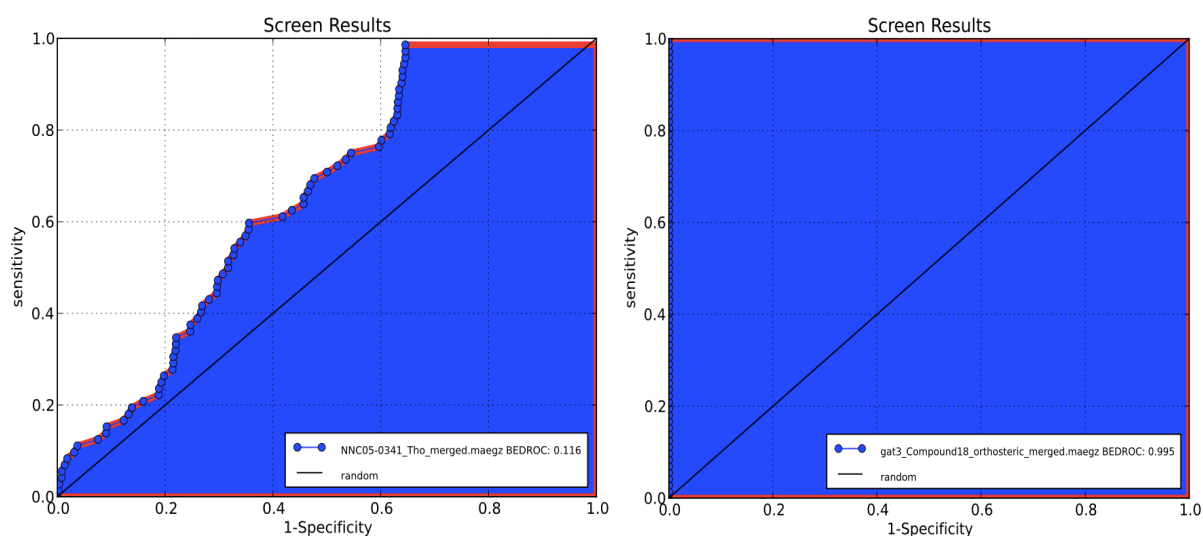


Figure 11: **Enrichment plot of 5I73 allosteric-based model-I (left) and 5I73 orthosteric-based model-II (right)**. Both are hSERT based models, but 5I73 orthosteric-based model have higher BEDROC score, which mean that the model have better distinguishing properties between active and decoys than 5I73 allosteric-based model and also than the other models.

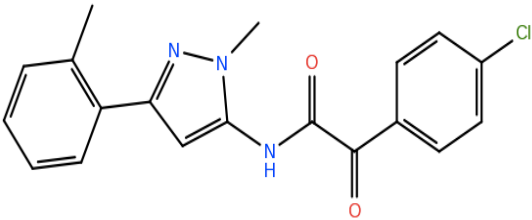
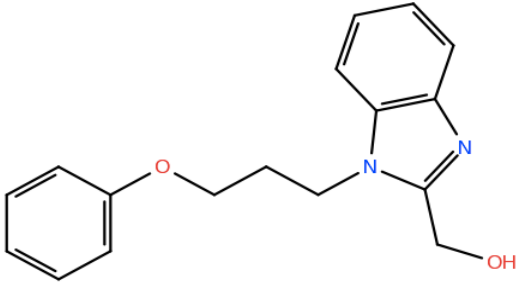
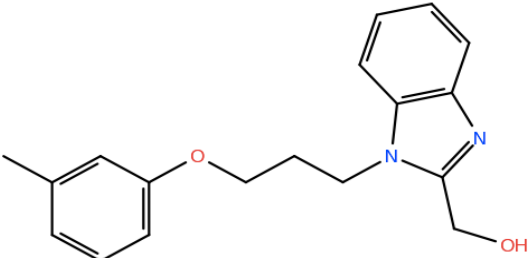
4.3.5 Docking the compounds obtained by the ligand based approach

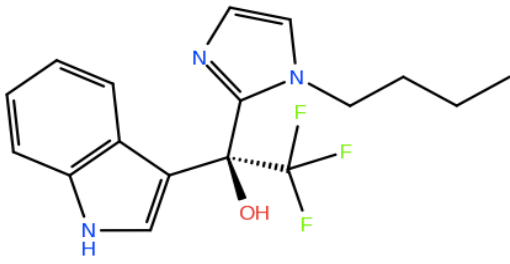
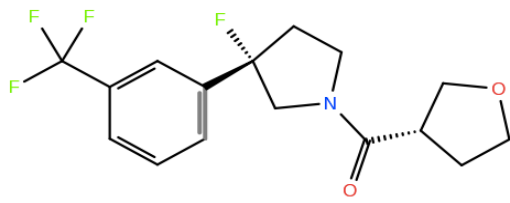
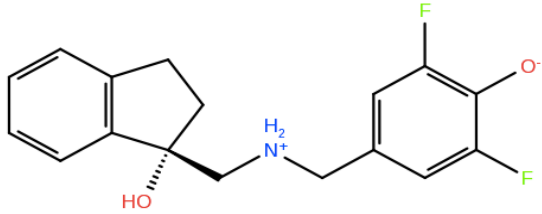
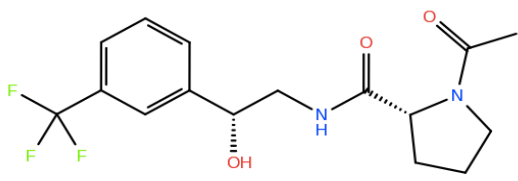
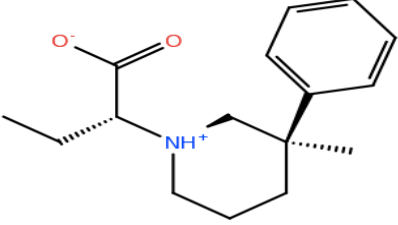
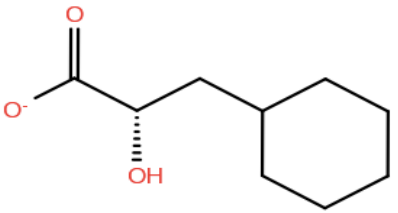
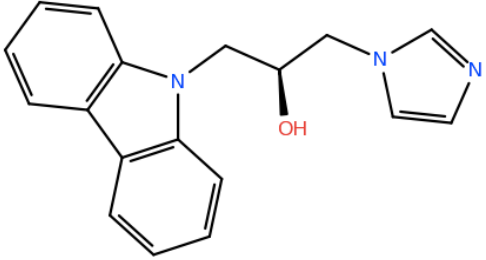
SBVS was performed with four selected models in order to identify compounds that may be GAT-3 inhibitors. These models were able to dock and created reliable ligand-receptor interactions for the active ligands. Of the selected models, three models were representing the orthosteric site (4XPA-, 4XP4- and 5I73-based model) and one model was representing the allosteric site (5I73-based model). Ligprep was used to prepare the 2151 compounds from ENAMINE database for docking. At most ten stereoisomers of each compound were generated, which gave a total of 5876 compounds. Negative value of docking score indicates good docking score, and a more negative value for the docking score represents a stronger binding of the ligand to the target. For each model, the 100 compounds with the best docking score were selected and structurally clustered into 10 structural clusters. The clustering was done in order to support the identification of compounds with favourable docking score and structural divergence for further experimental testing (Tables 12-15). The aim was that selected compounds should as much as possible represent the “conformational space” of GAT-3 ligands. For each model, one compound from each of the 10 clusters was selected and compared with the 72 active reference ligands.

4.3.5.1 4XPA-based model-I

The 100 compounds in 4XPA-based model had a docking score in the range of -7.97 to -9.78. IFD was performed for compound Z1428205595 from cluster 7, which had the highest docking score of the ENAMINE compounds in the model. The IFD improved the docking score to -11.21. For the 4XPA-based model-I, 60 out of the 100 selected compounds were from cluster 7 of the 10 clusters, while cluster 1 and cluster 4 had fewest compounds. Of the 10 clustered compounds after SBVS (Table 12), one compound (Z1315871150) was similar to the original cluster 1 of the 72 actives, while one compound (Z1836338785) was similar to initial cluster 3 compounds, while the other eight compounds were similar to cluster 2 compounds of the 72 active compounds.

Table 12: Clustering of ENAMINE compounds in 4XPA-based model-I: 100 compounds were clustered into 10 clusters and the highest scoring ligand in each cluster is shown together with docking score

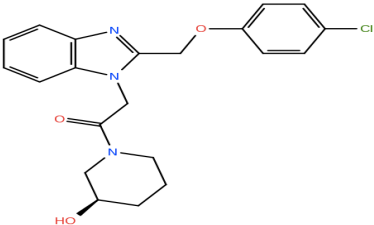
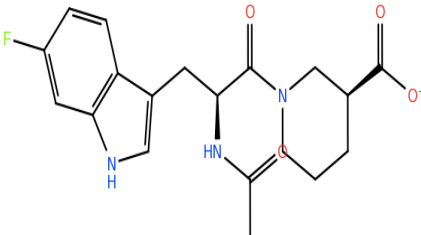
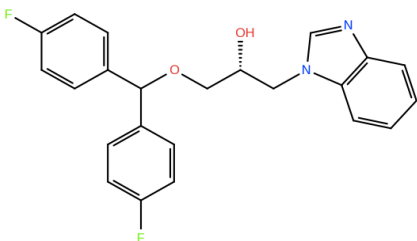
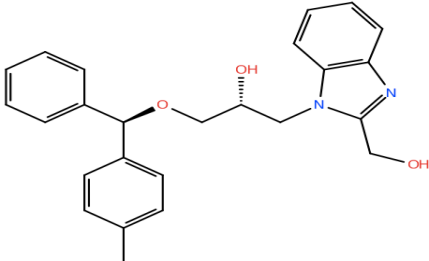
Ligand Cluster	ENAMINE compounds	Structure	Docking score
Cluster 1	Z1315871150		-8.24
Cluster 2	Z57171798		-9.62
Cluster 3	Z57300275		-8.78

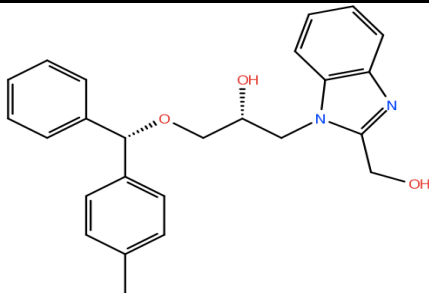
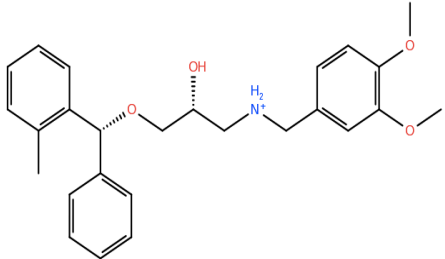
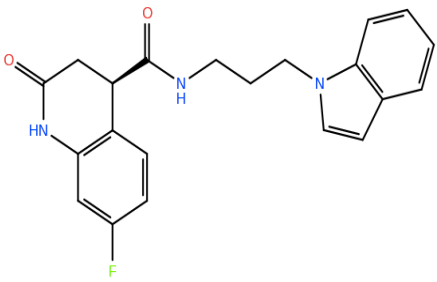
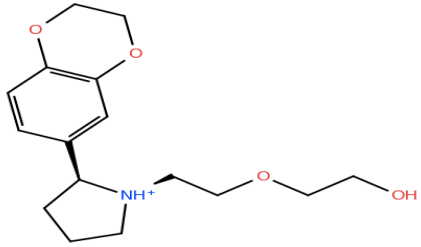
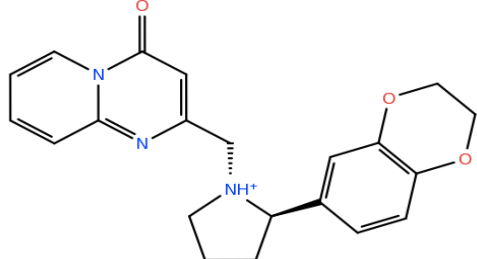
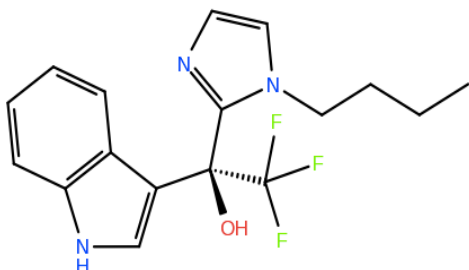
Cluster 4	Z227830852		-8.54
Cluster 5	Z2012456444		-8.80
Cluster 6	Z1499005107		-8.79
Cluster 7	Z1428205595		-9.78
Cluster 8	Z1838185187		-8.29
Cluster 9	Z1836338785		-8.04
Cluster 10	Z286789012		-8.70

4.3.5.2 4XP4-based model-II

The 100 best compounds in 4XP4-based model-II had a docking scores in the range of -10.13 to -12.00. Compound Z31336192 from cluster 5 had the best docking score and IFD was performed for this compound, but the docking score did not improve after IFD (reduced from -12.00 to -11.40). Among the 100 selected, 22 were from cluster 6 and 21 were from cluster 5. Cluster 8 and cluster 9 had the fewest compounds (1 compound) among the 100 compounds. All the 10 selected compounds were similar to initial cluster 2 of the 72 active compounds (Table 13).

Table 13: ENAMINE compounds in 4XP4-based model-II: 100 compounds were clustered into 10 clusters and the best scoring ligand in each cluster is shown together with docking score.

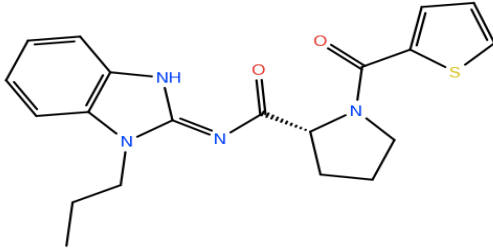
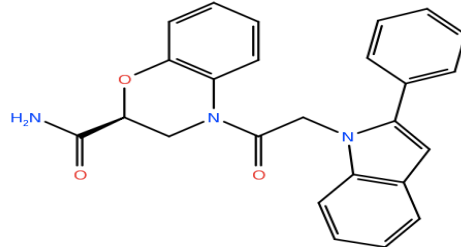
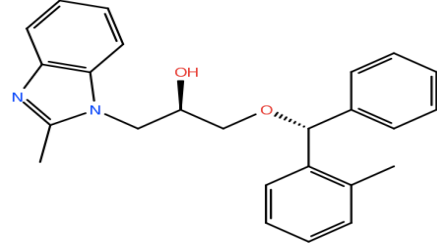
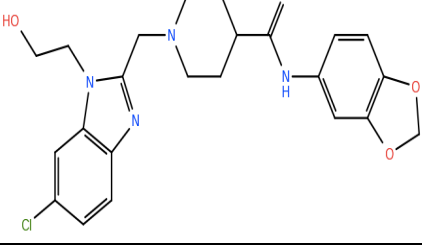
Ligand cluster	ENAMINE compound	Structure	Docking score
Cluster 1	Z245181750		-10.90
Cluster 2	Z1443622037		-11.25
Cluster 3	Z31371297		-11.21
Cluster 4	Z31336192		-11.88

Cluster 5	Z31336192		-12.00
Cluster 6	Z31381824		-11.28
Cluster 7	Z225290664		-10.55
Cluster 8	Z1203859376		-10.21
Cluster 9	Z105892372		-10.21
Cluster 10	Z227830852		-10.53

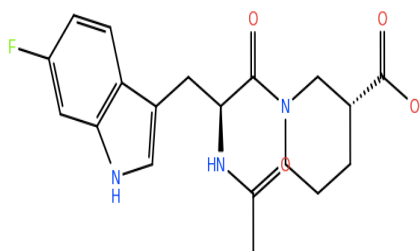
4.3.5.3 5I73 based model-I, allosteric binding site

The 100 selected compounds from the screening with the 5I73 allosteric-based model had docking scores in the range of -8.64 to -10.71. Z31371528 in cluster 3 was the compound showing best docking score and the scoring improved to -12.08 after IFD. Of the 100 compounds, 60 were from cluster 3, which was the cluster with the most compounds, while fewest compounds were from cluster 2 (1 compound), cluster 8 (1 compound) and cluster 10 (1 compound). All of the 10 selected compounds were most similar to cluster 2 from the initial clustering of active compounds (Table 14).

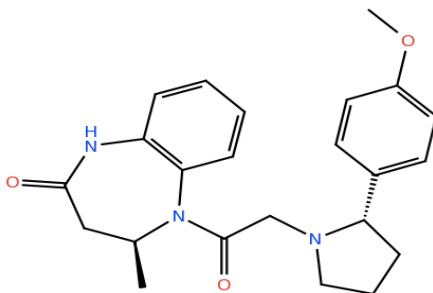
Table 14: ENAMINE compounds in 5I73 allosteric-based model-I: 100 compounds were clustered into 10 clusters and the highest docking score in each cluster is shown together with docking score.

Ligand cluster	ENAMINE compound	Structure	Docking score
Cluster 1	Z29078160		-9.37
Cluster 2	Z51885185		-9.71
Cluster 3	Z31371528		-10.71
Cluster 4	Z106741288		-10.46

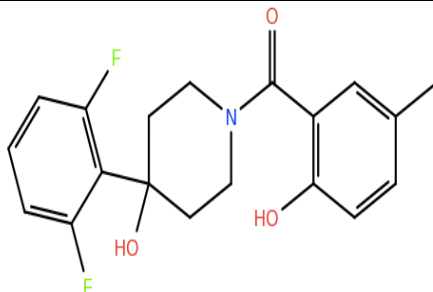
Cluster 5 Z1443622037 -8.84



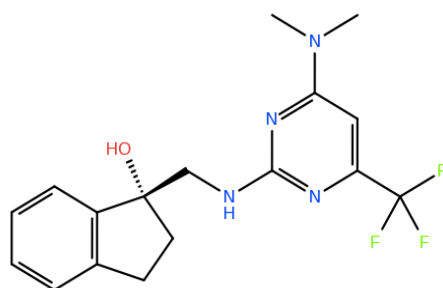
Cluster 6 Z103858668 -9.45



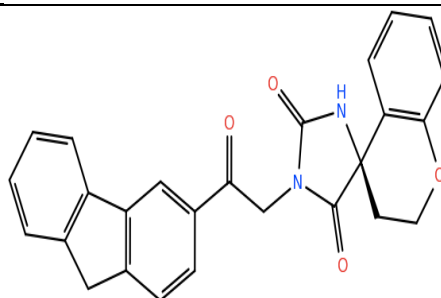
Cluster 7 Z1748616736 -9.08



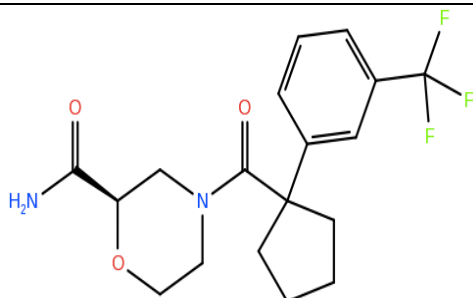
Cluster 8 Z1498356784 -8.96



Cluster 9 Z16336862 -10.35



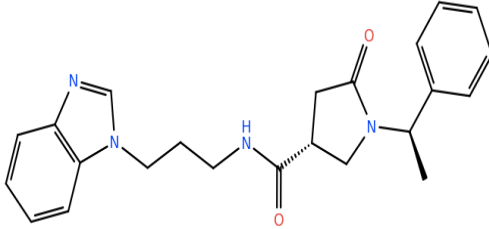
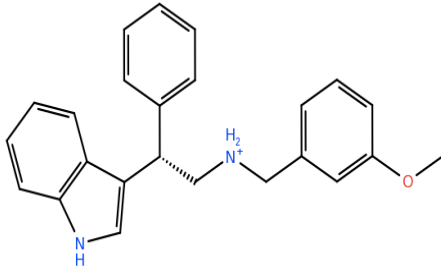
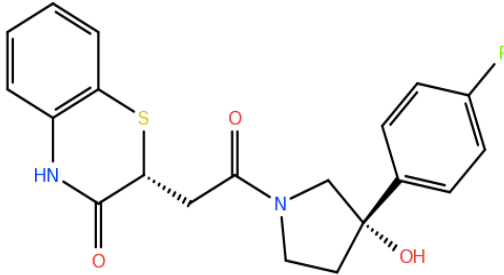
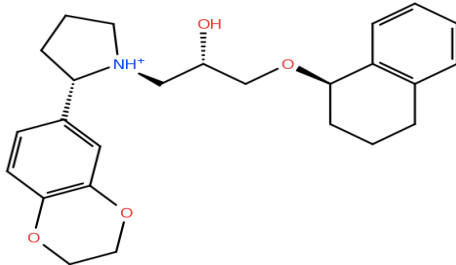
Cluster 10 Z961016376 -8.72



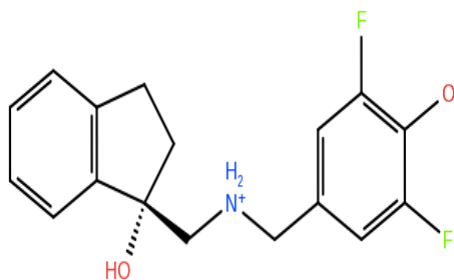
4.3.5.4 5I73 based model- II, orthosteric binding site

The 5I73 based model-II (orthosteric binding site) had a docking scores in range of -9.17 to -11.08 and Z103085506 from cluster 6 had the best docking score (Table 15). The docking score of Z103085506 was improved from -11.08 to -12.19 with IFD. Among the top 100 hits, 24 were from cluster 6 and that was the cluster with third most compounds among the 100. Cluster 3 was the cluster with most compounds, with 34 out of the 100 compounds, while cluster 7 and cluster 8 were those with fewest compounds (1 compound each). All of the 10 compounds (Table 15) were similar to cluster 2 from the clustering of actives.

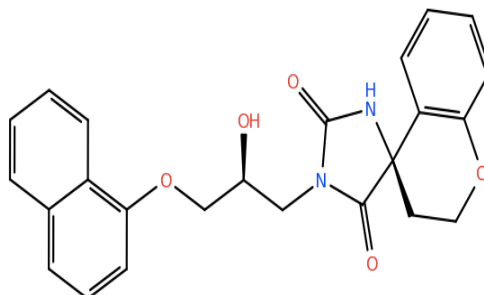
Table 15: ENAMINE compounds in 5I73 orthosteric-based model: 100 compounds were clustered into 10 clusters and the highest docking score in each cluster is shown together with docking score.

Ligand cluster	ENAMINE compound	Structure	Docking score
Cluster 1	Z102451230		-9.78
Cluster 2	Z126620738		-10.28
Cluster 3	Z1608250902		-10.06
Cluster 4	Z168960054		-9.66

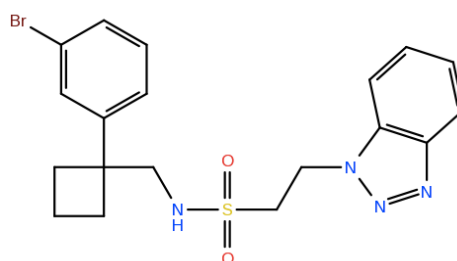
Cluster 5 Z1499005107 -9.69



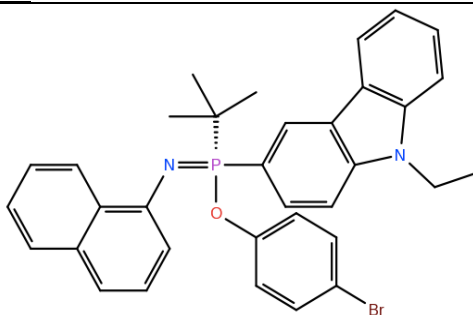
Cluster 6 Z103085506 -11.08



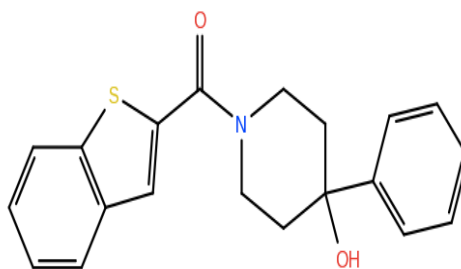
Cluster 7 Z785615210 -9.17



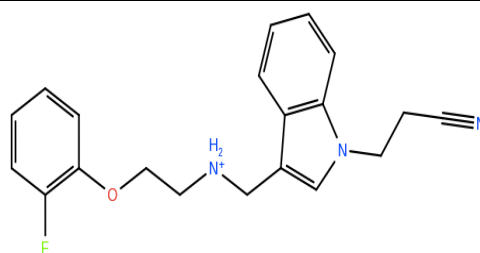
Cluster 8 Z56783608 -9.18



Cluster 9 Z1704011730 -10.00



Cluster 10 Z237524118 -9.51



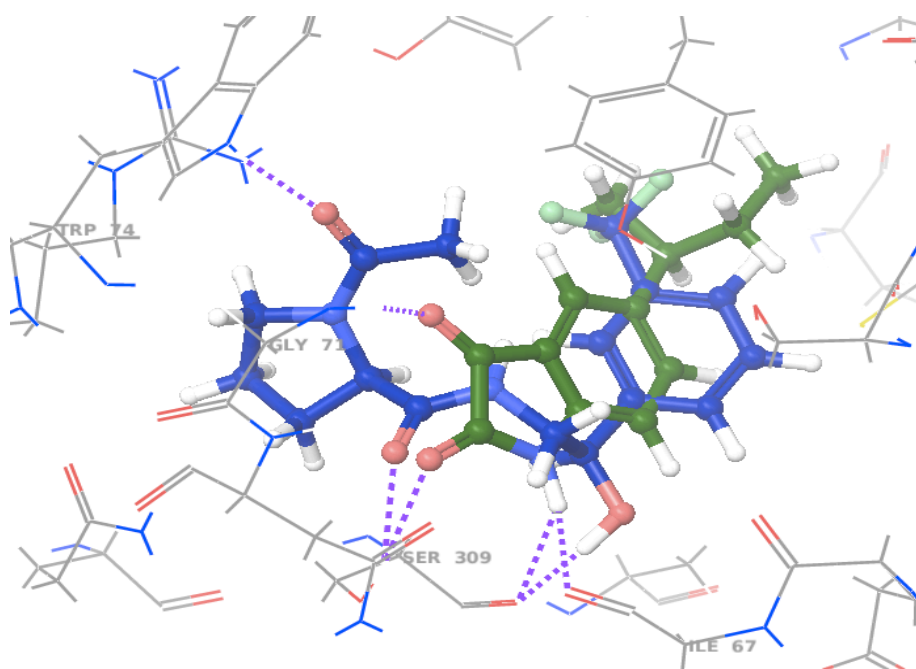


Figure 12: **The binding mode of ENAMINE Z1428205595 (blue carbons) and the active ligand Compound 16 (Dark green carbons) in the binding site of 4XPA-based model-I.** Hydrogen bonds interactions are represented in purple dashed lines and aromatic interactions in green dashed lines. Residues interactions within 5Å sphere radius around the ligands are displayed.

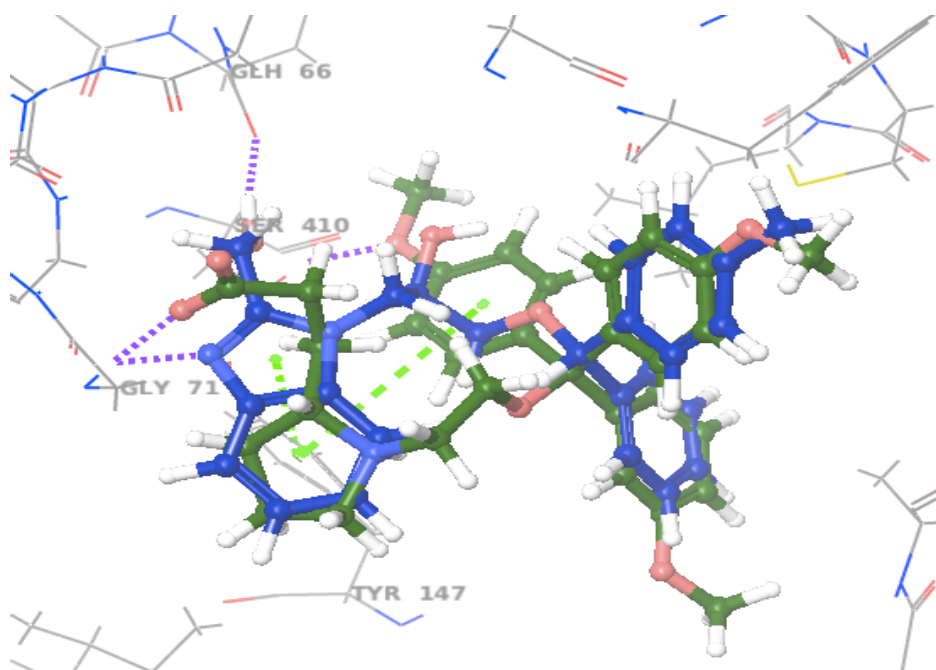


Figure 13: **The binding mode of ENAMINE Z31336192 (blue carbons) and the active ligand (R)-5d (dark green carbons) in the binding site of 4XP4-based model-II.** Hydrogen bonds interactions are represented in purple dashed lines and aromatic interactions in green dashed lines. Residues interactions within 5Å sphere radius around the ligands are displayed.

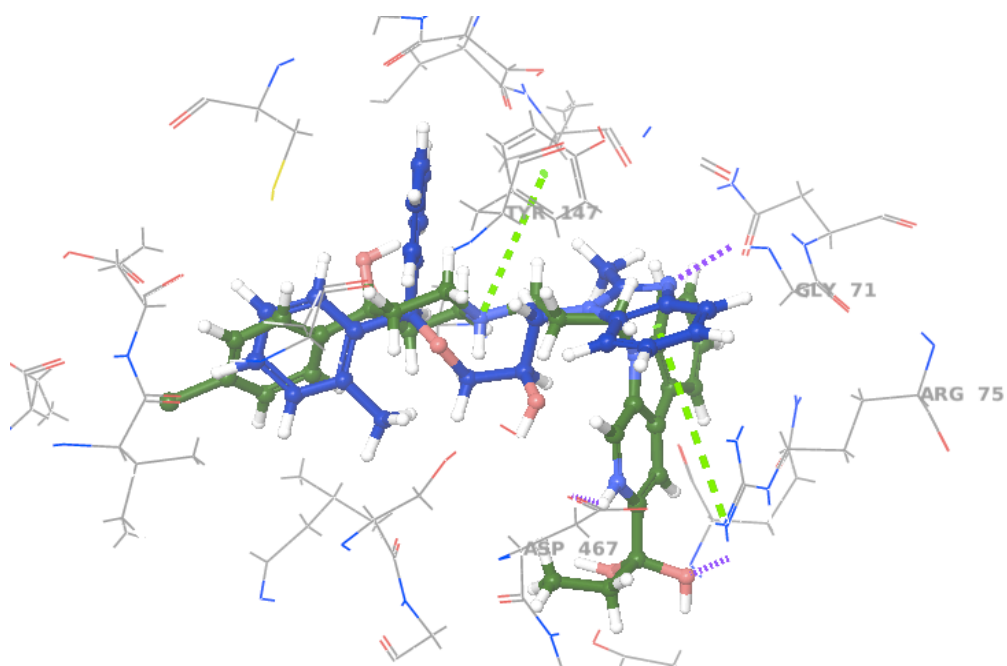


Figure 14: **The binding mode of ENAMINE Z31371528 (blue carbons) and the active ligand NNC05-0341 (dark green) in the binding site of 5I73 allosteric-based model-I.** Hydrogen bonds interactions are represented in purple dashed lines and aromatic interaction in green dashed lines. Residues interactions within 5Å sphere radius around the ligands are displayed.

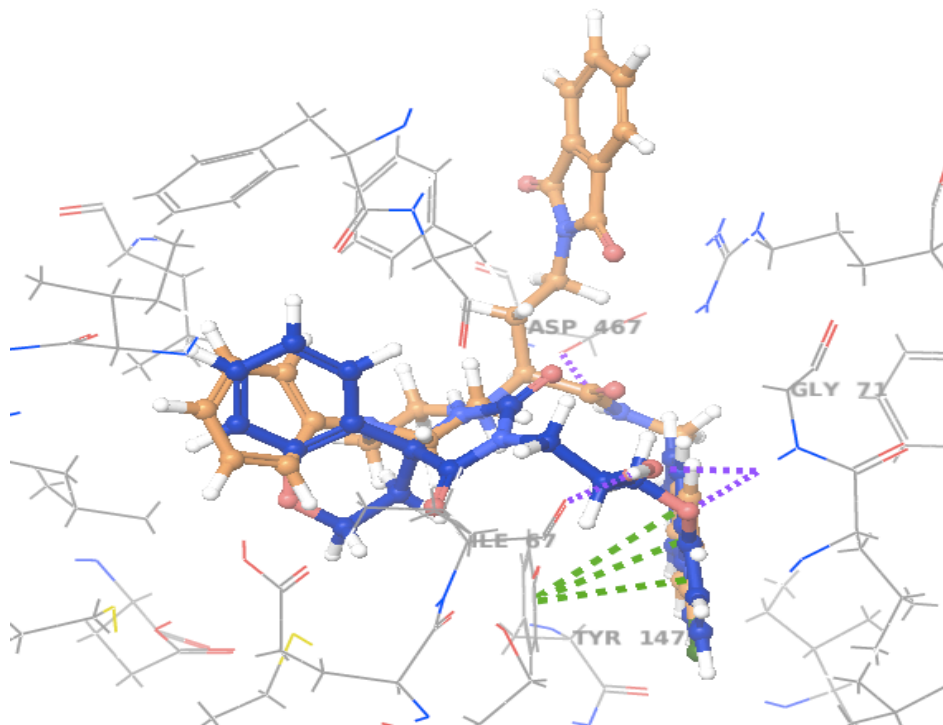


Figure 15: **The binding mode of ENAMINE Z103085506 (blue carbons) and the active ligand Compound 18 (green carbons) in the binding site of 5I73 orthosteric-based model-II.** Hydrogen bonds interactions are represented in purple dashed lines and aromatic interactions in green dashed lines. Residues interactions within 5Å sphere radius around the ligands are displayed.

5 DISCUSSION

In the present project, we were aiming to predict GAT-3 inhibitors. GAT-3 inhibitors may target a newly discovered abnormality in AD, the increased tonic inhibition in DG of hippocampus as a result of elevated GABA levels. The high GABA content is released from astrocytes and increase the GABA content in DG cells, and enhanced tonic inhibition of DG (8). GAT-3 inhibitors will decrease GABA in DG cells, and have a putative therapeutic value in AD. GAT-3 is one of the GABA transporters that has an essential role in the reuptake of the neurotransmitter GABA from the synaptic cleft into glial cells (21).

In general, experimental investigation of a new drug in the first stage can both be time consuming, and costly and requires huge resources. However, the use of computational drug design such as virtual screening that can screen large compound databases in a time and cost effective manner may reduce the time range and economical costs of the preclinical phase of drug development. In this study, a combination of LBVS and SBVS was performed to predict GAT-3 inhibitors. In the LBVS, 2D fingerprints was used for clustering of 72 active ligands (reference compounds) and for screening of 294 995 drug like compounds from the ENAMINE database. The hits from the LBVS were then used for structure-based virtual screening by docking the hits into selected GAT-3 homology models that could enrich the known 73 active ligands (Figure 6). Four homology models were selected based on BEDROC score and docking score of the active ligands.

5.1 Ligand-based virtual screening

5.1.1 Clustering of active ligands

In this study, 2D fingerprints of 72 known active ligands were used to predict new putative GAT-3 inhibitors. The selected 72 active GAT-3 binders and were clustered into five different groups based on their structural similarities. Generally, each molecule has unique set characteristics of physicochemical and biological properties. It is more uncomplicated to find information about structural molecules of compounds than their biological activities.

Clustering of compounds that share mutual properties can help us to identify and understand groups of compounds with structural similarities that are assumed to share common biological activity without testing all of them. The principle with clustering of compounds is also to minimize searching of compounds with undisclosed biologically activity, but we also should take into account that slight structural change of a compound can lead to huge changes in their biological activities (35).

There are many fingerprint types that can be used to cluster compounds. In this investigation, the radial fingerprint type as based of the Kelly criterion was used, which gave an ideal number of clusters of the input compounds. Before choosing the radial fingerprint, all the fingerprint types were tested to find the type which was best suitable for the 72 active ligands in this study. When other fingerprint types were chosen, the 72 active ligands were divided into several clusters and some of the clusters with quite a few compounds. The radial fingerprint type gave less number of clusters and several compounds in each cluster, and the merging distance was greater than of the other fingerprint types. In addition, the similarity matrix showed that the active ligands were not wrongly clustered. Cluster 5 was a cluster with only three compounds (Table 4). These three active ligands were outliers with different structural and physicochemical properties and were not similar to the other clusters.

5.1.2 Use of known active ligands in searching for GAT-3 inhibitors

Choosing the wrong fingerprint type can affect the result of a study. Previous studies have investigated 2D virtual screening in Canvas by using different fingerprint types and similarity metrics to find the most appropriate fingerprint. These studies concluded that MOLPRINT2D is the best to use when you are uncertain about the correct fingerprint type for your set of

ligands (36, 37). These studies also suggested that none of the fingerprint types is ideal for all the active compounds and they recommended using several fingerprint types for the set of ligands (36, 37). In this study Linear, Dendritic, Radial and MOLPRINT2D fingerprint types were used to screening the active ligands against decoys and the to determine cut-off values for the ENAMINE compounds. From Table 6 we can see that dendritic fingerprint gave a higher number of hits than the others. Most of the active compounds were clustered in cluster 2 and that may be the reason for that the hits from the screening of ENAMINE was most similar to cluster 2. Cluster 5 contained only three compounds with different structures and this may be the reason that compounds similar to cluster 5 were not identified in the screening. The ADMET-filtering was not performed for 2151 screened compounds (Table 6), because all of them were from ENAMINE and therefore within the rule of five for CNS drugs (Table 1).

The use of known ligand structures in the LBVS approach will increase the hit rate, but it is of course a possibility that the hits will be too similar to the compounds already known for target.

5.2 Homology modeling

The 3D structure of GAT-3 is unknown and homology modelling was used to build 3D models of this target. Most often X-ray structures are used in SBVS approaches, but the homology models are often sufficient for SBVS (51). Due to the protein folding, it is not possible to determine the entire 3D structure of a protein based on the primary or secondary structure, and thus virtual screening has to rely on homology models using experimentally determined structures as templates when the structure is unknown. X-ray crystallography is the most widely used technique, and in this study X-ray structures of dDAT and hSERT X-ray were used as templates. GAT-3 is a GABA transporter that belongs to the SLC6 family of transporters (21). dDAT and hSERT also belong to the SLC6 family and therefore they were used as templates to build the homology models. The amino acid sequence of GABA transporters; GAT-1, GAT-2 and BGT-1 were also included in multiple alignments to increase accuracy of the homology models. Both dDAT and hSERT based models have a typical SLC6 structures.

The sequence identity between the target and template were calculated. The sequence identity of GAT-3 with dDAT was 40% and 38% with hSERT. Studies have concluded that the quality of the obtained model depend on high sequence identity between the target and template. An acceptable sequence identity for highly reliable models is believed to be more than 30%. Amino acid sequence identity below 20% can cause some problems such as gaps in the alignment, but due to high conservation of important regions, still quite reliable models may be constructed (33). The sequence identities of GAT-3 with dDAT and hSERT indicate high quality models. The homology models were also evaluated by PROCHECK, ERRAT and Verify (Table 7). The evaluation results also indicate that we have high quality homology models, and that the models may be used to study GAT-3 ligand binding interactions.

5.3 Structure-based virtual screening

5.3.1 Evaluation of the selected final models, BEDROC score

The selected models were evaluated based on docking score and BEDROC score. The BEDROC score was calculated for each selected model to see if the model is able to distinguish between active GAT-3 ligands and decoys. The active ligands that were used as a reference were substrates or inhibitors and they all were clustered into five groups based on their structural similarity (from LBVS). In addition, we do not know for all of them if they are allosteric or orthosteric (or both) and that can affect our results. But, we assume most of them are orthosteric binders, and therefore we also calculated the BEDROC and included that as an evaluation of our models. The generated decoys of the active ligands have the same physiochemical properties, but structurally different from the active ligands and they are assumed to be non-binders.

The active ligands in this study had quite different molecular structures, some of them were quite large molecules (cluster 2 and 5) and some of them were quite small molecules (cluster 1, 3 and 4). The final four selected models had BEDROC score from 0.116 to 0.995 (Figures 10 and 11). Not all the active ligands could be docket into all selected models and this may be

because of the size of the defined binding pocket. However, all active ligands could be enriched by one or more of the selected models. Some of the models were quite specific for a particular cluster. Tables 10 and 11 show that the models that were able to dock all the active ligands had higher BEDROC score than the others. The 4XPA-based model-I was able to dock only 38 of active ligands and the binding site of the model was defined by Compound 16 from cluster 1. The model was able only to dock active ligands from cluster 1, cluster 3 and cluster 4 and which are quite small molecules. If the binding site of this model was defined by a large active ligand then the model most probably also could dock some or all of the other active ligands, but also the non-binders (decoys). The 4XP4-based model-I and 5I73 orthosteric-based model-II was able to dock all the 72 active ligands and the binding site was defined by active ligand ((R)-5d and Compound18) from cluster 2 and cluster 5, which is large molecular structures, but 5I73 orthosteric had higher BEDROC score. The 5I73 allosteric-based model-I model from cluster 2 was able to dock almost all the active ligands except Compound 18, which was too large to fit into the binding site. The binding site of the model was defined by the active ligand (NNC05-0341) that is a large molecule from cluster 2, but the model had the lower BEDROC score than the four selected models. This may be because of the large defined binding pocket of the model, but this was not surprising, since most of the active ligands most probably are orthosteric binders. The larger the binding site, the more active ligands the model will adopt, but it will also be easier for non-binders (decoys) to be docked into the model. If the models had a closed binding site, larger active ligands would not be fit into the binding site. Studies have shown that the binding pockets of this type of transporters are quite open. Therefore, in our theoretical studies it may be difficult to distinguish between binders and non-binders and that can lead to many false positives result (51). The disadvantage of BEDROC metric is that it does not have power to check if the defined binding pocket is optimal or not, and the metric is also not capable for a binding site with only few docked active ligands, the binding site of the model should be able to have at least 50 known active ligands. A high BEDROC score is automatically considered as a good model, but it may be a good model even if the BEDROC is quite bad.

5.3.2 Docking the compounds obtained by the ligand based approach

The better the fit of the ligand is to the binding pocket, the better is the affinity of the ligand to the target and the ligand may be a good drug candidate. We assume that similar molecules have similar interaction partners in the target. From the LBVS we found that 2151 screened compounds have quite similar structure to the active ligands. After Ligprep the number of compounds increased to 5876 that were docket into the selected homology models. 10 compounds from each model were studied more in detail (Tables 12-15). The problem with SBVS is that you may get many false positives that are not binders, because the model is not in a condition to distinguish between binders and non-binders. Thus, we can use the knowledge about the structure of known binders to obtain a structurally more narrow library prior to the docking by using a ligand based approach. Most probably that will reduce the number of false positives. The binding poses of the models were defined by the structure of active ligands, and thus were adopted to the structure of the active ligands. When it comes to docking score, none of the ENAMINE compounds had higher docking score than known compound. However the models were not trained for the ENAMINE compounds, and IFD of the ENAMINE compounds would most probable improve their scoring values.

It is essential to identify the residue interactions in the binding sites to know how well a ligand binds to the intended molecular target. We assumed that good binders must bind in the same manner as the known binders, but we also have to take into account that the binding sites of the constructed homology models are optimized for the known active ligands and not for the hit compounds, and we also need to have the difference in compound size in mind.

The 10 selected hit compounds in 4XPA-based model-I were large molecules (Table 12). In 4XPA-based model-I, only small active ligands from cluster 1, 3 and 4 were able to dock. The 4XPA-based model-I was the only model among the four final models that selected ENAMINE compounds similar to the initial clusters 1, 2 and 3. The active ligands from cluster 1 were small molecules and the selected ENAMINE compound Z1315871150 (cluster 1) is much larger than the active ligands of cluster 1 (Appendix 1). The active ligands of cluster 3 are mainly substrates and quite small molecules (Appendix 1), and that is reflected in that quite small ENAMINE compounds also were selected. A representative is Z1836338785 (ENAMINE) with a structure corresponding to the initial cluster 3. Eight of the selected ENAMINE compounds for this model corresponded to the initial cluster 2. However, in

Figure 13 the docking pose of the selected ENAMINE compound Z1428205595 is compared with the known active Compound 16 (cluster1), since Compound 16 was the active with best scoring in the 4XPA-based model-I. The Figure indicates that their binding mode is quite similar. Figure 12 show that the ENAMINE compound is much larger than Compound 16, but both compounds formed H-bond with residue Ser309, although they have different size and belong to different cluster groups. In addition, Compound 16 formed H-bond with residue Gly71 and Ile67, while Z1428205595 formed H-bond with Trp74.

The 10 selected ENAMINE compounds in 4XP4-based model-II (Table 13) were similar to the initial cluster 2. The docking pose of the selected Z31336192 ENAMINE compound (cluster 5) with the best docking score in 4XPA-based model-II was compared with the active ligand (R)-5d, from cluster 2. Figure 13 shows that Z31336192 is smaller than (R)-5d, but both compounds have similar H-bonds and aromatic interactions with the binding site. Compound Z31336192 and (R)-5d both have H-bonding interaction with residue Gly71 and aromatic interaction with residue Tyr147. In addition, Z31336192 has H-bonding interaction with Glh66, while (R)-5d has H-bond interaction with Ser410.

All the 10 selected ENAMINE compounds in 5I73 allosteric-based model-I (Table 14) were corresponded to the initial cluster 2. The binding mode of the selected Z31371528 ENAMINE compound (cluster 3) and the active ligand NNC05-0341 (cluster 2) with the highest docking scores was compared (Figure 14). Z31371528 is smaller than NNC05-0341, but both formed aromatic interaction with Arg75. In addition, Z31371528 formed H-bonding with Gly71, while NNC05-0341 formed a H-bond with Arg75 and Asp467.

The 10 selected ENAMINE compounds in 5I73 orthosteric-based model-II (Table 15) were similar to the initial cluster 2. The docking pose of the selected Z103085506 ENAMINE compound (cluster 6) and the active ligand Compound 18 from cluster 5 of the clustering of actives were compared (Figure 15). Compound 18 were the compound with the best docking score in the model. Both compounds formed aromatic interaction with Tyr147, in spite of that Z103085506 was not similar to the initial cluster 5. Z103085506 also formed H-bonding interactions with Gly71 and Ile67, while Compound 18 formed a H-bond with Asp467.

Z103085506 had more favourable interactions with the model than the known active compound.

The comparisons of the compounds shows that all selected ENAMINE compounds bind in the same area as the known binders, but because of different compound size and not optimized binding site of the ENAMINE compounds there are differences in the binding interactions. Some compounds had quite similar interactions in spite of that they did not correspond to the same initial structural cluster. According to the comparison of known active ligands and the selected ENAMINE compounds, the residues that may be most important at the orthosteric binding site are Ser309, Gly71 and Tyr147, while Arg75 seems most important at the allosteric site (Figures 12-15). We also need to take into account that this comparison was only based on four known active ligands and the four ENAMINE compounds with high docking score. But other active ligands (Table 10) and the 40 hit compounds (Tables 12-15) also have other similar beneficial H-bonding interactions with Phe308, Arg75 and Asp467 in the orthosteric site and Gly71 and Phe308 in the allosteric site (results not shown). However, there are no available mutagenesis data showing the importance of GAT-3 amino acids for ligand binding. The present study may be used to design site directed mutagenesis studies identifying important amino acids for ligand binding.

6 CONCLUSION

Homology modeling was used to build seven 3D models of GAT-3 based on dDAT and hSERT. The evaluations indicate that the models were of high quality and four of them were selected for SBVS based on the BEDROC score. LBVS and SBVS methods were used to identify 40 hit compounds with high docking score. A general problem with computational screening is that the results depend on the geometry/nature of the binding site and the ligands. An open binding site may adopt several false positive hits and not distinguish between binders and not binders. Based on the screening we have identified putative hits from the ENAMINE dataset (Table 12-15) that need to be tested experimentally in order to check if they are GAT-3 inhibitors.

7 FUTURE DIRECTIONS

In this study, only docking and scoring and molecular mechanics were used to assist in the selection of compounds, which often is the first step in drug discovery and development process. We have identified 40 hit compounds and the next step is to identify lead compounds. The 40 hit compounds should be evaluate by experimental methods, to confirm if the hit compounds have the binding affinity to the target as demonstrated in this study. The 40 hit compounds have drug-like properties and to develop a drug from lead compounds the affinity of the leads must be improved without hampering the ADMET properties of the leads from the ENAMINE database. The studied hit compounds were treated as semi-rigid in the binding site by the selected models that are were trained for the known active ligands. By performing IFD with the hit compounds, the interactions between the hits and the target could be improved, and that may support future selection for experimental verification.

8 REFERENCE

1. Yiannopoulou KG, Papageorgiou SG. Current and future treatments for Alzheimer's disease. *Therapeutic Advances in Neurological Disorders*. 2013;6(1):19-33.
2. Rees G. World Alzheimer Report 2016 [Internet] London: Alzheimer's Disease International (ADI); 2016 [2016 Sep 05]. Available from: <https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf>.
3. Zhagn L, Li Z. [Alzheimer and the discovery of Alzheimer's disease]. *Zhonghua yi shi za zhi* (Beijing, China : 1980). 2014;44(5):288-90.
4. Alzheimer's Disease Fact Sheet USA: The National Institute on Aging at the National Institutes of Health (NIH); 2016 [updated 2016 Aug 16. Available from: <https://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-fact-sheet - changes>.
5. H.P.Rang JMR, R.J.Flower,G.Henderson. Rang&Dale's Pharmacology. 8th ed: Elseveier Churchill Livingstone; 2016. p. 487-9.
6. Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E. Alzheimer's disease. *Lancet* (London, England). 2011;377(9770):1019-31.
7. Salomone S, Caraci F, Leggio GM, Fedotova J, Drago F. New pharmacological strategies for treatment of Alzheimer's disease: focus on disease modifying drugs. *British journal of clinical pharmacology*. 2012;73(4):504-17.
8. Wu Z, Guo Z, Gearing M, Chen G. Tonic inhibition in dentate gyrus impairs long-term potentiation and memory in an Alzheimer's [corrected] disease model. *Nature communications*. 2014;5:4159.
9. Pardridge WM. The Blood-Brain Barrier: Bottleneck in Brain Drug Development. *NeuroRx*. 2005;2(1):3-14.
10. Pajouhesh H, Lenz GR. Medicinal Chemical Properties of Successful Central Nervous System Drugs. *NeuroRx*. 2005;2(4):541-53.
11. Richard W. Olsen G-DL. Chapter 18-GABA: Academic Press; 2012 [Scott T. Brady, George J. Siegel, R. Wayne Albers and Donald L. Price:[Available from: <http://www.sciencedirect.com/science/article/pii/B9780123749475000183>.
12. Broer S, Gether U. The solute carrier 6 family of transporters. *British journal of pharmacology*. 2012;167(2):256-78.
13. Hack S, Worlein B, Hofner G, Pabel J, Wanner KT. Development of imidazole alkanolic acids as mGAT3 selective GABA uptake inhibitors. *European journal of medicinal chemistry*. 2011;46(5):1483-98.

14. Damgaard M, Al-Khawaja A, Vogensen SB, Jurik A, Sijm M, Lie ME, et al. Identification of the First Highly Subtype-Selective Inhibitor of Human GABA Transporter GAT3. *ACS chemical neuroscience*. 2015;6(9):1591-9.
15. Madsen KK, Clausen RP, Larsson OM, Krogsgaard-Larsen P, Schousboe A, White HS. Synaptic and extrasynaptic GABA transporters as targets for anti-epileptic drugs. *Journal of neurochemistry*. 2009;109 Suppl 1:139-44.
16. Kowalczyk P, Kulig K. GABA system as a target for new drugs. *Current medicinal chemistry*. 2014;21(28):3294-309.
17. Broer S. The SLC6 orphans are forming a family of amino acid transporters. *Neurochemistry international*. 2006;48(6-7):559-67.
18. Nelson N. The family of Na⁺/Cl⁻ neurotransmitter transporters. *Journal of neurochemistry*. 1998;71(5):1785-803.
19. Beuming T, Shi L, Javitch JA, Weinstein H. A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na⁺ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. *Molecular pharmacology*. 2006;70(5):1630-42.
20. Chen NH, Reith ME, Quick MW. Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6. *Pflugers Archiv : European journal of physiology*. 2004;447(5):519-31.
21. Kristensen AS, Andersen J, Jorgensen TN, Sorensen L, Eriksen J, Loland CJ, et al. SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacological reviews*. 2011;63(3):585-640.
22. 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.
23. Rudnick G. Bioenergetics of neurotransmitter transport. *Journal of bioenergetics and biomembranes*. 1998;30(2):173-85.
24. Nakada K, Yoshikawa M, Ide S, Suemasa A, Kawamura S, Kobayashi T, et al. Cyclopropane-based conformational restriction of GABA by a stereochemical diversity-oriented strategy: identification of an efficient lead for potent inhibitors of GABA transporters. *Bioorganic & medicinal chemistry*. 2013;21(17):4938-50.
25. Jin X-T, Galvan A, Wichmann T, Smith Y. Localization and Function of GABA Transporters GAT-1 and GAT-3 in the Basal Ganglia. 2011;5(63):10.
26. Kim DU, Kim MK, Cho YW, Kim YS, Kim WJ, Lee MG, et al. Association of a synonymous GAT3 polymorphism with antiepileptic drug pharmacoresistance. *Journal of human genetics*. 2011;56(9):640-6.

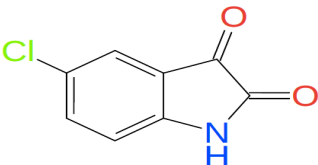
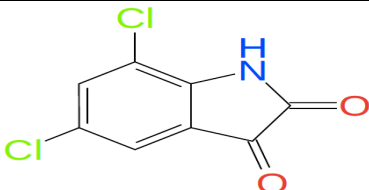
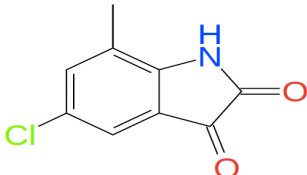
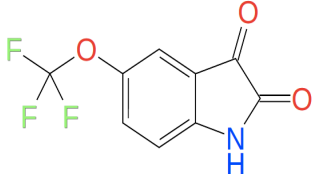
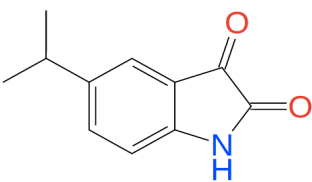
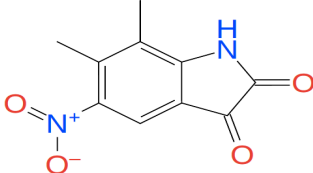
27. Patrick GL. An introduction to medicinal chemistry. 5th ed. United Kingdom: Oxford university press; 2013.
28. Gabrielsen M. Structure, function and inhibition of the serotonin transporter studied by molecular docking, -dynamics and virtual screening Tromsø: University of Tromsø. Faculty of health sciences department of medical biology 2011 [Available from: <http://munin.uit.no/bitstream/handle/10037/6403/thesis.pdf?sequence=6&isAllowed=y>].
29. Gabrielsen M. Structure, function and inhibition of the serotonin transporter studied by molecular docking, -dynamics and virtual screening Tromsø University of Tromsø 2011.
30. Lindert S, Maslennikov I, Chiu E, Pierce LC, McCammon JA, Choe S. Drug screening strategy for human membrane proteins: from NMR protein backbone structure to in silico- and NMR-screened hits. *Biochem Biophys Res Commun.* 2014;445(4):724-33.
31. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Current opinion in structural biology.* 2008;18(5):581-6.
32. Breda A, Valadares NF, Souza ONd, Garratt RC. Chapter A06 Protein Structure, Modelling and Applications: National Center for Biotechnology Information (US); 2008.
33. Ravna AW, Sylte I. Homology modeling of transporter proteins (carriers and ion channels). *Methods in molecular biology (Clifton, NJ).* 2012;857:281-99.
34. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug discovery today.* 2006;11(23-24):1046-53.
35. László A Zahoránszky GYK, Péter Hári, András Málnási-Csizmadia, Katharina A Zweig, and Gergely Zahoránszky-Köhalmi. Breaking the hierarchy - a new cluster selection mechanism for hierarchical clustering methods. *Algorithms Mol Biol.* 2009.
36. Duan J, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *Journal of molecular graphics & modelling.* 2010;29(2):157-70.
37. Sastry M, Lowrie JF, Dixon SL, Sherman W. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *Journal of chemical information and modeling.* 2010;50(5):771-84.
38. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules (Basel, Switzerland).* 2015;20(7):13384-421.
39. 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.
40. Canvas, version 2.4, Schrödinger, LLC,. New York, NY 2012.

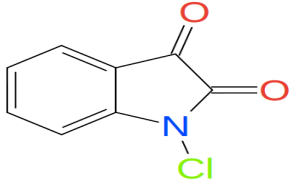
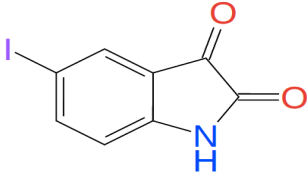
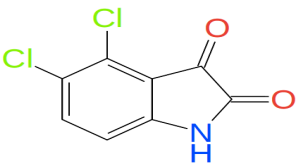
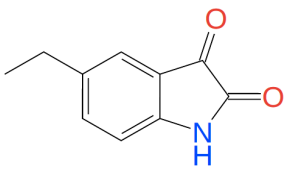
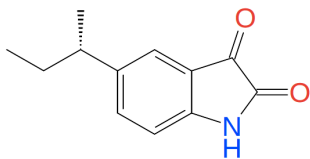
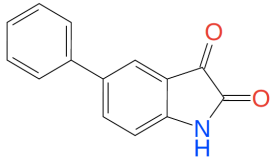
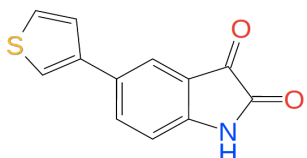
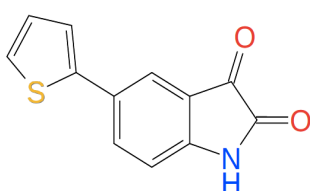
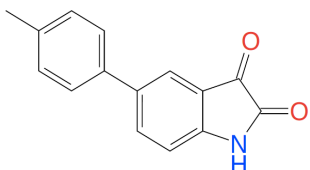
41. RCBS PDB. About the PDB Archive and the RCSB PDB RCBS PDB [cited 2017 Nov 28]. Available from:
http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/index.html
42. LigPrep, version 2.3, Schrödinger, LLC. New York, NY, 2009.
43. Glide, version 5.5, Schrödinger, LLC,. New York, NY 2009.
44. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins*. 1992;12(4):345-64.
45. Fulep GH, Hoesl CE, Hofner G, Wanner KT. New highly potent GABA uptake inhibitors selective for GAT-1 and GAT-3 derived from (R)- and (S)-proline and homologous pyrrolidine-2-alkanoic acids. *European journal of medicinal chemistry*. 2006;41(7):809-24.
46. Sitka I, Allmendinger L, Fulep G, Hofner G, Wanner KT. Synthesis of N-substituted acyclic beta-amino acids and their investigation as GABA uptake inhibitors. *European journal of medicinal chemistry*. 2013;65:487-99.
47. Thomsen C, Sorensen PO, Egebjerg J. 1-(3-(9H-carbazol-9-yl)-1-propyl)-4-(2-methoxyphenyl)-4-piperidinol, a novel subtype selective inhibitor of the mouse type II GABA-transporter. *British journal of pharmacology*. 1997;120(6):983-5.
48. Kragler A, Hofner G, Wanner KT. Novel parent structures for inhibitors of the murine GABA transporters mGAT3 and mGAT4. *European journal of pharmacology*. 2005;519(1-2):43-7.
49. Gerile, Sogawa C, Ohyama K, Masuko T, Kusama T, Morita K, et al. Inhibitory action of antidepressants on mouse Betaine/GABA transporter (BGT1) heterologously expressed in cell cultures. *International journal of molecular sciences*. 2012;13(3):2578-89.
50. Salat K, Wieckowska A, Wieckowski K, Hofner GC, Kaminski J, Wanner KT, et al. Synthesis and pharmacological properties of new GABA uptake inhibitors. *Pharmacological reports* : PR. 2012;64(4):817-33.
51. Gabrielsen M, Kurczab R, Siwek A, Wolak M, Ravna AW, Kristiansen K, et al. Identification of novel serotonin transporter compounds by virtual screening. *Journal of chemical information and modeling*. 2014;54(3):933-43.
52. Dhar TG, Borden LA, Tyagarajan S, Smith KE, Branchek TA, Weinshank RL, et al. Design, synthesis and evaluation of substituted triarylnipeptotic acid derivatives as GABA uptake inhibitors: identification of a ligand with moderate affinity and selectivity for the cloned human GABA transporter GAT-3. *Journal of medicinal chemistry*. 1994;37(15):2334-42.
53. Faust MR, Hofner G, Pabel J, Wanner KT. Azetidine derivatives as novel gamma-aminobutyric acid uptake inhibitors: synthesis, biological evaluation, and structure-activity relationship. *European journal of medicinal chemistry*. 2010;45(6):2453-66.

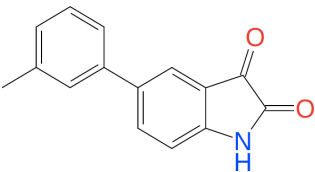
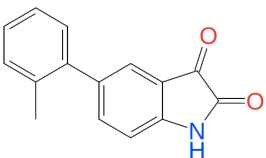
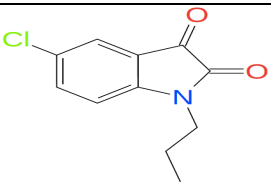
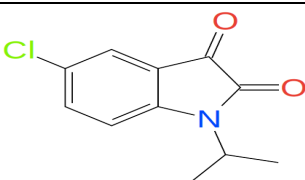
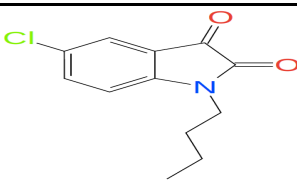
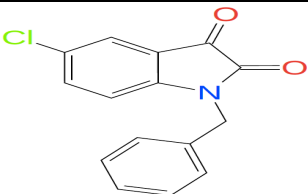
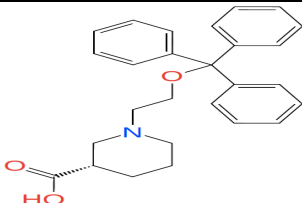
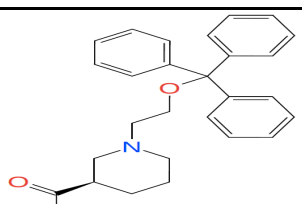
54. Schaffert ES, Hofner G, Wanner KT. Aminomethyltetrazoles as potential inhibitors of the gamma-aminobutyric acid transporters mGAT1-mGAT4: synthesis and biological evaluation. *Bioorganic & medicinal chemistry*. 2011;19(21):6492-504.
55. Zhao X, Pabel J, Hofner GC, Wanner KT. Synthesis and biological evaluation of 4-hydroxy-4-(4-methoxyphenyl)-substituted proline and pyrrolidin-2-ylacetic acid derivatives as GABA uptake inhibitors. *Bioorganic & medicinal chemistry*. 2013;21(2):470-84.
56. Kerscher-Hack S, Renukappa-Gutke T, Hofner G, Wanner KT. Synthesis and biological evaluation of a series of N-alkylated imidazole alcanoic acids as mGAT3 selective GABA uptake inhibitors. *European journal of medicinal chemistry*. 2016;124:852-80.
57. Vogensen SB, Jorgensen L, Madsen KK, Jurik A, Borkar N, Rosatelli E, et al. Structure activity relationship of selective GABA uptake inhibitors. *Bioorganic & medicinal chemistry*. 2015;23(10):2480-8.

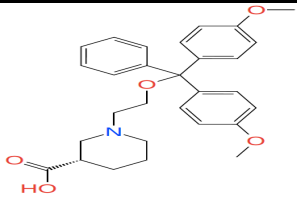
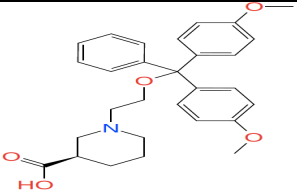
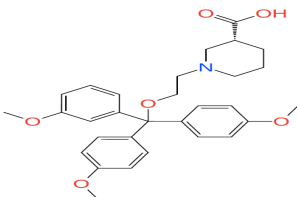
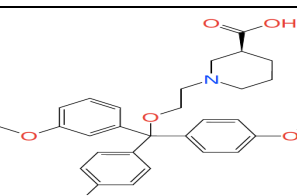
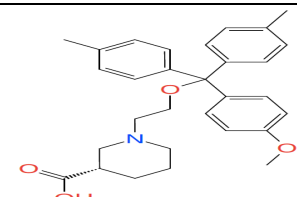
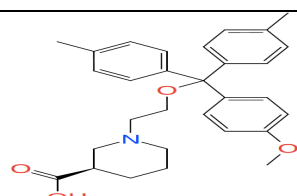
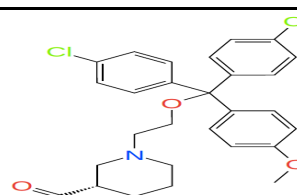
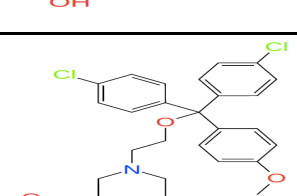
9 APPENDIX

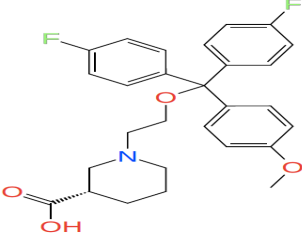
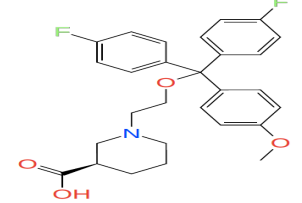
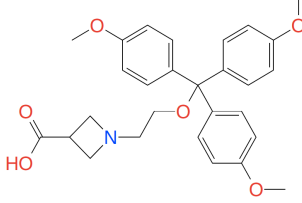
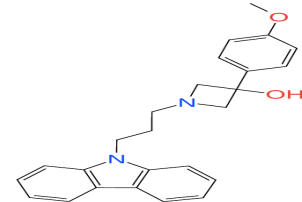
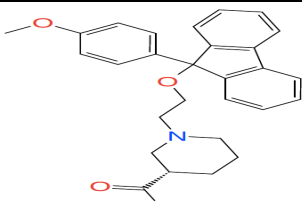
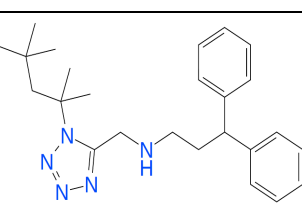
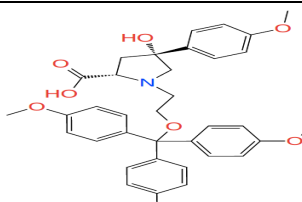
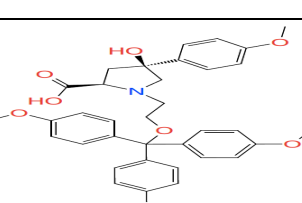
Appendix 1: Overview of the 72 known active ligands

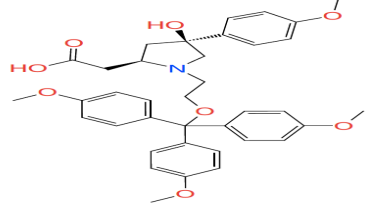
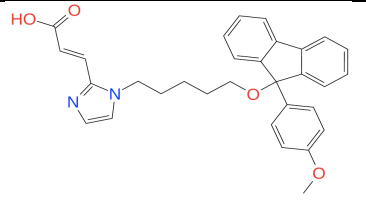
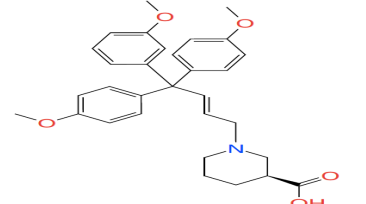
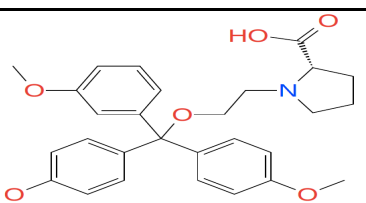
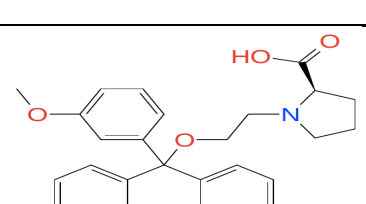
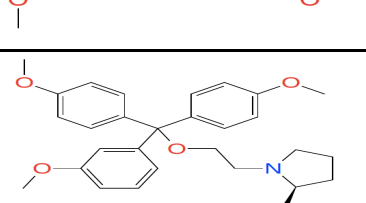
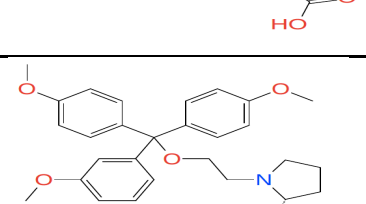
Name	Structure	Uptake inhibition	GAT3/GAT1 selectivity ratio	Reference
Cluster 1				
Compound 6		18 μ M	11 μ M	(14)
Compound 7		17 μ M	6.7 μ M	(14)
Compound 8		18 μ M	60 μ M	(14)
Compound 9		16 μ M	11 μ M	(14)
Compound 10		25 μ M	8.7 μ M	(14)
Compound 11		55 μ M	4.2 μ M	(14)

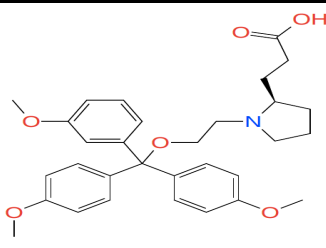
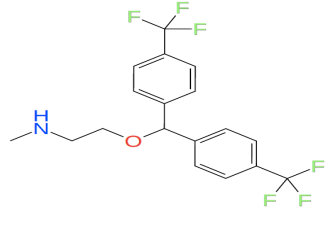
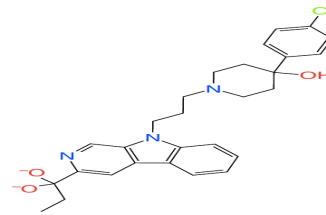
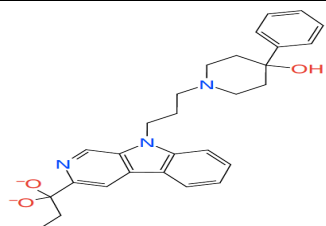
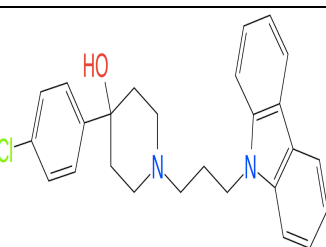
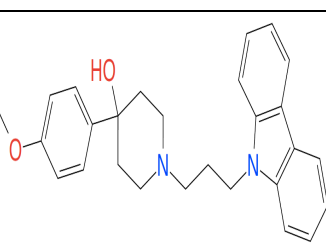
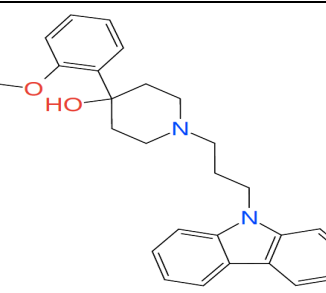
Compound 12		22 μ M	17 μ M	(14)
Compound 13		9 μ M	5.1 μ M	(14)
Compound 14		10 μ M	8.5 μ M	(14)
Compound 15		49 μ M	11 μ M	(14)
Compound 16		15 μ M	16 μ M	(14)
Compound 17		8 μ M	110 μ M	(14)
Compound 19		17 μ M	4.8 μ M	(14)
Compound 20		6 μ M	10 μ M	(14)
Compound 21		22 μ M	3.5 μ M	(14)

Compound 22		35 μ M	> 29 μ M	(14)
Compound 23		24 μ M	> 42 μ M	(14)
Compound 31		52 μ M	4.8 μ M	(14)
Compound 32		46 μ M	10 μ M	(14)
Compound 33		53 μ M	3.5 μ M	(14)
Compound 34		29 μ M	> 34 μ M	(14)
Cluster 2				
(S)-Compound 1		17 μ M	0.1 μ M	(52)
(R)-Compound 1		17 μ M	0.1 μ M	(52)

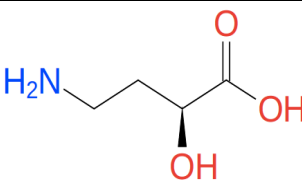
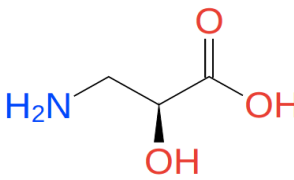
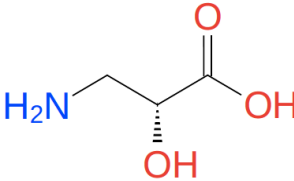
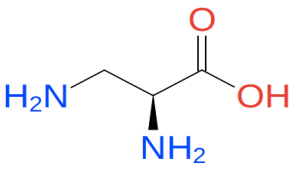
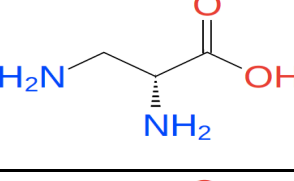
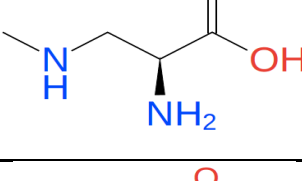
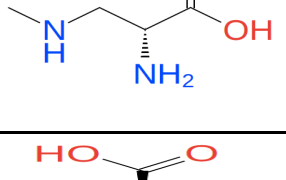
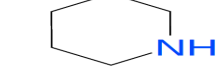
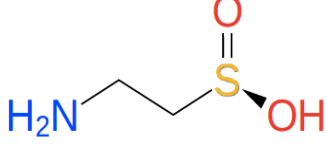
(S)-Compound 3		25 μ M	1.7 μ M	(52)
(R)-Compound 3		25 μ M	1.7 μ M	(52)
Compound 4 (R)-SNAP-5114		86 μ M	5.1 μ M	(52)
Compound 4 (S)-SNAP-5114		5 μ M	77.6 μ M	(52)
(S)-Compound 5		20 μ M	7.1 μ M	(52)
(R)-Compound 5		20 μ M	7.1 μ M	(52)
(S)-Compound 9		29 μ M	4.2 μ M	(52)
(R)-Compound 9		29 μ M	4.2 μ M	(52)

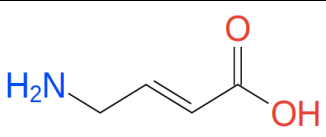
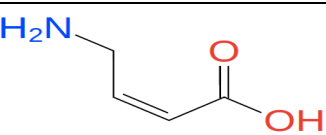
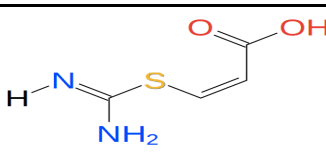
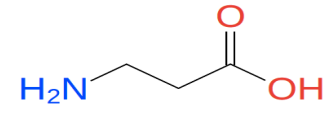
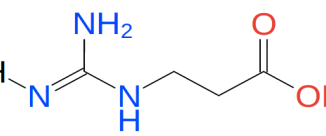
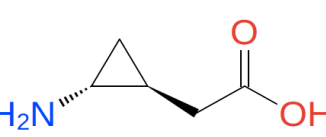
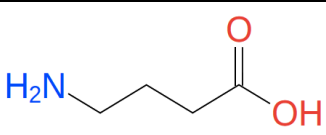
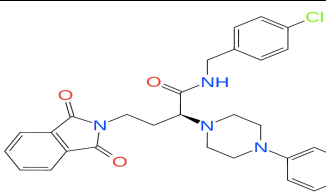
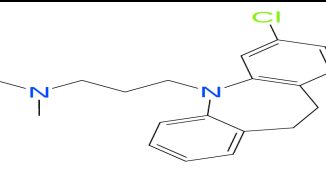
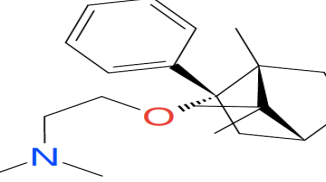
(S)-Compound 10		25 μ M	0.8 μ M	(52)
(R)-Compound 10		25 μ M	0.8 μ M	(52)
Compound 12d		15.3 μ M	76 μ M	(53)
Compound 18e		31 μ M	31.3 μ M	(53)
SNAP-5294		142 μ M	0.9 μ M	(13)
Compound 45		5.09 pIC ₅₀	13.8 pIC ₅₀	(54)
(2S,4R)-10d		29.7 μ M	1.6 μ M	(55)
(2R,4S)-10d		38 μ M	1.5 μ M	(55)

(2R,4S)-9d		19 μ M	5 μ M	(55)
Comp12e		5.13 pIC ₅₀	0.9 pIC ₅₀	(56)
DDPM-1457		5.87 pIC ₅₀	0.7 pIC ₅₀	(46)
(S)-3d		57.7 μ M	2.1 μ M	(45)
(R)-3d		18.5 μ M	7.7 μ M	(45)
(S)-4d		28.7 μ M	1.2 μ M	(45)
(R)-4d		3.1 μ M	21.9 μ M	(45)

(R)-5d		11.2μM	9μM	(45)
Comp9d		58μM	2.8μM	(57)
NNC05-0341		2.8	7.1K _i	(47)
NNC05-1965		2.8	8.2K _i	(47)
NNC05-1973		5.2	8.3K _i	(47)
NNC05-2045		6.1	4.4K _i	(47)
NNC05-2090		15	1.3K _i	(47)

Cluster 3

(S)-(-)-4-amino-2-hydroxybutyric-acid		20μM	0.2μM	(48)
(S)-Isoserine		4.3μM	581.4μM	(48)
(R)-Isoserine		4.3μM	581.4μM	(48)
(S)-2,3-diaminopropionic-acid		11μM	29.1μM	(48)
(R)-2,3-diaminopropionic-acid		11μM	29.1μM	(48)
(S)-2-amino-3-methylaminopropionic-acid		35μM	7.4μM	(48)
(R)-2-amino-3-methylaminopropionic-acid		35μM	7.4μM	(48)
(R)-nipecotic-acid		14μM	0.2μM	(48)
Hypotaaurine		4.9μM	34.7μM	(48)

Cluster 4				
Trans-4-aminocrotonic-acid		7.2 μ M	0.5 μ M	(48)
Cis-4-aminocrotonic-acid		9.5 μ M	29.5 μ M	(48)
ZAPA		51 μ M	0.2 μ M	(48)
Beta-alanine		12 μ M	55 μ M	(48)
3-guanidino-propionic-acid		3.6 μ M	7.2 μ M	(48)
Ila		13.9 μ M	> 7.2 μ M	(24)
GABA		1.4 μ M	1.8 μ M	(14)
Cluster 5				
Compound18		8.5	4.5 μ M	(50)
Clomipramine		40.2 μ M	2.3 μ M	(49)
EGYT-3886		46 μ M	0.6 μ M	(52)

