



# DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

“Homology Modelling of the Taurine Transporter“

verfasst von / submitted by

Eman Samir Allam

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Pharmazie (Mag. pharm.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on  
the student record sheet:

A 449

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Diplomstudium Pharmazie

Betreut von / Supervisor:

Univ. Prof. Mag. Dr. Gerhard Ecker

Mitbetreut von / Co-Supervisor:

Dr. Claire Colas

## Table of Contents

Acknowledgements.....	4
1.0 Introduction.....	6
1.1 Taurines.....	6
1.2 Function of Taurine.....	6
1.3 Taurine Transporter.....	8
1.4 Taurine Transport Mechanism.....	12
2.0 Aim of the Thesis.....	14
3.0 Methods and Materials.....	15
3.1 Homology Modelling.....	15
3.2 Steps to Create the Homology Model.....	15
3.2.1 Finding a Template.....	17
3.2.2 Alignment and Modification.....	17
3.2.3 TM10 and the Preserved Glycine in all the GABA Members.....	18
3.2.4 Options for TM10 Location.....	18
3.2.5. Modeller.....	19
3.3 Docking.....	30
3.3.1 Settings.....	30
3.4 Clustering.....	34
3.4.1 Conformer Clustering.....	34
3.4.2 Volume Overlap Clustering.....	34
3.5 Knime.....	35
3.6 MM GBSA.....	35
4.0 Results and Discussion.....	36
5.0 Outlook and Conclusion.....	44
6.0 Bibliography.....	45
7.0 Appendix.....	54

7.1 Table of Figures .....	54
7.2 Abstract.....	55
7.3 Zusammenfassung.....	56
7.4 List of Abbreviations .....	57

## Acknowledgements

First of all, I want to thank Prof. Dr. Gerhard Ecker for giving me the chance to work on my thesis under his supervision and for providing us with a fabulous learning and training environment in which we could contribute to scientific research and develop our strengths.

Working on my diploma thesis was the most challenging and difficult stage throughout my whole academic career. I haven't just come across hardships and challenges, but undoubtedly also curiosity and happiness made an appearance now and then.

The methods I worked with were completely foreign to me. However, Dr. Prof. Gerhard Ecker encouraged me by always saying: "Every beginning is difficult, but everything can be learned over time and with much patience, you can do it." He always supported me with his wise advice and motivation. Thank you so much, Prof. Dr. Gerhard Ecker, for your trust.

My deepest gratitude is dedicated to my supervisors Dr. Claire Colas and PhD student Stefanie Kickingner, who walked with me this long road of scientific research full of ups and downs. Thank you for being so patient with me and always giving me advice, that made me feel supported. Although Dr. Claire had always been very busy, she never failed to take time for me and to answer my questions. I appreciate your precious time a lot!

A considerable part of the time I spent was with Stefanie, one of the most dedicated and humble persons I got to know. Her patience and competence helped me understanding and getting through the process of my diploma thesis- she is the reason why I never quitted. When you work on such a new project, you often tend to give up, especially when it feels like a never-ending hardship. But Stefanie gave me ease and hope; she was my energy and motivation source. For all of this and much more, I want to say: Thank you, I would have never come so far without you.

I would also like to express my thanks to Dr. Riccardo Martini. My seat neighbour, who brightened every single day of these five months. Thank you for always telling a joke and having a positive mind even on the hardest of days. You had a solution for every technical and thematic difficulty. Sitting next to you calmed my mind when I had thousands of thoughts rushing through my head. Working next to someone else would have never been that funny and entertaining.

Finally, my special gratitude is dedicated to my beloved ones: I want to thank my parents, Samir and Asmaa, and my siblings Amany, Ahmed and Salma for never letting me down. You always had the right words and knew how to motivate me when I needed it the most.

Last but not least, I want to thank all my friends, especially my closest friends Ola and Amina for all of their words of encouragement and mental support.

## 1.0 Introduction

Cell fate, cell migration, cycle regulation, cellular metabolism, and intracellular signalling are profoundly affected by cell volume. Vital cellular processes such as increase in cell volume is often as a result of cell migration and cell mitosis (Hoffmann et al., 2009), which is triggered by a shift in cell volume, while cell apoptosis triggers cell shrinkage. Often, cell volume preservation incorporates active uptake of ions and passive release of organic osmolytes that is influenced by the net locomotion of osmolytes, which is osmotically followed by water.

The process of changing intracellular ionic surroundings or membrane potential through rapid extracellular osmolarity changes is useful in avoiding cellular dysfunction due to accumulated organic osmolytes and ions during cell shrinkage. This process is known as RVI or regulatory volume increase. Alternatively, the RVD or regulatory volume decrease is the process due to cell swelling that results in the release of the osmolytes (Lambert et al., 2008).

### 1.1 Taurine

Taurine can be defined as one of the primary semi-essential organic osmolytes containing a  $\beta$ -amino sulfonic acid that is used in the mammalian cell volume maintenance. Hence, its cellular content is as a result of a balance between active uptake, synthesis, and passive release of ions and organic osmolytes. The progressive accumulation and passive release of taurine in the cells occurs through the taurine transporter (TauT) and unidentified swelling induced release pathway, respectively. For example, in human adults, taurine is manufactured or synthesized by essential amino acid methionine and non-essential amino acid cysteine, mainly in the liver. On the other hand, the primary source of taurine for human neonates is dietary uptake (Lourenco & Camilo, 2002).

The estimated bulk concentration of Taurine in normal plasma is 10-100  $\mu$ M. Approximately more than half of the taurine component is made of the free pool of amino acid in tissues like the retina and heart with 40mM and 6mM concentrations, respectively (Gaulton et al., 2017). Since Taurine is part of the osmotic regulation, it is believed to be implicated in the control of cholesterol, insulin signalling, neuromodulation, detoxification, and antioxidative defence (Schaeffer et al., 2012).

### 1.2 Function of Taurine

Taurine is essential for the growing of the central nervous system, retina, cardiovascular muscle, and skeletal muscle (Kubo, 2016; Huxtable, 1992). Taurine is assumed to be a

biosynthetic precursor to the bile salts, namely sodium taurocholate and sodium taurochenodeoxycholate. As such, taurine acts as an antioxidant that suppresses the physiologically produced hypobromite and hypochlorite toxicity. The reaction between taurine and these halogen agents' results in the formation of N-Bromo- and N-chlorotaurine, whose toxicity is less compared to their precursor's hypohalides (Marcinkiewicz & Kontny, 2014).

It has also been shown that taurine lowers the secretions of lipids and apolipoprotein B100 in HepG2 cells (Yanagita et al., 2008). High amounts of apolipoprotein B100 and serum lipids as a vital structural constituent of LDL and VLDL can lead to serious risk factors of coronary heart disease and atherosclerosis. Therefore, the supplementation of taurine may help in the prevention of these health conditions (Xu et al., 2008)

The normal skeletal muscle development and functioning require the secretion of taurine (Warskulat et al., 2004). For example, mice with taurine deficiency show almost complete levels of cardiac muscle and skeletal muscle depletion, as well as more than 80% reduction of exercise ability than the control mice. Thus, taurine has the capacity of possibly reversing or influencing defects in the nerve sensory, motor nerve conduction velocity and nerve blood flow thresholds in rats according to the results of the experimental diabetic neuropathic (Li et al., 2006; Pop-Busui et al., 2001).

Salimaki et al. (2003) found that taurine can cross the blood-brain barrier, thus inhibiting a wide physiological phenomena array, namely the neurotransmission inhibition. Also, taurine induces long-term hippocampus/striatum potentiation (Dominy et al., 2004); thus, can prevent obesity, and function as an adipose tissue regulator, membrane stabilization, macrophage/neutrophil respiratory burst feedback inhibition, (Tsuboyama-Kasaoka et al., 2006). Taurine is also crucial in the prevention of epileptic seizures, protection against glutamate excitotoxicity, recovery from osmotic shock, and calcium homeostasis (El Idrissi et al., 2003; Leon et al., 2008; Stummer et al., 1995; Foos, & Wu, 2002). It has also been shown that taurine has the potential of preventing tubulointerstitial injury in diabetic nephropathy and preventing diabetes-associated microangiopathy (Verzola et al., 2002). Additionally, taurine causes an anxiolytic effect according to animal studies that can act as an anti-anxiety or a modulator agent in the CNS through glycine receptor activation (Chen et al., 2004; Kong et al., 2006).

Taurine is an inhibitor of glycation in taurine-treated diabetic rats. The application of taurine for rats with diabetic conditions resulted in a reduction of the Advanced Glycation End-products (AGEs) content and formation. For example, according to the US Department of Agriculture, there is a close connection between taurine, folate, lower levels of vitamin B6 in the elderly diets, and the cataract development (USDA, 2012).

Additionally, taurine potentially lowers blood sugar and weight in diabetic rats (Nakaya et al., 2000) by removing fatty deposits in the liver, thus preventing liver disease as well as lowers liver cirrhosis (McCall, 2005). The pieces of evidence available show that taurine lowers blood pressure in male rats. The supplementation of taurine in the diet is known to help animals such as cats that have no enzymatic machinery to be able to produce taurine (Knopf, 2011). There is a possibility that cats may suffer from retinal degeneration and even blindness due to taurine deficiency. Lack of taurine as an essential amino acid in a diet causes reproductive failure and dilated cardiomyopathy in female cats, leave alone tooth decay and hair loss. The Association of American Feed Control Officials (AAFCO) now incorporates taurine as a dietary requirement that any wet or dry food product must label; for example, the feeds must label at least 0.2% and 0.1% of taurine in wet and dry foods, respectively (AAFC Cat Food Nutrient Profiles, 2015).

### 1.3 Taurine Transporter

The human taurine transporter, which is a membrane protein, is a part of the Solute Carrier 6 family (SLC6) transporters. The SLC6 family comprises neurotransmitter transporters (NTT), which are also known as neurotransmitter sodium-dependent symporters (NSS). The SLC6 family is divided into four subgroups, namely the monoamine transporter subfamily, amino acid transporter subfamily (I), amino acid transporter subfamily (II), and the GABA transporter subfamily. SLC members mediate movement of solutes across the cellular membrane. This thesis project focuses on the SLC6A6, the taurine transporter (TauT), which is a secondary active transporter that uses an electrochemical gradient for the transport. TauT is a symporter that couples the transport of taurine as the substrate to the transport of sodium and chloride in the same direction. The GABA transporter subfamily comprises transporters for GABA, betaine, taurine, and creatine. In the brain, the inhibition of GABA transporters reduces clearance after synaptic release, thus enhancing the action of inhibitory synapses.

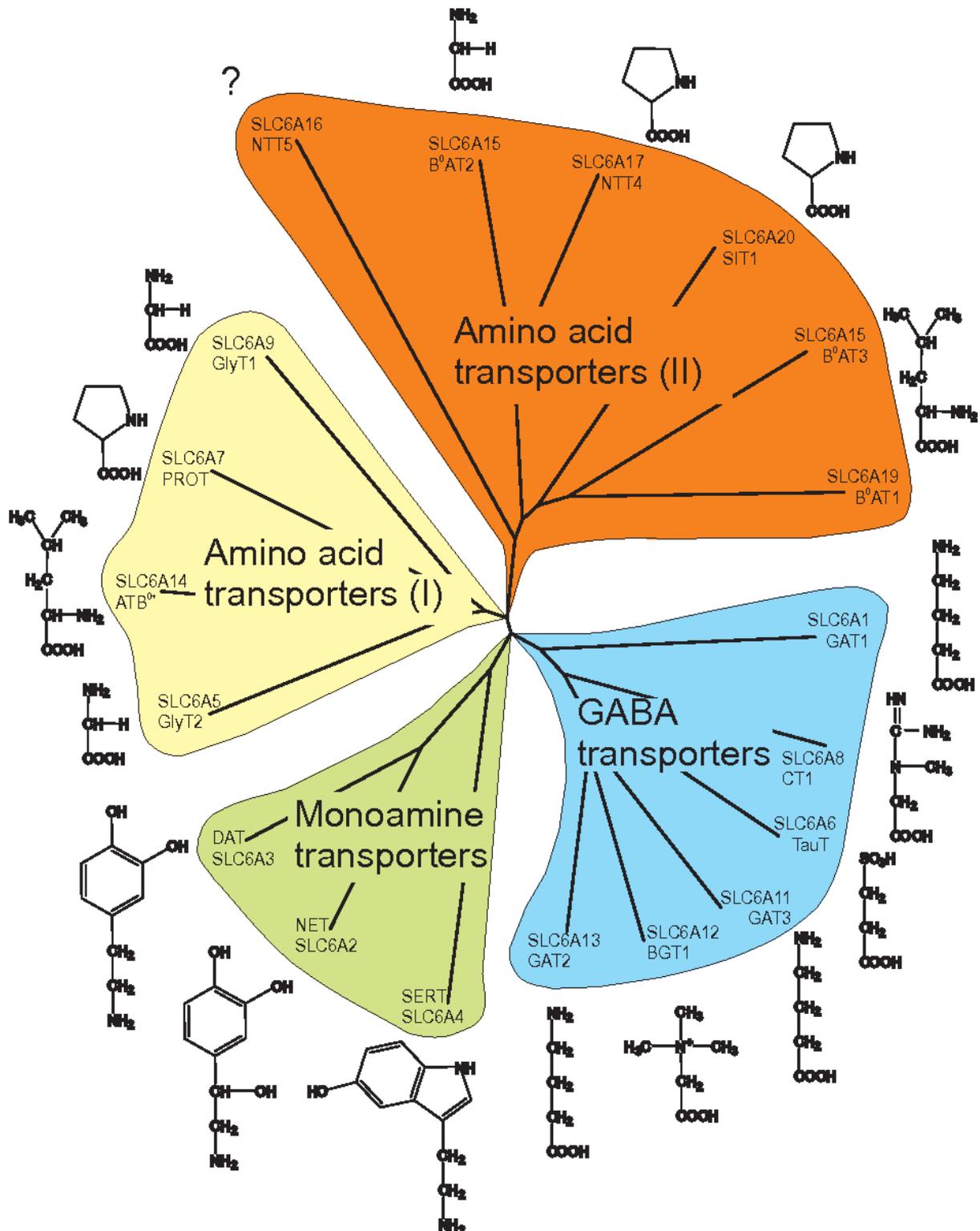


Fig. 1: Showing four Subfamilies of the SLC6 family (Bröer & Gether, 2012)

The accumulation of taurine by cells occurs through the low capacity yet high-affinity Na<sup>+</sup>-dependent taurine transporter TauT alongside the high capacity proton-coupled yet n-dependent  $\beta$ -amino acid transporter PAT1 (SLC36A1), which results in the release of taurine through the

volume-sensitive leak paths and volume-insensitive paths (Anderson et al., 2009). The TauT regulation and the volume-sensitive leak paths for organic osmolytes take place, as shown in Fig. 2, respectively.

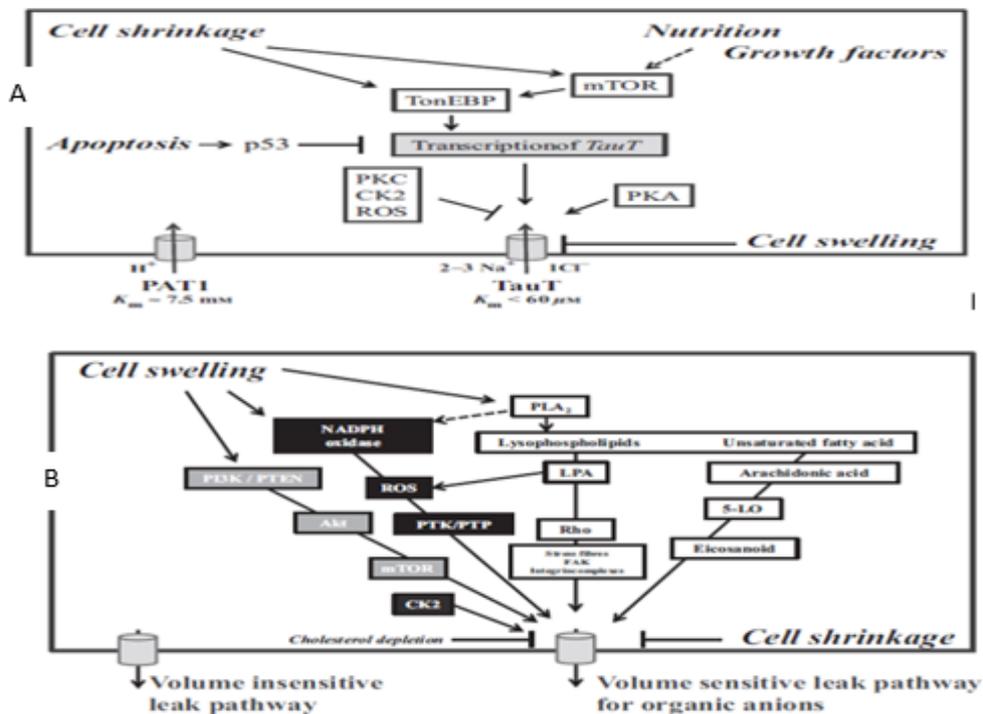


Fig. 2: (A): Secondary active Taurine Transporters (B): Release paths for volume-sensitive taurine (Lambert et al., 2015).

Taurine transport by TauT is both Na<sup>+</sup> and Cl<sup>-</sup>-dependent; to be precise, and taurine transport is insignificant when there is no extracellular Na<sup>+</sup> as well as when there is a considerable reduction of extracellular Cl<sup>-</sup> (Han et al. 2006). TauT comprises 12 hydrophobic transmembrane domains (TMs), the N and C terminus are exposed to the cytosolic compartment.

Also, based on various hypotheses, in the fourth site (S4) location, serine-322 (Ser-322), which is highly conserved, modulates the TauT function through PKC phosphorylation. Based on several studies, tissues of various species, such as pig kidney cells, human placenta, human thyroid cells, mouse brain, and rat brain have been cloned for the taurine transporter (Han et al., 1999). When the mutagenesis of the site was carried out for all these species, Ser-322 came out to be a critical PKC phosphorylation site. Taurine transport activity is found to be three times higher in an oocyte expression system compared with control (wild-type pNCT) when the pNCT Ser-322 was changed to alanine (S322A) as illustrated in Fig. 3.

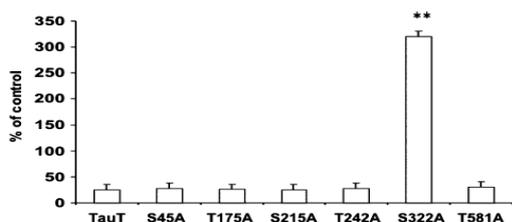


Fig. 3: Showing actions of PKC activation on Taurine uptake upon mutagenesis site-direction (Han et al., 2006).

In figure 3 above, a Ser-322 as the PKC phosphorylation critical site where the injection of Oocytes with 30 ng of wild-type pNCT takes place. The specific mutant pNCT-expressing oocytes or the Wild-type were brooded using 100 mm of active phorbol-ester 12-myristate 13-acetate (PMA) for half an hour to allow measurement of taurine uptake. The  $**P < 0.01$  versus control.

However, acidification acutely down-regulates the TauT activity, causes osmotic swelling of the cell, exposes oxygen reactive species (Lambert et al., 2015; Hansen et al. 2012), and activates the secretions of threonine/serine protein kinase (PKC) (Voss et al. 2004). The process of PKC activation inhibits the maximal transport activity of TauT of various cells (Lambert et al., 2015; Jacobsen et al., 2008; Voss et al., 2004). The TauT stoichiometry by Hill plot indicates a ratio of 2 Na ions: 1 Cl ions: 1 taurine, as illustrated in Fig. 4. (Han et al., 2006).

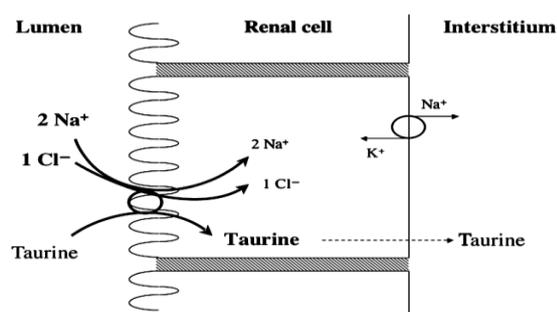


Fig. 4: 2 Na ions: 1 Cl ions: 1 taurine transporter model system across a rat's renal cells (Han et al., 2006).

The taurine transport comprises twelve (12) transmembrane (TMs) in which ten (10) constitute the core of the transporter. The first 10 helices are arranged in a pseudo-two-fold-symmetric pattern of named 5+5 inverted repeat fold. Subsequently, the first 5 helices can be superimposed onto the second 5 helices by a simple symmetry operation which is a common feature among several families of secondary active transporters (Han et al., 2006). Therefore,

helices 1 and 6, which contain unwound regions, which are in contacts with the sodium ions and substrate.

#### 1.4 Taurine Transport Mechanism

The taurine transporter follows a so-called alternating-access mechanism that includes rocker switch, elevator, and gated pore mechanisms. Alternating-access mechanism refers to a general framework that is used in understanding the mechanisms of membrane transporters. This mechanism postulates that there is alternate exposure of substrate on either side of the membrane through membrane transporters conformational changes (Jardetzky, 2016).

A rocker-switch is a major facilitator superfamily (MFS) transport mechanism that is thought to be triggered through the alternating-access mechanism (Huang et al., 2003; Abramson et al., 2003). It has a substrate-binding site, which is relatively immobile, especially within the membrane. It is at this site where the rearrangement confirmation of a membrane transporter takes place (Colas et al., 2016). In the rocker-switch mechanism, the transporter opens to either the cytoplasm or extracellular space and, at the same time, closes the opposing transporter face, thereby blocking the continuous movement across the membrane.

The elevator can be defined as the domain that contains the substrate-binding site (transport domain) that moves along the axis perpendicular to the membrane with a static oligomerization domain. The substrate-binding site in an elevator mechanism, on the other hand, is moved across the membrane through the protein conformational change (Lee et al., 2015). These changes enable alternate access to the sites of the substrate binding. Using rocker-switch principle, the two domains rock against each other to enable binding site access from either side of the membrane. Generally, helices 1, 2, 6, and 7 moves during the transport cycle while the remaining helices form a scaffold.

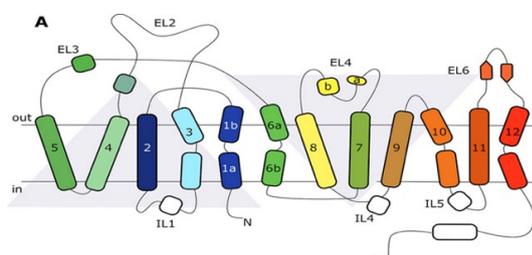


Fig. 5: Overview of LeuT-like fold of TauT. (Kicking et al., 2019)

The gating-pore mechanism involves the enclosure of the transporter binding site by two gates (Forrest & Rudnick, 2009). That is, the gate facing outward opens to allow for the substrate-

binding, in turn, released into the occluded cell through the second gate opening that faces the cytosol, as shown in Fig. TauT follows such a gating-pore mechanism which can be studied by the structure of hSERT and the homolog leucine transporter (LeuT) (Ferrari et al, 2004). According to Ferrari et al., the first site is the mutant L99A, which is caused by the Leu99 substitution  $\rightarrow$  Ala in the protein core, which is nearly hydrophobic while the second site cavity is L99A/M102Q, which is created by the Leu99 double mutant to  $\rightarrow$  Ala and then from Met102 to  $\rightarrow$  Gln. These two sites form one polar residue into the cavity (Ferrari et al., 2004). These sites face small but significant conformational changes on ligand binding, making them good systems for soft docking (Ferrari et al., 2004).

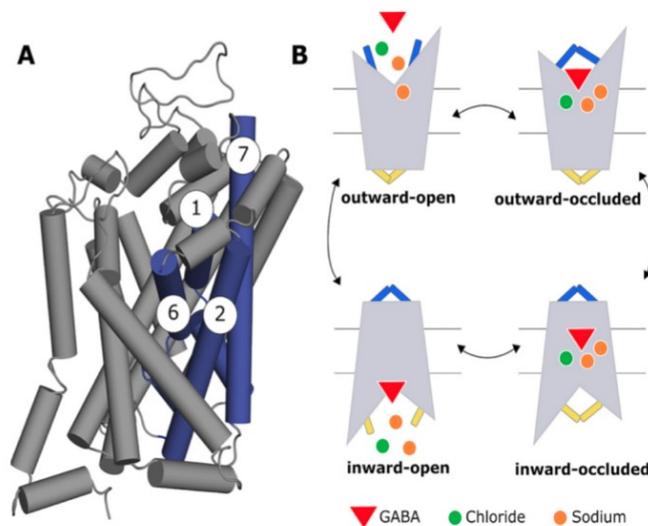


Fig. 6: Taurine transport mechanism: (Stefanie Kicking, et al., 2019)

Generally, the assumption is that the  $\text{Na}^+$  is primarily used for the taurine binding, which in turn triggers the transport cycle. Therefore, it is during the transport cycle that the  $\text{Na}^+$  ions are most likely released into the intracellular compartment as more current is produced by extra taurine when it is added to *Xenopus* oocytes, which in turn overexpress in Bergmann glia and murine TauT (Lambert et al., 2015). However, in EATC, the uptake of TauT-mediated taurine is electroneutral apparently because of the recycling of  $\text{Na}^+$  to the extracellular compartment through the  $\text{Na}^+$ -ATPase (Lambert et al., 2015). Similarly, it is assumed that the  $\text{Cl}^-$  ions also facilitate the second  $\text{Na}^+$  binding to TauT.

## 2.0 Aim of the Thesis

The taurine transporter represents an understudied member of the GABA-subfamily. Since no taurine transporter crystal or cry-electron-microscopy (cry-EM) structure was resolved yet, the specific ligand-transporter interactions that drive activity and selectivity are still elusive. The aim of this thesis was to gain more profound insights into the underlying mechanisms of ligand-transporter interactions on a molecular basis by protein homology modelling and ligand docking.

## 3.0 Methods and Materials

### 3.1 Homology Modelling

Homology Modelling is a methodology used in many studies; however, its usage presumes that proteins that are functionally linked usually share common structural properties implying that they depict the same fold (Centeno et al., 2005). Generally, the number of possible folds is limited (Alberts et al., 2002). Therefore, only two proteins with about 30-40% sequence identity would share similar shapes to create a foundation that can be compared to the transport mechanism (Venclovas & Margelevičius, 2005). Some constituents of solute carrier family (SLC) transporters are an exception, and this enables drawing of structural conclusions as low as 10% of sequence identity, which is attributed to evolutionary conservation of fold-motif (Colas et al., 2016). Therefore, homology modelling is a reliable structure prediction method that results in a 3-D calculated structure using a known template. The template used is based on the fact that the protein fold can be referenced from its primary sequence of the amino acid (Mohammed & Aki-Yalcin, 2019). This process is always referred to as comparative modelling since the structure used for prediction is based on the homologs of the probed structures. This process involves four steps as summarised in section 2.2, Fig. 7. A homology model based upon the homologous structure of the human SERT was created and was further utilised for molecular docking studies. Different clustering techniques as well as MM-GBSA calculations were employed to identify the most plausible binding modes of known inhibitors. Finally, a binding hypothesis was obtained that could explain to some extent the structure-activity-relationship (SAR) observed for known inhibitors. These insights will guide in future the design of new TauT inhibitors.

### 3.2 Steps to create the homology model

The steps for homology modelling are summarised as below:

- Selecting a model template
- Performing alignment between the protein target and the template
- Creating a homology model
- Model validation

The first step in the homology modelling was selecting a template. The template was selected using a known structure of closely related target sequence using the Basic Local Alignment Search Tool (BLAST). This step was done for all the available Protein Data Bank structures

(PDBs) to evaluate and rank the significant number of proteins that match. The PDB was used due to its ability to relay vital information regarding the crystal structures of the template, for example, mutations or ligands. After ranking, it was assumed that the proteins are suitable for selection as template structures. The next step involved running the BLAST to select a query and identity of coverage area (Alstchul et al., 1990). Followingly, a sequence search was performed using conserved motifs to discover the target templates (Sali et al., 1995). The aim was to ensure that the identified sequence between the target and the template is ideally 40% or higher. The reason for this was to make sure that the modelled protein atoms differed only with 1 Å (Angstrom) RMSD due to considerable correspondence between the selected protein and the X-ray structure. However, there was also a possibility of sharing one similar fold among several SLC transporters irrespective of their lower identities, for instance, (below 30% to 10%). This was one of the challenges for this process of modelling.

In step two, I performed an alignment of template and target sequence using PROMALS3D to create the homology model; this was the most crucial step in homology modelling. The step involved combining known structural and sequence data to acquire accurate sequence alignment predictions. The alignment predicted was then utilised to generate real homology models using MODELLER, thereby producing a 3-D output structure that meets all spatial requirements as accurately as possible. The last step involved model refinement, in which the alignment was adjusted manually by aligning the functionally essential residues or minimizing gaps until the most suitable model was produced (Sali et al., 1995).

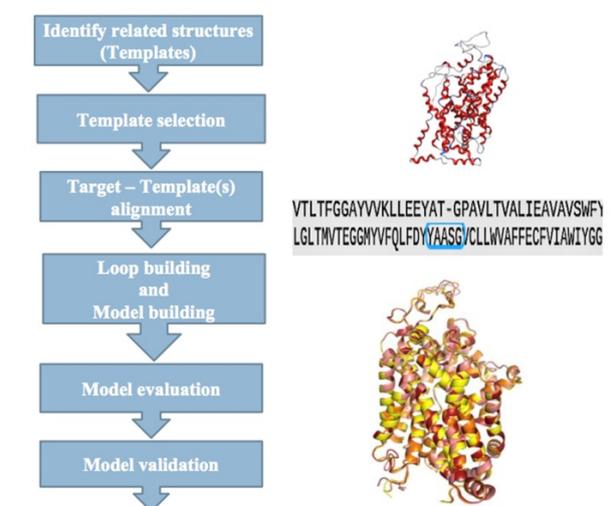


Fig.7: Steps of Homology Modelling



### 3.2.3 TM10 and the preserved glycine in all the GABA members

Vogensen et al. (2015), in their study, postulated that the entry pathway of the GATs could be rendered narrow by the insertion in TM10, which can contribute to various activities obtained by  $\beta$ -alanine and betaine. They also found that  $\beta$ -alanine shows high activity for GAT2 and GAT3, both containing the same amino acid insertion motif in TM10 (AlaAlaSer). The presence or absence of conservative substitutions of amino acids in a particular region of the sequence or within the protein family meant that the region was significant function-wise. From the literature, it can be stated that the TM10 is conserved in all GABA-members (Dayan et al., 2017). All GABA-members have an extra residue in TM10, as such some difficulties were met while creating the homology model; this is because the extra residue is expected to form a "pi-Helix", which is likely to be a requirement for stringent gating and tight coupling of ion and substrate fluxes in the GABA transporter family (Dayan et al., 2017). Regarding where the TM10 could be located, three distinct options were considered.

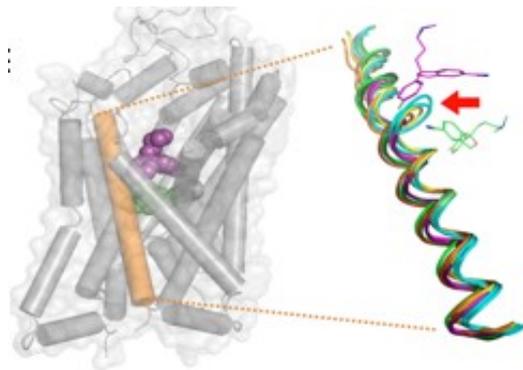


Fig.9: Conserved TM10 and Glycine in GABA-members (Kicking et al., 2019).

### 3.2.4 Options for TM10 Location

The Gap was shifted three times because, after the alignment, as one cannot be sure where the insertion is, on either the Serine(S), Glycine (G) or Alanine (A). Due to the importance of the region around the conserved amino acid glycine, I set up 3 options where the pi-helix could be located by shifting the gap as represented in the figure below. **Option 1** shows the gap aligned to the Serine. In **Option 2**, the gap is aligned to Alanine, while in **Option 3**, the gap is aligned to the conserved Glycine. The main objective of setting the three options was to gain reference material coverage in creating the best homology model.

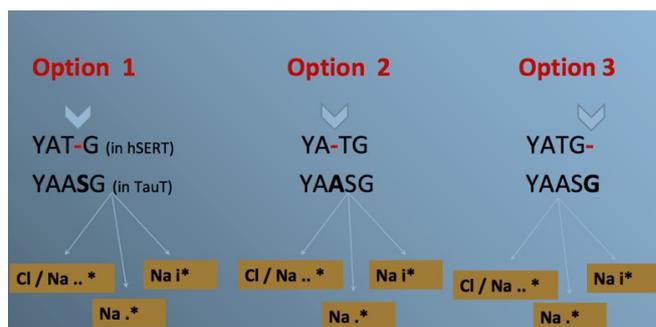


Fig.10: Showing three Options of gap shifting in TM10

**NACL.. \*** means that sodium and chloride ions stayed in the used PDB. **NA.. \*** means that the chloride ion was deleted, and only the sodium ion was allowed to stay in the PDB. **Nai\*** means that also the Chloride ion was deleted. Only the Sodium ion was allowed to stay in the PDB. The PIR file was terminated using \*. For PDB that contained ligands, **ligands.. \*** was used to terminate the PIR file. If the PDB contained ions, the PIR file was terminated using **i \***. This was a clear indication that the PDB was modified three times per option, and the researcher created 100 models for each modification that totaled 300 models per option. Towards the end, it turned out that all modifications indicated the same result, and there were no differences.

### 3.2.5. MODELLER

MODELLER is a program that is used to calculate and generate a model that contains all non-hydrogen atoms. This program works by automatically calculating the atomic coordinate of the template within a few minutes. This program can also carry out other tasks, such as alignment of two protein profiles or sequences, fold assignment (Marti-Renom et al., 2004), multiple protein structures or sequences alignment (Mudhusudhan et al., 2009), de novo loop modelling in protein sequences and phylogenetic calculation (Fiser et al., 2000). I used MODELLER to generate 100 models per Option using a python script *build.py*, thus resulting in 300 models of the transporters. Once the 300 models were generated, the best 10 PDBs were picked by looking at the dope score. This process was carried out automatically using the python script *extract\_score\_final\_python3*. The dope score is a distance-dependent atomic statistical potential used to assess how good the protein modeled is; that is, the lower the value, the better the model (Eramian, Eswar, Shen and Sali, 2008, p. 1890).

### 3.2.5.1 Loop Modelling

Loop modelling requires predictions of loop regions' conformations in proteins by either using a basic template or not. Within a particular protein fold or structural motif, loops often depict highly variable sequences, which correspond to regions that may not be aligned in sequence alignments. Additionally, they are frequently located at the solvent-exposed globular protein surfaces that make them more conformationally flexible. As a result, modelling them using standard homology modeling methods is difficult, hence there is the need for more constrained loop modeling versions in the data fitting steps to solve a protein structure. In this study, loop modeling was performed using a MODELLER to generate 100 different conformations for the loop, with both 4 and 5 loop-defining amino acid for the three Options. For Option 1, 2, and 3, one amino acid around the insertion G was selected, for a total of 5 amino acids. That is, from amino acids 462-466. Similarly, only one amino acid was selected for the insertion without any other amino acid around the insertion, for a total of 4 amino acids (from amino acid 462-465) using the script `loop_modeling.py`. Once this was complete, the best 10 PDBS were automatically selected, and again the best 3 ones were examined visually in MOE.

Option 1	Option 2	Option 3
YAT-G (in hSERT)	YA-TG	YATG-
YAASG (in taut)	YAASG	YAASG

Fig.11: The three Options as visually examined in MOE

For Option 2 and Option 3 after the Loop modelling, 10 loops appear per Option with their normalized dope score. In the case of this study about 10 loops of the normalized dope score marked one in the Top3 Models was selected. Hence, the pi-helix region in TM10 looks fine, as we can see (coloured grey in Option3 and coloured red in Option1). In the second image, the Top3 modelled loops of Option3 appear aligned and superimposed. The colours have been used to describe each loop, for example:

#### Option3

Yellow: template

Other colours: Top3 loops regarding their normalized dope score

Filename	Normalized Dope score
taut.BL99990180.pdb	-1.09456
taut.BL99990171.pdb	-1.09074
taut.BL99990176.pdb	-1.08782

Fig. 12: Showing the selected Top3 modelled loops regarding their normalized dope score.

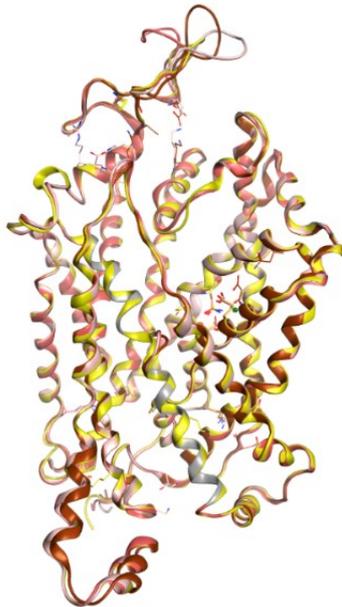


Fig. 13: Showing the selected Top3 modelled loops in the TM10

### Option1

Yellow: template

Other colours: Top3 Loops regarding their dope score

Filename	Normalized Dope score
taut.B99990128.pdb	-1.07879
taut.B99990142.pdb	-1.07692
taut.B99990116.pdb	-1.06649
taut.B99990040.pdb	-1.06577

Fig. 14: Showing the blue selected Top 3 modelled loops regarding their normalized dope score.

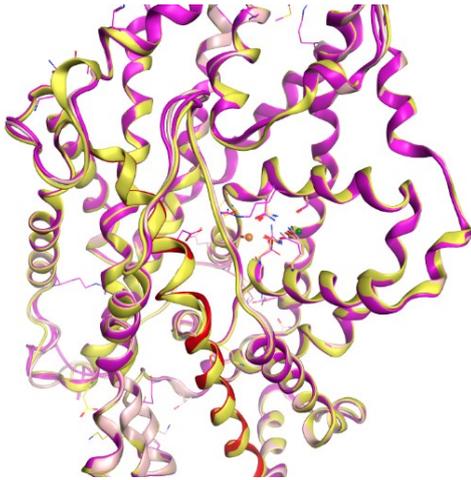


Fig. 15: Showing the selected Top 3 modelled loops in the TM10

### Loop Refinement:

One of the great, challenging problems in the field of computational biology is the prediction of a 3-D protein structure from amino acid sequence. The predicted structure must be of high quality, especially for practical uses that involve design or functional studies (Giorgetti et al., 2005; Zhang, 2009). Template-based modelling can be used in several cases to provide reliable predictions about protein structures through experimental studies (Baker & Sali, 2001). When the templates do not give sufficient structural information, then it is crucial to perform refinement of *ab initio* protein structure for further structural refining. The mainstream method for the refinement problem is based on (MD) molecular dynamics simulations (Lee et al., 2001; Chen & Brooks, 2007; Ishitani et al., 2008); that uses various force fields of molecular mechanics, namely OPLs-AA, CHARMM, and AMBER among others.

In this study, the loop refinement was done after loop modelling, in which Top3 refined model of every modelled loop of every Option were selected as follows:

### Option3

Modelled Loop 180:

Yellow: template

Other colours: Loops regarding their normalized dope score as seen in table and image of refined loops Top3. The Top3 refined loops with the best dope score were not chosen because its **pi-helix** does not look round, as we can see in the first image. Instead, the blue-coloured

modelled loops in the table were chosen because of their pi-helix have a fine look regarding the defined literary pi-helix.

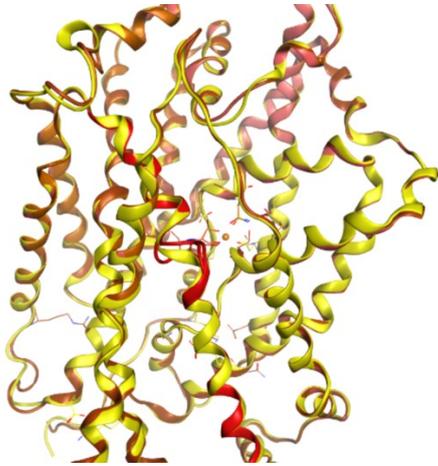


Fig.16: Showing the Top3 modelled loops in the TM10 that were not selected

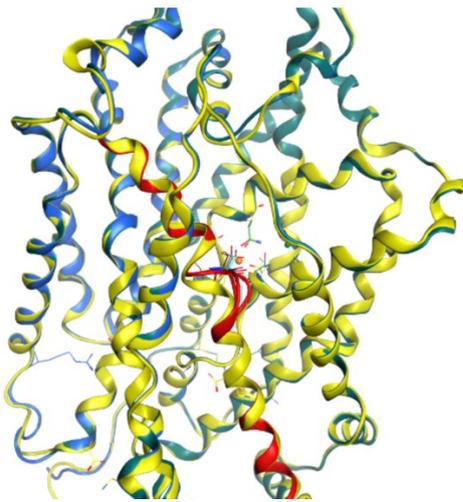


Fig.17: Showing the Top3 modelled loops in the TM10 that were selected

Filename	Normalized Dope score
taut.BL00600001.pdb	-1.10675
taut.BL00240001.pdb	-1.10534
taut.BL00890001.pdb	-1.10523
taut.BL00960001.pdb	-1.10483
taut.BL00090001.pdb	-1.10404
taut.BL00820001.pdb	-1.10383

Fig.18: Showing the Top3 modelled loops selected in blue

Modelled Loop 171:

Yellow: template

Other colors: refined loops regarding their dope score

The refined loops Top3 can be seen from the table and image. These refined loops were chosen with regard to their dope score because they have an insignificant difference from each other in the pi-helix region, as we can see in the image.

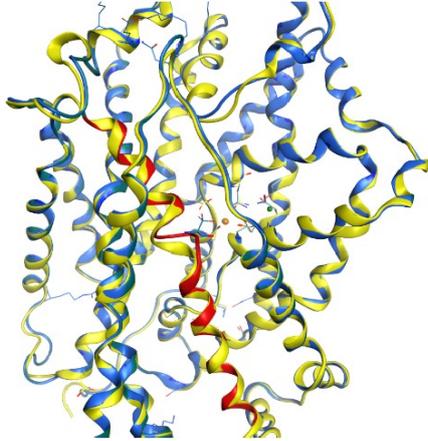


Fig. 19: Refined loops Top3 chosen regarding their normalized dope score due to their insignificant difference from each other in the pi-helix region

Filename	Normalized Dope score
taut.BL00300001.pdb	-1.1006
taut.BL00530001.pdb	-1.10047
taut.BL00580001.pdb	-1.10033

Fig.20: Showing the Top3 modelled loops selected in dark blue

Modelled Loop 176:

Yellow: template

other colors: refined loops regarding their normalized dope score

The refined loops Top3 can be again seen from the table and image. These refined loops were chosen for the same reason as mentioned above.

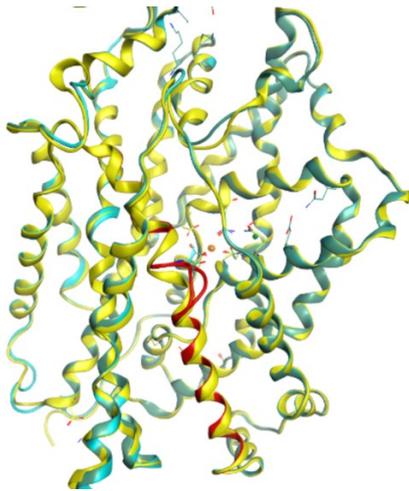


Fig.21: Showing the Top3 modelled loops aligned with each other

Filename	Normalized Dope score
taut.BL00010001.pdb	-1.10428
taut.BL00230001.pdb	-1.10235
taut.BL00490001.pdb	-1.10203

Fig.22: Showing the Top3 modelled loops selected

### OPTION 1

After Loop modelling for Option 1, the 10 modelled Loops with their normalized dope score appear. As such, based on their normalized dope score, the ones marked as Top3 Models were selected the marked; hence, the pi-helix region in TM10 had a fine look depicted by TM10 as we can see in the first image.

Modelled Loop **128**:

Yellow: template.

Other colors: Top3 modelled loops regarding their normalized dope score

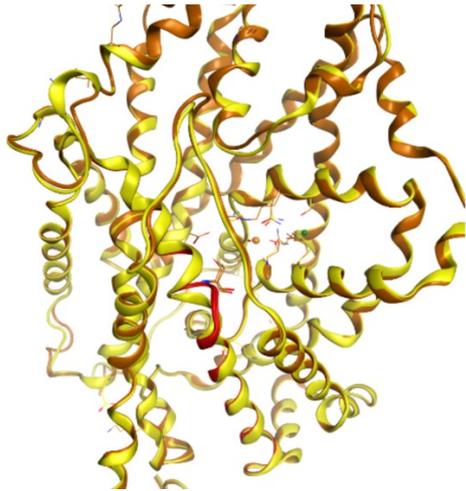


Fig. 23: Top3 modelled loops appearing aligned and superimposed

#### Table of Normalized Dope Score of the Top3 refined Loops

Filename	Normalized Dope score
taut.BL00440001.pdb	-1.09828
taut.BL00580001.pdb	-1.0976
taut.BL00800001.pdb	-1.09724

Modelled Loop 116:

Yellow: template

Other colors: Top3 refined loops regarding their dope score

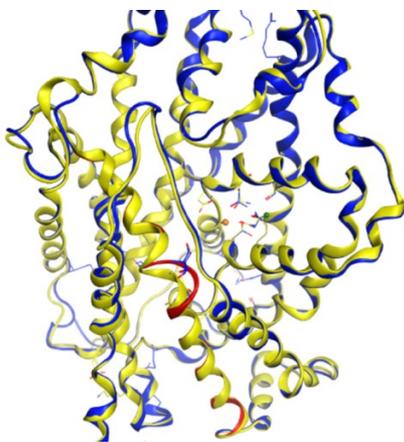


Fig.24: Showing Top3 loops chosen based on their normalized dope score

**Table of Normalized Dope Score of the Top3 refined Loops**

Filename	Normalized Dope score
taut.BL00800001.pdb	-1.1006
taut.BL00530001.pdb	-1.10047
taut.BL00580001.pdb	-1.10033

Modelled Loop 40:

Yellow: template

Other colours: Top3 loops regarding their dope score



Fig.25: Showing Top3 loops chosen based on their normalized dope score

**Table of Normalized Dope Score of the Top3 refined Loops**

Filename	Normalized Dope score
taut.BL00060001.pdb	-1.08315
taut.BL00410001.pdb	-1.08198
taut.BL00090001.pdb	-1.07962

**3.2.5.2 DSSP**

Dialog System for Structured Programming (DSSP) refers to a standard program that is used to assign secondary structure to a protein's amino acids as well as defining geometrical characteristics and solvent exposure of proteins, provided the proteins' atomic-resolution coordinates are known. So, DSSP is a secondary structure assignment database for all entries of protein made in the PDB. This program works by calculating the possibility of secondary structure assignments, provided a 3D protein structure. It carries out this function by reading all atoms' position in a protein and then calculating the energy of hydrogen bond between all

atoms (Kabsch & Sander, 1995). By doing so, the algorithm is said to discard any hydrogen found in the protein input structure and subsequently calculate the positions of all optimal hydrogen from the backbone N in the opposite direction from the backbone C=O bond by placing them at 1.000 Å. Lastly, the best hydrogen bonds for each atom are then used to examine the secondary structure that is most likely for each residue in the protein (Kabsch & Sander, 1995).

In this study, the DSSP program was used to recognise pi-helices. However, they were hard to identify because the pi-helices appeared inside an alpha-helix; however, the program always gives priority to alpha-helices. A Perl script, which can better define pi-helices, was used as the script pi-hunt.

The last analysis step was taken, and this involved taking the PDB of all models with the selected refined Top3 loops and dragged them into MOE to add the PDB template. Subsequently, all the PDBs were aligned and superimposed for comparison. Each of the two top refined loops per Option were visually examined. The top2 refined loops from Option 1 can be seen in orange and the top2 loops from Option 3 in turquoise aligned and superimposed in the first image. The script evaluate-model.py was used for the 4 loops seen in the picture in orange and turquoise to select the best model with the best-refined loop to get the best homology model at the end, regarding their dope scores. In the second image, we can see in orange the two best models of Option1 with the best-refined loops and their dope score and in turquoise the two best models of Option3 with the best-refined loops and their dope score. In the end, I chose model 176 with the refined loop 23, which has the most negative score. As we know, the more negative the dope score, the better the model is. The model is illustrated in Fig.27.

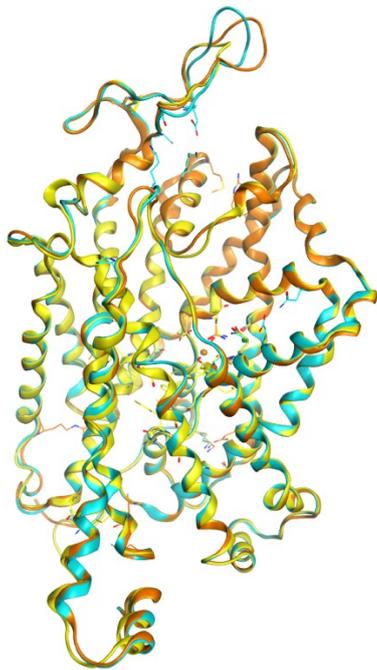


Fig.26: Showing coloured orange the Top2 models of Option1 and coloured turquoise of Option 3 aligned with the yellow template

	Best Model	Best refined Loop of the Model	Dope score of the best refined Loop after Evaluation
Option1	116	43	-1.08936
	116	58	-1.08877
Option3	176	23	-1.10235
	176	49	-1.10203

Fig.25: Showing results of the dope scores after evaluation

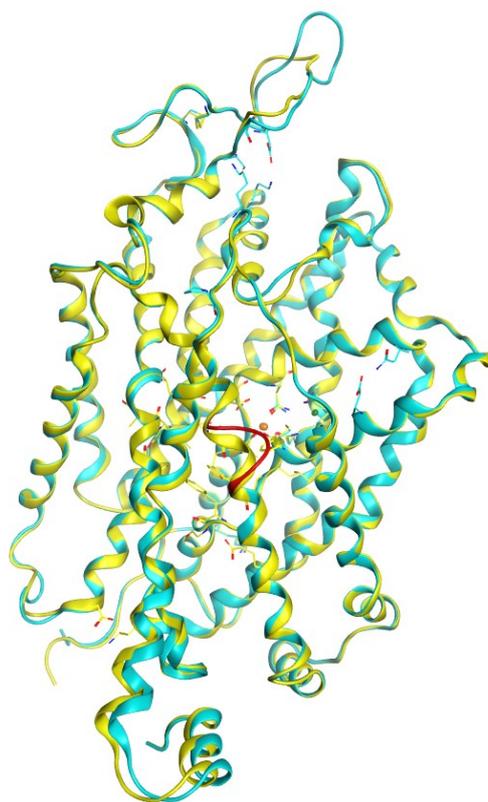


Fig.27: Showing the **final homology model 176** with the dope score **-1.10235** colored turquoise. The template is colored yellow.

### 3.3 Docking

Docking or molecular docking refers to a procedure or method of predicting the preferred interactions of a particular molecule with a target. Alternatively, it can be defined as a technique for modelling molecules that are used to predict the interaction of the protein with ligands (molecules) (Sali et al., 2000). Often, docking is performed to predict how ligands are oriented in a protein's binding site to determine the information stated by the interactions. Hence, it is possible to determine the complexity of a ligand-receptor interaction. As a result, it is possible to create a binding hypothesis for the ligand and the protein under investigation. In this project, the docking was performed using Glide (Schrödinger Release, 2019).

#### 3.3.1 Settings

Ligand and protein preparation settings were determined before running the docking by setting the grid and adjusting the docking conditions.

### *3.3.1.1 Protein Preparation*

#### *3.3.1.1.1 PDB-file*

The PDB-file accession code 5i6x was used in this project in the outward open conformation and complexed with paroxetine in both allosteric and orthosteric sites. The X-ray had a resolution of 3.14 Å.

#### *3.3.1.1.2 Protein Preparation Wizard*

Protein preparation started by uploading the PDB-file that contain the protein of interest. Subsequently, an algorithm was added to make the structure suitable for the use with Glide as well as to add the missing information in the PDB-file. The program was selected due to more accurate results that it generates when fed with information concerning reoriented side chains, ionization states, and bond orders for available options naming for the protein preparation. Moreover, optimization of the hydrogen bonding network was carried out to minimize the entire structure of the protein (Sastry et al., 2013). The last step involved undertaking a problem report to check the final protein structure, which was not encountered in my project.

#### *3.3.1.2 Ligand Preparation*

The preparation of ligand was carried out to ensure structures suitability for Glide use; this procedure involves variation generation and structures correction and optimization. A minimized 3D structure preparation was generated by adding hydrogens to all the ligands. Here, the preparation of ligand was carried out using a LigPrep panel (Sastry et al., 2013), which produced various pH ionization states with a range of 7.0 +/- 0.5 for physiological conditions imitations. Epik was utilised for the generation of ionization states, in which both tautomerization and ionization were performed at the same time as per (Shelley et al., 2007). The penalty of Epik state was calculated in kcal/mol, which made it directly fit for use in GlideScore for docking. As a result, it was possible to find out the effect of adding the Epik state to the GlideScore (Sastry et al., 2013). In Glide, the total sum of the Epik state penalty and the GlideScore is known as DockingScore, which is used for calculating final ranking and enrichment (Epik, 2011). Moreover, there was a treat metal-binding states mode in Epik that was used to increase the range of pH for the steps of state generation while reducing the penalty in the docking stage with over 100 likely states. Here, stereoisomers were generated by retaining chirality's from the input structures, in that way only varying stereochemistry for chiral centres where the chirality is undetermined. The researcher altered this setting for the various experiments that they performed, as illustrated in Figures 27 and 28 below.

Ligand	Structure	Activity ( $\mu\text{M}$ )	Activity type
Hypotaurine		10.0	IC50
$\beta$ -Alanine		40.0	IC50
L-diaminopropionic acid		66.1	IC50
Guanidinoethanesulfonic acid		89.1	IC50
Guanidinopropionic acid		117.5	IC50
JK-70-3E		249.3	IC50
3-Amino-1-propanesulfonic acid (homotaurine)		631.0	IC50
IAA		658.6	Ki
$\gamma$ -Aminobutyric acid		1070.0	IC50
Nipecotic acid		2020.0	IC50

Fig. 28: Ligands with known IC50 value used for the LigPrep

Ligand	Structure	Activity (%)	Activity type
Taurine		14.3	Effect (% of control)
L- $\alpha$ -Alanine		70.5	Effect (% of control)
Creatine		75.6	Effect (% of control)

Fig. 29: Ligands with a known percentage of inhibition used for the LigPrep

### 3.3.1.3 Generation of Receptor Grid

Docking of the prepared ligands took place inside the receptor grid, thereby representing the active site properties and shape of the protein. The receptor grid was set up using the Receptor Grid Generation panel that enables the uploading of the receptor structure; thus, one can determine the position and size of the binding site. In the event of necessary constraints, either metal or H-bond positional constraints were used. Besides, a prepared protein structure was used for the grid generation.

### 3.3.1.4 IFD-Induced Fit Docking

From Koshland's (1960) hypothesis, the protein and ligand interaction are assumed to be a binding site adjustment continual process. Numerous possible binding site conformations due

to the dependence of the binding process on the ligand's ruling properties. The false ligand ranking can, therefore, be evaded through assuming that both the protein-receptor and ligand are flexible during the docking process. This will, as a result, minimize the generation of erroneous results through docking and simultaneously achieve extra protein conformation (Meng et al., 2011). Hence, the three steps below can be used as a procedure of induced docking protocol.

- i. Computation of various ligand poses through regular ligand docking to give the initial protein structure by use of Glide.
- ii. Actual induced-fit priming through adjusting every binding-site to its specific Output ligand.
- iii. Redocking of ligand to create a newly obtained receptor conformations that can then be assessed and graded based on GlideScore for Primer energy and redocking (Schrödinger, 2011).

It was hard to validate the process since only 16 ligands for the taurine transporter were available for this project. Therefore, I decided to do an IFD, meaning that the protein's side chains were rendered flexible to give more variability. The amino acids chosen for the induce-fit docking grid were: Leu468, Ala463, Glu406, Gln403, Ser402, Leu306, Ala303, Tyr302, Ser301, Phe300, Tyr138, Asn135, Leu134, Val131, Phe58, Gly57.

Ligand docking is a computational tool that is vastly used for identification, prediction, and depiction of relevant interactions between different proteins and ligands. The concept of ligand docking can be used to illustrate and even explain some ideal protein-ligand interactions, thus allowing the estimation of different ligands binding affinities. It also creates a pathway for screening for new unknown ligands, which is based on rational docking calculations (Seeliger and de Groot, 2010; Feinstein and Brylinski, 2015). Schrodinger's docking program **Glide** was used for calculations (Schrodinger, 2019).

As discussed in Induced Fit Docking, IFT was used when more precise output was required. This process involved using the standard docking protocol for Glide, including preparation of protein, ligand, generation of the grid then docking. The acquired ligand pose that contains specification and spatial orientation of ligand-based on the binding of protein was acquired according to (Repasky et al., 2012). Therefore, the calculated ligand poses are sieved and narrowed down through the hierarchical filters sequence process to rate ligand poses based on

different values, for example, GlideScore, which in turn is used for binding affinity prediction and rank-ordering. The selection aim of the hierarchical filter sequence process is to terminate ligand conformations that are not relevant and thus allow the provision of accurate ligand poses for further examination (Stenger et al., 2006).

### 3.4 Clustering

A cluster refers to a conformere set that deviates from the central conformer below the allowed threshold. This deviation is the maximum distance calculated between the matched atoms, which include all heavily superimposed atoms based on an individual atoms' subset that the user indicates or the graph automorphism mappings (Vesterman et al., 1996). The central conformer can be used as a significant representative of a cluster that meaningfully decreases efforts of molecular modelling without features of the system is lost.

#### 3.4.1 Conformere clustering

Conformere Clustering is done after IFD using the algorithm of divide-and-conquer to create feasible conformations of molecules. This method involves three steps, namely dividing the input molecules into pieces (breaking exo-cyclic rotatable bonds), obtaining conformations for all the pieces obtained and lastly, building an entire molecules' conformations by reconnecting the pieces in different ways (Schrödinger Suite 2009; Watts et al., 2010).

#### 3.4.2 Volume overlap clustering

Volume Overlap Clustering is also done after performing IFD using the Clustering Based on Volume Overlap panel in Maestro (Schrödinger Suite 2009). Its matrix is calculated using a single linkage with a fixed atom radius of 0.5 Å based on a Simplified Molecular Input Line Entry System (SMARTS $\ddagger$ ) of the common atoms, which is a framework of additional common carbon atoms). In Volume Overlap Clustering, groups are clustered based on their standard binding pose; this is done to enable impact comparison of various profiles of hydroxylation on the binding. As a result, the selection of a single representative binding pose from each populated cluster that attained the required common pose with the best possible rank score is carried out. The selected clustered poses can thus be used to build structure-based pharmacophores for depicting and rationalizing the specificity of the substrate specificity as well as for screening new compounds databases.

The researcher found it was quite challenging to create a SMART that included all common substructures because all ligands were very different in their functional groups. Therefore, the

Conformere Clustering did not generate any outcome with SMART. However, after several trial and error methods, the researcher managed to create a SMART that was used to perform Volume Overlap Clustering; hence, my project used Volume Overlap Clustering.

### 3.5 KNIME

Konstanz Information Miner (KNIME) is a modular platform that is used to build and execute workflows through predefined components known as nodes (Schrödinger Suite 2009). It also refers to a comprehensive, user-friendly, open-source platform that is used to integrate, process, analyse, and explore data. I used KNIME to filter results after Clustering; that is, it was possible to quickly and easily automatically calculate the standard deviation and mean values of all the scores.

### 3.6 MM GBSA

Molecular Mechanics with generalised Born and surface-area (MM-GBSA) solvate calculation method used to calculate Prime uses of free bonding affinities ( $\Delta G$ ). The binding affinities are estimated in kcal/mol according to protein and ligand interaction properties. This operation is performed by subtracting the sum of the calculated receptor and ligand energy from the complex energy calculated (Schrödinger, 2015; Mulakala & Viswanadhan, 2013). Then, the obtained data were compared to determine whether the calculated binding energies were matching with the measured values, which was, in turn, used to confirm the docking poses used in my project and hence reinforcing the substrate affinity theory if necessary. Similarly, it is essential to note that the homology model development accuracy is one of the predictions limiting factors. Most studies rely on correctness and quality. In the project, the researcher decided to use a suitable prediction tool to confidently examine binding poses, for example, for unknown substrates or inhibitors.

## 4.0 Results and Discussion

To gain more detailed results on the method's efficiency, the volume overlap clustering, and the conformere clustering was compared with each other. It is in this sense notable that the conformere clustering approach failed in meeting the aspired outcomes, whereas the volume overlap clustering could deliver results. The latter approach can be described as follows: the aim was to find an order (SMART) that enables the volume overlap clustering to define substructures based on SMILES extensions (vgl daylight). Multiple tries were carried out until the aforementioned SMART could be found. This step is followed by the production of the volume overlap clustering's results: This method delivered 26 clusters. To gain a better overview of the most popular clusters, KNIME was used. A workflow was firstly designed for the data analysis, which is considered as an automatized process. It delivered the standard deviation of the glide gscore, glide emodel score, and the IFD score and their respective means. Further, we aspired to gain an overview of the most popular clusters with the most poses and to define the unique counts within these clusters. The following picture serves as an exemplification.

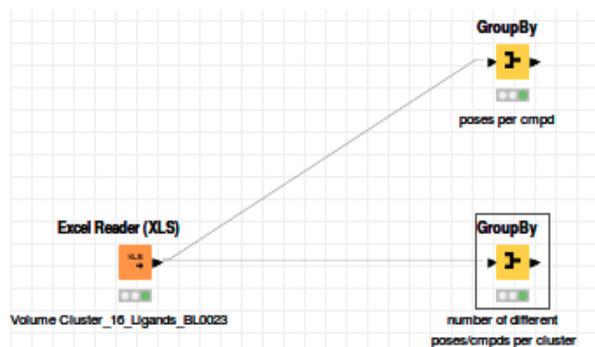


Fig. 30: KNIME workflow

Row ID	Canvas Cluster Index	Count(Title)	Unique count(Title)
Row6	7	84	13
Row8	9	40	12
Row13	14	22	10
Row11	12	20	7
Row19	20	10	3
Row4	5	8	3
Row14	15	8	3
Row2	3	4	3
Row10	11	4	4
Row0	1	3	2
Row9	10	3	1
Row23	24	3	2
Row1	2	2	1
Row16	17	2	1
Row22	23	2	1
Row25	26	2	2
Row3	4	1	1
Row5	6	1	1
Row7	8	1	1
Row12	13	1	1
Row15	16	1	1
Row17	18	1	1
Row18	19	1	1
Row20	21	1	1
Row21	22	1	1
Row24	25	1	1

Fig. 31: Cut out of the KNIME workflow to show the two most popular clusters

As the data shows, the most popular cluster is number 7, constituting 84 poses and 13 unique counts as well as the second most popular cluster, number 9, constituting 40 poses and 12 unique counts. In this context, the aim was to find out which of the 13 ligands could be identified within the most popular cluster, if any of them could not be found and if there are any findings within the second most popular cluster. It could be determined that Creatine, Guanidinopropionic acid, and IAA were not to be found within the most popular cluster but within the second most popular cluster. The following table serves as an exemplification.

Ligand	Cluster number 7 (most popular cluster)	Cluster number 9 (2 <sup>nd</sup> popular cluster)
Hypotaurine	✓	
JK-70-3 E-R	✓	
JK-70-3 E-S	✓	
β-Alanine	✓	
L-diaminopropionic acid	✓	
Guanidinoethanesulfonic acid	✗	✓
Guanidinopropionic acid	✓	
3-Amino-1-propanesulfonic acid	✓	
IAA	✗	✓
Taurine	✓	
L-α-Alanine	✓	
Creatine	✗	✓
Guvacine	✓	
Nipecotic-acid-R	✓	
Nipecotic-acid-S	✓	
γ-Aminobutyric acid	✓	

Fig. 32: To show which Ligand is in which cluster

Hypotaurine, as it can be seen, has the best IC<sub>50</sub> activity, and 3-amino-1-propanesulfonic acid has the worst IC<sub>50</sub> activity. It was, in this sense, interesting to investigate whether the best poses of each ligand are to be found within the most popular cluster. Unfortunately, this was met with disappointment, because it was expected that the Hypotaurine's pose would lie within the most popular cluster – which is not the case. The reason for this can probably be linked to the binding energy of Hypotaurine. In addition, an MM GBSA analysis was carried out to compare Hypotaurine (with the highest IC<sub>50</sub> activity) and 3-amino-1-propanesulfonic acid (with a lower IC<sub>50</sub> activity).

prime_mmgsa_hypotaurine-out1 (2)							
Hypotaurine_protein		9375654	-6.359	-6.359	-35.503	-916.15	
Hypotaurine_ligand		9375654	-6.359	-6.359	-35.503	-916.15	-33.79
prime_mmgsa_propanesulfonic-out1 (2)							
3-amino-1-propanesulfonic-acid_protein		9375654	-7.158	-7.158	-46.581	-915.84	
3-amino-1-propanesulfonic-acid_ligand		9375654	-7.158	-7.158	-46.581	-915.84	-20.75

Fig. 33: Cut out of the MM GBSA Panel in Maestro

Throughout the research, it could be recognized that the MM GBSA energy of the 3-amino-1-propanesulfonic acid was at -20,75 and therefore scored higher than Hypotaurine, which was -33,79. Similarly and expectedly, the Ligand Strain energy of 3-amino-1-propanesulfonic acid showed a value of 0,205711, while the Ligand Strain energy of Hypotaurine was at 0,0816880 – again smaller than what the 3-amino-1-propanesulfonic acid delivered.

Further, based on the MM GBSA analysis, it could be determined that the MM GBSA binding energy failed at meeting the expectations. It was expected that the MM GBSA binding energy

of Hypotaurine would be lower (demonstrating a higher IC<sub>50</sub> score) than that of the 3-amino-1-propanesulfonic acid (demonstrating a lower IC<sub>50</sub> score). Therefore, it was assumed that other ligands would propose similar results within the MM GBSA analysis. Ligands with a higher IC<sub>50</sub> score were supposed to propose lower values of the binding energy than ligands with a lower IC<sub>50</sub> score – which was not the case.

For this reason, a further MM GBSA analysis for the three ligands constituting of a Guanidino group was carried out, namely Creatine & Guanidinopropionic acid & Guanidinoethanesulfonic acid. Furthermore, another analysis, taking into account ten other ligands out of the most popular cluster (cluster number seven), was carried out.

The step following the analysis was marked by establishing two tables: the first table should serve as a demonstration for the two most popular clusters as well as all ligands, which proposed an IC<sub>50</sub> activity. Further, the table also depicts the MM BGSA binding energy and the Ligand strain energy of each ligand. The respective emodel score, which was used for the MM GBSA analysis, as well as the docking score and the best emodel score from the Induced Fit Docking, are also observable. The second table consists of all ligands, of which the percentages of inhibition could be accessed. The same parameters of the first table were used for the second.

Ligand	Cluster number	IC <sub>50</sub> (μM)	MM GBSA (binding energy)	Ligand strain energy	emodel score (used for MM GBSA)	best emodel score of IFD
Hypotaurine	7	10.00	-35.61	0.260889	-34.727	-35.503
β-Alanine	7	40.00	-17.18	0.156512	-30.058	-33.341
L-diaminopropionic acid	7	66.10	-14.37	0.374997	-27.220	-41.089
Guanidinoethanesulfonic acid	9	89.10	-33.58	1.329466	-33.038	-33.038
Guanidinopropionic acid	9	117.5	-37.25	0.844175	-35.949	-35.949
3-Amino-1-propanesulfonic acid	7	631.0	-20.75	0.205711	-46.581	-46.581
γ-Aminobutyric acid	7	1070	-27.29	0.080976	-41.491	-42.136
Nipecotic acid S	7	2020	-25.16	0.174705	-39.991	-41.665
Nipecotic acid R	7	2020	-25.27	0.020364	-44.074	-47.907
Guvacine	7	4190	-21.37	0.089586	-40.506	-46.327

Fig. 34: Ligands with a known IC<sub>50</sub> value

Ligand	Cluster number	Effect (% of control)	MM GBSA (binding energy)	Ligand strain energy	emodel score (used for MM GBSA)	best emodel score of IFD
Taurine	7	14.30	-16.04	0.072017	-33.019	-36.555
L- $\alpha$ -Alanine	7	70.50	-17.73	0.621192	-17.541	-20.661
Creatine	9	75.60	-21.28	2.519927	-30.091	-30.091

Fig. 35: Ligands with a known percentage of inhibition

Yellow marked ligands in the table represent ligands that carry a Guanidino group. When thus comparing the ligands – the Guanidinoethanesulfonic acid, which is marked orange within the cluster pose and the Guanidinopropionic acid which is marked yellow, it becomes evident that both ligands are not only folded but constraint too. This is precisely what the Strain energy demonstrates. The higher the MM GBSA binding energy, the higher the Ligand strain energy will be.

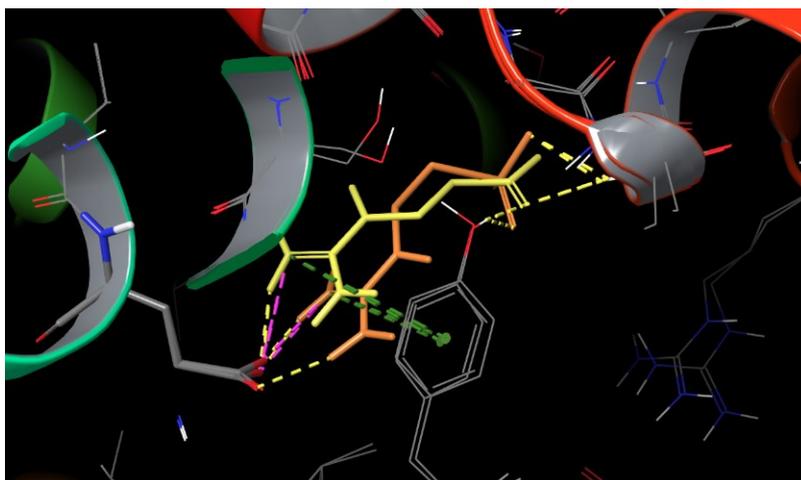


Fig. 36: Folded Guanidinopropionic acid and Guanidinoethanesulfonic acid

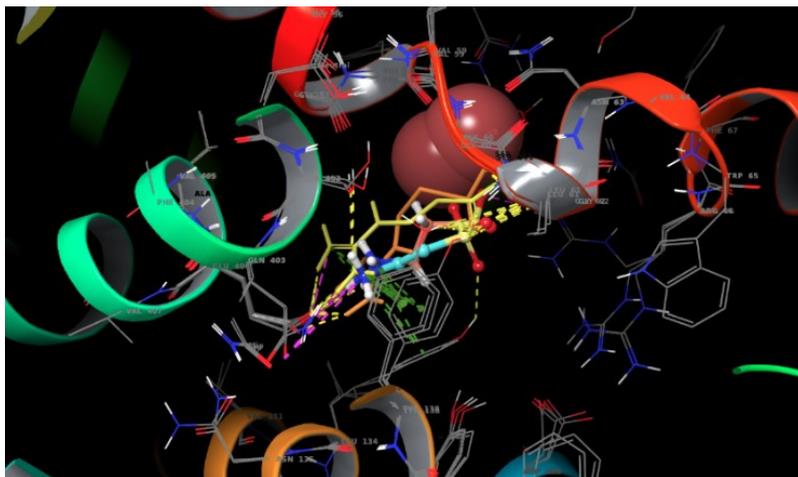


Fig. 37: Two ligands which hold a Guanidino group are clustered with Hypotaurine and Taurine to display that the Guanidino carriers are folded

What the table also depicts, is that the Ligand strain energy of Creatine is very high. The reason for this might be that the Guanidino group might be too high for the conformation that is available in the transporter. Therefore, the ligand is very constrained and precisely what can be derived from the Ligand strain energy.

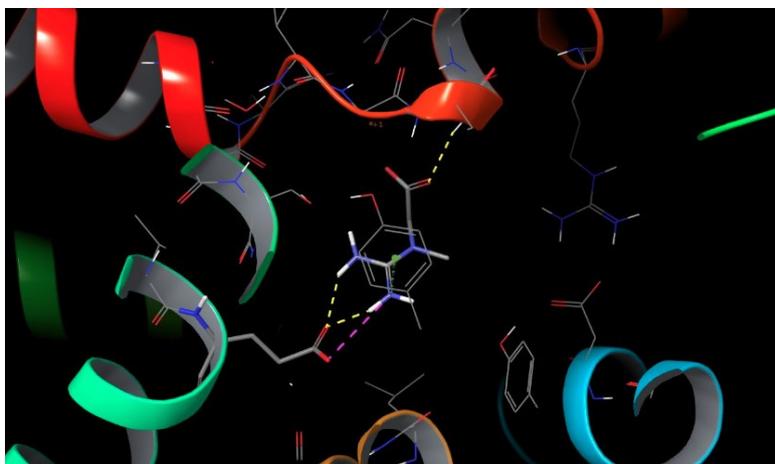


Fig. 38: Interaction between Creatine and other residues in the pocket

Looking at Hypotaurine in this context, an interaction with Arg66 becomes evident, which generally would not occur if carboxylate is available. This is the reason why the MM GBSA of Hypotaurine is better.

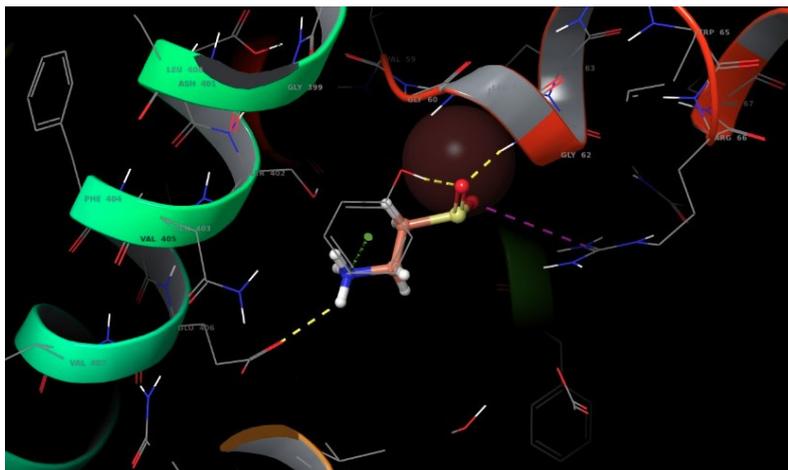


Fig. 39: Interaction between Hypotaurine and Arg66

It was expected to find similar values of the MM GBSA binding energy and the Ligand strain energy for both Hypotaurine and Taurine because the two ligands demonstrate very similar values of IC50 activity and the effect of inhibition. Though, a very remarkable difference between the MM GBSA binding energy of both ligands can be observed in the above table. While Hypotaurine shows a value of -35,61, Taurine is at -16,04.

Besides, an interaction between Hypotaurine and Arg66 is observable; however, such an interaction would not occur in the presence of a Carboxylate group. Yet, the interaction could be taken as an example to explain the good MM GBSA binding energy of Hypotaurine.

Taking into account the same binding pose, it is evident that Taurine does not interact with Arg66 as it occurs with Hypotaurine, although both ligands demonstrate an identical carbon length. This can be justified by that the Sulfo group of Hypotaurine provides the ligand with a larger surface that eventually enables Hypotaurine to reach Arg66. This process can be observed in the following two pictures.

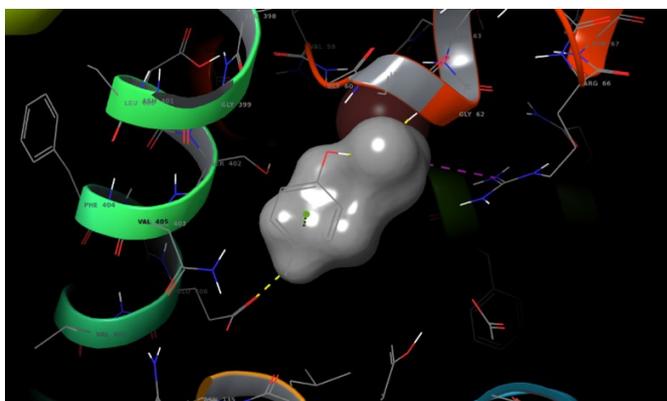


Fig. 40: Surface of Hypotaurine

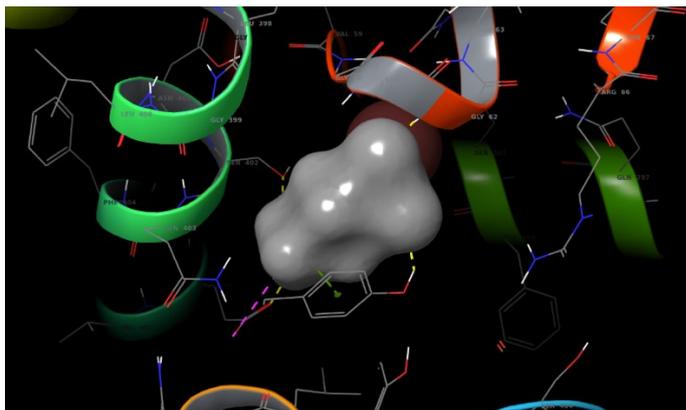


Fig. 41: Surface of Taurine

## 5.0 Outlook and Conclusion

By analyzing and comparing the above-described ligands, a pattern concerning the three ligands carrying a Guanidino group emerged. All demonstrate high Ligand strain energy. Although a better percentage of inhibition characterizes Taurine, no result could be determined for it, since its MM GBSA binding energy is higher than it should initially be. As well, a considerable difference between Taurine and L- $\alpha$ -Alanin in respect of their MM GBSA binding energy could be identified. Yet, there is room for further exploration to gain knowledge of the still unknown.

In this sense, there is room for further analyses. Yet, the restricted access to the conformations of the transporter posed an obstacle in terms of finding more suitable methods which might have delivered a more comprehensive range of data on the particular transporter.

Carrying out this MM GBSA analysis primarily aimed to set the building blocks for a better understanding of the Taurine transporter. Based on the first elements that this research delivers; further analyses are required to provide more detailed knowledge of the complex characteristics and functions of the Taurine transporter. Due to a time restriction of five months, a more in-depth discussion of the topic was not feasible, yet, it is with confidence argued, that the delivered findings in the course of this exploration provide essential elements which can guide further analyses.

## 6.0 Bibliography

- Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., & Iwata, S. (2003). "Structure and mechanism of the lactose permease of *Escherichia coli*". *Science*. 301 (5633): 610–5.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Anderson CM, Howard A, Walters JR, Ganapathy V, Thwaites DT (2009). Taurine uptake across the human intestinal brush-border membrane is via two transporters: H<sup>+</sup>-coupled PAT1 (SLC36A1) and Na<sup>+</sup>- and Cl<sup>-</sup>-dependent TauT (SLC6A6). *J Physiol*; 587:731-744
- Baker, D. & Sali, A. (2001). Protein structure prediction and structural genomics. *Science*; 294:93–96
- Baldwin, R., 2020. *Learn To Program Using Python: Writing And Using Scripts - Developer.Com*. [online] Developer.com. Available at: <<https://www.developer.com/lang/other/article.php/626311/learn-to-program-using-python-writing-and-using-scripts.htm>> [Accessed 7 April 2020].
- Baloria, U., Akhoun, B.A., Gupta, S.K., Sharma, S. & Verma, V., (2012). "In silico proteomic characterization of human epidermal growth factor receptor 2 (HER-2) for the mapping of high affinity antigenic determinants against breast cancer". *Amino Acids*. 42 (4): 1349–60. doi:10.1007/s00726-010-0830-x. PMID 21229277.
- Barker, W., 2000. The Protein Information Resource (PIR). *Nucleic Acids Research*, 28(1), pp.41-44.
- BLAST: Basic Local Alignment Search Tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- Bröer, S., & Gether, U. (2012). The solute carrier 6 family of transporters. *British journal of pharmacology*, 167(2), 256-278.
- Bully, S. & Shen, W., (2010). Reciprocal regulation between taurine and glutamate response via Ca<sup>2+</sup>-dependent pathways in retinal third-order neurons. *J Biomed Sci*. 2010 Aug 24; 17 Suppl 1():S5. [PubMed] [Ref list]

- Canul-Tec, J. C., Assal, R., Cirri, E., Legrand, P., Brier, S., Chamot-Rooke, J., et al. (2017). Structure and allosteric inhibition of excitatory amino acid transporter 1. *Nature* 544, 446–451. doi: 10.1038/nature22064
- César-Razquin, A., Snijder, B., Frappier-Brinton, T., Isserlin, R., Gyimesi, G., Bai, X., et al. (2015). A call for systematic research on solute carriers. *Cell* 162, 478–487. doi: 10.1016/j.cell.2015.07.022
- Centeno, N. B., Planas-Iglesias, J., & Oliva, B. (2005). Comparative modelling of protein structure and its impact on microbial cell factories. *Microbial Cell Factories*, 4(1), 20.
- Chakraborty, P., & Di Cera, E. (2017). Induced fit is a special case of conformational selection. *Biochemistry*, 56(22), 2853-2859.
- Chen, J.H & Brooks, C.L. (2007). Can molecular dynamics simulations provide high-resolution refinement of protein structure? *Proteins*; 67:922–930. Wiley Online Library CAS PubMed Web of Science®Google Scholar
- Colas, C. et al. (2015). Ligand Discovery for the Alanine-Serine-Cysteine Transporter (ASCT2, SLC1A5) from Homology Modelling and Virtual Screening. *PLOS Comput. Biol.* 11, e1004477.
- Colas, C., Masuda, M., Sugio, K., Miyauchi, S., Hu, Y., Smith, D. E., et al. (2017). Chemical modulation of the human oligopeptide transporter 1, hPepT1. *Mol. Pharm.* 14, 4685–4693. doi: 10.1021/acs.molpharmaceut.7b00775
- Colas, C., Ung, P. M. U., & Schlessinger, A. (2016). SLC transporters: structure, function, and drug discovery. *Medchemcomm*, 7(6), 1069-1081.
- Console, L., Scalise, M., Tarmakova, Z., Coe, I. R., and Indiveri, C. (2015). N-linked glycosylation of human SLC1A5 (ASCT2) transporter is critical for trafficking to membrane. *Biochim.Biophys.Acta* 1853,1636–1645.doi: 10.1016/j.bbamcr.2015.03.017
- Dayan, O., Nagarajan, A., Shah, R., Ben-Yona, A., Forrest, L.R. and Kanner, B.I., 2017. An Extra Amino Acid Residue in Transmembrane Domain 10 of the  $\gamma$ -Aminobutyric Acid (GABA) Transporter GAT-1 Is Required for Efficient Ion-coupled Transport. *Journal of Biological Chemistry*, 292(13), pp.5418-5428.

- DSSP. Swift.cmbi.umcn.nl. (2020). Retrieved 1 April 2020, from [https://swift.cmbi.umcn.nl/gv/dssp/DSSP\\_3.html](https://swift.cmbi.umcn.nl/gv/dssp/DSSP_3.html).
- Dominy Jr, J., Thinschmidt, J. S., Peris, J., Dawson Jr, R., & Papke, R. L. (2004). Taurine-induced long-lasting potentiation in the rat hippocampus shows a partial dissociation from total hippocampal taurine content and independence from activation of known taurine transporters. *Journal of neurochemistry*, *89*(5), 1195-1205.
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* *5*:113. doi: 10.1186/1471-2105-5-113
- Eramian, D., Eswar, N., Shen, M. Y., and Sali, A. (2008). How well can the accuracy of comparative protein structure models be predicted? *Protein Sci.* *17*, pp. 1881–1893. doi: 10.1110/ps.036061.108
- Feinstein, W.P. and Brylinski, M., 2015. Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets. *Journal of cheminformatics*, *7*(1), pp.1-10.
- Ferrari, A. M., Wei, B. Q., Costantino, L., & Shoichet, B. K. (2004). Soft docking and multiple receptor conformations in virtual screening. *Journal of medicinal chemistry*, *47*(21), 5076-5084.
- Fiser, A. & Sali, A., (2003). "ModLoop: automated modelling of loops in protein structures". *Bioinformatics*. *19* (18): 2500–1. doi:10.1093/bioinformatics/btg362. PMID 14668246.
- Forrest, L. R., Tang, C. L., and Honig, B. (2006). On the accuracy of homology modelling and sequence alignment methods applied to membrane proteins. *Biophys. J.* *91*, 508–517. doi: 10.1529/biophysj.106.082313
- Forrest, L.R & Rudnick, G. (2009). *Physiology* (Bethesda). *24*:377–386.
- Fuchs, B. C., Finger, R. E., Onan, M. C., and Bode, B. P. (2007). ASCT2 silencing regulates mammalian target-of-rapamycin growth and survival signaling in human hepatoma cells. *Am. J. Physiol. Cell Physiol.* *293*, C55–63. doi: 10.1152/ajpcell.00330.2006
- Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids, Res.* *45*, D945–D954. doi: 10.1093/nar/gkw1074

- Giorgetti, A., Raimondo, D., Miele, A.E., & Tramontano, A. (2005). Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics*; 21(Suppl 2):ii72–76
- Han, X., Patters, A. B., Jones, D. P., Zelikovic, I., & Chesney, R. W. (2006). The taurine transporter: mechanisms of regulation. *Acta physiologica*, 187(1-2), 61-73.
- Han, X., Budreau, A. M., & Chesney, R. W. (1999). Ser-322 Is a Critical Site for PKC Regulation of the MDCKCell Taurine Transporter (pNCT). *Journal of the American Society of Nephrology*, 10(9), 1874-1879.
- HHpred | Bioinformatics Toolkit. <https://toolkit.tuebingen.mpg.de/tools/hhpred#>.
- Hoffmann EK, Lambert IH, Pedersen SF (2009). Physiology of cell volume regulation in vertebrates. *Physiol Rev*;89:193- 277
- Huang, Y., Lemieux, M.J., Song, J., Auer, M. & Wang, D.N. (2003). "Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli". *Science*. 301 (5633): 616–20.
- Ishitani, R., Terada, T. & Shimizu, K. (2008). Refinement of comparative models of protein structure by using multicanonical molecular dynamics simulations. *Mol Simulat*; 34:327–336.
- Jardetzky O. (2016). Simple allosteric model for membrane pumps. *Nature*; 211:969–970.
- Kickinger, S., Hellsberg, E., Schousboe, A., Ecker, G. F., & Wellendorph, P. (2019). Structural and molecular aspects of betaine-GABA transporter 1 (BGT1) and its relation to brain function. *Neuropharmacology*.
- Koshland, D. E. (1960). The active site and enzyme action. *Adv. Enzymol. Relat. Subj. Biochem.* 22, 45–97.
- Knopf, F., Hammond, C., Chekuru, A., Kurth, T., Hans, S., Weber, C. W., ... & Weidinger, G. (2011). Bone regenerates via dedifferentiation of osteoblasts in the zebrafish fin. *Developmental cell*, 20(5), 713-724.
- Lambert IH, Hoffmann EK, & Pedersen S.F (2008): Cell volume regulation: physiology and pathophysiology. *Acta Physiol (Oxf)*.194:255-282

- Lambert, I. H., Kristensen, D. M., Holm, J. B., & Mortensen, O. H. (2015). The physiological role of taurine—from organism to organelle. *Acta Physiologica*, 213(1), 191-212.
- Lane, A. N., and Fan, T. W. (2015). Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids, Res.* 43, 2466–2485. doi: 10.1093/nar/gkv047
- Lee, M.R., Tsai, J., Baker, D. & Kollman, P.A. (2001). Molecular dynamics in the endgame of protein structure prediction. *J Mol Biol*; 313:417–430. Crossref CAS PubMed Web of Science®Google Scholar
- Lee, Y., Nishizawa, T., Yamashita, K., Ishitani, R., & Nureki, O. (2015). Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nat Commun*; 6:6112.
- Li, Y., Wang, F., Lee, J. A., & Gao, F. B. (2006). MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes & development*, 20(20), 2793-2805.
- Lourenco, R., & Camilo, M. E. (2002). Taurine: a conditionally essential amino acid in humans? An overview in health and disease. *Nutr Hosp*, 17(6), 262-270.
- McGann, M. (2011). FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 51, 578–596. doi: 10.1021/ci100436p
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., et al. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids, Res.* 41, W597–W600. doi: 10.1093/nar/gkt376
- Meng, X.Y., Zhang, H.X., Mezei, M. & Cui, M. (2011). Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* 7, 146–157.
- Muhammed, M. T. & Aki-Yalcin, E. (2019). Homology modelling in drug discovery: Overview, current applications, and future perspectives. *Chem. Biol. Drug Des.* 93, 12–20 (2019).
- Mulakala, C. & Viswanadhan, V. N. (2013). Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies? *J. Mol. Graph. Model.* 46, 41–51.

- Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55, 6582–6594. doi: 10.1021/jm300687e
- Pei, J., and Grishin, N. V. (2014). PROMALS3D: Multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.* 1079, 263–271. doi: 10.1007/978-1-62703-646-7\_17
- Pei, J., Kim, B.-H. & Grishin, N. V. (2008). PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36, 2295–2300.
- Perraud, A.L., Takanishi, C.L., Shen, B.,... & Scharenberg, A.M., (2005). "Accumulation of free ADP-ribose from mitochondria mediates oxidative stress-induced gating of TRPM2 cation channels". *The Journal of Biological Chemistry.* 280 (7): 6138–48. doi:10.1074/jbc.M411446200. PMID 15561722.
- Pop-Busui, R., Sullivan, K. A., Van Huysen, C., Bayer, L., Cao, X., Towns, R., & Stevens, M. J. (2001). Depletion of taurine in experimental diabetic neuropathy: implications for nerve metabolic, vascular, and functional deficits. *Experimental neurology*, 168(2), 259-272.
- PROMALS3D Documentation  
[http://prodata.swmed.edu/promals3d/info/promals3d\\_help.html](http://prodata.swmed.edu/promals3d/info/promals3d_help.html)
- Ray, G. and Cook, J., 2005. Molecular modelling of heme proteins using MOE: Bio-inorganic and structure-function activity for undergraduates. *Biochemistry and Molecular Biology Education*, 33(3), pp.194-201.
- Ripps, H., & Shen, W. (2012). Taurine: a “very essential” amino acid. *Molecular vision*, 18, 2673.
- Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Šali, A., Potterton, L., Yuan, F., van Vlijmen, H., & Karplus, M. (1995). Evaluation of comparative protein modelling by MODELLER. *Proteins: Structure, Function, and Bioinformatics*, 23(3), 318-326.

- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design*, 27(3), 221-234.
- Schaffer SW, Azuma J, Mozaffari M (2009). Role of antioxidant activity of Taurine in diabetes. *Can J Physiol Pharmacol*;87:91-99
- Schaeffer, D., Reis, F. P., Johnson, S. J., Arraiano, C. M., & van Hoof, A. (2012). The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. *Nucleic acids research*, 40(18), 9298-9307.
- Schlessinger, A., Wittwer, M. B., Dahlin, A., Khuri, N., Bonomi, M., Fan, H., et al. (2012). High selectivity of the gamma-aminobutyric acid transporter 2 (GAT-2, SLC6A13) revealed by a structure-based approach. *J. Biol. Chem.* 287, 37745–37756. doi: 10.1074/jbc.M112.388157
- Schrödinger Release 2019-1: Glide, Schrödinger, LLC, New York, NY, 2019.
- Schrödinger S. (2009). Induced Fit Docking protocol; Glide version 5.5, Schrödinger, LLC, New York, NY, 2009; Prime version 2.1, Schrödinger, LLC, New York, NY, 2009.
- Schrödinger, S., (2015). Can I relate MM-GBSA energies to binding affinity? | Schrödinger. <https://www.schrodinger.com/kb/1647>
- Scopelliti, A. J., Font, J., Vandenberg, R. J., Boudker, O., and Ryan, R. M. (2018). Structural characterization reveals insights into substrate recognition by the glutamine transporter ASCT2/SLC1A5. *Nat. Commun.* 9, 1–12. doi: 10.1038/s41467-017-02444-w
- Seeliger, D. and de Groot, B.L., 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of computer-aided molecular design*, 24(5), pp.417-422.
- Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M (2007) *J Comput Aided Mol Des* 21(12):681
- Shen, M. Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507–2524. doi: 10.1110/ps.062416606
- Singh, K., Tanui, R., Gameiro, A., Eisenberg, G., Colas, C., Schlessinger, A., et al. (2017). Structure-activity relationships of benzylproline-derived inhibitors of the glutamine

- transporterASCT2. *Bioorg.Med.Chem.Lett.* 27,398–402. doi: 10.1016/j.bmcl.2016.12.063
- Singh, S.K., Piscitelli, C.L. & Yamashita, A. (2008). *Science*. 322:1655–1661.
- Stenger, B., Thayananthan, A., Torr, P.H. and Cipolla, R., 2006. Model-based hand tracking using a hierarchical bayesian filter. *IEEE transactions on pattern analysis and machine intelligence*, 28(9), pp.1372-1384.
- Tastesen HS, Holm JB, Moller J, Poulsen KA, Moller C, Sturup S, Hoffmann EK, Lambert IH, (2010). Pinpointing differences in cisplatin-induced apoptosis in adherent and non-adherent cancer cells. *Cell Physiol Biochem*;26:809-820
- Uniprot Consortium, T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 46:2699. doi: 10.1093/nar/gky092
- Venclovas, Č., & Margelevičius, M. (2005). Comparative modelling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment. *Proteins: Structure, Function, and Bioinformatics*, 61(S7), 99-105.
- Vesterman, B., Golender, V., Golender, L., & Fuchs, B. (1996). Conformer clustering algorithm and its application for crown-type macrocycles. *Journal of Molecular Structure: THEOCHEM*, 368, 145-151.
- Vogensen, S.B., Jørgensen, L., Madsen, K.K., Jurik, A., Borkar, N., Rosatelli, E., Nielsen, B., Ecker, G.F., Schousboe, A. & Clausen, R.P., (2015). Structure activity relationship of selective GABA uptake inhibitors. *Bioorganic & medicinal chemistry*, 23(10), pp.2480-2488.
- Voss J.W., Pedersen SF, Christensen ST, & Lambert, I.H (2004). Regulation of the expression and subcellular localization of the taurine transporter TauT in mouse NIH3T3 fibroblasts. *Eur J Biochem*; 271:4646-4658
- Watts, K. S., Dalal, P., Murphy, R. B., Sherman, W., Friesner, R. A., & Shelley, J. C. (2010). ConfGen: a conformational search method for efficient generation of bioactive Conformere. *Journal of chemical information and modeling*, 50(4), 534-546.
- Weaver, T. M. (2000). The  $\pi$ -helix translates structure into function. *Protein Science*, 9(1), 201-206.

- Webb, B. & Sali, A. (2016). Comparative Protein Structure Modelling Using MODELLER. *Curr. Protoc. Bioinforma.* Ed. Board Andreas Baxevanis A1 54, 5.6.1-5.6.37.
- Webb, B., and Sali, A. (2017). Protein structure modelling with MODELER. *Methods Mol. Biol.* 1645, 39–54.
- Xu, Y. J., Arneja, A. S., Tappia, P. S., & Dhalla, N. S. (2008). The potential health benefits of taurine in cardiovascular disease. *Experimental & Clinical Cardiology*, 13(2), 57.
- Yahara, T., Tachikawa, M., Akanuma, S. I., Kubo, Y., & Hosoya, K. I. (2014). Amino acid residues involved in the substrate specificity of TauT/SLC6A6 for taurine and  $\gamma$ -aminobutyric acid. *Biological and Pharmaceutical Bulletin*, 37(5), 817-825.
- Yamashita, A., Singh, S.K., Kawate, T., Jin, Y. & Gouaux E. (2005). *Nature*. 437:215–223.
- Zhang, Y. (2009). Protein structure prediction: when is it useful? *Curr Opin Struct Biol*; 19:145–155.

## 7.0 Appendix

### 7.1 Table of Figures

<a href="#">Fig. 1: Showing four divisions of taurine transporters</a> .....	10
<a href="#">Fig. 2: (Top): Secondary active Taurine Transporters (Bottom): Release paths for volume-sensitive taurine</a> .....	10
<a href="#">Fig. 3: Showing actions of PKC activation on Taurine uptake upon mutagenesis site-direction</a> .....	11
<a href="#">Fig. 4: 2 Na ions: 1 Cl ions: 1 taurine transporter model system across a rat's renal cells</a> .....	12
<a href="#">Fig. 5: Showing predicted LeuT-like fold</a> .....	13
<a href="#">Fig. 6: Mechanisms of Taurine Transport</a> .....	15
<a href="#">Fig.7: Steps of Homology Modelling</a> .....	19
<a href="#">Fig.8: SLC6 Family (FASTA SEQUENCE) plus 3 structural hSERT (SLC6A4), secondly --&gt; 4xP9, PDB code for DAT (SLC6A3) and thirdly --&gt; 2A65, PDB code for LeuT (SLC6A2)</a> .....	20
<a href="#">Fig.9: Conserved TM10 and Glycine in GABA-members</a> .....	21
<a href="#">Fig.10: The three Options as visually examined in MOE</a> .....	25
<a href="#">Fig. 11: Showing the selected Top 3 modelled Loops regarding their dope score</a> .....	25
<a href="#">Fig. 12: Showing the selected Top 3 modelled Loops in the TM10</a> .....	26
<a href="#">Fig. 13: Showing the blue selected Top 3 modelled Loops regarding their dope score</a> .....	26
<a href="#">Fig. 14: Showing the selected Top 3 modelled Loops in the TM10</a> .....	27
<a href="#">Fig.15: Showing the Top3 modelled Loops in the TM10 that were not selected</a> .....	28
<a href="#">Fig.16: Showing the Top3 modelled Loops in the TM10 that were selected</a> .....	29
<a href="#">Fig.17: Showing the Top3 modelled Loops selected in blue</a> .....	29
<a href="#">Fig. 18: Refined loops Top3 chosen regarding their dope score due to their insignificant difference from each other in the pi-helix region</a> .....	30
<a href="#">Fig.19: Showing the Top3 modelled Loops selected in dark blue</a> .....	30
<a href="#">Fig.20: Showing the Top3 modelled Loops aligned with each other</a> .....	31
<a href="#">Fig.21: Showing the Top3 modelled Loops selected</a> .....	31

<a href="#">Fig. 22: Top3 modelled Loops appearing aligned and superimposed</a>	32
<a href="#">Fig.23: Showing Top3 loops chosen based on their dope score</a>	33
<a href="#">Fig.24: Showing colored orange the Top2 models of Option1 and colored turquoise of Option 3 The template is colored yellow template</a>	36
<a href="#">Fig. 25: Showing results of the dope scores after evaluation</a>	36
<a href="#">Fig.26: Showing the <b>final homology model 176</b> with the dope score <b>-1.10235</b> colored turquoise. The template is colored yellow</a>	37
<a href="#">Fig. 27: Ligands with known IC50 value used for the LigPrep</a>	40
<a href="#">Fig. 28: Ligands with known percentage of inhibition used for the LigPrep</a>	40
<a href="#">Fig. 29: KNIME workflow</a>	45
<a href="#">Fig. 30: Cutout of the KNIME workflow to show the two most popular clusters</a>	46
<a href="#">Fig. 31: To show which Ligand is in which cluster</a>	47
<a href="#">Fig. 32: Cutout of the MM GBSA Panel in Maestro</a>	47
<a href="#">Fig. 33: Ligands with a known IC50 value</a>	49
<a href="#">Fig. 34: Ligands with a known percentage of inhibition</a>	49
<a href="#">Fig. 35: Folded Guanidinopropionic acid and Guanidinoethanesulfonic acid</a>	50
<a href="#">Fig. 36: Two ligands which hold a Guanidino group are clustered with Hypotaurine and Taurine to display that the Guanidino carriers are folded</a>	50
<a href="#">Fig. 37: Interaction between Creatine and othe reisdues in the pocket</a>	51
<a href="#">Fig. 38: Interaction between Hypotaurine and Arg66</a>	51
<a href="#">Fig. 39: Surface of Hypotaurine</a>	52
<a href="#">Fig. 40: Surface of Taurine</a>	52

## 7.2 Abstract

The GABA transporter family is part of the SLC6A family and encloses also the taurine transporter (SLC6A6). The Taurine transporter's function is to secure the supply of taurine to the brain, liver, and the retina. Hence, the effect of the aminosulphonic acid bears on multiple organs: taurine affects not only the development of the retina, the central nervous system, and

the cardiac muscle, but also the development of the skeletal muscles. In addition, it serves as an antioxidant and accordingly assists in removing toxins from the body. Due to its ability to reduce the secretion of lipids, the susceptibility to suffer from diseases such as arteriosclerosis and coronary heart diseases can be minimized. In this context, the essential role of the taurine transporter is marked by the transport of the acid to the respective location. The aim of this thesis was to gain insights into the underlying mechanisms of the interaction of taurine with its transporter by means of protein homology modelling and ligand docking.

In a first attempt, an intensive analysis was carried out to explore the definition of the transmembrane domains in conjunction with the enclosed helices. To establish a connection, theoretical knowledge and the results based on prediction tools and sequence alignment were combined. Finally, a template was selected taking into account not only diverse sequence alignments of the provided crystal structures, but also the PDB codes for LeuT (SA65), dDAT (4xP9), and hSERT (SLC6A4). The latter was then selected for further steps of modelling because it demonstrated a high sequence identity. The MODELLER software was used for the establishment of 3D models, and the final selection was based on the examination of Ramachandran plots by the aid of MOE. The validation process is marked by an analysis of electrostatic potentials and transmembrane helices and their respective polar amino acid residues. Finally, in the course of a docking study, active ligands were docked on the created 3D model. Subsequently, the arising poses were clustered into groups – a process that allows conclusions on binding poses and amino acids, which are involved in the binding mechanism. Unfortunately, no clear picture could be obtained which allows to link the activity differences of the ligands docked to their docking scores.

### 7.3 Zusammenfassung

Die GABA Transporter Familie ist Teil der SLC6A Familie und umschließt ebenso den Taurin Transporter SLC6A6. Die Funktion des TauT besteht darin, die Taurinzufuhr zum Gehirn, zur Leber und zur Retina sicherzustellen. Der Effekt der Aminosulfonsäure erstreckt sich auf mehrere Organe: Taurin beeinflusst sowohl die Entwicklung der Netzhaut, des Zentralnervensystems und die des Herzmuskels als auch die der Skelettmuskulatur. Darüber hinaus dient Taurin als Antioxidans und filtert dementsprechend Toxine aus dem Körper. Durch seine Fähigkeit die Absonderung von Lipiden zu verringern, kann mithilfe dieser Aminosulfonsäure die Anfälligkeit an Arteriosklerose sowie an koronaren Herzerkrankungen zu erkranken, minimiert werden. Die essenzielle Rolle des Taurin Transporters besteht in

diesem Sinne darin, die Säure zur jeweiligen Stelle zu befördern. Das Ziel dieser Dissertation ist es daher, anhand von Homology Modelling und Ligand Docking Kenntnisse über die Interaktionsmechanismen des Taurins mit seinem Transporter zu erlangen.

In einem ersten Schritt wurde eine intensive Analyse unternommen, welche die Definition der transmembranen Domänen in Verbindung mit den umschlossenen Helices, anstrebte. Um einen Zusammenhang herstellen zu können, wurden hierfür sowohl Theorie als auch Resultate basierend auf Vorhersagertools und Sequenzalignments miteinander verknüpft. Das Template wurde schlussendlich nicht nur auf Basis verschiedener Sequenzalignments der gebotenen Kristallstrukturen der SLC Anhänger, sondern auch auf Basis der PDB codes für LeuT (2A65), dDAT (4xP9) und des hSERT (SLC6A4), selektiert. Letzteres wurde für das weitere Modelling herangezogen, da hier eine sehr hohe Sequenzidentität festgestellt werden konnte. Die MODELLER Software wurde für die Erstellung der 3D Modelle eingesetzt, während die Untersuchung von Ramachandran-Plots mithilfe von MOE die letzte Selektion gewährleistete. Der Validierungsprozess charakterisiert sich durch eine Analyse elektrostatischer Potenziale und transmembraner Helices und ihren polaren Aminosäureresten. Im Zuge einer Dockingstudie wurden schlussendlich aktive Liganden an das erstellte 3D Modell gedockt. Schlussendlich wurden daraus hervorgegangen Posen in Gruppen geclustert- ein Prozess, welcher erlaubt, elementare Schlüsse über Bindungsposen und Aminosäuren, welche im Bindungsmechanismus involviert sind, zu ziehen. Unglücklicherweise konnte kein deutliches Bild gewonnen werden, welches erlaubt, die Leistungsunterschiede der angedockten Liganden mit ihren Andockwerten in Zusammenhang zu bringen.

#### 7.4 List of Abbreviations

**AAFCO-** Association of American Feed Control Officials

**AGEs-** Advanced Glycation End-products

**AMBER-** Assisted Model Building with Energy Refinement

**BLAST-** Basic Local Alignment Search Tool

**CHARMM-** Chemistry at Harvard Macromolecular Mechanics

**CNS-** Central Nervous System

**DAT-** Dopamine Active Transporter

**DOPE**-Discrete Optimized Protein Energy

**DSSP**- Dialogue System for Structured Programming

**EAAT**- Excitatory Amino Acid Transporter

**EATC**- Ehrlich Ascites Tumour Cells

**GABA**- Gamma Aminobutyric Acid

**GAT**- Glycine Alanine Transporter

**GES**- Guanidinoethyl Sulfonate

**HHPerd**- Remote Protein Homology detection and structure Prediction

**hNTCP**- Human Sodium Taurocholate Cotransporting Polypeptide

**hSERT**- Human Serotonin Transporter

**IF**- Induced Fit

**LDL**- Low-Density Lipoproteins

**LeuT**-Leucine Transporter

**MDS**- Molecular Dynamics Simulations

**MFS**- Major Facilitator Superfamily

**MOE**- Molecular Operating Environment

**MSA**- Multiple Sequence Alignment

**NSS**- Neurotransmitter Sodium-dependent Symporters

**NTT**- Neurotransmitter Transporters

**OPLs**- Optimized Potentials for Liquid Simulations

**PAT1**- Proton-Assisted Amino Acid Transporter

**PCK**- Protein Kinase C

**PDB**- Protein Data Bank

**PIR**- Protein Information Resource

**PMA**- Phorbol Myristate Acetate

**pNCT**- Madin—Darby Canine Kidney Cell Taurine Transporter

**PROMALS3D**- PROfile Multiple Alignment with Predicted Local Structures and 3D

**RMSD**- Root Mean Square Deviation

**RVD**- Regulatory Volume Decrease

**RVI**- Regulatory Volume Increase

**SLC Transporters**-Solute Carrier Transporters

**SLC1**- Solute Carrier Family 1

**SLC6**- Solute Carrier6

**SMATS**- SMILES Arbitrary Target Specification

**TauT**- Taurine Transporter

**TM**- Transmembrane

**TM10**-alanine-serine-cysteine transporter 2

**VLDL**- Very Low-Density Lipoproteins