Expression, Purification, and Small Molecule Binding to Cysteine Sulfinic Acid Decarboxylase

Aiza Ashfaq

Supervisors: Jan Haavik, Anne Baumann, Kazi Alam

A thesis submitted in partial fulfillment of the requirements for the degree of

Master in Biomedical Sciences

Department of Biomedicine

University of Bergen

Norway

May 2023

Acknowledgments

Table of contents

LIST OF ABBREVIATIONS

ADME	Absorption, distribution, metabolism, and excretion
ATF6	Activation transcription factor 6
ATP	Adenosine triphosphate
Bax	Bcl-2-associated X protein
Bcl-xl	B-cell lymphoma-extra large
CA	Cysteine acid
CDO	Cysteine dioxygenase
CNS	Central nervous system
CSA	Cysteine sulfinic acid
CSAD	Cysteine sulfinic acid decarboxylase
CV	Column volume
DMSO	Dimethyl sulfoxide
DSF	Differential Scanning Fluorimetry
DTT	Dithiothreitol
E. Coli	Escherichia coli
ECFP4	Extended connectivity fingerprint, up to four bonds
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FDA	Food and drug administration
GABA	γ-aminobutyric acid
GADL1	Glutamic acid decarboxylase-like protein 1
GTP	Guanosine-5'-triphosphate
His$_6$-tag	Hexahistidine tag
HPLC	High-performance liquid chromatography
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRE1	Inositol-requiring enzyme 1
K$_D$	Dissociation constant
lacO	Lactose operon

LB	Luria broth
MS	Mass Spectrometry
NADH	Nicotinamide adenine dinucleotide hydrogen
Ni-NTA	Nickel-nitrilotriacetic
Nrf2	Nuclear factor (erythroid-derived 2)-like 2
OD	Optical density
OPA	O-phthalaldehyde
PBS	Phosphate-buffered saline
PERK	PKR-like endoplasmic reticulum
PKR	Protein kinase R
PL	Pyridoxal
PLP	Pyridoxal 5'-phosphate
PM	Pyridoxamine
PMSF	Phenylmethylsulphonyl fluoride
PN	Pyridoxine
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TB	Terrific broth
TCEP	Tris-(2-carboxyethyl) phosphine
TEV	Tobacco Etch Virus
T_m	Melting temperature
UPR	Unfolded protein response

ABSTRACT

Cysteine Sulfinic acid decarboxylase (CSAD) is a PLP-dependent enzyme involved in the biosynthesis of taurine. It converts cysteine sulfinic acid (CSA) into hypotaurine. Hypotaurine is later oxidized into taurine, one of the mammals' most common free amino acids. Numerous vital physiological functions are associated with taurine, including cell volume regulation, Ca^{2+} homeostasis, inflammation modulation, mitochondrial translation, and reduction of apoptosis in the central nervous system. In addition, taurine is usually considered a modulator of neurotransmission as it activates an inhibitory neurotransmitter, GABA, and inhibits an excitatory neurotransmitter, glutamate. Although taurine is critical in various physiological processes, limited research exists on taurine biosynthesis, particularly on CSAD. The discovery of inhibitors and activators of CSAD may be used to alter taurine production in the human body, resulting in increased or decreased oxidative stress, mitochondrial function, inflammation, excitotoxicity, and apoptosis. All these mechanisms can impact neurotransmission and potentially be used as a therapeutic agent against neurological disorders like Parkinson's, Alzheimer's, and Huntington's disease. Therefore, this project aimed to optimize protein expression and purification of human and mouse CSAD, optimize the CSAD enzymatic activity assay, and investigate inhibitors and activators of CSAD. In this project, human and mouse CSAD was recombinantly expressed and purified. Due to low protein yield just mouse CSAD was used to screen a compound library of 344 small drug-like molecules using differential scanning fluorimetry (DSF). A total of 17 compounds with stabilizing or destabilizing effects on CSAD were further validated by a concentration-dependent assay. The hits were further tested by enzymatic activity assay. The results from the enzymatic activity of CSAD indicated that [5-Chloro-1-(cyclopropylmethyl)-6-oxo-1,6-dihydro-4-pyridazinyl](2-pyridinyl)acetonitrile (a compound from well F21 from Prestwick Original Molecules library) can be a weak inhibitor of CSAD activity, whereas, (3-Chloro-6-[3-(4-morpholinyl)propyl]amino-4-pyridazinyl)(1,3-dihydro-2H-benzimidazol-2-ylidene)acetonitrile (a compound from well D18) can be an activator or stabilizer of CSAD. Furthermore, molecular docking of these compounds against the active site of CSAD indicated that these compounds may not bind to the active site but possibly to another (allosteric) site of the CSAD.

1. INTRODUCTION

Cysteine sulfinic acid decarboxylase (CSAD) plays a crucial role in the biosynthesis of taurine, which serves multiple purposes in the human body. One of these is the regulation of neurotransmission, which is essential for a normal functioning nervous system. Neurotransmission controls muscle movement, sensory perception, thoughts, mood, emotions, and various cognitive and autonomic functions. Any abnormalities in neurotransmission can lead to neurological/behavioral disorders. Therefore, it is important to understand the mechanisms of neurotransmission to develop new therapies and treatments for such disorders.

1.1 Neurotransmitters

Neurotransmitters are chemicals that enable communication between neurons and target cells throughout the body [1]. Neurons are specialized cells for receiving information, conducting electrical impulses, and impacting other neurons or effector tissues [2]. Signals passing across neurons are the source of sensations, movements, thoughts, memories, and feelings. Typically, a neuron consists of the axon, cell body, and dendrites. Most of the molecules that a neuron requires to survive and operate are produced in the cell body. Dendrites spread from the cell body and have the purpose of receiving signals. Passing through the cell body, these signals move towards the axon and are transmitted to other neurons, muscle cells, or cells in other organs through the axon terminal [3]. The transmission of signals from one neuron to another cell occurs at specialized contact sites called synapses [4]. This allows the nervous system to perform tasks by releasing chemicals from presynaptic neural cells to postsynaptic receptors [1].
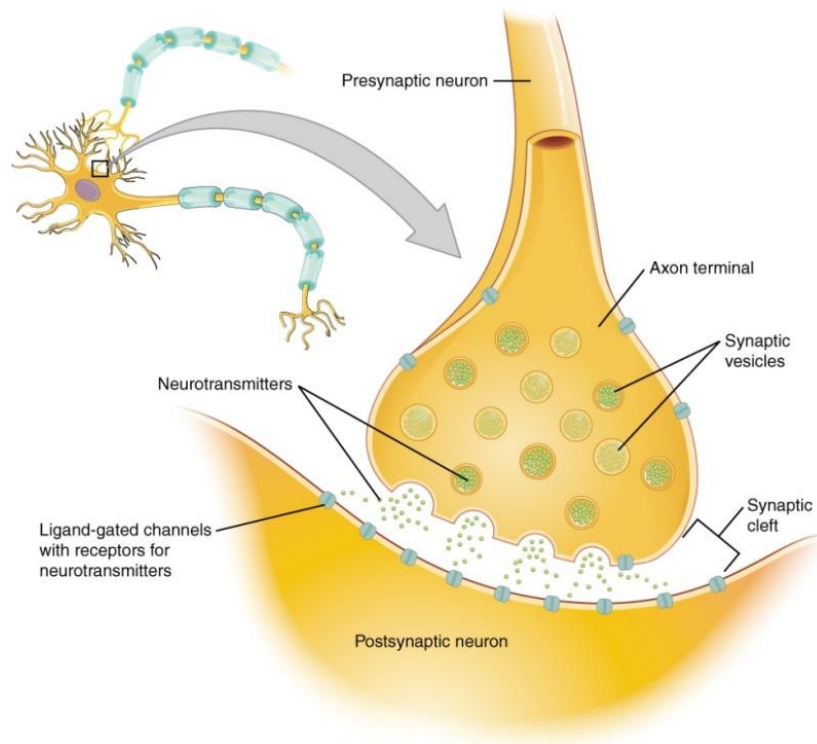
Figure 1: Transmission of neurotransmitters from the presynaptic neuron to the postsynaptic neuron. Figure taken from an open article by Christman et al., 2022 [5], licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

As Fig. 1 shows, classical neurotransmitters are stored in vesicles at presynaptic nerve terminals. The release of neurotransmitters occurs when the presynaptic nerve terminal is stimulated by an action potential [4]. An action potential depolarizes the presynaptic terminal, leading to calcium channel opening. Due to the subsequent calcium influx into the cell, synaptic vesicles undergo fusion with the presynaptic plasma membrane, and the neurotransmitters are released into the synaptic cleft. The transmitters travel across the synaptic cleft and bind to specific receptors on the postsynaptic membrane, which changes the postsynaptic cell electrochemically and biochemically [2].

Neurotransmitters can be excitatory or inhibitory depending on their release location, the receptors they bind to, and the ionic circumstances they encounter. Excitatory neurotransmitters activate cation channels, resulting in an inflow of sodium or calcium ions, depolarizing the postsynaptic membrane and causing an action potential. Glutamate and serotonin are examples of excitatory neurotransmitters. In contrast, inhibitory neurotransmitters activate anion channels, such as chloride or potassium channels, suppressing the firing of an action potential, such as γ-aminobutyric acid (GABA) and glycine. However, these ionotropic ligand-gated ion channels

discussed above are not responsible for all signaling in the nervous system. Most neurotransmitter molecules, including a wide range of neuropeptides, produced by nerve terminals bind to metabotropic receptors, also referred to as G-protein-coupled receptors [4, 6].

G-protein-coupled receptors utilize an intermediate called GTP-binding protein to regulate the intracellular processes. When a neurotransmitter binds with this receptor, it triggers activation of G-proteins. These G-proteins then dissociate from the receptor and activate ion channels either through direct interaction or through the binding with effector proteins. The postsynaptic actions generated by G-protein-coupled receptors can last from a few milliseconds to several hours or even days. In contrast, the postsynaptic actions generated by ionotropic receptors are usually rapid. β-adrenergic receptor is one of the well-known examples of G-protein-coupled receptor. The structure of this receptor is changed upon the binding of norepinephrine and epinephrine, which triggers structural changes in G-protein. This results in a chain of reactions that enables the G-protein to control the intracellular signaling cascades. Other examples of G-protein-coupled receptors includes metabotropic glutamate receptors, odorant receptors in the olfactory system, and muscarinic acetylcholine receptors [7].

1.2 Taurine

Taurine, also known as 2-amino-ethanesulfonic acid, is one of the most common free amino acids in mammals. It is found in high amounts in the liver, brain, spinal cord, leukocytes, heart and muscle cells, retina, and almost all other tissues. It was initially discovered and isolated from the ox (*Bos taurus*) bile, from where it received its name. Taurine differs from other amino acids as it has a sulfonate group instead of carboxyl group. It is considered a "non-essential" amino acid since it is not involved in protein synthesis [8]. In the liver, taurine is a component in formation of bile acids such as taurochenodeoxycholic acid and in conjugation of xenobiotics [9, 10]. Free taurine is involved in many important physiological functions, such as cell volume regulation, Ca^{2+} homeostasis, modulation of inflammation, mitochondrial translation, and reduction of apoptosis in the central nervous system [11]. In addition, it has been proposed that taurine increases membrane stability and stabilizes membrane proteins [12].

1.2.1 Biosynthesis of taurine

Taurine biosynthesis occurs mainly in the liver, kidney, and brain. The major pathway in mammalian tissues is through cysteine sulfinic acid (CSA). The first step includes cysteine

oxidation into cysteine sulfinic acid (CSA) by an enzyme cysteine dioxygenase (CDO). CSA is decarboxylated by cysteine sulfinic acid decarboxylase (CSAD) into hypotaurine, which is later oxidized either non-enzymatically or by hypotaurine dehydrogenase into taurine [13]. An alternative process involves oxidizing CSA to form cysteic acid (CA), which can then be decarboxylated by CSAD, resulting in the production of taurine [14].

1.2.2 Role of taurine in neurotransmission

Taurine acts as an inhibitory agent in the brain, inducing hyperpolarization and inhibiting neuronal activity. It is a close structural analog of γ-aminobutyric acid (GABA) [15]. GABA is the major inhibitory neurotransmitter in the nervous system. Previous studies suggest that a lack of GABA production in some brain areas causes hyperactivation of neurons and produces anxiety and other negative emotions [16]. Three postsynaptic receptors are present in GABAergic synapses: the ionotropic $GABA_A$ and $GABA_C$, that are permeable to chloride ions, and the metabotropic receptor $GABA_B$. Taurine interacts with both $GABA_A$ and $GABA_B$ receptors [17], activating them but less potently than GABA itself [15]. This leads to increased chloride conductance, which hyperpolarizes the cell [18]. Moreover, taurine inhibits the function of glutamate, which is an excitatory neurotransmitter. Excitatory glutamate stimulation may raise intracellular calcium levels, causing cell death [19]. Taurine may protect neurons from such glutamate excitotoxicity by preventing the depolarization of the membrane, which prevents the increase in glutamate-induced calcium influx [15].

Neurotransmitters can be classified as classical (conventional) transmitters or unconventional transmitters, such as the endocannabinoids or gaseous transmitters [20]. To be considered a classical neurotransmitter, a substance must meet the following five criteria:

- it must be present in the suspected neuron, preferably concentrated at the nerve terminal.
- it must be released upon stimulation in a calcium-dependent manner.
- it must elicit the appropriate physiological response.
- it must bind to a specific receptor.
- it must have an inactivation mechanism [21]

By this definition, taurine is usually considered a modulator of neurotransmission, but not as a classical neurotransmitter. From previous studies, it is verified that CSAD is localized in the neurons in rabbits' and rats retinas [21-23], and the release of taurine is calcium-dependent [22].

Taurine has been found to cause neuronal hyperpolarization, mostly likely by activating the chloride channel in the guinea pig cerebellum [21, 24] and in rat hippocampus [25]. Moreover, it is reported that taurine has specific receptors in the rabbit brain [21, 26, 27]. Thus, there is some evidence that taurine is a neurotransmitter but only in limited regions of the brain, and only in some developing animals. More research is required to validate taurine as a neurotransmitter [28].

1.2.3 Taurine and mitochondrial dysfunction

The mitochondria, the cell's "powerhouse" is important in energy production, biosynthesis, calcium signaling, cell apoptosis and production and sequestration of reactive oxygen species (ROS). Neurons mainly relay on mitochondria to fulfill their energy requirements [29]. Taurine modulates intracellular calcium homeostasis and prevents glutamate-induced excitotoxicity. Generally, glutamate leads to increase in intracellular calcium which can result in a collapse in mitochondrial membrane potential, causing a decrease in energy production and cell death. Taurine improves energy production through the maintenance of mitochondrial function, promoting neuronal survival [19].

In mitochondria, pH regulation is required to maintain normal mitochondrial metabolism in neurons and astrocyte for proper brain function. After glutamate is released from synapses, astrocytes and neurons take in glutamate that leads to change in pH in mitochondria matrix. Taurine has shown to preserve mitochondrial physiology by regulating the extreme pH fluctuations [17, 30, 31].

Mitochondrial tRNAs are conjugated with taurine-containing modified uridines [17, 32]. In the absence of taurine, the levels of specific mitochondria-encoded proteins decline, such as NADH-ubiquinone oxidoreductase chain 6. This leads to loss of the respiratory chain complex I subunit synthesis, causing reduced oxygen consumption, increased glycolysis and lactate production, and decline in ATP production [33, 34].

Taurine can act as an antioxidant in mitochondria to protect it from oxidative stress. However, taurine is probably not a free radical scavenger, as argued by Aruoma et al. [35], hence it is unknown how taurine prevents oxidative stress in the mitochondria [19]. Taurine itself cannot scavenge classical ROS. However, taurine can eliminate hypochlorous acid (HOCl), which is produced from hydrogen peroxide (H_2O_2) in the presence of chloride ions, forming N-chlorotaurine [19, 35]. The major function of N-chlorotaurine is to regulate inflammatory response. It has been

demonstrated that N-chlorotaurine activates nuclear factor (erythroid-derived 2)-like 2 (Nrf2), a transcription factor that regulated the transcription of several antioxidant genes and hence suppresses inflammation [19, 36].

1.2.4 Taurine and endoplasmic reticulum stress

Endoplasmic reticulum (ER) stress, caused by a buildup of misfolded proteins, disrupts neuronal signaling and causes neuronal cell death. The neuronal stress caused by these misfolded proteins activates unfolded protein response (UPR), which further activates three signaling pathways: PKR-like endoplasmic reticulum kinase (PERK), inositol-requiring enzyme 1 (IRE1), or activation transcription factor 6 (ATF6) to restore normal cellular function. The primary role of these components is to initiate signaling events to mitigate ER stress, but if the stress condition becomes severe, UPR system fails to restore normal cellular function, causing apoptosis [37]. Stroke causes increased ER stress, resulting in unfolded/misfolded protein accumulation which activates UPR system, and can cause neuronal damage. Taurine has shown to have a neuroprotective effect under ER stress inhibiting apoptosis [38].

1.2.5 Taurine and apoptosis

Taurine has been identified to inhibit apoptosis in response to a wide variety of toxic stimuli. The decrease in apoptotic rates and the improvement of neurological outcomes following brain ischemia were shown to be taurine's most notable neuroprotective benefits. One of the proposed mechanisms was the reduction of ER and mitochondrial stress [17].

Taurine can reduce mitochondria- dependent cell death. This may be accomplished by activating the antioxidant mechanism, preventing energy charge dampening, hindering the decrease of anti-apoptotic Bcl-xl and the increase of pro-apoptotic Bax, which prevents cytochrome C to release from mitochondria, and then inhibiting the activation of calpain and caspase-3 (Fig. 2). Taurine has also been shown to inhibit apoptosis by suppressing the UPR via inhibiting ATF6, PERK and IRE1 pathways [17].
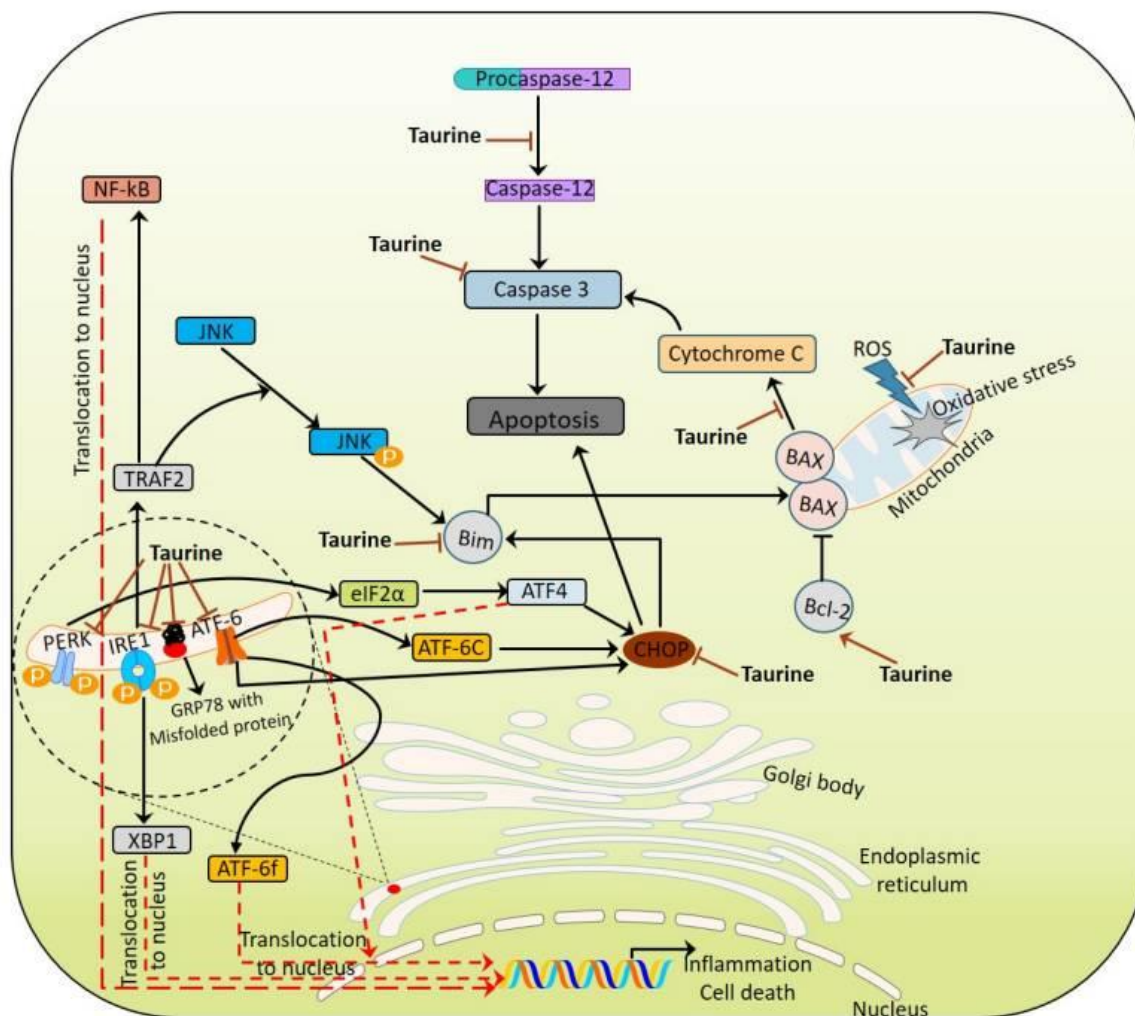
Figure 2: Schematic overview of taurine inhibiting mitochondrial- and endoplasmic stress-induced apoptosis. Figure taken from an article by Bhat et al., 2020 [39], licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

1.2.6 Taurine as a therapeutic agent

Neurodegenerative diseases like Alzheimer's disease, Parkinson's disease, and Huntington's disease share several common mechanisms, which includes oxidative stress, mitochondrial dysfunction, inflammatory changes, excitotoxicity, and apoptosis. As discussed above, taurine may help in reducing all of these mechanisms. Furthermore, it is reported that taurine levels in these diseases are reduced, suggesting that taurine could be a therapeutic agent against these conditions. However, additional research is required to prove taurine as a therapeutic agent [40].

1.3 Pyridoxal 5'-phosphate (PLP) dependent enzymes

Pyridoxal 5'-phosphate (PLP) is the active form of vitamin B6. Vitamin B6 refers to six chemically related pyridine compounds that all are water-soluble, including pyridoxine (PN), pyridoxamine

(PM) and pyridoxal (PL), in addition to their phosphorylated forms (PNP, PMP and PLP). The sole source of vitamin B6 is food, mainly meat, vegetables, and cereals [41].

PLP-dependent enzymes are involved in the metabolism of numerous neurotransmitters, including dopamine, serotonin, glycine, epinephrine, norepinephrine, D-serine, L-glutamate, gamma-aminobutyric acid and histamine [42]. PLP serves as a cofactor in enzyme reactions in several metabolic pathways. It can function as an enzyme-bound or free catalyst for transamination, decarboxylation, racemization, or elimination [43]. This versatility results from PLP's ability to form a covalent bond with the substrate, and serves as an electrophilic catalyst, stabilizing various carbonionic reaction intermediates [44].

PLP-dependent enzymes have gained interest due to their versatility as catalyst and widespread involvement in cellular processes. Many of these enzymes have been discovered as pharmacological targets in cancer and neurological diseases, further emphasizing their importance [45]. For instance, serine hydroxy methyltransferase has been discovered as a therapeutic target for cancer [46, 47], L-dihydroxyphenylalanine decarboxylase inhibitors are used in the treatment of Parkinson's disease [47, 48], and GABA aminotransferase inhibitors are used to treat epilepsy [47, 49].

1.4 Cysteine sulfinic acid decarboxylase

Cysteine sulfinic acid decarboxylase (CSAD) is a PLP-dependent enzyme involved in the biosynthesis of taurine, converting CSA to hypotaurine [14, 50]. It is considered a rate limiting enzyme in taurine production, which consequently also determines the demand for dietary taurine supplementation. Studies have shown that certain organisms, including cats have low levels of CSAD and CDO, resulting in a reliance on dietary taurine supplies. On the other hand, rodents have high CSAD levels, and taurine is not required in their diet [51, 52]. In mammals, CSAD can be found in liver, kidney, and brain as a dimer with subunit molecular weights around 43 to 55 kDa [50, 53].

Although CSAD has been reported to be found in neurons [54], it has been mainly been detected in astrocytes in hippocampus and cerebellum [55]. CDO and CSAD are found in different tissues, leading to the hypothesis that taurine biosynthesis begins in neurons but ends in astrocytes [53, 56]. However, another study showed that both murine neurons and astrocytes were capable of biosynthesizing taurine from cysteine, demonstrating that both cell types contain the complete

enzymatic machinery for taurine synthesis [57]. Still it is unknown which enzymes in either cell types are responsible for taurine production [53].

Generally, sequence alignment involves comparing two biological sequences (either nucleic acid or protein) to identify regions of similarity that may reveal structural, functional, and/or evolutionary connections. Sequence alignment of mouse and human CSAD shown in Fig. 3 indicates high sequence similarity. In Fig. 3, the stars (*) indicate that the amino acid found in both human and mouse CSAD is identical, while the colon (:) and period (.) denote different amino acids. Both variants of the sequence consist of 493 amino acids. After performing a pairwise sequence alignment using EMBOSS Needle ClustalW algorithm, it was found that the similarity between the CSAD sequence of mouse and human is 95.3% [58].

```
Human    MADSEALPSLAGDPVAVEALLRAVFGVVVDEAIQKGTSVSQKVCEWKEPEELKQLLDLEL    60
Mouse    MADSKPLRTLDGDPVAVEALLQDVFGIVVDEAILKGTSASEKVCEWKEPEELKQLLDLEL    60
         ****: * :* **********: ***.****** ****.*:******************

Human    RSQGESQKQILERCRAVIRYSVKTGHPRFFNQLFSGLDPHALAGRIITESLNTSQYTYEI    120
Mouse    QSQGESREQILERCRTVIHYSVKTGHPRFFNQLFSGLDPHALAGRIITESLNTSQYTYEI    120
         :*****::*******:**:*****************************************

Human    APVFVLMEEEVLRKLRALVGWSSGDGIFCPGGSISNMYAVNLARYQRYPDCKQRGLRTLP    180
Mouse    APVFVLMEEEVLKKLRALVGWNSGDGVFCPGGSISNMYAMNLARFQRYPDCKQRGLRALP    180
         ************:*******.****:*************:****:***********:**

Human    PLALFTSKECHYSIQKGAAFLGLGTDSVRVVKADERGKMVPEDLERQIGMAEAEGAVPFL    240
Mouse    PLALFTSKECHYSITKGAAFLGLGTDSVRVVKADERGRMIPEDLERQIILAEAEGSVPFL    240
         ************** *********************:*:********  :*****:****

Human    VSATSGTTVLGAFDPLEAIADVCQRHGLWLHVDAAWGGSVLLSQTHRHLLDGIQRADSVA    300
Mouse    VSATSGTTVLGAFDPLDAIADVCQRHGLWFHVDAAWGGSVLLSRTHRHLLDGIQRADSVA    300
         ***************:************:************:*****************

Human    WNPHKLLAAGLQCSALLLQDTSNLLKRCHGSQASYLFQQDKFYDVALDTGDKVVQCGRRV    360
Mouse    WNPHKLLAAGLQCSALLRDTSNLLKRCHGSQASYLFQQDKFYDVALDTGDKVVQCGRRV    360
         ****************:*****************************************

Human    DCLKLWLMWKAQGDQGLERRIDQAFVLARYLVEEMKKREGFELVMEPEFVNVCFWFVPPS    420
Mouse    DCLKLWLMWKAQGGQGLERRIDQAFALTRYLVEEIKKREGFELVMEPEFVNVCFWFVPPS    420
         ************* ***********.*:*.*****:***********************

Human    LRGKQESPDYHERLSKVAPVLKERMVKEGSMMIGYQPHGTRGNFFRVVVANSALTCADMD    480
Mouse    LRGKKESPDYSQRLSQVAPVLKERMVKKGTMMIGYQPHGTRANFFRMVVANPILAQADID    480
         ****:***** :***:***********.*:*********.****:**** *: **:*

Human    FLLNELERLGQDL          493
Mouse    FLLGELELLGQDL          493
         ***.*** *****
```

Figure 3: Amino acid sequence alignment of human (UniProt: Q9Y600-1) and mouse (UniProt: Q9DBE0) CSAD using Clustal Omega Multiple Sequence Alignment [59]. The colors correspond to: red = hydrophobic residue; blue = acidic residue; green = hydroxyl + sulfhydryl + amine + G; magenta = basic residue. The asterisk (*) indicates a single, fully conserved residue; a colon (:) indicates conservation between groups of strongly similar properties (i.e., scoring > 0.5 in the Gonnet PAM 250 matrix), and a period (.) indicates conservation between groups with weakly similar properties (i.e., scoring = < 0.5 in the Gonnet PAM 250 matrix).

In terms of structural similarity, glutamic acid decarboxylase-like protein 1 (GADL1) is the closest homologue to CSAD [53, 60]. GADL1 is involved in the production of the dipeptides anserine and carnosine by decarboxylating aspartic acid (Asp) to β-alanine [60, 61]. Both CSAD and GADL1 can decarboxylate cysteic acid, aspartic acid and glutamic acid, whereas CSA is their preferred substrate [53, 60].
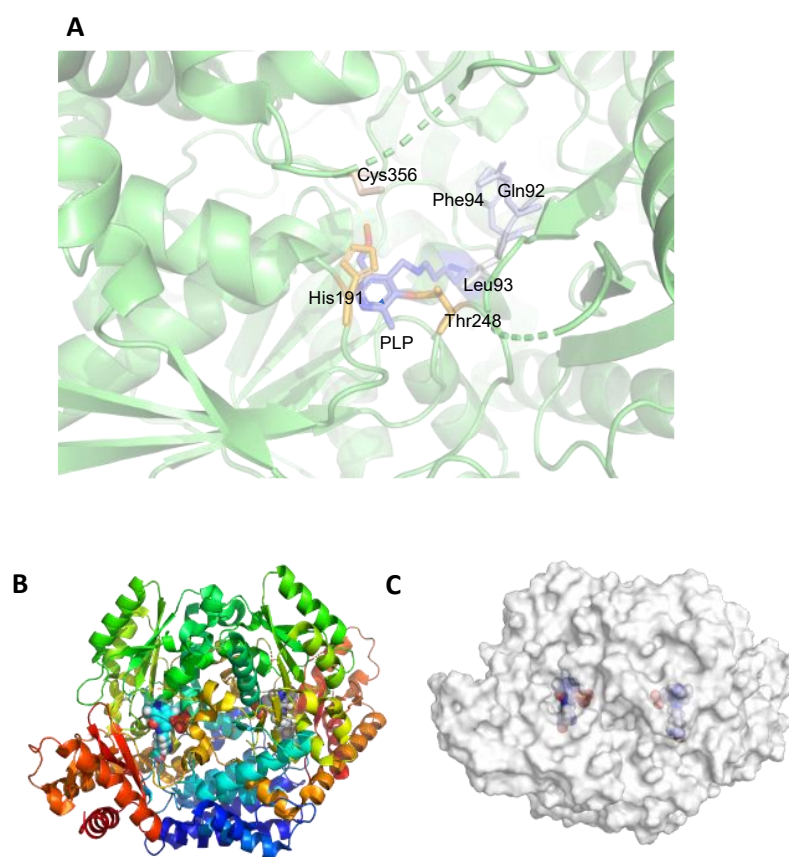
*Figure 4: **A)** Key residues at mouse CSAD (PDB Id: 6ZEK) binding site, **B)** Mouse CSAD dimer from the side, **C)** Surface representation of mouse CSAD dimer with PLP covalently bond to Lys305 shown as spheres. The figures were created by using Pymol software.*

Fig. 4 shows the structure of mouse CSAD (PDB Id: 6ZEK). The active site of CSAD is occupied with PLP which is covalently linked to Lys 305 as an internal aldimine. The key residues at the binding site of CSAD are Cys 356, His 191, Thr 248, Leu 93, Phe 94 and Gln 92 (Fig. 4A). It is possible that the strong positive charge in the active-site cavity, which can be seen on the electrostatic potential surface of CSAD in Fig. 4C, plays a key role attracting negatively charged substrates. CSAD favors substrates with short acidic side chains, such as Asp, CA and CSA, with CSA being the preferred substrate [60].

Park et al., 2014 [51] reported that taurine levels in the plasma in CSAD knockout mice were decreased by 83%. Moreover, most of the 2nd generation's offspring of these CSAD deficient mice died right after birth, demonstrating an essential physiological role of CSAD [51, 53].

1.5 Methodology

The basic principles behind the methods used in this project are described below:

1.5.1 Recombinant protein expression

Recombinant proteins have numerous applications in the fields of biology and biomedicine. They can be used to produce proteins in large quantities for studying their structure and function. Additionally, they can be used in the production of vaccines, drugs, and antibodies [62, 63]. To create a recombinant protein, the gene of interest can be cloned into an expression vector, which is then transformed into a preferred host. After this, the protein is induced and can finally be purified [64].

There are different types of hosts that can be used to achieve recombinant protein, such as bacteria, yeast, insect cells, mammalian cells, transgenic animals, and transgenic plants. Among them, *Escherichia coli (E. coli)* is the most preferred host due to its cost-effectiveness, well-studied biochemistry and genetics, rapid growth, and high productivity [65].

To produce recombinant protein, the target gene is cloned into the multiple cloning site of an expression vector [65]. An expression vector is typically a circular plasmid composed of double-stranded DNA that serves the purpose of introducing a particular gene into a host cell. Expression vectors contain several key components, such as replicons, promotors, selection markers, multiple cloning sites, and affinity tags. These components make it easier to express proteins and purify them. One of the most widely utilized plasmids for generating recombinant proteins in *E. coli* is the pET series [66]. The vectors used in this project contain a T7 *lacO* promoter system, specific antibiotic resistance gene, a hexahistidine (his_6)- tag at the N-terminal, and a tobacco etch virus (TEV) protease site [67]. Once the target gene is cloned into an expression vector, the vector is transformed into an *E. coli* strain. During this step, antibiotics specific to antibiotic resistance gene present in vector are added. As a result, the cells that have vectors survive and multiply as they will be resistant to those specific antibiotics [64, 67].

After transformation, *E. coli* cells containing an expression vector are cultured into a nutrient-rich medium at optimal temperature for producing cells in large amount. T7 *lacO* promoter system can be induced by isopropyl β-D-1-thiogalactopyranoside (IPTG). The Lac repressor and Lacl binds to the lac promotor in the absence of IPTG, repressing transcription and preventing the production of T7 RNA polymerase. IPTG triggers the release of Lacl from the lac promotor, resulting in transcription and translation of the T7 RNA polymerase gene. Afterwards, this gene starts transcription of the T7 promotor present in the vector, and subsequently it is translated into the target protein [65].

1.5.2 Affinity chromatography

Affinity chromatography utilizes the strong and specific interaction between two molecules to effectively purify individual molecules or groups of molecules from complex mixtures [68]. It uses "affinity ligand" as a stationary phase, which can be obtained from a biological source, such as enzyme, antibodies, and carbohydrate binding proteins, or it can be synthetic ligands such as metal ions, biomimetic dyes and boronates. Immobilized metal ion affinity chromatography (IMAC) uses specific interaction between immobilized metal ions, such as Ni^{2+}, Cu^{2+}, Zn^{2+}, Fe^{3+} and Co^{2+} and targets, such as protein, nucleic acids, peptides and amino acids. IMAC is most commonly used for purifying proteins that contain a polyhistidine motif, most commonly His_6-tag. For this purification process, the sample is passed through an immobilized metal ion column. This results in binding of target protein with His_6-tag to the metal ion. The rest of the sample is washed away by using a washing buffer. The addition of elution buffer consisting of a competing agent such as imidazole or a shift in pH is used to elute the target protein from the column [69]. In this project, TALON Superflow resin was used, which contained immobilized Co^{2+} ions [70].

1.5.3 Size exclusion chromatography

Size exclusion chromatography (SEC), or gel filtration, is used to separate molecules based on their size when a solution is passed through a column packed with a porous resin, containing a porous matrix of beads that are neither reactive nor adsorptive. After adding the sample to the column, the larger molecules that are unable to diffuse through the beads will elute first. The extent to which a molecule can penetrate the pore is determined by its size. If the molecules are smaller in size, they tend to enter the pores, resulting in a slower elution process. The SEC method is utilized to track

the protein quality, determine protein stability, and identify any aggregates or oligomers present in the protein sample [63, 71].

1.5.4 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a widely used high throughput screening method in drug discovery projects due to its affordability and high efficiency in identifying binders for therapeutic targets [72]. It is used to calculate the melting temperature (T_m) of a protein. The T_m refers to the temperature where the concentration of folded and unfolded protein becomes equal. To determine T_m of a protein, a solution is mixed with a dye, such as SYPRO Orange and then heated in qPCR instrument. The DSF dye exhibits minimal fluorescence when it is in solution. However, it exhibits bright fluorescence when it is attached to the hydrophobic areas of unfolded proteins. When subjected to heat, the protein gradually unfolds and reveals binding sites for the dye, resulting in an increase in fluorescence that is directly proportional to the amount of unfolded protein [73, 74]. The DSF method can be used for a series of dose-response assays, from which a preliminary dissociation constant (K_D) can be determined. However, one needs to keep in mind that one obtains the K_D, which is temperature dependent, at the melting temperature of the protein with the ligand. Thus, the K_D acquired is unlikely to be physiological and needs to be validated using other methods e.g. surface plasmon resonance, bio-layer interferometry or isothermal titration calorimetry. For this project, HTSDSF Explorer was used, a software for analyzing high throughput DSF screening, to determine both T_m and K_D values [72].

1.5.5 High-performance liquid chromatography (HPLC)

High-performance liquid chromatography (HPLC) is used in biochemical and analytical chemistry for pharmaceutical, clinal and research purposes. The sample is analyzed both qualitatively and quantitatively by separating various compounds. Chromatography operates on the principle of the sample components having an affinity for either the mobile phase or the stationary phase. Components that have a greater affinity for the mobile phase move through the column quicker, whereas components that have a stronger affinity for the stationary phase move slowly through the column. This results in different peaks on a chromatogram [75].

In this project, HPLC is used to measure enzymatic activity assay. There are two main reasons for conducting an enzymatic activity assay. One is to determine the presence or absence of a specific enzyme, while the other is to quantify the amount of enzyme present in the sample. Several crucial

factors affect enzymatic assays, including temperature, ionic strength, pH, and the concentration of substrate and enzyme. The quantity of a product produced by an enzyme can be utilized to quantify the activity of an enzyme. The first step in preparing an activity assay typically involves adding the enzyme to a buffer. This buffer creates the optimal conditions for enzyme activity, and other cofactors are added if an enzyme requires a cofactor for proper functioning. To initiate the reaction, a substrate is carefully added under regulated conditions, such as precise temperature and duration. The reaction is then halted, and the amount of product produced is analyzed using a separation method like HPLC [76]. This type of assay is referred as a stopped assay. The formula used to calculate the specific activity is:

$$Specific\ activity\left(\frac{\frac{nmol}{min}}{mg}\right) = \frac{n_{product}\ (nmol)}{t\ (min) * m_{enzyme}\ (mg)}$$

where n is number of moles in nmol, t is reaction time in min, and m is mass of enzyme in mg. The number of moles of product present in the sample is calculated by using:

$$n_{product}\ (nmol) = c_{product}\ (M) * V\ (L) * 10^6$$

where c is concentration in M and V is total volume of assay in L.

For determining the concentration of a product created by an enzyme in a sample using HPLC chromatography, the formula is as follows:

$$c_{product}\ (M) = \frac{area\ of\ product\ formed\ in\ sample}{slope\ of\ standard\ curve}$$

To conduct these assays, the standards containing reaction products of known concentrations are utilized to create a standard curve. The curve is then fitted to form a linear equation y = mx, where m is the slope of the standard curve.

1.5.6 Choice of molecule library for hit screening

In previous pilot experiments performed in our laboratory, Dr. Sunil K. Pandey and Dr. Kazi Alam performed a DFS screening of the Prestwick FDA approved chemical library consisting of 1280 drugs against purified mouse CSAD. No consistent binding data were observed, but a few promising hits were observed (Sunil K. Pandey, unpublished observations). To supplement this

experiment, the current project focused on screening of CSAD against Prestwick Original Molecules library. This library consists of 344 compounds and is created to offer a large range of chemical diversity, making it practical for subsequent structural modifications. The library has a high therapeutic potential score and possesses drug-like properties (*Lipinski*) and lead-like properties (*Opera*), along with good ADME-Tox properties. The small drug-like molecules have a mean molecular weight of 395 g/mol [77].

2. AIMS

While taurine plays a critical role in various physiological processes, limited research exists on the biosynthesis of taurine, particularly on CSAD. CSAD is a rate-limiting enzyme involved in taurine biosynthesis. Identifying inhibitors and activators of CSAD using CSA as a substrate effects hypotaurine concentration, which could significantly impact taurine levels in the body, making it a potential treatment for neurological disorders and neurodegenerative conditions such as Alzheimer's, Parkinson's, and Huntington's disease [40]. Therefore, this project aims to:

1) Optimize protein expression and purification of human and mouse CSAD
2) Optimize the CSAD enzymatic activity assay
3) Investigate inhibitors and activators of CSAD

3. MATERIALS AND METHODS

3.1 Expression and purification of human CSAD

The majority of the chemicals used in this project came from Sigma-Aldrich or Merck, unless otherwise specified in the text.

3.1.1 Transformation

Codon optimized human CSAD gene was synthesized from Gene Script contained pET-28b(+) plasmid was transformed into *E. coli* BL21- CodonPlus (DE3)-RIPL cells (Stratagene). The transformation was performed by adding 2 µL of His$_6$-CSAD-pET-28b(+) plasmid into 50 µL of competent expression cells. The cells were incubated on ice for 30 min, heat-shocked at 42°C for 45 s and returned on ice for 5 min, 500 µL of Luria Broth (LB) media (peptone 10 g/L, sodium chloride 10 g/L, yeast extract 5 g/L) was added and cells were cultured for 60 min at 400 rpm. After incubation, 100 µL of the mixture was spread on 100 µg/mL kanamycin and 34 µg/mL chloramphenicol containing LB agar plate. The agar plate was incubated upside down overnight at 37°C.

3.1.2 Overnight pre-culture

A single bacterial colony was added to 10 mL LB media containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol and incubated overnight at 37°C, 200 rpm.

3.1.3 Protein expression with varying media, incubation time and incubation temperature

Different media, incubation time, and incubation temperatures were tested to find the best growth conditions for human CSAD. Terrific Broth (TB) media, LB media and autoinduction media were prepared as shown in Table 1.

Table 1: Components required to prepare different media.

Media	Reagents
TB media	24 g/L yeast extract, 20 g/L peptone, 4 mL/L glycerol, 0.017 M KH$_2$PO$_4$ and 0.072 M K$_2$HPO$_4$
LB media	10 g/L peptone, 10 g/L sodium chloride and 5 g/L yeast extract

Auto-induction media	1% peptone, 0.5% yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 2 mM $MgSO_4$ and 0.2x trace elements.

A 100 µL overnight pre-culture was added to 100 mL of LB, TB or auto-induction media containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. All batches were started at 37°C and 200 rpm except autoinduction media. The auto-induction media was incubated, as shown in Fig. 5. In LB and TB media, optical density (OD at 600 nm) was measured after every 30-60 min. When the OD reached over 0.5, 2 mM Pyridoxine (PN) and 0.5 mM isopropylthio-β-galactoside (IPTG) was added, and then the media was divided and incubated, as shown in Fig. 5. Samples for sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) were taken for each time condition shown in Fig. 5.
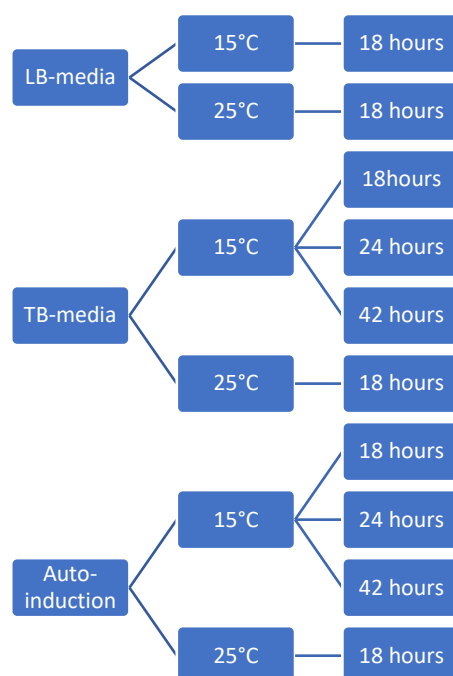

Figure 5: Overview of conditions that were tested for protein expression. Samples for SDS-PAGE were taken for each condition.

3.1.4 Cell harvest and lysis

After incubation, cells were collected by centrifugation at 6000 rpm for 20 min at 4°C. The supernatant was discarded, and the protein was extracted from the pellet by sonication.

Lysis buffer containing 50 mM sodium phosphate, 500 mM NaCl, 1 mM MgCl, 0.2 mg/mL lysozyme, 1 mM phenylmethylsulphonyl fluoride (PMSF), 1 µL benzonase and 1 cOmplete Mini, EDTA-free protease inhibitor cocktail tablet per 50 mL lysis buffer was added to the pellet. To test whether imidazole affects the protein yield, 20 mM imidazole was also added to lysis buffer in one experiment. Samples were incubated for 30 min on a rotating wheel at 4°C and thereafter sonicated for five cycles using Sonics Vibra Cell instrument with an output power of 20 Watts, 9 s pulser (one cycle: 30 s sonication and 30 s wait).

3.1.5 TALON crude

The column for TALON crude was prepared by adding 2 mL TALON superflow resin into a 20 mL gravity flow column. The column was washed with 10 column volumes (CV) Milli-Q water and then equilibrated with 10 CV of 50 mM sodium phosphate, 500 mM NaCl, pH 7.4. The cell lysate was loaded onto the column and was incubated for 1 hour on a rotating wheel at 4°C. A fraction of the flow through was collected for SDS-PAGE, and then the column was further washed with wash buffer. Two types of wash buffer were tested, one with 20 mM imidazole in binding buffer (50 mM sodium phosphate, 500 mM NaCl, pH 7.4) and the other one with 10 mM imidazole in binding buffer. Fractions of the washes were collected for SDS-PAGE. The protein was eluted using an elution buffer, prepared by adding 250 mM imidazole into the binding buffer. A PD10 column was used to exchange the buffer to 20 mM Hepes, 200 mM NaCl and 0.5 mM tris- (2-carboxyethyl) phosphine (TCEP).

3.2 Expression and purification of mouse CSAD

3.2.1 Transformation

His-CSAD mouse contained pTH27 plasmid and was transformed into *E. coli* BL21- CodonPlus (DE3)-RIPL (Stratagene) expression cells by adding 2 µL of His_6-CSAD-pTH27 plasmid into 50 µL of competent expression cells. The cells were incubated on ice for 30 min, heat-shocked at 42°C for 45 s and again placed on ice for 5 min. 500 µL of LB media (peptone 10 g/L, sodium chloride 10 g/L, yeast extract 5 g/L) was added, and cells were grown for 60 min at 400 rpm. After 60 min, 100 µL of the mixture was spread on 100 µg/mL ampicillin and 34 µg/mL chloramphenicol containing LB agar plate. The agar plate was incubated upside down overnight at 37°C.

3.2.2 Overnight pre-culture

LB media containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol was inoculated with a single colony and was grown overnight at 37°C, 200 rpm.

3.2.3 Glycerol stock

A glycerol stock was prepared from pre-culture that was incubated overnight by adding 50% ice-cold glycerol and pre-culture in equal amounts. The glycerol stock was stored at -80°C and used for future experiments to prepare overnight pre-culture.

3.2.4 Large-scale protein expression

As a growth medium for the bacterial clones, autoclaved LB medium was used. Ampicillin with 100 µg/mL, chloramphenicol with 34 µg/mL final concentration and overnight pre-culture (10 mL per 1 L LB media) was added to LB medium and was placed in an incubator at 37°C, 200 rpm. The OD of the medium was continuously measured every 30- 60 min until the density reached approximately 0.8. Then the incubation temperature was reduced to 15°C and 2 mM PN and 0.5 mM IPTG was added. After overnight protein expression, the cells were centrifuged at 4000 rpm for 15 min at 4°C. The supernatant was removed, and the pellet was washed with 20 mM imidazole, 50 mM sodium phosphate, 500 mM NaCl, pH 7.4 buffer to remove LB media. The cells were centrifuged again for 15 min at 4000 rpm at 4°C. The supernatant was removed, and the pellet was mixed with phosphate-buffered saline (PBS) and stored at -20°C. Another experiment was performed to test the effect of PN by not adding PN to media.

3.2.5 Cell harvest and lysis

The cells were centrifuged at 14000 rpm for 30 min at 4°C to remove PBS. After centrifugation, the supernatant was discarded, and lysis buffer was added. Lysis buffer was prepared by adding 1 mM $MgCl_2$, 0.2 mg/mL lysozyme, 1 mM PMSF, 0.5 mM dithiothreitol (DTT), 20 mM imidazole, 1 µL benzonase and 1 cOmplete Mini, EDTA-free protease inhibitor cocktail tablet was added in 50 mL 50mM sodium phosphate buffer, pH 7.4. In the pellet from 1L bacterial culture, 20 mL lysis buffer was added and then the cells were incubated for 30 min at 4°C on a rotating wheel. Sonication was performed on Sonics Vibra Cell instrument for 10 min with an output power of 20 Watts every 2 s while keeping the sample on ice.

3.2.6 TALON Crude

The column for TALON crude was prepared by adding TALON Superflow resin into a 20 mL gravity flow column. The column was washed with 10 CV of Milli-Q water and then equilibrated with 10 CV of 50 mM sodium phosphate, 500 mM NaCl, pH 7.4. The cell lysate was centrifuged at 4000 rpm for 15 min at 4°C, and the supernatant was loaded onto the column and was incubated for 1 hour on a rotating wheel at 4°C. A fraction of the flow through was collected for SDS-PAGE. Then the column was washed with wash buffer containing 50mM sodium phosphate buffer with 500 mM NaCl, pH 7.4. The protein was eluted using elution buffer, which was prepared by adding 300 mM imidazole in the binding buffer (50 mM sodium phosphate, 500 mM NaCl, pH 7.4). PD10 column was used to exchange the buffer to gel filtration buffer (20 mM Hepes, 200 mM NaCl and 0.5 mM TCEP).

3.2.7 TEV protease cutting tests

TEV protease was used to cleave the His_6-tag from mouse CSAD. Two samples were prepared with varying TEV protease (1 mg/mL) to His_6-CSAD ratios. Those ratios were 1:100 and 1:200. The samples were incubated overnight at 4°C and then added onto a nickel-nitrilotriacetic (Ni-NTA) column. The flowthrough was found to contain cleaved CSAD. To confirm the protein was indeed cleaved, this sample was sent for mass spectrometry (MS) analysis at the Biocenter Oulu core facility in Finland.

3.2.8 Size exclusion chromatography

For further protein purification, ÄKTA Pure was used for size exclusion chromatography using a Superdex 200 Increase 10/300 GL column to separate proteins by size at 4°C. The column was washed with filtered and degassed water and then equilibrated with filtered and degassed gel filtration buffer (20 mM Hepes, 200 mM NaCl and 0.5 mM TCEP). After equilibration, the His_6-taged sample was injected at a flow rate of 0.5 mL/min into the column. A chromatogram with one sharp peak was received. Fractions (each 0.5 mL) around that peak were collected, pooled together and upconcentrated using an Amicon centrifugal filter (30 000 Da molecular weight cut-off) to around 500 µL. The concentration was measured spectrophotometrically at 280 nm using a Nanodrop. Aliquots of protein were prepared, snap froze with liquid nitrogen and stored at -80°C.

3.2.9 Gel-electrophoresis

The samples were prepared for SDS-PAGE by adding 5 µL of loading dye to 15 µL of sample. The loading dye was prepared by mixing BIO-RAD 4x Laemmle sample buffer with β-mercaptoethanol. These samples were then incubated for 5 min at 95˚C. The pre-casted gel from BIO-RAD (Mini Protean 4-20% TGX Gels, 15 µL, 15 well) was used to run the SDS-PAGE electrophoresis. The gel tank was filled with SDS-running buffer (Tris-HCL 25 mM, glycine 200 mM, SDS 0.1%). Marker (#26616 Pageruler Prestained by Thermo Fisher Scientific) and the samples were loaded into the wells. The gel tank was then connected to BIO-RAD Power Pac 300 and was run at 200 V for about 30 min. The gel was washed for 5-10 min with water and then kept in Instant Blue for 30-60 min for the visibility of the bands and detained again in water. The image of SDS-PAGE was taken by using BIO-RAD CemiDoc^{TM} XRS+ with Image Lab^{TM} Software.

3.3 Finding T_m of mouse CSAD by DSF

A His$_6$-CSAD aliquot was thawed and centrifugated at 10000 rpm, 4°C for 10 min, the protein concentration was remeasured using Nanodrop. A small validation assay was performed to find optimal protein and SYPRO Orange concentration. For that gel filtration buffer containing 20 mM Hepes, 200 mM NaCl (pH 7.4) was used to prepare 0.1 mg/mL, 0.075 mg/mL and 0.05 mg/mL CSAD concentrations. These samples also contained either 5x or 10x SYPRO Orange from Thermo Fisher Scientific. Triplicates of each condition (6 in total) were prepared. Instrument Light Cycler 480 II from Roche with 384-well plate was used to detect SYPRO Orange fluorescence intensity (λex = 465 nm, λem = 580 nm) while heating from 20°C to 99°C at 2°C per min.

Protein expression was performed with and without the addition of PN (2 mM). The purified protein from both conditions was subsequently analyzed using 0.1 mg/mL His$_6$-CSAD with 5x SYPRO Orange under identical instrumental settings as described above. The purpose of this experiment was to determine whether the addition of PN during protein expression has any impact on the T_m of CSAD.

3.4 HPLC enzymatic activity assay to find optimal conditions for CSAD

HPLC was used to determine the best activity assay condition of His$_6$-CSAD by testing varying concentrations of CSA, i.e., 0.5 mM, 1 mM and 2 mM for 1, 5, 15 and 30 min at 37°C. For each incubation time and each CSA concentration, dublicates of samples were prepared.

3.4.1 Sample preparation

A 100 µL reaction mixture was prepared by adding 60 mM K-phosphate, 1 mM DTT, 50 µM PLP, and 0.035 mg/mL CSAD. CSA with concentrations of 0.5 mM, 1 mM and 2 mM was added to the mixture. After adding CSA, the samples were incubated at 37°C for 1, 5, 15 and 30 min. The reaction was stopped by adding ice-cold 5% acetic ethanol. After incubating the samples at -20°C for a minimum of 30 min, the samples were centrifuged at 13500 rpm for 10 min at 4°C before being added to a microtiter plate. As a standard, 0.05 mg/mL hypotaurine in 50 mM Na- phosphate buffer was used. To detect fluorescence, 1 mL O-phthalaldehyde (OPA) reagent was mixed with 10 µL β-mercaptoethanol.

Activity assay was also performed on CSAD with and without PN. For this assay, triplicates were prepared with 2 mM CSA concentration. The reaction was stopped after 1 min of incubation. All the other assay conditions were the same as described above.

3.4.2 Instrumental settings

The samples were analyzed on an Agilent 11100 Series HPLC by using column Zorbax Eclipse XDBC-18. The mobile phase was set at 0.7 mL/min, containing Milli-Q water, 100% ethanol, and 0.1 M Na-phosphate pH 6.0 in a ratio of 27:30:50. The column thermostat was set to 40°C. The injector was programmed to wash the needle first, then the sample was added to 50 mM Na-phosphate buffer in a ratio of 1:1. Afterwards the needle was washed with 50 mM Na-phosphate buffer, and 4.4% of OPA was added. The sample was then mixed, and 20 µL of that sample was injected into the column. The fluorescence signal was detected by using the excitation wavelength (λex) of 360 nm and the emission wavelength (λem) was set to 455 nm.

3.5 High throughput screening of the Prestwick Original Molecules library

To find the stability effect of different compounds on His$_6$-CSAD, T$_m$ was determined using DSF. For this assay, 0.1 mg/mL His$_6$-CSAD in 20 mM Hepes, 200 mM NaCl (pH 7.4), 5x SYPRO Orange, and 250 µM compound (5% Dimethyl sulfoxide (DMSO)) from Prestwick Original Molecules library was used. For reference, two columns in 384- well plate were used for adding 100% DMSO instead of compounds leading to 5% final DMSO. All the other instrumental conditions were the same as described in 3.3.

3.6 Hit validation

The primary hits from the screening of Prestwick Original Molecule library and hits obtained previously by co-workers from the Prestwick FDA approved chemical library were further validated by concentration-dependent binding assays. The compounds were tested for 12 concentrations. The initial concentration of the compounds was 250 µM with a dilution factor of 0.5 dilutions per well to a concentration of 0.12207 µM. In this experiment, duplicates for each sample were prepared. All the other assay conditions were the same as described above.

3.7 Enzymatic activity assay using a fluorescence plate reader (Tecan Spark)

To measure the enzymatic activity of CSAD, five types of samples were prepared in triplicates: 1) sample with only CSAD, (2) sample with both CSAD and CSA, (3) sample with only CSA. Additional to those tests, (4) mesna was tested, (5) and a DMSO control (both last two samples contained 0.035 mg/mL CSAD and 2mM CSA). To prepare the samples, 60 mM K-phosphate, 1 mM DTT, and 50 µM PLP were added. The total volume, which should be 100 µL including K-phosphate, DTT, PLP, CSAD, CSA, OPA, and mesna, was adjusted using Milli-Q water. OPA reagent was prepared by adding 10 µL β-mercaptoethanol per 1 mL OPA. From that OPA reagent, 8.9% OPA was added to the samples. 100 µM mesna was added to the fourth sample and an equal volume of DMSO was added to the fifth sample. Then CSAD was added in all the samples, except the third sample where CSAD was not supposed to be added. The samples were incubated for 5 min at 37°C. The reaction was started by adding CSA, except for the first sample. Right after adding CSA, the fluorescence intensity was measured every minute for 20 min by using Tecan Spark.

In addition to the samples, standards were prepared. Hypotaurine with 0 µM, 5 µM, 10 µM, 50 µM, 100 µM, 250 µM, 500 µM, and 1000 µM concentrations in 50 mM Na-phosphate buffer (pH 6) was used as standards. An equal amount of 5% acetic ethanol was added to the standards and was analyzed on Tecan Spark.

3.8 Enzymatic activity of CSAD in the presence of stabilizer and destabilizers using HPLC

A 100 µL reaction mixture was prepared by adding 60 mM K-phosphate, 1 mM DTT, 50 µM PLP, 0.035 mg/mL CSAD and compounds from hit validation with concentrations of 0 µM (DMSO control), 1 µM, 10 µM, 100 µM, 250 µM, and 500 µM. After adding 2 mM CSA, the samples were

incubated at 37°C for 1 min. The reaction was stopped by adding ice-cold 5% acetic ethanol. The samples were incubated at -20°C for a minimum of 30 min. Afterwards they were centrifuged at 13500 rpm for 10 min at 4°C and were added to a microtiter plate. Hypotaurine with concentrations of 0 µM, 25 µM, 50 µM, 75 µM, 100 µM, 200µM, and 400 µM in 50 mM Na-phosphate (pH 6) was added to equal amounts of 5% acetic ethanol, was used as standards. OPA reagent was prepared by adding 2.5 µL of β-mercaptoethanol per 1 mL OPA. The instrumental settings of HPLC were the same as described in section 3.4.2.

3.9 Structural similarity and molecular docking

The structural similarity evaluation between the structure of CSA and hit compounds, as well as molecular docking was performed by Dr. Kazi Alam. The structures were compared using morgan fingerprints, specifically Extended connectivity fingerprint, up to four bonds (ECFP4), along with the tanimoto coefficient. Additionally, Schrodinger was utilized to analyze the interaction of the hit compounds against the protein.

4. RESULTS

4.1 Expression and purification of human CSAD

Human CSAD was expressed in different media using varying incubation times and temperatures to find the best growth conditions. After sonication, the samples were centrifuged, and then samples from the supernatant (soluble fraction) and the pellet (insoluble fraction) were separated using SDS-PAGE, see Fig. 6. The SDS-PAGE from Fig. 6A shows that LB-media with 18 hours of incubation at 25°C produced the highest levels of CSAD, as the soluble fraction has the strongest band around 57.5 kDa (lane 4), corresponding to the molecular weight of human CSAD with a His$_6$-tag. The autoinduction media produced not the protein of interest (see Fig. 6A, lane 8-15), whereas the TB media had the strongest bands in the insoluble fractions (see Fig. 6B).
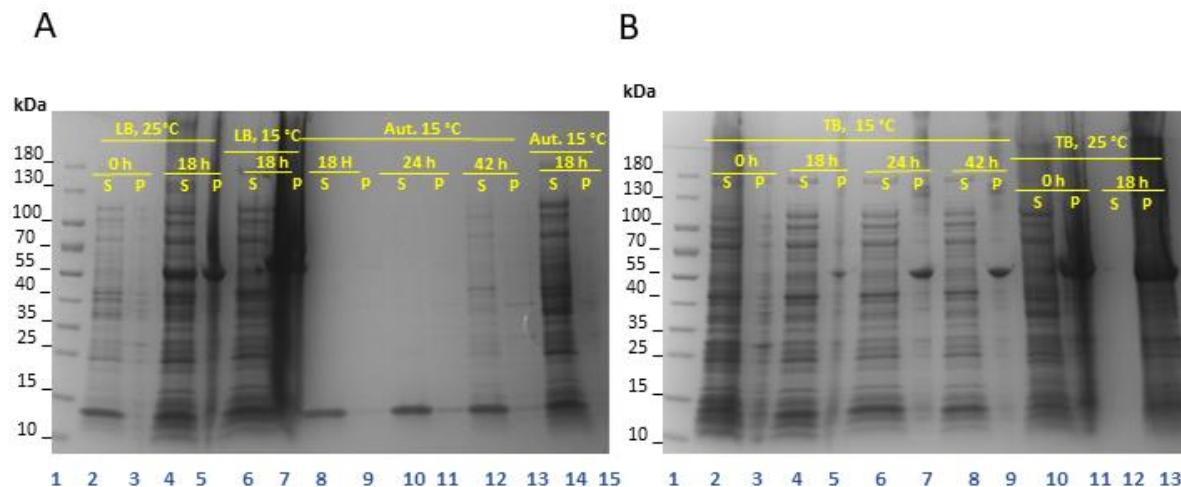

*Figure 6: SDS-PAGE of samples from the small-scale expression of human CSAD. In both SDS-PAGE soluble fractions are marked with S and insoluble fractions (pellet) are market with P. **A)** Lane **1**: Marker with known molecular weight (kDa); **2**: LB media before induction, soluble fraction, 25°C; **3**: LB media before induction, insoluble fraction, 25°C; **4**: LB media, 18 hours, soluble, 25°C; **5**: LB media, 18 hours, insoluble, 25°C; **6**: LB media, 18 hours, soluble, 15°C; **7**: LB media, 18 hours, insoluble, 15°C; **8**: Autoinduction, 18 hours, soluble, 15°C; **9**: Autoinduction, 18 hours, insoluble, 15°C; **10**: Autoinduction, 24 hours, soluble, 15°C; **11**: Autoinduction, 24 hours, insoluble, 15°C; **12**: Autoinduction, 24 hours, soluble, 15°C; **13**: Autoinduction, 42 hours, insoluble, 15°C; **14**: Autoinduction, 18 hours, soluble, 25°C; **15**: Autoinduction, 18 hours, insoluble, 25°C. **B)** lane **1**: Marker with known molecular weight (kDa); **2**: TB media before induction, soluble fraction, 15°C; **3**: TB media before induction, insoluble fraction, 15°C; **4**: TB media, 18 hours, soluble, 15°C; **5**: TB media, 18 hours, insoluble, 15°C; **6**: TB media, 24 hours, soluble, 15°C; **7**: TB media, 24 hours, insoluble, 15°C; **8**: TB, 42 hours, soluble, 15°C; **9**: TB, 42 hours, insoluble, 15°C; **10**: TB, uninduced, soluble, 25°C; **11**: TB, uninduced, insoluble, 25°C; **12**: TB, 18 hours, soluble, 25°C; **13**: TB, 18 hours, insoluble, 25°C*

To obtain higher protein yield, different imidazole concentrations were tested during purification using TALON crude chromatography (see Table 2). An imidazole concentration of 20 mM in the wash buffer and 250 mM in the elution buffer, gave the highest protein yield of 0.38 mg from 4 L culture.

Table 2: Effect of imidazole concentration in different buffers on protein yield.

	Binding buffer	Wash buffer	Elution buffer	Protein yield of human CSAD from 4 L culture
1.	0	20 mM	250 mM	0.38 mg
2.	0	10 mM	250 mM	0.11 mg
3.	20 mM	20 mM	250 mM	0.15 mg

4.2 Expression and purification of mouse CSAD

Optimizing protein expression and purification of human CSAD was difficult and time consuming since the protein yield was low, even after trying different concentrations of imidazole. Therefore, mouse CSAD was expressed and purified for further testing. For mouse CSAD, there was an established protocol for expression and purification [60]. That same method was used for both human CSAD and mouse CSAD to see the difference in protein yield (see Fig. 7, A and B).
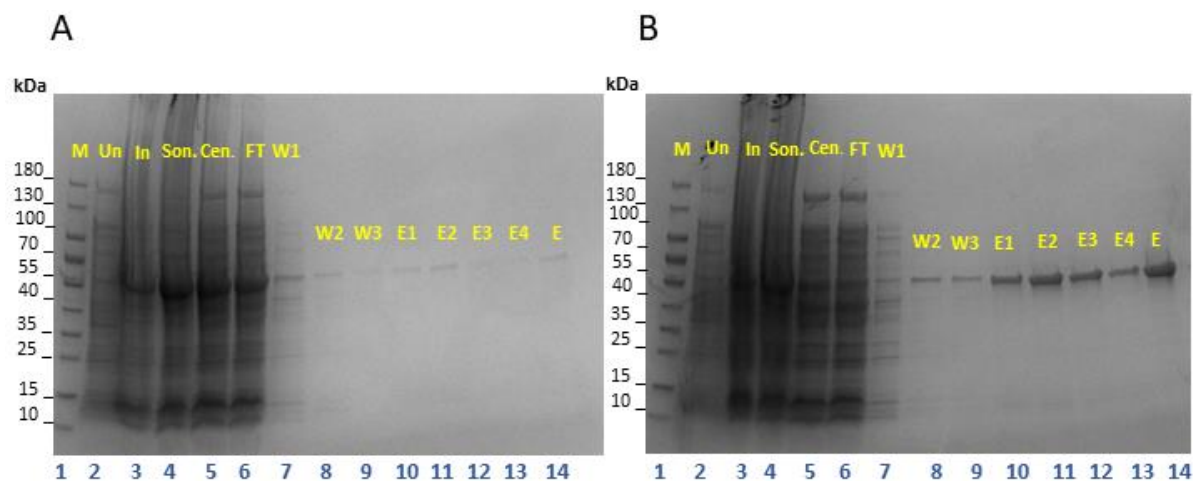

Figure 7: SDS-PAGE for CSAD expression and purification. A) shows the expression and purification of human His$_6$-CSAD, and B) shows the expression and purification of mouse His$_6$-CSAD. For both A) and B) Lane 1: Marker with known molecular weight in kDa, 2: Uninduced sample, 3: induced sample, 4: sample taken after sonication, 5: supernatant after centrifugation, 6: flow-through from the TALON, 7: first wash, 8: second wash, 9: third wash, 10: first elution, 11: second elution, 12: third elution, 13: forth elution, 14: sample where all the elution fraction were collected together and were up concentrated.

The theoretical molecular weight of human His$_6$-CSAD is 57.5 kDa, whereas mouse His$_6$-CSAD is 58.3 kDa. Fig. 7 shows that both human CSAD (Fig. 7A) and mouse CSAD (Fig. 7B) were pure, but the protein yield of human CSAD was very low. Fig. 7A shows that most of human CSAD was lost in the flow through. The protein yield of human His$_6$-CSAD from 2 L culture was 0.025 mg, whereas the corresponding yield of mouse His$_6$-CSAD was 0.235 mg. Although mouse CSAD had

a tenfold higher protein yield, it was still insufficient for further experiments. Some of the protein was lost in washes, and to limit this loss, the imidazole concentration in buffers was adjusted. In Fig. 7B, the imidazole concentration was 20 mM in the wash buffer and 100 mM in the elution buffer. In later experiments, imidazole was not added to the wash buffer, and the concentration in the elution buffer was increased to 300 mM. After these changes, the protein yield from the 3 L culture became 3.5 mg, and the protein loss in washes was reduced (Fig. 8, lane 6-8).
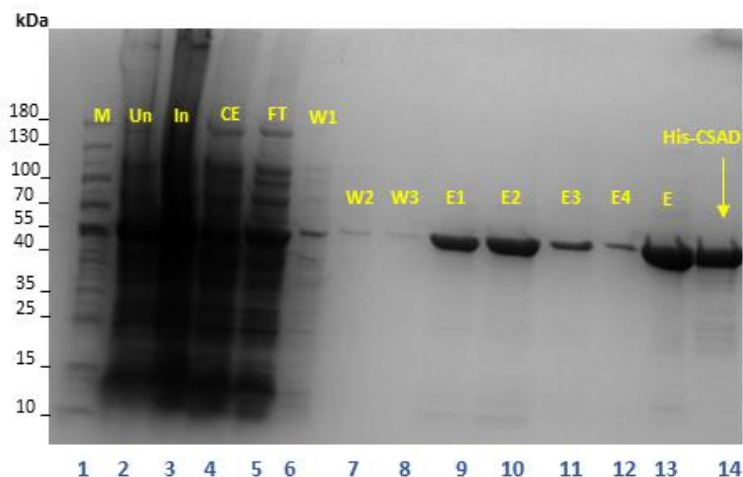

*Figure 8: SDS-PAGE of Mouse His$_6$-CSAD. Lane **1**: marker, **2**: uninduced sample, **3**: induced, **4**: crude extract, **5**: flow-through, **6**: first wash 1, **7**: second wash, **8**: third wash, **9**: first elution, **10**: second elution, **11**: third elution, **12**: fourth elution, **13**: His$_6$-CSAD where all the elution fractions were pooled together and were up concentrated, **14**: His$_6$-CSAD after gel filtration.*

4.3 TEV protease cutting

Several samples with different TEV protease to CSAD ratio were sent for mass spectrometry analysis to the Biocenter Oulu core facility in Finland for accurate mass measurements of CSAD. The sample without TEV protease with mass around 58 kDa is shown in Fig. 9A. The His$_6$-tag was effectively cleaved in the samples with TEV protease to CSAD ratios of 1:100 (results not presented) and 1:200 (shown in Fig. 9B), resulting in a protein mass of 55 kDa, which matched the theoretical molecular weight of mouse CSAD without His$_6$-tag.

A

B
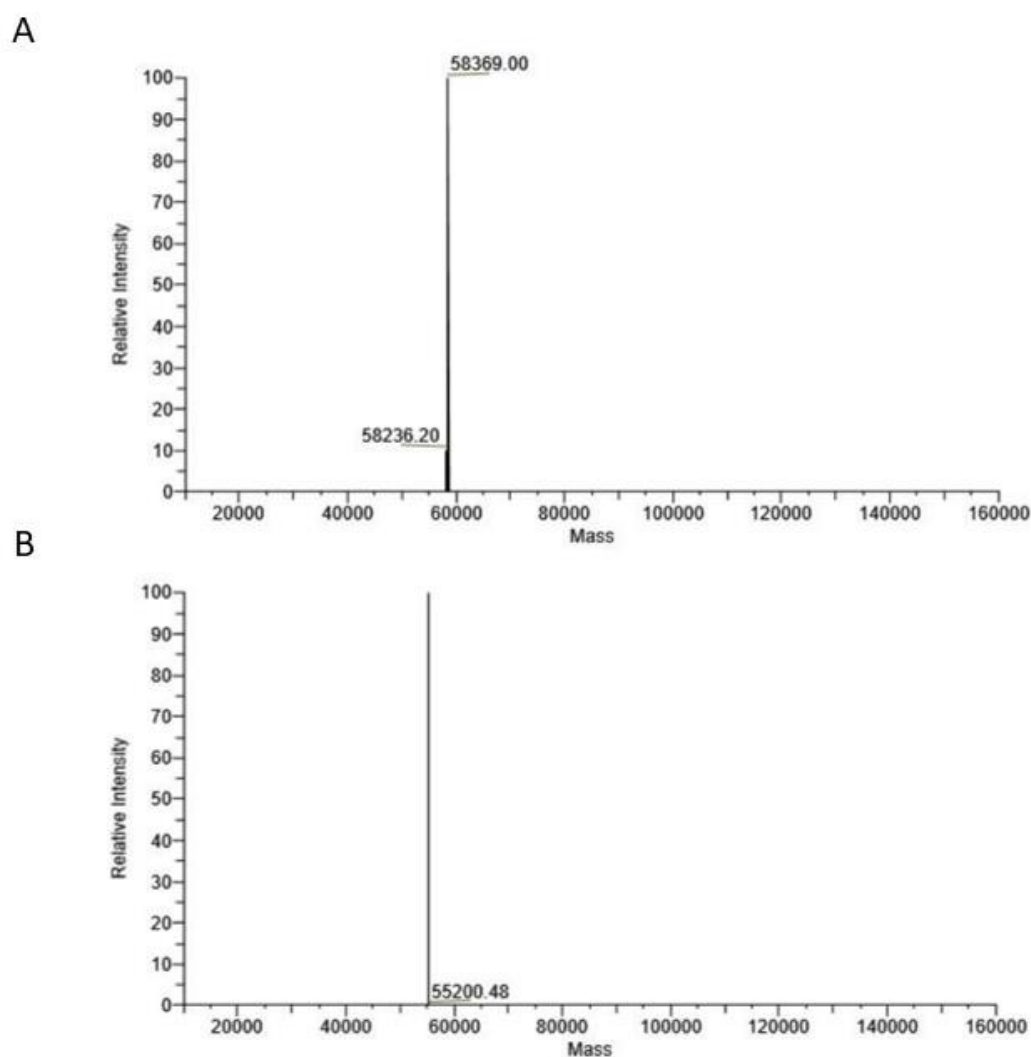
*Figure 9: Mass spectrometry of mouse CSAD after TEV protease cutting. **A)** Mass spectrum of mouse CSAD with His$_6$-tag. **B)** Cleaved mouse CSAD using 1:200 TEV to His$_6$-CSAD.*

4.4 T_m of CSAD using DSF

The T_m of CSAD was measured using DSF with varying concentrations of CSAD and SYPRO Orange, as shown in Fig. 10. CSAD concentration with 0.1 mg/mL and 5x SYPRO Orange was chosen for further experiments since the peak was sharp, without any background noise (Fig. 10, blue undotted line). The T_m with these concentration conditions was $60.0 \pm 0.38°C$, calculated by finding the extrema of arbitrary fluorescence unit derivative (Fig. 10, lower traces).

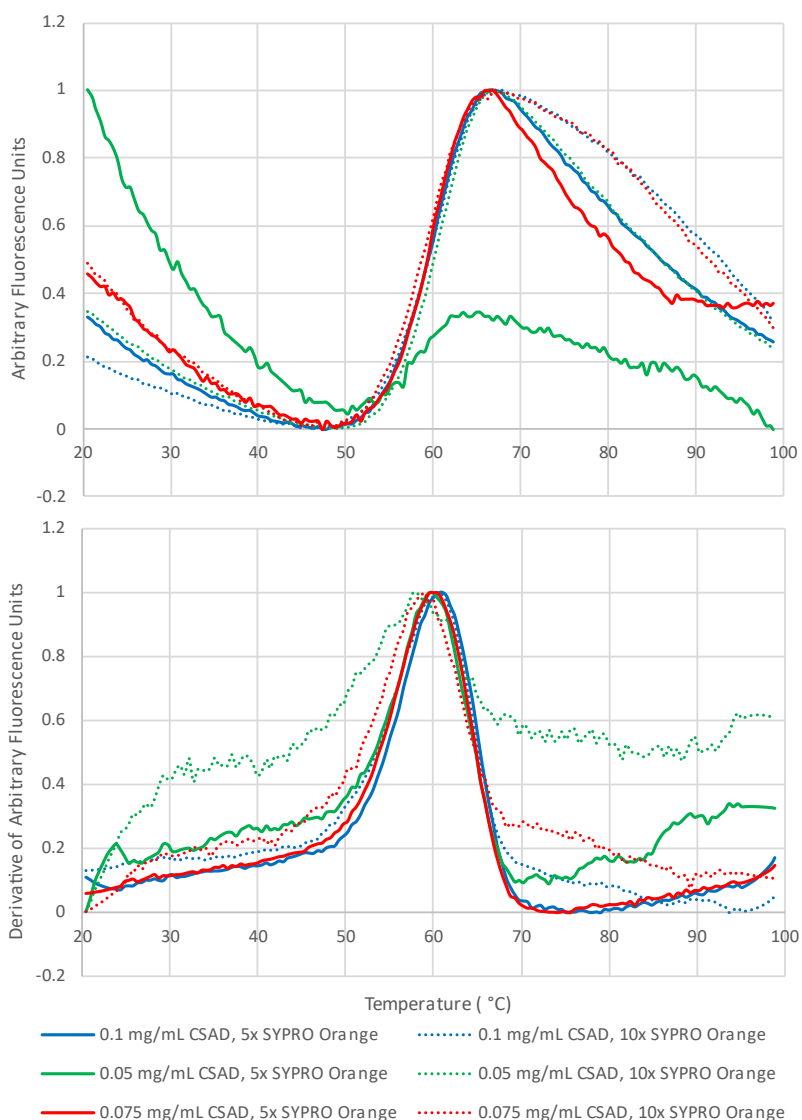*Figure 10: Thermal stability validation of CSAD denaturation curves (upper traces) and first derivative (lower traces) of varying protein and SYPRO Orange concentrations. All signals are normalized between 0 and 1.*

4.5 Activity assay with varying CSA concentrations and incubation times

The Activity assay of CSAD was analyzed by using HPLC by using CSA as a substrate. CSAD converts CSA into hypotaurine. As CSAD, CSA, and hypotaurine are not fluorescent by themselves, OPA is used to measure fluorescence intensity. A known concentration of hypotaurine was used as a standard.

Fig. 11 shows the reproducibility of fluorescence signal of the OPA-hypotaurine conjugate after 1, 3, and 6 hours of OPA preparation. The graph shows that OPA-hypotaurine conjugate was not so stable, and therefore several standards were needed while analyzing samples.
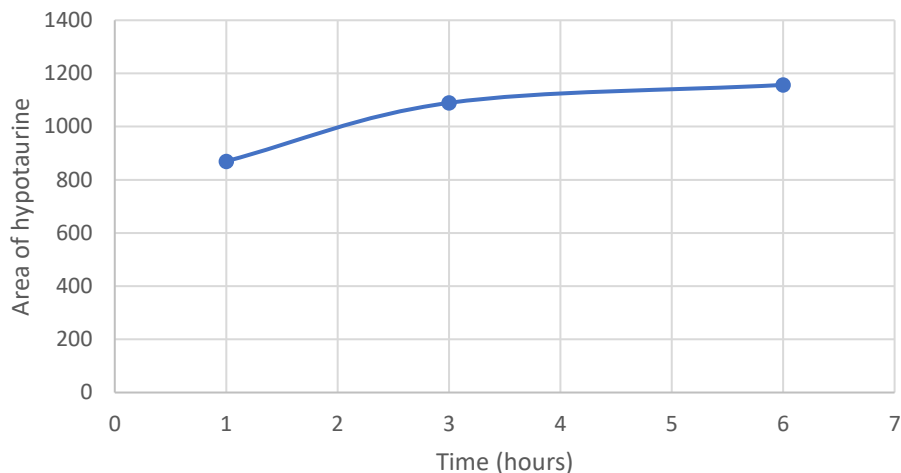
Figure 11: Area of OPA-hypotaurine conjugate over 6 hours. OPA was prepared by adding 10 µL of β-mercaptoethanol in 10 mL OPA reagent.

To find the best conditions for the specific activity of CSAD, varying CSA concentrations and incubation times were tested. Fig. 12 shows that the specific activity was highest when 2 mM CSA was added and the sample was incubated for 1 min. Longer incubation time reduced the specific activity.

Figure 12: Specific activity of CSAD representing mean ± SD when 0.5 mM CSA, 1 mM CSA, and 2 mM CSA were added for 1, 5, 15, and 30 min.

4.6 Effect of PN on CSAD

4.6.1 Comparison of T_m of CSAD in the presence and absence of PN

To determine if the addition of PN during protein expression affected the T_m of CSAD, triplicates of CSAD in the presence and absence of PN were analyzed using DSF (Fig. 13). Fig. 13A and 13B (upper traces) show how the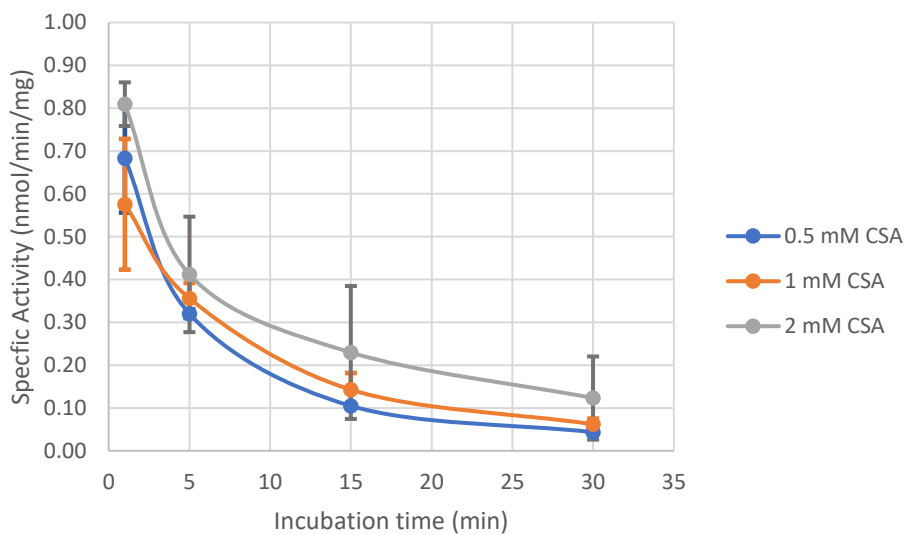 arbitrary fluorescence units change when the temperature increases. Whereas lower traces in Fig.13 are plotted by derivating arbitrary fluorescence units from Fig. 13A and 13B.

The T_m of CSAD in the absence of PN was $60.6 \pm 0.03°C$, whereas the T_m of CSAD in the presence of PN was $61.3 \pm 0.54°C$. This indicated that the presence or absence of PN did not significantly ($p=0.16 > 0.05$, two tailed t-test) affect the T_m of CSAD.
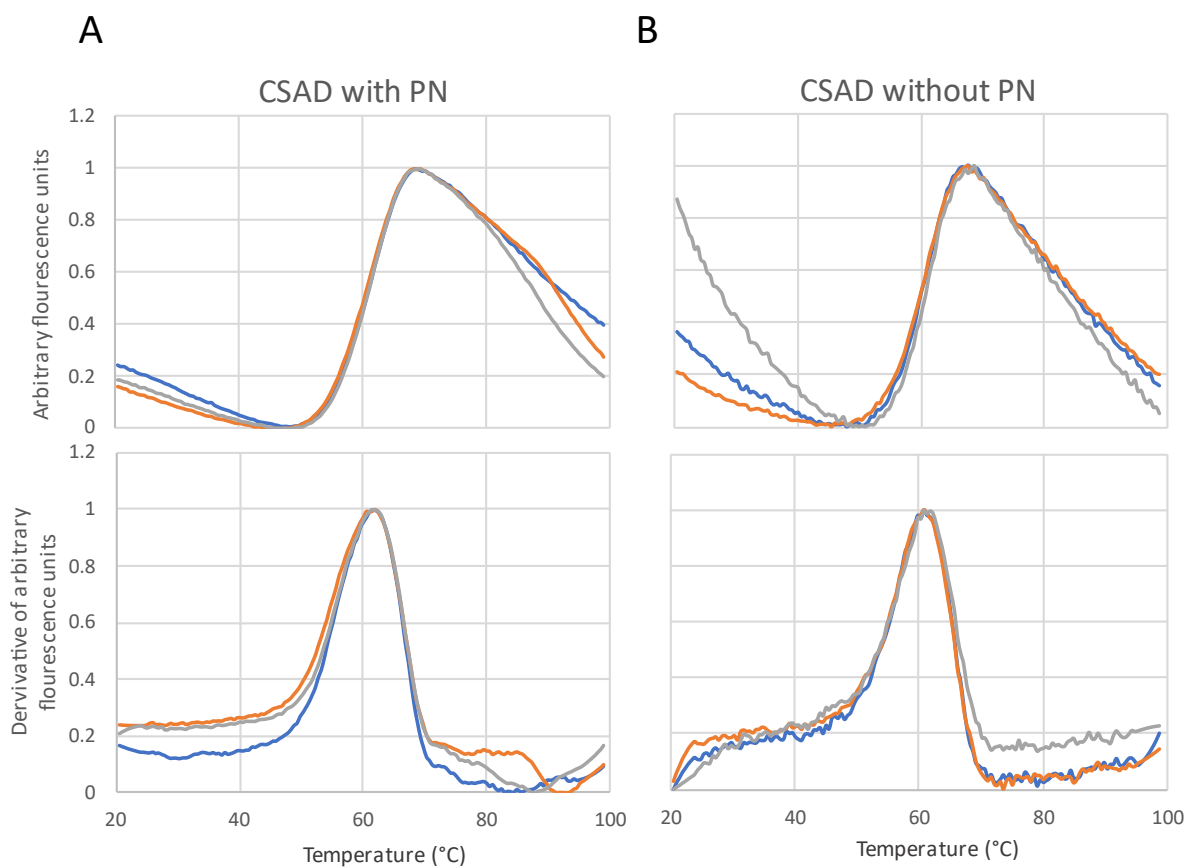

*Figure 13: Thermal stability of CSAD in the presence **A)** and absence **B)** of PN. DSF denaturation curves (upper traces) and first derivative (lower traces) are normalized between 0 to 1. Different colors represent experimental replicates.*

4.6.2 Comparison of specific activity of CSAD in the presence and absence of PN

Activity assay was performed on triplicates of both CSAD in the presence and absence of PN during protein expression to see if the molecule had any effect on the activity of CSAD (Fig. 14). The area of 0.46 mM hypotaurine was used as a standard to calculate the specific activity. This standard was analyzed before and after the samples. The specific activity of CSAD in the presence of PN was 0.48 ± 0.072 nmol/min/mg, whereas, in the absence of PN, it was 0.33 ± 0.13 nmol/min/mg (p=0.19, two tailed t-test).
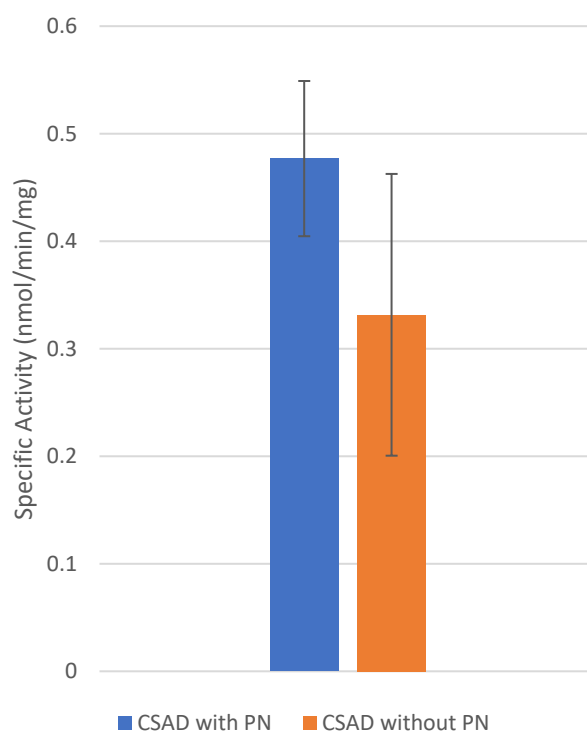
Figure 14: Specific activity of CSAD in the presence and absence of PN. The average specific activity of CSAD with PN was 0.48 ± 0.072 nmol/min/mg, whereas CSAD without PN was 0.33 ± 0.13 nmol/min/mg (p=0.19). Data are presented as mean ± SD of three experiments.

4.7 High throughput DSF screening

High throughput DSF screening was performed on Prestwick Original Molecule library to find the effect of compounds on the thermostability of CSAD. The graph in Fig. 15 shows ΔT_m of all the compounds, and DMSO controls which were used as a reference to calculate ΔT_m ($\Delta T_m = T_{m(compound)} - T_{m(DMSO\ control)}$). CSAD in 20 mM Hepes, 200 mM NaCl pH 7.4 with 5% DMSO (DMSO control) had a $T_m = 58.23 \pm 1.18°C$ (mean ± standard deviation).
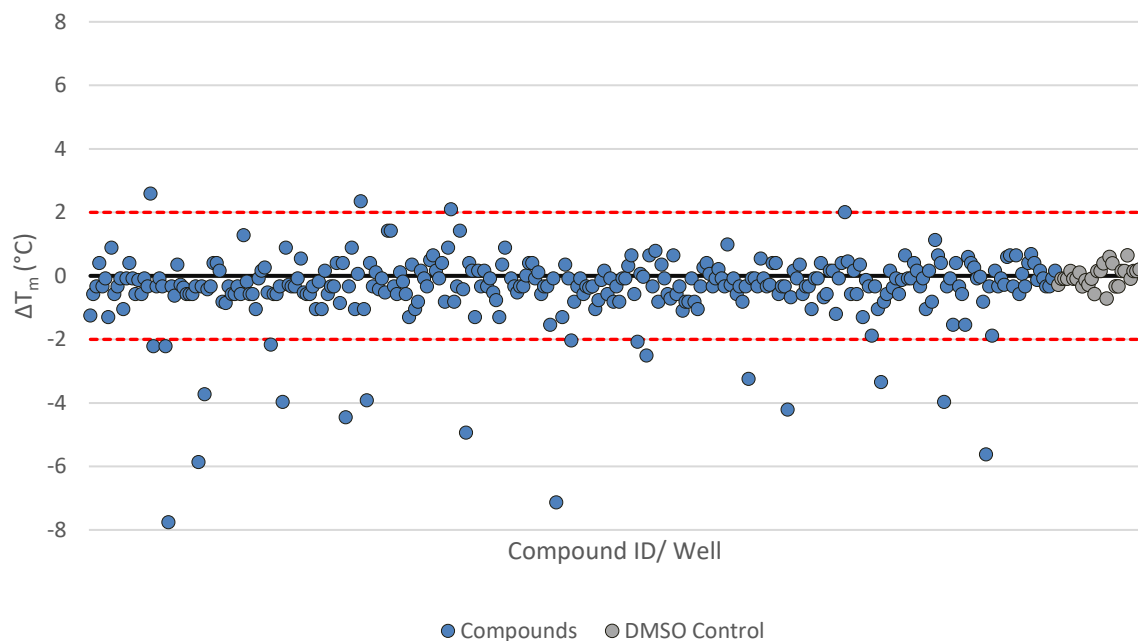
Figure 15: Overview of compounds from Prestwick Original Molecular library, where the blue circles are compounds and the gray circles are DMSO control (Reference samples). The black line at ΔT$_m$ = 0 shows the average of DMSO control, whereas the red dotted lines show the threshold of ΔT$_m$ = 2°C. The compounds with ΔT$_m$ above 2°C or below -2°C were used for serial dilution.

The threshold of $\Delta T_m = 2°C$ was set to select the compounds for further hit validation. Not all compounds with $|\Delta T_m| > 2$ were selected. Since some of those compounds did not have any peaks, the peaks were not sharp, or there were multiple peaks (indicating two/multiple-state transition). For these kind of graphs, it was challenging to find the melting temperature since the derivative of arbitrary fluorescence unit had many or no extrema at all.

All selected hits from the initial screening are shown in Fig. 16. DMSO control are shown in gray. In total 17 compounds were selected for further validation, 4 of them showed an increasing effect on the T_m-value for CSAD ($+\Delta T_m$, stabilizer) whereas 13 had a decreasing effect ($-\Delta T_m$, destabilizer).
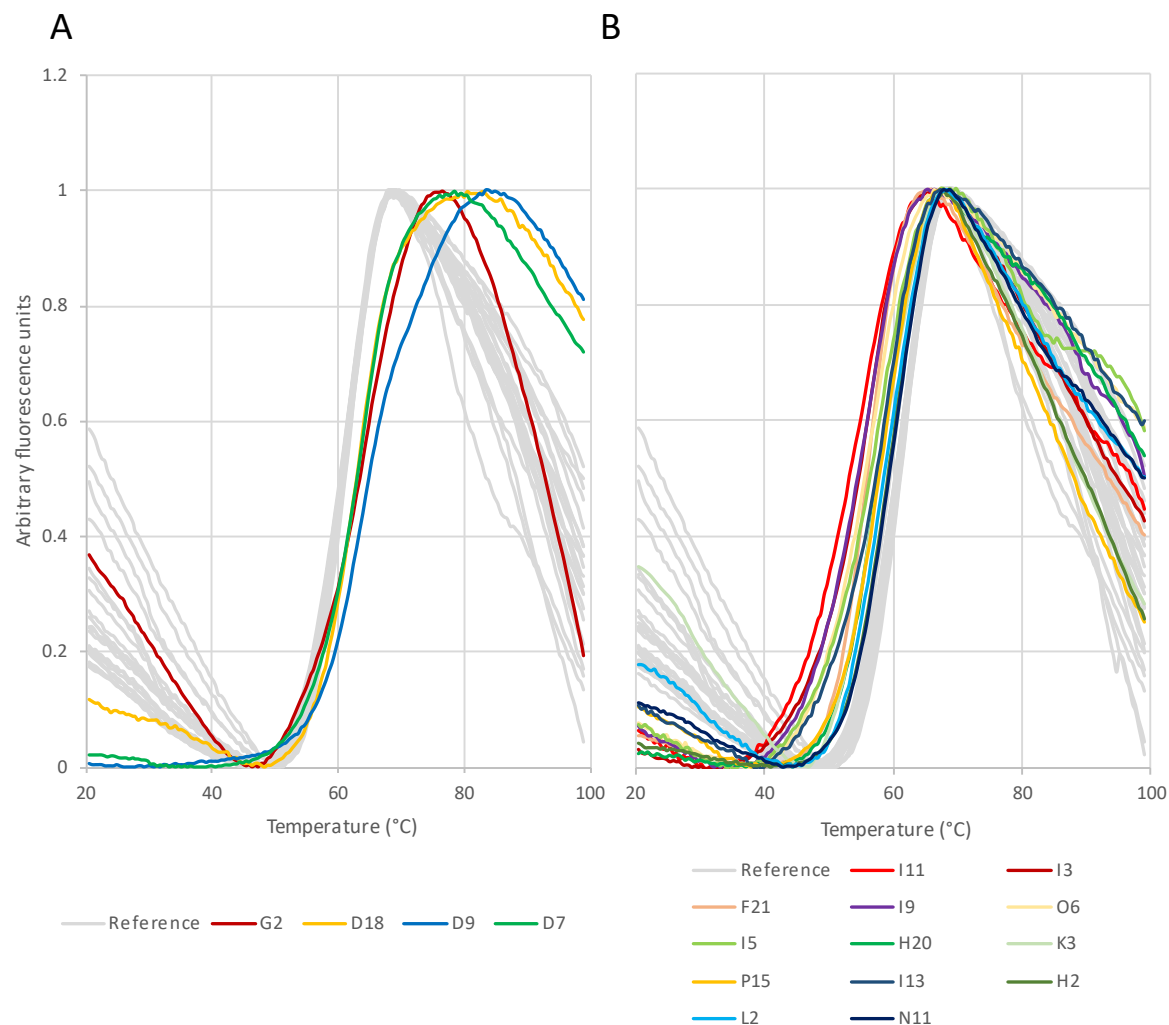
*Figure 16: Thermal melting curves of the selected hits from DSF screening of Prestwick Original Molecules library. **A)** stabilizers and **B)** destabilizers.*

4.8 Hit validation

A total of 17 primary hits from the Prestwick Orignal Molecules library (see Fig. 16) and one hit (previously found by a Dr. Sunil K. Pandey, unpublished data) from the Prestwick FDA-approved chemical library (Plate 4, well H5 containing methanethiolate/ 2-mercaptoethanesulfonate/ mesna) were further validated by concentration-dependent analysis.
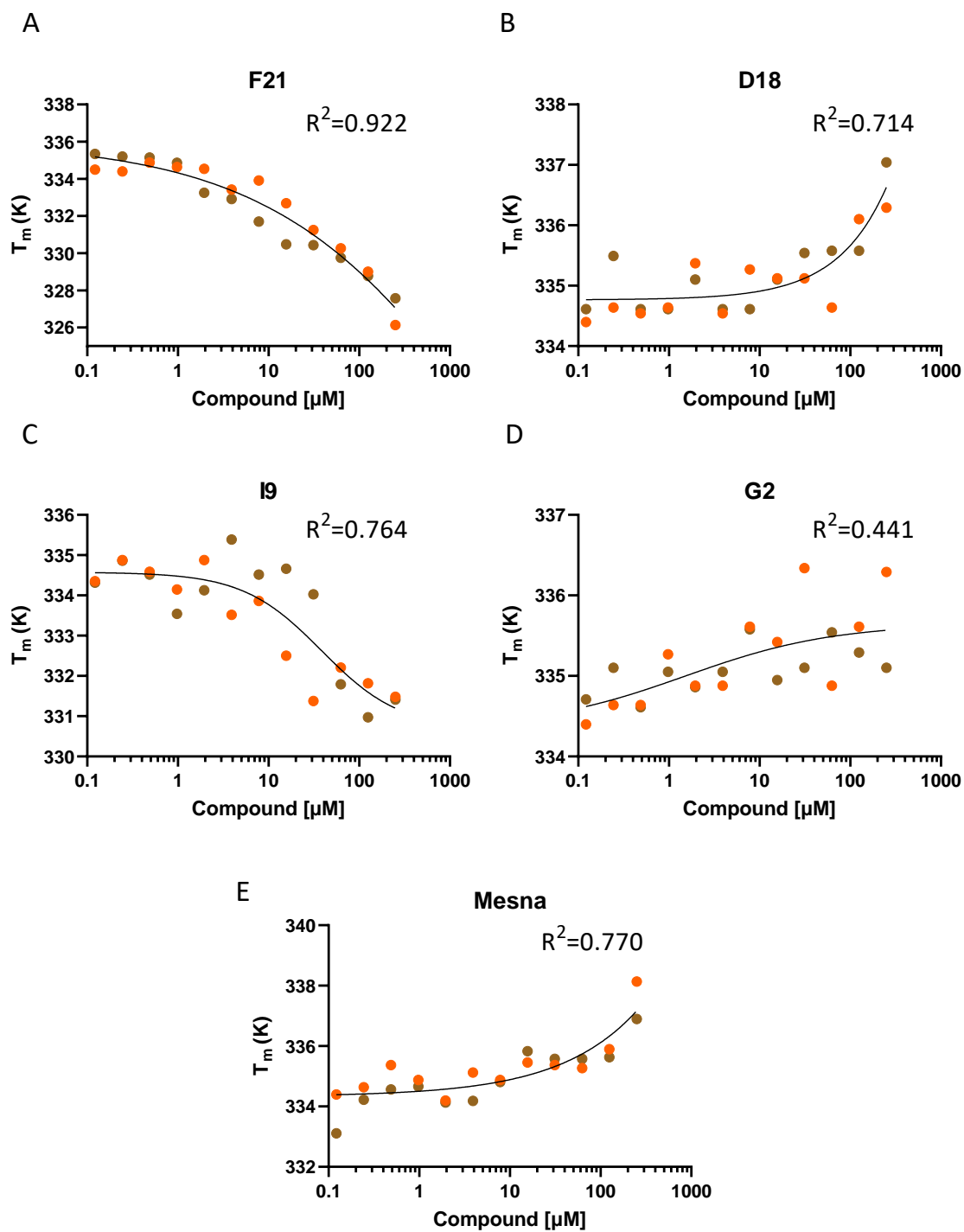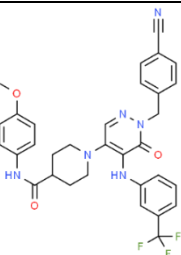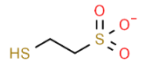
Figure 17: Graphs of the compounds selected for further experiments, where the compound concentration is plotted against the melting temperatures. The duplicates of this experiment are represented by two different colors of the circles.

The melting temperature versus the compound concentration was plotted and curve fitting was performed. The program HTSDSF Explorer was used to calculate the melting temperature and the dissociation constant (K_D). Fig. 17 shows the data for the compounds that were selected for further experiments. This selection was based on graphs with low K_D (see Table 2) and a high coefficient of determination (R^2). Wells F21 and I9 had decreasing melting temperatures with increasing concentrations and are classified as destabilizers, whereas mesna and compounds from wells D18 and G2 are stabilizers. Table 2 shows chemical names of the compounds that were excessively long, therefore compounds are instead referred to by their well number, with the exception of mesna.
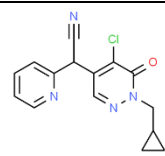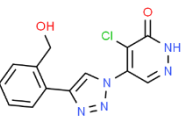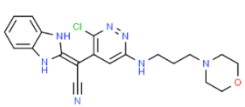
Table 3: Chemical name, structure, ΔT_m, dissociation constant and structural similarity of selected hits compared to CSA. Structures are adapted from website ChemSpider.

Compound	Chemical Name	Structure	ΔT_m (°C)	K_D (µM)	Similarity compared to CSA
Well F21[a]	[5-Chloro-1-(cyclopropylmethyl)-6-oxo-1,6-dihydro-4-pyridazinyl](2-pyridinyl)acetonitrile		-5.62	2.78	3%
Well I9[a]	4-Chloro-5-{4-[2-(hydroxymethyl)phenyl]-1H-1,2,3-triazol-1-yl}-3(2H)-pyridazinone		-4.94	7.14	3.7%
Well D18[a]	(3-Chloro-6-{[3-(4-morpholinyl)propyl]amino}-4-pyridazinyl)(1,3-dihydro-2H-benzimidazol-2-ylidene)acetonitrile		2.01	211.18	1%
Well G2[a]	1-[1-(4-Cyanobenzyl)-6-oxo-5-{[3-(trifluoromethyl)phenyl]amino-1,6-dihydro-4-pyridazinyl]-N-(4-methoxyphenyl)-4-piperidinecarboxamide		2.59	$1.10e^{-1}$	3.5%
Well H5[b]	2-mercaptoethanesulfonate (mesna)		6.43[c]	138.87	8.3%

[a] Prestwick Original Molecules library

[b] Prestwick FDA approved chemical library plate 4

[c] measures by co-worker

4.9 Activity assay using a fluorescence plate reader (TECAN SPARK)

Activity assay using HPLC was a time-consuming method since one sample took 15 min to analyze, and analysis of several samples took several hours. This resulted in the instability of OPA, and therefore a fluorescence plate reader (Tecan Spark) has been tried to measure the specific activity of CSAD. Hypotaurine with known concentrations was analyzed on Tecan Spark for a standard curve (see Fig. 18A). The graph shows linearity till 250 µM; therefore, only the linear part of the graph was selected for calculating the specific activity of CSAD.

A   B
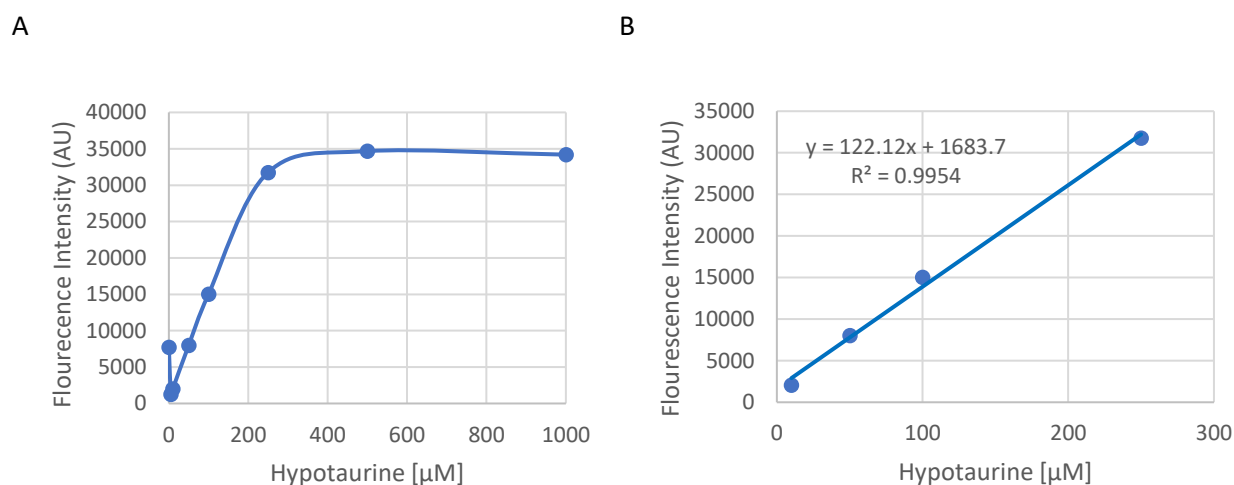
*Figure 18: **A)** standard curve of hypotaurine was prepared by analyzing 0 µM, 5 µM, 10 µM, 50 µM, 100 µM, 250 µM, 500 µM, and 1000 µM hypotaurine in 50 mM Na-phosphate buffer (pH 6) on Tecan Spark. **B)** Linear part of Figure 18A.*

The disadvantage of not using the HPLC is the lack of separation of the product from the substrate and all other remaining reagents of the reaction mix. Thus, further tests needed to be performed to make sure that the fluorescence signal was due to OPA interacting with hypotaurine. Therefore, the fluorescence intensity of the samples described in section 4.5 was measured every minute for 20 min.
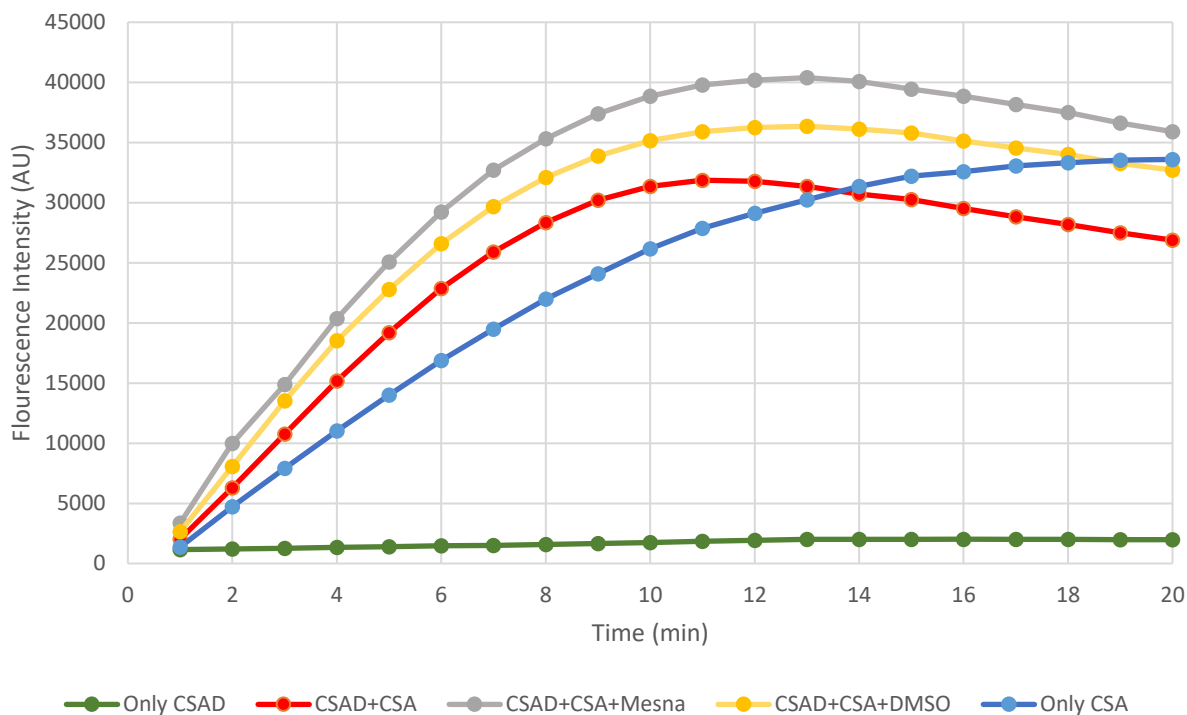
Figure 19: Fluorescence intensity measured every minute after starting the reaction of triplicates of five different samples. The average fluorescence intensity of triplicates was plotted.

Fig. 19 shows that mesna had the highest fluorescence intensity of all the other samples. The fluorescent intensity produced by the sample that contained only CSAD was the lowest. Moreover, CSA had relatively high signals. These high fluorescent signals were difficult to eliminate from the samples where both CSA and CSAD were present. Therefore, it was difficult to calculate the fluorescence intensity produced by hypotaurine. Another problem with this experiment was that the fluorescence signals in Fig. 19 were out of range of the linear part of the standard curve shown in Fig. 18B. Therefore, higher concentrations of hypotaurine were needed.

4.10 OPA-hypotaurine stability test (HPLC)

The activity assay using Tecan Spark showed a high background from non-enzymatic increase in fluorescence intensity. Therefore, it was decided to optimize the HPLC-based assay protocol instead. Fig. 11 showed that the OPA-hypotaurine complex was not stable as the area from the same hypotaurine concentration was constantly changing. Therefore, OPA was prepared differently by adding 0.5 mL β-mercaptoethanol in 200 µL OPA.

A

B
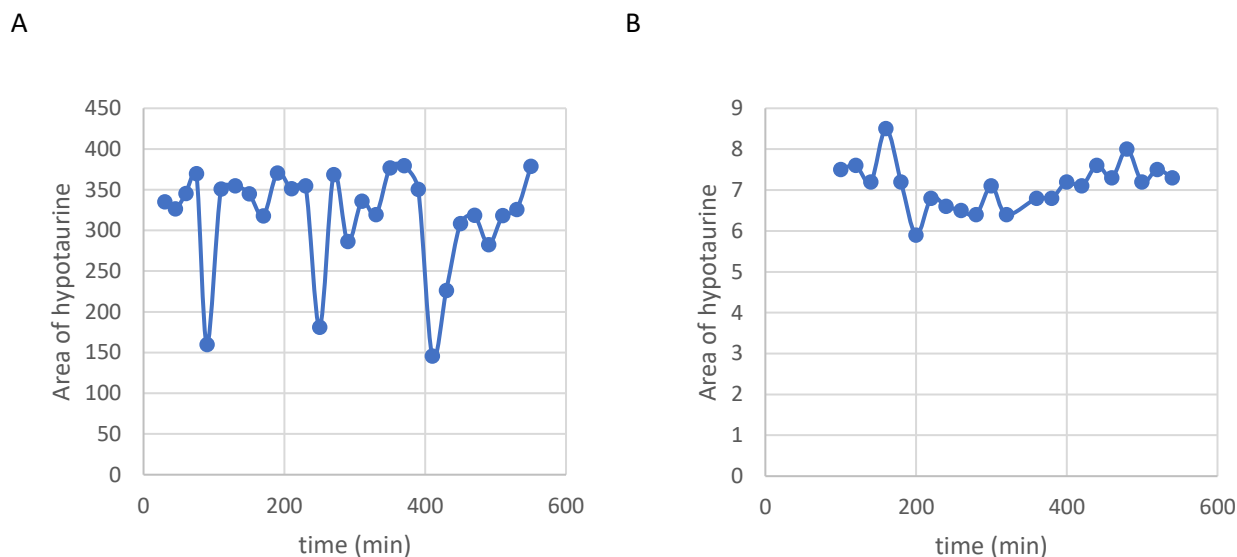
*Figure 20: OPA-hypotaurine conjugate stability test for 9 hours. 0.5 mL β-mercaptoethanol was added to 200 μL of OPA, and hypotaurine was prepared by diluting 5 mg/mL by 1: 500 in 50 mM Na-phosphate (pH 6). In **A)** acetic ethanol is not added, whereas **B)** contains acetic ethanol in equal amounts of hypotaurine.*

Fig. 20A shows the change in fluorescence signal of the OPA-hypotaurine conjugate over 9 hours. For this experiment, hypotaurine was prepared by diluting 5 mg/mL 1: 500 in 50 mM Na-phosphate (pH 6). The graph was quite stable but with few dips after some intervals. Fig. 20B shows the change in the hypotaurine area over 9 hours, where hypotaurine was prepared the same way as in Fig. 20A, but acetic ethanol was added in an equal amount of hypotaurine. The purpose of adding acetic ethanol was to prepare the standard the same way as the samples. As the graph in Fig. 20B looked more stable than Fig. 20A, it was decided to add acetic ethanol to the standards for further experiments.

4.11 Enzymatic activity assay in the presence of selected compounds

Enzymatic activity assay was performed on the selected compounds from hit validation using HPLC. The standard curve was prepared using the known concentrations of hypotaurine, which was later used for calculating the specific activity of CSAD. Fig. 21 shows the results from the activity assay. Three experiments were performed for each compound, and all the results from all three experiments were plotted, and then GraphPad was used to do the curve fitting. As shown in Fig. 21 compound F21 could be an inhibitor, whereas D18 could be an activator. The R^2 values of F21 and D18 obtained from curve fitting on all the three experiments was 0.2 and 0.4, respectively. However, R^2 of each experiment was much better. All the other compounds did not have much effect on the specific activity of CSAD.
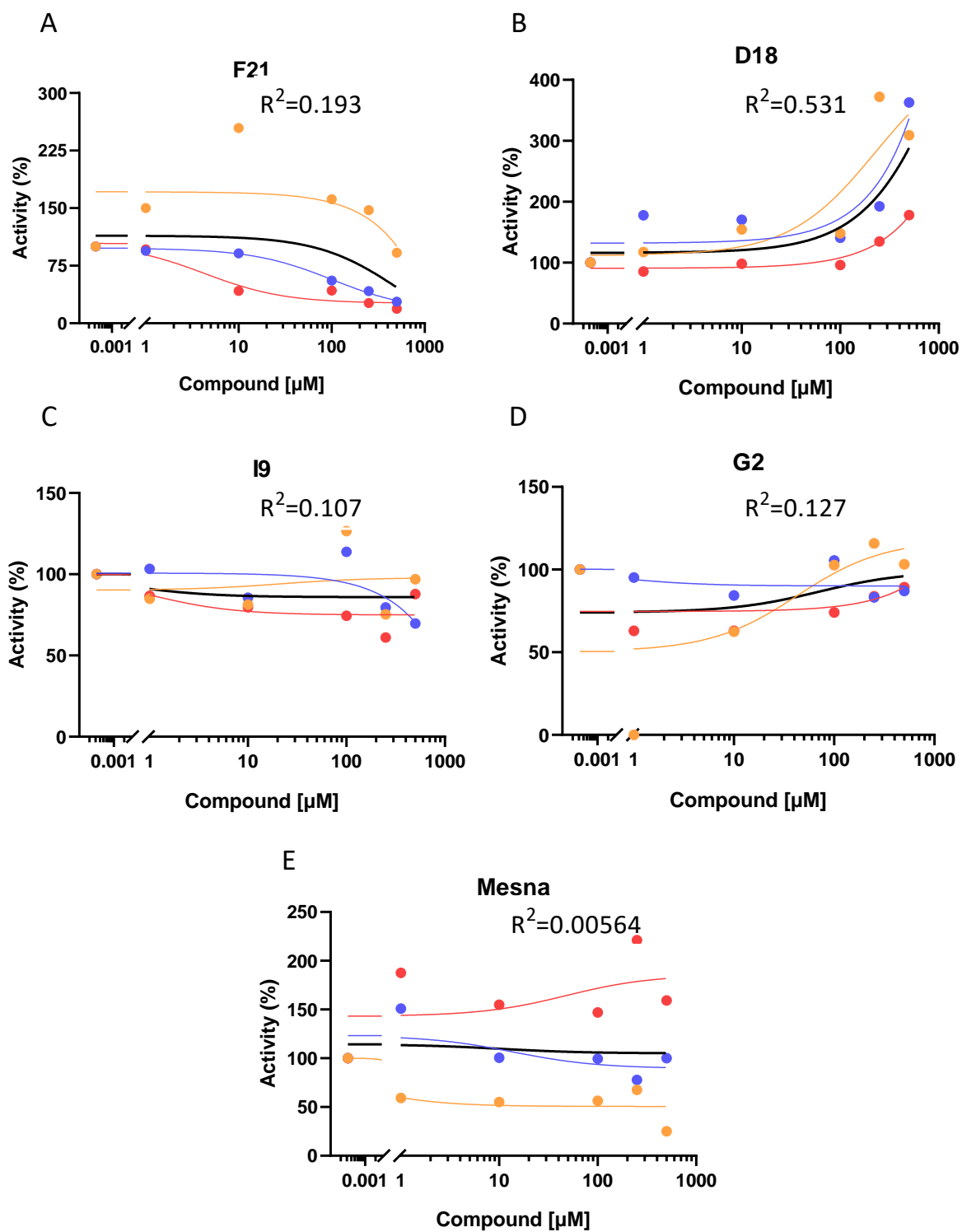
Figure 21: The activity of CSAD was tested in the presence of various compounds, and the data was analyzed using GraphPad Prism (Version 9.5.1). Three experiments were conducted for each compound, with each experiment represented by dots and circles of orange, blue, and red. The black line on the graph shows the curve fitting from all three experiments combined.

4.12 Structural similarity and molecular docking

Based on the data presented in Table 3, it appears that the structural similarity with CSA was quite low. The glide score from molecular docking of D18, F21, I9, G2 and mesna was -3.58 kcal/mol, -4.04 kcal/mol, -4.63 kcal/mol, -3.23 kcal/mol and -5.17 kcal/mol respectively. The results from Fig. 22 show that D18 interacts with CSAD by binding to Arg 105 and Phe 94, but F21 attaches to residue Arg 105. Compound I9 binds to CSAD on residues Leu 93, Gln 92 and Gln 312. The interaction between Mensa and CSAD occurs at Gln 92, Hie 304 and Llp 305, whereas G2 binds to CSAD at Hie 191 and Gln 92.

A

Compound F21
Glide score: -4.04 kcal/mol

B

Compound D18
Glide score: -3.58 kcal/mol

Figure 22: Molecular docking of compounds with CSAD.
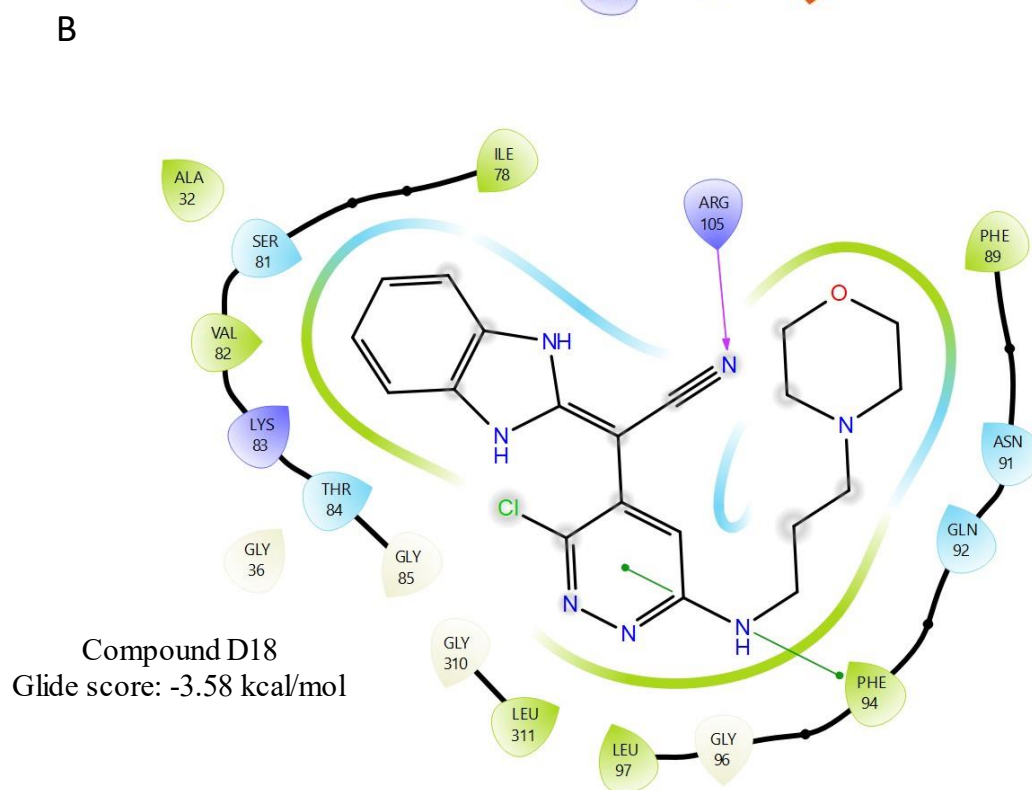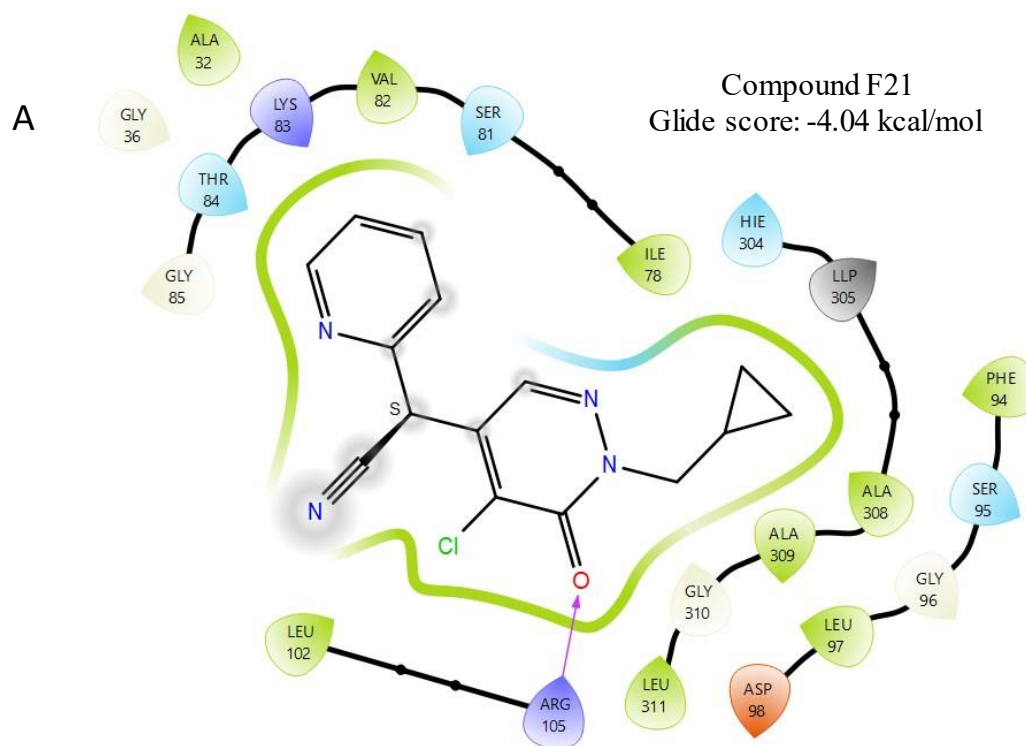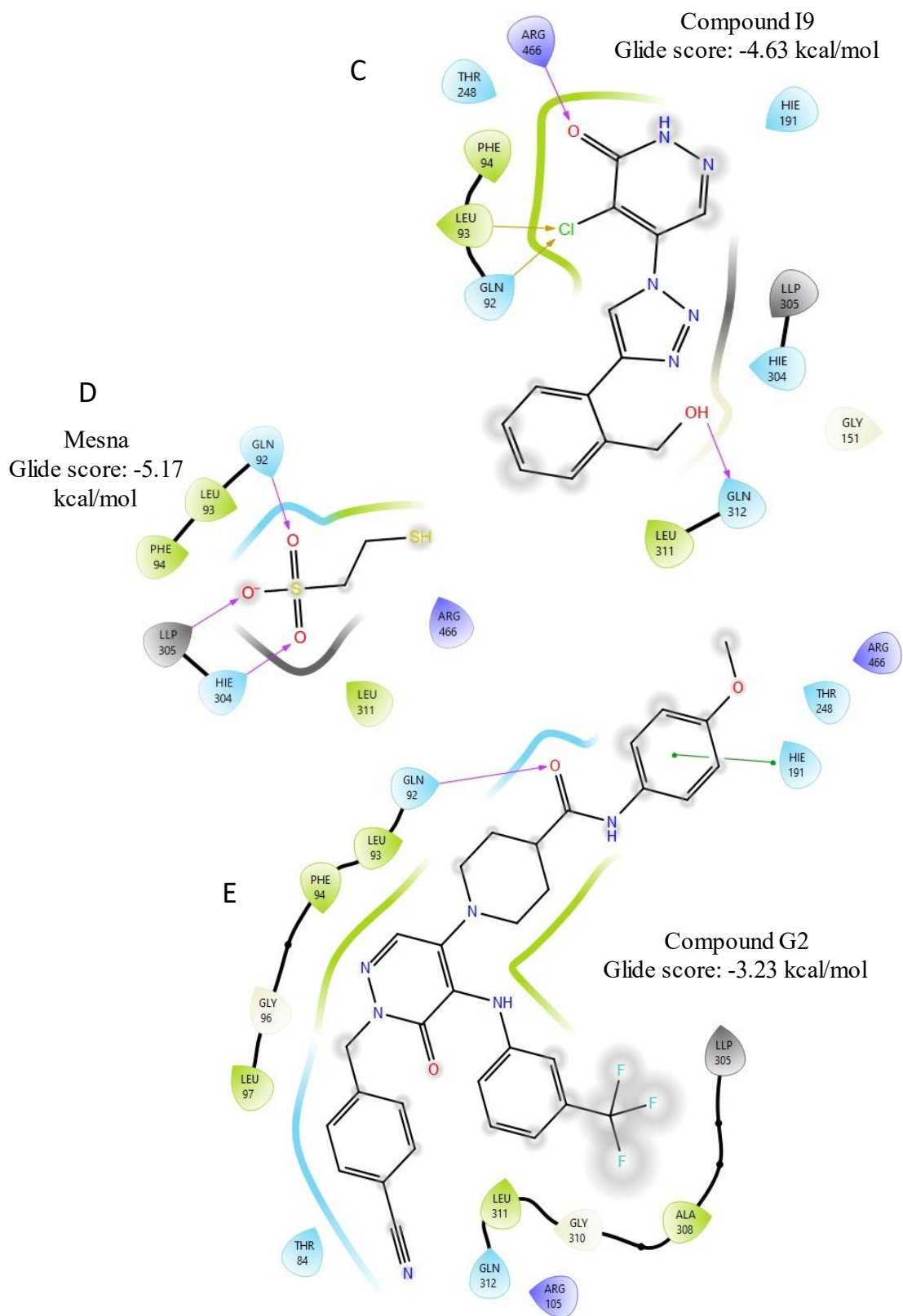
5. DISCUSSION

5.1 Challenging optimization of human CSAD expression and purification

After small-scale expression of human CSAD using different media, incubation time, and temperature, the SDS-PAGE (Fig. 6) showed that the TB media had most of the protein in the insoluble fraction, and more time for sonication was required. The autoinduction media did not produce CSAD at all, the reason is most likely that the used overnight pre-culture was incubated for more than 20 hours at 37°C. This increased incubation time could have led to the overgrowth of bacteria, resulting in the consumption of all the nutrients from the media and the secretion of toxic metabolites, which could have caused the whole bacteria population to die. LB media with 18 hours of incubation at 25°C (Fig. 6A, lane 4) were the best growth conditions for human CSAD, and these conditions were later used for large-scale protein expression.

After large-scale expression, human CSAD was purified using TALON affinity chromatography. Different imidazole concentrations in binding, washing, and elution buffers were tried. Still, the protein yield was insufficient for further experiments (Table 2). Imidazole is added to the elution buffer to elute the protein. In contrast, the purpose of adding imidazole in binding and wash buffer is to remove protein that has weak binding with the column resin. So, theoretically, the protein yield after reducing the imidazole concentration from the wash buffer was supposed to become higher, but instead, it decreased. This could indicate that the protein expression or sonication was not optimal. When testing the small-scale protein expression, sonication was performed for just 45 s because the culture had a small volume, whereas, during large-scale expression, 5 min were used in total for sonication. There is a possibility that the protein during large-scale expression was not sonicated long enough. Thus, bacteria cell membranes were not adequately broken apart to release the protein, or proteins were thermally denatured during sonication, in both cases leaving proteins in the pellet. It is also possible that overexpressed protein formed insoluble aggregates, also called inclusion bodies. Another problem could be aggregation of the protein during purification, preventing it from binding to the column. The possible reason for it can be the incorrect buffer condition, such as the ionic strength of the buffers. Reducing agents, detergents, or urea could have been used to eliminate the aggregation.

An established protocol for protein expression and purification of mouse CSAD was received from a co-worker [60]. That protocol was used for both human CSAD and mouse CSAD (Fig. 7). Even

though the protein sequence of these variants of CSAD has 95.3% sequence similarity, the results from protein purification from these had a big difference. The protein yield of human CSAD was much lower than mouse CSAD. Since there was only limited time for lab experiments, it was decided to work with just mouse CSAD.

5.2 Successful protein expression and purification of mouse CSAD

Using the protocol obtained from Dr. Elaheh Mahootchi who previously worked in the research group, mouse CSAD was purified using IMAC chromatography, but the protein yield was still too low in order to perform further experiments. There were two known reasons behind the poor protein yield. The first reason was that the protein was lost during washing the TALON resin. The other one could have been that the imidazole concentration in the elution buffer was too low, leading to the presence of His_6-tagged protein attached to the column even after elution. Therefore, adjustments in imidazole concentration were made in the buffers. Imidazole was not added in the wash buffer, whereas in the elution buffer, the concentration was increased. This resulted in a much higher protein yield, and the SDS-PAGE in Fig. 8 (lane 13) confirms that the protein was pure. However, some weak bands could be seen at lanes 13 and 14 due to the protein overloading while adding the sample into the well during electrophoresis.

Only a small amount of His_6-CSAD was tested for TEV cutting after purifying CSAD by using TALON crude. As shown in Fig. 9, the TEV cutting worked successfully by cutting the His_6-tag from CSAD. A study by Winge, *et al*., 2015, showed that the fusion protein and cleaved protein had similar K_m values, but the V_{max} values were highest for the cleaved protein [53]. Since the protein yield becomes low after TEV, it was decided not to cut the His_6-tag from the protein. Another thought behind not cutting the TEV was that it was unknown if the cleaved protein was stable. Therefore, size-exclusion chromatography was performed right after the TALON crude to remove aggregates and to get pure protein.

5.3 Melting temperature and activity assay of CSAD under varying conditions

DSF was used to find the melting temperature of CSAD. Optimal DSF conditions were determined by using varying concentrations of CSAD and SYPRO Orange. The selection of the best condition is based on the sharpness of the peak. Well-folded proteins with cooperative unfolding transitions generally exhibit sharp peaks. Broad peaks, on the other hand, may indicate that the protein has undergone misfolding transitions or being just partially unfolded. The graphs with high arbitrary

fluorescence units at the start of the reaction are also not selected because those samples indicate that the protein was already unfolded from the start. Moreover, different SYPRO Orange concentrations could also affect the peaks. Higher concentrations can increase the signal-to-noise ratio, whereas low concentrations may not give adequate signals to determine T_m-values. For all these reasons, T_m measurement based on 0.1 mg/mL CSAD with 5x SYPRO Orange (Fig. 10) was chosen because the peak was sharp, starting with low fluorescence intensity and without background noise. Therefore, the T_m value from that peak seems more accurate and reliable than the others.

After DSF, varying CSA concentration and incubation time were tested to determine the optimal condition for CSAD's enzymatic activity assay. Generally, too low substrate concentration and too short incubation time can cause the reaction rate to be slow and incomplete, whereas high concentration and long incubation times can cause saturated reactions. As Fig.12 shows the CSAD specific activity was highest when CSA concentration was highest but rapidly decreased with increasing incubation time Therefore, it was decided to use 2 mM CSA and 1 minute of incubation time for further experiments. The standard deviation between the duplicates was quite high sometimes, it could be possible because of unstable OPA. However, standards were analyzed several times during the experiment to undergo this problem. As Fig. 11 shows the area of hypotaurine varies quite a lot with increasing time and one sample was supposed to take around 15 min which could have led to unstable OPA over time.

5.4 No significant effect of addition of PN during protein expression on CSAD's melting temperature and specific activity

PN is a precursor to PLP and was added during protein expression to ensure that sufficient PLP was available to support the proper folding, stability and catalytic activity of CSAD. In the presence of PN, the melting temperature was slightly higher than without PN (T_m with PN = 61.3 ± 0.54°C; T_m without PN = 60.6 ± 0.03°C) (Fig. 13). However, the effect of PN on the melting temperature was not significant (p=0.16). The specific activity was also tested with and without adding PN during the protein expression. Notably, in both conditions an excess of PLP was added during the enzyme activity assay. Based on the results, it appears that PN did not have a significant impact on CSAD stability or yield. This also suggests that the concentration of PLP that was added during the assay preparation was enough to achieve the optimal activity of CSAD. The experiments on the

effects of PN need to be replicated before any firm conclusions can be made. However, based on the preliminary evidence, it was decided to add PN for further experiments to improve CSAD's stability and enzymatic activity.

5.5 Screening and hit validation of Prestwick Original Molecules library identified five compounds of interest

DSF was chosen for screening and hit validation of compounds from the Prestwick Original Molecule library. The selection of this library is discussed in 1.5.6. Another reason for choosing this library was the low CSAD protein yield which made the screening of larger library impossible.

When a compound binds to a protein, it may change its thermal stability. Therefore, stabilizers and destabilizers can be potential candidates as enzyme activators or inhibitors. The reason behind using DSF for compound screening was that this method can perform high throughput screening, analyzing a complete compound library on a 384 well plate in just one hour. The method is relatively cheap and sensitive, as the ligand-protein binding can be detected as a shift in thermal stability of the protein even in the presence of low concentrations of protein and low molecular weight compounds. On the other hand, a disadvantage can be the false signals, if the compound interacts with SYPRO Orange instead of protein, causing a change in T_m. Fluorescent compounds can also be problematic as they will interfere with the dye's signal. In addition, this method cannot identify specific or nonspecific ligand-protein binding [78].

Using different concentrations of compounds, it was possible to calculate K_D-values, which can indicate how strongly the ligand is binding to the protein. However, most of the hit validation results had R^2 lower than 0.9, indicating that the T_m of the samples was not proportional with increasing compound concentration. The possible reason can be errors while pipetting, since very small volumes (such as 0.2 µL) of buffer, CSAD and compound were added. Additionally, there could be bubbles or particles that could have affected the fluorescence intensity and noise levels, or instability of protein. Another reason can be the light sensitivity of SYPRO Orange, causing degradation if exposed to light and leading to reduced fluorescence intensity. Although the data was noisy, a clear concentration dependent trend of increased or decreased melting temperature was observed for several compounds and those were selected for activity assays (Fig. 17).

5.6 Method development and optimization of CSAD enzymatic activity assay

As discussed earlier, OPA-hypotaurine conjugate was quite unstable over extended periods of time and as HPLC is a time-consuming method, it was decided to test a fluorescence plate reader (Tecan Spark) to measure enzymatic activity. The advantage of using Tecan Spark is that it measures fluorescence signals from several samples at once, and it requires only a few seconds to analyze the samples. The Tecan Spark instrument is highly sensitive, as it is capable of detecting even the low levels of fluorescence [79]. Additionally, this method is less expensive compared to HPLC, as it does not require the use of buffers for the mobile phase. However, the results from Fig. 19 showed that CSA itself had high fluorescence intensity, making it challenging to distinguish the signals created by hypotaurine. Consequently, the attempt to determine the enzymatic activity of CSAD with Tecan Spark was unsuccessful, and HPLC was utilized to measure CSAD's activity instead. One of the major benefits of using HPLC is that it can separate molecules, making it easier to distinguish between the substrate, product and other reagents present in the assay. However, a major problem with the HPLC based acitiviy assay was the instability of OPA.

During the preparation of OPA, the amount of β-mercaptoethanol was reduced. When β-mercaptoethanol is added to OPA, the intention is to reduce its oxidation rate, which causes instability over time. The decision to reduce the amount of β-mercaptoethanol was made based on information from the data sheet of OPA P0532 provided by Sigma-Aldrich. That data sheet stated that excessive amounts of β-mercaptoethanol can lead to a decrease in reagent sensitivity [80]. Moreover, 5% acidic ethanol was added to the standards, so that the standards are handled in the same manner as the samples. Although the addition of acidic ethanol decreased the signals drastically, the signals remained quite consistent and stable (Fig. 20).

5.7 Compound F21 as an inhibitor and D18 as an activator of CSAD

The results from the enzymatic activity of CSAD indicated that [5-Chloro-1-(cyclopropylmethyl)-6-oxo-1,6-dihydro-4-pyridazinyl](2-pyridinyl)acetonitrile, a compound from well F21 from Prestwick Original Molecules library can be a weak inhibitor of CSAD activity. Moreover, (3-Chloro-6-[3-(4-morpholinyl)propyl]amino-4-pyridazinyl)(1,3-dihydro-2H-benzimidazol-2-ylidene)acetonitrile, a compound from well D18 from Prestwick Original Molecules library was an activator of CSAD. The R^2 values of both F21 and D18 were quite low, but there was a noticeable concentration dependent trend in Fig. 21. These results also showed that HPLC activity

assay did not have good reproducibility. Although the signal of the OPA derivative of hypotaurine was quite stable in Fig. 20, there were still some dips in the graph, indicating that either OPA was the problem or some other unknown error.

The activity assay was performed three times, but with different protein batches of CSAD, which could also have impacted the reproducibility. All these protein batches were prepared the same way but there can still be some factors like aggregation and degeneration causing protein stability, which impacted specific activity. Moreover, small volumes of reagents, compounds, substrate, and protein were pipetted while preparing the assay, which could have led to experimental errors. During the assay preparation, CSA was added to the samples and incubated for 1 min. Acetic ethanol was then added. However, there may have been slight timing variations in each experiment, which could have resulted in variation in observed specific activity.

The samples were centrifuged before analyzing it on HPLC. If cell debris or particles were still present in the sample, it could have caused blockage of injection needle of HPLC, reducing the reproducibility of the results. The presence of these particles might affect the fluorescence readings as well.

Based on the Morgen fingerprints analysis, it appears to be little structural similarity between CSA and the chosen compounds (Table 3). Molecular docking also revealed low glide score, indicating that the compounds have low affinity towards CSAD. The binding site of CSAD was found to have Gln 92 as a key residue, and the molecular docking of I9, G2 and mesna showed binding at the same residue, suggesting that these molecules can bind to the CSAD binding site (Fig. 22). However, mesna had the highest structural similarity with CSA, lowest glide score among the other compounds and a binding at Gln 92, which predicted that mesna could bind to CSAD a bit stronger than the other compounds. However, in the concentration range tested mesna had no clear effect on the specific activity of CSAD.

Furthermore, it should be noted that neither compound F21 nor D18 appear to bind to the same residue where CSA binds. However, both compounds have an acetonitrile group and a pyridazine ring, which can indicate that these compounds may not bind to the active site but to another (allosteric) site of the enzyme. This hypothesis may be tested in future experiments using X-ray crystallography or spectroscopic techniques.

6. CONCLUSION AND FUTURE RESEARCH

Most of the aims of the project have been achieved. The protein expression and purification of human CSAD was tried to optimize by using small-scale expression and by trying different imidazole concentration in buffers. However, the protein yield was insufficient to perform for further experiments. As there was a limited time frame for this project, a decision to proceed with mouse CSAD was made. Furthermore, the sequence alignment between the mouse CSAD and human CSAD revealed a high degree of similarity at 95.3%. This suggests that compounds capable of binding to mouse CSAD may also bind to human CSAD.

Mouse CSAD was successfully expressed and purified. However further research is needed to understand the effect of adding PN during protein expression. In addition, the protein stability and enzymatic activity needs to be tested without His_6-tag.

Protein stability was determined by using DSF in the presence of compounds from the Prestwick Original Molecules library. Dose-response series indicated five compounds that partially stabilized or destabilized CSAD. Enzymatic activity assays showed a weak inhibitory effect of [5-Chloro-1-(cyclopropylmethyl)-6-oxo-1,6-dihydro-4-pyridazinyl](2-pyridinyl)acetonitrile (a compound from well F21 from Prestwick Original Molecules library). Moreover, (3-Chloro-6-[3-(4-morpholinyl)propyl]amino-4-pyridazinyl)(1,3-dihydro-2H-benzimidazol-2-ylidene)acetonitrile (a compound from well D18 from Prestwick Original Molecules library) was a weak activator or stabilizer of CSAD. Molecular docking of these compounds against the active site of CSAD indicated that these compounds may not bind to the active site but possibly to another site of the CSAD. Furthermore, R^2 from the curve fitting of specific activity was very low, making it impossible to calculate IC_{50} and EC_{50} values. Therefore, the experiment will need to be repeated to confirm that these compounds are inhibitors or activators. In addition, the interaction between CSAD and these compounds can be studied using X-ray crystallography or spectroscopic techniques in the future.

Access to inhibitors and activators of CSAD, maybe used to alter the production of taurine in the human body. As previously discussed, taurine is known to play important roles in neurotransmission, and a change in its level can lead to increased or decreased oxidative stress, mitochondrial function, inflammation, excitotoxicity, and apoptosis. Thus, an activator or an inhibitor of CSAD can indirectly also impact neurotransmission and potentially be used as a

therapeutic agent against metabolic and neurological disorders like Parkinson's, Alzheimer's, and Huntington's disease [40].

REFERENCES

1. Pillarisetty, Z.M.S., R. Vamsi, and S. Leela, *Physiology, Neurotransmitters.* 2022.

2. Haines, D.E. and G.A. Mihailoff, *Fundamental neuroscience for basic and clinical applications* 2018: Elsevier Health Sciences.

3. *Brain Basics: Know Your Brain.* 2023; Available from: https://www.ncbi.nlm.nih.gov/pubmed/.

4. Alberts, B., J.H. Wilson, and T. Hunt, *Molecular biology of the cell.* 6th ed. 2015, New York: Garland Science.

5. Christman, O.R.f.N., E. Kimberly, and Elizabeth, *Chapter 7 Depressive Disorders.* 2022.

6. E. Hyman, S., *Neurotransmitters.* Current Biology, 2005. **15**(5).

7. Purves, D., et al., *Neuroscience.* International sixth edition. ed. 2018, New York: Sinauer Associates.

8. Ripps, H. and W. Shen, *Review: taurine: a "very essential" amino acid.* Mol Vis, 2012. **18**: p. 2673-86.

9. Cabrera, D., J.P. Arab, and M. Arrese, *UDCA, NorUDCA, and TUDCA in Liver Diseases: A Review of Their Mechanisms of Action and Clinical Applications*, in *Bile Acids and Their Receptors*, S. Fiorucci and E. Distrutti, Editors. 2019, Springer International Publishing: Cham. p. 237-264.

10. Quistad, G.B., L.E. Staiger, and D.A. Schooley, *Xenobiotic conjugation: a novel role for bile acids.* Nature, 1982. **296**(5856): p. 462-464.

11. Kilb, W. *Putative Role of Taurine as Neurotransmitter During Perinatal Cortical Development.* in *Taurine 10.* 2017. Dordrecht: Springer Netherlands.

12. Kilb, W. and A. Fukuda, *Taurine as an Essential Neuromodulator during Perinatal Cortical Development.* Front Cell Neurosci, 2017. **11**: p. 328.

13. Wu, J.-Y., et al., *Regulation of Taurine Biosynthesis and Its Physiological Significance in the Brain*, in *Taurine 3: Cellular and Regulatory Mechanisms*, S. Schaffer, J.B. Lombardini, and R.J. Huxtable, Editors. 1998, Springer US: Boston, MA. p. 339-345.

14. Tramonti, A., et al., *A Novel, Easy Assay Method for Human Cysteine Sulfinic Acid Decarboxylase.* Life, 2021. **11**(5): p. 438.

15. Oja, S.S. and P. Saransaari, *Taurine and epilepsy.* Epilepsy Res, 2013. **104**(3): p. 187-94.

16. Shao, F.-b., et al., *Anxiolytic effect of GABAergic neurons in the anterior cingulate cortex in a rat model of chronic inflammatory pain.* Molecular Brain, 2021. **14**(1): p. 139.

17. Rafiee, Z., A.M. García-Serrano, and J.M.N. Duarte, *Taurine Supplementation as a Neuroprotective Strategy upon Brain Dysfunction in Metabolic Syndrome and Diabetes.* Nutrients, 2022. **14**(6): p. 1292.

18. Ramírez-Guerrero, S., et al., *Taurine and Astrocytes: A Homeostatic and Neuroprotective Relationship.* Frontiers in Molecular Neuroscience, 2022. **15**.

19. Jong, C.J., P. Sandal, and S.W. Schaffer, *The Role of Taurine in Mitochondria Health: More Than Just an Antioxidant.* Molecules, 2021. **26**(16).

20. Leonelli, M., A.S. Torrão, and L.R. Britto, *Unconventional neurotransmitters, neurodegeneration and neuroprotection.* Braz J Med Biol Res, 2009. **42**(1): p. 68-75.

21. Wu, J.Y. and H. Prentice, *Role of taurine in the central nervous system.* J Biomed Sci, 2010. **17 Suppl 1**(Suppl 1): p. S1.

22. Lin, C.T., G.X. Song, and J.Y. Wu, *Is taurine a neurotransmitter in rabbit retina?* Brain Res, 1985. **337**(2): p. 293-8.

23. Lin, C.T., G.X. Song, and J.Y. Wu, *Ultrastructural demonstration of L-glutamate decarboxylase and cysteinesulfinic acid decarboxylase in rat retina by immunocytochemistry.* Brain Res, 1985. **331**(1): p. 71-80.

24. Okamoto, K., H. Kimura, and Y. Sakai, *Taurine-induced increase of the Cl-conductance of cerebellar Purkinje cell dendrites in vitro.* Brain Research, 1983. **259**(2): p. 319-323.

25. Taber, K.H., et al., *Taurine in hippocampus: localization and postsynaptic action.* Brain Res, 1986. **386**(1-2): p. 113-21.

26. Wu, J.Y., X.W. Tang, and W.H. Tsai, *Taurine Receptor: Kinetic Analysis and Pharmacological Studies*, in *Taurine: Nutritional Value and Mechanisms of Action*, J.B. Lombardini, S.W. Schaffer, and J. Azuma, Editors. 1992, Springer US: Boston, MA. p. 263-268.

27. Frosini, M., et al., *A specific taurine recognition site in the rabbit brain is responsible for taurine effects on thermoregulation.* Br J Pharmacol, 2003. **139**(3): p. 487-94.

28. Saransaari, P. and S.S. Oja, *Taurine in Neurotransmission*, in *Handbook of Neurochemistry and Molecular Neurobiology: Neurotransmitter Systems*, A. Lajtha and E.S. Vizi, Editors. 2008, Springer US: Boston, MA. p. 325-342.

29. Zhao, X.-Y., et al., *Mitochondrial Dysfunction in Neural Injury.* Frontiers in Neuroscience, 2019. **13**.

30. Abad, M.F., et al., *Mitochondrial pH monitored by a new engineered green fluorescent protein mutant.* J Biol Chem, 2004. **279**(12): p. 11521-9.

31. Azarias, G., et al., *Glutamate transport decreases mitochondrial pH and modulates oxidative metabolism in astrocytes.* J Neurosci, 2011. **31**(10): p. 3550-9.

32. Suzuki, T., et al., *Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases.* Embo j, 2002. **21**(23): p. 6581-9.

33. Schaffer, S.W., et al., *Role of taurine in the pathologies of MELAS and MERRF.* Amino Acids, 2014. **46**(1): p. 47-56.

34. Baliou, S., et al., *Protective role of taurine against oxidative stress (Review).* Mol Med Rep, 2021. **24**(2).

35. Aruoma, O.I., et al., *The antioxidant action of taurine, hypotaurine and their metabolic precursors.* Biochem J, 1988. **256**(1): p. 251-5.

36. Kim, C. and Y.-N. Cha, *Taurine chloramine produced from taurine under inflammation provides anti-inflammatory and cytoprotective effects.* Amino Acids, 2014. **46**(1): p. 89-100.

37. Kumari, N., H. Prentice, and J.-Y. Wu. *Taurine and Its Neuroprotective Role*. in *Taurine 8*. 2013. New York, NY: Springer New York.

38. Menzie, J., H. Prentice, and J.Y. Wu, *Neuroprotective Mechanisms of Taurine against Ischemic Stroke.* Brain Sci, 2013. **3**(2): p. 877-907.

39. Bhat, M.A., et al., *Expedition into Taurine Biology: Structural Insights and Therapeutic Perspective of Taurine in Neurodegenerative Diseases.* Biomolecules, 2020. **10**(6).

40. Menzie, J., et al., *Taurine and central nervous system disorders.* Amino Acids, 2014. **46**(1): p. 31-46.

41. Oppici, E., et al., *Pyridoxamine and pyridoxal are more effective than pyridoxine in rescuing folding-defective variants of human alanine:glyoxylate aminotransferase causing primary hyperoxaluria type I.* Human Molecular Genetics, 2015. **24**(19): p. 5500-5511.

42. L., d.S.M., M.K. Safo, and Contestabile, *Biomedical aspects of pyridoxal 5'-phosphate availability.* Frontiers in bioscience (Elite edition), 2012. **4**(3).

43. Abosamak, N.R. and V. Gupta, *Vitamin B6 (Pyridoxine)*. 2023, StatPearls: Treasure Island (FL).

44. Percudani, R. and A. Peracchi, *A genomic overview of pyridoxal-phosphate-dependent enzymes.* EMBO Rep, 2003. **4**(9): p. 850-4.

45. Eliot, A.C. and J.F. Kirsch, *Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations.* Annual Review of Biochemistry, 2004. **73**(1): p. 383-415.

46. Daidone, F., et al., *In silico and in vitro validation of serine hydroxymethyltransferase as a chemotherapeutic target of the antifolate drug pemetrexed.* European Journal of Medicinal Chemistry, 2011. **46**(5): p. 1616-1621.

47. Paiardini, A., et al., *PLP-Dependent Enzymes.* BioMed Research International, 2014. **2014**: p. 856076.

48. Daidone, F., et al., *Identification by Virtual Screening and In Vitro Testing of Human DOPA Decarboxylase Inhibitors.* PLOS ONE, 2012. **7**(2): p. e31610.

49. Sarup, A., O.M. Larsson, and A. Schousboe, *GABA Transporters and GABA-Transaminase as Drug Targets.* 2003.

50. Heinämäki, A.A., A.K. Perämaa, and R.S. Piha, *Characterization of cerebral cysteine sulfinic acid decarboxylase. Molecular parameters and inhibition studies.* Acta Chem Scand B, 1982. **36**(5): p. 287-90.

51. Park, E., et al., *Development of a Novel Cysteine Sulfinic Acid Decarboxylase Knockout Mouse: Dietary Taurine Reduces Neonatal Mortality.* Journal of Amino Acids, 2014. **2014**: p. 346809.

52. Sturman, J.A., *Taurine in development.* Physiological reviews, 1993. **73**(1): p. 119-147.

53. Winge, I., et al., *Mammalian CSAD and GADL1 have distinct biochemical properties and patterns of brain expression.* Neurochemistry International, 2015. **90**: p. 173-184.

54. Chan-Palay, V., S.L. Palay, and J.-Y. Wu, *Sagittal cerebellar microbands of taurine neurons: Immunocytochemical demonstration by using antibodies against the taurine-synthesizing enzyme cysteine sulfinic acid decarboxylase.* Proceedings of the National Academy of Sciences, 1982. **79**(13): p. 4221-4225.

55. Reymond, I., K. Almarghini, and M. Tappaz, *Immunocytochemical localization of cysteine sulfinate decarboxylase in astrocytes in the cerebellum and hippocampus: a*

quantitative double immunofluorescence study with glial fibrillary acidic protein and S-100 protein. Neuroscience, 1996. **75**(2): p. 619-633.

56.   Dominy, J., S. Eller, and R. Dawson, *Building Biosynthetic Schools: Reviewing Compartmentation of CNS Taurine Synthesis.* Neurochemical Research, 2004. **29**(1): p. 97-103.

57.   Vitvitsky, V., S.K. Garg, and R. Banerjee, *Taurine Biosynthesis by Neurons and Astrocytes*.* Journal of Biological Chemistry, 2011. **286**(37): p. 32002-32010.

58.   *Pairwise Sequence Alignment Tools*. 2023; Available from: https://www.ebi.ac.uk/Tools/psa/.

59.   *Clustal Omega*. [cited 2023; Available from: https://www.ebi.ac.uk/Tools/msa/clustalo/.

60.   Mahootchi, E., et al., *Structure and substrate specificity determinants of the taurine biosynthetic enzyme cysteine sulphinic acid decarboxylase.* bioRxiv, 2020: p. 2020.09.25.308866.

61.   Mahootchi, E., et al., *GADL1 is a multifunctional decarboxylase with tissue-specific roles in β-alanine and carnosine production.* Science Advances, 2020. **6**(29): p. eabb3713.

62.   Nosaki, S. and K. Miura, *Chapter Nine - Transient expression of recombinant proteins in plants*, in *Methods in Enzymology*, W.B. O'Dell and Z. Kelman, Editors. 2021, Academic Press. p. 193-203.

63.   Gräslund, S., et al., *Protein production and purification.* Nat Methods, 2008. **5**(2): p. 135-46.

64.   Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Frontiers in Microbiology, 2014. **5**.

65.   Overton, T.W., *Recombinant protein production in bacterial hosts.* Drug Discovery Today, 2014. **19**(5): p. 590-601.

66.   Shilling, P.J., et al., *Improved designs for pET expression plasmids increase protein production yield in Escherichia coli.* Communications Biology, 2020. **3**(1): p. 214.

67.   Francis, D.M. and R. Page, *Strategies to optimize protein expression in E. coli.* Curr Protoc Protein Sci, 2010. **Chapter 5**(1): p. 5.24.1-5.24.29.

68. Urh, M., D. Simpson, and K. Zhao, *Chapter 26 Affinity Chromatography: General Methods*, in *Methods in Enzymology*, R.R. Burgess and M.P. Deutscher, Editors. 2009, Academic Press. p. 417-438.

69. Hage, D.S., et al., *Pharmaceutical and biomedical applications of affinity chromatography: Recent trends and developments.* Journal of Pharmaceutical and Biomedical Analysis, 2012. **69**: p. 93-105.

70. *Talon Superflow histidine-tagged protein purification resin.* [cited 2023; Available from: https://www.cytivalifesciences.com/en/us/shop/chromatography/resins/affinity-tagged-protein/talon%C2%AE-superflow%E2%84%A2-histidine-tagged-protein-purification-resin-p-06007.

71. Kostanski, L.K., D.M. Keller, and A.E. Hamielec, *Size-exclusion chromatography—a review of calibration methodologies.* Journal of Biochemical and Biophysical Methods, 2004. **58**(2): p. 159-186.

72. Martin-Malpartida, P., et al., *HTSDSF Explorer, A Novel Tool to Analyze High-throughput DSF Screenings.* Journal of Molecular Biology, 2022. **434**(11): p. 167372.

73. Miotto, M., et al., *Insights on protein thermal stability: a graph representation of molecular interactions.* bioRxiv, 2018: p. 354266.

74. Wu, T., et al., *Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments.* bioRxiv, 2020: p. 2020.03.22.002543.

75. *What is HPLC (High Performance Liquid Chromatography) ？.* 2023 [cited 2023; Available from: https://www.shimadzu.com/an/service-support/technical-support/analysis-basics/basic/what_is_hplc.html.

76. Bisswanger, H., *Enzyme assays.* Perspectives in Science, 2014. **1**(1): p. 41-55.

77. *Prestwick Original Molecules Library - Prestwick Chemical Libraries.* 2023; Available from: https://www.prestwickchemical.com/screening-libraries/prestwick-original-molecules-library/.

78. Gao, K., R. Oerlemans, and M.R. Groves, *Theory and applications of differential scanning fluorimetry in early-stage drug discovery.* Biophysical Reviews, 2020. **12**(1): p. 85-104.

79. Ag, T.T. *Multimode microplate reader, Live cell assays.* 2023; Available from: https://lifesciences.tecan.com/multimode-plate-reader.

80.   *Phthaldialdehyde Reagent Solution Complete 643-79-8*. 2023; Available from: http://www.sigmaaldrich.com/.

APPENDIX

Table 4: Selected compounds from screening of Prestwick Original Molecules library

Well	Effect	T_m (°C)	ΔT_m (°C)	$T_{m \text{ (DMSO control)}}$ (°C)	$\pm SD_{\text{(DMSO control)}}$
I11	Destabilizing	54.47	-7.13	61.6	0.31
I3	Destabilizing	55.74	-5.86	61.6	0.31
F21	Destabilizing	55.98	-5.62	61.6	0.31
I9	Destabilizing	56.66	-4.94	61.6	0.31
O6	Destabilizing	57.15	-4.45	61.6	0.31
I5	Destabilizing	57.63	-3.97	61.6	0.31
H20	Destabilizing	57.63	-3.97	61.6	0.31
K3	Destabilizing	57.87	-3.73	61.6	0.31
P15	Destabilizing	58.36	-3.24	61.6	0.31
I13	Destabilizing	59.09	-2.51	61.6	0.31
H2	Destabilizing	59.38	-2.22	61.6	0.31
L2	Destabilizing	59.38	-2.22	61.6	0.31
N11	Destabilizing	59.57	-2.03	61.6	0.31
D18	Stabilizing	63.61	2.01	61.6	0.31
D9	Stabilizing	63.7	2.1	61.6	0.31
D7	Stabilizing	63.95	2.35	61.6	0.31
G2	Stabilizing	64.19	2.59	61.6	0.31