Design, identification, and validation of novel C-terminal binding protein modulators

Submitted in partial fulfilment

of the requirements of the

Degree of Master of Pharmacy

Nicole Xuereb

Department of Pharmacy

2023

Abstract

C-terminal binding proteins (CtBP) are a prospective therapeutic target in a number of cancers. These proteins are oncogenic transcriptional co-regulators that can decrease cell migration, eliminate cellular invasion, and enhance DNA repair, if they are inhibited. The aim of the study is to identify high affinity CtBP modulators with a propensity to *in vivo* bioavailability, using the experimental molecules 4-methylthio 2-oxobutyric acid (MTOB) and hydroxyimino-3-phenylpropanoic acid (HIPP), as the lead molecules. During virtual screening, MTOB and HIPP pharmacophores were modelled and the consensus pharmacophore was created by superimposition of the two pharmacophores. Hits were identified and drug- and lead-like filters were applied. A protomol was modelled and the hits were docked into it and ranked in order of affinity to the protomol, in order to identify the highest affinity lead-like molecule. Through virtual screening, 674 Lipinski rule compliant hit structures were acquired. 2D topology maps were then generated to guide seed structure modelling during the *de novo* design process where two successful seeds were created. The molecules generated by the seeds were categorized into pharmacophorically similar families in order of ligand binding affinity and then filtered according to Lipinski's rules. The 5 optimal molecules with regards to their log P and binding affinity were then assessed for toxicity. In conclusion, CtBP in isoform 1 and 2, has a role in future cancer therapy. The structural approaches used helped in the development of novel antineoplastic CtBP inhibitors because the leads used were viable.

Acknowledgments

I would like to convey my sincere gratitude and appreciation to the tutor of this dissertation, Dr. Claire Shoemake for her continuous assistance and support during the entire study.

I would also like to thank Prof. Lilian M. Azzopardi, Prof. Anthony Serracino Inglott and all the staff at the pharmacy department for their guidance throughout the past five years.

Finally, I would like to thank my family and friends, and all of the staff at the St. John's Pharmacy, especially the managing pharmacist, for their everlasting support and encouragement.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

CtBP-C-terminal Binding Protein

E1A - Adenovirus early region

PXDLS - Pro-X-Asp-Leu-Ser

NADH – Nicotinamide Adenine Dinucleotide Hydride

EMT – Epithelial-to-Mesenchymal Transition

BRCA – breast cancer

NAD – Nicotinamide Adenine Dinucleotide

NBD – Nucleotide Binding Domain

CD – Catalytic Domain

MTOB – 4-methylthio-2-oxybutyric acid

HIPP – 2-hydroxyimino-3-phenylpropionic acid

CP61 – cyclo - SGWTVVRMY

PDB – Protein Data Bank

Chapter 1

Introduction

1.1. C-terminal binding protein as a novel anti-cancer target

C-terminal Binding Proteins (CtBPs) are a potential therapeutic target in cancer as they promote multiple pro-oncogenic activities. These proteins can act as oncogenes in various cancers, such as osteosarcoma, melanoma, pancreatic ductal adenocarcinoma, colon, prostate, breast, ovarian cancers, and leukemia (Chen, 2021). The first discovery of CtBP was due to its function as a phosphoprotein that interacted with the C-terminal protein sequences encoded in exon 2 of the oncogenic adenovirus 2/5 E1A protein (Byun & Gardner, 2013). This binding is predominantly mediated by a conserved peptide motif in E1A, Pro-X-Asp-Leu-Ser (PXDLS). The latter has been observed in many CtBP binding transcription factors (Blevins *et al.*, 2017). Histone deacetylases, histone methyltransferases, and histone demethylases, in addition to numerous different classes of sequence-specific DNA binding proteins and chromatin associated complexes, are examples of binding partners for CtBP (Byun *et al*., 2019).

Later on, the CtBP was established to function as a transcriptional co-repressor when its targets included gene promoters through the C-terminal sequences of E1A (Byun & Gardner, 2013). Cancer cell survival and migration/invasion is promoted through these CtBP transcriptional co-repressors (Straza *et al.*, 2010). The transcriptional coregulator, CtBP, represses tumour suppressor genes and activates proto-oncogenes. Multiple tumour suppressor genes are repressed by CtBP, comprising E-cadherin, PTEN, the apoptosis genes Bik, Noxa, Puma and Perp and the breast cancer susceptibility gene BRCA1 (Korwar *et al*., 2016).

1.2. Role of CtBP in cancer

CtBP is elevated in many forms of cancer since CtBP has a role in both tumorigenesis and tumour progression. Processes important for oncogenesis, tumour maintenance and progression/metastasis, are all regulated by CtBP (Straza *et al*., 2010). Alongside CtBP's role in repression of apoptotic pathways and activation of growth and metastasis, this protein is also upregulated in a number of cancer tissues. The types of cancer where CtBP is dominant include colorectal cancer, melanoma, metastatic prostate cancer, oesophageal squamous cell carcinoma, ovarian cancer and breast cancer (Bellesis *et al*., 2018). High levels of CtBP, reflect a poor chance of survival. Remarkably, increased levels of CtBP in tumour tissues have been interconnected with a poorer chance of survival in breast cancer, ovarian cancer and hepatocellular carcinoma (Bellesis *et al*., 2018). Since CtBPs are prominent in multiple cancers, pharmacological inhibition of these proteins is a revolutionary treatment of cancer. There are only a few transcription factors besides CtBP, that harbour an intrinsic enzymatic and druggable, functional domain (Straza *et al*., 2010). This means that targeting CtBP is an indispensable concept in drug design.

Tumour cells survive under unique conditions such as hypoxia and/or high nicotinamide adenine dinucleotide hydride (NADH). NADH is a central product of carbohydrate metabolism that can function as a secondary messenger to control the activity of multiple different epigenetic regulatory complexes in human cells (Byun & Gardner, 2013). The repressor activity of CtBP does not rely on the dehydrogenase activity of CtBP but rather on NADH/NAD+ binding. Activation of CtBP as a transcriptional co-regulator is only possible under conditions of hypoxia or glycolysis. The latter is responsible for the

elevated NADH levels, as might be found in hypoxic tumours or those undergoing aerobic glycolysis (Korwar *et al*., 2016).

CtBP remains in an inactive and monomeric form if the concentration of NADH is low. This can lead to retention of CtBP1 in the cytoplasm. However, if NADH levels are elevated, this results in the dimerization of two CtBP subunits, so as to form the protein's transcriptionally active form. This links enzymatic function and transcriptional activity. Transcriptional repression of a wide network of genes that acknowledge the tumour cell phenotype, is caused by the NADH-dependent CtBP dimerization (Birts *et al*., 2013). Elevated concentrations of NADH escalate oligomer formation, subsequently influencing CtBP's ability to relate with transcription factors (Dcona *et al*., 2017). CtBP activity is reliant upon its ability to form dimers and higher order oligomers (Byun *et al*., 2019). Recent studies have shown that CtBPs may be implicated in chemotherapy-induced drug resistance and that CtBPs are tremendously expressed in inflammatory diseases such as acute lung injury, chronic renal failure, traumatic brain injury and osteoarthritis (Chen, 2021).

1.3. CtBP receptor subtypes

CtBP1 and CtBP2 are the two genes from which CtBP is expressed (Byun & Gardner, 2013). Both of these genes have the ability to homodimerize and heterodimerize in the nucleus to impact various distinct epigenetic nuclear events. This homodimerization or heterodimerization is achieved by recruiting a diverse array of chromatin – modifying complexes. Consequently, CtBP in its dimeric form, has the comprehensive prospects of reshaping the landscape of epigenetic regulation throughout the nucleus (Byun *et al*., 2019). CtBP1 and CtBP2 are two closely related, paralogous genes whose overexpression

is seen across the spectrum of solid human tumours, consisting of breast, ovarian, prostate, colon and gastric cancer. Crystal lattices of CtBP1 and CtBP2 revealed a lot of similarities between the two proteins. They both share a very identical tetrameric arrangement, they both bind to NADH by the dehydrogenase homology domain, and they can use NADH/NAD+ as a cofactor (Han *et al*., 2022). Despite the fact that CtBP1 and CtBP2 occupy high sequence homology and structural similarities, they have been shown to exert both unique and duplicative effects (Stankiewicz *et al*., 2014).

1.3.1. Differences between CtBP1 and CtBP2

A broad expression of both CtBP1 and CtBP2 is seen in mammalian tissues including the brain (Cousins & Stephenson, 2018). However, one main difference is that CtBP1 is mainly expressed from embryo to adult while CtBP2 appears to be expressed earlier in development (Acosta-Baena *et al*., 2022). With respect to their subcellular distribution, CtBP2 has a nuclear localization signal and accumulates within the nucleus while CtBP1 is observed in both the cytoplasm and in the nucleus (Cousins & Stephenson, 2018).

Furthermore, CtBP1 possesses a PDZ-binding domain at its C-terminus which is not seen in CtBP2, thus allowing for cytoplasmic functions with particular proteins such as neuronal nitric oxide synthase (Acosta-Baena *et al*., 2022). Another difference is that the C-terminal region of CtBP1 is more prone to proteolytic cleavage than that of CtBP2. CtBP2 also contains a nuclear localization signal at its N-terminus that is not present in CtBP1 and it possesses an additional twenty amino acids at its N-terminus. This plays a compelling role in the regulation of the function of CtBP2.

1.3.2. CtBP1

Overexpression of CtBP1 is considered to be pro-tumorigenic and it influences the regulation of gene networks affiliated with “cancer hallmarks” and malignant behavior, including enhanced cell survival, proliferation, migration, invasion and the epithelial mesenchymal transition (EMT) (Blevins *et al*., 2017). The two most important cancer hallmarks associated with CtBP1 overexpression are an elevated EMT and the evasion of cell death. The evasion of growth or tumour suppressor functions and an increment in cell proliferation, are further hallmarks with which CtBP is correlated (Blevins *et al*., 2017). CtBP1 expression is elevated in metastatic prostate cancer, melanomas, colon cancer, leukemia, ovarian cancer and breast cancer. The breast cancer genes BRCA1 and BRCA2 function in DNA repair and produce tumour suppressor proteins. If there are any mutations in these genes, this leads to breast cancer. BRCA1 and BRCA2 are repressed by CtBPs, resulting in cancer. Studies in head and neck squamous cell carcinomas exposed that BRCA1 is repressed through CtBP1 binding to its promoter (Stankiewicz *et al*., 2014).

1.3.3. CtBP2

CtBP2 performs in opposition to significant tumour suppressors such as E-cadherin, $p16^{INK4A}$, $p15^{INK4B}$, PTEN, HIPK2, INK 4a/Arf and APC, and intensifies cell proliferation, migration and invasion (Yang *et al*., 2017). CtBP2 is raised in ovarian cancer and in breast cancer tissues and cell lines. This makes breast cancer cells phenotypically more malignant and lowers the patient’s overall survival rate. In ovarian cancer cell lines, CtBP2 represses the BRCA1 gene while in breast cancer, CtBP2 lowers the $p16^{INK4A}$ levels

as these are inversely related. This results in a curtailed transition of the cell cycle in breast cancer cells (Yang *et al*., 2017).

1.4. Physiological role of CtBP

Transgenic studies involving CtBP revealed that CtBP contribute to a vast variety of developmental functions (Byun & Gardner, 2013). High levels of CtBP are exhibited in the course of development and the proteins also participate in axial patterning, cellular proliferation, and differentiation within many organs, such as the eyes, heart, brain, placenta vasculature, and muscle (Blevins *et al*., 2017). CtBP plays a fundamental role in normal embryogenesis due to its co-transcriptional function, since it is responsible for two main functions. It acts as a regulator of the EMT and is vital in proper foetal cell differentiation (Bellesis *et al*., 2018). EMT is a striking way via which transformed epithelial cells gain the potential to invade, resist apoptosis and disseminate. Thus, EMT is a crucial phenomenon in cancer, throughout which cells are deprived of their epithelial status but instead acquire mesenchymal traits (Dcona *et al*., 2017).

Another physiological role of CtBPs is seen in their responsiveness to metabolic signals where the development of brown and white adipose tissue is regulated (Jack *et al*., 2011). In adipose tissue, CtBP has displayed its ability to induce the phenotypic conversion of white adipocytes to a brown phenotype through the transcriptional repression of a set of white fat genes (Blevins *et al*., 2017). Therefore, besides having a role in human cancer, CtBPs have two functional and physiological roles which are, to co-repress epithelial genes in order to permit EMT, and to modulate the growth of both white and brown adipocytes, which is a process known as adipogenesis.

1.5. Structure of CtBP

CtBP1 and CtBP2 are thoroughly preserved and they split a notable degree of sequence and structural homology (Stankiewicz *et al*., 2014). A conserved functional domain is shared within all invertebrate and vertebrate CtBP proteins (Jack *et al*., 2011). CtBP is made up of 3 functional domains: an N-terminal transcription factor – binding domain, a central dehydrogenase domain, and a C-terminal extension. The latter is additionally involved in interactions with transcription factors and further proceeds to stabilize receptor binding (Dcona *et al*., 2017). These domains, all have specific features and functions.

1.5.1. The N-terminus

The N-terminus consists of a hydrophobic cleft known as the PXDLS binding site which is a receptor interaction cleft that recognizes and binds to PXDLS-like motifs in CtBP partner proteins. These partner proteins encompass a selection of transcription factors and transcriptional coregulators that can target CtBP proteins to gene regulatory sites in DNA (Jack *et al*., 2011). This hydrophobic cleft is vital for CtBP function as it is the transcription factor – binding domain. One of the specific roles of the PXDLS binding site is its recruitment of the core machinery of the co-repressor complex, including histone deacetylases (HDACs), histone methyltransferases (HMTases) and transcriptional repressors. This recruitment is essential for CtBP to execute its proper co-repressor functions (Stankiewicz *et al*., 2014). The PXDLS domain also facilitates direct binding of CtBPs to proteins in order to repress their function.

1.5.2. The dehydrogenase domain

The dehydrogenase domain of CtBP accommodates an NADH or nucleotide-binding domain (NBD), a catalytic domain (CD), and an RRT motif that binds to some transcription factors (Dcona *et al*., 2017). The NBD is able to bind to both NAD+ and NADH. Nonetheless, it has been found that CtBPs capability to bind to NADH is needed so as these proteins form dimers with one another and apply transcriptional repression (Stankiewicz *et al*., 2014). Thus, the dehydrogenase domain region contributes to dimerization in response to high NADH levels. In the presence of NADH, CtBP is subjected to a conformational change that aligns the receptor and NBD into close proximity, forming a "closed state", which is an indispensable step in the formation of CtBP oligomers (Dcona *et al*., 2017). The oligomerization of CtBP is what allows CtBP itself, to act as a transcriptional co-repressor. This implies that when there is increased concentration of NADH, oligomer formation is increased as well and this influences association with transcription factors. Along with the NBD, CD encodes an enzymatic function that catalyzes reduction of α – keto acids to 2-hydroxy acids in the presence of NADH (cofactor) (Dcona *et al*., 2017).

The dehydrogenase domain also contains an RRT binding-motif. Like the PXDLS hydrophobic cleft, the RRT-binding pocket of CtBPs is primarily utilized for binding and recruitment of members of the co-repressor complex. Several CtBP interacting proteins that bind at this site also possess PXDLS-binding motifs (Stankiewicz *et al*., 2014).

1.5.3. The C-terminus

The C-terminus is a potential site for attracting additional protein partners to the co-repressor complex. Due to the fact that the C-terminus owns a proline/glycine – rich

sequence this makes it unstructured. The latter contributes to a conformational disorder and makes CtBP unable to be crystallized full length. From evidence obtained by chromatographic and X-ray scattering experiments, it has been shown that the C-terminus encourages the formation of CtBP tetramers (Dcona *et al*., 2017). The tetramer is important for CtBP function. A PXDLS-binding motif is also found in the C-terminus of CtBP1.

1.6. CtBP inhibition

CtBP inhibition is mainly accomplished through four different manners, including disruption of CtBP dimerization, reduction of cell migration, termination of cellular invasion, and improvement of DNA repair (Byun *et al*., 2019). Inhibition of CtBP's transcriptional coregulator activity has been proven to reverse the neoplastic phenotype (Korwar *et al*., 2016). Even though, therapeutically targeting transcription factors is generally regarded as a demanding endeavour, a considerable amount of progress has been achieved when it comes to manipulating different modes of transcriptional regulation. The various modes comprehend protein-protein interaction, enzymatic activity, DNA binding, and epigenetic alterations (Blevins *et al*., 2017). This implies that there are several methods or approaches to inhibit CtBP activity. Therefore, these targeting agents, which typically are small molecules, work via inhibition of enzymatic activity, interference with dimerization or, inhibition of the interaction of CtBP with other co-regulators (Dcona *et al*., 2017).

1.6.1. Targeting the dehydrogenase activity of CtBP

CtBP transcription repressors serve as prospectively appealing cancer drug targets because they encode a potentially "druggable" dehydrogenase domain. These repressors are silenced by RNAi, thus leading to anti-cancer effects, namely apoptosis and abolishment of cancer cell migration/invasion (Straza *et al*., 2010). This catalytically active dehydrogenase domain is an attractive objective for small molecule intervention, thus bringing about pharmacological inhibition of CtBP's dehydrogenase activity which in turn, reverses the cancer phenotype and induces apoptosis.

The CtBP dehydrogenase substrate 4-methylthio-2-oxybutyric acid (MTOB) can function as a CtBP inhibitor when present at high concentrations, and is cytotoxic to cancer cells (Straza *et al*., 2010). The intermediate in the methathionine salvage pathway MTOB, preferentially kills cancer cells meaning that it has limited or no toxicity in normal cells and tissues. Investigation of cell death originating from CtBP inhibition with MTOB demonstrated that this treatment assists cell death in various cancer cell lines through displacement of CtBP from the pro-apoptotic Bik promoter (Stankiewicz *et al*., 2014). Bik has been involved as a human tumour suppressor, with epigenetic silencing, deletions or mutations detected in renal, colon, brain and lymphatic malignancies (Straza *et al*., 2010).

Out of all of the receptor – based inhibitors amongst the CtBP family, the compound, MTOB, and a few of its derivatives are the most meticulously examined (Blevins *et al*., 2017). A potential explanation for the mechanism of inhibition of MTOB is that it adjusts the NADH/NADp ratio by consuming NADH in the dehydrogenase reaction, thereby altering the monomeric and dimeric state of CtBP1, consequently regulating its

transcriptional activities (Blevins *et al*., 2017). The reduction of the ketone in MTOB to the alcohol in the product 4-methylthio-2-hydroxybutyric acid (MTHB), is catalyzed by the dehydrogenase active site, whilst converting NADH to NAD+ in the process (Korwar *et al*., 2016). Since MTOB inhibits the dehydrogenase activity of the CtBPs, it can decrease tumour burden in colon cancer and has the potential to modulate the immune response to head injury (Li *et al*., 2020).

MTOB can also reduce tumour size and it is specific, which implies that it is a superior substrate for CtBP's dehydrogenase domain than other α–ketoacids that are similar in structure. In order for MTOB to inhibit the dehydrogenase domain, it must bind to CtBP1 and CtBP2's active site cleft within their substrate-binding domain (SBD) in the N-terminus. Overall tertiary or quaternary change in the protein's structure is not induced by MTOB (Hilbert *et al*., 2014).

As a result of its mode of action, MTOB needs to be used at high concentrations to prompt a cellular response (Birts *et al*., 2013). This response intercepts the recruitment of CtBPs to target promoters and antagonizes CtBP transcriptional regulation. However, for this procedure to be successful, there must be treatment with high concentrations of MTOB such as 4 and 10mM (Blevins *et al*., 2015). Even though high MTOB concentrations (mM) are requisite for receptor inhibition of CtBP, MTOB's clear inhibitory effect on cancer cells issues proof of concept that small molecules could be refined to successfully target and treat cancers that are particularly regulated by CtBP activity (Hilbert *et al*., 2014).

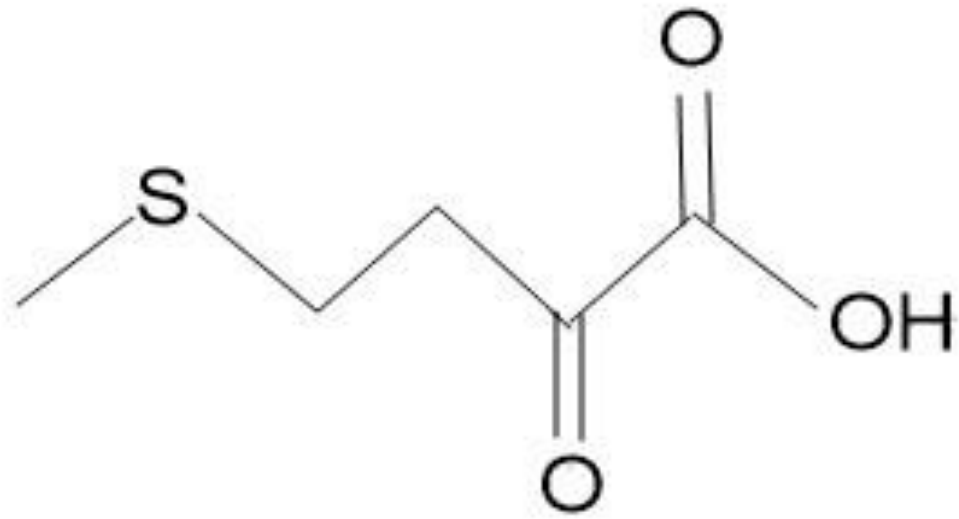


Fig 1.1: 2D Structure of MTOB. Rendered in BIOVIA® Draw[1]

The difference between CtBP family proteins and further D2-dehydrogenases is that they accommodate a unique active site tryptophan residue within their active site (Trp 318/324 for CtBP1/2) (Dcona *et al*., 2017). The tryptophan residue together with a hydrophilic channel/cavity contribute to the unique active site features of CtBP and are useful for the design of highly selective anti-neoplastic CtBP inhibitors. In the future, it may be viable to look into the development of inhibitors that exploit the use of this open cavity and displace all the water molecules to increase favourable binding interactions (Dcona *et al*., 2017). The incorporation of additional groups that expand into this hydrophilic cavity may bestow an appropriate approach in designing potent and specific CtBP inhibitors. A specific CtBP inhibitor is 2-hydroxyimino-3-phenylpropionic acid (HIPP), which is a potent substrate competitive inhibitor that can fully occupy the active site of CtBP.

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

HIPP projects its hydroxyimino hydroxyl into a hydrophilic channel and it has a high binding affinity. This binding arrangement is favourable since a substantial increase in affinity is noticed with HIPP and, in addition, the hydrophilic cavity is accessible to an inhibitor without substantial penalty for displacing ordered water molecules (Hilbert *et al*., 2015). The strong inhibitor, HIPP, substantially inhibits CtBP catalysis and it binds to CtBP with an affinity greater than 1000-fold over that of MTOB. HIPP can be used besides other inhibitors as a scaffold for attaching chemical moieties which intrude on NADH binding.

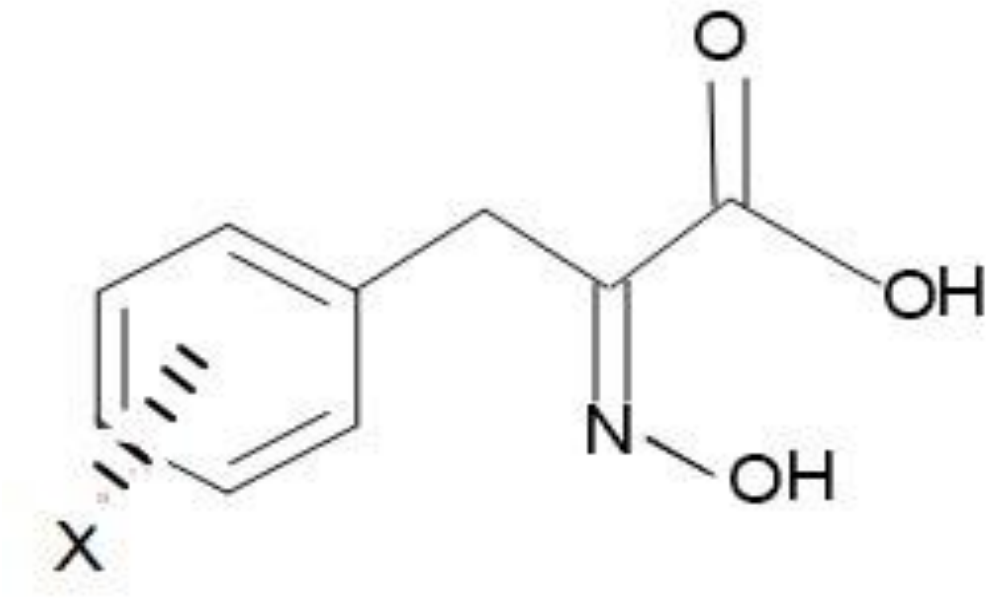


Fig 1.2: 2D Structure of HIPP. Rendered in BIOVIA® Draw[1]

1.6.2. Targeting the dimerization of CtBP

Another type of CtBP antagonist takes advantage of the fact that the co-repressive functions of CtBP rely upon NADH-dependent dimerization. CtBP1 homodimerizes and heterodimerizes with CtBP2 through a predominantly large hydrophobic interface in its NBD.

---

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

This homodimerization and heterodimerization allows for the connection of DNA-associated transcription factors with epigenetic regulators (Blevins *et al*., 2017). The mechanism of action of these CtBP antagonists, in order to inhibit CtBP activity, is to target the CtBP dimerization interface. Cyclic peptides have also been utilized to inhibit CtBP's activity and these macromolecular structures operate by inhibiting CtBP's ability to dimerize (Dcona *et al*., 2017). A cyclic nona-peptide, CP61 (cyclo-SGWTVVRMY), inhibits heterodimer formation between CtBP1 and CtBP2, *in vitro* and in cells by binding to an allosteric site (Korwar *et al*., 2016). Thus, this restricts the ability of CtBP1 to colocalize into the nucleus. CP61 is cyclized using NH_2-terminal serine and COOH-terminal tyrosine, thus forming a loop-like structure (Dcona *et al*., 2017).

Other functions of CP61 include regulating cellular function, thereby reducing mitotic fidelity, proliferation, and colony forming potential in cancer cells. This is all achieved by inhibiting the protein-protein interaction of a transcription factor such as CtBP, which provides insight into the fundamental processes of cell cycle control. Thus, CP61 further exhibits the potential of CtBPs as targets for the development of anti-cancer therapeutics (Birts *et al*., 2013).

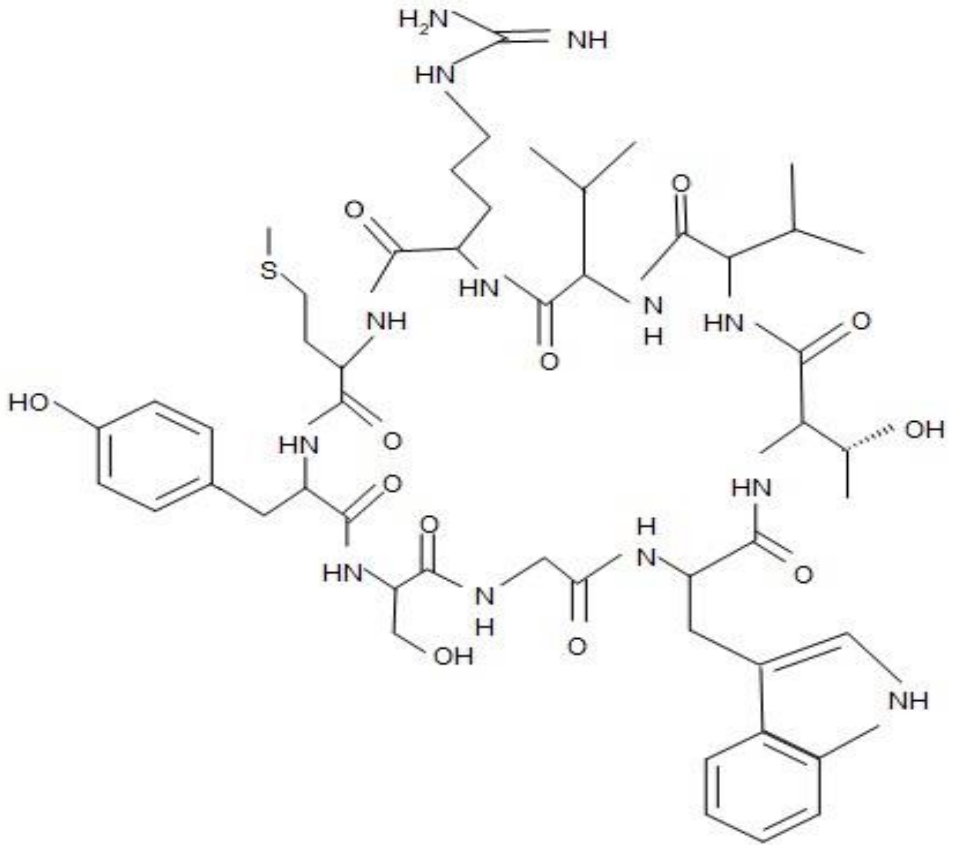


Fig 1.3: 2D Structure of CP61. Rendered in BIOVIA® Draw[1]

1.6.3. Targeting the interaction between CtBP and its protein partners

A different approach to decrease CtBP activity is to disrupt protein-protein interactions such as the CtBP/transcription factor interaction via small molecule intervention. These highly selective small-molecule inhibitors can bind to the hydrophobic cleft in CtBP to perturb CtBP-mediated transcriptional repression. In gene repression and tumorigenesis, there are CtBP1 transcription factor interactions, that are demonstrated through single amino acid mutations in the PXDLS sequence or the CtBP1-binding cleft. This suggest that small molecule inhibitors targeting such a protein interaction could be a beneficial therapeutic approach such as that seen with NSC95397 (Blevins *et al*., 2017).

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

NSC95397 was found to specifically target the CtBP-E1A interaction in secondary protein-protein interaction, enzymatic, and *in vitro* assays. However, NSC95397 does have other known activities (Blevins *et al*., 2017). NSC95397 also retains features of a weak substrate of CtBP dehydrogenase activity and does not inhibit the other dehydrogenase, LDH (Blevins *et al*., 2015). When compared to MTOB, NSC95397 was a weaker CtBP inhibitor but it was just a weak a substrate, as MTOB. However, NSC95397 was competent enough to inhibit interaction between CtBP1 and E1A proteins, better than MTOB. NSC95397 is more effective than MTOB, in disrupting the interaction between CtBP and E1A, because it is larger than MTOB and this produced a conformational change in CtBP and locks it in a conformation that intercepts its interaction with transcription partners (Blevins *et al*., 2015).

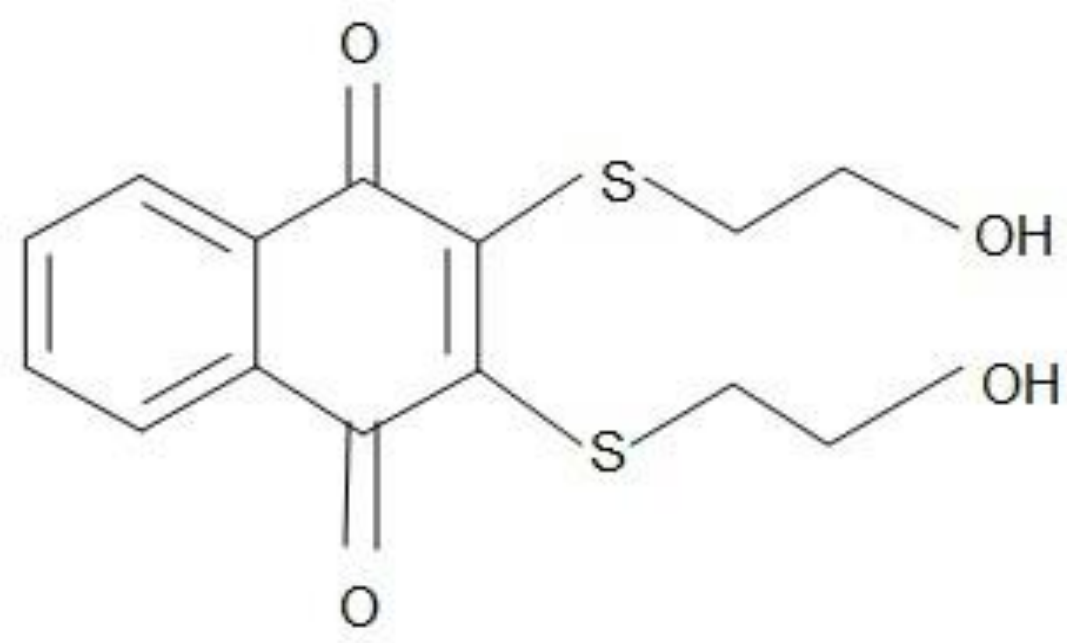


Fig 1.4: 2D Structure of NSC95397. Rendered in BIOVIA® Draw[1]

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

1.6.4. Advantages of CtBP inhibitors

The four CtBP inhibitors, MTOB, HIPP, CP61, and NSC95397, all take different routes to tackle CtBP inhibition. These various methods to antagonize CtBP, stimulate the development of new anti-cancer treatments that specifically target CtBP. Up until now a considerable advancement has been done to comprehend the role of the CtBP transcriptional complex in tumorigenesis. It endures as an attractive target due to its limited expression in most adult tissues and its potential to reactivate developmental programs critical for tumorigenesis/metastasis when aberrantly re-expressed (Blevins *et al*., 2017). Reports of unpleasant side effects regarding therapy targeting CtBP are limited since CtBP expression levels are low or absent in many adult tissues, except for the brain tissue, adipose tissue, and skeletal muscle. Screening of CtBP inhibitors and CtBP-partner interaction inhibitors is crucially needed to continue exploring the mechanistic role of CtBP in cancer development (Chen, 2021). In fact, to date, a number of small-molecule pools are commercially available.

The identification of CtBP inhibitors can be a rigorous project. Nevertheless, CtBP has brought about exceptional interest in the context of being a co-repressor that targets multiple tumour suppressors, and also an activator of key oncogenes (Dcona *et al*., 2017). Therefore, CtBP maintains tremendous promise as a wholly novel target for future cancer therapeutic development.

1.7. The protein data bank

The protein data bank (pdb) (Bernstein et al., 1977) archive is the single worldwide repository of information regarding the 3D structure of large biological molecules, such as proteins and nucleic acids. Each protein in the pdb has extra descriptions and

annotations allocated to it[2]. The specific information pertaining to each protein includes, the method used to resolve the protein crystal (can be either X-ray diffraction or Nuclear Magnetic Resonance (NMR)), the resolution, the R-value and the free R-value. This repository is essential for drug development since the shape of a molecule must be acknowledged before any molecule is accepted as an agonist or antagonist.

1.7.1. pdb crystallographic deposition 4LCE

The pdb file chosen as a template for this study was 4LCE (Hilbert *et al*., 2014). This file describes the bound coordinates of the CtBP1 small molecule antagonist MTOB complex. The analytical technique used to resolve the protein crystal is X-ray diffraction. X-ray crystallography is an analytical tool used to elucidate protein structure and it has no size limitations for the molecule or complex to be studied.

The pdb file was valid because it had a good resolution value of 2.38Å. Resolution is the minimum distance between adjacent atoms that can be reliably read since the atoms are distinguished from each other. The units of resolution are Ångström (Å), where the lower the value of Å, the higher is the resolution.

The R-value is one element that is used to assess the quality of a model and to evaluate the progress in the refinement of that model from X-ray crystallographic data. The R value is a calculation of discrepancy between the observed intensities from the diffraction pattern and the predicted intensities that are calculated from the model. A model is reliable if the R-value is less than or equal to 0.20. The molecule used for this project had an R-value of exactly 0.200, making it suitable with regards to its R value.

[2]RCSB PDB [Internet]. About the PDB Archive and the RCSB PDB. 2020. Available from: https://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/index.html.

Free R is also a more reliable tool for judging the model than the R-value, since the free R value comes from data that is not used in the experiment itself, implying that, there is reduced bias. The value of Free R is erroneous if it larger than 0.40 and this raises concern about the model. In the case of the molecule utilized in this project, the Free R value was 0.239. Although this Free R value did not exceed the R-value by more than 0.05, it was still acceptable since it was smaller than 0.40.

The target and ligand protein complex below shows the CtBP1 in complex with MTOB. The secondary structure of this CtBP1 protein represents the 3D arrangement of the protein in space. This secondary structure can be either an alpha helix or a beta pleated sheet. In this case it is that of an extended spiral spring, hence an alpha helix. Its structure is maintained by the hydrogen bonds between the neighbouring CO and NH groups. The purple, green, blue, yellow and red strands, represent the receptor, while the inner turquoise structure represents the ligand.

Fig 1.5: pbd crystallographic deposition 4LCE (Hilbert *et al.*, 2014), showing CtBP1 in complex with MTOB. Rendered in BIOVIA Discovery Studio Visualizer[3].

[3]Dassault Systèmes BIOVIA. BIOVIA Discovery Studio Visualizer. San Diego: Dassault Systèmes; 2019.

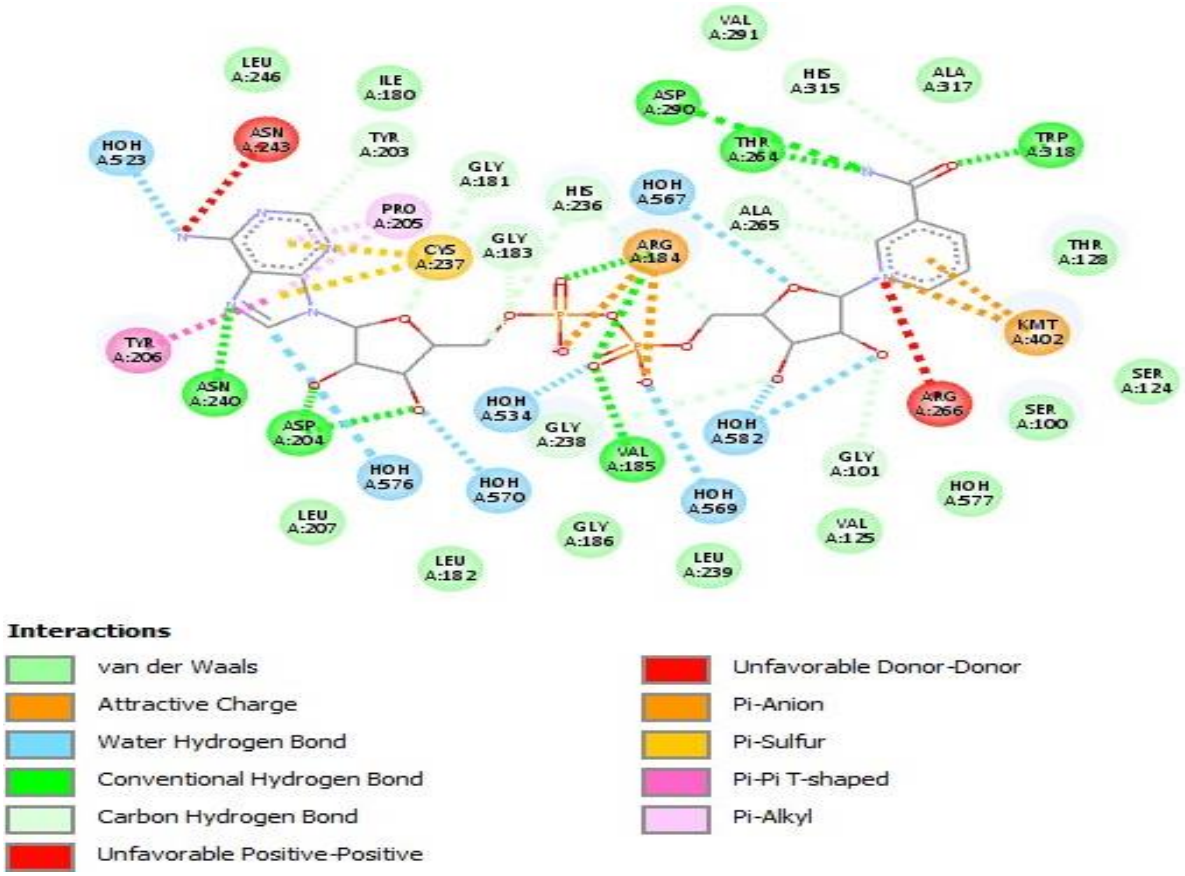


Fig 1.6: A 2D topology map of the receptor-ligand interaction. Rendered in BIOVIA Discovery Studio Visualizer[3].

1.7.2. pdb crystallographic deposition 4U6Q

A second pdb file was chosen as a template for this study. This was pdb crystallographic deposition 4U6Q (Hilbert *et al.*, 2015). This file describes the bound coordinates of the CtBP1 small molecule inhibitor HIPP. The analytical technique used to resolve the protein crystal was X-ray diffraction. The pdb file was also valid as well because it was well resolved to good resolution value of 2.30 Å. The model was also considered reliable because the R-value was 0.204 - only slightly exceeding the cut-off value of 0.20. The Free R value was also acceptable (0.249) and lower than 0.40. The target and ligand protein complex below shows the CtBP1 in complex with HIPP. The secondary structure of this CtBP1 protein represents the 3D arrangement of the protein in space.

[3]Dassault Systèmes BIOVIA. BIOVIA Discovery Studio Visualizer. San Diego: Dassault Systèmes; 2019.

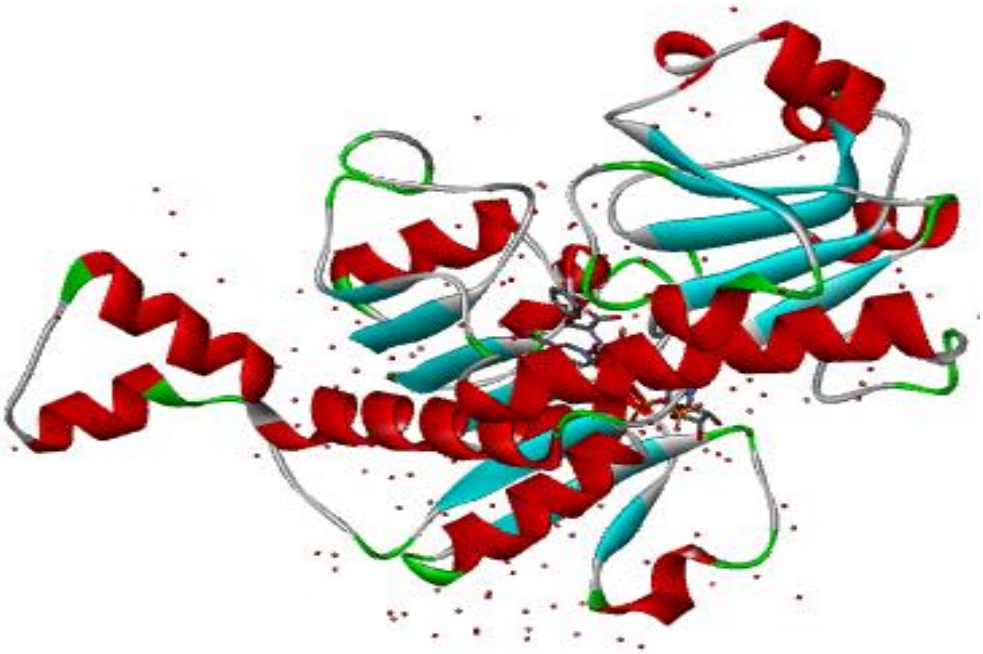


Fig 1.7: pbd crystallographic deposition 4U6Q (Hilbert *et al*., 2015), showing CtBP1 in complex with receptor HIPP. Rendered in BIOVIA Discovery Studio Visualizer[3].

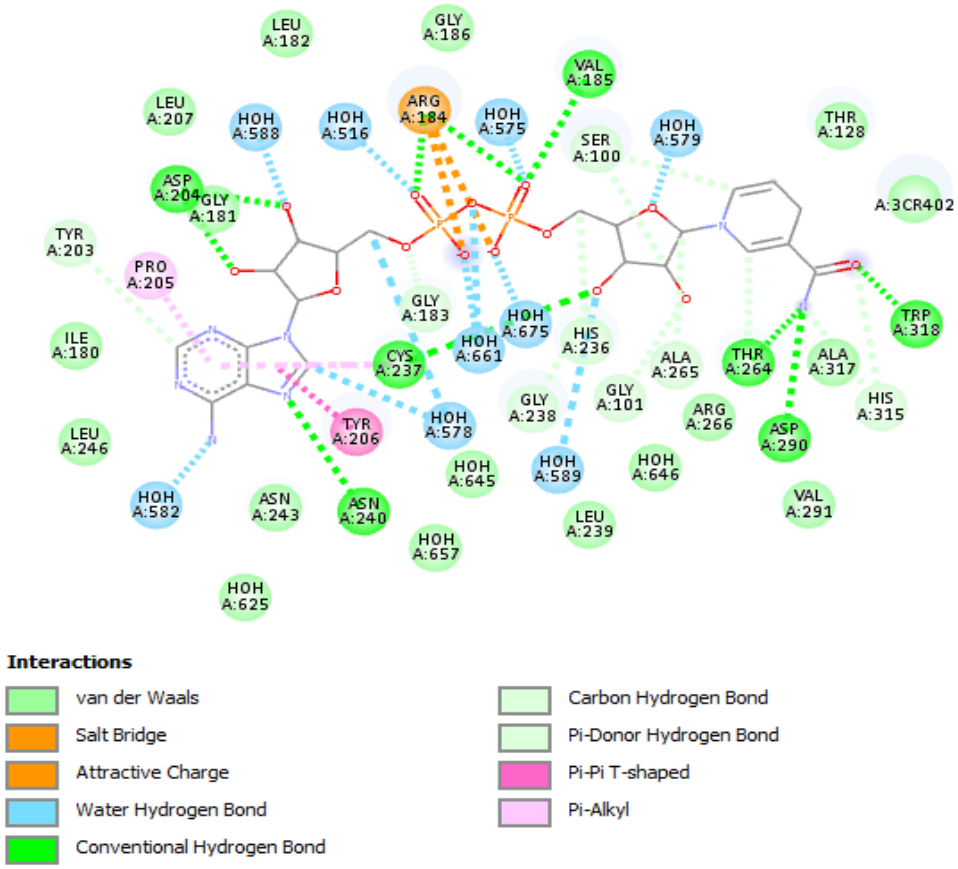


Fig 1.8: A 2D topology map of the ligand interaction with the receptor. Rendered in BIOVIA Discovery Studio Visualizer[3].

[3]Dassault Systèmes BIOVIA. BIOVIA Discovery Studio Visualizer. San Diego: Dassault Systèmes; 2019.

1.8. Computational software utilized in drug design

Development of new pharmaceutical drugs depends upon drug discovery software to test whether a newly created drug will be efficacious in treating a particular disease. Computational medicinal chemists contribute to the discovery and optimization of biologically active compounds by taking advantage of all kinds of software and resources in the computer aided drug design (CADD) field (Liao et al, 2011). Computational tools aid in simplifying the determination of the 3D structure of proteins, ligands, and the protein ligand complex which have the potential to become a novel target and drug. This makes the CADD an indispensable tool in the pharmaceutical industry. The list of software used for this project have been listed below.

1.8.1. BIOVIA® Discovery Studio Visualizer

BIOVIA® Discovery Studio Visualizer is a visualization tool utilized to model data and view, share and analyze proteins. It offers a visualization and collaboration framework that includes a comprehensive science portfolio[4]. Molecular visualization using this software, enables a mechanistic understanding of a molecule's structure to be visualized such as the visualization of pdb files in 3D. The latter allows for a straightforward process when it comes to *in silico* investigation via the modelling of structures and biological processes. This program can produce high performance publication quality graphics and can supply information regarding the interaction formed between the ligand and the receptor in both 2D and 3D formats.

[4]OmicX [Internet]. BIOVIA® Discovery Studio. 2015. Available from: https://omictools.com/biovia-discovery-studio-tool.

1.8.2. BIOVIA Draw®

BIOVIA Draw® provides structure and query drawings, permits the registration, search and report on chemically modified peptide or nucleotide sequences and allows for the addition of structure drawing and display to user's application. It also lets the user customize according to their organizational workflows[5].

1.8.3. LigandScout®

LigandScout® is used for precise virtual screening based on 3D chemical feature pharmacophore models and it provides seamless workflows, starting from ligand- and structure-based pharmacophore modeling. This program incorporates the latest high-performance alignment algorithms[6].

1.8.4. LigBuilder® Version 1.2

LigBuilder® can automatically build ligand molecules within the binding pocket and eventually these molecules are screened. All of this is based on the 3D structure of the target protein. This multi-purpose program was developed for structure-based *de novo* drug design and optimization. Analysis of the synthesis accessibility of designed compounds can be achieved with LigBuilder®. A cavity detection procedure is applied to distinguish the positions and shapes of the binding sites on the surface of a given protein structure and to quantitatively estimate druggability (Yvan et al, 2011).

---

[5]OmicX [Internet]. BIOVIA® Draw. 2015. Available from: https://omictools.com/biovia-drawtool.

[6]Bioinformatics Software [Internet]. LigandScout® 3.12 – Pharmacophore 3D modelling. 2020.

1.8.5. Mona®

MONA® is used for the preparation, inspection and conversion of small-molecule datasets and it primarily works on molecules rather than the molecule's occurrences in a particular dataset (Hilbig *et al*., 2013).

1.8.6. PoseView®

The tool PoseView® automatically creates two-dimensional diagrams of complexes with specified three-dimensional structures whilst complying with conventions of chemical structure diagram generation[7]. The 3D structures of protein-ligand complexes are entered as input in PoseView®. Such structures may come directly from crystal structures or be computed for example by a docking program[8]. The advantages of this software are that it automatically generates publication-quality graphics and it can export to a variety of graphic formats.

1.8.7. SYBYL- X®

SYBYL-X® is a comprehensive suite of computer-aided design tools manufactured to accelerate drug design and other molecular discovery projects, from high throughput screening to late lead optimization[9]. This program permits the creation of molecular modelling from sequence via lead optimization. SYBYL-X® has the potential to model and simulate both small molecules and macromolecules along with cheminformatics and lead recognition.

---

[7]ZBH Universität Hamburg [Internet]. PoseView®. 2018. Available from: https://www.zbh.unihamburg.de/en/forschung/amd/server/poseview.html.

[8]BioSolveIT [Internet]. PoseView®. 2020. Available from: https://www.biosolveit.de/PoseView/.

[9]Scientific Computing World [Internet]. SYBYL-X® 2.1. 2013. Available from: https://www.scientific-computing.com/press-releases/sybyl-x-21.

1.8.8. UCSF Chimera® Version 1.12

The program UCSF Chimera® provides an interactive visualization and analysis of molecular structures and related data, which include density maps, trajectories, and sequence alignments[10]. Another feature of UCSF Chimera® is the 3D visualization and structural analysis of molecular entities. It also enhances researcher workflow with novel extension features[11]. pbd files are visualized and rendered in a 3D format using this program.

1.8.9. ZincPharmer®

ZincPharmer® is an extensive molecular database that is sectioned in other separate databases. It aids in the construction and refinement of hypotheses straight from molecular structure and it allows for an interactive search of purchasable chemical space (Koes & Camacho, 2012).

1.9. Aims and Objectives

The project aimed to identify through virtual screening, high affinity CtBP modulators with a propensity for *in vivo* bioavailability. Experimental molecules, MTOB and HIPP, identified from the literature review, were used as lead molecules for the study.

---

[10]RBVI [Internet]. UCSF Chimera®. 2018. Available from: https://www.cgl.ucsf.edu/chimera/.

[11]USCF Innovation Ventures [Internet]. UCSF Chimera®: Molecular Modelling Software for Visualization and Analysis of Molecular Structures. 2006.

Chapter 2

Methodology

2.1. Background

This study was registered and acknowledged by the Faculty of Medicine & Surgery Research Ethics Committee (FREC) below the Application ID MED-2022-00397. The application was sent in for record purposes.

2.2. Virtual Screening

2.2.1. Generation of pharmacophores

pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB, and pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015) describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP were used to model the pharmacophores. The modelling of the pharmacophores and the consensus pharmacophore was done using LigandScout®[12].

2.2.1.1. MTOB pharmacophore

pbd crystallographic deposition 4LCE (Hilbert *et al*., 2014) was read into LigandScout®[12] and the bioactive conformation of MTOB was isolated. This was viewed in the structure-based modelling perspective, which is one of the five perspectives offered by LigandScout®[12] to analyze interactions between the ligand and macromolecule.

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

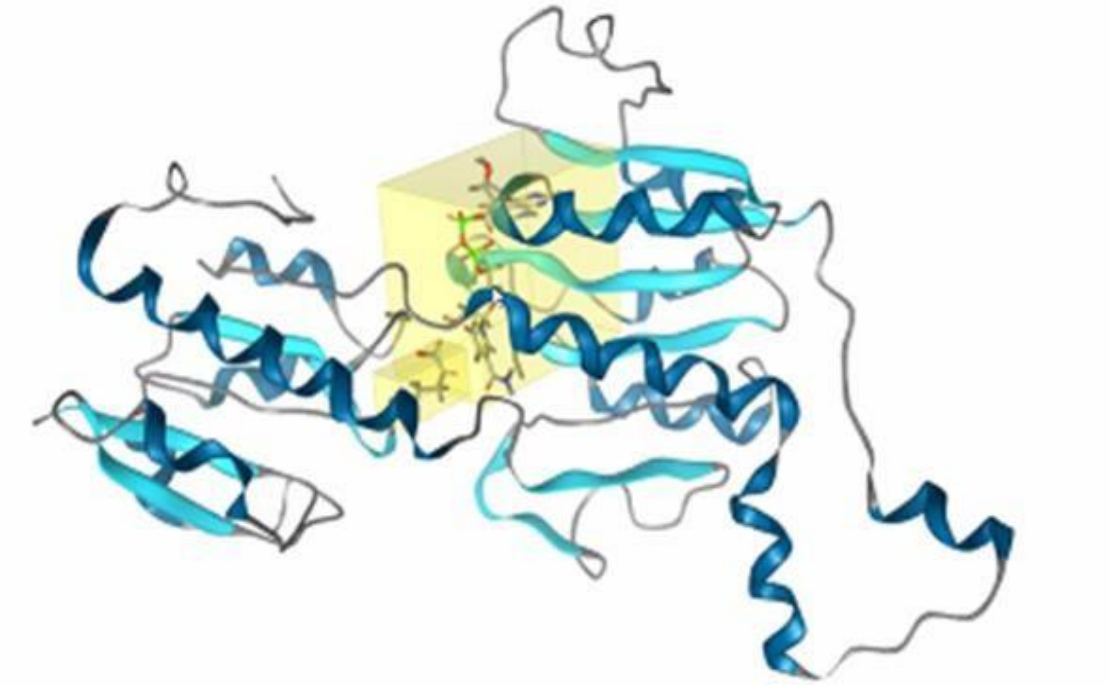


Fig 2.1: pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB. Modelled in LigandScout®[12]

Fig 2.2 below shows the bioactive conformation of the small molecule inhibitor MTOB as described in pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014) describing the bound coordinates of CtBP1 with small molecule antagonist MTOB.

Fig 2.2: Bioactive conformation of MTOB. Modelled in LigandScout®[12]

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

The bioactive conformation of MTOB was used to create a pharmacophore that described the most important binding moieties on the small molecule. This was done by clicking on the "Pharmacophore" tab and then choosing "Generate pharmacophore". The pharmacophore structure modelled was colour coded with yellow areas denoting hydrophobicity, red arrows denoting hydrogen bond acceptors and the red star denoting a negative ionizable area. A 2D map showing the interactions between the amino acids lining the CtBP ligand binding pocket and MTOB was also generated. The MTOB pharmacophore and ligand were then copied into alignment perspective using the copyboard widget, which is a tool that allows the transfer of chemical data from one perspective to another. The alignment perspective examines recurrent properties of pharmacophore models small molecules by aligning them and generating shared or merged feature pharmacophores.

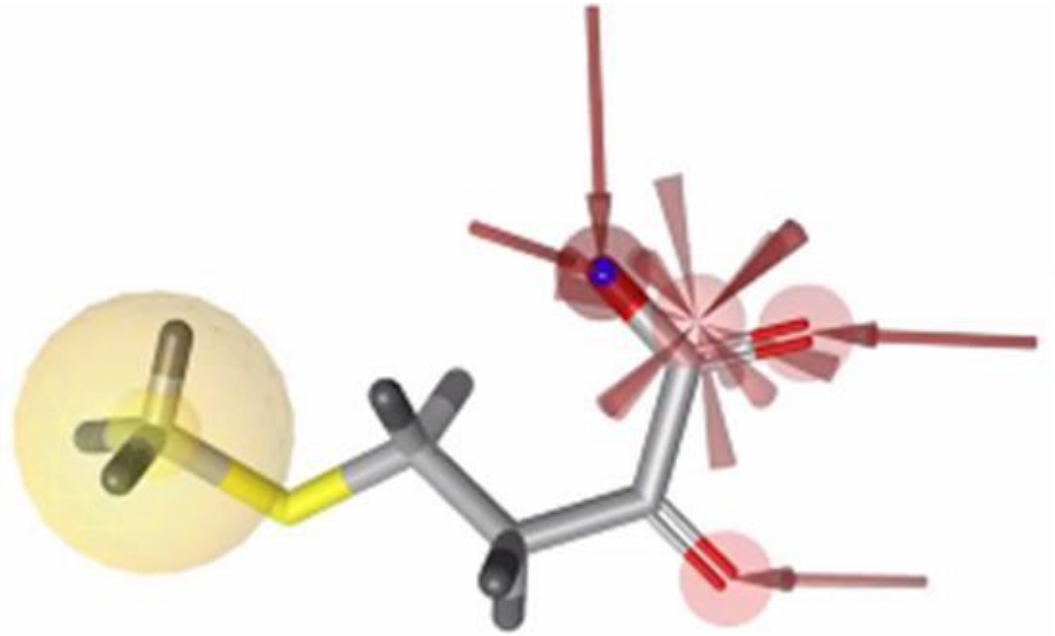


Fig 2.3: MTOB pharmacophore superimposed on the bioactive conformation as described in pdb ID 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB. Modelled in LigandScout®[12]

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

Fig 2.4: MTOB pharmacophore based on its bioactive conformation as described in pdb ID 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB. Modelled in LigandScout®[12]

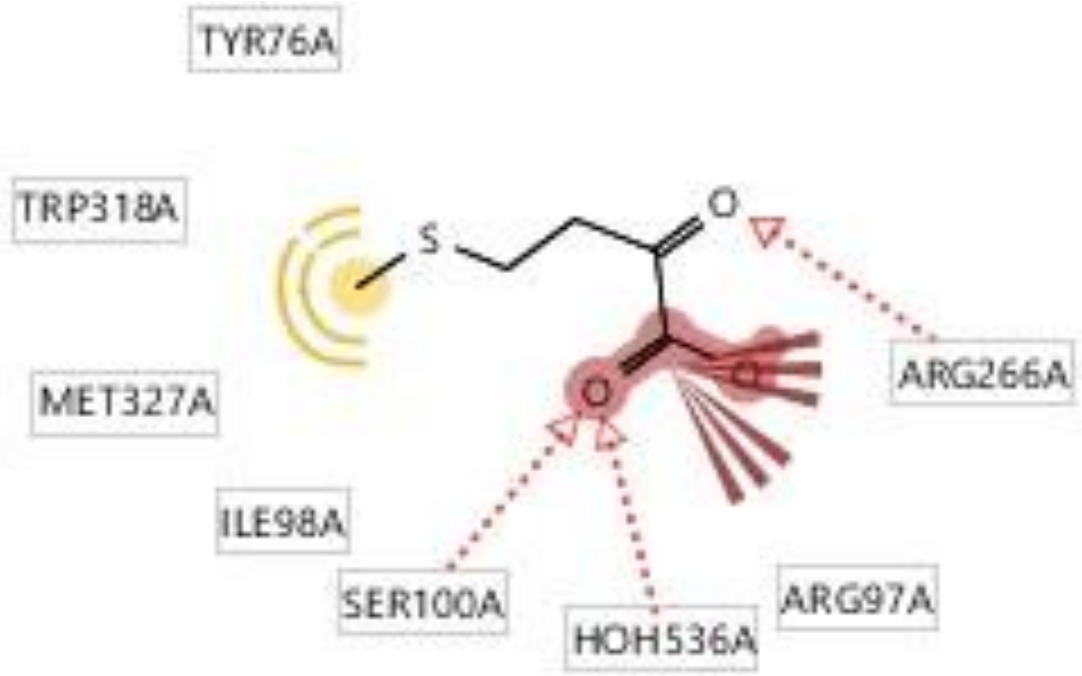


Fig 2.5: 2D map showing the interactions between the amino acids lining the CtBP ligand binding pocket and MTOB as described in pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB. Modelled in LigandScout®[12]

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

2.2.1.2. HIPP pharmacophore

pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015) was read into LigandScout®[12] and the bioactive conformation of HIPP was isolated. This was also viewed in the structure-based modelling perspective.

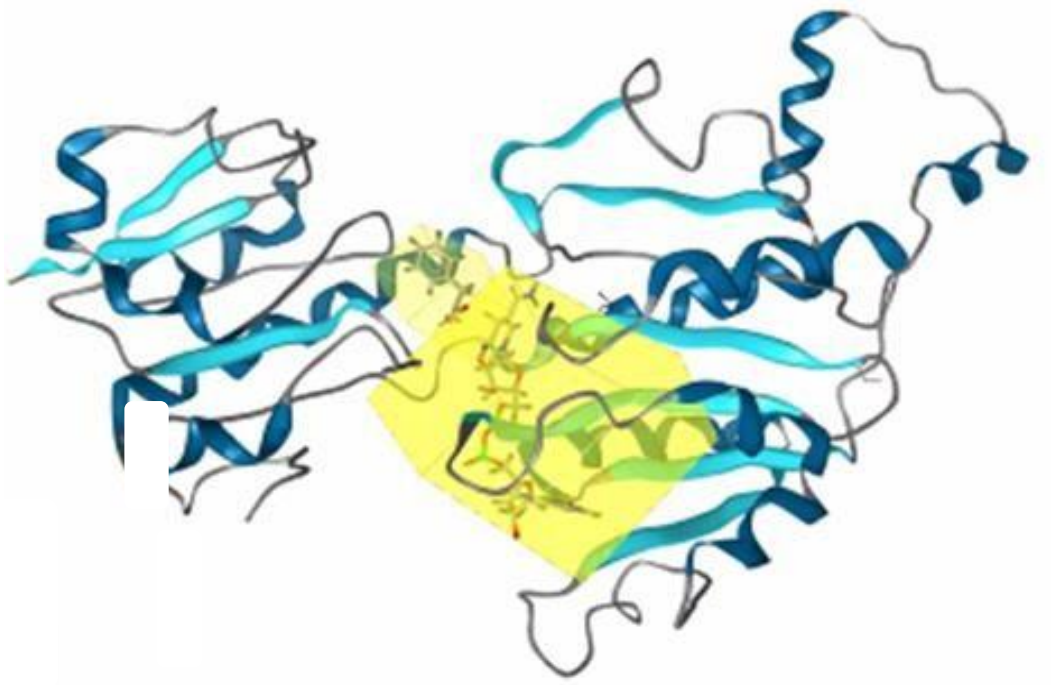


Fig 2.6: pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in LigandScout®[12]

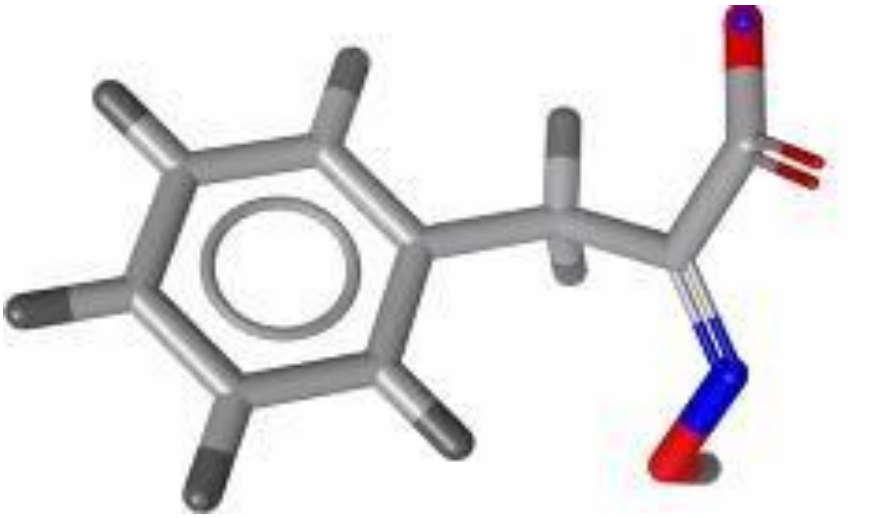


Fig 2.7: Bioactive conformation of HIPP. Modelled in LigandScout®[12]

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

Fig 2.7 above shows the bioactive conformation of the small molecule inhibitor HIPP extracted from pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), which describes the bound coordinates of CtBP1 with small molecule inhibitor HIPP.

Once the bioactive conformation of HIPP was acquired, the pharmacophore of HIPP, that described the most important binding moieties on the small molecule, was created. The pharmacophoric structures as modelled were colour coded with yellow areas denoting hydrophobicity, red arrows denoting hydrogen bond acceptors, green arrows denoting hydrogen bond donors, the red star denoting a negative ionizable area and the blue ring denoting an aromatic ring.

The HIPP pharmacophore and the ligand were also copied into alignment perspective using the copyboard widget, thus leaving 4 entities in the alignment perspective. These 4 entities were aligned using the alignment algorithm. A 2D map showing the interactions between the amino acids lining the CtBP ligand binding pocket and HIPP was also generated.

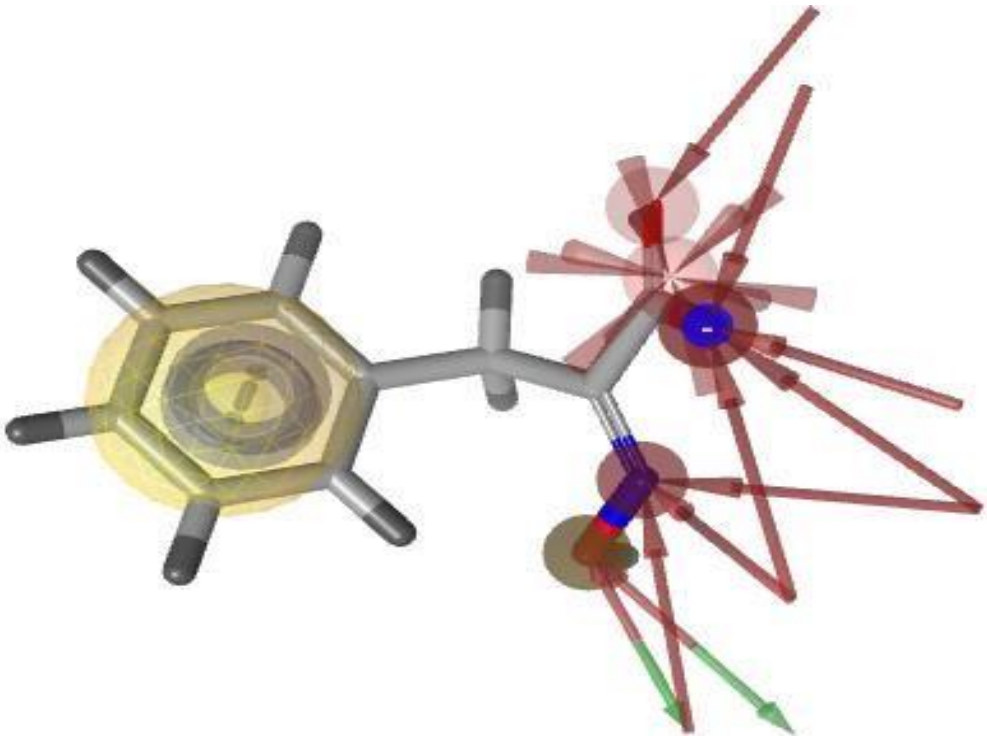


Fig 2.8: HIPP pharmacophore superimposed on the bioactive conformation as described in pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in LigandScout®[12]

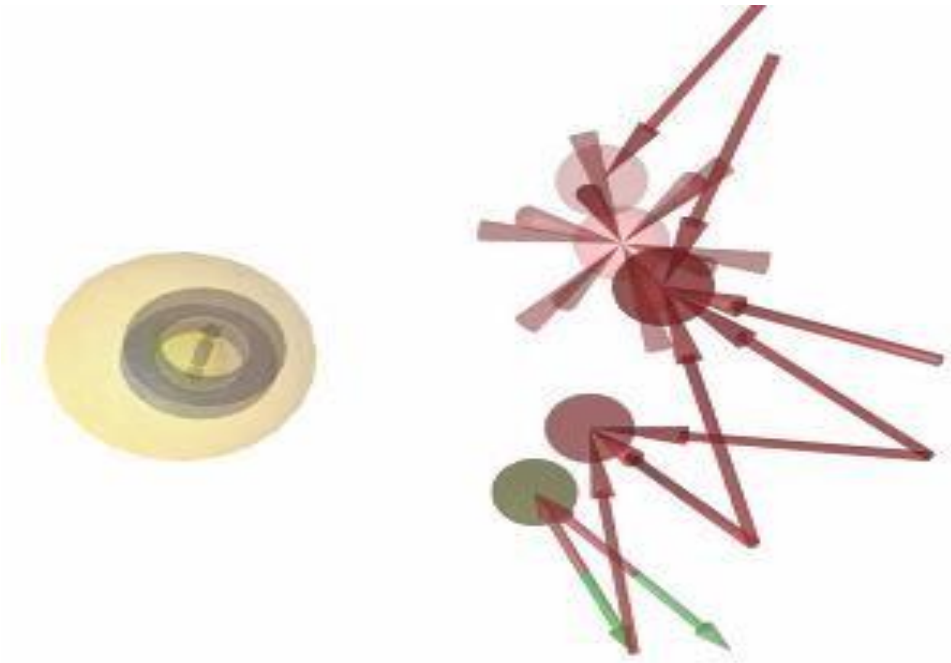


Fig 2.9: HIPP pharmacophore based on its bioactive conformation as described in pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in LigandScout®[12]

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

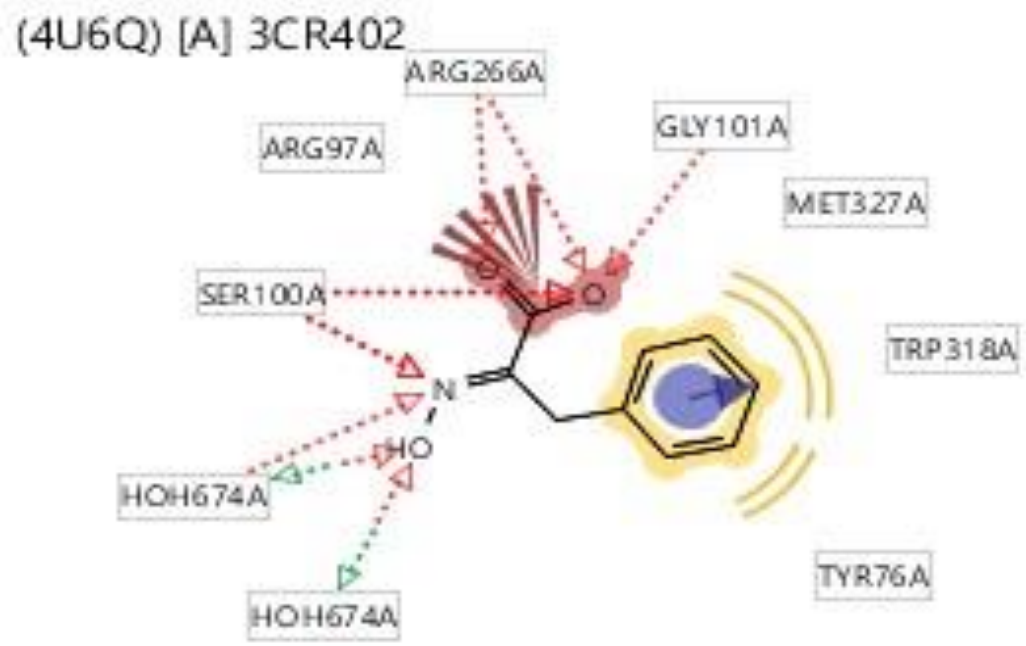


Fig 2.10: 2D map showing the interactions between the amino acids lining the CtBP ligand binding pocket and HIPP as described in pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in LigandScout®[12]

The MTOB core molecule was set as reference for comparison by clicking the "Set reference" icon. The MTOB pharmacophore was superimposed onto that of HIPP such that a consensus pharmacophore representing the critical interactions forged between the two molecules was created. This was achieved using the "Generate shared feature pharmacophore" icon. The shared pharmacophore can then be exported to other programs such as the ZINCPharmer®[13] database.

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

[13]Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database [Internet]. Nucleic Acids Research. 2012 Jul 01; 40(W1): W409-W414

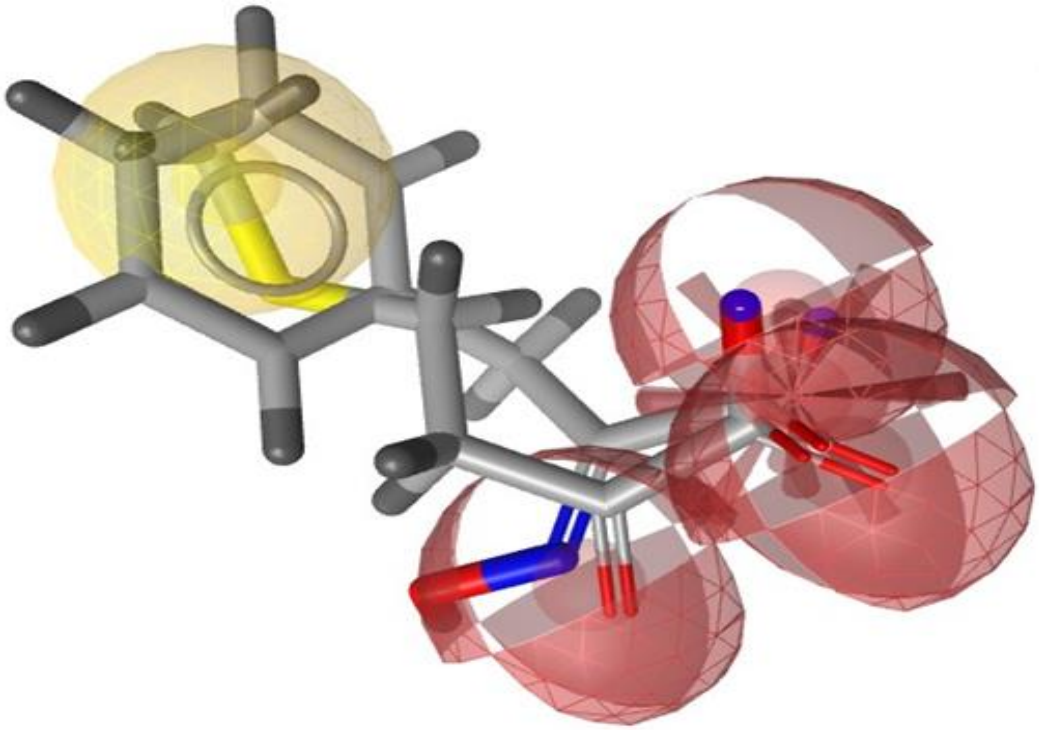


Fig 2.11: Consensus pharmacophore representing an average of the critical interactions forged between MTOB and HIPP and the CtBP receptor as described in pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB and in pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in LigandScout®[12]

2.2.2. Identification of hits

After obtaining the coordinates of the shared pharmacophore using LigandScout®[12], this pharmacophore was used to probe the ZINCPharmer®[13] database. This was done by uploading the consensus pharmacophore into ZINCPharmer®[13] by clicking on "Load features". Filters were applied during the virtual screening exercise in order to ensure the lead-likeness of the obtained hit cohort. Lead-likeness is important owing to the fact that it allows for the addition of structural moieties during optimization without violation of Lipinski's rules (Lipinski *et al*., 2001).

[12]Bendix F, Ibis G, Kainrad T, Langer M, Seidel T, Wieder O, et al. LigandScout 4.4.5. Inte:Ligand GmbH (19992021).

[13]Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database [Internet]. Nucleic Acids Research. 2012 Jul 01; 40(W1): W409-W414

When the filters in ZINCPharmer®[13] were applied, the two parameters regulated were the number of protein triple bonds and the molecular weight. The applied filters were consequently Rule of 3 (Congreve *et al*., 2003) compliant with the exception of hydrogen bond donor and acceptor count which are not included as filter variables on ZINCPharmer®[13]. The output from ZINCPharmer®[13] consisted of six files and these were saved in ".sdf" format. The files' titles were Zinc purchasable, Zinc Drug Database, Zinc In Man, Zinc Drug Database Metabolites, Zinc Natural Derivatives and Zinc Natural Products.

2.2.3. Inspection of molecules for Lipinski rule compliance

The 6 files obtained from ZINCPharmer®[13] were uploaded into the molecule set section of Mona®[14] where they were entered as a union set. The fact that hydrogen bond donor and acceptor count were not taken into account during the lead filtration process was compensated for during a second filtration process carried out using Mona®[14], where the fact that all accepted hits had hydrogen bond donor and acceptor count compliant with Lipinski's rules (Lipinski *et al*., 2001) was ensured. This was attained by applying the filters according to the Lipinski rule (Lipinski *et al*., 2001) chemical properties in order to create a filtered set. The latter included molecules within the following parameters:

---

[13]KoesDR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database [Internet]. Nucleic Acids Research. 2012 Jul 01; 40(W1): W409-W414

[14]Hilbig M, Urbaczek S, Groth I, Heuser S, Rarey M. MONA. Zentrum fur Bioinformatik: Universitat Hamburg; 2013.

- H bond acceptors between 1 and 10
- H bond donors between 1 and 5
- Lipinski acceptors between 1 and 10
- Lipinski donors between 1 and 5
- log P between 1 and 5
- Molecular weight between 0 and 500 Daltons

The filtered set created was then saved in "mol2" format as it originally was in "sdf" format.

2.2.4. Creation of protomol and upload of hits into the protomol

The pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014) describing the bound coordinates of CtBP1 with small molecule antagonist MTOB, was uploaded into SYBYL-X®[15] and any repeating subunits or unwanted molecules such as water molecules were removed. The protomol was modelled to identify the virtual extended ligand binding pocket of CtBP and the protomol represents the total vacant space inside the CtBP receptor. SYBYL-X®[15] has the ability to identify the unsatisfied areas in the protein in order to construct the protomol. Then the hit structures acquired from ZINCPharmer®[13] were docked into the protomol by uploading the file of the filtered molecules obtained from Mona®[14].

---

[13]Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database [Internet]. Nucleic Acids Research. 2012 Jul 01; 40(W1): W409-W414

[14]Hilbig M, Urbaczek S, Groth I, Heuser S, Rarey M. MONA. Zentrum fur Bioinformatik: Universitat Hamburg; 2013.

[15]Tripos LP. SYBYL-X 1.1. Rohm and Haas Company; 1991.

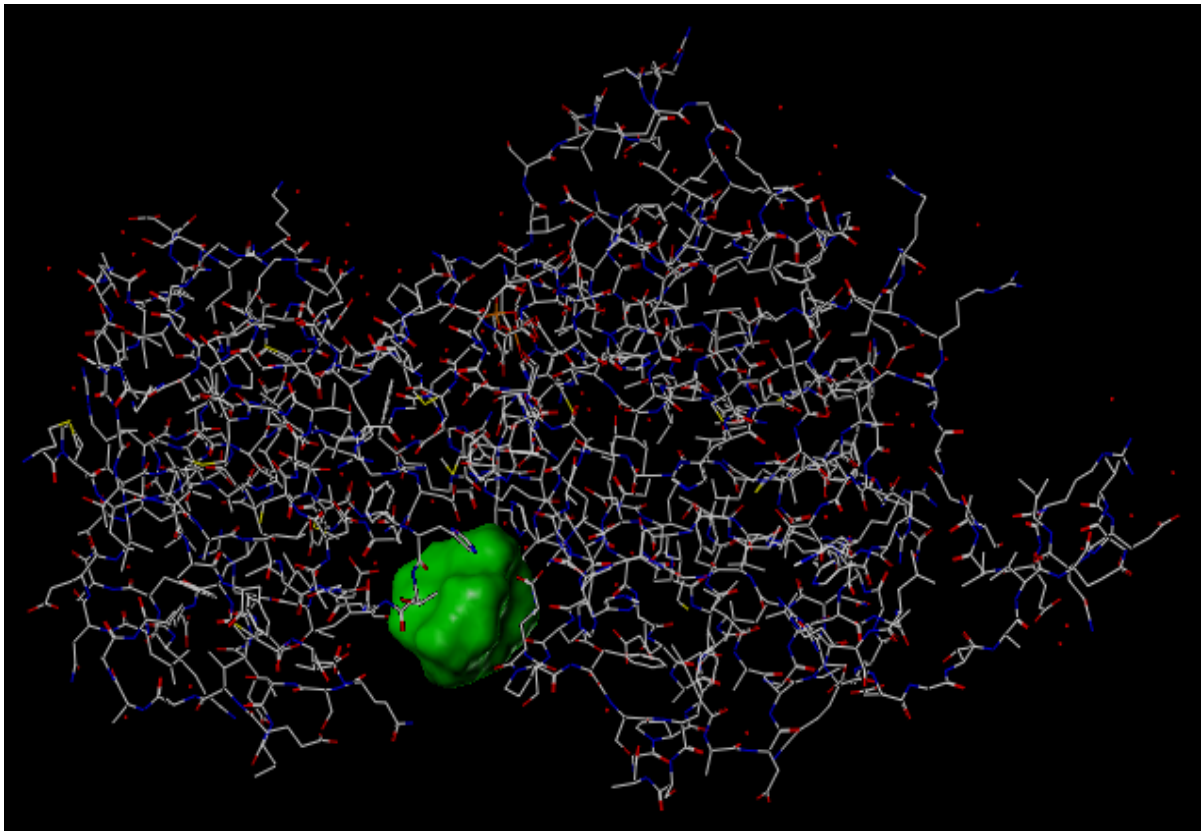


Fig 2.12: Representation of the protomol. Modelled in SYBYL-X®[15]

2.2.5. Ranking of molecules in order of affinity to protomol

A table with a total score was procured and this table was saved in SYBYL-X®[15]. This total score function identified the highest affinity lead-like molecule via ranking of the molecules in order of affinity to the protomol. The high affinity lead-like molecules were then proposed for further validation. The top 3 highest affinity structures were then assessed for toxicity using ProTox-II®[16].

[15]Tripos LP. SYBYL-X 1.1. Rohm and Haas Company; 1991.

[16]Charite Universitatsmedizin Berlin. ProTox-II. 2023

2.3. *De novo* design

2.3.1. Generation of 2D topology maps

Poseview®[17] was used to create 2D topology maps describing the critical interactions between CtBP and the small molecule inhibitors MTOB and HIPP respectively. This was done in order to guide seed structure modelling during the *de novo* design process since these topology maps highlight the areas of the molecule that are critical to binding. These are areas of the molecule that make critical contacts with the receptor. pdb crystallographic depositions 4LCE (Hilbert *et al*., 2014) describing the bound coordinates of CtBP1 with small molecule antagonist MTOB, and 4U6Q (Hilbert *et al*., 2015) describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP, were read sequentially into Poseview®[17] and the ligand binding pocket was mapped. The red areas denote the areas of interaction.

[17]Stierand K, Rarey, M. PoseView [Internet]. Zentrum fur Bioinformatik: Universitat Hamburg; 2010.

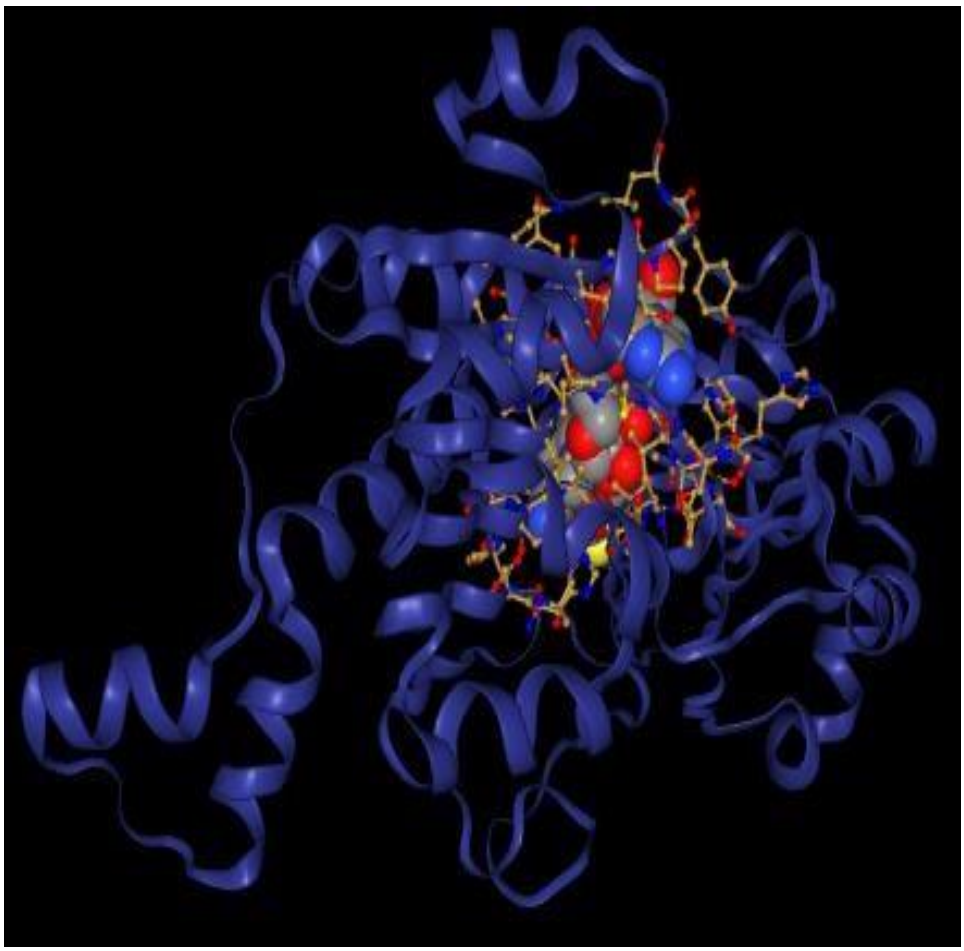


Fig 2.13: Illustration of the CtBP ligand binding pocket bound to small molecule antagonist MTOB as described in pdb crystallographic depositions 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB. Modelled in PoseView®[17]

[17]Stierand K, Rarey, M. PoseView [Internet]. Zentrum fur Bioinformatik: Universitat Hamburg; 2010.

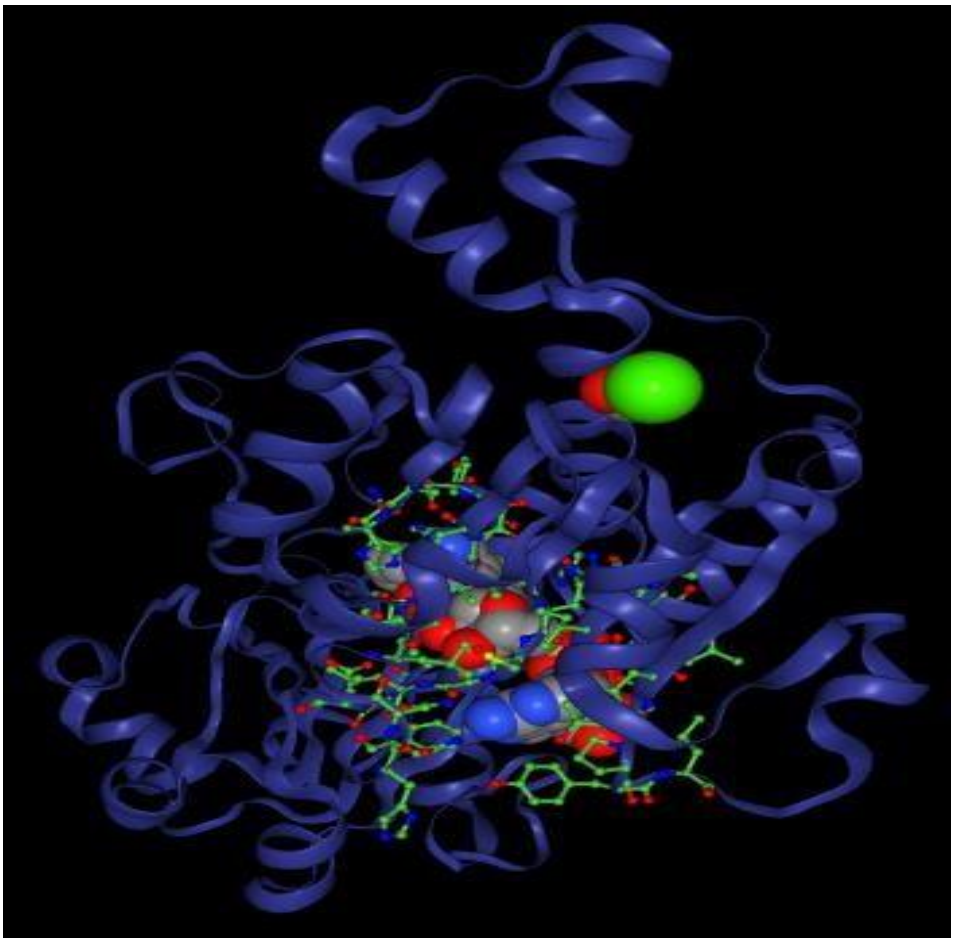


Fig 2.14: Illustration of the CtBP ligand binding pocket bound to small molecule antagonist HIPP as described in pdb crystallographic deposition 4U6Q (Hilbert *et al.*, 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in PoseView®[17]

These maps below show the interactions that MTOB and HIPP forged with the CtBP ligand binding pocket as described in pdb crystallographic deposition 4LCE (Hilbert *et al*., 2014) describing the bound coordinates of CtBP1 with small molecule antagonist MTOB and in pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015) describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP, on an atomic level. The dashed lines represent the hydrogen bonds while the spline segments represent the hydrophobic contacts between the ligand and the protein.

[17]Stierand K, Rarey, M. PoseView [Internet]. Zentrum fur Bioinformatik: Universitat Hamburg; 2010

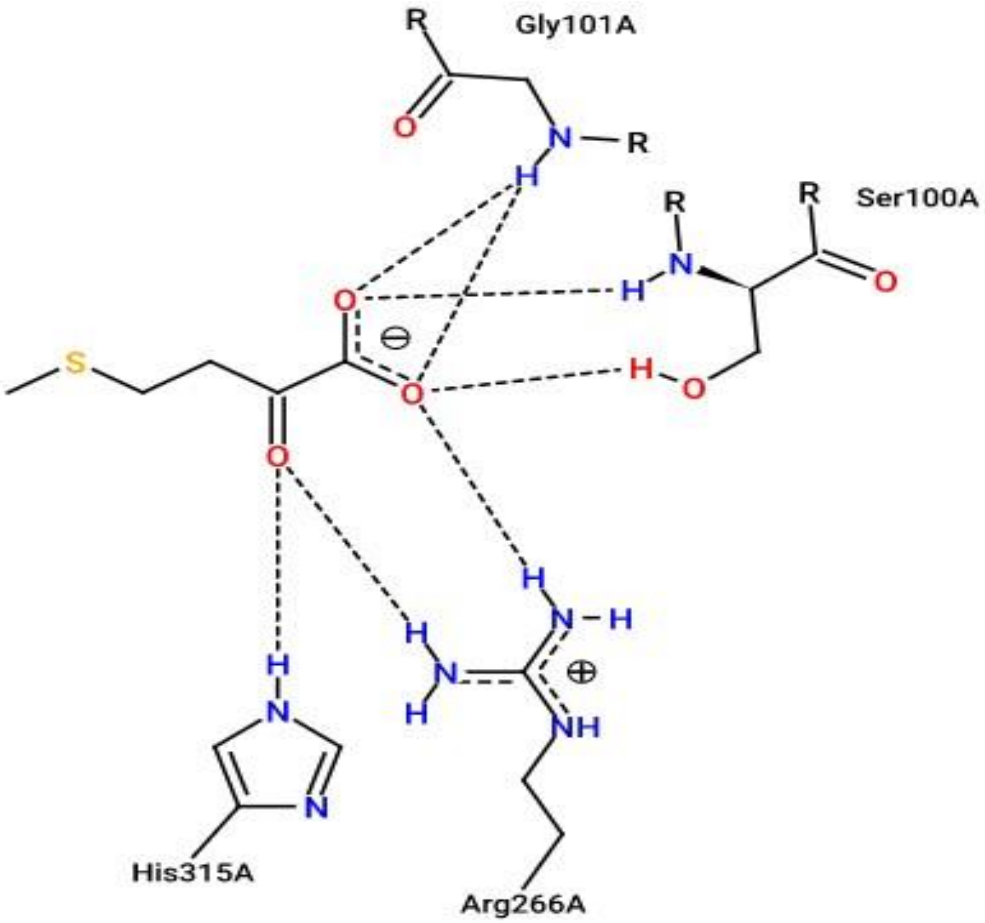


Fig 2.15: 2D map of the ligand-protein interactions between CtBP and MTOB as described in pdb crystallographic depositions 4LCE (Hilbert *et al*., 2014), describing the bound coordinates of CtBP1 with small molecule antagonist MTOB. Modelled in PoseView®[17]

[17]Stierand K, Rarey, M. PoseView [Internet]. Zentrum fur Bioinformatik: Universitat Hamburg; 2010.

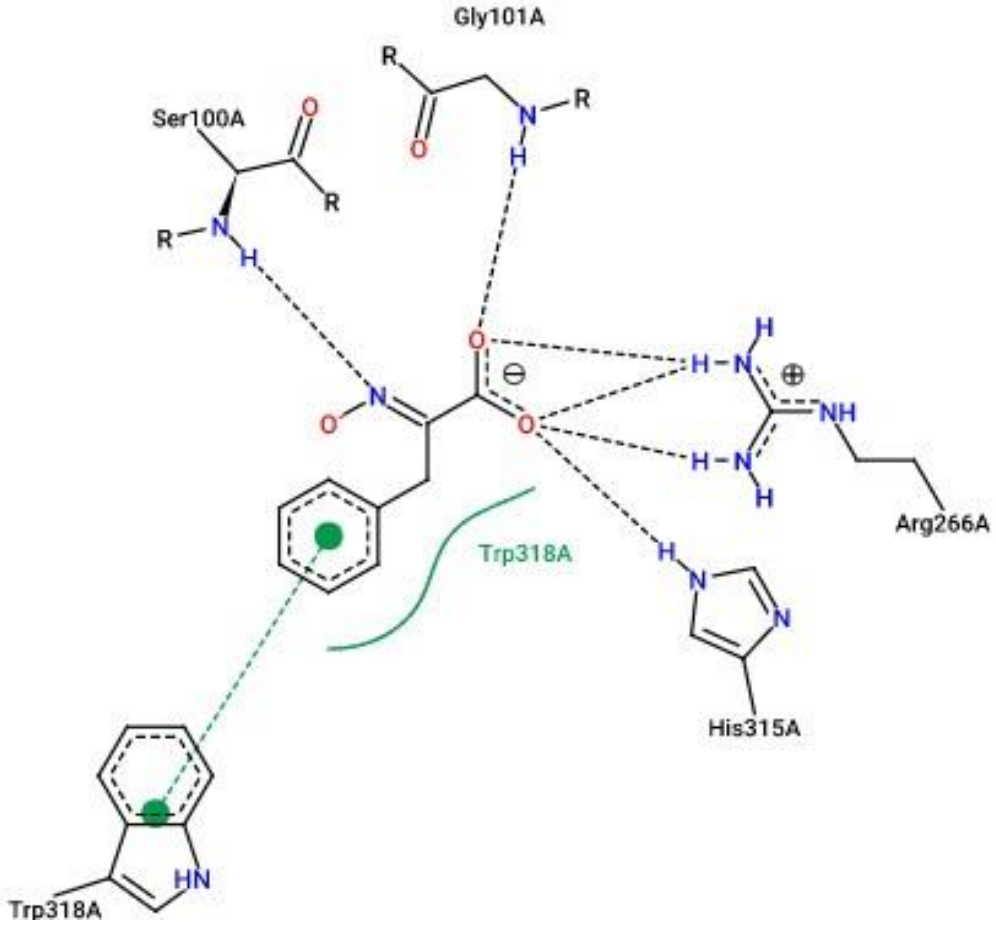


Fig 2.16: 2D map of the ligand-protein interactions between CtBP and HIPP as described in pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015), describing the bound coordinates of CtBP1 with small molecule inhibitor HIPP. Modelled in PoseView®[17]

2.3.2. Modelling of the ligand binding pocket

LigBuilder®[18] was used to create a 3D map of the ligand binding pocket using the module called "Pocket". LigBuilder®[18] has 4 main modules in all, including "Pocket", "Grow", "Link", and "Process". The ligand binding pocket is the area within which docking can take place and where the prospective seeds will grow. It is also a bioactive space delineated by the molecules.

[17]Stierand K, Rarey, M. PoseView [Internet]. Zentrum fur Bioinformatik: Universitat Hamburg; 2010.
[18]LigBuilder v1.1. Wang R, Gao Y, Lai L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design

This was created using pdb crystallographic deposition 4U6Q (Hilbert *et al*., 2015). There are 3 other modules in LigBuilder®[18] which include the fragment.mdb, forbidden.mdb and toxicity.mdb. The fragment database is used to construct the molecules and to add fragments. The forbidden database consists of fragments that cannot be added due to synthetic feasibility and the toxicity database consists of fragments that are known to be toxic so need to be excluded.

2.3.3. Extraction of ligand and modelling of seed structures

The HIPP ligand was extracted in "mol2" format from the CtBP receptor using SYBYL-X®[15] in order to model the seed structures upon. The seed structures were created in SYBYL-X®[15] by removing parts of the molecule that from the 2D topology maps were found to be ineffective for binding and then modifying the atom type to "H.spc". This atom type was detected by LigBuilder®[18] as a site from which it can modify and grow. Then the bond type is chosen as that of a single bond so as there is more freedom in growing and the bond length was chosen as 1. Different areas of the molecule were modified at different attempts to get seed structures. The "Grow" and "Link" modules in LigBuilder®[18] were used to determine the successful seeds and to generate new molecules out of the optimal seeds by running commands through the Ubuntu®[19] terminal.

---

[15]Tripos LP. SYBYL-X 1.1. Rohm and Haas Company; 1991.

[18]LigBuilder v1.1. Wang R, Gao Y, Lai L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design

[19]Canonical. Ubuntu. Canonical ltd. 2023

The “Grow” module sustains new growth as there is a unidirectional growth from a terminus, whilst the “Link” module links two termini together. Within the “Grow” and “Link” modules, copies of the index files for each seed were saved as FILE1 for seed 1 and so on. An example of the commands inputted though the Ubuntu®[19] terminal for the “Grow” module is:

- cd Software
- cd LigBuilderv1.2
- cd grow
- ./grow FILE1

The successful seeds had up to 20 generations on the Ubuntu®[19] terminal. Then the “Process” module was used to copy the index file and this was saved as “Process1” and then the following commands were entered in the Ubuntu®[19] terminal:

- cd Software
- cd LigBuilderv1.2
- cd process
- ./process PROCESS1

If the command works, then there would be a certain number of molecules in the “results.mdb” folder that is found in the “Process” module. The maximum total output of molecules for each working seed is 200 molecules. The “Process” module was then utilized to collect all the molecules and organize these new molecules into the pharmacophorically similar families.

[19]Canonical. Ubuntu. Canonical ltd. 2023

These molecules were then filtered for Lipinski rule (Lipinski *et al*., 2001) compliance using Mona®[14] and the optimal molecules with regards to binding affinity and log P were distinguished. This means that there needs to be a compromise between the value of the log P and that of the binding affinity not just the highest binding affinity or the log P that is closest to 5.

These molecules were assessed for toxicity using ProTox-II®[16] by entering the molecular formula in Canonical SMILES format and their predicted toxicity class were noted and compared to the toxicity of the original lead molecule, HIPP. Besides the predicted toxicities in classes, the ProTox-II®[16] output also gave the predicted Lethal Dose 50 (LD_{50}), which is the median dose predicted to kill 50% of a given test population. It is expressed in milligrams of substance per kilogram of body weight (mg/kg) where high values correlate to a lower toxicity and *vice versa*.

[14]Hilbig M, Urbaczek S, Groth I, Heuser S, Rarey M. MONA. Zentrum fur Bioinformatik: Universitat Hamburg; 2013.

[16]Charite Universitatsmedizin Berlin. ProTox-II. 2023

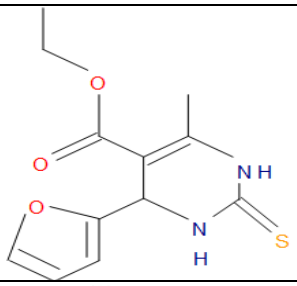
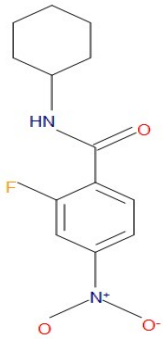
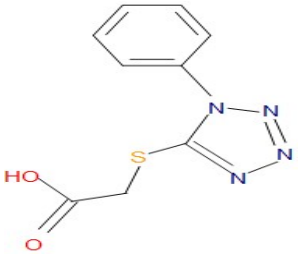
Chapter 3

Results

3.1. Hit structures

Through virtual screening, 674 Lipinski rule (Lipinski *et al*., 2001) compliant hit structures were acquired from the total score table using Sybyl-X®[15]. Lipinski's rule of 5 (Lipinski *et al*., 2001) entails recommendations that foresee the drug-likeness of new synthetic compounds, in this case the hits. The highest affinity lead-like molecule was identified as ZINC56650138 which is named as 3-(((2S,4S)-4-Sulfanylpyrrolidine-2-carbonyl) amino) benzoic acid with a total score of 5.02.

Table 3.1: Top 3 highest affinity hits

ZINC ID, Molecular formula	**Total Score**	**Molecular weight (Da)**	**log P**	**H bond donors**	**H bond acceptors**	**2D structures** (Modelled in Biovia Draw®[1])
ZINC56650138 **$C_{12}H_{14}N_2O_3S$**	**5.02**	**266**	**1.22**	**2**	**5**	
ZINC92955252 **$C_{13}H_{15}FN_2O_3$**	**4.82**	**266**	**1.50**	**1**	**5**	
ZINC08550911 **$C_9H_8N_4O_2S$**	**4.40**	**236**	**1.98**	**1**	**5**	

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

[15]Tripos LP. SYBYL-X 1.1. Rohm and Haas Company; 1991.

These 3 molecules are all Lipinski rule (Lipinski *et al*., 2001) compliant since the molecular weight is less than 500 Daltons, the log P and hydrogen bond donors are less than 5, and the hydrogen bond acceptors are less than 10. The log P value is a constant that forecasts the partition coefficient which is a measure of hydrophobicity. It is usually a negative number if the drug is hydrophilic or a positive number if the drug is lipophilic. For ideal oral and intestinal absorption of the drug, the log P should fall within the range of 1.35 to 1.80, so the top 3 hits obtained are all suitable. This means that here lipophilicity is prominent as this is a crucial parameter that dictates the efficacy and success of the drug. The hydrogen bond acceptor is the atom, ion or molecule component of a hydrogen bond that does not provide the shared hydrogen atom, while a hydrogen bond donor is bond or molecule that provides the hydrogen atom of a hydrogen bond. When the hits were screened for toxicity, these toxicity values were obtained:

Table 3.2: Toxicity screening for hits. Rendered in ProTox-II®[16]

Formula	**Chemical name**	**LD50 (mg/kg)**	**Toxicity class**
$C_{12}H_{14}N_2O_3S$	3-(((2S,4S)-4-Sulfanylpyrrolidine-2-carbonyl) amino) benzoic acid	3160	Class V – slightly toxic
$C_{13}H_{15}FN_2O_3$	1-[2-(3-Fluorophenoxy)acetyl]-1,4-diazepan-5-one	1000	Class IV – slightly toxic
$C_9H_8N_4O_2S$	N-Nitrosobenzthiazuron	1000	Class IV – slightly toxic

[16]Charite Universitatsmedizin Berlin. ProTox-II. 2023

From the toxicity screening, it was found that all of the three hits were slightly toxic. The toxicities of these 3 hits molecules obtained through virtual screening were compared to the toxicity of the lead molecule MTOB. The latter was found to be a non-toxic class V molecule with an LD50 of 12,870mg/kg. Therefore, the MTOB lead had the best toxicity profile when compared to the hits.

3.2. Seed structures

Through *de novo* design two working seeds were created. Seed 1 generated a total of 200 molecules using the "Process" module in LigBuilder®[18] and these were categorized into 4 pharmacophorically similar families that were ranked in order of ligand binding affinity. The latter refers to the magnitude of interaction of ligands with their binding site. These were then filtered down to 9 molecules using Mona®[14] in order to ensure Lipinski rule (Lipinski *et al*., 2001) compliance. The filters applied included the following ranges for each chemical property:

- H bond acceptors between 1 and 10
- H bond donors between 1 and 5
- Lipinski acceptors between 1 and 10
- Lipinski donors between 1 and 5
- log P between 1 and 5
- Molecular weight between 0 and 500 Daltons

---

[14]Hilbig M, Urbaczek S, Groth I, Heuser S, Rarey M. MONA. Zentrum fur Bioinformatik: Universitat Hamburg; 2013.

[18]LigBuilder v1.1. Wang R, Gao Y, Lai L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design.

Table 3.3: The nine filtered molecules obtained from seed 1

Formula	H bond accep-tors	H bond donors	Lipinski acceptors	Lipinski donors	log P	Molecular weight (Da)	Binding affinity
$C_{21}H_{23}NO_6$	6	2	7	2	4.48	385	6.76
$C_{23}H_{29}NO_6$	6	2	7	2	4.74	415	6.58
$C_{22}H_{29}NO_6$	6	2	7	3	5.72	403	6.39
$C_{23}H_{26}N_2O_6$	6	3	8	3	4.35	426	6.76
$C_{26}H_{28}N_2O_6$	6	3	8	4	5.60	464	6.65
$C_{18}H_{18}NO_6$	6	2	7	2	4.35	344	6.55
$C_{19}H_{18}NO_6$	6	2	7	2	4.96	356	6.55
$C_{21}H_{22}NO_6$	6	2	7	2	5.52	384	6.38
$C_{20}H_{22}NO_6$	6	2	7	2	5.02	372	6.38

There were three molecules that were considered as the optimal molecules with regards to their ligand binding affinity and log P. These were:

- $C_{21}H_{23}NO_6$ with a log P of 4.48 and a binding affinity of 6.76 as depicted in Fig 3.1 below:

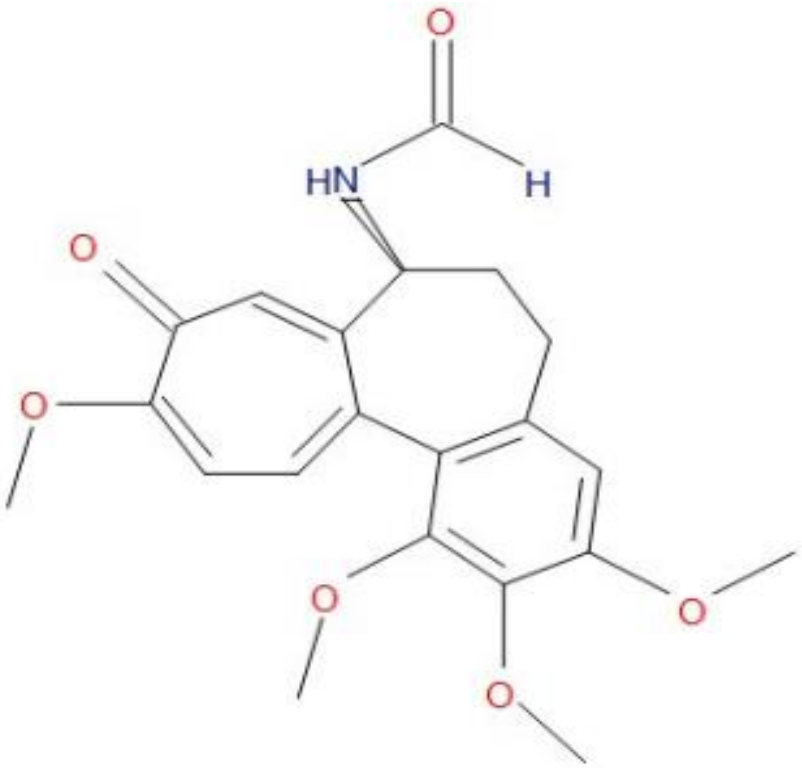


Fig 3.1: Gloriosine ($C_{21}H_{23}NO_6$) obtained through seed 1. Modelled in BioviaDraw®[1]

- $C_{23}H_{26}N_2O_6$ with a log P of 4.35 and a binding affinity of 6.76 as depicted in Fig 3.2.

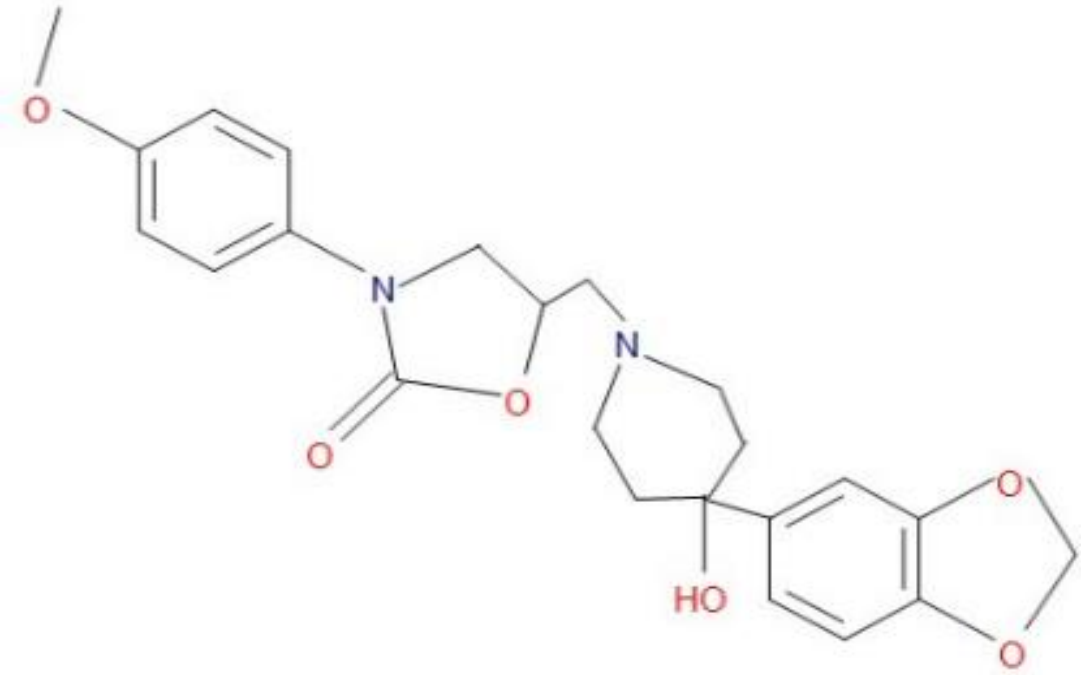


Fig 3.2: $C_{23}H_{26}N_2O_6$ (Panamesine) obtained through seed 1. Modelled in BioviaDraw®[1]

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

- $C_{18}H_{18}NO_6$ with a log P of 4.35 and a binding affinity of 6.55 as depicted in Fig 3.3.

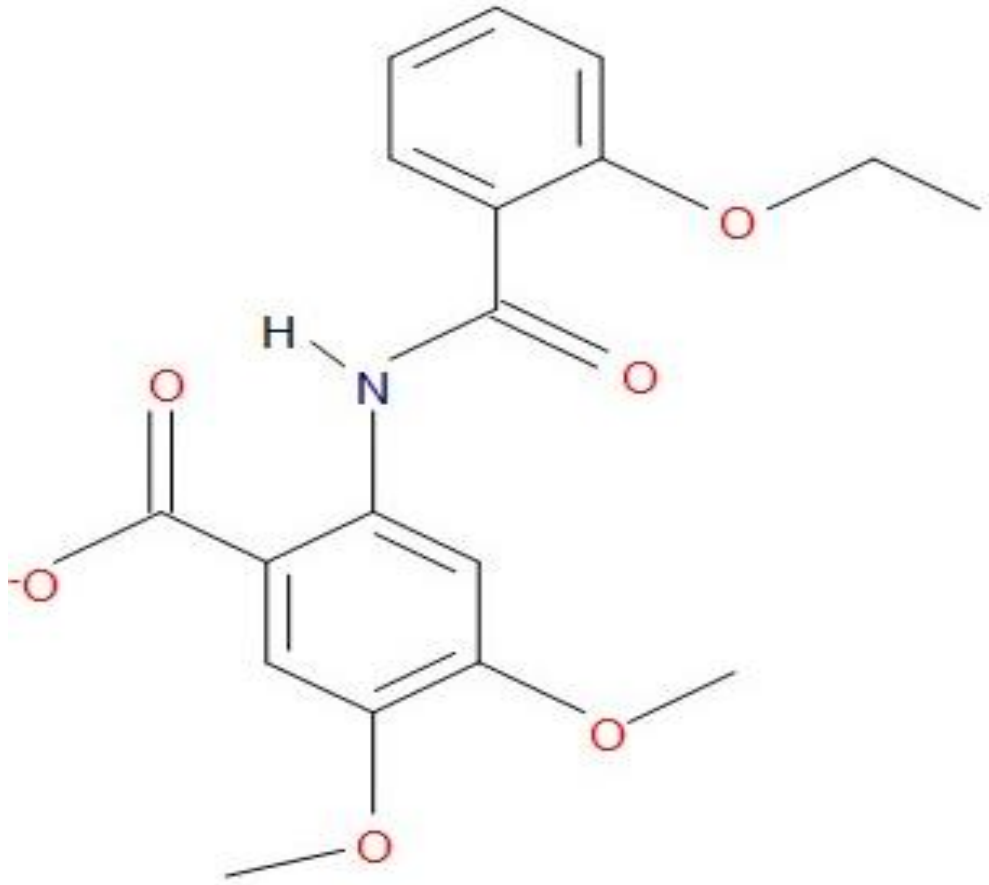


Fig 3.3: $C_{18}H_{18}NO_6$ (2-{[(2-ethoxyphenyl)carbonyl] amino}-4,5-dimethoxybenzoic acid) obtained through seed 1. Modelled in BioviaDraw®[1]

The log P values of the HIPP ligand are slightly better since this has a value of 1.17 which is ideal for oral and intestinal absorption as it is close to the optimal range of 1.35 to 1.80. In this scenario, the HIPP ligand still remains the best option for oral administration since the molecules obtained through *de novo* design have higher log P values.

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

Table 3.4: The four filtered molecules obtained from seed 2

Formula	H bond acceptors	H bond donors	Lipinski acceptors	Lipinski donors	log P	Molecular weight (Da)	Binding affinity
$C_{20}H_{28}NO_3$	3	1	4	1	3.67	330	5.31
$C_{17}H_{22}NO_4$	4	2	5	2	3.19	304	5.27
$C_{19}H_{27}O_4$	4	1	4	1	4.64	319	5.01
$C_{18}H_{24}NO_5$	2	4	6	4	4.04	334	5.05

The second seed generated a total of 5 molecules out of the "Process" module and these were categorized into 2 pharmacophorically similar families, that were also ranked in order of ligand binding affinity. These were filtered down to 4 molecules using Mona®[14]. There were two molecules that were considered the optimal molecules with regards to their ligand binding affinity and log P. The log P values are better than that of the HIPP ligand since this has a negative value of – 0.19. These molecules were:

- $C_{20}H_{28}NO_3$ with a log P of 3.67 and a binding affinity of 5.31 as depicted in Fig 3.4 below:

[14]Hilbig M, Urbaczek S, Groth I, Heuser S, Rarey M. MONA. Zentrum fur Bioinformatik: Universitat Hamburg; 2013.

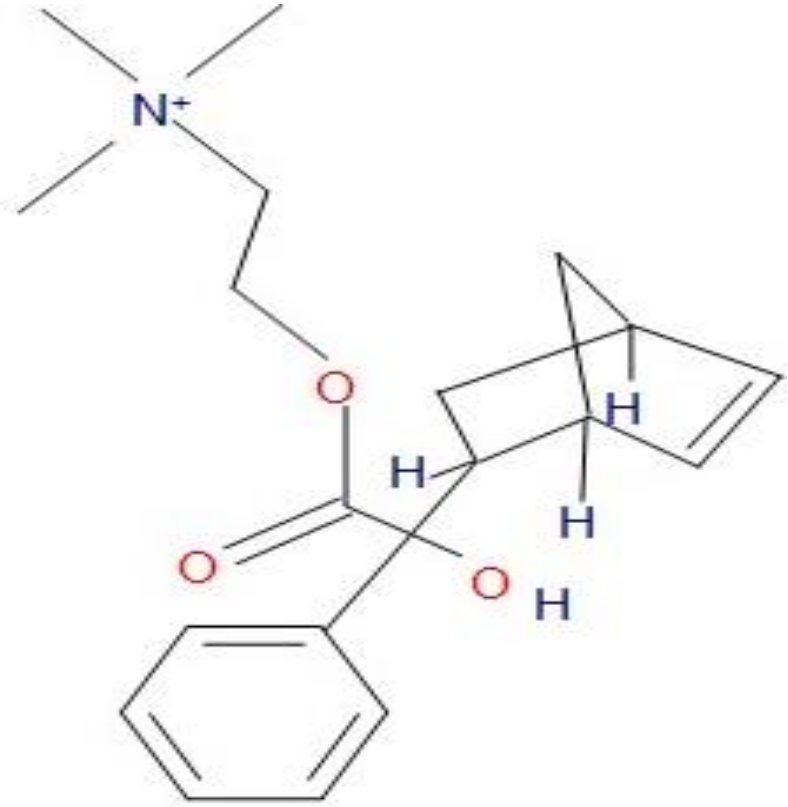


Fig 3.4: $C_{20}H_{28}NO_3$ (Endobenzyline) obtained through seed 1. Modelled in BioviaDraw®[1]

- $C_{18}H_{24}NO_5$ with a log P of 4.04 and a binding affinity of 5.05 as depicted in Fig 3.5.

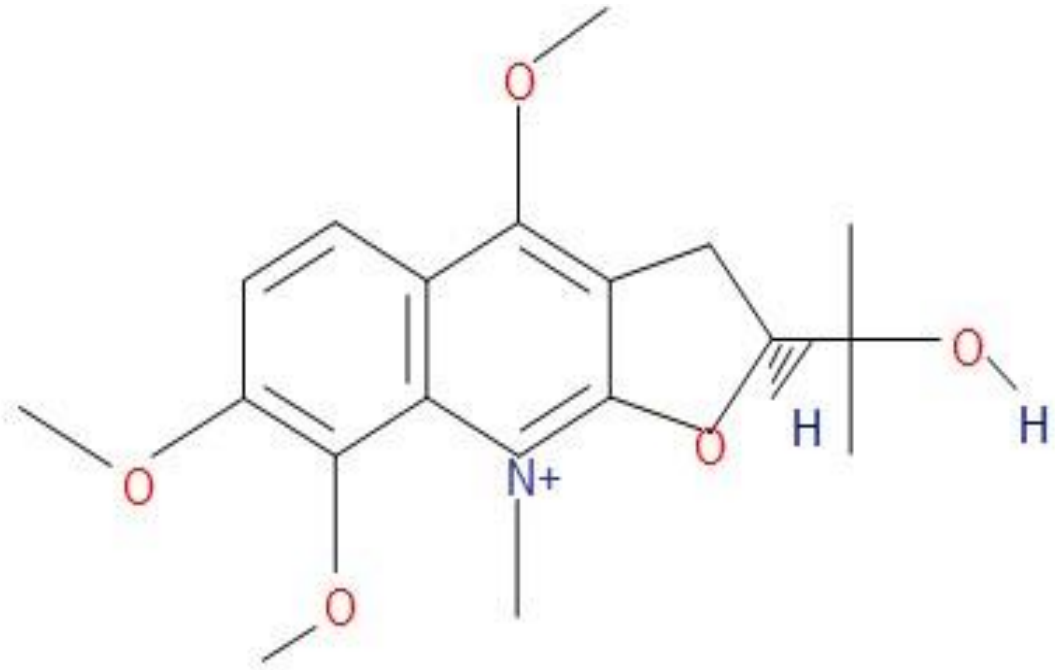


Fig 3.5: $C_{18}H_{24}NO_5$ (Veprisinium) obtained through seed 1. Modelled in BioviaDraw®[1]

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

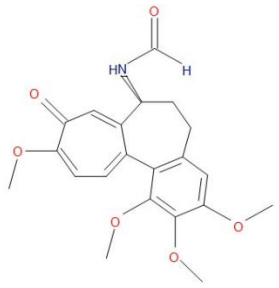
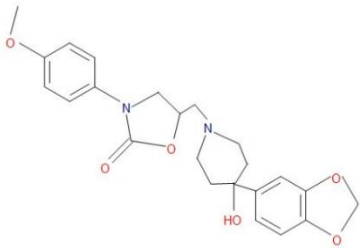
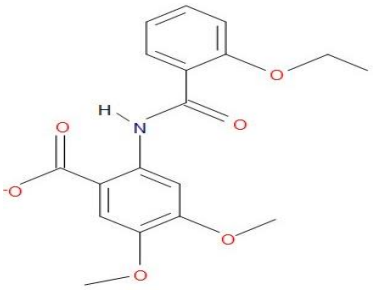
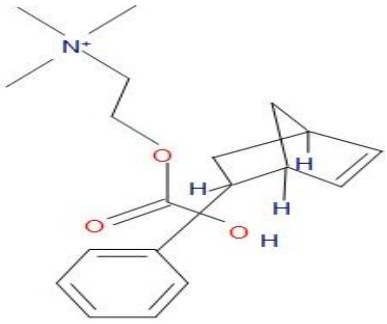
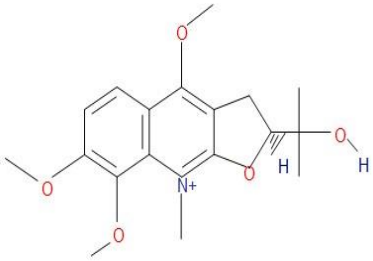
3.3. Toxicity

When the 5 optimal molecules obtained through the *de novo* design process were screened for toxicity using ProTox-II®[16], it was found that three of them were slightly toxic while the other two were moderately toxic and extremely toxic respectively. The toxicity classes ranged from class I to VI with class I being the most fatal and class VI being non-toxic. These classes are determined in accordance with the globally harmonized system of classification of labelling of chemicals. Each class also has a specific range of LD50 values allocated to it, where the higher the LD50 is, the lower is the toxicity.

The toxicities of the 5 molecules obtained through *de novo* design were compared to the toxicity of the lead molecule HIPP. The latter was found to be a slightly toxic class V molecule which may be harmful if swallowed and with an LD50 of 2800mg/kg. Out of the toxicities of the 5 molecules listed in Table 3.4 below, the molecule $C_{18}H_{18}NO_6$ called 2-{[(2-ethoxyphenyl) carbonyl] amino}-4,5-dimethoxybenzoic acid), was the only molecule that had a slightly better toxicity profile when compared to HIPP, because it had an LD50 of 3000mg/kg. However, it still remains a slightly toxic molecule, because the ideal molecule would require an LD50 greater than 5000mg/kg. The other molecules had a worse toxicity profile than HIPP especially $C_{18}H_{24}NO_5$ called veprisinium, which had a moderate toxicity and $C_{21}H_{23}NO_6$ called gloriosine, which was found to be extremely toxic.

[16]Charite Universitatsmedizin Berlin. ProTox-II. 2023

Table 3.5: Predicted toxicities. Rendered in ProTox-II®[16]

Formula	Chemical name	LD 50 (mg/ kg)	Toxicity class	2D structures (Modelled in Biovia Draw®[1])
$C_{21}H_{23}NO_6$	Gloriosine	6	Class 2 – extremely toxic	
$C_{23}H_{26}N_2O_6$	Panamesine	550	Class 4 – slightly toxic	
$C_{18}H_{18}NO_6$	2-{[(2-ethoxyphenyl)carbonyl] amino}-4,5-dimethoxybenzoic acid)	3000	Class 5 – slightly toxic	
$C_{20}H_{28}NO_3$	Endobenzyline	860	Class 4 – slightly toxic	
$C_{18}H_{24}NO_5$	Veprisinium	150	Class 3 – moderately toxic	

[1]Dassault Systèmes BIOVIA. BIOVIA Draw. San Diego: Dassault Systèmes; 2019.

[16]Charite Universitatsmedizin Berlin. ProTox-II. 2023

Chapter 4

Discussion

4.1. Discovery of novel CtBP inhibitors

This study focused on the identification and design of novel small molecules that could be useful for CtBP inhibition due to the fact that literature shows that CtBP transcription repressors represent potentially attractive cancer drug targets. This is because these transcription repressors encode a potentially "druggable" dehydrogenase domain which leads to apoptosis and abrogation of cancer cell migration/invasion (Straza *et al*., 2010). These small molecules could consequently serve as future options for the management of malignant disease.

4.2. Advantages and disadvantages of virtual screening

Virtual screening, which is also known as *in-silico* screening, was undertaken as a computational technique for the identification of molecules that are in possession of the required properties – both spatial and physicochemical, from a large candidate selection resident on molecule libraries. In this study, this approach was utilized to identify new lead candidates in the drug development process, that, due to their homology with MTOB and HIPP, could be considered as potential CtBP inhibitors.

During the virtual screening approach, the generation of shared feature pharmacophores was a powerful tool to combine the knowledge gained from multiple pharmacophores. In this study, a consensus pharmacophore that was essentially an average of the pharmacophoric features of the small molecule antagonists MTOB and HIPP, was modelled. This shared pharmacophore allowed for a more representative and broader result to be obtained from virtual screening since the search was based upon the critical binding features of two and not one, high affinity molecules. The greater the number of molecules that the consensus

pharmacophore incorporates, the more robust it is. The submission of an intact pharmacophore search meant that the search was not based on atoms but on atom types such as hydrophilic, acceptor, donor etc. The robustness of the consensus pharmacophore increases with the incorporation of more molecules. In virtual screening, because of the pharmacophore, anything that conforms to that pharmacophore could be obtained.

Submission of the consensus pharmacophore into the ZINCPharmer®[13] database in order to find similarities, gave ZINCPharmer®[13] the ability to display molecules which are structurally different from the ones that were submitted as a query. Thus, docking the molecules obtained from virtual screening is an innovative process since the actual location of where they are going to be placed in the bioactive space is not known. Upon submission, lead-like filters were applied. Lead-like implies adherence to the stringent Rule of 3 (Congreve *et al*., 2003). Adherence to the Rule of 3 (Congreve *et al*., 2003) is important in the early stages of drug design. In the case of virtual screening, it ensures that the retrieved structures are suitable for further optimization which inevitably includes the insertion of additional moieties. The fact that Rule of 3 (Congreve *et al*., 2003) compliant molecules are small, envisages that the products of their optimization remain Lipinski Rule compliant (Lipinski *et al*., 2001).

[13]Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database [Internet]. Nucleic Acids Research. 2012 Jul 01; 40(W1): W409-W414

The virtual screening process as carried out was consequently pro-active, working towards ensuring Lipinski (Lipinski *et al*., 2001) and Veber Rule (Mahgoub *et al*., 2022) compliance when the optimization process was terminated. Lipinski's rules (Lipinski *et al*., 2001) state that the molecular weight should be less than 500 Daltons, the hydrogen bond acceptor groups less than 10, hydrogen bond donor groups less than 5, and finally the calculated log P less than 5. A compound would be regarded as a high-risk compound in terms of oral bioavailability if it violates more than one of these rules. On the other hand, Veber's rules (Mahgoub *et al*., 2022) increase the criteria for bioavailability as they state that a drug molecule with a good oral bioavailability should consists of less than 10 rotatable bonds and a polar surface area that is less than 140Å. Veber's rules (Mahgoub *et al*., 2022) highlight the fact that a low molecular weight is not needed for oral bioavailability because polarity and molecular flexibility could substantially manage the drugs' bioavailability. The latter could be positively impacted by a reduction in the number of rotatable bonds and the polar surface area of the molecule.

The advantage of finding a lead-like hit as opposed to a drug-like hit is that when moieties are added to the molecule, Lipinski's rule of 5 are still being followed. Another advantage of virtual screening is that there are no restrictions being placed upon the hit structures obtained except for the filters. Thus, in virtual screening, structurally diversified hits are acquired.

After obtaining molecules that are similar to the pharmacophoric structure, the ligand binding pocket of the lead molecule was identified via generation of a protomol, which defined the entire pharmacophoric space available inside the receptor for

binding. This space may be defined as the active site of a receptor in its entirety, and which is ready to bind to small molecules in order to overcome its energetic deficiencies and consequently become more stable. This is the principle behind endogenous molecule and drug binding to protein targets.

In this study, the designed protomol highlighted all the energetically unsatisfied zones in the CtBP receptor due to imperfections in the protein folding process. These imperfections are a consequence of the primary amino acid sequence of the protein and of the hydrophobic effect that drives protein folding. In fact, ligand binding pockets are created via imperfections in the protein folding process which is driven by the hydrophobic effects since the protein will preferably fold such that the hydrophilic amino acids are on the outside while the hydrophobic amino acids are on the inside but this preferred folding does not always occur. Therefore, the quaternary or tertiary structure of a protein are a function of its folding.

The use of the protomol in this study therefore, was synonymous with the exploration of all of unknown zones of the protein implying again that innovation is a major advantage of the virtual screening process. More clearly, the virtual screening process as performed also ensured innovation from the point of view of docking area and of a more complete exploration of pharmacophoric space. Another major advantage of virtual screening is speed with the process being designed to ensure a highly rapid turn-over time.

A major disadvantage of this approach, and the use of the protomol as a docking perimeter centers around predictions of bioactivity. The protomol is by definition all of the energetically unsatisfied space at the core of the protein. There can be no associated claim of bioactivity. Consequently, high affinity to a molecule to the protomol may or may not be associated with bioactivity. The possibility that a novel therapeutic use for a specific small molecule is discovered must also be entertained which once again supports the notion that one of the major strengths of virtual screening is innovation. The molecules identified in this study through virtual screening consequently require thorough *in vitro* analysis.

4.3. Advantages and disadvantages of *de novo* drug design

De novo drug design refers to the creation of novel chemical entities from the very beginning and it is dependent upon the interactions of a biologically active small molecule used as a lead, with the active site of its target receptor. *De novo* design is split into two vital approaches, which include structure-based and ligand-based drug design. Structure-based pertains to the generation of interaction sites for the ligand since the 3D structure of the receptor is known, whilst ligand-based drug design involves the generation of a direct similarity search since the 3D structure of the receptor is undetermined (Mouchlis *et al*., 2021). This structure is crucial since the molecular shape, physical, and chemical properties of the active site are essential for tight and specific binding of a ligand. In this study, a structure-based approach was undertaken since the 3D structure of the CtBP receptor was known.

In the *de novo* structure-based design process as adopted in this study, portions of the lead molecule (seeds) were docked into the target ligand binding pocket, which was crystallographically defined as being bioactive. As opposed to the situation in the virtual screening study where a protomol was used, in this *de novo* approach, there is a greater pre-disposition towards the bioactivity of the designed molecules, and this may be considered as one of the major strengths of this approach.

The major shortcoming of this approach was its lack of ability to truly innovate. Firstly, the seed structures were user modelled, and incorporated, to varying extents fragments of the lead molecules selected for on the basis of their ability to bind to the ligand binding pocket efficiently. These seed structures were, therefore, the primary pharmacophores of the resultant *de novo* modelled structures.

4.4. Molecular cohort evaluation

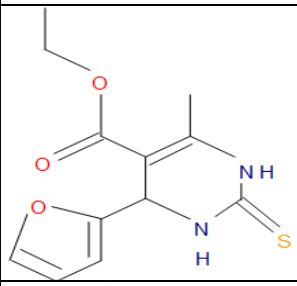
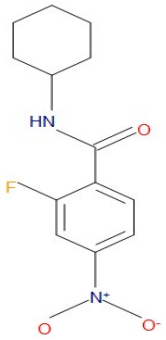
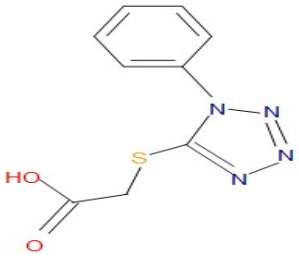
This study adopted the dual approach of virtual screening and *de novo* design owing to the fact that they are complementary in nature with each one compensating for the deficiencies of the other. As previously outlined, while virtual screening was strong on speed and innovation, the main strength of the *de novo* approach was the higher propensity for the designed molecules for bioactivity.

In the virtual screening part of the study, the three molecules with the highest-ranking affinity were identified as depicted in Table 4.1 below. These molecules all have a low molecular weight (236-266 Da) and a low log P (1.22-1.98). The latter means that they are very water soluble, implying a low lipophilicity. Their H bond acceptor and donor values are also all within the required range so all the parameters fall under the requirements for lead-likeness. Out of the 3 hits, the one best suited

for optimization is ZINC08550911 (highlighted in red font in Table 4.1). Of the 3 selected structures this had the lowest affinity for the protomol and the highest log P. While all three molecules have a low log P, (at lead stage the log P value should be around 3) ZINC08550911 has the highest log P value (1.99). In choosing ZINC08550911 for optimization, a compromise is being made between the affinity and log P. This is in accordance with the principles of rational drug design which while recognizing the importance of binding affinity, gives significant importance to physicochemical parameters that have a direct bearing on bioavailability. This reasoning may be defended from the point of view of solubility and permeability.

Molecules with a very low log P are so hydrophilic that they become very intimately bound with water to the extent that dissociation from the intra-luminal water molecules becomes energetically unfeasible. This results in a scenario in which the drug remains in solution in the intestinal lumen and is excreted unchanged in the faeces without ever being absorbed or having the opportunity to exert clinical activity. Of the 3 highest affinity hits resulting from virtual screening, it would probably be easiest to effect structural changes to raise the log P up to 3 with ZINC08550911.

Table 4.1: Top 3 highest affinity hits

ZINC ID, Molecular formula	**Total Score**	**Molecular weight (Da)**	**log P**	**H bond donors**	**H bond acceptors**	**2D structures** (Modelled in Biovia Draw®[1])
ZINC56650138 $C_{12}H_{14}N_2O_3S$	**5.02**	**266**	**1.22**	**2**	**5**	
ZINC92955252 $C_{13}H_{15}FN_2O_3$	**4.82**	**266**	**1.50**	**1**	**5**	
ZINC08550911 $C_9H_8N_4O_2S$	**4.40**	**236**	**1.98**	**1**	**5**	

In the *de novo* design part of the study, the five optimal molecules obtained from the seed structures, had log P values ranging from 3.67 to 4.48 meaning that the molecules are more lipophilic when compared to the hits obtained from virtual screening. In this case there is however the possibility that optimization could skew the logP value to a value exceeding Lipinski's cut-off value of 5, making the molecule too lipophilic and not sufficiently water soluble.

When comparing toxicity profiles obtained from virtual screening and from *de novo* design, it was evident that virtual screening yielded hits that were more toxic than the lead molecule, MTOB as seen in Table 3.2 in page 49. This implies that the toxicity profile of the molecules emanating from this approach should be closely monitored and the moieties responsible for the toxicity identified. A bioisosteric approach in which the toxic moieties were replaced with biologically equivalent less toxic ones may represent a way forward in this case. The molecules deriving from the *de novo* design approach generated three molecules that were slightly toxic while the other two were moderately toxic and extremely toxic respectively. Out of the toxicities of the 5 molecules listed in Table 3.5 in page 57, the molecule $C_{18}H_{18}NO_6$ called 2-{[(2-ethoxyphenyl) carbonyl] amino}-4,5-dimethoxybenzoic acid), was the only molecule that had a slightly better toxicity profile when compared to HIPP. This meant that the *de novo* design approach was more feasible with regards to toxicity profiles when compared to virtual screening.

4.5. Limitations of the study

There are a number of limitations inherent to this study. The most pertinent is that both of the adopted approaches are static and do not take into account the motion inherent to protein:ligand complexes. The use of molecular dynamics simulation studies would give deeper insight into the nuances of the interactions between the designed small molecules and the target receptor. Another limitation stems from the fact that in the case of the molecules deriving from virtual screening it is unclear as to how these would interact with the bioactive active site described in the pdb. A

docking study with subsequent conformational analysis and in silico binding affinity calculation would be an appropriate first step in answering this question.

4.6. Conclusion

CtBP has initiated immense attention and interest due to its function as a co-repressor that targets multiple tumour suppressors, as well as an activator of key oncogenes. The study contributed to drug design through identification of 2 molecular cohorts with the propensity to have good oral bioavailability and to bind with high affinity to the target. Since CtBPs are prominent in multiple cancers, targeting CtBP via novel molecules is an indispensable concept in drug design which could impact further development of new therapies. A high level of progress has been reached to acknowledge the role of CtBP transcriptional complex in tumorigenesis and there are various methods to antagonize CtBP. Therefore, CtBP shows signs of being a revolutionary treatment in future cancer therapy.

References

Acosta-Baena, Tejada-Moreno JA, Arcos-Burgos M, Villegas-Lanau CA. CtBP1 and CtBP2 mutations underpinning neurological disorders: a systematic review. Neurogenetics. 2022; 23(4): 231-240. DOI: 10.1007/s10048-022-00700-w

Bellesis AG, Jecrois AM, Hayes JA, Schiffer CA, Royer UE. Assembly of human C-terminal binding protein (CtBP) into tetramers. Journal of Biological Chemistry. 2018 Jun 08; 293(23): 9101-9112. DOI: 10.1074/jbc.RA118.002514

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rogers JR, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. Journal of Molecular Biology. 1977; 112(3): 535–42. DOI: 10.1111/j.1432-1033.1977.tb11885.x.

Birts CN, Nijjar SK, Mardle CA, Hoakwie F, Duriez P, Blaydes JP, *et al*. A cyclic peptide inhibitor of C-terminal binding protein dimerization links metabolism with mitotic fidelity in breast cancer cells. Chemical Science. 2013 Aug 01; 4(8): 3046-3057. DOI: 10.1039/c3sc50481f

Blevins MA, Huang M, Zhao R. The Role of CtBP1 in Oncogenic Processes and Its Potential as a Therapeutic Target. Molecular Cancer Therapeutics. 2017 Jun; 16(6): 981-990. DOI: 10.1158/1535-7163

Blevins MA, Kouznetsova J, Krueger AB, King R, Mathews Griner L, Hu X, *et al.* Small molecule, NSC95397, inhibits the CtBP1-protein partner interaction and CtBP1-mediated transcriptional repression. Journal of Biomolecular Screening. 2015 Jun; 20(5): 663-672. DOI: 10.1177/1087057114561400

Byun JS, Gardner K. C-Terminal Binding Protein: A Molecular Link between Metabolic Imbalance and Epigenetic Regulation in Breast Cancer. International Journal of Cell Biology. 2013; 2013: 1-14. DOI: 10.1155/2013/647975

Byun JS, Park S, Yi DI, Shin JH, Hernandez SG, Hewitt SM, *et al*. Epigenetic re-wiring of breast cancer by pharmacological targeting of C-terminal binding protein. Cell death and Disease. 2019 Oct 01; 10(10): 689. DOI: 10.1038/s41419-019-1892-7

Chen Z. The transrepression and transactivation roles of CtBPs in the pathogenesis of different diseases. Journal of Molecular Medicine. 2021; 99(10): 1335-1347. DOI: 10.1007/s00109-021-02107-w

Congreve M, Carr R, Murray C, Jhoti H. A 'Rule of Three' for fragment-based lead discovery? Drug Discovery Today. 2003 Oct 01; 8(19): 876-877. DOI: 10.1016/s1359-6446(03)02831-9

Cousins SL, Stephenson FA. Identification of C-Terminal Binding Protein 1 as a Novel NMDA Receptor Interactor. Neurochemical Research. 2018; 44(6): 1437-1445. DOI: 10.1007/s11064-018-2633-5

Dcona MM, Morris BL, Ellis KC, Grossman SR. CtBP – an emerging oncogene and novel small molecule drug target: Advances in the understanding of its oncogenic action and identification of therapeutic inhibitors. Cancer Biology and Therapy. 2017 Jun 03; 18(6): 379391. DOI: 10.1080/15384047.2017.1323586

Han L, Feng Z, Xingjian C, Ping Z, Xiaobo H, Wei X, *et al*. C-terminal binding protein 2 promotes high-glucose-triggered cell proliferation, angiogenesis and cellular adhesion of human retinal endothelial cell line. International Ophthalmology. 2022; 42(10): 2975-2985. DOI: 10.1007/s10792-022-02283-9.

Hilbert BJ, Grossman SR, Schiffer CA, Royer WE. Crystal structures of human CtBP in complex with substrate MTOB reveal active site features useful for inhibitor design. Federation of European Biochemical Societies. 2014 May 12; 588(9): 1743-1748. DOI: 10.1016/j.febslet.2014.03.026

Hilbert BJ, Grossmann SR, Schiffer CA, Royer WE. CtBP1 in complex with substrate MTOB. 2014; *FEBS Lett* 588: 1743-1748.

Hilbert BJ, Morris BL, Ellis KC, Paulsen JL, Schiffer CA, Grossman SR, *et al*. Structure-guided design of a high affinity inhibitor to human CtBP. ACS Chemical Biology. 2015 Apr 17; 10(4): 1118-1127. DOI: 10.1021/cb500820b

Hilbert BJ, Morris BL, Ellis KC, Paulsen JL, Schiffer CA, Grossman SR, et al. (2015) CtBP1 bound to inhibitor 2-(hydroxyimino)-3-phenylpropanoic acid. *ACS Chem Biol* 10: 1118-1127

Hilbig M, Urbaczek S, Groth I, Heuser S, Rarey M. MONA – Interactive manipulation of molecule collections. Journal of Cheminformatics. 2013; 5(1): 38. DOI: 10.1186/1758-2946-5-38.

Jack B H.A., Pearson RC, Crossley M. C-terminal binding protein: A metabolic sensor implicated in regulating adipogenesis. International Journal of Biochemistry and Cell Biology. 2011 May; 43(5): 693-696. DOI: 10.1016/j.biocel.2011.01.017

Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database. Nucleic acids research. 2012; 40: 409-414. DOI: 10.1093/nar/gks378.

Korwar S, Morris BL, Parikh HI, Coover RA, Doughty TW, Love IM, *et al*. Design, synthesis, and biological evaluation of substrate-competitive inhibitors of C-Terminal Binding Protein (CtBP). Bioorganic and Medicinal Chemistry. 2016 Jun 15; 24(12): 2707-2715. DOI: 10.1016/j.bmc.2016.04.037

Liao C, Sitzmann M, Pugliese A, Nicklaus MC. Software and resources for computational medicinal chemistry. Future Medicinal Chemistry. 2011 Jun; 3(8): 1057-1085. DOI: 10.4155/fmc.11.63

Li H, Zhang C, Yang C, Blevins M, Norris D, Zhao R, *et al*. C-terminal binding proteins 1 and 2 in traumatic brain injury-induced inflammation and their inhibition as an approach for anti-inflammatory treatment. International Journal of Biological Sciences. 2020; 16(7): 1107-1120. DOI: 10.7150/ijbs.42109.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2001 Mar 01; 46(1-3): 3-26. DOI: 10.1016/s0169-409x(00)00129-0

Mahgoub RE, Atatreh N, Ghattas MA. Using filters in virtual screening: A comprehensive guide to minimize errors and maximize efficiency. In: Caballero J, editor. Annual Reports in Medicinal Chemistry. Virtual Screening and Drug Docking. 59. USA: Elsevier; 2022.p.2-252.

Mouchlis VD, Afantitis A, Serra A, Fratello M, Papadiamantis AG, Aidinis V, *et al*. Advances in de novo drug design: From conventional to machine learning methods. International Journal of Molecular Sciences. 2021; 22(4):1676. DOI: 10.3390/ijms22041676.

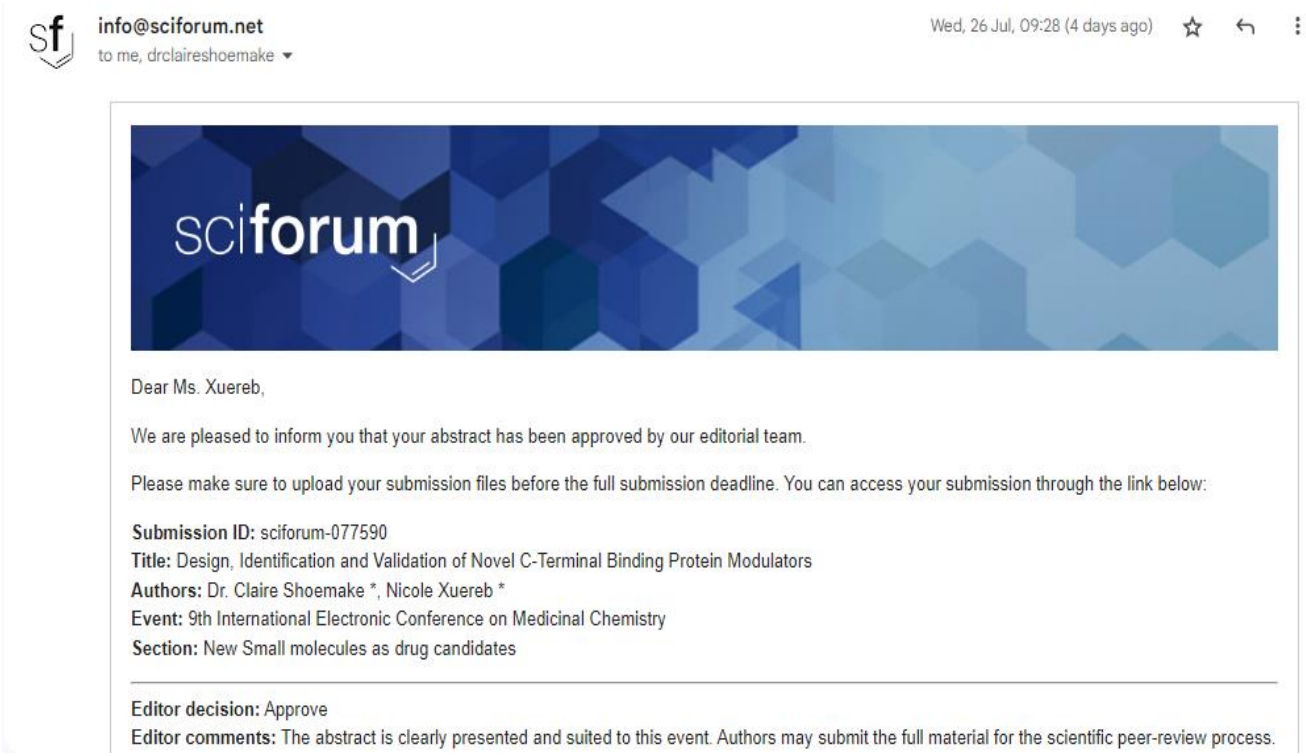
Stankiewicz TR, Gray JJ, Winter AN, Linseman DA. C-terminal binding proteins: Central players in development and disease. Biomolecular Concepts. 2014 Dec 5; 5(6): 489-511. DOI: 10.1515/bmc-2014-0027

Straza MW, Paliwal S, Kovi RC, Rajeshkumar B, Trenh P, Parker D, *et al*. Therapeutic targeting of C-terminal binding protein in human cancer. Cell Cycle. 2010 Sep 15; 9(18): 3740-3750. DOI: 10.4161/cc.9.18.12936

Yang X, Sun Y, Li H, Shao Y, Zhao D, Yu W, *et al*. C-terminal binding protein-2 promotes cell proliferation and migration in breast cancer via suppression of $p16^{INK4A}$. Oncotarget. 2017 Apr 18; 8(16): 26154-26168. DOI: 10.18632/oncotarget.15402

Yuan Y, Pei J, Lai L. LigBuilder 2: A practical *de novo* Drug Design Approach. Journal of Chemical Information and Modelling. 2011; 51(5): 1083-1091. DOI: 10.1021/ci100350u

List of Publications and Abstracts



info@sciforum.net
to me, drclaireshoemake

Wed, 26 Jul, 09:28 (4 days ago)

sciforum

Dear Ms. Xuereb,

We are pleased to inform you that your abstract has been approved by our editorial team.

Please make sure to upload your submission files before the full submission deadline. You can access your submission through the link below:

Submission ID: sciforum-077590
Title: Design, Identification and Validation of Novel C-Terminal Binding Protein Modulators
Authors: Dr. Claire Shoemake *, Nicole Xuereb *
Event: 9th International Electronic Conference on Medicinal Chemistry
Section: New Small molecules as drug candidates

Editor decision: Approve
Editor comments: The abstract is clearly presented and suited to this event. Authors may submit the full material for the scientific peer-review process.

Appendix I: Ethics Approval

The status of your REDP form (MED-2022-00397) has been updated to Acknowledged Inbox ×

form.urec@um.edu.mt Tue, 14 Mar, 10:57

to me

Dear Nicole Xuereb,

Please note that the status of your REDP form (MED-2022-00397) has been set to *Acknowledged*.

This status change was accompanied by the following explanation/justification: *Dear applicant, Your research ethics application has been received. As indicated in the Research Ethics Review Procedures, submissions which have no self-assessment issues are kept for record and audit purposes only, so research may commence. Kindly note that FREC will not issue any form of approval as the responsibility for the self-assessment part lies exclusively with the researcher. Regards, MED FREC*

You can keep track of your applications by visiting: https://www.um.edu.mt/research/ethics/redp-form/frontEnd/.

This email has been automatically generated by URECA. Please do not reply. If you wish to communicate with your F/REC please use the respective email address.