Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 8(1) – Year: 2024 – Pages: 40-53

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.1249159

Received: 08.02.2023 ***Accepted***: 15.03.2023 *Research Article*

Discovery of Repurposable Drugs in the Combination Therapy of Breast Cancer: A Virtual Drug Screening Study

Ebru Kırmızıay, Rümeysa Demir, Ceren Öğütçü, Hüseyin Saygın Portakal [1]

Izmir University of Economics, Genetics and Bioengineering Department

Abstract: Cathepsin D (Cat D) is a lysosomal aspartic acid protease encoded by CTSD gene and has significant biological roles such as degradation of extracellular and intracellular proteins, regulation of apoptosis, hormone processing, antigen processing etc. Furthermore, it is overexpressed by breast cancer cells and it acts a role in many processes affecting the cancer prognosis such as metastasis, angiogenesis, invasion, and drug resistance through regulation of the metabolic pathways and digesting the extracellular matrix (ECM) proteins. Due to that there is no drug targeting Cat D in clinical trial phases, a virtual drug screening in order to reveal possible drugs with high Cat D inhibitory activity from a library composed of 12,111 ligands is carried out with this study. Results have demonstrated that ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), ZINC000003786250 (YM022), and ZINC000150338819 (Ledipasvir) have high binding affinity to Cat D. Among these chemical ligands, YM022 from Drugs in Clinical Trials dataset has been evaluated as most promising one that might be repurposed in the treatment of breast cancer due to its high affinity, convenient ADME and Toxicity properties, and highest bioactivity profiles. However, the possible activity of YM022 should be analyzed with further molecular dynamics (MD) simulations, in vitro and in vivo studies.

Keywords: Breast Cancer, Cathepsin D, Cathepsin Inhibitors, Protease Inhibitors, Virtual Drug Screening, Drug Repurposing

1. Introduction

Due to the fact that treatment of cancers with chemotherapy applications are limited by several cancer cells' strategies such as metastasis, invasion, angiogenesis, etc., combination of several drug molecules targeting distinct parameters is one of the major approaches [1]. Strikingly, regulation and degradation of extracellular matrix (ECM) are the primary reasons providing these strategies, particularly metastasis and invasion [2]. ECM consists of collagen, fibronectin, and laminin proteins as well as many digestive enzymes such as serine proteases, matrix metalloproteases, and cathepsin derivatives that are responsible for ECM regulation [3].

While proteases are classified into three groups according to their active amino acid residues as cysteine proteases (B, C, H, K, L, O, S, V, X, W), aspartic acid proteases (D and E), and serine proteases (A and G), Cathepsin D (Cat D) is a lysosomal aspartic acid protease encoded by CTSD gene and a member of pepsin superfamily [4–7]. Cat D is synthesized as inactive form (pre-proenzyme) and in order to gain activity it undergoes various post-translational modifications such as glycolisation and ubiquitination [8]. Due to its proteolysis activity, Cat D acts significant roles in many physiological processes such as extracellular and intracellular protein degradation, antigen processing, activation of enzyme precursors, hormone processing, cytoskeletal protein degradation, ECM remodeling, and regulation of apoptosis [9–14]. Considering the functions of Cat D, the connections between many pathologies such as osteoarthiritis, acute pancreatitis, neurodegenerative disorders

[1] Corresponding Authors

e-mail: saygin.portakal@ieu.edu.tr

(Alzheimer, Neuronal Ceroid Lipofuscinosis (NCL), Niemann Pick type C (NPC), etc.), and overexpression or mutations of Cat D have been reported in literature over many years [15–17]. In addition, while many cancers express Cat D in order to manipulate their ECM, overexpression of it in breast cancer has most attention of the studies [18], [19]. Principally, released Cat D digests extracellular C terminal domain of SPARC in order to provide metastasis and invasion to breast cancer cells and it causes to proliferation through releasing the growth factors to tumor microenvironment (TME) [20,21]. Today, the expression pattern and role in cancer prognosis make Cat D as biomarker for diagnosis and attractive drug target in the treatment of breast cancer [22].

Despite the significance in breast cancer as well as other pathologies, any drug molecule against Cat D has not reached to clinical trial phases. Nonetheless, Pepstatin A's inhibitory activity has been reported with many publications and it's used as model drug targeting Cat D [23]. Furthermore, few ligands' Cat D inhibitory activity have been shown with few data [24]. As such, to discover and to design novel therapeutics are quite promising in Cat D mediated diseases. Virtual drug screening is one of the novel approaches in order to reveal possible ligands that might inhibit target protein. This novel low cost and timesaving technique has been carried out for many diseases and its efficiency has been proved [25]. In this study, a virtual drug screening strategy based on the molecular docking of totally 12,111 chemical ligands composed of FDA-Approved Drugs, World-not-FDA Approved Drugs, Drugs in Clinical Trials, and Non-human Metabolites datasets from ZINC15 database to crystal structure of Cat D has been performed to reveal possible inhibitors. Results have shown that ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), ZINC000003786250 (YM022), and ZINC000150338819 (Ledipasvir) have highest binding affinities within the created library. Furthermore, due to its convenient ADME, toxicity, and bioactivity profiles, gastrin/CCKB antagonist YM022 is evaluated as most promising one to be repurposed for inhibition of metastasis, invasion, and angiogenesis of breast cancer through inhibition of Cat D activity. While the potential of YM022 in combination therapy of breast cancer is declared in this study, yet its stability, and efficiency should be tested with molecular dynamics (MD) simulations as well as in vitro and in vivo studies.

2. Computational Method

2.1. Protein Preparation

Cathepsin D chemical structure including N-(3,4-dimethoxybenzyl)-Nalpha-{N-[(3,4-dimethoxyphenyl) acetyl] carbamimidoyl}-D-phenylalaninamide (2RZ) inhibitor was retrieved from Protein Data Bank (PDB) (PDB ID: 4OD9). The protein file's resolution, R-value (observed), R-value (free) parameters are 1.90 Å, 0.185, and 0.216, respectively. The protein was prepared by Dock Prep module of UCSF Chimera software (version of 1.16). During the preparation of the protein, heteroatoms, ligands, and water molecules were removed, hydrogen atoms and partial charges were added, and side chains were replaced with Dunbrack 2010 rotamer library [26].

2.2. Ligand Library Preparation

A library consists of 12,111 drug molecules were designed by retrieving World-not-FDA Approved Drugs (4,288 ligands), FDA-Approved Drugs (1,615 ligands), Drugs in Clinical Trials (3.897 ligands), and Non-human Metabolites (2,313 ligands) datasets from ZINC15 database. Once to separately import the datasets to PyRx Virtual Screening tool, the ligands were geometrically minimized by the energy minimization module of the software [27].

2.3. Virtual Drug Screening

AutoDock Vina package of PyRx Virtual Screening Tool were utilized for molecular docking based virtual drug screening [28]. As such, by considering the protein region interacting with 2RZ, grid box were created with the coordinates x= -4.4045, y= 16.2359, and z= 34.4130 and dimensions 20 x 20 x 20. The .csv files of the data demonstrating the docked ligands' binding affinities, rmsd/ub, and rmsd/lb values were exported. 30 hit drugs with highest binding affinity scores from each datasets were investigated and the modes of them with 0 rmsd/ub, and 0 rmsd/lb values were exported in .pdb format. Consequently, their interactions with

the protein were investigated in Biovia Discovery Studio Visualiser software.

2.4. Validation

In order to validate the efficiency of the developed molecular docking strategy, 2RZ inhibitor found in the chemical structure of the protein were exported as a separate file and docked to protein by following the same protocol. Besides, a novel inhibitor library consists of totally 8 ligands whose Cat D inhibitory activity had been demonstrated in literature were designed by retrieving the chemical structures from PubChem database with the Compound CIDs: 19098, 231736, 3080772, 5478883, 24847649, 25066229, 44449299, 56950381, 103091773, and investigated with the same strategy.

2.5. Protein-Protein Docking with HADDOCK

Cat D-Transglutaminase 2 (Tgase 2) complexes were created by protein-protein docking applications with HADDOCK online tool (version 2.4) [29]. As such, crystal structure of Tgase 2 was retrieved from PDB with the PDB ID: 3LY6, and prepared by following protein preparation protocol. Possible interacting amino acid residues of both proteins were predicted with SPPIDER II online tool [30]. Prepared proteins were submitted to HADDOCK tool and predicted interacting residues were used as input data. Considering the docking score and Z-score, best cluster's highest score complex were retrieved, and the interactions created in protein interface were analyzed with PyMol visualizer software.

2.6. Toxicity, ADME, and Bioactivity Analysis

ADME (absorption, distribution, metabolism, and excretion) and toxicity properties of the four best hit drugs were analyzed with both OSIRIS Property Explorer software [31] and swissADME online server [32]. As such, solubility, lipophilicity, physico-chemical, and pharmacokinetics properties as well as possible toxicities such as mutagenicity, tumorigenicity, irritant effects and reproductive effects were investigated. During this study, ZINC000150338819 (Ledipasvir) was not analyzed since that it had been approved by FDA previously. In addition, since that Cat D is a protease enzyme, protease inhibitory and enzyme inhibitory activities of the five best hit drugs as well as three best scored inhibitors from literature were examined with Molinspiration Cheminformatics v2020 online server by using the SMILE formats of the molecules as an input [33].

3. Results and discussion

In order to reveal possible drug molecules that might target Cat D, a molecular docking based virtual drug screening strategy has been carried out with 12,111 ligands from FDA-Approved Drugs, World-not-FDA Approved Drugs, Drugs in Clinical Trials, and Non-human Metabolites datasets of ZINC15 database. Binding affinities of best 10 hit ligands and the interacting amino acid residues of Cat D with these drugs are demonstrated in Table 1. Results put forward that ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), ZINC000003786250 (YM022) ligands from Drugs in Clinical Trials and ZINC000150338819 (Ledipasvir) from FDA-Approved Drugs data sets have highest binding affinities to Cat D with -11.7 kcal/mol and -11.4 kcal/mol scores. Chemical structures of the five best hit drugs are demonstrated in Figure 1.

Research validation study including to re-dock of 2RZ ligand found in crystal structure of the protein have shown that 2RZ has quite affinity to Cat D with -9.7 kcal/mol binding affinity score. Furthermore, the interactions between 2RZ and Cat D have been revealed in both retrieved crystal structure and re-docked form. As such, it's revealed that 2RZ interacts with Cat D's ASP 33, ASP 231, and THR 234 residues through conventional hydrogen bonds, ALA 13, SER 80, THR 125, TYR 205, GLY 233, ASP 310, ACT 401 residues through van der Waals interactions, carbon-hydrogen and pi-donor bonds, VAL 31, TYR 78, GLY 79, ALA 129, MET 309, and ILE 311 residues through pi-pi stacked, pi-pi T-shaped, amide-pi stacked, alkyl, and pi-alkyl interactions in crystal structure of retrieved protein. In addition, it might interact with Cat D's ASP 33, GLY 79, ASP 231, THR 234, SER 235 residues through conventional hydrogen bonds and attractive charge interactions, ALA 13 and GLY 233 residues through carbon-hydrogen bond, VAL 31, ALA129, PHE 131, TYR 205, and ILE 311 residues through pi-pi stacked, pi-

pi T-shaped, alkyl, pi-alkyl interactions in the re-docked form. The results demonstrating common interacting amino acids and interaction styles of the ligand in crystal structure and re-docked form prove the adequacy of developed molecular docking strategy (Figure 2).

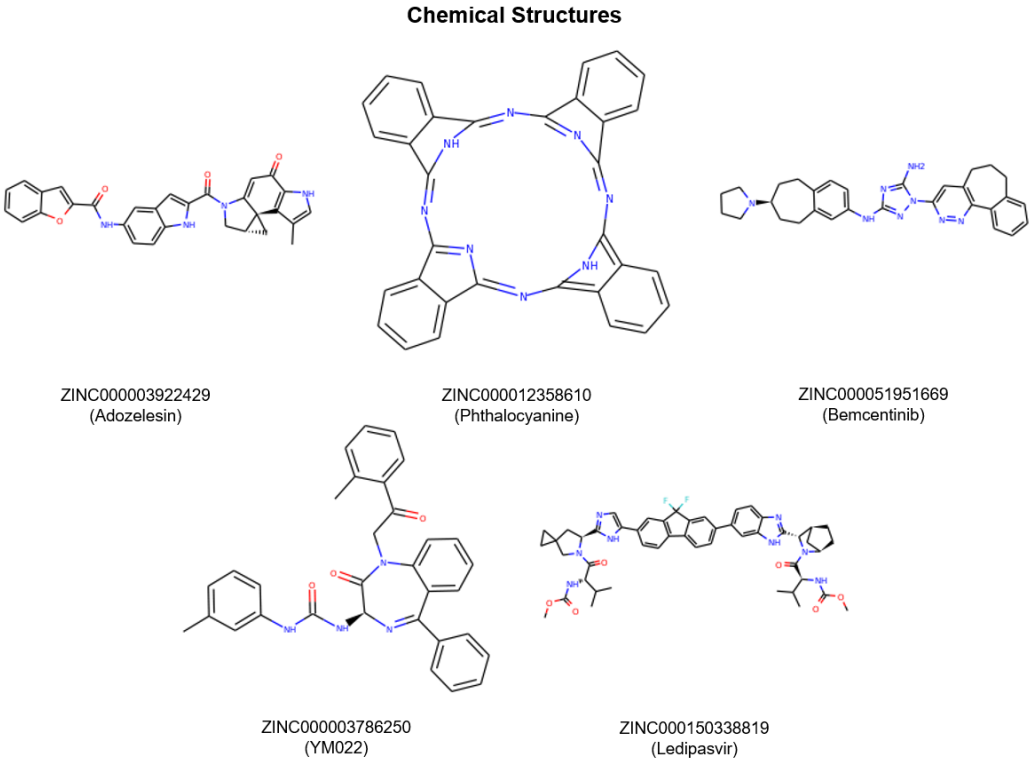

Figure 1. Chemical structures of five hit drugs having the highest binding affinity to Cat D.

Table 1. Best 10 hit drugs' binding affinity scores, datasets, and the interacted amino acid residues of Cat D protein.

10 Hit Ligands			
Ligand Name	**Score (kcal/mol)**	**Dataset**	**Receptor Residues Interacting with Ligands**
ZINC000003922429	-11.7	Drugs in Clinical Trials	ALA 13, VAL 31, ILE 76, PHE 131, ASP 231, THR 234
ZINC000012358610	-11.7	Drugs in Clinical Trials	ALA 13, ILE 134, ASP 231, MET 307, MET 309
ZINC000051951669	-11.7	Drugs in Clinical Trials	ASP 33, TYR 78, SER 80, PHE 131, ASP 231, THR 234, VAL 238, MET 307, MET 309
ZINC000003786250	-11.4	Drugs in Clinical Trials	GLY 35, TYR 78, GLY 79, SER 80, ILE 142, GLY 233, VAL 238, ILE 311, ILE 320
ZINC000150338819	-11.4	FDA-Approved Drugs	ALA 13, TYR 78, GLY 79, SER 80, ALA 129, TYR 205, GLY 233, MET 307, MET 309
ZINC000027990463	-11.3	FDA-Approved Drugs	ALA 13, GLN 14, TYR 15, VAL 31, GLY 79, ALA 129, ILE 134, TYR 205, ILE 229, ASP 231, THR 232, MET 307, ILE 311, LEU 318, ILE 320, ASP 323
ZINC000052955754	-11.3	FDA-Approved Drugs	VAL 31, GLY 79, SER 80, ILE 142

ZINC000053073961	-11.3	World-not-FDA Approved Drugs	VAL 31, ASP 33, TYR 78, SER 80, PHE 131, ILE 134, TYR 205, ILE 229, ASP 231, VAL 238, MET 307, ILE 311, LEU 318, ILE 320
ZINC000141846530	-11.3	Drugs in Clinical Trials	SER 36, ILE 142, VAL 144, ASP 231, LEU 236, MET 307
ZINC000096928979	-11.2	Drugs in Clinical Trials	VAL 31, GLY 35, ASN 38, TYR 78, SER 80, ALA 129, ILE 134, TYR 205, ILE 320

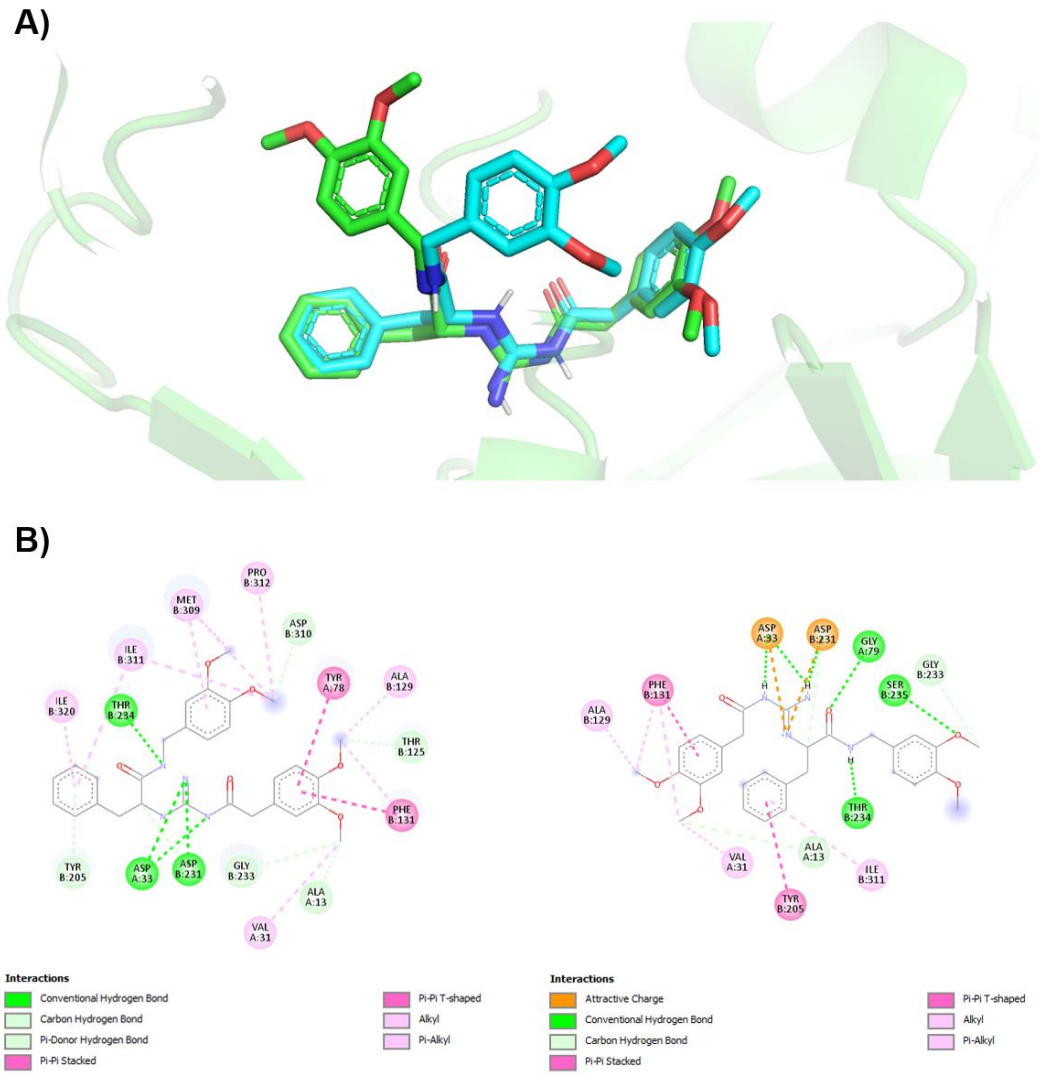

Figure 2. A) To superimpose of 2RZ in re-docked form (Cyan) with the ligand in crystal structure (Green), B) Interactions created between Cat D and 2RZ in crystal structure and re-docked form, respectively.

The role of Cat D in various diseases is based on its interactions with other cellular proteins. For instance, Kim and his colleagues have demonstrated that interactions between Cat D and Transglutaminase 2 (Tgase 2) cause to depletion of Cat D and reduce apoptosis level [34]. Blocking this interactions via inhibitory ligands promises to stop the role of Cat D in reduced apoptosis based diseases such as cancer. In order to analyze the possible inhibitory effect of novel hit drugs, Cat D-Tgase 2 complex has been created by HADDOCK protein-protein docking tool. While totally 45 complexes within 8 clusters have been created, the interactions between the proteins of the best cluster

with -30.4 docking and -1.6 Z scores have been analyzed with PyMol software (Figure 3-A). Findings have put forward that PRO 2, ILE 3, ASP 12, GLN 14, ASP 33, SER 36, ASP 50, SER 80, SER 82 residues of Cat D interact with THR 16, ARG 19, GLU 29, ASP 97, HIS 134, GLN 599, LYS 600, ARG 601, ALA 687 residues of Tgase 2. In addition, due to the fact that any ligand's Cat D inhibitory activity had not been proved with sufficient data to date, except Pestatin A which is frequently used as model ligand in the Cat D inhibition studies [23], a novel inhibitor library composed of 8 ligands whose possible activity had been demonstrated in literature was created and analyzed by promoting the same protocol (Table 2). Among these possible inhibitors N'-(3,4-dimethylacridin-9-yl)-3-methylbenzohydrazide, N'-(3,4-dimethylacridin-9-yl)-2-hydroxybenzohydrazide, and 1-(4-Phenylphenyl)-2-[2-(2-pyridyl)benzimidazol-1-yl]ethanone have exhibited highest binding affinities with -9.5 kcal/mol, -9.5 kcal/mol, and -9.4 kcal/mol affinity scores, respectively. In addition, the affinity score of Pepstatin A to Cat D has been computed as -8.1 kcal/mol. Ligand-receptor interaction analysis has revealed that; i) N'-(3,4-dimethylacridin-9-yl)-3-methylbenzohydrazide interacts with Cat D's THR 234 residue through conventional hydrogen bonds, GLY 79 residue through van der Waals interactions, ASP 33 and ASP 231 residues through salt bridge and attractive charge interactions, TYR 78 residue through amide-pi stacked interaction, and ILE 134 residue through pi-alkyl interaction, ii) N'-(3,4-dimethylacridin-9-yl)-2-hydroxybenzohydrazide interacts with Cat D's THR 234 residue through conventional hydrogen bonds, GLY 79 residue through van der Waals interactions, ASP 33 and ASP 231 residues through salt bridge and attractive charge interactions, and TYR 78 residue through amide-pi stacked interaction, iii) 1-(4-Phenylphenyl)-2-[2-(2-pyridyl)benzimidazol-1-yl]ethanone interacts with Cat D's THR 234 residue through carbon hydrogen bonds, ASP 33 and GLY 79 residues through van der Waals interactions, TYR 78 and GLY 233 residues through pi-pi stacked and amide-pi stacked interactions, ILE 142 and VAL 238 residues through pi-alkyl interactions and MET 307 residue through pi-sulfur interaction (Figure 3-B). Considering the interactions between Cat D and the ligands from inhibitor library, common amino acid residues of Cat D have been discovered as ALA 13, ASP 33, TYR 78, GLY 79, SER 80, PHE 131, ASP 231, GLY 233, THR 234, SER 235, ILE 311, and ASP 323. Moreover while GLN 14, ASP 33, and SER 80 have been described as significant in the interactions with Tgase 2, they are also found within common interacting amino acid residues. The frequencies of these amino acids are shown in Figure 3-C.

In order to reveal possible drug molecules targeting Cat D, virtual drug screening have been carried out and consequently several hit drug molecules having higher binding affinity then Pepstatin A, 2RZ, and other ligands with Cat D inhibitory activities were discovered. In particular, four ligands from Drugs in Clinical Trials dataset and one ligand from FDA-Approved Drugs dataset which are ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), ZINC000003786250 (YM022), and ZINC000150338819 (Ledipasvir) have been assessed as promising now that they have highest affinity scores as -11.7 kcal/mol and -11.4 kcal/mol. It's known that Adozelesin is an alkylating agent and it may interact with minor grooves of DNA molecules, sequence specifically [42]. Through this mechanism, Adozelesin is able to inhibit DNA replication and induces apoptosis. In addition, it's antitumor activity against distinct cancer types have been proved over years since that it decelerates cell proliferation by blocking them in G2 phase [43]. Despite the antitumor activity of Adozelesin, any evidence pointing its possible interactions with Cat D had not been reported. The ligand-protein interaction analysis has demonstrated that Adozelesin could interact with Cat D's ALA 13, ASP 231, and THR 234 residues through conventional hydrogen bonds and carbon hydrogen bonds, VAL 31 and ILE 76 residues through pi-alkyl interactions, and PHE 131 residue through pi-pi stacked interactions. Phthalocyanine is an organic compound and frequently used in chemical dye synthesis and photoelectricity researches. Due to its proper aromatic rings created by nitrogen atoms, it's frequently used to stain several biomolecules and it has quite absorption and emission properties [44]. Furthermore, novel approaches are using Phthalocyanine in the

phototherapy treatment of tumors [45]. In addition, while there is no evidence pointing the interactions of Phthalocyanine with Cat D in literature, ligand-protein interaction analysis has shown that Phthalocyanine might interact with Cat D's ALA 13, ILE 134, MET 307, and MET 309 residues through pi-sulfur and pi-alkyl interactions, ASP 231 residue through pi-anion interactions. Bemcentinib is another hit ligand with highest affinity score and its AXL kinase inhibition activity have been declared with various researches for many years [46]. Due to this inhibitory activity, the anticancer property of Bemcentinib over several cancers such as non-small-cell lung cancer, triple negative breast cancer, acute myeloid leukemia, and melanoma has been shown and it has reached to Phase II clinical trials to be used [47], [48]. Although any data pointing the Cat D inhibitory activity of Bemcentinib has not been observed, ligand-protein interaction analysis has shown that it could interact with Cat D's ASP 33 residue through conventional hydrogen bonds, SER 80 and THR 234 residues through carbon hydrogen bonds, pi-donor hydrogen bonds, and pi-lone pair interactions, ASP 231 residue through pi-anion interaction, TYR 78 and PHE 131 residues through pi-pi stacked and pi-pi T-shaped interactions, ASP 231 residue through pi-anion interaction, MET 307 and MET 309 residues through pi-sulfur interactions, and VAL 238 residue through pi-alkyl interaction. YM022 is a recently developed gastrin/CCKB antagonist ligand and inhibits the secretion of gastric acid via gastrin induced pathway [49, 50]. It's revealed that YM022 might interact with Cat D's GLY 35, GLY 79, SER 80 residues through conventional hydrogen bonds, TYR 78 and GLY 233 residues through carbon hydrogen bonds, and ILE 142, VAL 238, ILE 311 and ILE 320 residues through alkyl and pi-alkyl interactions. Any data indicating the Cat D inhibitory activity of YM022 is not exist in literature. Finally, Ledipasvir is an FDA-approved drug and frequently used in the treatment of hepatitis C virus (HCV) infection [51]. While Ledipasvir inhibits the viral replication [52], to date any findings about the interactions of it with Cat D have not been reported. Protein ligand interaction analysis has revealed that Ledipasvir might interact with Cat D's TYR 78, TYR 205 and GLY 233 residues through conventional hydrogen bonds and pi-sigma interactions, ALA 13 and SER 80 residues through carbon hydrogen bonds, GLY 79 residue through halogen interaction, ALA 129, MET 307 and MET 309 residues through alkyl and pi-alkyl interactions. Therefore, it's also revealed that discovered hit drugs might interact with the amino acid residues that have been described as common and the others adjacent them.

Table 2. The results of ligands having Cat D inhibitory activity and re-docked 2RZ.

Inhibitors and Re-docked 2RZ			
Ligand Name	**Binding Affinity (kcal/mol)**	**Receptor Residues Interacting with Ligand**	**Ref.**
2RZ	-9.7	ALA 13, VAL 31, ASP 33, GLY 79, ALA 129, PHE 131, TYR 205, ASP 231, GLY 233, THR 234, SER 235, ILE 311	[35]
N'-(3,4-dimethylacridin-9-yl)-3-methylbenzohydrazide	-9.5	ASP 33, TYR 78, GLY 79, ILE 134, ASP 231, THR 234	[36]
N'-(3,4-dimethylacridin-9-yl)-2-hydroxybenzohydrazide	-9.5	ASP 33, TYR 78, GLY 79, ASP 231, THR 234	[36]
1-(4-phenylphenyl)-2-(2-pyridin-2-ylbenzimidazol-1-yl)ethanone	-9.4	ASP 33, TYR 78, GLY 79, ILE 142, GLY 233, THR 234, VAL 238, MET 307	[37]
3-hydroxy-4-[(1-hydroxynaphthalen-2-yl)diazenyl]naphthalene-1-sulfonic acid	-8.7	TYR 78, PHE 131, ASP 231, THR 234, ILE 311, ILE 320	[38]
Antipain Dihydrochloride	-8.2	ASP 33, GLY 79, SER 80, TYR 205, THR 232, GLY 233, THR 234, SER 235, ILE 311, ASP 323	[39]
Pepstatin A	-8.1	ALA 13, GLN 14, TYR 15, GLY 35, TYR 78, SER 80, PHE 131, ILE 134, ILE 142, SER 235, ASP 323	[23]
2-[(2-amino-1-cyano-2-oxoethyl)diazenyl]benzoic acid	-6.8	ASP 33, SER 80, GLY 233, LEU 236	[40]
(2S)-2-amino-3-(4-azidophenyl)propanoic acid	-6.1	ALA 13, TYR 15, VAL 31, ASP 33, PHE 131, ASP 323	[38]
4-[(Z)-3-amino-2-methylbut-1-enyl]-2-bromo-6-methoxyphenol	-5.6	ASP 33, TYR 78, SER 80, ASP 231, GLY 233	[41]

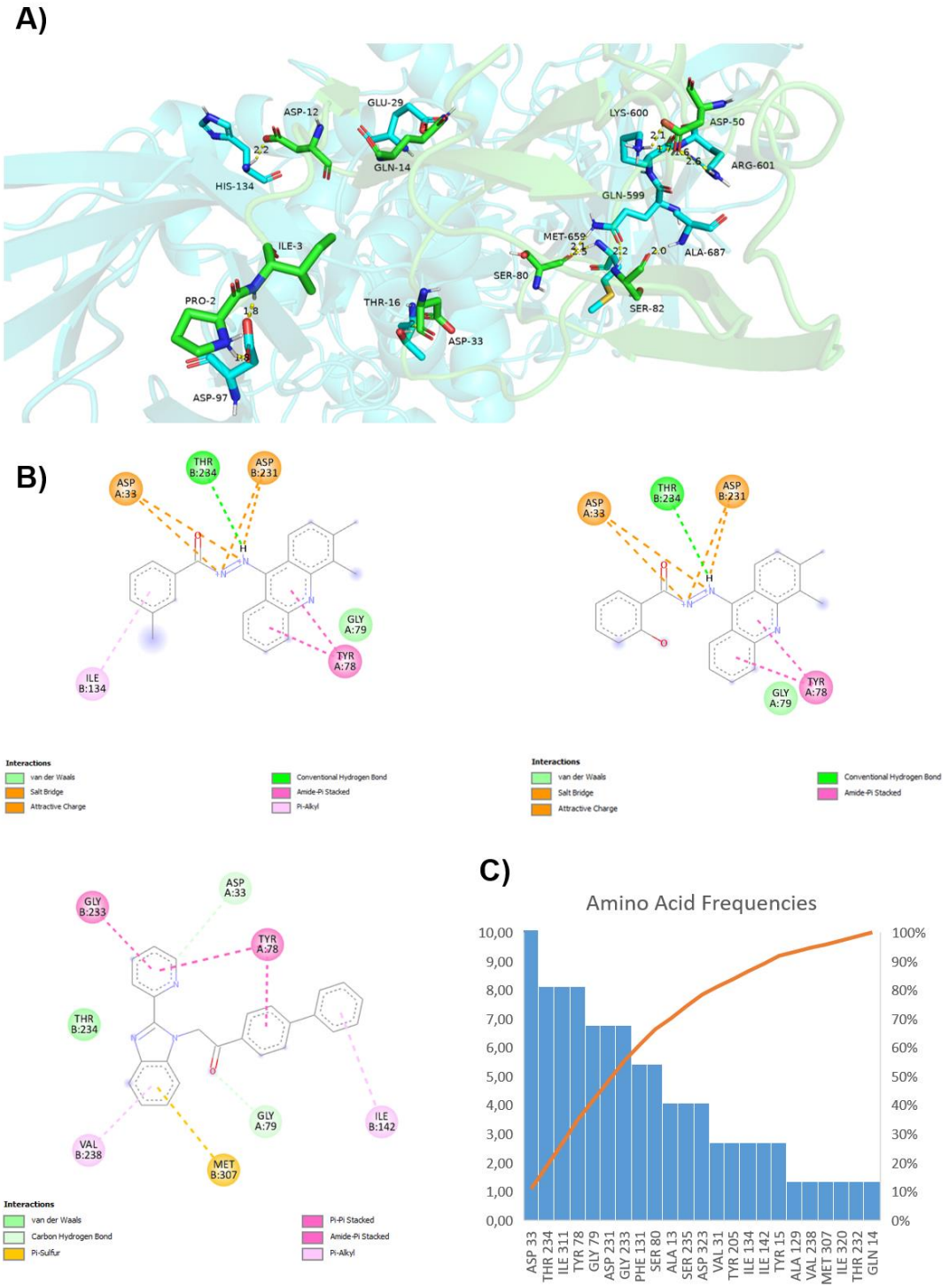

Figure 3. A) HADDOCK results demonstrating the interactions in the interface of Cat D-Tgase 2 complex, B) The interactions between Cat D and three inhibitors with the highest binding affinity which are N'-(3,4-dimethylacridin-9-yl)-3-methylbenzohydrazide, N'-(3,4-dimethylacridin-9-yl)-2-hydroxybenzohydrazide, and 1-(4-Phenylphenyl)-2-[2-(2-pyridyl)benzimidazol-1-yl]ethanone, respectively, C) The frequencies of the common interacting amino acids of Cat D with the ligands in inhibitor library.

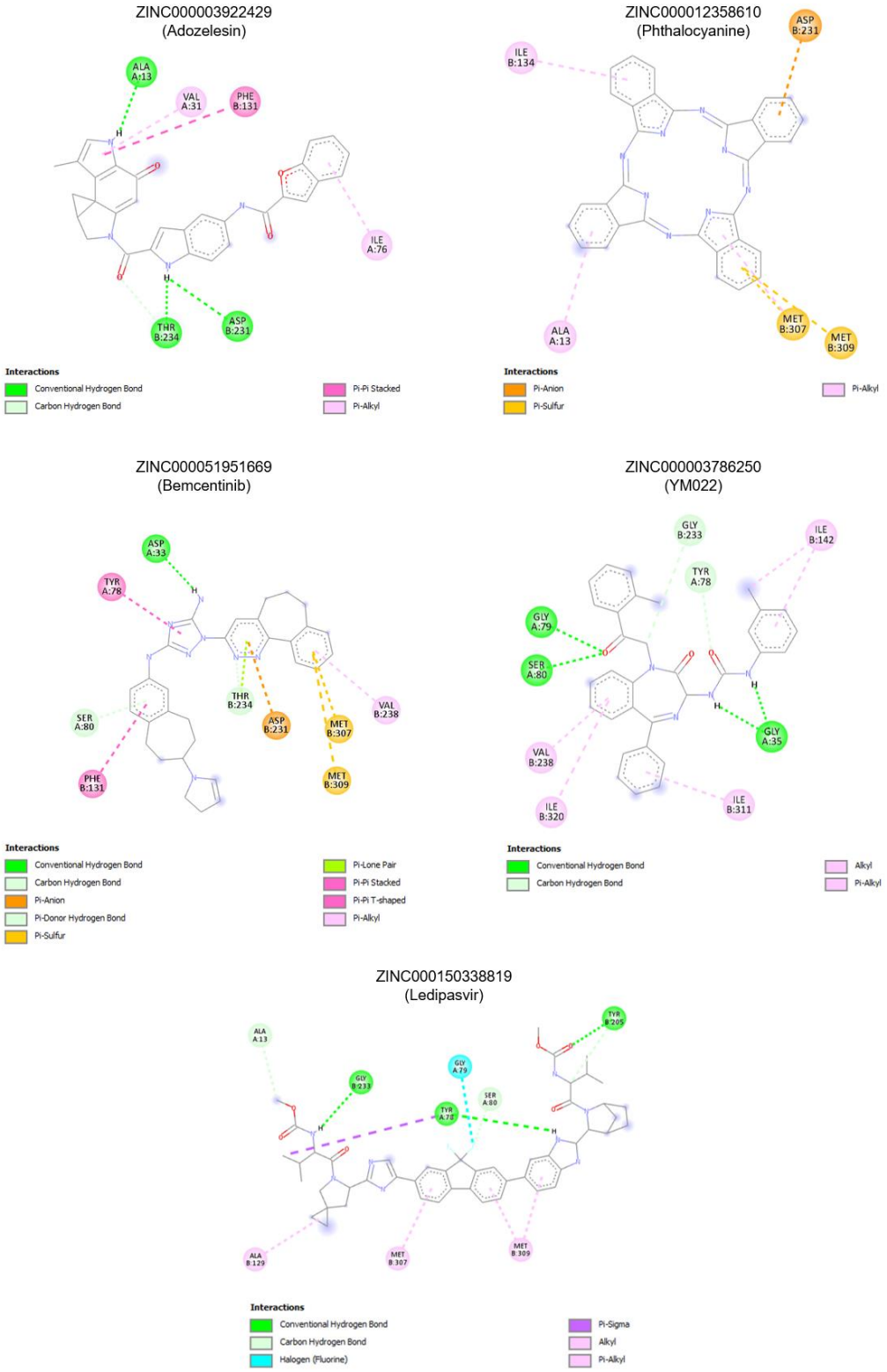

Figure 4. The interactions between Cat D and five hit drugs having the highest binding affinities that are ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), ZINC000003786250 (YM022), and ZINC000150338819 (Ledipasvir), respectively.

Table 3. Toxicity and ADME analysis of three inhibitors with the highest affinity, ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), and ZINC000003786250 (YM022) ligands from Drugs in Clinical Trials dataset.

ADME Properties and Toxicity Profiles								
Properties		**Best Scored Inhibitors**			**Best Scored Ligands**			
Physico-chemical properties	**Ligand Name**	N'-(3,4-dimethylacridin-9-yl)-3-methylbenzohydrazide	N'-(3,4-dimethylacridin-9-yl)-2-hydroxybenzohydrazide	1-(4-Phenylphenyl)-2-[2-(2-pyridyl)benzimidazol-1-yl]ethanone	ZINC000003922429 (Adozelesin)	ZINC000012358610	ZINC000051951669	ZINC000003786250 (YM022)
	Formula	C23H21N3O	C22H19N3O2	C26H19N3O	C30H22N4O4	C32H22N8	C30H34N8	C32H28N4O3
	Molecular Weight (g/mol)	355.43	357.41	389.45	502.52	518.57	506.64	516.59
	Molar Refractivity	111.16	108.22	119.38	145.58	173.55	155.1	159.61
	TPSA (topological polar surface area)	54.02 Å²	74.25 Å²	47.78 Å²	111.20 Å²	111.94 Å²	97.78 Å²	90.87 Å²
Lipophilicity	**Log $P_{o/w}$ (iLOGP)**	3.28	2.67	3.00	2.85	2.78	4.51	3.76
	Log $P_{o/w}$ (XLOGP3)	5.77	5.60	5.20	4.45	7.99	5.55	5.57
	Log $P_{o/w}$ (WLOGP)	4.88	4.28	5.65	4.73	5.79	4.51	4.57
	Log $P_{o/w}$ (MLOGP)	4.23	3.46	3.57	1.97	5.64	4.48	3.99
	Log $P_{o/w}$ (SILICOS-IT)	4.87	3.87	5.28	5.08	7.12	3.58	5.39
	Consensus Log *P*o/w	4.61	3.97	4.54	3.82	5.87	4.53	4.65
Solubility	**Log *S* (SILICOS-IT)**	-9.06	-8.10	-9.55	-9.24	-13.48	-8.78	-10.79
	SILICOS-IT Solubility (mg/ml)	3.07e-07	2.84e-06	1.10e-07	2.90e-07	1.70e-11	8.35e-07	8.40e-09
	SILICOS-IT Solubility (mol/l)	8.63e-10	7.94e-09	2.83e-10	5.77e-10	3.27e-14	1.65e-09	1.63e-11
	Solubility Class	Poorly Soluble	Poorly Soluble	Poorly Soluble	Poorly soluble	Insoluble	Poorly soluble	Insoluble
Druglikeness	**Druglikeness**	0.07	1.67	2.71	5.05	1.17	3.12	6.33
	Drug-score	0.1	0.16	0.44	0.21	0.25	0.3	0.28
Pharmacokinetics	**GI absorption**	High	High	High	High	Low	High	High
	BBB permeant	Yes	Yes	Yes	No	No	No	No
	P-gp substrate	No	No	Yes	Yes	No	Yes	No
	CYP1A2 inhibitor	Yes	Yes	Yes	Yes	No	Yes	No
	CYP2C19 inhibitor	Yes	Yes	Yes	Yes	No	Yes	Yes
	CYP2C9 inhibitor	Yes	Yes	Yes	Yes	No	Yes	Yes
	CYP2D6 inhibitor	Yes	Yes	No	No	No	Yes	No
	CYP3A4 inhibitor	Yes	Yes	Yes	Yes	No	Yes	Yes
Toxicity	**Mutagenicity**	Yes	Yes	No	Yes	No	No	No
	Tumorigenicity	Yes	Yes	No	No	No	No	No
	Irritant Effects	No	No	No	No	No	No	No
	Reproductive Effects	No	No	No	No	No	No	No

As such, the related hit drugs might inhibit the activity of Cat D, block its interactions with Tgase 2, and induce apoptosis. The ligand-protein interaction analysis of best five hit ligands are demonstrated in Figure 4.

Toxicity and ADME properties of ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), and ZINC000003786250 (YM022), as well as the three inhibitors with highest affinity scores have been

analyzed with swissADME server and OSIRIS Property Explorer tool (Table 3). Since that ZINC000150338819 (Ledipasvir) had been approved by FDA previously, there was no require to analyze its ADME and toxicity properties. Results have revealed that Adozelesin is poorly soluble, has no inhibitory activity over CYP2D6, high gastrointestinal (GI) absorption performance, and no toxicity except low mutagenicity. Phthalocyanine is insoluble, has no CYP isoforms inhibitory activity, low GI absorption, and no toxicity. Bemcentinib is poorly soluble, has high GI absorption, and no toxity. YM022 is insoluble, has high GI absorption, no CYP isoform inhibition potential except CYP2C19 and CYP3A4, and no toxicity. Comparing ADME and toxicity properties of the hit drugs with inhibitors it's revealed that that while various properties exhibit similarities, drug scores of hit drugs are higher than the inhibitors. Moreover, hit drugs are more proper in the Lipinski's rule of five parameters comparing the inhibitors. Considering the analysis, it's evaluated that the best hit drugs discovered with recently developed virtual drug screening strategy have quite potentials in order to be used in the blocking of Cat D activations. In particular, YM022 among the hit drugs is the most promising one due to its high drug score (0.28), convenient ADME properties, and toxicity profiles.

Bioactivity analysis of discovered hit drugs as well as the ligands whose Cat D inhibitory activity had been shown previously was performed in line with two parameters which are protease inhibitory activity and enzyme inhibitory activity due to that Cat D is a protease enzyme (Table 4). According to Molinspiration Cheminformatics v2020 online tool's computational analysis, ligands with the score within 0.00 – 0.5 possess high bioactivity, ligands with the score within -0.5 – 0.00 possess moderate bioactivity, and those with the score less than -0.5 are inactive [53]. Comparison the bioactivities of the drugs with the ligands from literature reveals that recently discovered drugs have higher potential in the inhibition activity for Cat D. Among all these molecules, YM022 has the highest enzyme inhibitory, and highest protease inhibitory activities. These findings prove that YM022 possess quite potential to be used in the Cat D activation based diseases.

Table 3. Bioactivity analysis of three inhibitors with the highest affinity, and five best hit drugs revealed from the study.

Bioactivity Analysis		
Ligand	**Protease Inhibitor**	**Enzyme Inhibitor**
ZINC000003922429 (Adozelesin)	-0.01	-0.03
ZINC000012358610 (Phthalocyanine)	-0.07	-0.04
ZINC000051951669 (Bemcentinib)	-0.20	0.19
ZINC000003786250 (YM022)	0.19	0.21
ZINC000150338819 (Ledipasvir)	-2.29	-3.29
N'-(3,4-dimethylacridin-9-yl)-3-methylbenzohydrazide	-0.19	-0.09
N'-(3,4-dimethylacridin-9-yl)-2-hydroxybenzohydrazide	-0.19	-0.04
1-(4-Phenylphenyl)-2-[2-(2-pyridyl)benzimidazol-1-yl]ethanone	-0.03	0.13

4. Conclusions

Cat D is an aspartic acid protease having significant functions within various physiological processes such as intracellular and extracellular protein degradation, regulation of programmed cell death, hormone and peptide processing, antigen presentation etc. In addition, the roles of Cat D in several pathologies such as osteoarthiritis, neurodegenerative diseases (Alzheimer, Neuronal Ceroid Lipofuscinosis (NCL), Niemann Pick type C (NPC), etc.), acute pancreatitis have been reported with many papers. However, the role of Cat D in breast cancer prognosis have taken the much attention of the studies. Due to the facts that Cat D is overexpressed and causes metastasis, invasion, releasing of growth factors, and angiogenesis through to digest ECM proteins, it's one of the most significant drug target in the treatment of breast cancer. In the meantime, to date any chemical ligand that can target Cat D has not achieve to clinical trial phases, and only Pepstatin A as well as few ligands' Cat D inhibitory activities have been demonstrated with the few data. As such, a virtual drug screening including to analyze the binding affinities of totally 12,111 drugs composed of FDA-Approved Drugs, World-not-FDA Approved Drugs, Drugs in Clinical Trials, and Non-human Metabolites datasets of ZINC15 database have been carried out with this study. While validation study including to re-dock 2RZ

ligand has proven the sufficiency of docking strategy, common interacting amino acid residues of the protein have been revealed via to dock the ligands whose Cat D inhibitory activities are shown previously and to create protein complex consisting with Cat D and its interacting protein Tgase 2. Consequently, results have put forward that totally five drug molecules might be repurposed to inhibit Cat D activation. While these drugs are ZINC000003922429 (Adozelesin), ZINC000012358610 (Phthalocyanine), ZINC000051951669 (Bemcentinib), ZINC000003786250 (YM022), and ZINC000150338819 (Ledipasvir), gastrin/CCKB antagonist YM022 is assessed as most promising one considering the high binding affinity, proper ADME and toxicity profiles, and high bioactivity performances such as protease and enzyme inhibitory activities. Thus, YM022 has quite potential to be used in order to inhibit metastasis, angiogenesis, invasion, and drug resistance of breast cancer cells. Nonetheless, the stability of YM022 in the binding to Cat D should be analyzed through molecular dynamics (MD) simulations, and its potential activities should be tested both in vitro and in vivo studies.

References

[1] M. Leary, S. Heerboth, K. Lapinska, and S. Sarkar, Sensitization of drug resistant cancer cells: A matter of combination therapy, Cancers 10 (2018) 1–18.

[2] T. Oskarsson, Extracellular matrix components in breast cancer progression and metastasis, Breast 22 (2013) 66–S72.

[3] B. Yue, Biology of the extracellular matrix: An overview, J. Glaucoma 23 (2014) 20–23.

[4] E. Di Cera, Serine proteases, IUBMB Life 61 (2009) 510–515.

[5] S. Verma, R. Dixit, and K. C. Pandey, Cysteine proteases: Modes of activation and future prospects as pharmacological targets, Front. Pharmacol. 7 (2016) 1–12.

[6] J. Tang, R. N. S. Wong, Evolution in the structure and function of aspartic proteases, J. Cell. Biochem. 33 (1987) 53–63.

[7] P. Benes, V. Vetvicka, and M. Fusek, Cathepsin D-Many functions of one aspartic protease, Crit. Rev. Oncol. Hematol. 68 (2008) 12–28.

[8] E. T. Baldwin et al., Crystal structures of native and inhibited forms of human cathepsin D: Implications for lysosomal targeting and drug design, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 6796–6800.

[9] T. Houben et al., Cathepsin D regulates lipid metabolism in murine steatohepatitis, Sci. Rep. 7 (2017) 1–10.

[10] P. Gan et al., Knockdown of cathepsin D protects dopaminergic neurons against neuroinflammation-mediated neurotoxicity through inhibition of NF-κB signalling pathway in Parkinson's disease model, Clin. Exp. Pharmacol. Physiol. 46 (2019) 337–349.

[11] J. Liu, L. Yang, H. Tian, and Q. Ma, Cathepsin D is involved in the oxygen and glucose deprivation/reperfusion-induced apoptosis of astrocytes, Int. J. Mol. Med. 38 (2016) 1257–1263.

[12] A. Eguchi, A. E. Feldstein, Lysosomal Cathepsin D contributes to cell death during adipocyte hypertrophy, Adipocyte 2 (2013) 170–175.

[13] N. Zaidi, A. Maurer, S. Nieke, and H. Kalbacher, Cathepsin D: A cellular roadmap, Biochem. Biophys. Res. Commun. 376 (2008) 5–9.

[14] C. E. Chwieralski, T. Welte, and F. Bühling, Cathepsin-regulated apoptosis, Apoptosis 11 (2006) 143–149.

[15] S. A. Abideen, M. Khan, M. Irfan, and S. Ahmad, Deciphering the dynamics of cathepsin D as a potential drug target to enhance anticancer drug-induced apoptosis, J. Mol. Liq. 361 (2022) 119677.

[16] A. A. Aghdassi et al., Cathepsin d regulates cathepsin b activation and disease severity predominantly in inflammatory cells during experimental pancreatitis, J. Biol. Chem. 293 (2018) 1018–1029.

[17] A. Amritraj, Y. Wang, T. J. Revett, D. Vergote, D. Westaway, and S. Kar, Role of Cathepsin d in u18666a-induced neuronal cell death potential implication in Niemann-Pick type c disease pathogenesis, J. Biol. Chem. 288 (2013) 3136–3152.

[18] E. Liaudet-Coopman et al., Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis, Cancer Lett. 237 (2006) 167–179.

[19] C. Zhang, M. Zhang, and S. Song, Cathepsin D enhances breast cancer invasion and metastasis through promoting hepsin ubiquitin-proteasome degradation, Cancer Lett. 438 (2018) 105–115.

[20] L. B. Alcaraz et al., A 9-kDa matricellular SPARC fragment released by cathepsin D exhibits pro-tumor activity in the triple-negative breast cancer microenvironment, Theranostics 11 (2021) 6173–6192.

[21] H. S. Anantaraju, M. B. Battu, S. Viswanadha, D. Sriram, and P. Yogeeswari, Cathepsin D inhibitors as potential therapeutics for breast cancer treatment: Molecular docking and bioevaluation against triple-negative and triple-positive breast cancers, Mol. Divers. 20 (2016) 521–535.

[22] D. E. Abbott et al., Reevaluating cathepsin D as a biomarker for breast cancer: Serum activity levels versus histopathology, Cancer Biol. Ther. 9 (2010) 23–30.

[23] N. Bidère et al., Cathepsin D triggers Bax activation, resulting in selective apoptosis-inducing factor (AIF) relocation in T lymphocytes entering the early commitment phase to apoptosis, J. Biol. Chem. 278 (2003) 31401–31411.

[24] R. Houštecká et al., Biomimetic Macrocyclic Inhibitors of Human Cathepsin D: Structure-Activity Relationship and Binding Mode Analysis, J. Med. Chem. 63 (2020) 1576–1596.

[25] A. Gimeno et al., The light and dark sides of virtual screening: What is there to know?, Int. J. Mol. Sci. 20 (2019) 1375.

[26] E. F. Pettersen et al., UCSF Chimera - A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[27] S. Dallakyan, A. Olson, Small-Molecule Library Screening by Docking with PyRx, NY: Springer New York, U.S.A, 2015, 243-250.

[28] O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, J. Comput. Chem. 17 (2011) 295-304.

[29] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin, HADDOCK: A protein-protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.

[30] A. Porollo, J. Meller, Prediction-Based Fingerprints of Protein-Protein Interactions, Proteins 66 (2007) 630-645.

[31] https://www.organic-chemistry.org/prog/peo/, January 2017, Accessed: 06.12.2022.

[32] A. Daina, O. Michielin, and V. Zoete, SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 1–13.

[33] https://www.molinspiration.com, January 1986, Accessed: 07.12.2022.

[34] S. J. Kim, K. H. Kim, E. R. Ahn, B. C. Yoo, and S. Y. Kim, Depletion of cathepsin D by transglutaminase 2 through protein cross-linking promotes cell survival, Amino Acids 44 (2013) 73–80.

[35] U. Grädler et al., Structure-based optimization of non-peptidic Cathepsin D inhibitors, Bioorganic Med. Chem. Lett. 24 (2014) 4141–4150.

[36] M. K. Azim, W. Ahmed, I. A. Khan, N. A. Rao, and K. M. Khan, Identification of acridinyl hydrazides as potent aspartic protease inhibitors, Bioorganic Med. Chem. Lett. 18 (2008) 3011–3015.

[37] Z. S. Saify et al., New benzimidazole derivatives as antiplasmodial agents and plasmepsin inhibitors: Synthesis and analysis of structure-activity relationships, Bioorganic Med. Chem. Lett. 22 (2012) 1282–1286.

[38] W. Ahmed, U. Jabeen, and S. Khaliq, New inhibitors of proteolytic enzymes Cathepsin D and Plasmepsin II, Pakistan J. Biochem. Mol. Biol. 47 (2014) 129–132.

[39] L. Gangoda et al., Inhibition of cathepsin proteases attenuates migration and sensitizes aggressive N-Myc amplified human neuroblastoma cells to doxorubicin, Oncotarget 6 (2015) 11175–11190.

[40] W. Ahmed, I. A. Khan, M. N. Arshad, W. A. Siddiqui, M. A. Haleem, and M. K. Azim, Identification of Sulfamoylbenzamide derivatives as selective cathepsin D inhibitors, Pak. J. Pharm. Sci. 26 (2013) 687–690.

[41] S. R. M. Ibrahim et al., Thiophenes—Naturally Occurring Plant Metabolites: Biological Activities and In Silico Evaluation of Their Potential as Cathepsin D Inhibitors, Plants 11 (2022) 1-64.

[42] P. R. Cao, M. M. McHugh, T. Melendy, and T. Beerman, The DNA minor groove-alkylating cyclopropylpyrroloindole drugs adozelesin and bizelesin induce different DNA damage response pathways in human colon carcinoma HCT116 cells, Mol. Cancer Ther. 2 (2003) 651–659.

[43] B. K. Bhuyan, K. S. Smith, E. G. Adams, T. L. Wallace, D. D. Von Hoff, and L. H. Li, Adozelesin, a potent new alkylating agent: cell-killing kinetics and cell-cycle effects, Cancer Chemother. Pharmacol. 30 (1992) 348–354.

[44] T. Furuyama, K. Satoh, T. Kushiya, and N. Kobayashi, Design, synthesis, and properties of phthalocyanine complexes with main-group elements showing main absorption and fluorescence beyond 1000 nm, J. Am. Chem. Soc. 136 (2014) 765–776.

[45] C. C. Rennie, R. M. Edkins, Targeted cancer phototherapy using phthalocyanine-anticancer drug conjugates, Dalt. Trans. 51 (2022) 13157–13175.

[46] A. Hoel et al., Axl-inhibitor bemcentinib alleviates mitochondrial dysfunction in the unilateral ureter obstruction murine model, J. Cell. Mol. Med. 25 (2021) 7407–7417.

[47] A. Garcia-Sampedro, G. Gaggia, A. Ney, I. Mahamed, and P. Acedo, The state-of-the-art of phase ii/iii clinical trials for targeted pancreatic cancer therapies, J. Clin. Med. 10 (2021) 1–45.

[48] D. Zdzalik-Bielecka, K. Kozik, A. Po'swiata, K. Jastrzębski, M. Jakubik, and M. Miaczyńska, Bemcentinib and Gilteritinib Inhibit Cell Growth and Impair the Endo-Lysosomal and Autophagy Systems in an AXL-Independent Manner, Mol. Cancer Res. 20 (2022) 446–455.

[49] H. Yuki et al., YM022, a potent and selective gastrin/CCK-B receptor antagonist, inhibits peptone meal-induced gastric acid secretion in Heidenhain pouch dogs, Dig. Dis. Sci. 42 (1997) 707–714.

[50] S. Attoub, L. Moizo, J. P. Laigneau, B. Alchepo, M. J. M. Lewin, and A. Bado, YM022, a highly potent and selective CCK(B) antagonist inhibiting gastric acid secretion in the rat, the cat and isolated rabbit glands, Fundam. Clin. Pharmacol. 12 (1998) 256–262.

[51] M. Charlton et al., Ledipasvir and Sofosbuvir Plus Ribavirin for Treatment of HCV Infection in Patients with Advanced Liver Disease, Gastroenterology 149 (2015) 649–659.

[52] C. C. Lo et al., Ledipasvir/sofosbuvir for HCV genotype 1, 2, 4–6 infection: Real-world evidence from a nationwide registry in Taiwan, J. Formos. Med. Assoc. 121 (2022) 1567–1578.

[53] M. Arshad, M. S. Khan, S. A. A. Nami, S. I. Ahmad, M. Kashif, and A. Anjum, Synthesis, characterization, biological, and molecular docking assessment of bioactive 1,3-thiazolidin-4-ones fused with 1-(pyrimidin-2-yl)-1H-imidazol-4-yl) moieties, J. Iran. Chem. Soc. 18 (2021) 1713–1727.