Combined Molecular and Machine Learning Models to Probe the Activation Potential of ACE2 in Pulmonary Arterial Hypertension (PAH)

By

Amna Tariq

(Registration No: 00000401565)

Department of Sciences

School of Interdisciplinary Engineering & Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(September,2024)

Combined Molecular and Machine Learning Models to Probe the Activation Potential of ACE2 in Pulmonary Arterial Hypertension (PAH)

By

Amna Tariq

(Registration No: 00000401565)

A thesis submitted to the National University of Sciences & Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in
Bioinformatics

Supervisor: Dr. Ishrat Jabeen

School of Interdisciplinary Engineering & Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(September,2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that the final copy of MS Thesis written by Miss. Amna Tariq (Registration No. 00000401565), of SINES (School/College/Institute) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/ Regulations/ Masters Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of Masters degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/ local evaluators of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor Dr. Ishrat Jabeen

Date: 12th Sept, 2024

For HOD Sci

Signature (HOD): _____

Date: 16/09/2024

Dr. Mian Ilyas Ahmad
HoD Engineering
Professor
SINES - NUST, Sector H-12
Islamabad

Signature (Dean/ Principal) _____

Dr. SYED IRTIZA ALI SHAH
Principal & Dean
SINES - NUST, Sector H-12
Islamabad

Date: 18 SEP 2024

AUTHOR'S DECLARATION

I hereby state that my MS thesis titled "**Combined Molecular and Machine Learning Models to Probe the Activation Potential of ACE2 in Pulmonary Arterial Hypertension (PAH)**" is my work and has not been submitted previously by me for taking any degree from the National University of Sciences and Technology, Islamabad, or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Amna Tariq

Date: _____

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in the thesis titled "**Combined Molecular and Machine Learning Models to Probe the Activation Potential of ACE2 in Pulmonary Arterial Hypertension (PAH)**" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and the National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred to/cited.

I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis even after the award of MS degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and NUST, Islamabad have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: _____

Name: Amna Tariq

DEDICATION

I dedicate this thesis to my exceptional parents, siblings, friends, and teachers whose unconditional love, support, and guidance led me to this world of accomplishment.

ACKNOWLEDGEMENTS

I begin by acknowledging the grace and guidance of **Allah Almighty**, whose unwavering support has been the bedrock of my academic pursuits. His blessings have illuminated my path, and I am deeply grateful for the strength and inspiration derived from His infinite mercy.

I extend my heartfelt appreciation to my supervisor, ***Dr. Ishrat Jabeen***, whose remarkable dedication, scholarly insight, and mentorship have been pivotal to the successful completion of this research. Dr. Ishrat's guidance has not only enriched the intellectual depth of my work but also shaped my growth as a researcher. I am indebted to her for the countless hours of constructive feedback, encouragement, and the invaluable lessons imparted during this journey. My sincere gratitude extends to ***Dr. Masood ur Rehman Kayani,*** Guidance and Examination Committee (GEC), and to ***Dr. Muhammad Tariq Saeed*** for their invaluable contributions. Their expertise, insightful feedback, and scholarly guidance have been instrumental in refining the quality and scope of this research.

To my family, the cornerstone of my support system, I offer my deepest appreciation. To my parents for their sacrifices, encouragement, and unwavering belief in my potential have fueled my determination. To my sister, Asma Tariq, your togetherness and understanding have made this journey more meaningful. Your collective support has been the driving force behind this accomplishment.

I extend my thanks to my colleagues, Minahil Tariq, Arisha Wasim and Muqadas Chaudhary whose encouragement provided a refreshing balance to the academic challenges. Their shared experiences and unwavering belief in my capabilities have been a source of strength.

Thank you all for being integral parts of this significant chapter in my academic life.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

LIST OF ABBREVIATIONS:

PH	Pulmonary Hypertension
AP	Arterial pressure
PAH	Pulmonary Arterial Hypertension
RAAS	Renin Angiotensin Aldosterone System
PVR	Pulmonary Vascular Resistance
JG	Juxtaglomerular
Ang	Angiotensin
ACE	Angiotensin converting enzyme
AT1R	Angiotensin type 1 receptor
AT2R	Angiotensin type 2 receptor
AT3R	Angiotensin type 3 receptor
AT4R	Angiotensin type 4 receptor
MasR	Mas G coupled protein receptors
PD	Peptidase Domain
ACEI	ACE Inhibitor

MCT	Monocrotaline-Induced
MAPKs	Mitogen-Activated Protein Kinases
IL	Interleukin
ARBs	Angiotensin Receptor Blockers
MR	Mineralocorticoid
IRAP	Insulin-Regulated Membrane Aminopeptidase
ROS	Reactive Oxygen Species
CVDs	Cardiovascular Diseases
PVR	Pulmonary Vascular Remodeling
DIZE	Diminazene aceturate
HIF-1α	Hypoxia Inducible Factor-1α
PASMCs	Pulmonary Artery Smooth Muscle Cells
ChEMBL	Chemical Database of European Molecular Biology Laboratory
GA	Genetic Algorithm
GOLD	Genetic Algorithm of Ligand Docking
PDB	Protein Data Bank

nM	Nanomolar
IC50	Half Maximal Inhibitory Concentration
MW	Molecular Weight
AMBER	Assisted Model Building with Energy Refinement
PLIF	Protein Ligand Interaction Fingerprints
MOE	Molecular Operating Environment
SMILES	Simplified Molecular Input Line Entry System
ML	Machine Learning
ANN	Artificial Neural Network
SVM	Support Vector Machine

ABSTRACT:

Pulmonary Arterial Hypertension (PAH) is a severe cardiovascular disorder, characterized by high blood pressure. Untreated PAH can lead to heart failure and disrupt lung functions. Many recent studies have suggested that an altered renin-angiotensin-aldosterone (RAAS) can be a causative factor in PAH pathogenesis. Therefore, an increased level of Angiotensin II (Ang II) has been associated with the development of PAH. Previously, various ACE inhibitors (ACEI) have been proposed as potential drug candidates to mitigate the detrimental effects of Ang II. ACE2, a recently discovered homolog of ACE, opposes the effect of Ang II by converting Ang II into Ang 1-7. Briefly, targeting the activation of ACE2 by ACE inhibitors may act as a counterbalance to the effect of Ang II. In this study, we employed molecular docking guided machine learning models to predict the binding potential of ACE inhibitors to activate ACE2 for PAH treatment. Predictive machine learning models were implemented on docked complexes of ACE inhibitors for the prediction of the activation potential of ACE2. Support Vector Machine (SVM) and Artificial Neural Network (ANN) models accurately classified ACE inhibitors and ACE2 activators with overall accuracies of 99.57% and 90.69%, respectively. Ligands with ChEMBL ids, CHEMBL273140 and CHEMBL10521, demonstrated the most effective dual functionality as both ACE inhibitor and ACE2 activator. Our findings aided in understanding the binding attributes of ACE2 activators at the molecular level, which can assist in developing novel pharmaceutical agents for the treatment of PAH.

Chapter 1 Introduction:

Pulmonary Hypertension (PH) is a hemodynamic condition that is followed by resting mean pulmonary artery pressure (PAP) of ≥25 mm Hg [1]. PH is a progressive disorder that results in premature death. There are majorly five categories of PH. Pulmonary Arterial Hypertension (PAH) is one of these five categories that causes pathological changes within pulmonary vasculature. The remaining four groups of PH are referred to as secondary groups as they are followed by other disorders [2]. Group 2 results from left heart diseases, group 3 results from lung disease or hypoxia, group 4 is Chronic Thromboembolic PH (CPEH), and group 5 is pulmonary hypertension of uncertain multifactorial mechanism as mentioned in **Table 1.1** [3].

Table 1.1: Classification of Pulmonary Hypertension (PH) by WHO [4].

Groups	Mechanism	Clinical Definition	Hemodynamic Definition	Examples
1	Vascular remodeling of pulmonary arteries	Precapillary	mPAP > 20 mmHg PAOP ≤ 15 mmHg PVR > 2 WU	Idiopathic, medications, HIV, connective tissue disorders
2	Left heart disease causing backup of blood flow	Postcapillary	mPAP > 20 mmHg PAOP > 15 mmHg	Left heart failure, aortic valve disease, mitral valve disease
		Isolated postcapillary	mPAP > 20 mmHg PAOP > 15 mmHg PVR ≤ 2 WU	
		Combined pre and postcapillary	mPAP > 20 mmHg PAOP > 15 mmHg PVR > 2 WU	
3	Chronic lung disease causing hypoxemia	Precapillary	mPAP > 20 mmHg PAOP ≤ 15 mmHg PVR > 2 WU	COPD, ILD, sleep apnea

4	CTEPH	Precapillary	mPAP > 20 mmHg PAOP ≤ 15 mmHg PVR ≥ 2 WU	PE
5	Unclear and multifactorial	Precapillary	mPAP > 20 mmHg PAOP ≤ 15 mmHg PVR > 2 WU	Sarcoidosis, chronic hemolytic anemia, thyroid disorders, sickle cell anemia, splenectomy, mediastinal tumors, chronic renal failure
		Postcapillary	mPAP > 20 mmHg PAOP > 15 mmHg	
		Isolated postcapillary	mPAP > 20 mmHg PAOP > 15 mmHg PVR ≤ 2 WU	
		Combined pre and postcapillary	mPAP > 20 mmHg PAOP > 15 mmHg PVR > 2 WU	

Because of limited treatment and low survival rate, PAH is the most important of all. Pulmonary Arterial Hypertension is characterized by high Pulmonary Arterial Pressure (PAP) that results in right ventricular dysfunction. Various forms of PAH itself share similar histopathological findings such as adventitial thickening, medial smooth muscle cells (SMC) enlargement, and intimal proliferation. The most prevalent subtypes of PAH are idiopathic and connective tissue disease-associated PAH in the Western World; PAH is being diagnosed in an aging population with comorbidities at an alarming rate [5]. The key characteristics of PAH set forth by World Health Organization (WHO) are pulmonary vascular remodeling in the small pulmonary arteries, increased PAP, and elevated pulmonary vascular resistance (PVR). Breathlessness, exhaustion, chest pain, abdomen fullness, lightheadedness, or syncope are the most frequent symptoms of PAH

that increase with the severity of the disease. These symptoms and the high mortality rate are due to decreased vascular compliance and an increase in PVR which lead to an increase in right ventricular (RV) afterload and progressive RV dysfunction. The worst outcomes in PAH are RV failure and morphological changes in RV [6]. With an incidence of 2.4 million per year, the prevalence of all PAHs is approximately 15-50 per million people. Although the outcomes of different types of PAH vary, incident patients have a 15% risk of dying within a year and a 30% risk of dying within 3 years [7]. The combination of different therapies is considered a potential treatment for PAH. Additionally, the early use of a combination of therapies at the time of diagnosis, especially in patients with severe disease represents the current paradigm of treatment.

1.1 PAH Pathogenesis:

Genetic mutation in BMPR2 and multiple pathways including potassium channels (KCNK3 and ABCC8), and transcription factors (TBX4 and SOX17) can be involved in PAH susceptibility and progression [8]. Elevated levels of cytokines, chemokines, and inflammatory mediators have been linked in PAH patients. Vascular and parenchymal cells, such as endothelial, smooth muscle, and fibroblasts, undergo phenotypic changes that alter their sensitivity to inflammatory triggers, resulting in active secretion of cytokines and chemokines [9]. Understanding the interactions between inflammatory mediators and vascular cells in PAH patients can be helpful for the development of therapy for PAH treatment. The level of EYA3 is also associated with PAH pathogenesis [10]. Renin Angiotensin Aldosterone System (RAAS) activation causes an increase in collagen deposition that can result in myocardial stiffness. This study predicts a direct linkage between RAAS activation and cardiovascular diseases [11]. RAAS activation promotes vasoconstriction, vascular remodeling, inflammation, and endothelial dysfunction which contribute to the pathogenesis of PAH.

1.2 Renin Angiotensin Aldosterone System (RAAS):

RAAS regulates extracellular volume, arterial blood pressure, and plasma sodium concentration. RAAS activation causes hypertension, cell proliferation, inflammation, and fibrosis, which affect every organ [12]. The alteration in RAAS is a causative factor in the development of PAH. The two main components of RAAS are renin and angiotensin as mentioned in **Figure 1.2**. Kidney granular cells secrete renin [13]. Prorenin, the precursor of renin, is a 406 amino acid protein that forms the active protein when processed [14]. Neuroendocrine convertase 1 (proprotein convertase 1) or cathepsin B can activate prorenin proteolytically in the kidney, while renin/prorenin receptor can non-proteolytically activate prorenin in various organs. Renin has 396 amino acids in its active form [15]. There are multiple isoforms of renin that have antagonizing functions [16]. Because of its signaling functions, this enzyme is also regarded as a hormone [17]. Its expression is caused by low sodium chloride, low arterial blood pressure, and sympathetic nervous system activity (beta-1 adrenoceptor activation). The process starts with renin, released in response to low blood pressure or blood volume. By acting on the bond between leucine (Leu) and valine (Val), it hydrolyzes the α-2-globulin protein angiotensinogen, which is produced by the liver (approximately 118aa, though the length can vary) to angiotensin I [18]. Angiotensinogen is a potential enzyme inhibitor as a member of the serpin family (SERPINA8) [19]. Corticosteroids, estrogen, thyroid hormone, and angiotensin II levels can all raise the plasma level of angiotensinogen [18]. Renin cleaves angiotensinogen, produced by the liver, and converts it into the inactive peptide angiotensin I. The endothelium-bound angiotensin-converting enzyme (ACE) further cleaves the decapeptide angiotensin I in the kidney epithelial cells, lung capillaries, and endothelial cells. Angiotensin I is converted into the peptide angiotensin II by this enzyme, a carboxypeptidase sometimes referred to as kininase II, peptidyl-dipeptidase A, or CD143 [20]. To

form angiotensin II, two of the C terminal amino acids of angiotensin I are removed. Angiotensin II due to its intracrine, autocrine and paracrine roles, affects almost all system [21]. Angiotensin II constricts smooth muscles and has a vasoactive effect on all blood vessels. It elevates prothrombotic potential by stimulating plasminogen activator inhibitor proteins PAI-1 and PAI-2 and boosting blood pressure and heart rate [22]. Moreover, it causes the adrenal gland cortex to release aldosterone [23]. By stimulating the kidney's proximal tubules to promote sodium reabsorption and so retain sodium while losing potassium, aldosterone preserves the sodium-potassium balance [24]. Hypothalamus, an important component of the central nervous system is stimulated by renin to trigger the thirst reflex (dipsogen) [25]. Osmoreceptors on the hypothalamus detect thirst feeling [26]. Urinary loss is thus decreased by the release of antidiuretic hormone (ADH)/vasopressin, a nonapeptide from the posterior pituitary gland [27]. Antidiuretic hormone (ADH) regulates the secretion of adrenocorticotropic hormone (ACTH) by acting on vasopressin receptors in the anterior pituitary. The cortisol production from the adrenal gland is regulated by the hormone ACTH [28]. Angiotensin II (Ang II) causes the release of prostaglandins which further results in renal vasoconstriction as shown in **Figure1.1**. PAH has been linked with the increased level of Ang II. Most of the functions of RAAS are facilitated by Ang II which activates angiotensin II type 1 receptor (AT_1 receptor), leading to vasoconstriction (narrowing of blood vessels), aldosterone release, and other functions that elevate blood pressure and induce hypertrophy.

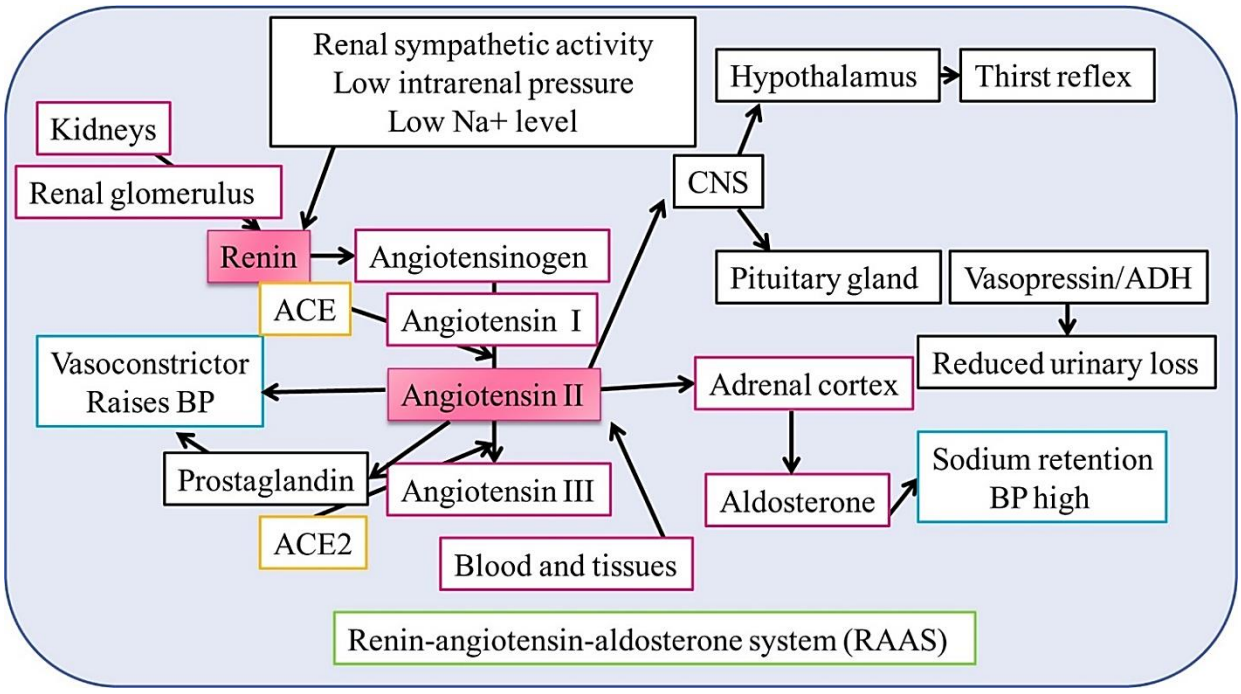

Figure 1.1: Schematic of the Renin-Angiotensin-Aldosterone System (RAAS) showing renin-triggered conversion of angiotensinogen to angiotensin I, then to angiotensin II by ACE, leading to vasoconstriction and aldosterone release. ACE2 mitigates these effects by converting angiotensin II to angiotensin III, while the pathway also regulates fluid balance via vasopressin and thirst mechanisms. [12].

AT1R and Ang II type 2 receptor (AT2R) are the primary receptors via which Ang II mediates its physiological effects [29]. AT1R controls blood pressure regulation, blood vessel contraction, electrolyte balance, aldosterone production, and adrenal cortex release [30]. Both receptors have different and opposite effects as AT1R mediates proliferation, vasoconstriction, fibrosis, and inflammation while AT2R mediates anti-inflammation, anti-fibrosis, and vasodilation [31]. AT1R and AT2R both are G-coupled protein receptors (GPCRs). They have a sequence identity of approximately 30%, yet they have the same binding affinity for their main ligand, Ang II [32]. The binding of Ang II with AT1R is involved in the pathology of many diseases. AT1R promotes activation of several signaling pathways that involve (ERK1/2), extracellular signal-regulated kinases, mitogen-activated protein kinase (MAPK), nicotinamide adenine dinucleotide phosphate

(NADPH) oxidase (NOX), nuclear factor κ-light-chain-enhancer of activated B cells (NF-κB), and signal transducer and activator of transcription 1 (STAT1) pathways [33]. In the lungs, the unbalanced signaling of AT1R is linked with pulmonary hypertension, fibrosis, bronchial hyperresponsiveness, and inflammation of the airways as shown in **Figure 1.2** [34]. The Ang type 4 (AT4R) is widely distributed and can be found in many different organs, such as the kidney, lung, heart, and adrenal gland. AT4R has a high-affinity binding site for Ang IV. Ang IV in the kidney reduces Na+ transport in isolated renal proximal tubules and increases renal cortical blood flow [35]. When AT4R binds with AT1R, it results in vasoconstriction, inflammation, and thrombosis [36]. Angiotensin-converting enzyme 2 (ACE2) is responsible for cleaving Ang II and producing Ang (1-7) that activates the G protein-coupled Mas (mitochondrial assembly-1) receptor producing vasodilation and lowering blood pressure.

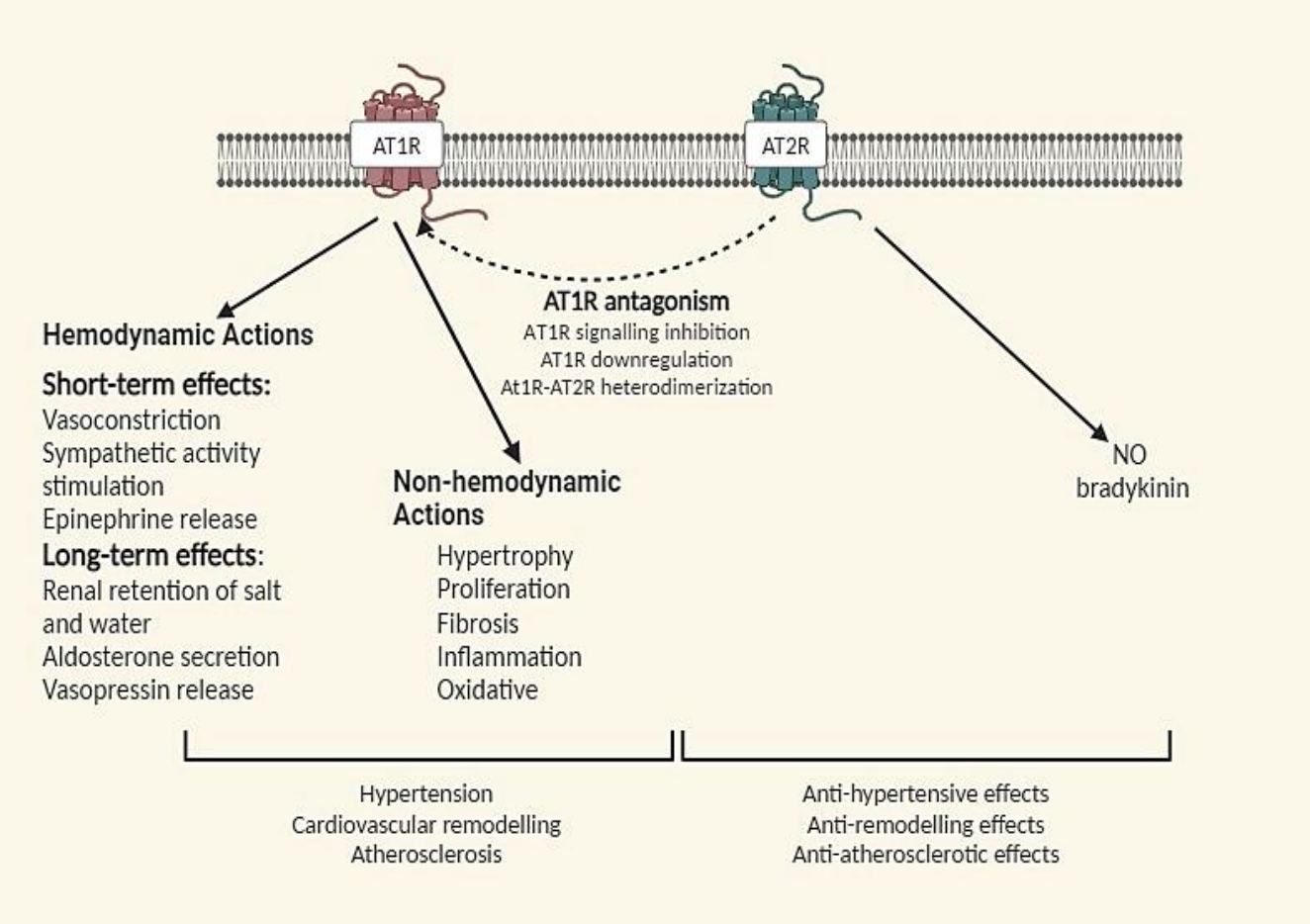

Figure 1.2: Angiotensin II Receptors. Ang II type 1 receptor (AT1R) and Ang II type 2 receptor (AT2R) are the two main receptor isoforms for Ang II [37].

1.3 Angiotensin Converting Enzyme (ACE 2):

ACE 2 is the human homologue of ACE and it adds more complexity to RAAS. Since its discovery in 2000, ACE2 has been identified as having a dual function in both healthy and pathological states, especially in relation to illnesses like pulmonary arterial hypertension (PAH) and viral infections like SARS-CoV-2. Upregulation of ACE2 expression is linked with therapeutic effects for lung disorders, diabetes, and cardiovascular and renal diseases [38]. ACE2 balances RAAS hypertensive actions by decreasing the overexpression of Ang I and II and acts as a natural ACE inhibitor (ACEI) and Ang Receptor Blockers (ARBs). ACE2 inactivates Ang II and negatively

regulates RAAS. ACE2 converts Ang II into Ang 1-7 [39]. However, the first crystal structure of ACE2 was determined in native (apo) with PDB ID 14R2 and the inhibitor-bound state with PDB ID 1R4L in 2004 [40]. ACE2 is divided into two subdomains (S1 and S2). The movement of subdomains depends upon the binding of the substrate. There are two confirmations of ACE i.e. open or close. Hinge movement of S1 happens to close the gaps upon binding with the substrate and adopts a closed conformation. In the closed conformational state, the protein's active site hides from the external environment. Subdomains move apart from each other and adopt an open state in the absence of any substrate [41]. ACE 2 counterbalances the detrimental effects of Ang II. ACE2 can hydrolyze Ang II and many other physiological substrates. As a type I integral membrane protein, ACE2 is expressed in the majority of human tissues and cell types [42]. There is highest expression level of ACE2 in the kidneys, testis, thyroids, heart, small intestine, and adipose tissue, whereas the lungs, colon, liver, bladder and adrenal glands have the intermediate level of expression and the blood, bone marrow, spleen, brain, blood vessels and muscles have the lowest expression of ACE2 [43]. With 805 amino acids, ACE2 is a type I transmembrane protein. Its structure consists of a C-terminal collectrin-like domain with an intracellular region and a transmembrane helix, as well as an N-terminal peptidase domain (PD) as mentioned in **Figure 1.3**. The zinc metallopeptidase domain of the PD, which catalyses the enzyme's action, has similarities to that of other metalloproteases in the M2 family. The N-terminal peptidase domain of ACE2 has seven possible N-linked glycosylation sites and has the characteristic HEXXH motif. There are two structurally distinct forms of ACE2: a soluble form and a full-length membrane-bound version. The membrane-bound form largely performs its enzymatic action at the cell surface. Metalloproteases such as ADAM17 shed the extracellular domain, resulting in the soluble form that circulates in the bloodstream while retaining its catalytic activity [44]. ACE2 has claw-like

crystal structure, as revealed by X-ray crystallography. It binds to its substrates, angiotensin II (Ang II) and angiotensin I (Ang I), converting them into angiotensin-(1-7) and angiotensin-(1-9), respectively [44]. ACE2 affects the RAAS pathway and its potential as a therapeutic target for the treatment of PAH.

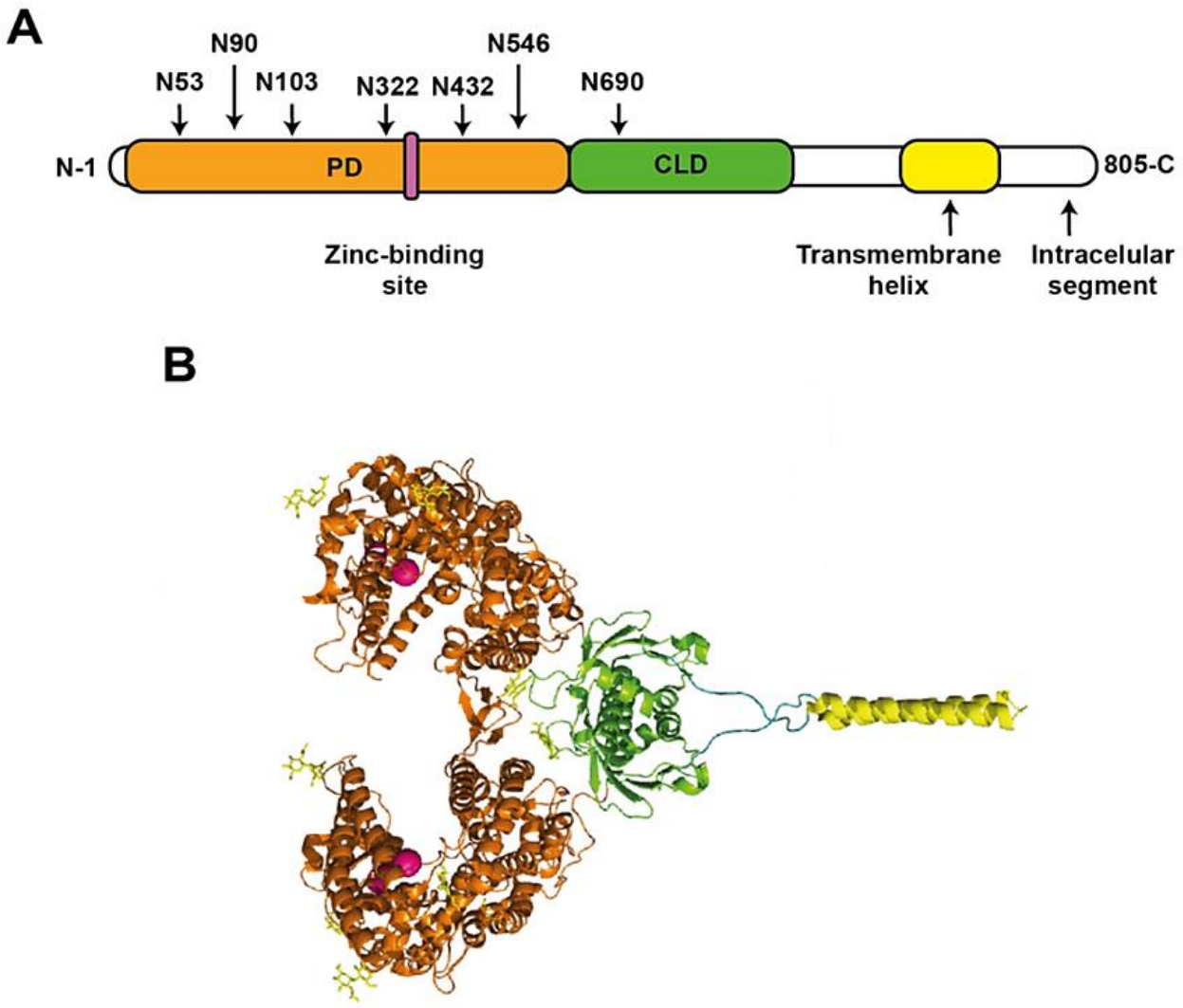

Figure 1.3: Structural Overview of ACE2. (A) Schematic of ACE2 protein showing domains: peptidase (PD) with a zinc-binding site, collectrin-like (CLD), transmembrane helix, and intracellular segment, along with N-glycosylation sites. (B) 3D structure of ACE2 highlighting the PD (orange), CLD (green), and transmembrane/intracellular regions (yellow), with the zinc-binding site marked in magenta. [45].

1.4 Role of ACE2 in RAAS:

The imbalance between angiotensin-(1–7) and Ang II is a major determinant in the pathogenesis of PAH. Increased pulmonary vascular resistance and pressure result from the promotion of pulmonary vasoconstriction, inflammation, and fibrosis by Ang II [46]. One of the main characteristics of PAH is the overactivity of Ang II, which aggravates the disease by causing vascular remodeling and endothelial dysfunction. The enzymatic activity of ACE2 suppresses these pathogenic processes. The activity of ACE2 in PAH has important therapeutic implications. In experimental models, it has been demonstrated that recombinant ACE2 or increasing ACE2 activity can treat PAH symptoms. This treatment approach improves right ventricular function, lowers pulmonary artery pressures, and prevents pulmonary vascular remodeling. Research has indicated that pulmonary hemodynamics and vascular structure can be significantly improved by ACE2 overexpression or by using ACE2 activators. ACE2 efficiently lowers Ang II levels while boosting the advantageous effects of angiotensin-(1-7) by converting Ang II to angiotensin-(1-7) [47]. This change in the pathway aids in reducing PAH symptoms. Recombinant ACE2 treatment, for instance, has been shown to raise angiotensin-(1-7) levels, which in turn reduces pulmonary artery inflammation and fibrosis. This therapy strategy emphasizes the possibility of treating PAH by focusing on the ACE2/angiotensin-(1-7)/Mas axis. Angiotensin-(1–7) leads to vasodilation and protects against endothelium damage and fibrosis by acting through the Mas receptor [46].

In short, ACE2 is essential to the pathophysiology of PAH by controlling the RAAS pathway. ACE2 aids in balancing the negative effects of Ang II by converting it to angiotensin-(1–7), which in turn promotes vasodilation, lowers inflammation and inhibits fibrosis. The therapeutic potential of ACE2 in PAH is encouraging, providing opportunities for novel therapeutic approaches targeted

at augmenting ACE2 activity or delivering recombinant ACE2 to slow down the advancement of PAH.

1.5 Problem Statement:

Angiotensin Converting Enzyme 2 (ACE2 has shown potential in counteracting the detrimental effects of Ang II on the cardio-pulmonary system. However, there is a lack of comprehensive data on the ability of ACE inhibitors to activate ACE2 and modulate its protective functions in the context of PAH.

1.6 Objectives:

- To collect and preprocess the existing data of the inhibitors of ACE.
- To prepare the binding complexes of the ACE and ACE2 with the collected and preprocessed data for the generation of the data matrix.
- To perform the Molecular Docking guided attribute computation.
- Development of ML models for the prediction of the activation potential of the ACE inhibitors against ACE2.

Chapter 2 Literature Review:

2.1 Detrimental effects of Ang II:

The primary effector and naturally occurring hormonal peptide of the renin-angiotensin-aldosterone system (RAAS), angiotensin II (Ang II), has both pathological and physiological effects. It affects blood vessel function and influences the heart, kidneys, and brain. Ang II mediates these effects. It is essential for controlling sodium (Na+), blood pressure (BP), fluid and sodium retention, inducing vasoconstriction, and encouraging cell division. On the other hand, long-term stimulation of Ang II causes damage to multiple organs through sympathetic overload, increased secretion of aldosterone, fibrosis in the brain, and muscle hypertrophy.

Aldosterone is a steroid hormone with mineralocorticoid (MR) action that is released from the kidney's adrenal cortex when Ang II and endothelin-1 act on it. When MR receptors are overactivated, this hormone can cause hypertension because it increases renal Na+ retention and potassium (K+) excretion at the distal tubule [48]. Leading roles in the development of cardiovascular diseases (CVDs), particularly hypertension, are played by both aldosterone and Ang II. Patients with pulmonary arterial hypertension (PAH) have been found to have elevated levels of adrenocorticotropic hormone, Ang II, endothelin-1, and plasma K+, all of which promote the synthesis of aldosterone. Research indicates that kidney and cardiovascular illnesses may be influenced by Ang II's activation of MR receptors. Renal fibrosis, vascular remodeling, and heart hypertrophy are all promoted by prolonged exposure to Ang II. Several signaling pathways are activated when abnormal RAAS activation enhances ACE, which in turn enhances Ang II levels. Mitogen-activated protein kinases (MAPKs), for example, are activated by Ang II, and extracellular signal-regulated kinase (ERK 1/2), JAK/STAT, platelet-derived growth factor

(PDGF), and focal adhesion kinase (FAK) receptors which are MAP, tyrosine, non-receptor tyrosine, and focal adhesion kinase receptors, respectively are stimulated by Ang-AT1R binding. Furthermore, AT1R triggers the activation of NADPH oxidase, which results in the generation of reactive oxygen species (ROS) and causes vascular inflammation, hypertension, and the activation of protein kinase C (PKC), which increases the production of NADPH and ROS more [49].

2.2 Ang II in PAH pathogenesis:

Numerous studies have focused on Ang II significance in PAH. Usui et al. examined the impact of neurohormonal variables in mice with PAH that were induced to produce monocrotaline (MCT). In MCT-induced rats, the RNase protection assay showed elevated plasma levels of Ang II and noradrenaline. Biventricular hypertrophy was shown in these rats, although short-term survival was greatly improved when ACE inhibitors such valsartan and carvingilol were administered [50]. In another study, Qiao et al. produced Ang II and cell proliferation in rats to examine the impact of Notch signaling inhibitors on Ang II-induced pulmonary vascular resistance (PVR). Vascular wall thickness and cell proliferation were markedly enhanced by Ang II stimulation; however, the effects of Ang II were decreased when treated with DAPT inhibitor, indicating that Ang II inhibition may be useful in lowering PVR [51]. Increased level of ACE2 counterbalances the detrimental effects of Ang II. In a study, recombinant ACE2 injections were injected to mice, and it was observed that ACE2 inactivated Ang II and negatively regulated RAAS. ACE 2 converts Ang II into Ang 1-7. Acute lungs injury causes reduction in ACE2 expression and increases Ang II level. Thus, recombinant ACE2 can be used as therapy for lung failure in mice [52]. Zhu et al. did genetic analysis of 13 polymorphisms in study group. Association analysis of all biomarkers with ACE concentration was performed and then ACE concentration and blood pressure level of study group was measured. ACE polymorphism was observed to be associated with variation in

ACE concentration and blood pressure level [53]. Various studies have demonstrated that the pulmonary artery smooth muscle cells (PASMCs) that are involved in the pathophysiology of hypoxia-induced PH are stimulated abnormally by ACE, leading to their proliferation and displacement. In a study, Zhang et al. concluded that hyperoxia could result in upregulation of ACE in hPASMCs, upregulation of ACE2 in early stages and downregulation of ACE2 later. HIF-1α, a transcripting factor activated during hyperoxia. HIF-1α caused the upregulation of ACE protein expression. The reduction of ACE2 protein in later stages of hyperoxia may contribute to pulmonary vascular remodeling in hypoxic pulmonary hypertension [54]. ACE2 expression plays an important role in the therapeutics of cardiovascular diseases (CVDs). To evaluate the effects of ACE2 on cardiovascular functions, Yamamoto et al. performed transverse aortic constriction (TAC) on ACE2−/y and WT mice. Some ACE2−/y mice received candesartan therapy. Measurements of Ang II content, left ventricular pressure, proximal aortic pressure, lung tissue histology, gene expression and protein analysis were done. ACE2 deletion leaded to hypertrophy and dysfunction in reaction to pressure overload by increasing conc. Of Ang II. ACE2−/y mice experienced more congestion and high rate of cardiac death incidents under pressure overload conditions. Thus, ACE2 plays an important role in reducing effects of pressure overload on heart [55]. Diabetes may lead to cardiovascular diseases by increasing Ang II level and lowering ACE2 expression. To predict association of diabetes and CVDs, Tonon et al. did case study that compared 33 Type-1 Diabetes (TID) patients to 30 controls. ACE and ACE2 gene expressions were evaluated. Diabetes patients had higher level of Ang II and ACE and lower level of ACE2. Thus, diabetes results in upregulation of Ang II and ACE level, and downregulation of ACE2 that may lead to cardiovascular diseases [56]. ACE2 deficiency results in severe ventricular contractility abnormalities and increased levels of Ang II, underscoring the enzyme's role in reducing Ang II's

detrimental effects in cardiovascular disorders. Another study revealed that ACE2 deletion mice were more susceptible to hypertension and renal diseases caused by higher Ang II, highlighting the protective role of ACE2 against the harmful effects of Ang II [57]. Wysocki et al. observed in a study that ACE2 deficiency increased the severity of kidney injury by decreasing Ang-(1-7) levels and increasing Ang II levels and it demonstrated the involvement of the ACE2 in maintaining renal function [58]. High level of Ang II leads to vasoconstriction and inflammatory effects that further results in PAH.

2.3 Ang II Receptors:

Angiotensin II (Ang II) exerts its effects primarily through its receptors, Angiotensin II type 1 receptor (AT1R), Angiotensin II type 2 receptor (AT2R), Angiotensin II type 3 receptor (AT3R), and Angiotensin II type 4 receptor (AT4R). These receptors mediate a variety of biological functions and are members of the G-protein-coupled receptor (GPCR) family.

2.3.1 Angiotensin II Type I Receptor (AT1R):

Most of the effects of Ang II, such as vasoconstriction, aldosterone secretion, sodium reabsorption, and sympathetic nervous system activation, are mediated by AT1R, a seven-transmembrane GPCR. AT1R activates multiple intracellular signaling pathways upon binding to Ang II, including phospholipase C, protein kinase C, and mitogen-activated protein Kinases (MAPKs [59]. The regulation of cardiovascular and renal function depends on AT1R. It helps to keep the fluid-electrolyte balance and blood pressure stable. Chronic activation of AT1R has been linked to kidney disorders, heart failure, cardiac hypertrophy, and hypertension. In pathological conditions, AT1R also stimulates oxidative stress, fibrosis, and inflammation, which adds to organ damage and vascular remodeling [60]. ARBs, such as candesartan, valsartan, and losartan, are often prescribed medications for hypertension. By specifically blocking Ang II's effects at the

AT1R, these medications lower blood pressure and guard against kidney and cardiovascular damage. Chronic kidney disease, heart failure, and hypertension have all been effectively treated with ARBs [61].

2.3.2 Angiotensin II Type 2 Receptor (AT2R):

Another seven-transmembrane GPCR, AT2R, exerts effects that are frequently in opposition to those of AT1R. Vasodilation, anti-inflammatory, anti-fibrotic, and anti-proliferative actions are often enhanced by AT2R activation. It promotes these actions using a variety of signaling channels, such as nitric oxide synthesis, inhibition of MAPK pathways, and activation of protein phosphatases [62]. AT2R plays a complicated role in both physiology and disease. Due to its high expression during fetal development, AT2R may have a role in development and growth. Its expression is elevated in pathological circumstances such as renal damage, heart failure, and myocardial infarction in adults. By encouraging vasodilation, lowering oxidative stress, and preventing cell division and inflammation, AT2R is thought to balance out the negative effects of AT1R [63]. Research on the therapeutic potential of targeting AT2R is ongoing. The potential benefits of compounds like C21, which selectively activate AT2R, are being studied for their potential benefits in cardiovascular and renal disorders. By boosting AT2R protective functions while avoiding the negative effects of AT1R activation, these AT2R agonists may present a novel therapeutic approach.

2.3.3 Angiotensin II Type 3 Receptor (AT3R):

AT3R is less well characterized compared to AT1R and AT2R. It is believed to play a role in Ang II signaling, but its exact physiological and pathological roles remain unclear. Some studies suggest that AT3R may participate in the modulation of blood pressure and fluid balance, like AT1R and AT2R [64]. The specific functions of AT3R in the cardiovascular and renal systems are

still under investigation. Preliminary research indicates that it may have roles in modulating vasoconstriction and sodium reabsorption, although more studies are needed to elucidate its exact mechanisms and effect [65].

2.3.4 Angiotensin II Type 4 Receptor (AT4R):

AT4R or Insulin-regulated aminopeptidase (IRAP), is the enzyme that binds peptides produced from Ang II, including Ang IV. Its structure and function set it apart from AT1R and AT2R, mainly impacting memory and cognitive functions [66]. In the brain, AT4R is abundantly expressed, especially in regions linked to learning and memory. It has been investigated for a possible involvement in neurodegenerative illnesses and is implicated in the control of cognitive functioning. Although its functions in the cardiovascular system are not fully understood, AT4R may also affect vascular function and blood flow [67]. Targeting AT4R may open new therapeutic options for the treatment of certain cardiovascular diseases as well as cognitive problems. Adjusting this receptor could improve learning and memory, providing chances for addressing conditions like Alzheimer's disease.

2.4 Therapeutic Effects of ACE2:

ACE2 counterbalances the detrimental effects of Ang II by converting it into Ang 1-7. Various studies have demonstrated that increased level of ACE2 has therapeutic effects on cardiovascular diseases. Batlle et al. examined the potential therapeutic advantages of soluble recombinant ACE2 (rhACE2) in mitigating the deleterious consequences of Ang II, particularly in the context of acute pulmonary damage. It was observed that administering rhACE2 efficiently raised angiotensin-(1-7) levels and decreased Ang II levels, hence offering protective effects against lung injury caused by elevated Ang II. According to the findings, rhACE2 may be a useful treatment for diseases like COVID-19 and CVDs that are marked by increased Ang II [68]. Diminazene aceturate (DIZE) can

be used as an ACE2 activator. In experimental animals, a study demonstrated that DIZE markedly improved myocardial fibrosis and hypertension, decreased Ang II levels, and boosted ACE2 activity. The results suggest that by modifying the ACE2/Ang-(1-7) axis, DIZE may prove to be a useful therapeutic drug for the treatment of cardiovascular disorders [69]. In another study, the effects of XNT (Xanthine Derivative), an ACE2 activator, on hypertensive rats were examined. It was discovered that administering XNT increased ACE2 activity, decreased Ang II levels, and resulted in cardiovascular protection. The findings imply that by focusing on the ACE2/Ang-(1-7) axis, XNT and related compounds may be created as novel antihypertensive medications [70]. Gurley et al. investigated the benefits of ACE2 activation on the cardiovascular system. In animal models, it was discovered that ACE2 activators dramatically decreased Ang II-mediated hypertension and the problems associated with it. The research underscored the significance of ACE2 in preserving cardiovascular well-being and proposed that ACE2 activators may have therapeutic utility in the management of hypertension and associated cardiovascular conditions [71]. Thus ACE2 inhibitors can be used for treating hypertension and cardiovascular diseases. ACE2 activation results in decreased Ang II levels and improved cardiac function. ACE2 activators has potential to reduce hypoxia-induced right ventricular hypertrophy and pulmonary hypertension and can be used as a new therapeutic approach for cardiovascular diseases [72]. ACE2 activators have demonstrated potential as a defense against renal damage caused by Ang II. Renal inflammation and fibrosis are reduced by ACE2 activators by decreasing Ang II and increasing Ang-(1–7). In diseases like diabetic nephropathy, where increased Ang II causes kidney damage, this preventive action is crucial. Ang-(1–7) has strong anti-inflammatory properties and is produced when ACE2 is activated. By lowering pro-inflammatory cytokine and chemokine levels, ACE2 activators aid in the mitigation of chronic inflammation linked to increased Ang II

levels. Numerous inflammatory conditions, including cardiovascular and renal disorders, benefit from this impact. There are different studies supporting Ang-(1-7) anti-fibrotic qualities. ACE2 activators reduce fibrosis in the kidneys, heart, and lungs by increasing ACE2 activity. This therapeutic impact plays a critical role in treating diseases that cause scarring and tissue remodeling [56].

2.5 Previously Discovered Drugs:

Prostanoid, endothelin receptor antagonists, and phosphodiester type 5 inhibitors are commonly used medications for the treatment of PAH [7]. Prostanoids work by mimicking the effects of prostacyclin which helps in vasodilation. Endothelin-receptor antagonists block the effects of endothelin and reduce pulmonary arterial pressure. PDE-5 inhibitors inhibit the PDE-5, an enzyme that cleaves cyclic guanosine monophosphate (cGMP), a molecule that dilates blood vessels. The combination of ACEI and natural endopeptidase (NEP) is an effective medication for PAH, but their combination can lead to angioedema. Serotonin, vasoactive peptides, oxidative stress, tyrosine kinase, Rho-kinase, and metabolic pathways, as well as anti-inflammatory agents can be used as therapy for treating PAH. Other already-developed drugs target endothelin-1, prostacyclin, and nitric oxide pathways for PAH treatment. These drugs do not ensure quality of life for PAH patients. PAH pathogenesis involves many molecular processes.

Genetic mutation in BMPR2 and multiple pathways including potassium channels (KCNK3 and ABCC8), and transcription factors (TBX4 and SOX17) can be involved in PAH susceptibility and progression [73]. Elevated levels of cytokines, chemokines, and inflammatory mediators have been linked in PAH patients. Vascular and parenchymal cells, such as endothelial, smooth muscle, and fibroblasts, undergo phenotypic changes that alter their sensitivity to inflammatory triggers, resulting in active secretion of cytokines and chemokines [9]. Understanding the interactions

between inflammatory mediators and vascular cells in PAH patients can be helpful for the development of therapy for PAH treatment. As the first licensed activin signaling inhibitor therapy, sotatercept is the first new treatment for PAH in over ten years. Activin signaling inhibitors are a new family of medications that regulate vascular cell proliferation underlying PAH by enhancing the balance between pro- and anti-proliferative signaling. Humbert et al. supported the efficacy and long-term safety of the clinical benefits of stotarcept for PAH [74]. The level of EYA3 is also associated with PAH pathogenesis. The level of EYA3 is elevated in pulmonary arterial smooth muscle cells in PAH. EYA3 inhibitors were used as potential drugs for PAH treatment [10]. RAAS activation promotes cardiovascular fibrosis and stiffness. RAAS activation causes an increase in collagen deposition that can result in myocardial stiffness. Drugs targeting RAAS components were used as a potential therapy for PAH treatment as available drugs are mentioned in **Figure 2.1**. Sex hormones, genetic abnormalities, elastase inhibition, metabolic dysfunction, cellular therapies, and anti-inflammatory approaches can be alternative approaches for the treatment of PAH [6]. Various studies have demonstrated that ACE inhibitors can be potential drug for the treatment of PAH. To predict the effect of AT1R blocker on pulmonary arterial pressure in ethanol-induced PAH model, Tanriverdi et al. administered ethanol to swine to induce PAH. Then losartan (Ang II receptor blocker) was injected to test its impact on PAH and concentration of renin-angiotensin system (RAS) ligands, oxidative stress levels, and P38 MAPK signaling were measured in the pulmonary artery. Losartan partially inhibited the elevated pulmonary arterial pressure and moderately reversed the pulmonary arterial remodeling in PAH. It worked by inhibiting oxidative stress and P38 MAPK signaling. Losartan did not affect the gene expression of AT2R [75]. Lisinopril, an ACE inhibitor is used as a therapy for PAH treatment. ACE Inhibitor increases the level of Ang 1-7 and ACE2 mRNA expression while AT1 receptor blocker increases

both levels of Ang 1-7 and Ang II. Their combined therapy has a similar effect to ACE inhibitors alone with decreased ACE2 mRNA expression [76]. Drugs targeting BMPR2 mutations, apoptosis of PAECs and proliferation of PASMCs, inflammation, epigenetic modifications, and metabolic pathways are also used as potential therapies for the treatment of PAH [77].

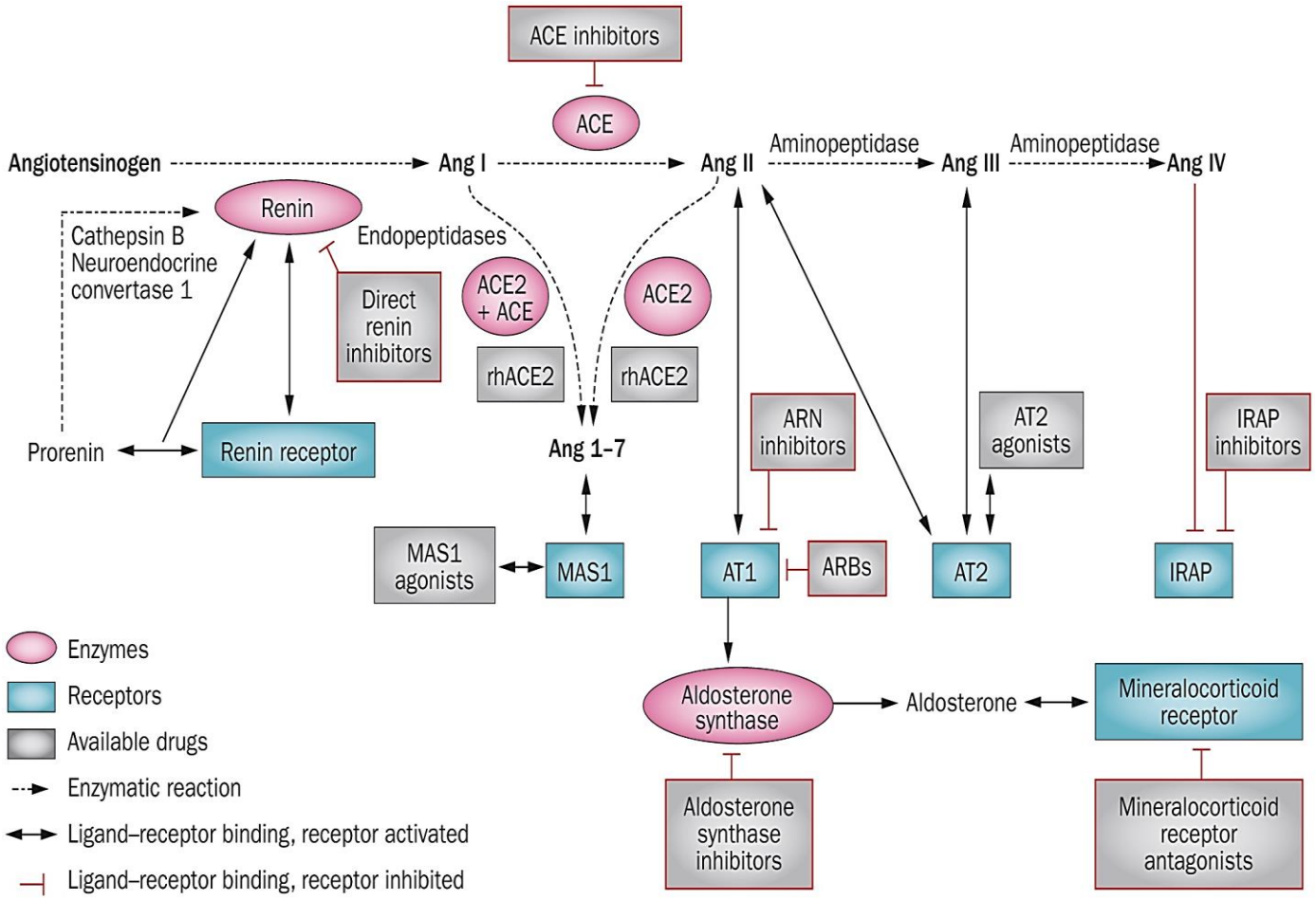

Figure 2.1: Schematic Overview of the Renin-Angiotensin System (RAS) and Therapeutic Interventions. This diagram illustrates the key enzymatic reactions and regulatory pathways within the RAS, highlighting the roles of ACE and ACE2 in converting angiotensinogen to various angiotensin peptides [78].

Chapter 3 Materials and Methods:

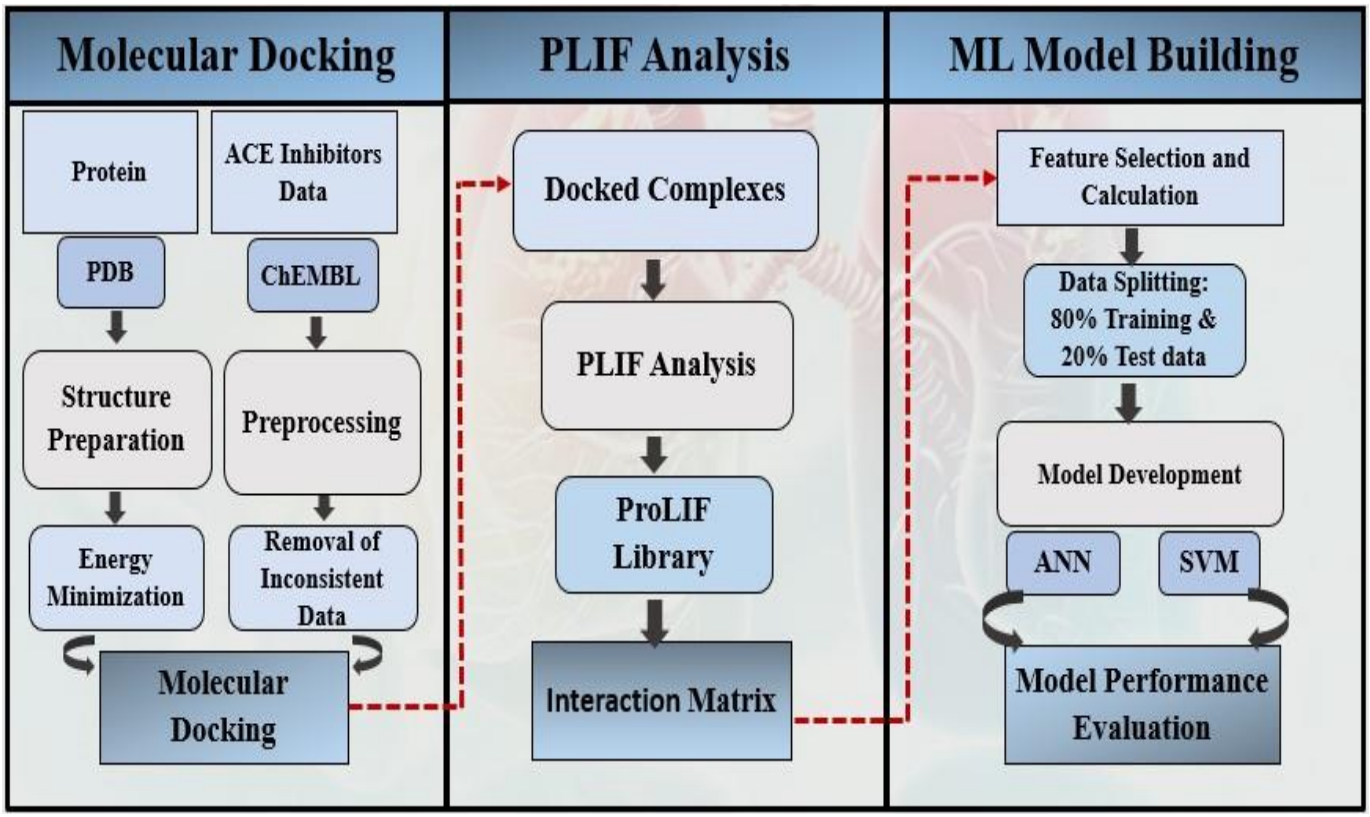

Figure 3.1: Overall workflow of research methodology. **Step 1:** Molecular Docking, **Step 2:** PLIF Analysis and **Step 3:** Machine Learning Model Building.

3.1 Data Collection:

From the ChEMBL database, a set of 835 compounds of ACE with their biological activity (IC_{50}) values ranging from 0.029 nM to 100 mM was downloaded. The initial list of compounds was reduced to 591 compounds after the preprocessing of dataset which involved removing duplicates and excluding compounds with unclear potency values or a molecular weight (MW) of 200 or below as mentioned in **Appendix 1**. An open-sourced software OpenBabel was used to generate the three-dimensional (3D) structures of these compounds from their 2D SMILES representations [79]. This process adjusted the atomic coordinates to represent the most energetically favorable

conformations of the molecules., Principal Component Analysis (PCA) was performed to probe the structural diversity within the dataset. PCA reduced the dimensionality of the data which distills it into principal components that explain the majority of variance within the dataset. This statistical approach clarified the underlying patterns in the structural data and aided in identifying key features correlating with biological activity.

3.2 Protein Structure Collection and Preparation:

3.2.1 ACE2:

In molecular modeling, we selected the most recent structure as this foundation significantly impacts the subsequent analyses. We used high-resolution structure of ACE2. We focused on the ACE2 protein due to its relevance in SARS-CoV-2 research where it acts as a receptor for the virus spike protein. The most recent structure of human ACE2 in the apo state was retrieved from the PDB with an ID of 7lo4 and a resolution of 2.46Å that highlights its clarity and detail [80]. The structure was processed thoroughly to ensure the structural integrity and suitability of our model by removing any heteroatoms or bound ligands. The 3D structure underwent energy minimization using AMBER 99 forcefield to improve their conformation.

3.2.2 ACE:

The PDB's most current X-ray crystallographic structure of human ACE (PDB ID: 6TT3) was selected with a resolution of 1.70 Å [81]. To ensure the reliability of this structural model for further computational tasks, a series of preprocessing steps were performed. Initially, the ACE structure was cleaned to remove any irrelevant bounded ligands that could interfere with subsequent processes. Subsequently, the water molecules were removed from the structure. Following this, the structure went through an energy minimization process to refine the conformation towards a more energetically favorable state, ensuring the stability of the model.

Energy was minimized by using AMBER 99 forcefield in Molecular Operating Environment (MOE).

3.3 Molecular Docking:

Molecular docking was employed to predict the binding hypothesis of ACE inhibitors with ACE and ACE2. The goal was to achieve optimized binding conformations and gain insights into the protein-ligand interactions. Firstly, energy minimized ACE2 along with all preprocessed 591 ACE inhibitors were imported to Genetic Optimization for Ligand Docking (GOLD suite v 5.3) to perform molecular docking. GOLD utilizes a Genetic algorithm (GA) that allows for a partially flexible receptor and full ligand flexibility [82]. The docking efficacy was assessed using the GOLD fitness score which incorporates several energy parameters including internal and exterior hydrogen bond energies (HB_int, HB_ext), Van der Waals forces (VDW_int, VDW_ext), and internal torsion(TOR_int). The equation for GOLD fitness score is stated below:

$$GOLDScore = \Delta G\,(HB\_int) + \Delta G\,(HB\_ext) + \Delta G(VDW\_int) + \Delta G\,(VDW\_ext) + \Delta G\,(TOR\_int)$$

The binding pocket for docking was carefully chosen based on previous studies identifying key residues in ACE2 that interact with bile acids [83]. The binding site was selected by including residues that surround the hinge region which is situated between two subdomains as its the previously noted site for ACE2 activator interaction as mentioned in **Figure 3.2**[84]. Docking parameters were set to encompass this region that ensures that all significant residues are included. Up to 10 genetic algorithm runs per ligand were executed. The results were analyzed in Molecular Operating Environment (MOE) to perform pose analysis that ensures the selection of the most suitable conformation for each ligand based on gold fitness score.

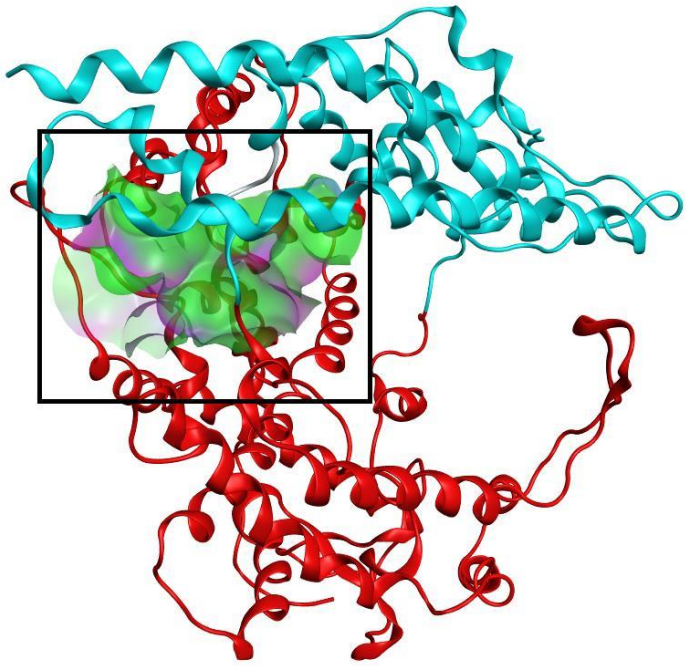

Figure 3.2: Binding site of ACE2. The Binding site is located in the hinge region between two subdomains.

Then the energy minimized ACE was imported in GOLD along with preprocessed ACE inhibitors. The same docking parameters were selected for this protein. All the ACE inhibitors bind to zinc ion present in the active site as mentioned in **Figure 3.3**. [85]. Ten GA runs per ligand were generated. A correlation plot was created to compare the pIC50 values, which represent the inhibitory potency of the ligands, against their GOLD scores. This analysis aimed to establish a relationship between the biochemical activity and the binding affinity of the ligands to ACE and ACE2, providing a comprehensive view of how these inhibitors may modulate enzyme activity.

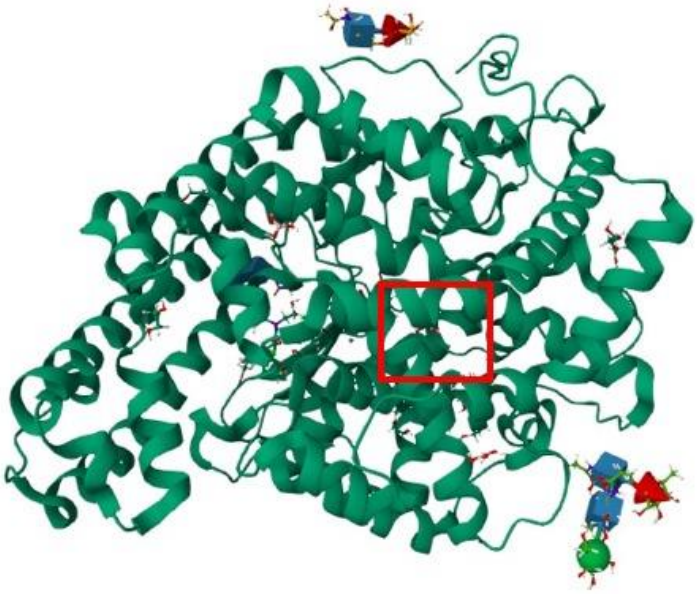

Figure 3.3: Binding Site of ACE. All previously reported ACE inhibitors bind to the zinc ion present in the binding site.

3.4 Protein-Ligand Interaction Fingerprint (PLIF):

Protein Ligand Interaction Fingerprint (PLIF) analysis is a computational technique used to detail the interactions between a protein and various ligand's conformations based on molecular docking results. PLIF determines the presence or absence of specific types of interactions such as hydrogen bonds, hydrophobic contacts and ionic bonds between protein residues and ligand atoms based on docking results. PLIF was done by a python script using prolif library [86]. For conducting PLIF analysis, computational environment was set up with all necessary tools and libraries. Initially, several python libraries which are pivotal for our analysis were installed. This included netCDF4 for handling and storing scientific data, MDAnalysisTests and MDAnalysisData for molecular dynamics analysis, prolif for generating protein-ligand interaction fingerprints, and rdkit, which is essential for cheminformatics and machine learning applications in drug discovery. MDAnalysis was used to handle and analyze molecular dynamics trajectories, pandas alongside numpy facilitated data manipulation and numerical operations which are critical for handling the output

from prolif. The core of our PLIF analysis involved processing the molecular structures of ACE and ACE2 proteins and the ligands, using the prolif library. This library enabled us to identify and record specific interactions such as hydrogen bonds, hydrophobic interactions, and ionic bonds, between the ligands and the amino acids in the protein. The binary format used to encode each interaction providing a detailed representation of the ligand binding patterns. Data generated from the PLIF analysis forms a comprehensive dataset that captures the complex details of protein-ligand interactions. This dataset is particularly suited for use in machine learning models as it provides a rich feature set that describes how different ligands interact with specific residues within the ACE and ACE2 proteins.

3.5 Machine Learning Models:

The dataset derived from PLIF analysis underwent a critical phase of preprocessing to optimize it for subsequent analyses. This included normalizing the data, managing any missing values, and converting categorical variables to a format suitable for machine learning algorithms.

Once the data was preprocessed, it was analyzed to determine the percentage of interactions each ligand had with both ACE and ACE2 proteins. This analysis was pivotal in establishing a classification system for the ligands based on their interaction profiles. Ligands were categorized based on their number of interactions as if a ligand exhibited a higher percentage of interactions with ACE, it was labeled as '1' which indicated it was an ACE inhibitor. As if a ligand interacted more frequently with ACE2, it was labeled as '0' which indicated it as an ACE2 activator. The ligands were classified and labelled into two binary classes as 1 and 0. This binary classification formed the foundation of the training dataset for the machine learning model. By applying this labeling approach, we were able to clearly define two classes of compounds which the machine learning model could then learn to predict. This approach not only leveraged the detailed

interaction data to classify compounds accurately but also facilitated a streamlined and effective methodology for predicting the potential therapeutic category of novel compounds based on their molecular interaction profiles. This robust classification system enhances our ability to quickly identify promising candidates for further drug development and therapeutic intervention. The model performance was evaluated by precision, accuracy, recall and F1-score. The equations of these parameters are stated below:

Precision = True positives/ (True positives + False positives)

Accuracy = (True positives + True Negatives)/ (True positives + True negatives + False positives + False negatives)

Recall = True Positive / (True Positive + False Negative)

F1-score = 2 * (Precision * Recall) / (Precision + Recall)

3.5.1 Artificial Neural Network (ANN):

Pandas, a python data manipulation library was used to load the dataset containing interaction frequencies between compounds and both ACE and ACE2 proteins. The features were labelled as 'Percentage (ACE)' and 'Percentage (ACE2)', and they were extracted along with the target variable 'Label'. Each compound is classified as an ACE inhibitor or ACE2 activator based on the interaction frequency where a binary label of '1' indicates an ACE inhibitor and '0' an ACE2 activator.

The dataset was split into training and testing subsets with 80% of the data for training and 20% for testing. The train_test_split function from scikit-learn was used to accomplish this split, guaranteeing that the model could be trained and then test on unseen data. To ensure reproducibility of the outcomes, a random seed (random_state=42) was incorporated into the

splitting procedure. To standardize the range of independent variables or features in the training data, feature scaling was done. A StandardScaler from scikit-learn was used to normalize the feature set, removing the mean and scaling to unit variance. This step is critical for many machine learning algorithms and especially significant when using neural networks to ensure that the gradient descent algorithm used in training converges more quickly.

Using the Keras API from TensorFlow, a sequential model with several layers designed to learn from the dataset was built. Using the ReLU activation function, the first layer, which consists of 20 dense neurons, adds non-linearity to the model. A Dropout layer with a rate of 0.5 follows, randomly setting input units to 0 during training to avoid overfitting. The pattern is continued by an additional dense layer with 10 neurons and ReLU activation. An additional Dropout layer with a rate of 0.5 helps to alleviate the overfitting issue even more. The last layer is a dense layer with a single neuron that outputs a probability indicating the chance that the input is an ACE inhibitor using a sigmoid activation function.

For binary classification tasks, the binary cross-entropy loss function and Adam optimizer were used in the model's compilation. The training process of the model comprised 50 epochs with a batch size of 10 which specify the number of training dataset iterations and the number of training samples per gradient update. The overall accuracy along with precision, recall, and F1-score for each class, was documented that provides insights into the model's ability to distinguish between ACE inhibitors and ACE2 activators effectively.

3.5.2 Support Vector Machine (SVM):

Support Vector Machine (SVM) was built to categorize ligands as either ACE inhibitors or ACE2 activators based on their interaction frequencies with these proteins. Using the Pandas library, the labelled dataset was first loaded where primary features included 'Percentage (ACE)' and

'Percentage (ACE2)' and the target variable labeled compounds accordingly. To ensure robustness in our dataset, missing values were addressed using a Simple Imputer from scikit-learn which replaced missing entries with the mean of respective features. We also enhanced our feature set by incorporating polynomial and interaction terms through the PolynomialFeatures function that expands the linear feature space to capture more complex relationships between features.

To avoid biases and overfitting, the consistency of our dataset was carefully examined for duplicates and constant features. We also evaluated our class balance to find any skewness that could impact the model performance. We split feature set into training and testing subsets, allocating 80% for training and 20% for testing, to build the SVM model. The reproducibility of our results was ensured by using a fixed random state to control this split. Hyperparameter tuning was a critical step to optimize the SVM's performance which was carried out by GridSearchCV, that explored a range of values for parameters such as the penalty parameter C, kernel type, and gamma. The best model from GridSearchCV was then thoroughly evaluated on the test set. To evaluate the model's efficacy, performance measures including recall, accuracy, precision, and F1 score were computed. The comprehensive analysis of these matrices for both classes in the classification report shed light on the model’s discriminative ability. Cross-validation was performed on the entire feature set to validate the model's robustness that provides an unbiased estimate of the model's performance across different subsets of the data. The efficacy of the model was evaluated by different parameters.

Chapter 4 Results:

4.1 Molecular Docking:

4.1.1 Molecular Docking of ACE2:

The binding site for ACE2 was carefully chosen for docking because of its previously recognized significance in ACE2 activator interactions. The specific coordinates set for the binding site were X=15.4040, Y=34.9380, and Z=-46.3260. This site encapsulates crucial residues including Lys94, Leu95, Gln98, Tyr196, Tyr202, Asp206, Val209, Asn210, Pro565, and Trp566, which are strategically located around the hinge region situated between two subdomains of the protein. This region is known for its critical role in ligand interaction, making it a prime target for docking studies aimed at exploring potential ACE2 activators [80].

A total of 591 ligands as mentioned in **Appendix 1** were imported to the GOLD for docking with each ligand generating 10 potential poses. This approach allowed for a comprehensive assessment of how each ligand could theoretically interact with the binding site under different configurations. The evaluation of these poses was based on their respective GOLD fitness scores. The docking results were highly encouraging. The highest GOLD score observed among all poses was 93, indicating a particularly favorable interaction between the ligand and the ACE2 binding site as shown in **Figure 4.1**. All ligand poses achieved positive GOLD scores that suggest that each ligand configuration presented a viable interaction with the protein. This uniform positivity across scores signifies a broad potential for these ligands to act as effective ACE2 activators based on their docking profiles. These findings provide valuable insights into the molecular dynamics at play within the ACE2 binding site and highlight the therapeutic potential of these ligands in modulating ACE2 activity.

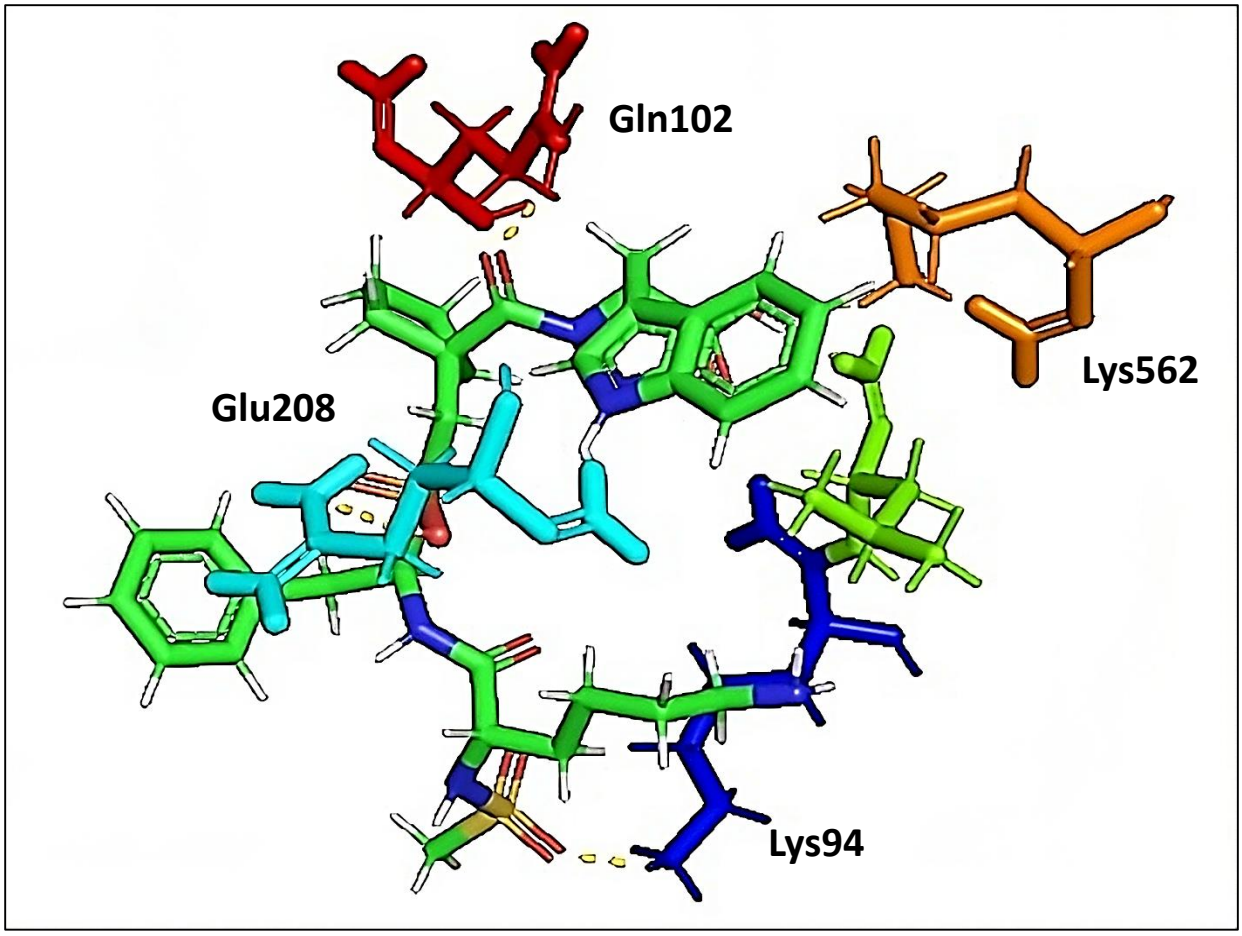

Figure 4.1: Best pose conformation of ligand with ChEMBL id (CHEMBL273140) having pIC_{50} value of 7.38

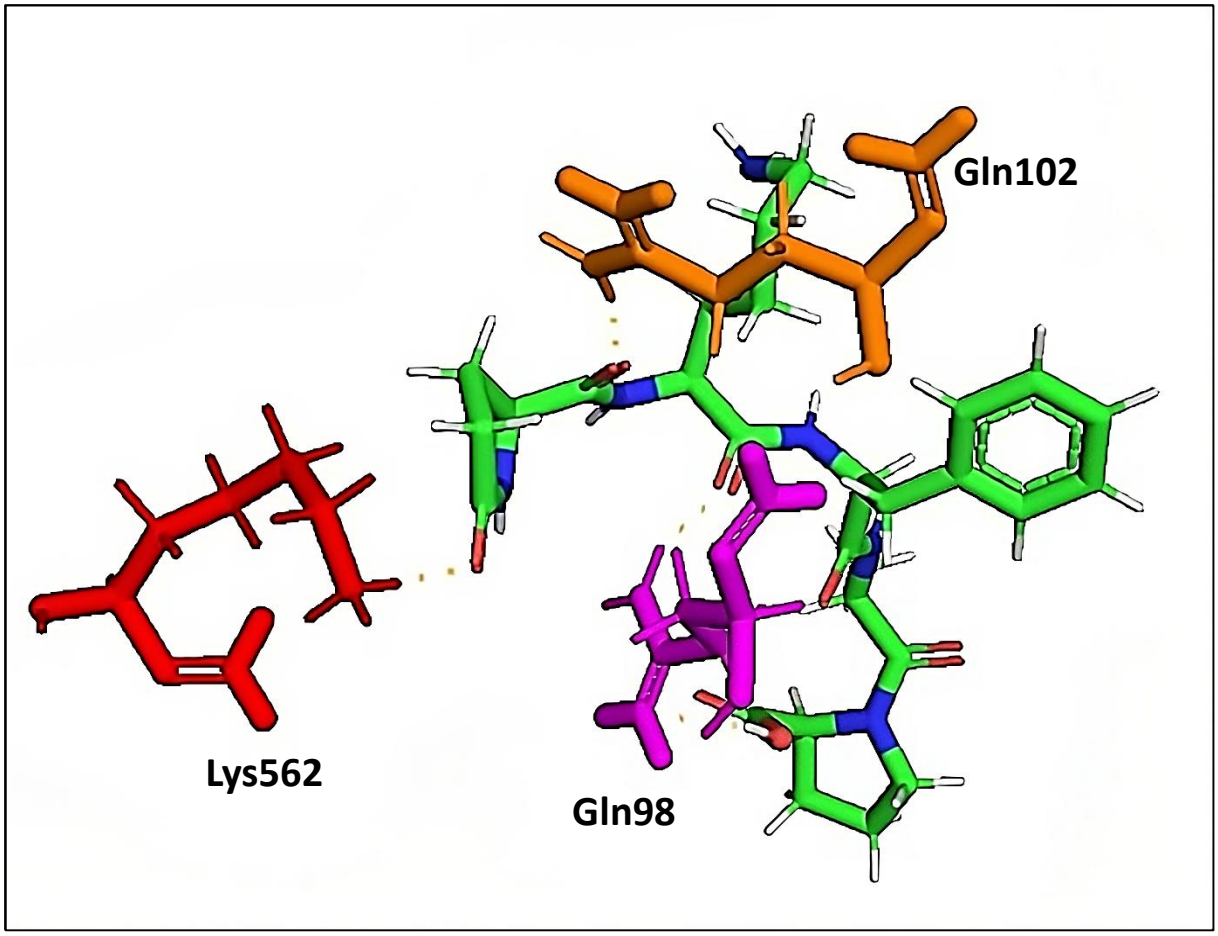

Figure 4.2: Second best pose conformation of ligand with ChEMBL id (CHEMBL10521) having pIC_{50} value of 8.52.

4.1.2 Molecular Docking of ACE:

In a detailed docking study aimed at investigating the interaction of ACE inhibitors with the protein binding site, a strategic approach was employed to identify potential binding efficiencies and stabilities. The binding site for ACE was precisely defined by its coordinates, set at X=2.9413, Y=-19.6702, and Z=-20.8413. This site includes critical residues that are integral to the ACE functionality, including His491, Ala334, Gln259, Lys489, His331, Phe435, Tyr501, His365, His361, Ser333, Phe490, Glu389, Arg500, Gly382, Pro385, and Thr496. These residues have been previously identified in the literature as key components of the ACE binding site which notably interacts with a zinc ion central to its catalytic mechanism.

For the docking, each of the 591 ACE inhibitors were docked into ACE by generating 10 distinct poses, utilizing ten Genetic Algorithm (GA) runs per ligand to ensure a thorough exploration of potential binding configurations. This methodical approach allowed for an extensive evaluation of how each inhibitor could engage with the binding site under various theoretical orientations.

The docking results were notably positive with the highest GOLD fitness score recorded at 83 that indicates an exceptionally favorable interaction between an inhibitor and the ACE binding site as shown in **Figure 4.3**. All generated poses across all ligands achieved positive GOLD fitness scores. This consistent result highlights not only the potential efficacy of these inhibitors in targeting the ACE enzyme but also underscores the robustness and relevance of the chosen binding site. The high GOLD fitness scores suggest that many of the studied ACE inhibitors have the potential to effectively interact with and potentially inhibit ACE activity which is crucial for developing therapeutic strategies against cardiovascular diseases that involve the renin-angiotensin system.

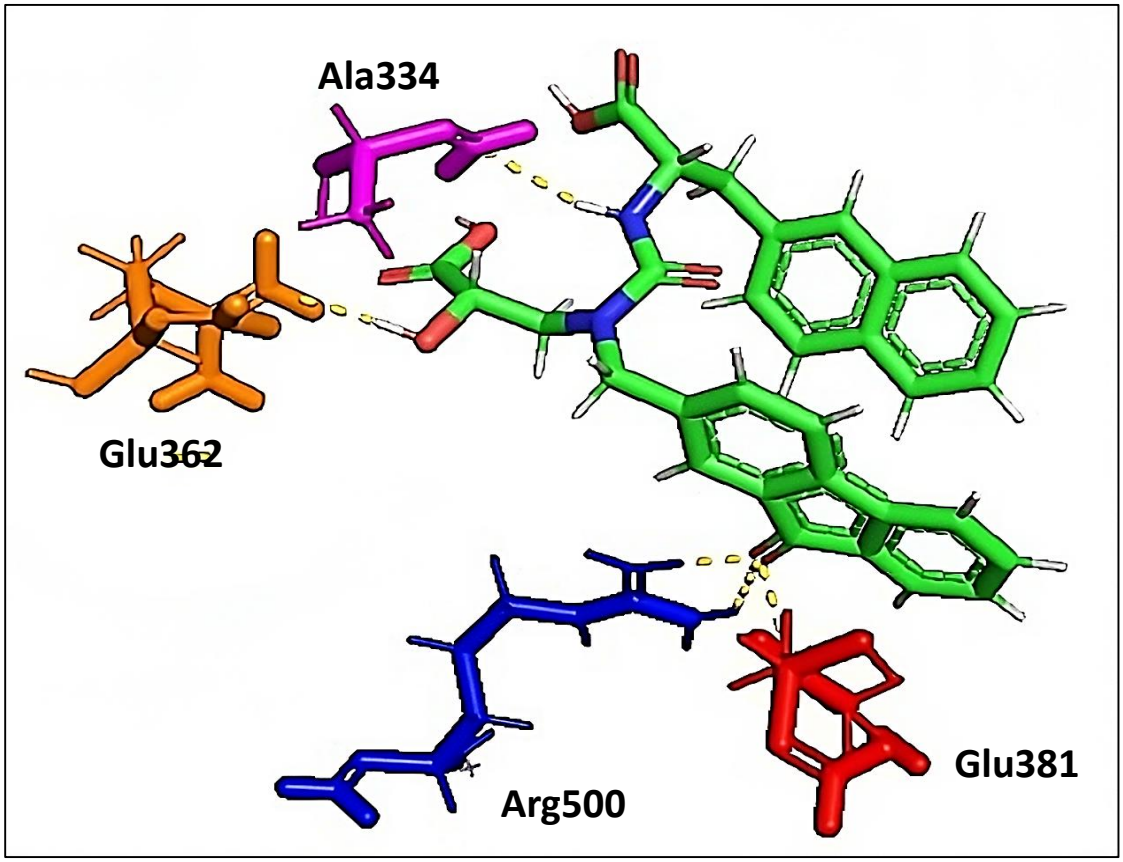

Figure 4.3: Best pose conformation of ligand with ChEMBL id (CHEMBL1233799) having pIC_{50} value of 7.95.

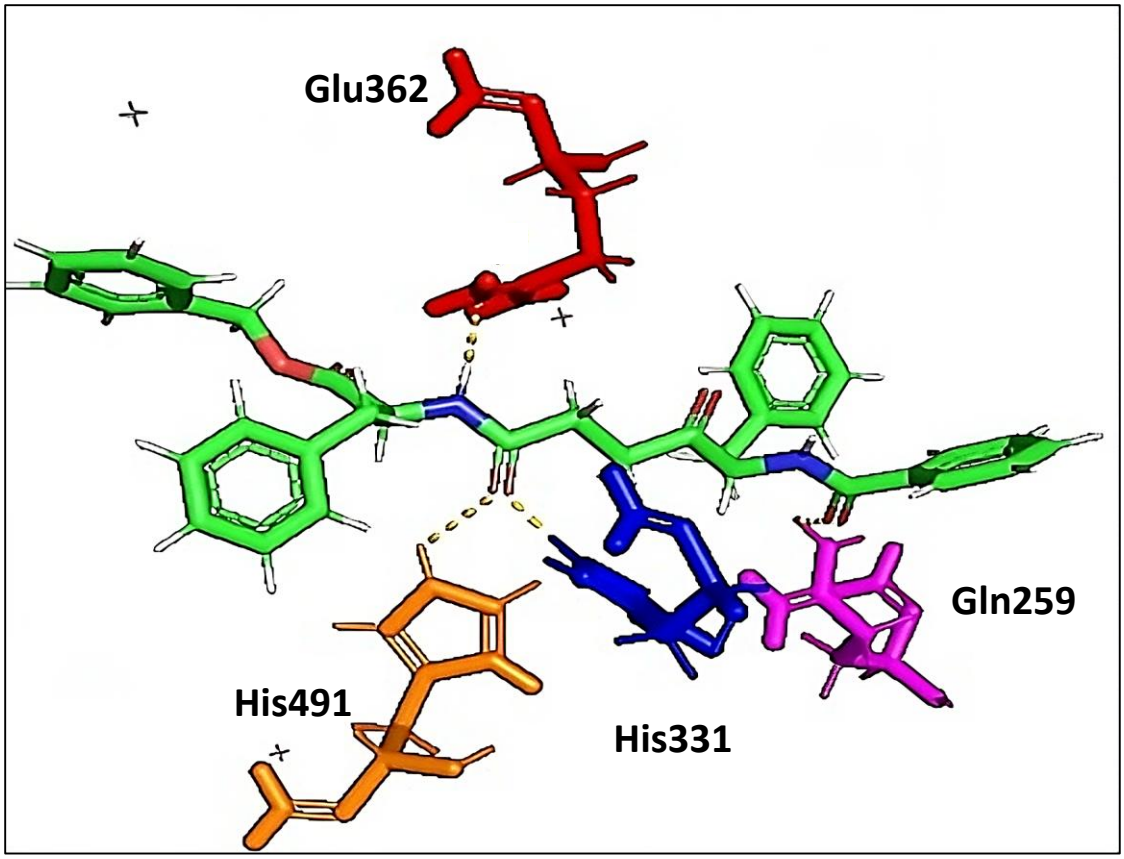

Figure 4.4: Second best pose conformation of ligand with ChEMBL id (CHEMBL418959) having pIC_{50} value of 6.383.

4.2 PLIF Analysis:

The PLIF analysis shows the frequency of interactions between 591 ligands and certain ACE protein residues. Comprehensive PLIF analysis was conducted to elucidate the interaction dynamics between a set of ligands and protein residues. The analysis processed ligands across binding residues of the protein that captures detailed interaction data such as hydrogen bonds, hydrophobic interactions, and ionic bonds. The results of these interactions were compiled into an Excel file that facilitates a structured exploration of interaction patterns.

4.2.1 ACE PLIF Analysis:

Several important residues that show notable interaction frequencies with the ligands have been identified, indicating their crucial involvement in ligand binding and the enzymatic activity of ACE. Arg500 stands out prominently with the highest interaction frequency at 51.4% that indicates that more than half of the ligands interact with this residue. The interactions involve both Van der Waals forces and hydrogen bonding, showcasing Arg500's pivotal role in stabilizing ligand binding through multiple non-covalent interactions. Arg500 constitutes sidechain hydrogen bond acceptor, hydrophobic, and arene interactions. Asp336, His331, and other residues also show notable interactions including Van der Wall contact, backbone acceptor, sidechain acceptor, and arene interactions. Trp335 forms hydrogen backbone acceptor and arene interactions as shown in **Figure 4.5**. These interactions might contribute to forming a binding pocket that is structurally and electrostatically complementary to a wide variety of ligand structures.

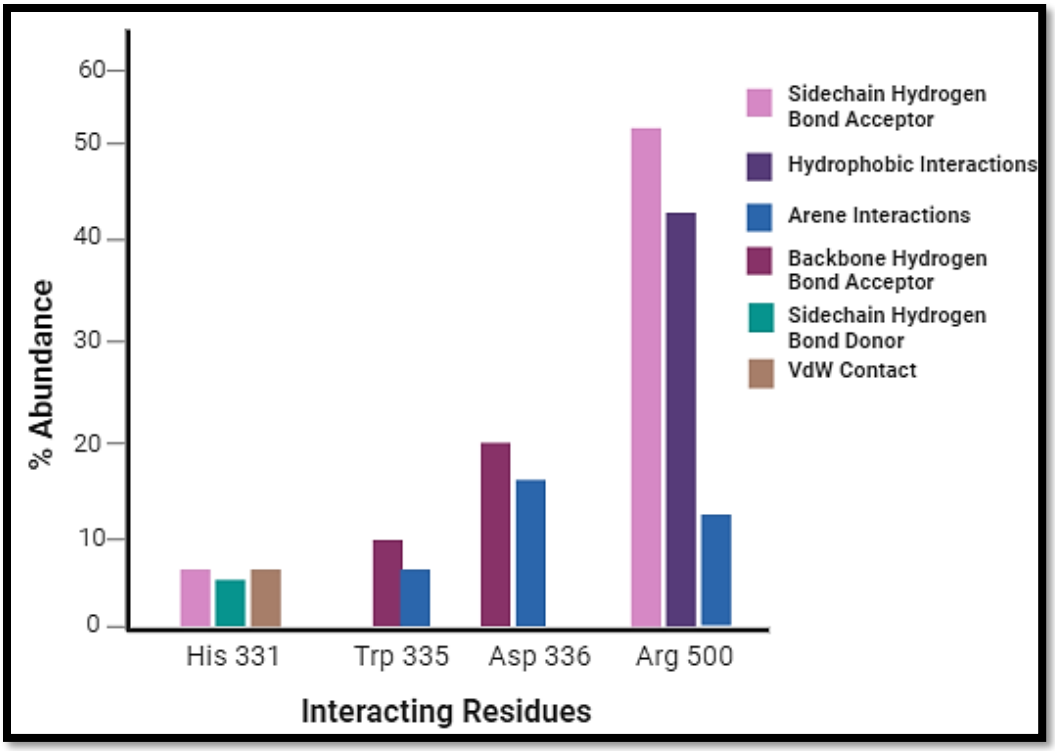

Figure 4.5: PLIF Analysis of ACE. Arg500 constitutes the highest frequency of interactions that includes sidechain acceptor, hydrophobic and arene interactions.

4.2.2 ACE2 PLIF Analysis:

Gln98 and His195 showed remarkably high contact rates compared to other residues. This implies that these residues play a crucial role in the recognition and binding of ligands that possibly enhances the stability and specificity of ligand-ACE2 complexes. Gln98 binds to ligands through sidechain donor, sidechain acceptor, Van der Waals forces and arene interactions while His195 binds to ligands through backbone donor interactions and Van der Waals forces. The other significant residues are Leu85 and Asn210 that show sidechain acceptor, backbone acceptor, VdW contact and arene interactions as mentioned in **Figure 4.6**.

The Protein-Ligand Interaction Fingerprint (PLIF) analysis conducted on ACE2 has revealed significant binding interactions between various ligands and specific residues within the ACE2 binding site. These interactions are not only crucial for understanding the molecular dynamics of ligand binding but also hold therapeutic potential. The PLIF analysis identified that several ligands, traditionally classified as ACE inhibitors, have binding affinities to critical residues within the ACE2 binding site.

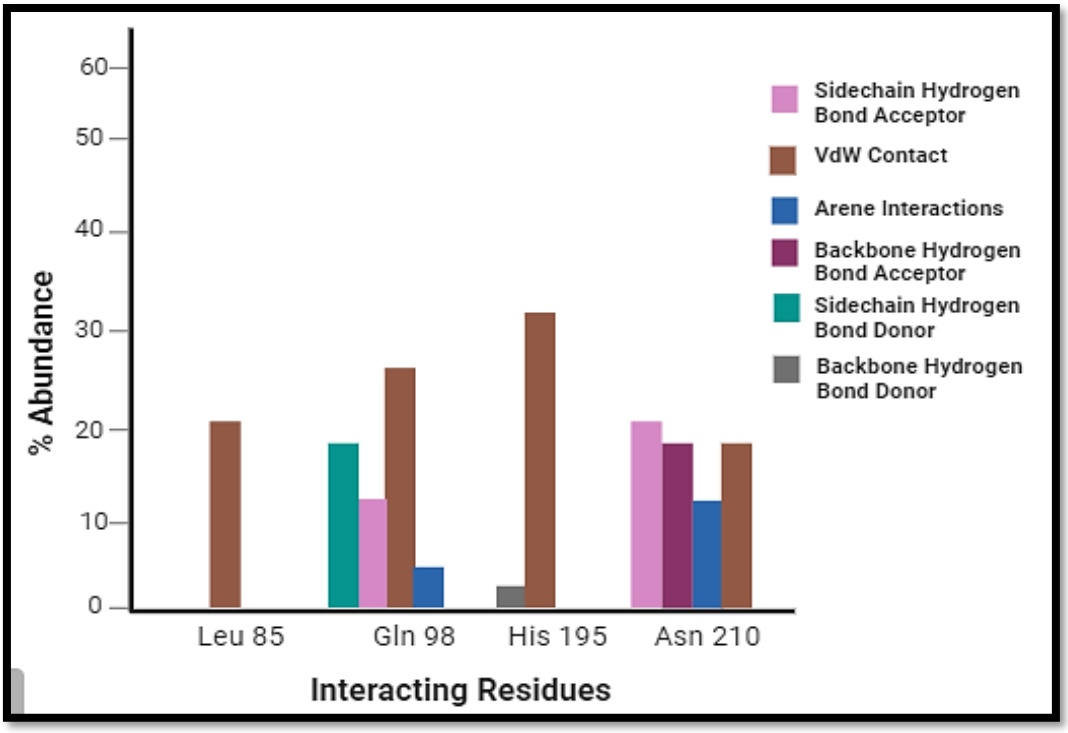

Figure 4.6: PLIF Analysis of ACE2. His195 constitutes the highest frequency of interactions that includes backbone donor and Van der Wall contacts.

4.3 Machine Learning Model:

4.3.1 Artificial Neural Network (ANN):

The Artificial Neural Network (ANN) model was developed and evaluated throughout a series of 50 training epochs. The results demonstrated a steady increase in model accuracy and a drop in loss values, indicating the model's ability to adjust and learn from the training data. Initially, the model started with an accuracy of 76.4% and a loss rate of 0.5298. These results significantly improved as training went on with the model lowering its loss to a negligible 0.0278 and attaining its maximum accuracy of 98.69% by the 49th epoch. This consistent improvement shows that the model's learning algorithm is both effective and capable of continuously improving its predictions. When tested against a test set, the model provided 90.69% overall accuracy as mentioned in **Table 4.1**. The performance metrics were computed for two classes that reveals insights into the model's

predictive strengths and weaknesses. With regard to Class 0, the model showed a precision of 0.87 that means that 87% of the predictions were accurate and a recall of 1.00 that means that the model successfully and flawlessly detected every incident of Class 0. The resulting F1-score of 0.93 for this class indicates that the model's predictions are very reliable and accurate.

On the other hand, the model's precision score of 1.00 for Class 1 indicates that it accurately identified positive instances with perfect accuracy. A gap in the model's capacity to recognize all real positive instances was highlighted by the recall for Class 1 which was comparatively lower at 0.76 and missed almost 24% of actual positives. An F1-score of 0.86 was obtained as a result that indicates that although the predictions are very accurate but the model's sensitivity to Class 1 might be increased.

Figure 4.7: Confusion Matrix of ANN.

The difference in recall between the two classes highlights a critical area that needs more fine-tuning. A more balanced and efficient model might result from increasing the model's sensitivity especially for Class 1 without sacrificing the high levels of precision attained. ANN model has shown promising capabilities in discriminating between ACE inhibitors (Class 0) and ACE2 activators (Class 1) with high accuracy and precision. The test evaluations demonstrated the accuracy of the model in recognizing ACE inhibitors which attains an almost perfect recall. However, the lower recall for ACE2 activators points to a critical area for further refinement. Enhancing the model's sensitivity to ACE2 activators without compromising the high precision observed for ACE inhibitors could lead to a more balanced and effective predictive tool.

4.3.2 Support Vector Machine (SVM)s:

To categorize things as either ACE inhibitors (Class 0) or ACE2 activators (Class 1), the SVM model underwent extensive training and testing on a dataset that had been preprocessed for best results. The first stage was to use mean imputation to fill in the missing values and polynomial transformation to enlarge the feature space to better capture non-linear interactions. The dataset showed no constant features or duplicates that indicates strong data quality and variability necessary for reliable machine learning model performance. Using GridSearchCV, the model's hyperparameters were carefully adjusted with gamma, kernel type, and penalty parameter C being the main areas of optimization. The optimal model employed a linear kernel with a regularization value of C=1 and ' scale' for gamma indicating that a relatively simple model without complex transformations provided the best results. This configuration achieved exceptionally high metrics on the test set that reflects the model's ability to accurately classify and distinguish between the two classes. Extensive performance measures demonstrated that the accuracy on the test set was roughly 99.58% with precision, recall, and F1-score all closely matching this value as mentioned

in **Table 4.1**. These high values indicate that the model performed remarkably well in detecting ACE inhibitors and ACE2 activators without significant bias or error. With an average accuracy of 99.66% and little variance, the cross-validation results supported these findings and further demonstrated the stability and generalizability of the model across various data subsets. The balanced class weights used in the SVM helped mitigate any potential bias due to this imbalance that contributes to the model's high performance across metrics.

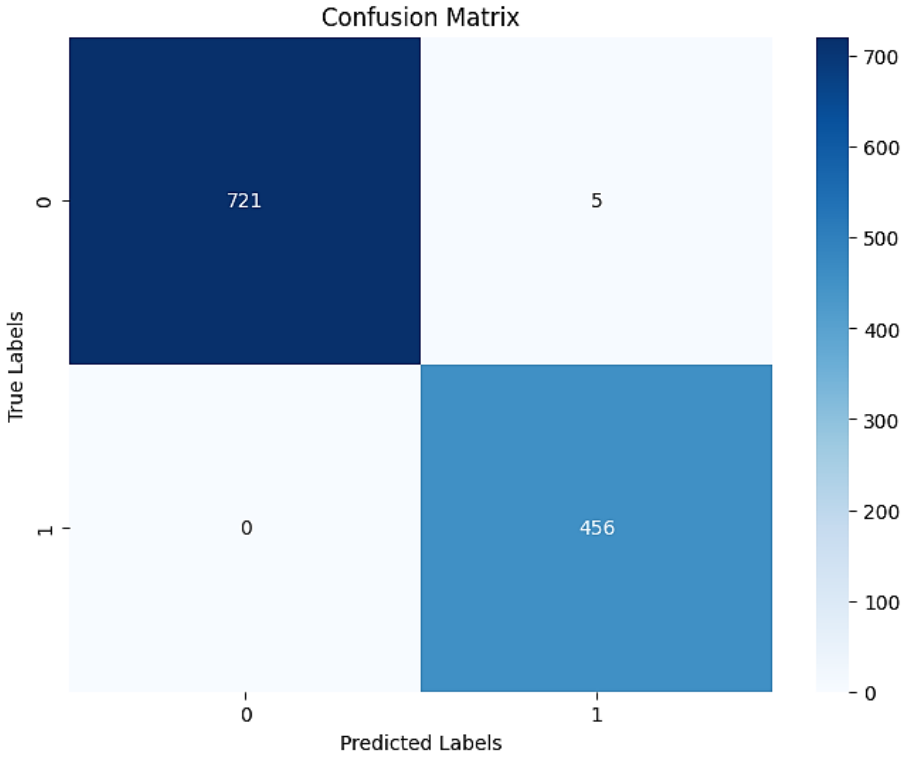

Figure 4.8: Confusion matrix of SVM.

The SVM model demonstrated outstanding capabilities in classifying compounds as ACE inhibitors or ACE2 activators with high reliability and accuracy. These results confirm the effectiveness of the selected machine learning techniques and model parameters and suggest the model's potential applicability in clinical or pharmacological settings where precise classification of such compounds is crucial

Table 4.1: Performance matrix of SVM and ANN.

Models	**Accuracy**	**Precision**	**F1-Score**	**Recall**
SVM	99.57%	99.58%	99.57%	99.57%
ANN	90.69%	91.91%	90.38%	90.6%

Chapter 5 Discussion:

PAH is a serious disorder characterized by elevated blood pressure in the pulmonary arteries leading to heart failure and reduced survival rate. An imbalance in the RAAS is the major cause of the disorder. Specifically, the hyperactivity of the ACE raises the levels of Ang II, which leads to vascular damage and vasoconstriction. By converting Ang II into vaso-protective peptides, the ACE2 counteracts these effects, indicating that increasing ACE2 activity may be a novel treatment strategy. ACE2 counterbalances the detrimental effects of Ang II by converting into Ang (1-7) and Ang I into Ang (1-9). Traditionally, ACE inhibitors are used to manage PAH by preventing the conversion of Ang I to Ang II that results in vasoconstriction and hypertension. However, these inhibitors alone are not fully effective as Ang II can also be produced through alternative, non-ACE pathways. ACE inhibitors have the potential to activate ACE2 that offers a more comprehensive approach to manage PAH. As no chemical data of ACE2 activators has yet been discovered, therefore, ACE inhibitors' data has been used to explore their potential to activate ACE2 for the development of new drugs. 835 ACE inhibitors from the ChEMBL database were downloaded. After the preprocessing, including Principal Component Analysis (PCA) to reduce data complexity, the data was shortlisted to 591 inhibitors based on their biochemical properties. All the duplicates and missing values were removed.

To investigate the binding potential of ACE inhibitors with ACE and ACE2, docking was performed. Each of the ACE inhibitors was docked into the active sites of ACE and ACE2 generating 10 poses to assess their binding affinities and interaction patterns, using the GOLD software. For ACE2, the docking was performed by selecting binding site with key residues. These residues are located around the hinge region, important for activator interaction. The highest GOLD score observed among all poses was 93 which indicates a particularly favorable interaction between the ligand and the ACE2 binding site. The highest GOLD score for ACE docking was 83 that suggests effective inhibition. These docking studies provide a deeper understanding of how ACE inhibitors can potentially interact with and activate ACE2, while also effectively inhibiting ACE.

Following the docking studies, PLIF analysis has been performed to further detail the interactions between the ACE inhibitors and the active sites of both ACE and ACE2. This analysis was important for understanding the specific molecular interactions. Using the prolif library, the PLIF analysis identified various types of interactions such as hydrogen bonds, hydrophobic contacts, and ionic bonds between the ligands and specific amino acids within the enzyme binding sites. Each interaction was captured in a binary format that indicates the presence or absence of a specific type of interaction within each ligand-protein complex. The binary interaction profiles generated from this PLIF analysis provided ligand-binding patterns. By understanding these interaction patterns, the analysis contributed to a deeper understanding of the potential for ACE inhibitors to interact favorably with ACE2 that suggests pathways through which these inhibitors might exert therapeutic effects beyond their traditional roles.

On the interaction data provided by PLIF analysis, two advanced machine learning models, an Artificial Neural Network and a Support Vector Machine were developed to classify compounds as either ACE inhibitors or potential ACE2 activators based on their specific interaction patterns with ACE and ACE2. The ANN model provided a good performance with an accuracy of 91%. It was particularly effective in identifying ACE inhibitors, achieving a high precision of 0.87 which indicates that 87% of the model's predictions for ACE inhibitors were correct. Furthermore, the recall for the class 0 was 1.00 which means that it identified all true instances of ACE inhibitors without any false negatives. The resulting F1-score of 0.93 for this class reflects the model's reliability and accuracy that indicates a balanced combination of precision and recall. On the other hand, the SVM model outperformed the ANN model in overall accuracy and precision. It achieved a remarkable accuracy of nearly 99.58% on the test set, with precision, recall, and F1-score values closely aligning with this high performance. These metrics not only indicate the SVM's ability to correctly identify and classify both ACE inhibitors and potential ACE2 activators but also highlight its effectiveness in handling the binary interaction data from the PLIF analysis without significant bias or error. Both models demonstrated significant capabilities in predicting the functionality of ACE inhibitors. The comparative accuracy of the SVM model demonstrated that it can handle non-linear relationships more effectively. The results from this extensive study strongly support the idea that ACE inhibitors have a dual role as they do not only inhibit ACE but they also enhance the activity of ACE2. This dual action could lead to significant improvements in treating PAH by better regulating RAAS. Furthermore, these insights pave the way for developing new drugs that specifically use these mechanisms potentially leading to more effective and personalized treatments for those suffering from PAH.

CONCLUSION:

In this research, we have demonstrated the dual functionality of ACE inhibitors such as ligands with ChEMBL ids, CHEMBL273140 and CHEMBL10521, showing that they can function both as ACE inhibitors and ACE2 activators. These findings pave the way to new therapeutics for the treatment of PAH which offers an approach to manage this complex condition by more effectively modulating the Renin-Angiotensin-Aldosterone System. Molecular docking was performed to examine interactions at molecular levels, detailed PLIF analysis to understand these interactions, and machine learning models to predict and classify the dual behavior of the compounds. The SVM model differentiated the ACE inhibitors and ACE2 activators with 99.57% accuracy, showing the potential of ACE inhibitors to activate ACE2. This research has laid a solid foundation for future studies that can aid value in developing new more effective treatments for patients suffering from diseases associated with RAAS dysregulation like PAH by illustrating that ACE inhibitors have the potential to activate ACE2 as well.

REFERENCES:

[1] M. M. Hoeper, H.-A. Ghofrani, E. Grünig, H. Klose, H. Olschewski, and S. Rosenkranz, "Pulmonary Hypertension," *Dtsch. Arzteblatt Int.*, vol. 114, no. 5, pp. 73–84, Feb. 2017, doi: 10.3238/arztebl.2017.0073.

[2] "Haemodynamic definitions and updated clinical classification of pulmonary hypertension | European Respiratory Society." Accessed: Jul. 02, 2024. [Online]. Available: https://erj.ersjournals.com/content/53/1/1801913

[3] N. Galiè *et al.*, "2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT)," *Eur. Heart J.*, vol. 37, no. 1, pp. 67–119, Jan. 2016, doi: 10.1093/eurheartj/ehv317.

[4] Rishi, "WHO Pulmonary Hypertension (PH) Groups," RK.MD. Accessed: Jul. 24, 2024. [Online]. Available: https://rk.md/2020/who-pulmonary-hypertension-ph-groups/

[5] E. M. T. Lau, E. Giannoulatou, D. S. Celermajer, and M. Humbert, "Epidemiology and treatment of pulmonary arterial hypertension," *Nat. Rev. Cardiol.*, vol. 14, no. 10, pp. 603–614, Oct. 2017, doi: 10.1038/nrcardio.2017.84.

[6] E. Spiekerkoetter, S. M. Kawut, and V. A. de Jesus Perez, "New and Emerging Therapies for Pulmonary Arterial Hypertension," *Annu. Rev. Med.*, vol. 70, pp. 45–59, Jan. 2019, doi: 10.1146/annurev-med-041717-085955.

[7] M. Humbert *et al.*, "Pulmonary arterial hypertension in France: results from a national registry," *Am. J. Respir. Crit. Care Med.*, vol. 173, no. 9, pp. 1023–1030, May 2006, doi: 10.1164/rccm.200510-1668OC.

[8] L. Southgate, R. D. Machado, S. Gräf, and N. W. Morrell, "Molecular genetic framework underlying pulmonary arterial hypertension," *Nat. Rev. Cardiol.*, vol. 17, no. 2, pp. 85–95, Feb. 2020, doi: 10.1038/s41569-019-0242-x.

[9] Y. Hu, L. Chi, W. M. Kuebler, and N. M. Goldenberg, "Perivascular Inflammation in Pulmonary Arterial Hypertension," *Cells*, vol. 9, no. 11, p. 2338, Oct. 2020, doi: 10.3390/cells9112338.

[10] Y. Wang *et al.*, "The EYA3 tyrosine phosphatase activity promotes pulmonary vascular remodeling in pulmonary arterial hypertension," *Nat. Commun.*, vol. 10, no. 1, p. 4143, Sep. 2019, doi: 10.1038/s41467-019-12226-1.

[11] G. Jia, A. R. Aroor, M. A. Hill, and J. R. Sowers, "Role of Renin-Angiotensin-Aldosterone System Activation in Promoting Cardiovascular Fibrosis and Stiffness," *Hypertens. Dallas Tex 1979*, vol. 72, no. 3, pp. 537–548, Sep. 2018, doi: 10.1161/HYPERTENSIONAHA.118.11065.

[12] S. Patel, A. Rauf, H. Khan, and T. Abu-Izneid, "Renin-angiotensin-aldosterone (RAAS): The ubiquitous system for homeostasis and pathologies," *Biomed. Pharmacother. Biomedecine Pharmacother.*, vol. 94, pp. 317–325, Oct. 2017, doi: 10.1016/j.biopha.2017.07.091.

[13] U. C. Kopp, "Neural Control of Renin Secretion Rate," in *Neural Control of Renal Function*, Morgan & Claypool Life Sciences, 2011. Accessed: Jul. 02, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK57240/

[14] Z. Wu, M. G. Cappiello, B. B. Scott, Y. Bukhtiyarov, and G. M. McGeehan, “Purification and characterization of recombinant human renin for X-ray crystallization studies,” *BMC Biochem.*, vol. 9, p. 19, Jun. 2008, doi: 10.1186/1471-2091-9-19.
[15] J. Bouhnik *et al.*, “Production and characterization of human renin antibodies with region-oriented synthetic peptides,” *J. Biol. Chem.*, vol. 262, no. 6, pp. 2913–2918, Feb. 1987.
[16] H. Wanka *et al.*, “Anti-necrotic and cardioprotective effects of a cytosolic renin isoform under ischemia-related conditions,” *J. Mol. Med. Berl. Ger.*, vol. 94, no. 1, pp. 61–69, Jan. 2016, doi: 10.1007/s00109-015-1321-z.
[17] M. J. Brown, “Renin: friend or foe?,” *Heart Br. Card. Soc.*, vol. 93, no. 9, pp. 1026–1033, Sep. 2007, doi: 10.1136/hrt.2006.107706.
[18] P. Verdecchia, F. Angeli, G. Mazzotta, G. Gentile, and G. Reboldi, “The renin angiotensin system in the development of cardiovascular disease: role of aliskiren in risk reduction,” *Vasc. Health Risk Manag.*, vol. 4, no. 5, pp. 971–981, Oct. 2008.
[19] C. Heit *et al.*, “Update of the human and mouse SERPINgene superfamily,” *Hum. Genomics*, vol. 7, no. 1, p. 22, Oct. 2013, doi: 10.1186/1479-7364-7-22.
[20] K. E. Bernstein *et al.*, “Angiotensin converting enzyme over expression in myelocytes enhances the immune response,” *Biol. Chem.*, vol. 395, no. 10, pp. 1173–1178, Oct. 2014, doi: 10.1515/hsz-2013-0295.
[21] M. J. Peach and D. E. Dostal, “The Angiotensin II Receptor and the Actions of Angiotensin II,” *J. Cardiovasc. Pharmacol.*, vol. 16, p. S25, 1990.
[22] E. P. Feener, J. M. Northrup, L. P. Aiello, and G. L. King, “Angiotensin II induces plasminogen activator inhibitor-1 and -2 expression in vascular endothelial and smooth muscle cells.,” *J. Clin. Invest.*, vol. 95, no. 3, pp. 1353–1362, Mar. 1995.
[23] J. Yatabe *et al.*, “Angiotensin III stimulates aldosterone secretion from adrenal gland partially via angiotensin II type 2 receptor but not angiotensin II type 1 receptor,” *Endocrinology*, vol. 152, no. 4, pp. 1582–1588, Apr. 2011, doi: 10.1210/en.2010-1070.
[24] A. Chiolero, M. Maillard, J. Nussberger, H.-R. Brunner, and M. Burnier, “Proximal Sodium Reabsorption,” *Hypertension*, vol. 36, no. 4, pp. 631–637, Oct. 2000, doi: 10.1161/01.HYP.36.4.631.
[25] A. E. Cuadra, Z. Shan, C. Sumners, and M. K. Raizada, “A Current View of Brain Renin-Angiotensin System: Is the (Pro)renin Receptor the Missing Link?,” *Pharmacol. Ther.*, vol. 125, no. 1, p. 27, Jan. 2010, doi: 10.1016/j.pharmthera.2009.07.007.
[26] M. Sahay and R. Sahay, “Hyponatremia: A practical approach,” *Indian J. Endocrinol. Metab.*, vol. 18, no. 6, pp. 760–771, 2014, doi: 10.4103/2230-8210.141320.
[27] J. M. Sands and H. E. Layton, “The Physiology of Urinary Concentration: an Update,” *Semin. Nephrol.*, vol. 29, no. 3, pp. 178–195, May 2009, doi: 10.1016/j.semnephrol.2009.03.008.
[28] C. T. Lim and B. Khoo, “Normal Physiology of ACTH and GH Release in the Hypothalamus and Anterior Pituitary in Man,” in *Endotext*, K. R. Feingold, B. Anawalt, M. R. Blackman, A. Boyce, G. Chrousos, E. Corpas, W. W. de Herder, K. Dhatariya, K. Dungan, J. Hofland, S. Kalra, G. Kaltsas, N. Kapoor, C. Koch, P. Kopp, M. Korbonits, C. S. Kovacs, W. Kuohung, B. Laferrère, M. Levy, E. A. McGee, R. McLachlan, M. New, J. Purnell, R. Sahay, A. S. Shah, F. Singer, M. A. Sperling, C. A. Stratakis, D. L. Trence, and D. P. Wilson, Eds., South Dartmouth (MA): MDText.com, Inc., 2000. Accessed: Jul. 03, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK279116/
[29] F. Moccia *et al.*, “COVID-19-associated cardiovascular morbidity in older adults: a position paper from the Italian Society of Cardiovascular Researches,” *GeroScience*, vol. 42, no. 4, pp. 1021–1049, Aug. 2020, doi: 10.1007/s11357-020-00198-w.
[30] S. C. Simões, A. L. Balico-Silva, L. T. Parreiras-E-Silva, A. L. B. Bitencourt, M. Bouvier, and C. M. Costa-Neto, “Signal Transduction Profiling of Angiotensin II Type 1 Receptor With Mutations

Associated to Atrial Fibrillation in Humans," *Front. Pharmacol.*, vol. 11, p. 600132, 2020, doi: 10.3389/fphar.2020.600132.
[31] P. K. Mehta and K. K. Griendling, "Angiotensin II cell signaling: physiological and pathological effects in the cardiovascular system," *Am. J. Physiol. Cell Physiol.*, vol. 292, no. 1, pp. C82-97, Jan. 2007, doi: 10.1152/ajpcell.00287.2006.
[32] M. A. Sparks, S. D. Crowley, S. B. Gurley, M. Mirotsou, and T. M. Coffman, "Classical Renin-Angiotensin System in Kidney Physiology," *Compr. Physiol.*, vol. 4, no. 3, pp. 1201–1228, Jul. 2014, doi: 10.1002/cphy.c130040.
[33] K. Husain, W. Hernandez, R. A. Ansari, and L. Ferder, "Inflammation, oxidative stress and renin angiotensin system in atherosclerosis," *World J. Biol. Chem.*, vol. 6, no. 3, pp. 209–217, Aug. 2015, doi: 10.4331/wjbc.v6.i3.209.
[34] W. S. D. Tan, W. Liao, S. Zhou, D. Mei, and W.-S. F. Wong, "Targeting the renin-angiotensin system as novel therapeutic strategy for pulmonary diseases," *Curr. Opin. Pharmacol.*, vol. 40, pp. 9–17, Jun. 2018, doi: 10.1016/j.coph.2017.12.002.
[35] S. Y. Chai *et al.*, "What's new in the renin-angiotensin system?," *Cell. Mol. Life Sci. CMLS*, vol. 61, no. 21, pp. 2728–2737, Nov. 2004, doi: 10.1007/s00018-004-4246-1.
[36] T. Watanabe, T. A. Barker, and B. C. Berk, "Angiotensin II and the endothelium: diverse signals and effects," *Hypertens. Dallas Tex 1979*, vol. 45, no. 2, pp. 163–169, Feb. 2005, doi: 10.1161/01.HYP.0000153321.13792.b9.
[37] S. Bernardi, A. Michelli, G. Zuolo, R. Candido, and B. Fabris, "Update on RAAS Modulation for the Treatment of Diabetic Cardiovascular Disease," *J. Diabetes Res.*, vol. 2016, no. 1, p. 8917578, 2016, doi: 10.1155/2016/8917578.
[38] I. Hamming *et al.*, "The emerging role of ACE2 in physiology and disease," *J. Pathol.*, vol. 212, no. 1, pp. 1–11, May 2007, doi: 10.1002/path.2162.
[39] M. A. Zaman, S. Oparil, and D. A. Calhoun, "Drugs targeting the renin-angiotensin-aldosterone system," *Nat. Rev. Drug Discov.*, vol. 1, no. 8, pp. 621–636, Aug. 2002, doi: 10.1038/nrd873.
[40] P. Towler *et al.*, "ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis," *J. Biol. Chem.*, vol. 279, no. 17, pp. 17996–18007, Apr. 2004, doi: 10.1074/jbc.M311191200.
[41] A. Bhattarai, S. Pawnikar, and Y. Miao, "Mechanism of Ligand Recognition by Human ACE2 Receptor," *J. Phys. Chem. Lett.*, vol. 12, no. 20, pp. 4814–4822, May 2021, doi: 10.1021/acs.jpclett.1c01064.
[42] D. W. Lambert *et al.*, "Tumor necrosis factor-alpha convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2)," *J. Biol. Chem.*, vol. 280, no. 34, pp. 30113–30119, Aug. 2005, doi: 10.1074/jbc.M505111200.
[43] M.-Y. Li, L. Li, Y. Zhang, and X.-S. Wang, "Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues," *Infect. Dis. Poverty*, vol. 9, no. 1, p. 45, Apr. 2020, doi: 10.1186/s40249-020-00662-x.
[44] A. J. Turner, J. A. Hiscox, and N. M. Hooper, "ACE2: from vasopeptidase to SARS virus receptor," *Trends Pharmacol. Sci.*, vol. 25, no. 6, pp. 291–294, Jun. 2004, doi: 10.1016/j.tips.2004.04.001.
[45] J. Zepeda-Cervantes *et al.*, "Implications of the Immune Polymorphisms of the Host and the Genetic Variability of SARS-CoV-2 in the Development of COVID-19," *Viruses*, vol. 14, Jan. 2022, doi: 10.3390/v14010094.
[46] "Pathology and pathobiology of pulmonary hypertension: state of the art and research perspectives | European Respiratory Society." Accessed: Jul. 03, 2024. [Online]. Available: https://erj.ersjournals.com/content/53/1/1801887

[47] C. Tikellis and M. C. Thomas, “Angiotensin-Converting Enzyme 2 (ACE2) Is a Key Modulator of the Renin Angiotensin System in Health and Disease,” *Int. J. Pept.*, vol. 2012, p. 256294, 2012, doi: 10.1155/2012/256294.
[48] P. J. Mulrow, “Angiotensin II and aldosterone regulation,” *Regul. Pept.*, vol. 80, no. 1–2, pp. 27–32, Mar. 1999, doi: 10.1016/s0167-0115(99)00004-x.
[49] R. M. Touyz, “Reactive oxygen species in vascular biology: role in arterial hypertension,” *Expert Rev. Cardiovasc. Ther.*, vol. 1, no. 1, pp. 91–106, May 2003, doi: 10.1586/14779072.1.1.91.
[50] S. Usui *et al.*, “NADPH oxidase plays a central role in cone cell death in retinitis pigmentosa,” *J. Neurochem.*, vol. 110, no. 3, pp. 1028–1037, Aug. 2009, doi: 10.1111/j.1471-4159.2009.06195.x.
[51] L.-N. Qiao, H.-B. Xu, K. Shi, T.-F. Zhou, Y.-M. Hua, and H.-M. Liu, “Role of notch signal in angiotensin II induced pulmonary vascular remodeling,” *Transl. Pediatr.*, vol. 2, no. 1, Art. no. 1, Jan. 2013, doi: 10.3978/j.issn.2224-4336.2012.05.03.
[52] “Angiotensin-converting enzyme 2 protects from severe acute lung failure | Nature.” Accessed: Jul. 03, 2024. [Online]. Available: https://www.nature.com/articles/nature03712
[53] X. Zhu *et al.*, “Linkage and association analysis of angiotensin I-converting enzyme (ACE)-gene polymorphisms with ACE concentration and blood pressure,” *Am. J. Hum. Genet.*, vol. 68, no. 5, pp. 1139–1148, May 2001, doi: 10.1086/320104.
[54] R. Zhang *et al.*, “Role of HIF-1alpha in the regulation ACE and ACE2 expression in hypoxic human pulmonary artery smooth muscle cells,” *Am. J. Physiol. Lung Cell. Mol. Physiol.*, vol. 297, no. 4, pp. L631-640, Oct. 2009, doi: 10.1152/ajplung.90415.2008.
[55] K. Yamamoto *et al.*, “Deletion of angiotensin-converting enzyme 2 accelerates pressure overload-induced cardiac dysfunction by increasing local angiotensin II,” *Hypertens. Dallas Tex 1979*, vol. 47, no. 4, pp. 718–726, Apr. 2006, doi: 10.1161/01.HYP.0000205833.89478.5b.
[56] G. Y. Oudit, M. A. Crackower, P. H. Backx, and J. M. Penninger, “The role of ACE2 in cardiovascular physiology,” *Trends Cardiovasc. Med.*, vol. 13, no. 3, pp. 93–101, Apr. 2003, doi: 10.1016/s1050-1738(02)00233-5.
[57] G. I. Rice, A. L. Jones, P. J. Grant, A. M. Carter, A. J. Turner, and N. M. Hooper, “Circulating activities of angiotensin-converting enzyme, its homolog, angiotensin-converting enzyme 2, and neprilysin in a family study,” *Hypertens. Dallas Tex 1979*, vol. 48, no. 5, pp. 914–920, Nov. 2006, doi: 10.1161/01.HYP.0000244543.91937.79.
[58] J. Wysocki *et al.*, “ACE and ACE2 activity in diabetic mice,” *Diabetes*, vol. 55, no. 7, pp. 2132–2139, Jul. 2006, doi: 10.2337/db06-0033.
[59] M. de Gasparo, K. J. Catt, T. Inagami, J. W. Wright, and T. Unger, “International union of pharmacology. XXIII. The angiotensin II receptors,” *Pharmacol. Rev.*, vol. 52, no. 3, pp. 415–472, Sep. 2000.
[60] S. J. Forrester *et al.*, “Angiotensin II Signal Transduction: An Update on Mechanisms of Physiology and Pathophysiology,” *Physiol. Rev.*, vol. 98, no. 3, pp. 1627–1738, Jul. 2018, doi: 10.1152/physrev.00038.2017.
[61] M. Burnier and H. R. Brunner, “Angiotensin II receptor antagonists,” *Lancet Lond. Engl.*, vol. 355, no. 9204, pp. 637–645, Feb. 2000, doi: 10.1016/s0140-6736(99)10365-9.
[62] R. M. Carey and H. M. Siragy, “Newly recognized components of the renin-angiotensin system: potential roles in cardiovascular and renal regulation,” *Endocr. Rev.*, vol. 24, no. 3, pp. 261–271, Jun. 2003, doi: 10.1210/er.2003-0001.
[63] T. Unger and B. Dahlöf, “Compound 21, the first orally active, selective agonist of the angiotensin type 2 receptor (AT2): implications for AT2 receptor research and therapeutic potential,” *J. Renin Angiotensin Aldosterone Syst.*, vol. 11, no. 1, pp. 75–77, Mar. 2010, doi: 10.1177/1470320309347792.

[64] “Recent Advances in Angiotensin Research | Hypertension.” Accessed: Jul. 03, 2024. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/hypertensionaha.117.08931
[65] “International Union of Basic and Clinical Pharmacology. XCIX. Angiotensin Receptors: Interpreters of Pathophysiological Angiotensinergic Stimuli [corrected] - PubMed.” Accessed: Jul. 03, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26315714/
[66] A. L. Albiston *et al.*, “Evidence that the angiotensin IV (AT(4)) receptor is the enzyme insulin-regulated aminopeptidase,” *J. Biol. Chem.*, vol. 276, no. 52, pp. 48623–48626, Dec. 2001, doi: 10.1074/jbc.C100512200.
[67] S. Y. Chai *et al.*, “Distribution of angiotensin IV binding sites (AT4 receptor) in the human forebrain, midbrain and pons as visualised by in vitro receptor autoradiography,” *J. Chem. Neuroanat.*, vol. 20, no. 3–4, pp. 339–348, Dec. 2000, doi: 10.1016/s0891-0618(00)00112-5.
[68] D. Batlle, J. Wysocki, and K. Satchell, “Soluble angiotensin-converting enzyme 2: a potential approach for coronavirus infection therapy?,” *Clin. Sci. Lond. Engl. 1979*, vol. 134, no. 5, pp. 543–545, Mar. 2020, doi: 10.1042/CS20200163.
[69] V. B. Patel, J.-C. Zhong, M. B. Grant, and G. Y. Oudit, “Role of the ACE2/Angiotensin 1–7 axis of the Renin-Angiotensin System in Heart Failure,” *Circ. Res.*, vol. 118, no. 8, pp. 1313–1326, Apr. 2016, doi: 10.1161/CIRCRESAHA.116.307708.
[70] J. A. Hernández Prada *et al.*, “Structure-based identification of small-molecule angiotensin-converting enzyme 2 activators as novel antihypertensive agents,” *Hypertens. Dallas Tex 1979*, vol. 51, no. 5, pp. 1312–1317, May 2008, doi: 10.1161/HYPERTENSIONAHA.107.108944.
[71] S. B. Gurley and T. M. Coffman, “Angiotensin-converting enzyme 2 gene targeting studies in mice: mixed messages,” *Exp. Physiol.*, vol. 93, no. 5, pp. 538–542, 2008, doi: 10.1113/expphysiol.2007.040014.
[72] “Overexpression of ACE2 prevents hypoxia-induced pulmonary hypertension in rats by inhibiting proliferation and immigration of PASMCs”.
[73] L. Southgate, R. D. Machado, S. Gräf, and N. W. Morrell, “Molecular genetic framework underlying pulmonary arterial hypertension,” *Nat. Rev. Cardiol.*, vol. 17, no. 2, pp. 85–95, Feb. 2020, doi: 10.1038/s41569-019-0242-x.
[74] M. Humbert *et al.*, “Sotatercept for the treatment of pulmonary arterial hypertension: PULSAR open-label extension,” *Eur. Respir. J.*, vol. 61, no. 1, p. 2201347, Jan. 2023, doi: 10.1183/13993003.01347-2022.
[75] “Novel aspects of the renin-angiotensin system for pulmonary arterial hypertension | Hypertension Research.” Accessed: Jul. 03, 2024. [Online]. Available: https://www.nature.com/articles/s41440-023-01211-3
[76] C. M. Ferrario *et al.*, “Effect of angiotensin-converting enzyme inhibition and angiotensin II receptor blockers on cardiac angiotensin-converting enzyme 2,” *Circulation*, vol. 111, no. 20, pp. 2605–2610, May 2005, doi: 10.1161/CIRCULATIONAHA.104.510461.
[77] Y. Xiao, P.-P. Chen, R.-L. Zhou, Y. Zhang, Z. Tian, and S.-Y. Zhang, “Pathological Mechanisms and Potential Therapeutic Targets of Pulmonary Arterial Hypertension: A Review,” *Aging Dis.*, vol. 11, no. 6, pp. 1623–1639, Dec. 2020, doi: 10.14336/AD.2020.0111.
[78] “Novel RAAS agonists and antagonists: clinical applications and controversies | Nature Reviews Endocrinology.” Accessed: Jul. 24, 2024. [Online]. Available: https://www.nature.com/articles/nrendo.2015.6
[79] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, “Open Babel: An open chemical toolbox,” *J. Cheminformatics*, vol. 3, no. 1, p. 33, Oct. 2011, doi: 10.1186/1758-2946-3-33.

[80] “SARS-CoV-2 Spike receptor-binding domain with a G485R mutation in complex with human ACE2 | bioRxiv.” Accessed: Jul. 05, 2024. [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.03.16.434488v1
[81] G. E. Cozier, L. Lubbe, E. D. Sturrock, and K. R. Acharya, “ACE-domain selectivity extends beyond direct interacting residues at the active site,” *Biochem. J.*, vol. 477, no. 7, pp. 1241–1259, Apr. 2020, doi: 10.1042/BCJ20200060.
[82] M. K. Annamala, K. K. Inampudi, and L. Guruprasad, “Docking of phosphonate and trehalose analog inhibitors into M. tuberculosis mycolyltransferase Ag85C: Comparison of the two scoring fitness functions GoldScore and ChemScore, in the GOLD software,” *Bioinformation*, vol. 1, no. 9, pp. 339–350, Jan. 2007, doi: 10.6026/97320630001339.
[83] B. Fiorillo *et al.*, “Discovery of Bile Acid Derivatives as Potent ACE2 Activators by Virtual Screening and Essential Dynamics,” *J. Chem. Inf. Model.*, vol. 62, no. 1, pp. 196–209, Jan. 2022, doi: 10.1021/acs.jcim.1c01126.
[84] L. V. Kulemina and D. A. Ostrov, “Prediction of off-target effects on angiotensin-converting enzyme 2,” *J. Biomol. Screen.*, vol. 16, no. 8, pp. 878–885, Sep. 2011, doi: 10.1177/1087057111413919.
[85] “ACE-domain selectivity extends beyond direct interacting residues at the active site | Biochemical Journal | Portland Press.” Accessed: Jul. 08, 2024. [Online]. Available: https://portlandpress.com/biochemj/article/477/7/1241/222416/ACE-domain-selectivity-extends-beyond-direct
[86] C. Bouysset and S. Fiorucci, “ProLIF: a library to encode molecular interactions as fingerprints,” *J. Cheminformatics*, vol. 13, no. 1, p. 72, Sep. 2021, doi: 10.1186/s13321-021-00548-6.

APPENDICES:

Appendix 1: Smiles of all 591 ACE inhibitors data.

Molecule ChEMBL ID	Index	Smiles	IC50 (nM)
CHEMBL84966	1	NC1CCCCCN(CC(=O)O)C1=O	440000
CHEMBL128223	2	O=C(/C=C\S)C1CC[C@@H](C(=O)O)C1	650
CHEMBL306556	3	O=C(O)[C@@H]1C=CCN1C(=O)CCS	645.65
CHEMBL541207	4	CC(C)C[C@H](N)C(=O)N[C@@H](C)C(=O)O	309029.54
CHEMBL90670	5	NC(=O)CC[C@H](N)C(=O)NCC(=O)O	7413102.41
CHEMBL330038	6	NCC(=O)N[C@@H](CCC(N)=O)C(=O)O	7079457.84
CHEMBL91561	7	NCCCC[C@H](N)C(=O)NCC(=O)O	3235936.57
CHEMBL420061	8	NCCCC[C@H](NC(=O)CN)C(=O)O	5370317.96
CHEMBL305938	9	O=C(O)[C@H]1CCCN1C(=O)CCS	200
CHEMBL10962	10	O=C(O)[C@@H]1CCCN1C(=O)CCS	1800000
CHEMBL323051	11	O=C(O)C1CCC(=O)N1CCCS	1000000000
CHEMBL11707	12	O=C(O)CN1CCCC(CS)C1=O	1000
CHEMBL442750	13	O=C(O)[C@@H]1CCC(=O)N1CCCS	645654.23
CHEMBL421361	14	N[C@@H](CCC(=O)O)C(=O)NCC(=O)O	10000000
CHEMBL17503	15	C[C@H](NC(=O)[C@@H](N)CC(=O)O)C(=O)O	3801893.96
CHEMBL314532	16	NCC(=O)N[C@@H](CCC(=O)O)C(=O)O	5370317.96
CHEMBL78812	17	CC(C)[C@H](NC(=O)CCS)C(=O)O	1584.89
CHEMBL330552	18	CSCC[C@H](NC(=O)CN)C(=O)O	1412537.54

CHEMBL90972	19	CSCC[C@H](N)C(=O)NCC(=O)O	4786300.92
CHEMBL80242	20	O=C(O)C[C@H](NC(=O)CS)C(=O)O	77624.71
CHEMBL309914	21	CC(S)C(=O)N[C@@H](CS)C(=O)O	7943.28
CHEMBL309333	22	O=C(CCS)N[C@@H](CS)C(=O)O	67608.3
CHEMBL93127	23	NCC(=O)N[C@@H](Cc1cnc[nH]1)C(=O)O	3090295.43
CHEMBL90452	24	N[C@@H](Cc1c[nH]cn1)C(=O)NCC(=O)O	6309573.44
CHEMBL439762	25	CC(C)[C@H](N)C(=O)N1CCC[C@H]1C(=O)O	416869.38
CHEMBL102044	26	C[C@H](N)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O	10000
CHEMBL310578	27	O=C(O)CCC(=O)N1CCC[C@H]1C(=O)O	33113.11
CHEMBL310764	28	CC(C)[C@H](NC(=O)CCC(=O)O)C(=O)O	1096478.2
CHEMBL93178	29	C[C@H](NC(=O)[C@@H](N)CCCCN)C(=O)O	380189.4
CHEMBL1560	30	C[C@H](CS)C(=O)N1CCC[C@H]1C(=O)O	9
CHEMBL162360	31	C[C@H](CS)C(=O)N1CCCC1C(=O)O	23
CHEMBL76577	32	C[C@@H](CS)C(=O)N1CCC[C@@H]1C(=O)O	10964.78
CHEMBL90074	33	C[C@H](NC(=O)[C@@H](N)CCC(=O)O)C(=O)O	10000000
CHEMBL412666	34	NC(CS)C(=O)N1CCC[C@H]1C(=O)O	63.1
CHEMBL299889	35	NCC(=O)N[C@@H](Cc1ccccc1)C(=O)O	602559.59
CHEMBL80830	36	CC(C)(S)C(=O)N[C@@H](CS)C(=O)O	7244359.6
CHEMBL94119	37	CC[C@H](C)[C@H](N)C(=O)N1CCC[C@@H]1C(=O)O	128824.96
CHEMBL1807684	38	CC[C@H](C)[C@H](N)C(=O)N1CCC[C@H]1C(=O)O	128824.96
CHEMBL277919	39	O=C(O)CCCC(=O)N1CCC[C@H]1C(=O)O	70000
CHEMBL78805	40	C[C@H](CC(=O)O)C(=O)N1CCC[C@H]1C(=O)O	10232.93
CHEMBL311102	41	O=C(CS)N[C@@H](Cc1c[nH]cn1)C(=O)O	257039.58

CHEMBL78435	42	O=C(O)[C@@H]1CCCN1C(=O)C1CCC1S	48.98
CHEMBL106193	43	O=C(O)C1CCCN1C(=O)C1CCC1S	158.49
CHEMBL291381	44	O=C(CCC(=O)N1CCCC1C(=O)O)NO	50000
CHEMBL311988	45	CC(C(=O)O)C1CCCCC(CS)C1=O	52.48
CHEMBL309782	46	CC(C)C[C@H](NC(=O)CCC(=O)O)C(=O)O	616595
CHEMBL79008	47	CCC(C)[C@H](NC(=O)CCC(=O)O)C(=O)O	1047128.55
CHEMBL367315	48	O=c1cc(N2CCOCC2)oc2ccccc12	13400
CHEMBL96806	49	NCC(=O)N[C@@H](CCCN=C(N)N)C(=O)O	3235936.57
CHEMBL311525	50	C[C@H](CS)C(=O)N1C(=O)CC[C@H]1C(=O)O	8.913
CHEMBL11783	51	CC(=O)SCC1CCN(CC(=O)O)C1=O	33000
CHEMBL110375	52	CC(CS)C(=O)N1C(=O)CCC1C(=O)O	31.62
CHEMBL446407	53	CC(C)(S)C(=O)N[C@@H](CC(=O)O)C(=O)O	3467368.5
CHEMBL59354	54	CC(CS)C(=O)N1CCc2ccccc2C1	48000
CHEMBL78953	55	C[C@@H](N)C(=O)N[C@@H](Cc1ccccc1)C(=O)O	190546.07
CHEMBL57338	56	C[C@H](N)C(=O)N[C@@H](Cc1ccccc1)C(=O)O	190546.07
CHEMBL80240	57	Cc1ccc(N(CC(=O)O)C(=O)C(C)N)cc1	2570.4
CHEMBL263556	58	O=C(O)[C@@H]1CCCN1C(=O)CP(=O)(O)O	8317.64
CHEMBL322069	59	CC(=O)SCC(C)C(=O)Nc1ccccc1	2000
CHEMBL89956	60	N[C@@H](Cc1ccc(O)cc1)C(=O)NCC(=O)O	1995262.31
CHEMBL53400	61	NCC(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	208929.61
CHEMBL322291	62	O=C(O)C1C=CCn2c(=O)n(CS)c(=O)n21	2511.89
CHEMBL80939	63	O=C(O)[C@@H]1C=CCn2c(=O)n(CS)c(=O)n21	707.95
CHEMBL78780	64	C[C@H](CCC(=O)O)C(=O)N1CCC[C@H]1C(=O)O	4900

CHEMBL152188	65	CC(CCC(=O)N1CCC[C@H]1C(=O)O)C(=O)O	260000
CHEMBL293213	66	CC(CCC(=O)N1CCCC1C(=O)O)C(=O)O	260000
CHEMBL62678	67	C[C@@H](CCC(=O)O)C(=O)N1CCCC1C(=O)O	4900
CHEMBL311239	68	O=C(CCS)N[C@@H](Cc1c[nH]cn1)C(=O)O	229086.77
CHEMBL431983	69	O=C(O)[C@@H]1CCCN1C(=O)C1CCCC1S	64.57
CHEMBL321012	70	O=C(O)C1CCCN1C(=O)C1CCCC1S	125.89
CHEMBL358439	71	C[C@@H](NCC(=O)O)C(=O)N1CCC[C@H]1C(=O)O	2400
CHEMBL65423	72	C[C@H](NCC(=O)O)C(=O)N1CCC[C@H]1C(=O)O	2398.83
CHEMBL11795	73	CC(=O)SCC1CCCN(CC(=O)O)C1=O	100000
CHEMBL55103	74	CCC(C)C1CC(Cc2ccccc2)SC1=O	2300
CHEMBL312550	75	O=C(O)[C@@H]1CCCN1C(=O)CCP(=O)(O)O	47863.01
CHEMBL417226	76	O=C(O)C1c2ccccc2CN1C(=O)CCS	58000
CHEMBL326801	77	CC(=O)SCC(C)C(=O)NCc1ccccc1	500000
CHEMBL94016	78	C[C@H](N)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	87096.36
CHEMBL316103	79	C[C@H](NC(=O)[C@@H](N)Cc1ccc(O)cc1)C(=O)O	457088.19
CHEMBL298827	80	O=C(O)CNC(=O)[C@@H](CS)Cc1ccccc1	140
CHEMBL312094	81	O=C(CCS)N[C@@H](Cc1ccccc1)C(=O)O	426.58
CHEMBL10247	82	O=C(O)CNC(=O)C(CS)Cc1ccccc1	316.23
CHEMBL38405	83	O=C(O)CC(=O)NC(CS)Cc1ccccc1	1000000000
CHEMBL80667	84	O=C(CCS)OC(Cc1ccccc1)C(=O)O	707.95
CHEMBL326902	85	O=C(O)Cn1c(=O)n2n(c1=O)C(C(=O)O)C=CC2	3.162
CHEMBL298647	86	O=C(CN1Cc2c(Cl)cccc2NC1=O)NO	90000
CHEMBL152758	87	CC(N[C@H](C)C(=O)N1CCC[C@H]1C(=O)O)C(=O)O	90

CHEMBL105790	88	CC1(CS)C(=O)N2CCCC(C(=O)O)N2C1=O	3162.28
CHEMBL325544	89	O=C(O)C1CCCN2C(=O)CC(CS)C(=O)N12	125.89
CHEMBL416339	90	CC(=O)SCC1CCCN(C(C)C(=O)O)C1=O	35000
CHEMBL311316	91	CC(C)(C)OC(=O)[C@@H]1CCCN1C(=O)CCS	38904.51
CHEMBL187265	92	O=c1c2cc(O)c(O)cc2oc2cc(O)cc(O)c12	530800
CHEMBL477921	93	O=c1c2ccc(O)c(O)c2oc2c(O)c(O)ccc12	238500
CHEMBL477740	94	O=c1c2cc(O)c(O)cc2oc2c(O)c(O)ccc12	35400
CHEMBL12700	95	O=c1c2cc(O)c(O)cc2oc2cc(O)c(O)cc12	769000
CHEMBL448040	96	O=c1c2ccc(O)c(O)c2oc2cc(O)cc(O)c12	69200
CHEMBL299759	97	NCC(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	30199.52
CHEMBL327588	98	N[C@@H](Cc1c[nH]c2ccccc12)C(=O)NCC(=O)O	5888436.55
CHEMBL4248427	99	C[C@H](CS)C(=O)N1C[Si](C)(C)C[C@H]1C(=O)O	43
CHEMBL8486	100	CC(C)[C@H](N)C(=O)N[C@@H](Cc1ccccc1)C(=O)O	52480.75
CHEMBL110870	101	CC(NP(=O)(O)O)C(=O)C1CCCC1C(=O)O	5.012
CHEMBL80101	102	O=C(O)CCC(=O)N[C@@H](Cc1ccccc1)C(=O)O	549540.87
CHEMBL62203	103	CC(CS)C(=O)N1Cc2ccccc2C1C(=O)O	14000
CHEMBL23849	104	CC(CS)C(=O)N1c2ccccc2CC1C(=O)O	12.59
CHEMBL611148	105	C[C@@H](CS)C(=O)N1c2ccccc2C[C@H]1C(=O)O	3.7
CHEMBL78443	106	C[C@H](NP(=O)(O)O)C(=O)N1CCC[C@H]1C(=O)O	776.25
CHEMBL53141	107	CCC(C)C(CC(S)Cc1ccccc1)C(=O)O	4000
CHEMBL308267	108	O=C(O)CCc1cc(O)c2n(c1=O)[C@H](C(=O)O)CC2	457.09
CHEMBL107635	109	O=C(O)CCc1cc(O)c2n(c1=O)C(C(=O)O)CC2	31.62
CHEMBL366937	110	CC(S)C(Cc1ccccc1)C(=O)NCC(=O)O	11000

CHEMBL80657	111	CN(C(=O)CCS)[C@@H](Cc1ccccc1)C(=O)O	30902.95
CHEMBL196902	112	O=C(O)CCNC(=O)C(CS)Cc1ccccc1	32000
CHEMBL90260	113	CC(C)C[C@H](NC(=O)[C@@H](N)Cc1c[nH]cn1)C(=O)O	3235936.57
CHEMBL310750	114	O=C(O)CCn1c(=O)n2n(c1=O)[C@H](C(=O)O)C=CC2	1
CHEMBL78903	115	C[C@@]1(CCS)C(=O)N2CC=C[C@@H](C(=O)O)N2C1=O	100
CHEMBL90575	116	NC(N)=NCCC[C@H](N)C(=O)N1CCC[C@@H]1C(=O)O	181970.09
CHEMBL79016	117	C[C@H](NC(=O)CCS)C(=O)N1CCC[C@H]1C(=O)O	2570.4
CHEMBL78971	118	CC[C@@H](C)[C@H](C)C(=O)N[C@@H](Cc1ccccc1)C(=O)O	933254.3
CHEMBL57202	119	CC(=O)SCC(C)C(=O)N1CCc2ccccc2C1	49000
CHEMBL312120	120	O=C(CS)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	758.58
CHEMBL91330	121	CC[C@H](C)[C@H](N)C(=O)N[C@@H](Cc1ccccc1)C(=O)O	933254.3
CHEMBL279737	122	CC(CS)C(=O)N1Cc2ccccc2CC1C(=O)O	18
CHEMBL79189	123	C[C@H](CS)C(=O)N1Cc2ccccc2C[C@H]1C(=O)O	19.95
CHEMBL61916	124	CC(CS)C(=O)N1CCc2ccccc2C1C(=O)O	58000
CHEMBL90069	125	CC(C)[C@H](N)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	21877.62
CHEMBL114552	126	O=C(O)CCCNCP(=O)(O)CCc1ccccc1	100000
CHEMBL263501	127	CCCCC(CC(=O)NO)S(=O)(=O)c1ccccc1	81000
CHEMBL1788203	128	CCCC[C@H](CC(=O)NO)S(=O)(=O)c1ccccc1	34000
CHEMBL79212	129	C[C@H](NC(CCN)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	323.59
CHEMBL416706	130	CCCC(C(=O)O)N1CCCC(CSC(C)=O)C1=O	30000
CHEMBL421581	131	C[C@H](N)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O	1500
CHEMBL858	132	O=C(O)CN(CCN(CC(=O)O)CC(=O)O)CC(=O)O	14000
CHEMBL309799	133	CC(S)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	64.57

CHEMBL61137	134	CC(=O)SCCC(=O)N1Cc2ccccc2C1C(=O)O	22000
CHEMBL91090	135	CC[C@H](C)[C@H](N)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	3715.35
CHEMBL340119	136	CC(C)C[C@H](N)C(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1C(=O)O	2300
CHEMBL340152	137	CCCC[C@H](N)C(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1C(=O)O	700
CHEMBL6246	138	O=c1oc2c(O)c(O)cc3c(=O)oc4c(O)c(O)cc1c4c23	5000000
CHEMBL3581907	139	N[C@@H](CCN[C@@H](CCN1CC[C@H]1C(=O)O)C(=O)O)C(=O)O	18700
CHEMBL91777	140	CC(C)[C@H](N)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	1584.89
CHEMBL139523	141	O=C(CCC(=O)N1CCCC1C(=O)O)CCc1ccccc1	2600000
CHEMBL80147	142	C[C@H](NCCCc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	5754.4
CHEMBL78489	143	CC(CS)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	33113.11
CHEMBL80981	144	CC(C)(S)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	38018.94
CHEMBL3581908	145	N=C(N)NCCC[C@H](NC(=O)C1CSSC1)C(=O)O	113000
CHEMBL57200	146	CC(=O)SCC(C)C(=O)N1Cc2ccccc2C1C(=O)O	24000
CHEMBL173287	147	CC(C)CC(S)CC(=O)NC(Cc1ccccc1)C(=O)O	10000
CHEMBL3581909	148	O=C(OC1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)C1CSSC1	347000
CHEMBL3037879	149	C[C@H](N[C@H](C(=O)O)[C@@H]1CCCN1)C(=O)N1CCC[C@H]1C(=O)O	700
CHEMBL110444	150	O=C(O)C1CSC(c2ccccc2O)N1C(=O)CCS	19.95
CHEMBL309962	151	O=C(O)[C@@H]1CS[C@H](c2ccccc2O)N1C(=O)CCS	3.715
CHEMBL451209	152	C[C@H](CS)C(=O)N[C@@H](CSCc1ccccc1)C(=O)O	34
CHEMBL78340	153	CC(C)CCC(N[C@@H](C)C(=O)N1CCC[C@H]1C(=O)O)C(=O)O	2.57
CHEMBL108606	154	CC(C)CCC(NC(C)C(=O)N1CCCC1C(=O)O)C(=O)O	7.943
CHEMBL56923	155	CC[C@H](C)[C@H](N)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	1995.26
CHEMBL2370859	156	C[C@H](NC(=O)CCc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	330000

CHEMBL148616	157	C[C@H](CS)C(=O)N1Cc2[nH]c3ccccc3c2CC1C(=O)O	500
CHEMBL27560	158	N[C@@H](Cc1ccccc1)C(=O)NCC(=O)N1CCC[C@H]1C(=O)O	20000
CHEMBL86891	159	O=C(O)CN1CCC(NC(CCc2ccccc2)C(=O)O)C1=O	12000
CHEMBL560685	160	O=C(CC(S)C(F)(F)F)NC(Cc1ccccc1)C(=O)O	830
CHEMBL92255	161	NC(N)=NCCC[C@H](N)C(=O)N[C@@H](Cc1ccccc1)C(=O)O	229086.77
CHEMBL291636	162	N=C(N)NCCC[C@H](NC(=O)[C@@H](N)Cc1ccccc1)C(=O)O	912010.84
CHEMBL58720	163	CC(=O)SCC(C)C(=O)N1Cc2ccccc2CC1C(=O)O	130
CHEMBL12006	164	CC(=O)SCC1CCCN(C(C(=O)O)c2ccccc2)C1=O	19000
CHEMBL342091	165	O=C1N[C@H](C(=O)O)CCCCCc2ccccc2C[C@@H]1CS	8000
CHEMBL148140	166	O=C1N[C@@H](C(=O)O)Cc2ccccc2CCCCC[C@@H]1CS	6000
CHEMBL357311	167	O=C(O)[C@@H]1CCCCCc2ccccc2C[C@@H](CS)C(=O)N1	2000
CHEMBL148840	168	O=C(O)[C@@H]1CCCCCc2cccc(c2)C[C@@H](CS)C(=O)N1	1400
CHEMBL147185	169	O=C1N[C@H](C(=O)O)CCCCCc2cccc(c2)C[C@@H]1CS	4
CHEMBL50480	170	C[C@H](S)C(=O)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	24
CHEMBL2110258	171	CC(C)C[C@H](S)[C@H](Cc1ccccc1)C(=O)N[C@@H](C)C(=O)O	10000
CHEMBL173919	172	CC(C)CC(S)[C@H](Cc1ccccc1)C(=O)N[C@@H](C)C(=O)O	4000
CHEMBL174237	173	CC(C)C[C@@H](S)[C@H](Cc1ccccc1)C(=O)N[C@@H](C)C(=O)O	5000
CHEMBL55286	174	CC(C)CC(S)C(Cc1ccccc1)C(=O)NC(C)C(=O)O	12000
CHEMBL2110347	175	CC(C)C[C@H](S)[C@@H](Cc1ccccc1)C(=O)N[C@@H](C)C(=O)O	16100
CHEMBL459960	176	Cc1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	240
CHEMBL107700	177	CC(NC(=O)C1CCCCC1S)C(=O)N1CCCC1C(=O)O	10
CHEMBL467755	178	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(F)cc1)C(=O)O	100
CHEMBL28361	179	C[C@H](NC(=O)[C@@H](N)Cc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	1400

CHEMBL128399	180	C[C@H](NC(=O)[C@@H](N)Cc1ccccc1)C(=O)N1CCC[C@@H]1C(=O)O	4200
CHEMBL44139	181	CCC[C@H](N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O)C(=O)O	6
CHEMBL314706	182	O=C(O)CN1CCCC(NC(CCc2ccccc2)C(=O)O)C1=O	430
CHEMBL78001	183	CCCCC(CC(=O)NO)S(=O)(=O)c1ccc2ccccc2c1	12400
CHEMBL357584	184	O=C1N[C@H](C(=O)O)CCCCCCc2cccc(c2)C[C@@H]1CS	175
CHEMBL147755	185	O=C1N[C@H](C(=O)O)Cc2ccccc2CCCCCC[C@@H]1CS	725
CHEMBL147526	186	O=C(O)[C@@H]1CCCCCCc2cccc(c2)C[C@@H](CS)C(=O)N1	118
CHEMBL149367	187	O=C1N[C@@H](C(=O)O)Cc2ccccc2CCCCCC[C@@H]1CS	470
CHEMBL540674	188	O=C(CC(S)C(F)(F)F)NC(Cc1ccc(O)cc1)C(=O)O	160
CHEMBL457796	189	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(C#N)cc1)C(=O)O	300
CHEMBL551145	190	O=C(CC(S)C(F)(F)F)NC(Cc1ccc(F)cc1)C(=O)O	300
CHEMBL498798	191	C[C@H](CS)C(=O)N1[C@@H](SCc2ccccc2)CC[C@H]1C(=O)O	2.4
CHEMBL511184	192	CCc1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	21
CHEMBL129432	193	C[C@H](NC(=O)[C@@H](N)CCCN=C(N)N)C(=O)N1CCC[C@@H]1C(=O)O	16000
CHEMBL511344	194	COc1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	100
CHEMBL172913	195	CC(NC(=O)c1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O	300
CHEMBL80193	196	C[C@H](NC(CCc1ccccc1)C(N)=O)C(=O)N1CCC[C@H]1C(=O)O	25703.96
CHEMBL111578	197	O=C(SC1CCCC1C(=O)N1CCCC1C(=O)O)c1ccccc1	63.1
CHEMBL444474	198	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(Cl)cc1)C(=O)O	250
CHEMBL40806	199	CC(C)C[C@H](N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O)C(=O)O	45
CHEMBL577	200	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	2.9
CHEMBL39538	201	C[C@H](NC(CCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	1.202
CHEMBL85175	202	O=C(O)CN1CCCCC(NC(CCc2ccccc2)C(=O)O)C1=O	700

CHEMBL41985	203	CCCC[C@H](N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O)C(=O)O	5
CHEMBL285897	204	CC(OC(CCc1ccccc1)C(=O)O)C(=O)N1CCCC1C(=O)O	110
CHEMBL77556	205	C[C@@H](NC(=O)C(S)Cc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	50.12
CHEMBL287896	206	C[C@H](NC(=O)[C@@H](S)Cc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	30
CHEMBL299169	207	O=C(O)CN1CCCC[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C1=O	22
CHEMBL417034	208	O=C(O)CN1NCCC[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C1=O	60
CHEMBL500408	209	Cc1ccccc1CS[C@H]1CC[C@@H](C(=O)O)N1C(=O)[C@H](C)CS	2.3
CHEMBL499612	210	Cc1cccc(CS[C@H]2CC[C@@H](C(=O)O)N2C(=O)[C@H](C)CS)c1	0.9
CHEMBL526298	211	C[C@H](CS)C(=O)N1[C@@H](SCCc2ccccc2)CC[C@H]1C(=O)O	0.18
CHEMBL309601	212	C[C@H](NP(=O)(O)CCc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	7.079
CHEMBL62955	213	CC(NP(=O)(O)CCc1ccccc1)C(=O)N1CCCC1C(=O)O	25.12
CHEMBL556775	214	Cl.NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O	25000
CHEMBL342652	215	CC(=O)N1CCCC1C(NC(C)C(=O)N1CCCC1C(=O)O)C(=O)O	9.4
CHEMBL461021	216	CCCc1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	210
CHEMBL461022	217	CC(C)c1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	300
CHEMBL476621	218	O=C(NNC(=S)Nc1ccc(Cl)c(Cl)c1)c1ccccc1O	0.7
CHEMBL79131	219	C[C@H](NP(=O)(O)OCc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	40.74
CHEMBL457317	220	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(N(C)C)cc1)C(=O)O	520
CHEMBL54942	221	CC(NC(=O)C(Cc1ccccc1)CC(S)c1ccccc1)C(=O)O	3500
CHEMBL457571	222	CCOc1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	340
CHEMBL456278	223	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc([N+](=O)[O-])cc1)C(=O)O	320
CHEMBL338446	224	N[C@@H](Cc1ccccc1)C(=O)N1CCC[C@H]1C(=O)N1CCC[C@@H]1C(=O)O	78000
CHEMBL511849	225	CSc1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	300

CHEMBL551622	226	O=C(CC(S)C(F)(F)F)NC(Cc1c[nH]c2ccccc12)C(=O)O	400
CHEMBL172917	227	CC(=O)NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O	330
CHEMBL91160	228	NC(N)=NCCC[C@H](N)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	15848.93
CHEMBL290328	229	COC(=O)[C@H](CC(C)C)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	1000
CHEMBL311363	230	C[C@@H]1CCCC(NC(CCc2ccccc2)C(=O)O)C(=O)N1CC(=O)O	3.02
CHEMBL110439	231	CC1CCCC(NC(CCc2ccccc2)C(=O)O)C(=O)N1CC(=O)O	10
CHEMBL86518	232	O=C(O)CN1CCCCCC(NC(CCc2ccccc2)C(=O)O)C1=O	1700
CHEMBL2114219	233	O=C(O)CN1CCCCC[C@H](N[C@@H](CCc2ccccc2)C(=O)O)C1=O	2
CHEMBL3350294	234	O=C(O)CN1CCCCC[C@@H](N[C@H](CCc2ccccc2)C(=O)O)C1=O	92
CHEMBL2304318	235	C[C@H](N[C@](C)(CCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	1.2
CHEMBL431052	236	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C(=O)NC[C@H]1C(=O)O	3.5
CHEMBL430554	237	NCCCC[C@H](N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O)C(=O)O	7
CHEMBL55613	238	CCC(C)C(CC(S)Cc1ccccc1)C(=O)N1CCCC1C(=O)O	2800
CHEMBL78726	239	C[C@H](NC(CCc1ccc(O)cc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	2.818
CHEMBL299181	240	C[C@H](NC(=O)[C@@H](CS)Cc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	4
CHEMBL299875	241	CCCC[C@H](S)C(=O)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	5.4
CHEMBL415932	242	CC(C)C[C@H](S)C(=O)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	6.3
CHEMBL1907762	243	CC(SC(CCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	580
CHEMBL63001	244	CN1CCC[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C(=O)N1CC(=O)O	16
CHEMBL416979	245	CSCC[C@H](N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O)C(=O)O	6.5
CHEMBL443353	246	C[C@H](CS)C(=O)N1[C@@H](SCCCc2ccccc2)CC[C@H]1C(=O)O	0.029
CHEMBL270576	247	CC(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	60
CHEMBL1581	248	CCC[C@H](N[C@@H](C)C(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)C(=O)OCC	1.5

CHEMBL463320	249	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(C(C)(C)C)cc1)C(=O)O	130
CHEMBL458238	250	CC(C)c1ccc(C(C)SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	200
CHEMBL558146	251	O=C(CC(S)C(F)(F)C(F)(F)F)NC(Cc1ccccc1)C(=O)O	1010
CHEMBL54601	252	CC(NC(=O)C(Cc1ccccc1)C(S)CCc1ccccc1)C(=O)O	4000
CHEMBL54477	253	O=C(O)CCNC(=O)C(Cc1ccccc1)CC(S)Cc1ccccc1	3500
CHEMBL293794	254	CC(NC(=O)C(Cc1ccccc1)CC(S)Cc1ccccc1)C(=O)O	3800
CHEMBL93562	255	CC(C)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](CCc1ccccc1)C(=O)O	100
CHEMBL4172457	256	CCCCCCC[C@H](N)C(O)C(=O)N[C@@H](C)C(=O)N[C@H](C(=O)O)C(C)C	460000
CHEMBL77446	257	O=C(O)C(CCc1ccccc1)N[C@H]1CCN2CCC[C@@H](C(=O)O)N2C1=O	56.23
CHEMBL4163389	258	CN1C(=O)/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc21	22.3
CHEMBL2442646	259	CCCC(N[C@@H](C)C(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)P(=O)(O)O	30
CHEMBL578	260	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1CCC[C@H]1C(=O)O	1.2
CHEMBL86915	261	O=C(O)CN1CCCCCCC(NC(CCc2ccccc2)C(=O)O)C1=O	8.1
CHEMBL40709	262	CCCC[C@H](N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O)C(=O)OCC	2600
CHEMBL2111941	263	C[C@H](NC(CNC(=O)c1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	40
CHEMBL99701	264	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O	1.7
CHEMBL2112767	265	C[C@H](N[C@@H](CNC(=O)c1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	17
CHEMBL284494	266	CCOC(=O)C(CCc1ccccc1)OC(C)C(=O)N1CCCC1C(=O)O	220
CHEMBL66254	267	O=C(N[C@H]1CCCN2CC[C@@H](C(=O)O)N2C1=O)[C@@H](S)Cc1ccccc1	12
CHEMBL114658	268	CC1(C)CCCC(NC(=O)[C@@H](S)Cc2ccccc2)C(=O)N1CC(=O)O	12
CHEMBL194651	269	CCCc1cnc(N2C[C@H](S)CC2CNCc2cc(F)ccc2F)nc1	31500
CHEMBL304881	270	CN1C(=O)CC[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C(=O)N1CC(=O)O	16
CHEMBL459961	271	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(C(F)(F)F)cc1)C(=O)O	340

CHEMBL499305	272	CC(C)c1ccc(CS[C@@H]2CC[C@@H](C(=O)O)N2C(=O)[C@H](C)CS)cc1	3.3
CHEMBL449792	273	CC(C)c1ccc(CS[C@H]2CC[C@@H](C(=O)O)N2C(=O)[C@H](C)CS)cc1	8.4
CHEMBL290802	274	O=C(O)CN1C(=O)[C@@H](N[C@@H](Cc2ccccc2)C(=O)O)CCc2ccccc21	8.5
CHEMBL383507	275	NCCC[C@H](NC(=O)[C@@H](N)Cc1ccsc1)C(=O)N1CCC[C@H]1C(=O)O	600
CHEMBL456723	276	CC(C)c1ccc(C(C)(C)SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	260
CHEMBL458447	277	CCC(SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)c1ccc(C(C)C)cc1	340
CHEMBL80646	278	C[C@H](NC(CCc1ccccc1)P(=O)(O)O)C(=O)N1CCC[C@H]1C(=O)O	51.29
CHEMBL258278	279	CC(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1ccc(O)c(O)c1)C(=O)O	23
CHEMBL457795	280	CCN(CC)c1ccc(CSC[C@H](NC(=O)[C@H](C)CS)C(=O)O)cc1	4600
CHEMBL101360	281	CCCCCCCC[C@H](N[C@@H](C)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O)C(=O)O	1.6
CHEMBL100966	282	CCCCCCCC[C@H](N[C@H](C)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O)C(=O)O	2400
CHEMBL525967	283	CSc1ccc(CS[C@H]2CC[C@@H](C(=O)O)N2C(=O)[C@H](C)CS)cc1	4.1
CHEMBL300016	284	C[C@H](NC(=O)[C@@H](S)Cc1ccccc1)C(=O)N(CC(=O)O)c1ccccc1	14
CHEMBL499610	285	C[C@H](CS)C(=O)N1[C@@H](SCc2ccc3ccccc3c2)CC[C@H]1C(=O)O	4.2
CHEMBL526896	286	C[C@H](CS)C(=O)N1[C@@H](SCc2cccc3ccccc23)CC[C@H]1C(=O)O	3.2
CHEMBL463520	287	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(-c2ccccc2)cc1)C(=O)O	280
CHEMBL325056	288	O=C(NC1CCC[C@H]2CCC[C@@H](C(=O)O)N2C1=O)[C@@H](S)Cc1ccccc1	5
CHEMBL78731	289	C[C@H](NC(CCC(=O)Nc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	5.248
CHEMBL255568	290	CC(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	103
CHEMBL305108	291	O=C(N[C@H]1CCC(=O)N2CC[C@@H](C(=O)O)N2C1=O)[C@@H](S)Cc1ccccc1	20
CHEMBL417169	292	CCOC(=O)[C@H](CCCCN)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	40
CHEMBL309308	293	NCCC[C@H](NC(CCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	2.188
CHEMBL65545	294	O=C(N[C@H]1CCCN2CCC[C@@H](C(=O)O)N2C1=O)[C@@H](S)Cc1ccccc1	8

CHEMBL458237	295	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(S(C)(=O)=O)cc1)C(=O)O	300
CHEMBL513671	296	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(Br)cc1)C(=O)O	280
CHEMBL340528	297	O=C(O)C(CCc1ccccc1)N[C@H]1CCCC2SC[C@@H](C(=O)O)N2C1=O	0.6
CHEMBL312224	298	O=C(O)C(CCc1ccccc1)NC1CCC[C@H]2SC[C@@H](C(=O)O)N2C1=O	0.6026
CHEMBL322266	299	O=C(O)C(CCc1ccccc1)NC1CCCC2SCC(C(=O)O)N2C1=O	1.995
CHEMBL347755	300	O=C(O)C(CCc1ccccc1)N[C@H]1CCC[C@H]2SC[C@@H](C(=O)O)N2C1=O	0.6
CHEMBL25996	301	CCOC(=O)C(CCc1ccccc1)SC(C)C(=O)N1CCCC1C(=O)O	2900
CHEMBL61566	302	CCCN1CCC[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C(=O)N1CC(=O)O	6
CHEMBL372307	303	O=C(N[C@H]1CCCSC2CCC[C@@H](C(=O)O)N21)[C@@H](S)Cc1ccccc1	5
CHEMBL53621	304	CCC(C)C(S)CC(C(=O)NC(Cc1ccc(O)cc1)C(N)=O)C(C)CC	4000
CHEMBL419649	305	C[C@H](CC(CCc1ccccc1)C(=O)O)C(=O)N1c2ccccc2C[C@H]1C(=O)O	4.786
CHEMBL53678	306	CCC(C)C(S)CC(C(=O)NC(Cc1ccc(O)cc1)C(=O)O)C(C)CC	3400
CHEMBL525948	307	CC(C)c1ccc(CCS[C@H]2CC[C@@H](C(=O)O)N2C(=O)[C@H](C)CS)cc1	19
CHEMBL512941	308	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(C2CCCCC2)cc1)C(=O)O	4000
CHEMBL500409	309	C[C@H](CS)C(=O)N1[C@@H](SCc2ccc(C(C)(C)C)cc2)CC[C@H]1C(=O)O	410
CHEMBL58480	310	CC(NC(CCc1ccccc1)C(=O)O)C(=O)N1Cc2ccccc2C1C(=O)O	3.1
CHEMBL80906	311	O=C(O)CN1C(=O)C(NC(CCc2ccccc2)C(=O)O)CCc2ccccc21	1.698
CHEMBL92949	312	CC(C)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)O	240
CHEMBL515834	313	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(OC(F)(F)F)cc1)C(=O)O	260
CHEMBL54457	314	O=C(O)C1CCCN1C(=O)C(Cc1ccccc1)CC(S)Cc1ccccc1	3600
CHEMBL458451	315	CC(C)c1ccc(C(SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)C(C)C)cc1	430
CHEMBL458450	316	CCCC(SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)c1ccc(C(C)C)cc1	420
CHEMBL78385	317	COP(=O)(O)C(CCc1ccccc1)N[C@@H](C)C(=O)N1CCC[C@H]1C(=O)O	436.52

CHEMBL80779	318	C[C@H](NC(CCc1cccc2ccccc12)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	1.096
CHEMBL105971	319	CC(NC(CCc1cccc2ccccc12)C(=O)O)C(=O)N1CCCC1C(=O)O	3.162
CHEMBL422944	320	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@H](S)Cc2ccccc2)CCc2ccccc21	284
CHEMBL299639	321	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](S)Cc2ccccc2)CCc2ccccc21	12
CHEMBL50315	322	O=C(O)CN1C(=O)[C@H](NC(=O)[C@@H](S)Cc2ccccc2)CCc2ccccc21	306
CHEMBL51576	323	O=C(O)CN1Cc2ccccc2C[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C1=O	33
CHEMBL100063	324	CCCCCCCC[C@H](N[C@@H](CC)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O)C(=O)O	2.4
CHEMBL54036	325	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](S)Cc2ccccc2)COc2ccccc21	28
CHEMBL204944	326	NCCC[C@H](NC(=O)CCc1nc2ccccc2[nH]1)C(=O)N1CCC[C@H]1C(=O)O	10000
CHEMBL409920	327	CC(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1cccc2ccccc12)C(=O)O	110
CHEMBL450935	328	CC(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1ccc2ccccc2c1)C(=O)O	52
CHEMBL78346	329	C[C@H](NC(CCc1ccccc1)C(=O)O)C(=O)N1CC2CCCCC2[C@H]1C(=O)O	2.188
CHEMBL2079670	330	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C2CCCCC2C[C@H]1C(=O)O	260
CHEMBL4167651	331	CC(=O)N1C(=O)/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc21	10.67
CHEMBL4171024	332	CCCN1C(=O)/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc21	11.8
CHEMBL299438	333	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](S)CC2CCCCC2)CCc2ccccc21	22
CHEMBL317094	334	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O	9900
CHEMBL99798	335	CC(C)C[C@H](N[C@@H](C)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O)C(=O)O	2.1
CHEMBL66481	336	O=C(N[C@H]1CCC(=O)N2CCC[C@@H](C(=O)O)N2C1=O)[C@@H](S)Cc1ccccc1	7
CHEMBL1237	337	NCCCC[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	4.7
CHEMBL105890	338	NCCCCC(NC(CCc1ccccc1)C(=O)O)C(=O)N1CCCC1C(=O)O	3.981
CHEMBL514268	339	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(Oc2ccccc2)cc1)C(=O)O	370
CHEMBL91218	340	C[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	660

CHEMBL1422	341	N[C@@H](CC(=O)N1CCn2c(nnc2C(F)(F)F)C1)Cc1cc(F)c(F)cc1F	11000
CHEMBL62417	342	CCCN1C(=O)CC[C@H](NC(=O)[C@@H](S)Cc2ccccc2)C(=O)N1CC(=O)O	10
CHEMBL442455	343	CC(C)CN(CP(=O)(O)O)C(=O)N[C@@H](Cc1ccc2ccccc2c1)C(=O)O	29
CHEMBL289556	344	O=C(N[C@H]1CCS[C@H]2CCC[C@@H](C(=O)O)N2C1=O)[C@@H](S)Cc1ccccc1	2
CHEMBL499611	345	CC(C)c1ccc(CCCS[C@H]2CC[C@@H](C(=O)O)N2C(=O)[C@H](C)CS)cc1	15
CHEMBL1733	346	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1Cc2ccccc2C[C@H]1C(=O)O	280
CHEMBL61811	347	CC(NC(CCc1ccccc1)C(=O)O)C(=O)N1CCc2ccccc2C1C(=O)O	5.8
CHEMBL91623	348	C[C@H](CS)C(=O)N1C[C@H](NC(=O)C(CS)Cc2ccccc2)C[C@H]1C(=O)O	18
CHEMBL328327	349	C[C@H](CS)C(=O)N1C[C@@H](NC(=O)C(CS)Cc2ccccc2)C[C@H]1C(=O)O	87
CHEMBL26226	350	CC(OC(CCc1ccccc1)C(=O)O)C(=O)N1Cc2ccccc2CC1C(=O)O	86
CHEMBL1907759	351	CCOC(=O)C(CCc1ccccc1)SC(C)C(=O)N1CSC[C@@H]1C(=O)O	4800
CHEMBL505067	352	CCCCC(SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)c1ccc(C(C)C)cc1	3600
CHEMBL42683	353	O=C(O)CN1C(=O)[C@@H](N[C@@H](COCc2ccccc2)C(=O)O)CCc2ccccc21	10
CHEMBL172262	354	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)c1ccco1	1.4
CHEMBL92632	355	CC(C)[C@H](N[C@@H](C)C(=O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	23
CHEMBL328737	356	CC[C@H](C)C(NCC(=O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	65
CHEMBL301364	357	O=C(O)CCN1C(=O)[C@@H](NC(=O)[C@@H](S)Cc2ccccc2)CCc2ccccc21	36
CHEMBL51498	358	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](CS)Cc2ccccc2)CCc2ccccc21	4.8
CHEMBL50814	359	O=C(O)CN1C(=O)C(NC(=O)[C@@H](S)Cc2ccccc2)CCCc2ccccc21	25
CHEMBL48534	360	CC(S)(Cc1ccccc1)C(=O)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	63
CHEMBL316865	361	CCCCCCCC[C@H](N[C@@H](C)C(=O)N1C(=O)N(C)C[C@H]1C(=O)O)C(=O)OCC	5100
CHEMBL342373	362	CC(NC(C(=O)O)C1CCCN1C(=O)NCC(=O)O)C(=O)N1CCCC1C(=O)O	6
CHEMBL53879	363	C[C@H]1Oc2ccccc2N(CC(=O)O)C(=O)[C@H]1NC(=O)[C@@H](S)Cc1ccccc1	13

CHEMBL290284	364	O=C(O)CN1C(=O)[C@@H](NC(CCc2ccccc2)C(=O)O)CSc2ccccc21	3
CHEMBL51389	365	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](CS)Cc2ccccc2)COc2ccccc21	4.4
CHEMBL50559	366	C[C@@H]1Oc2ccccc2N(CC(=O)O)C(=O)[C@H]1NC(=O)[C@@H](S)Cc1ccccc1	36
CHEMBL38613	367	O=C(O)CN1C(=O)[C@@H](N[C@@H](CSc2ccccc2)C(=O)O)CCc2ccccc21	4
CHEMBL301751	368	O=C(O)CCN1C(=O)[C@@H](NC(=O)[C@@H](S)Cc2ccccc2)COc2ccccc21	31
CHEMBL171850	369	CC(C)CC(S)C(Cc1ccccc1)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	710
CHEMBL172787	370	CC(C)C(C(=O)NC(Cc1ccc(O)cc1)C(=O)O)C(S)CCc1ccccc1	4500
CHEMBL328030	371	CC(C)Oc1ccc(C[C@H](NC(=O)[C@@H](NCP(=O)(O)O)C(C)C)C(=O)O)cc1	92
CHEMBL368784	372	O=C(CCC(=O)N1CCC[C@H]1C(=O)O)C(Cc1ccccc1)NC(=O)C1CCCO1	20
CHEMBL273142	373	CC(C)(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1ccc2ccccc2c1)C(=O)O	191
CHEMBL1168	374	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1[C@H](C(=O)O)C[C@@H]2CCC[C@@H]21	4
CHEMBL298388	375	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](S)Cc2ccccc2)CSc2ccccc21	13
CHEMBL312431	376	C[C@H](NC(C(=O)O)C1CCCN1C(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	2.884
CHEMBL335997	377	CC(NC(C(=O)O)C1CCCN1C(=O)c1ccccc1)C(=O)N1CCCC1C(=O)O	2.9
CHEMBL4159339	378	CC(=O)CN1C(=O)/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc21	112
CHEMBL80055	379	Cc1ccc(CSC[C@@H](NC(=O)[C@@H](CS)Cc2ccccc2)C(=O)O)cc1C	460
CHEMBL4174467	380	O=C(O)CN1C(=O)/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc21	2.846
CHEMBL430689	381	CC(C)CC[C@H](N[C@@H](C)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O)C(=O)O	3.3
CHEMBL317304	382	CCCCN1C[C@@H](C(=O)O)N(C(=O)[C@H](C)N[C@@H](CCc2ccccc2)C(=O)O)C1=O	6.7
CHEMBL1907755	383	CC(SC(CCc1ccccc1)C(=O)O)C(=O)N1C2CCCCC2C[C@H]1C(=O)O	140
CHEMBL310200	384	COc1ccc(CSC[C@@H](NC(=O)[C@H](CS)Cc2ccccc2)C(=O)O)cc1	290
CHEMBL1242210	385	CN1CCN(c2cccc3nc(CN(C)[C@H]4CCCc5cccnc54)c(CO)n23)CC1	193
CHEMBL44077	386	O=C(O)CN1C(=O)[C@@H](N[C@@H](Cc2c[nH]c3ccccc23)C(=O)O)CCc2ccccc21	13

CHEMBL368640	387	NC(=O)[C@@H]1CCCN1C(=O)CCC(=O)C(Cc1ccccc1)NC(=O)c1ccccc1	150
CHEMBL502817	388	C[C@H](CS)C(=O)N1[C@@H](SCc2ccc(C3CCCCC3)cc2)CC[C@H]1C(=O)O	280
CHEMBL435360	389	O=C(N[C@@H](Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	12
CHEMBL416147	390	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	70
CHEMBL10465	391	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCCC1C(=O)O)c1ccccc1	10
CHEMBL191	392	CCCCc1nc(Cl)c(CO)n1Cc1ccc(-c2ccccc2-c2nnn[nH]2)cc1	19
CHEMBL32032	393	O=C(N[C@@H](Cc1ccccc1)C(=O)NCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	9400
CHEMBL176688	394	O=C(NC(Cc1cccnc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	5.7
CHEMBL456045	395	C[C@H](CS)C(=O)N[C@@H](CSC(C)(C)c1ccc(C2CCCCC2)cc1)C(=O)O	3000
CHEMBL838	396	CCOC(=O)[C@H](CCc1ccccc1)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	1.7
CHEMBL311471	397	CCN(C[C@H](CCc1ccccc1)C(=O)O)C(=O)N1Cc2ccccc2C[C@H]1C(=O)O	6.457
CHEMBL78341	398	C[C@H](NC(CCc1ccccc1)C(=O)O)C(=O)N(CC(=O)O)C1Cc2ccccc2C1	39.81
CHEMBL2115478	399	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2ccccc2[C@H]1C(=O)O	44
CHEMBL330316	400	C[C@H](CS)C(=O)N1C[C@@H](NC(=O)NC(CS)Cc2ccccc2)C[C@H]1C(=O)O	147
CHEMBL94173	401	CC[C@H](C)C(N[C@@H](C)C(=O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	22
CHEMBL430630	402	NCCCCC(N)C(=O)N[C@H](CCC(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)C(=O)O	12.02
CHEMBL51438	403	O=C(O)CCN1C(=O)[C@@H](NC(=O)[C@@H](CS)Cc2ccccc2)CCc2ccccc21	36
CHEMBL3954917	404	CC(C)C[C@H](S)C(=O)N1CCC[C@H]1C(=O)N1CCC[C@H]1C(=O)N(C)[C@@H](C)C(=O)O	640
CHEMBL409720	405	CC(C)CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	35
CHEMBL289267	406	O=C(O)CN1C(=O)[C@@H](N[C@@H](CSCc2ccccc2)C(=O)O)CCc2ccccc21	2.9
CHEMBL52733	407	C[C@H]1Oc2ccccc2N(CCC(=O)O)C(=O)[C@H]1NC(=O)[C@@H](S)Cc1ccccc1	18
CHEMBL54796	408	CCC(C)C(S)CC(Cc1ccccc1)C(=O)NC(Cc1ccc(O)cc1)C(N)=O	3500
CHEMBL378864	409	CN[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@@H](CCCN)C(=O)N1CCC[C@H]1C(=O)O	10000

CHEMBL331378	410	C[C@H](CSC(=O)c1ccccc1)C(=O)N1C[C@@H](Sc2ccccc2)C[C@H]1C(=O)O	0.4
CHEMBL55583	411	CCC(C)C(CC(S)Cc1ccccc1)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	800
CHEMBL55721	412	CCC(C)C(S)CC(Cc1ccccc1)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	4400
CHEMBL55520	413	CCC(C)C(S)CC(Cc1ccccc1)C(=O)N[C@H](Cc1ccc(O)cc1)C(=O)O	8000
CHEMBL51780	414	O=C(O)CN1C(=O)[C@@H](NC(=O)[C@@H](CS)Cc2ccccc2)CSc2ccccc21	17
CHEMBL1519	415	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1[C@H](C(=O)O)C[C@H]2CCCC[C@@H]21	0.93
CHEMBL299441	416	O=C(O)CCN1C(=O)[C@@H](NC(=O)[C@@H](S)Cc2ccccc2)CSc2ccccc21	14
CHEMBL2079671	417	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1C2CCCCC2C[C@H]1C(=O)O	100
CHEMBL381557	418	NCCC[C@H](NC(=O)[C@@H](N)Cc1c[nH]c2ccc(O)cc12)C(=O)N1CCC[C@H]1C(=O)O	4000
CHEMBL284898	419	CCOC(=O)C(CCc1ccccc1)OC(C)C(=O)N1C(C(=O)O)CC2CCCCC21	24
CHEMBL516325	420	CC(C)c1ccc(C(SC[C@H](NC(=O)[C@H](C)CS)C(=O)O)c2ccccc2)cc1	1700
CHEMBL89563	421	CC(C)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	25
CHEMBL311268	422	C[C@H](NC(CCCNC(=O)OCc1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	2.884
CHEMBL327799	423	C[C@H](O)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	665
CHEMBL173376	424	Cc1ccccc1C(=O)NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O	46
CHEMBL177394	425	COC(=O)[C@@H]1CCCN1C(=O)CCC(=O)C(Cc1ccccc1)NC(=O)c1ccccc1	82
CHEMBL172783	426	O=C(NC(Cc1ccccc1)C(=O)CCCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	42
CHEMBL328537	427	C[C@@H](CC(=O)[C@@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	3200
CHEMBL95564	428	C[C@H](CC(=O)[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	1
CHEMBL309941	429	C[C@H](CC(=O)[C@@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	3.02
CHEMBL2370854	430	C[C@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	2700
CHEMBL274553	431	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCCC1C(=O)NO)c1ccccc1	1600
CHEMBL131552	432	C[C@@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	3.2

CHEMBL366727	433	O=C(NC(Cc1ccccc1)/C(CCC(=O)N1CCC[C@H]1C(=O)O)=N/O)c1ccccc1	1200
CHEMBL4167252	434	O=C1/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc2N1c1ccccc1	18.97
CHEMBL2442647	435	C[C@H](NC(CCc1ccccc1)P(=O)(O)O)C(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21	40
CHEMBL173822	436	O=C(NC(Cc1ccc(O)cc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	4.7
CHEMBL172600	437	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CC(O)C[C@H]1C(=O)O)c1ccccc1	540
CHEMBL2115288	438	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2ccccc2C[C@@H]1C(=O)O	6100
CHEMBL1592	439	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2ccccc2C[C@H]1C(=O)O	8.3
CHEMBL2114322	440	CCOC(=O)[C@@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2ccccc2C[C@H]1C(=O)O	200
CHEMBL198316	441	O=C(N[C@H]1Cc2ccccc2C2CCC[C@H](C(=O)O)N2C1=O)[C@H](S)Cc1ccccc1	0.11
CHEMBL579	442	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1CC2(C[C@H]1C(=O)O)SCCS2	0.8
CHEMBL459959	443	C[C@H](CS)C(=O)N[C@@H](CSCc1ccc(I)cc1)C(=O)O	140
CHEMBL101409	444	C[C@H](N[C@@H](Cc1ccccc1)C(=O)O)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O	11
CHEMBL25782	445	CCOC(=O)C(CCc1ccccc1)OC(C)C(=O)N1Cc2ccccc2CC1C(=O)O	14
CHEMBL3960222	446	CC(C)C[C@H](S)C(=O)N1CCC[C@H]1C(=O)N1CCC[C@H]1C(=O)N1CC[C@H](C(=O)O)C1	49000
CHEMBL3932740	447	CC(C)C[C@H](S)C(=O)N1CCC[C@H]1C(=O)N1CCC[C@H]1C(=O)N1CCC[C@@H]1C(=O)O	12000
CHEMBL36503	448	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	36.31
CHEMBL93167	449	CC(C)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(C2CCCCC2)cc1)C(=O)O	122
CHEMBL42976	450	O=C(O)CN1C(=O)[C@@H](N[C@@H](CSCCc2ccccc2)C(=O)O)CCc2ccccc21	7
CHEMBL277270	451	O=C(O)[C@@H]1CC2(CN1C(=O)CP(=O)(O)CCCCc1ccccc1)SCCS2	3.981
CHEMBL108988	452	O=C(O)C1CC2(CN1C(=O)CP(=O)(O)CCCCc1ccccc1)SCCS2	12.59
CHEMBL54823	453	CCC(C)C(CC(S)CCc1ccccc1)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	2900
CHEMBL310869	454	NC(Cc1ccccc1)C(=O)N[C@H](CCC(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)C(=O)O	100

CHEMBL310139	455	N[C@@H](Cc1ccccc1)C(=O)N[C@H](CCC(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)C(=O)O	218.78
CHEMBL10973	456	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCCC1c1nn[nH]n1)c1ccccc1	22000
CHEMBL1907760	457	CCOC(=O)C(CCc1ccccc1)SC(C)C(=O)N1C2CCCCC2C[C@H]1C(=O)O	2000
CHEMBL90545	458	CC[C@@H](C)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	50
CHEMBL313419	459	CCCC[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	731
CHEMBL78996	460	CC(NC(=O)OCc1ccccc1)C(=O)N[C@H](CCC(=O)N1CCC[C@H]1C(=O)O)C(=O)O	389.05
CHEMBL296331	461	CCOC(=O)[C@H](Cc1c[nH]c2ccccc12)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	540
CHEMBL176970	462	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)OCc1ccccc1	240
CHEMBL100413	463	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O	1.7
CHEMBL100586	464	C[C@@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O	2400
CHEMBL430431	465	Cc1ccc(CSC[C@@H](NC(=O)[C@H](CS)Cc2ccc3ccccc3c2)C(=O)O)cc1	3100
CHEMBL1907754	466	CCOC(=O)C(CCc1ccccc1)SC(C)C(=O)N1Cc2ccccc2C[C@H]1C(=O)O	460
CHEMBL35419	467	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@@H](O)C[C@H]1C(=O)O	15.85
CHEMBL33025	468	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@H](O)C[C@H]1C(=O)O	0.871
CHEMBL4110944	469	CC(C)C[C@H](S)C(=O)N1C[C@@H](F)C[C@H]1C(=O)N1CCC[C@H]1C(=O)N1CC[C@@H](C(=O)O)C1	68000
CHEMBL10690	470	O=C(NC(Cc1ccccc1)C(=O)CCC(=O)N1CCCC1P(=O)(O)O)c1ccccc1	100000
CHEMBL309798	471	O=C(NC(CCc1ccccc1)P(=O)(O)CC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	72.44
CHEMBL310841	472	O=C(NC(Cc1ccccc1)CP(=O)(O)CC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	10
CHEMBL101469	473	CCCCCCCC[C@H](N[C@@H](C)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O)C(=O)O	1.5
CHEMBL40420	474	CC(C)(C)OC(=O)NCCCC[C@H](N[C@H]1CCc2cccc2N(CC(=O)O)C1=O)C(=O)O	4
CHEMBL301703	475	O=C(NC(Cc1ccc(O)cc1)C(=O)O)C(Cc1ccccc1)CC(S)Cc1ccccc1	2400
CHEMBL1907752	476	CCOC(=O)C(CCc1ccccc1)[S+]([O-])C(C)C(=O)N1C2CCCCC2C[C@H]1C(=O)O	220

CHEMBL175110	477	O=C(O)C(Cc1ccc(O)cc1)NC(=O)C(Cc1ccccc1)C(S)CCc1ccccc1	1600
CHEMBL4163792	478	O=C1/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc2N1C(=O)c1ccccc1	8.841
CHEMBL412893	479	CCOC(=O)C(CCc1ccccc1)NC(C)C(=O)N1Cc2ccccc2CC1C(=O)OCC	440
CHEMBL100826	480	CC[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O	2.3
CHEMBL2112769	481	C[C@H](N[C@H](C(=O)O)[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	7.7
CHEMBL2111942	482	C[C@H](N[C@@H](C(=O)O)[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	1400
CHEMBL4175518	483	COc1ccc(N2C(=O)/C(=N/C(=S)NCc3cc4ccccc4cc3O)c3ccccc32)cc1	9.82
CHEMBL4160421	484	O=C1/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc2N1CCBr	4.11
CHEMBL58042	485	COc1cc2c(cc1OC)CN(C(=O)C(C)NC(CCc1ccccc1)C(=O)O)C(C(=O)O)C2	3
CHEMBL2097001	486	C[C@H](C[Si](O)(O)[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	3.8
CHEMBL328814	487	C[C@@H](C[Si](O)(O)[C@@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCC[C@H]1C(=O)O	72
CHEMBL4163798	488	O=C1/C(=N\C(=S)NCc2cc3ccccc3cc2O)c2ccccc2N1c1ccccc1Cl	15.11
CHEMBL2153745	489	O=C(N[C@@H](CCc1ccccc1)CP(=O)(O)CC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	12
CHEMBL366455	490	O=C(CCC(=O)C(Cc1ccccc1)NC(=O)c1ccccc1)N[C@H](Cc1ccccc1)C(=O)O	28
CHEMBL367174	491	CC(C)CC(S)CC(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	2800
CHEMBL176960	492	CC(C)C(NC(=O)C(Cc1ccccc1)C(C)S)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	1300
CHEMBL29031	493	Cc1ccc(S(=O)(=O)N[C@@H](Cc2ccccc2)C(=O)NCC(=O)N2CCC[C@H]2C(=O)O)cc1	67000
CHEMBL80503	494	O=C(N[C@H](CCC(=O)N1CCC[C@H]1C(=O)O)C(=O)O)C1CCCN1C(=O)OCc1ccccc1	2511.89
CHEMBL328378	495	CCCCCCCC[C@H](N[C@@H](CC)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O)C(=O)O	2.1
CHEMBL284345	496	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N(CC(=O)O)c1ccccc1	1.072
CHEMBL3264007	497	CN[C@H](C(=O)N[C@@H](Cc1ccccc1)C(=O)N[C@@H](Cc1ccc(O)cc1)P(=O)(O)O)C(C)C	220
CHEMBL4166152	498	Cc1ccc(C(=O)N2C(=O)/C(=N\C(=S)NCc3cc4ccccc4cc3O)c3ccccc32)cc1	24.22
CHEMBL2112768	499	C[C@H](N[C@@H](C[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	3

CHEMBL2111940	500	C[C@H](N[C@H](C[C@H](Cc1ccccc1)NC(=O)c1ccccc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	4.8
CHEMBL328847	501	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O	7900
CHEMBL4174072	502	COc1ccc(CN2C(=O)/C(=N\C(=S)NCc3cc4ccccc4cc3O)c3ccccc32)cc1	7.64
CHEMBL93180	503	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](Cc1ccccc1)NCP(=O)(O)O	117
CHEMBL173849	504	COc1ccc(CC(NC(=O)c2ccccc2)C(=O)CCC(=O)N2CCC[C@H]2C(=O)O)cc1OC	1.4
CHEMBL285935	505	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N(CC(=O)O)C1CCCCC1	0.2951
CHEMBL560409	506	O=C(CC(S)C(F)(F)F)NC(Cc1ccccc1)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	280
CHEMBL3264009	507	CC[C@H](C)[C@H](NC(C)=O)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](Cc1ccc(O)cc1)P(=O)(O)O	76
CHEMBL433442	508	CS[C@H]1C[C@@H](C(=O)O)N(C(=O)[C@H](CCCCN)OP(=O)(O)CCCCc2ccccc2)C1	1.905
CHEMBL289022	509	CS[C@@H]1C[C@@H](C(=O)O)N(C(=O)[C@H](CCCCN)OP(=O)(O)CCCCc2ccccc2)C1	0.5248
CHEMBL172978	510	CCC(C)C(NC(=O)C(Cc1ccccc1)C(C)S)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	300
CHEMBL174066	511	CCCCC(NC(=O)C(Cc1ccccc1)C(C)S)C(=O)NC(Cc1ccc(O)cc1)C(=O)O	6100
CHEMBL89547	512	C[C@H](CS)C(=O)N1C[C@@H](NC(=O)C(CS)Cc2ccc(-c3ccccc3)cc2)C[C@H]1C(=O)O	696
CHEMBL109178	513	CC(OP(=O)(O)C(CCc1ccccc1)NC(=O)c1ccccc1)C(=O)N1CCCC1C(=O)O	39.81
CHEMBL330507	514	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](Cc1cccs1)NCP(=O)(O)O	385
CHEMBL284734	515	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1c2ccccc2C[C@H]1C(=O)O	0.2951
CHEMBL328811	516	CCCCCCCC[C@H](N[C@@H](C)C(=O)N1C(=O)N(Cc2ccccc2)C[C@H]1C(=O)O)C(=O)OCC	2200
CHEMBL3264008	517	CN[C@H](C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)N[C@@H](Cc1ccc(O)cc1)P(=O)(O)O)C(C)C	54
CHEMBL295690	518	NCCCC[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1CC2(C[C@H]1C(=O)O)SCCS2	0.5
CHEMBL91399	519	Cc1ccccc1C[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	498
CHEMBL92126	520	C[C@H](N[C@@H](Cc1ccccc1)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)P(=O)(O)O	44
CHEMBL329530	521	C[C@H](c1ccccc1)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	138

CHEMBL432689	522	O=C(O)CN1C(=O)[C@@H](N[C@@H](CCCCNC(=O)OCc2ccccc2)C(=O)O)CCc2ccccc21	13
CHEMBL3350318	523	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2cc(OC)c(OC)cc2C[C@@H]1C(=O)O	90000
CHEMBL2115075	524	CCOC(=O)[C@@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2cc(OC)c(OC)cc2C[C@H]1C(=O)O	1100
CHEMBL1165	525	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N1Cc2cc(OC)c(OC)cc2C[C@H]1C(=O)O	2.6
CHEMBL136312	526	CC(NC(C(=O)O)C1CCCN1C(=O)NC(Cc1ccccc1)C(=O)O)C(=O)N1CCCC1C(=O)O	5.4
CHEMBL420391	527	NCCCCC(NC(=O)OCc1ccccc1)C(=O)N[C@H](CCC(=O)N1CCC[C@H]1C(=O)O)C(=O)O	12.02
CHEMBL93785	528	COc1ccccc1C[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	160
CHEMBL172015	529	O=C(NC(Cc1cccnc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)OCc1ccccc1)c1ccccc1	15
CHEMBL35309	530	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@@H](c2ccccc2)C[C@H]1C(=O)O	0.1318
CHEMBL284272	531	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@H](c2ccccc2)C[C@H]1C(=O)O	338.84
CHEMBL329464	532	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](Cc1ccc(Cl)cc1)NCP(=O)(O)O	473
CHEMBL80665	533	C[C@H](NC(CCC(=O)Nc1ccc(I)cc1)C(=O)O)C(=O)N1CCC[C@H]1C(=O)O	0.2291
CHEMBL110925	534	CC(NC(CCC(=O)Nc1ccc(I)cc1)C(=O)O)C(=O)N1CCCC1C(=O)O	0.7943
CHEMBL3264010	535	CC[C@H](C)[C@H](NC(C)=O)C(=O)N[C@@H](Cc1ccccc1)C(=O)N[C@@H](Cc1ccc(O)cc1)P(=O)(O)O	7
CHEMBL425383	536	Cc1ccc(S(=O)(=O)N(C)NC(=O)C2C[C@@H](S)CN2S(=O)(=O)c2ccc3ccccc3c2)cc1	14600
CHEMBL35561	537	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N(CC(=O)O)c1ccc2c(c1)OCO2	0.1148
CHEMBL287518	538	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@H](C2CCCCC2)C[C@H]1C(=O)O	3.236
CHEMBL286339	539	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@@H](C2CCCCC2)C[C@H]1C(=O)O	0.4898
CHEMBL415881	540	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](CCCCN)C(=O)N1CC2(C[C@H]1C(=O)O)SCCS2	200
CHEMBL256895	541	O=C(O)[C@@H](O)CN(Cc1ccc2c(c1)Cc1ccccc1-2)C(=O)N[C@@H](Cc1ccc2ccccc2c1)C(=O)O	150
CHEMBL289618	542	CCOC(=O)[C@H](CCCCNC(=O)OCc1ccccc1)N[C@H]1CCc2ccccc2N(CC(=O)O)C1=O	15

CHEMBL328463	543	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](Cc1ccc2c(c1)OCO2)NCP(=O)(O)O	147
CHEMBL366503	544	C[C@H](OCc1ccccc1)[C@H](NCP(=O)(O)O)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O	656
CHEMBL60702	545	CCOC(=O)C(CCc1ccccc1)NC(C)C(=O)N1Cc2cc(OC)c(OC)cc2CC1C(=O)OCC	520
CHEMBL176729	546	Cc1ccccc1C(=O)NC(Cc1ccccc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)OCc1ccccc1	1600
CHEMBL4161811	547	CCCCCCC[C@H](N)[C@H](O)C(=O)N(C)[C@@H](CC[S+](C)[O-])C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O	31000
CHEMBL172801	548	O=C(NC(Cc1ccc(OCc2ccccc2)cc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)O)c1ccccc1	10
CHEMBL78353	549	O=C(N[C@H](CCC(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)C(=O)O)C1CCCN1C(=O)OCc1ccccc1	15.14
CHEMBL284843	550	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@@H](Cc2ccccc2)C[C@H]1C(=O)O	0.8913
CHEMBL34650	551	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1CC2(C[C@H]1C(=O)O)SCCS2	0.631
CHEMBL92326	552	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](Cc1ccc2ccccc2c1)NCP(=O)(O)O	53
CHEMBL445561	553	CSC1(SC)C[C@@H](C(=O)O)N(C(=O)[C@H](CCCCN)OP(=O)(O)CCCCc2ccccc2)C1	2.57
CHEMBL1233799	554	CC[C@H](C)[C@H](NC(C)=O)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)N[C@@H](Cc1ccc(O)cc1)P(=O)(O)O	11
CHEMBL35682	555	COc1ccc(N(CC(=O)O)C(=O)[C@H](CCCCN)OP(=O)(O)CCCCc2ccccc2)cc1OC	0.2884
CHEMBL258130	556	O=C1c2ccccc2-c2ccc(CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc3ccc4ccccc4c3)C(=O)O)cc21	39
CHEMBL177000	557	COC(=O)c1ccccc1NC(=O)CN(C)C(=O)C1C[C@@H](S)CN1S(=O)(=O)c1ccc2ccccc2c1	36600
CHEMBL88328	558	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@H](NC(=O)C(CS)Cc2ccccc2)C[C@H]1C(=O)O	253
CHEMBL312918	559	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@@H](NC(=O)C(CS)Cc2ccccc2)C[C@H]1C(=O)O	42
CHEMBL41289	560	CC(C)CC(NP(=O)(O)OC1OC(C)C(O)C(O)C1O)C(=O)NC(Cc1c[nH]c2ccccc12)C(=O)O	5.012
CHEMBL431707	561	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@H](Sc2ccccc2)C[C@H]1C(=O)O	0.7762

CHEMBL34832	562	NCCCC[C@H](OP(=O)(O)CCCCc1ccccc1)C(=O)N1C[C@@H](Sc2ccccc2)C[C@H]1C(=O)O	0.5248
CHEMBL310624	563	NCCCCC(NC(=O)OCc1ccccc1)C(=O)N[C@H](CCC(=O)N1c2ccccc2C[C@H]1C(=O)O)C(=O)O	2511.89
CHEMBL329801	564	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@@H](NC(=O)NC(CS)Cc2ccccc2)C[C@H]1C(=O)O	43
CHEMBL422419	565	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@H](OC(=O)NC(CS)Cc2ccccc2)C[C@H]1C(=O)O	506
CHEMBL418959	566	O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@H](Cc1ccc(-c2ccccc2)cc1)NCP(=O)(O)O	414
CHEMBL311757	567	NCCCCC(NC(=O)OCc1ccccc1)C(=O)N[C@H](CCC(=O)N1[C@H](C(=O)O)C[C@@H]2CCCC[C@@H]21)C(=O)O	3.467
CHEMBL369590	568	O=C(CCC(=O)C(Cc1ccccc1)NC(=O)c1ccccc1)NC(Cc1ccccc1)C(=O)OCc1ccccc1	180
CHEMBL3039598	569	CCC(=O)O[C@@H](O[P@](=O)(CCCCc1ccccc1)CC(=O)N1C[C@H](C2CCCCC2)C[C@H]1C(=O)O)C(C)C	1
CHEMBL273140	570	O=C1c2ccccc2C(=O)c2cc(CN(C[C@H](O)C(=O)O)C(=O)N[C@@H](Cc3ccc4ccccc4c3)C(=O)O)ccc21	40
CHEMBL271718	571	O=C(O)[C@@H](O)CN(CCN1C(=O)c2cccc3cccc(c23)C1=O)C(=O)N[C@@H](Cc1ccc2ccccc2c1)C(=O)O	1980
CHEMBL2370850	572	C[C@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)[C@H](CCCCN)NC(=O)[C@@H]1CCC(=O)N1)C(=O)N1CCC[C@H]1C(=O)O	50
CHEMBL78629	573	O=C(NC(Cc1ccccc1)C(=O)N[C@H](CCC(=O)N1c2ccccc2C[C@H]1C(=O)O)C(=O)O)OCc1ccccc1	0.2512
CHEMBL4067185	574	C[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@@H](NC(=O)C[C@H](N)Cc2cc(F)c(F)cc2F)C[C@H]1C(=O)O	158
CHEMBL369809	575	CC(C)(C)OC(=O)[C@@H]1CCCN1C(=O)CCC(=O)C(Cc1ccc(OCc2ccccc2)cc1)NC(=O)c1ccccc1	8600
CHEMBL58340	576	CCOC(=O)C(CCc1ccccc1)NC(C)C(=O)N1Cc2cc(OC)c(OC)cc2CC1C(=O)OCc1ccccc1	300
CHEMBL4086264	577	CC[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@@H](NC(=O)C[C@H](N)Cc2cc(F)c(F)cc2F)C[C@H]1C(=O)O	8.6
CHEMBL434617	578	CC(NC(C(=O)O)C1CCCN1C(=O)NC(CCCCN)C(=O)N1C(=O)CCC1C(=O)O)C(=O)N1CCCC1C(=O)O	8.5
CHEMBL4078112	579	CCC[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@@H](NC(=O)C[C@H](N)Cc2cc(F)c(F)cc2F)C[C@H]1C(=O)O	51

CHEMBL264538	580	C[C@H](NC(=O)[C@H](Cc1c[nH]c2ccccc12)NC(=O)[C@H](CCCCN)NC(=O)[C@@H]1CCC(=O)N1)C(=O)N1CCC[C@H]1C(=O)O	730
CHEMBL172356	581	O=C(NC(Cc1ccc(OCc2ccccc2)cc1)C(=O)CCC(=O)N1CCC[C@H]1C(=O)OCc1ccccc1)c1ccccc1	86
CHEMBL4090635	582	CCCC[C@H](N[C@@H](CCc1ccccc1)C(=O)O)C(=O)N1C[C@@H](NC(=O)C[C@H](N)Cc2cc(F)c(F)cc2F)C[C@H]1C(=O)O	279
CHEMBL327401	583	C[C@H](NC(=O)[C@H](Cc1c[nH]c2ccccc12)NC(=O)[C@H](CCCCN)NC(=O)[C@@H](N)CCC(=O)O)C(=O)N1CCC[C@H]1C(=O)O	60
CHEMBL449392	584	O=C(O[C@@H]1O[C@@H]2COC(=O)c3cc(O)c(O)c(O)c3-c3c(cc(O)c(O)c3O)C(=O)O[C@H]([C@H]1O)[C@@H]2O)c1cc(O)c(O)c(O)c1	3700000
CHEMBL274550	585	NCCCCC(NC(=O)C1CCC1)C(=O)NC(Cc1ccccc1)C(=O)CCC(=O)N1CCCC1C(=O)O.O=C(O)C(F)(F)F	7
CHEMBL263460	586	CC(NC(C(=O)O)C1CCCN1C(=O)NC(Cc1ccccc1)C(=O)NC(Cc1ccccc1)C(=O)O)C(=O)N1CCCC1C(=O)O	2.9
CHEMBL10521	587	CC(CC(=O)C(Cc1ccccc1)NC(=O)C(CCCCN)NC(=O)C1CCC1)C(=O)N1CCCC1C(=O)O.O=C(O)C(F)(F)F	3
CHEMBL275702	588	NCCCCC(NC(=O)C1CCC1)C(=O)NC(Cc1ccccc1)C(=O)CCC(=O)N1CCCC1C(=O)NO.O=C(O)C(F)(F)F	11
CHEMBL43370	589	CS(=O)(=O)N[C@@H](CCCCN)C(=O)N[C@@H](Cc1ccccc1)P(=O)(O)CC1(C(=O)N[C@@H](Cc2c[nH]c3ccccc23)C(=O)O)CCCC1	2.5
CHEMBL4294217	590	O=C(N[C@H]1CCS[C@H]2CCC[C@@H](C(=O)O)N2C1=O)[C@H](Cc1ccccc1)SS[C@@H](Cc1ccccc1)C(=O)N[C@H]1CCS[C@H]2CCC[C@@H](C(=O)O)N2C1=O	2200
CHEMBL506069	591	O=C1O[C@H]2[C@@H]3OC(=O)c4cc(O)c(O)c(O)c4-c4c(cc(O)c(O)c4O)C(=O)OC[C@H]2O[C@@H](OC(=O)c2cc(O)c(O)c(O)c2)[C@@H]3OC(=O)c2cc(O)c(O)c3c2[C@@H]2C1=CC(=O)[C@](O)(O3)C2(O)O	400000