APPLICATION OF MACHINE LEARNING METHODS ON VARIOUS HISTONE DEACETYLASE 10 (HDAC10) INHIBITORS: IDENTIFICATION OF MOLECULAR FINGERPRINTS AND BINDING MODE ANALYSIS

Submitted by

ARIJIT BHATTACHARYA

EXAM ROLL NO.: M4PHC23004

CLASS ROLL NO.: 002111402018

REG. NO.: 160252 of 2021-2022

Department of Pharmaceutical Technology
Jadavpur University

Session- ***2021-2023***

Under The Guidance Of

Dr. Shovanlal Gayen

Laboratory of Drug Design and Discovery
Department of Pharmaceutical Technology
Jadavpur University, Kolkata-700032

Thesis submitted in partial fulfilment of the requirements for the Degree of Master of Pharmacy
Department of Pharmaceutical Technology
Faculty of Engineering and Technology
Jadavpur University, Kolkata

2023

Jadavpur University

Jadavpur, Kolkata-700032

CERTIFICATE OF APPROVAL

This is to certify that ***Arijit Bhattacharya*** (Exam Roll No. **M4PHC23004**, Reg. No. **160250** of **2021-2022**) has sincerely carried out the research work on the subject entitled **"*Application of Machine Learning Methods on Various Histone Deacetylase 10 (HDAC10) Inhibitors: Identification of Molecular Fingerprints and Binding Mode Analysis*"** under the supervision of **Dr. Shovanlal Gayen**, Laboratory of Drug Design and Discovery, Department of Pharmaceutical Technology of Jadavpur University. He has incorporated his findings in this thesis submitted by him in partial fulfilment of the requirements for the degree of **Masters of Pharmacy** (Pharmaceutical Technology) of Jadavpur University. He has carried out the research work independently and sincerely with proper care and attention to our entire satisfaction.

27/9/23.

Head of the Department

Department of Pharmaceutical Technology

Jadavpur University

Kolkata-700032

Head
Dept. of Pharmaceutical Technology
Jadavpur University
Kolkata-700032, W.B. India

Shovanlal Gayen

Dr. Shovanlal Gayen

Laboratory of Drug Design and Discovery

Department of Pharmaceutical Technology

Jadavpur University

Kolkata-700032

DR. SHOVANLAL GAYEN
Assistant Professor
Dept. of Pharmaceutical Technology
Jadavpur University
Kolkata - 700 032, INDIA

Saswati Mazumder 26.9.23

Dean

Faculty of Engineering and Technology

Jadavpur University

Kolkata-700032

DEAN
Faculty of Engineering & Technology
JADAVPUR UNIVERSITY
KOLKATA-700 032

Acknowledgement

The outcome of this thesis required a lot of guidance and assistance from many people. I am extremely fortunate to have had these all along the completion of my work. Whatever I have done is only due to such guidance and assistance and I would not forget to thank them.

I am highly obliged and like to express my deep gratitude and profoundness to my respected supervisor **Dr. Shovanlal Gayen**, Department of Pharmaceutical Technology, Jadavpur University, Kolkata for his excellent and constant guidance and help, endless encouragement, thoughtful and freedom and cooperation throughout the term paper till its successful completion. I am greatly indebted to his motivation, fruitful suggestions, and inspiration. This feat was possible only because of the unconditional support provided by Sir, a person with an amicable and positive belief.

I owe my deep respect to **Prof. Amalesh Samanta**, Head of the Department, Department of Pharmaceutical Technology, Jadavpur University, Kolkata for all the necessary help and encouragement. I would like to convey my sincere gratitude to All India Council for Technical Education (AICTE) and Jadavpur University for their financial and equipment support for my M. Pharm course.

I also acknowledge **Prof. Tarun Jha**, Natural Science Laboratory, Jadavpur University, India for the continuous encouragement, necessary help, and support to perform my work.
I express my sincere thanks to **Prof. Biswajit Mukherjee, Prof. Lakshmi Kanta Ghosh, Prof. Sanmoy Karmakar, Prof. Pallab Kanti Halder, Prof. Kunal Roy, Prof. Tapan Kumar Maity, Prof. Subhas Chandra Mandal, Prof. Saikat Dewanjee, Dr. Emdad Hossain, Dr. Ketosetuo Kuotsu, Dr. Manas Bhowmik, Dr. Probir Kumar Ojha, Dr. Tapan Kumar Giri, Dr. Kajal Ghoshal, Dr. Nilanjan Ghosh, Dr. Sanchita Mandal, Dr. Ranu Biswas, Dr. Nilanjan Adhikari, Dr. Kajal Ghoshal,** and **Mr. Koushik Mukherjee** sir for their co-operation, help and support.

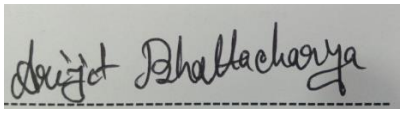
I am extremely grateful to **Dr. Abdul Amin, Ms. Samima Khatun, Mr. Suvankar Bannerjee, Mr. Sandip Baidya,** and **Mr. Sanjib Das** for their guidance and support which assisted me in gathering knowledge about the different aspects of this work. I have learnt a lot from them. I also like to thankall the research scholars during my stay in the department especially **Mr**. **Arnab Bhattacharjee, Mr. Avik Maji, Mr. Avik Paul**, and **Mrs. Ajeya Samanta** for their support. I would express my sincere thanks to my laboratory colleague **Mr.**

Sourav Sardar and my juniors **Totan Das, Rinki Prasad Bhagat** and **Indrasis Das Gupta** from the Laboratory of Drug Design and Discovery**,** Department of Pharmaceutical Technology, Jadavpur University, Kolkata-700032. I would like to thank **Mr. Pravat Kumar**, Department of Computer Science, Institute of Science, Banaras Hindu University, Varanasi, 221005, India for his necessary help.

I express my wholehearted thanks to my friends **Pankaj Nahata, Vinayak Ghosh, Chiranjib Bhattacharyya, Rahul Jana, Jigme Sangay Dorjay Tamang, Sourav De, Souvik Pore, Pankaj Paul, Rideb Chakraborty, Sumi Barman, Sourin Nahar,** and **Naureen Afrose** for always being there and motivating me.

Nobody has been more important to me in the pursuit of this project than the members of my family**.** I would like to express my deep respect to my father **Mr. Ashok bhattacharya**, mother **Mrs. Samapti Bhattacharya** and my sister **Ms. Arundhati Bhattacharya,** my friends and relatives for their continuous help, love, encouragement, and moral support throughout the period of my work.

Finally, I would like to thank one and all who were directly or indirectly there to help me for the successful completion of my thesis, as well as expressing my apology that I could not mention personally one by one.



[Arijit Bhattacharya]

Date: 20.09.2023

Place: Department of Pharmaceutical Technology, Jadavpur University,
Kolkata.

ABBREVIATIONS

BA - Biological activity

CADD - Computer aided drug design

DS - Discovery studio

ECFP - Extended-connectivity of fingerprint

FBDD - Fragment based drug design

FCFP - Functional-class fingerprint

FN - False negative

FP - False positive

HDAC -Histone deacetylase

HTS - High throughput screening

LBDD - Ligand-based drug design

LDA - Linear discriminant analysis

ML - Machine learning

MW – Molecular weight

QSAR - Quantitative structure activity relationship

ROC - Receiver operating characteristics

RF – Random Forest

RP - Recursive Partitioning

SBDD - Structure based drug design

TN - True negative

TP - True positive

Declaration of Originality and Compliance Of Academic Ethics

I hereby declare that this thesis contains a literature survey and original research work performed by me (Arijit Bhattacharya) as a part of my Masters of Pharmacy studies. All the information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have cited and referenced the materials and results that are not original to this work.

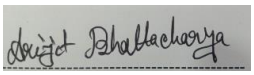
Name: Arijit Bhattacharya

Exam Roll Number: M4PHC23004

Class Roll Number: 002111402018

Registration Number: 160250 *of 2021-2022*

Thesis Title: "*APPLICATION OF MACHINE LEARNING METHODS ON VARIOUS HISTONE DEACETYLASE 10 (HDAC10) INHIBITORS: IDENTIFICATION OF MOLECULER FINGERPRINTS AND BINDING MODE ANALYSIS*".



20.09.23

Signature with Date ***(ARIJIT BHATTACHARYA)***

Dedicated to

My Parents, Teachers, Seniors

and

Friends

Contents

Preface

One of the hottest topics to mankind in the modern era is "oncology," which refers to a collection of cancer-related ailments that cause abnormal and unchecked cellular proliferation in living things. Currently, cancer is one of the leading causes of death, and the number of people affected by it is growing quickly. Additionally, cancer patients require expensive and complicated treatments, which puts our society and economy in danger.

There are several internal and external variables that contribute to the development and spread of cancer. Numerous external influencers, such as pollution, malnutrition, radiation, lifestyle choices, etc. are linked to the transformation of the pathophysiological system in addition to internal issues such as genetic mutation, improper hormone regulation, etc.

Histone deacetylase 10 (HDAC10) is a polyamine deacetylase type class IIB member of the HDAC family that plays a key role in the pathogenesis of several illnesses linked with tumours and non-tumour diseases, such as renal cell carcinoma, prostate cancer, immunoglobulin A nephropathy (IgAN), intracerebral haemorrhage, human immunodeficiency virus (HIV) infection and schizophrenia. Although there are many pan-HDAC and HDAC10 inhibitors available, the absence of selective HDAC10 inhibitors that are effective anticancer agents is a significant drawback for the treatment of malignancies and other disorders linked to HDAC10. The inclusion of a capping group, a linker moiety, and a zinc-binding group are three structural properties that the majority of HDAC10 inhibitors share in order to create an effective HDAC10 inhibitor. In contrast, the hydroxamic acid moiety has been demonstrated to be a more effective zinc binding group for HDAC.

Additionally, the quantitative structure-activity relationship (QSAR) methodology quantifies the biological activity of molecules in relation to their molecular structure through the development of mathematical correlation to predict the key structural factors influencing the activity of these compounds. As a result, a comparative molecular modelling technique is used in this study to examine a set of reported HDAC10 inhibitors that have a wide range of inhibitory activity and contain the hydroxamate group as the zinc-binding motif.

This study has two main goals. Firstly, to develop some classical machine learning based models which can be further utilized for classify the unknown compounds as active or inactive HDAC10 inhibitors. Along with that, our second motto is to investigate the significant structural attributes of the potent HDAC10 inhibitors through different statistically validated

machine learning-based QSAR modelling like Bayesian classification and Recursive partitioning study. The molecular docking study was also performed to check the binding mode analysis of important structural attributes that are essential for controlling the HDAC10 inhibitory activity. Some of the key structural variables along with their interacting binding residue inside the active site of HDAC10 protein responsible for regulating the activity were discovered in this work.

(Arijit Bhattacharya)

Introduction

Histone deacetylases (HDACs) are called the "REMOVER" enzymes as these eliminate the -acetate moiety from acetylated ε-amino groups of histone lysine and other non-histone proteins [1-4]. They belong to the category of Zn2+ or nicotinamide adenine dinucleotide (NAD+)-dependent proteolytic enzymes, involved in transcriptional subduing as well as chromatin condensation. HDACs execute post-translational changes, such as ubiquitination and methylation, and can exert effects on gene transcription by increasing the interactivity between DNA and histone [5-7]. The deacetylation nature of the HDACs takes part as a modulator or regulator in several types of bio-signaling pathways [8]. Thus, they have a crucial pathophysiological role in several deadliest maladies including neurodegenerative dysfunctions, inflammation, metabolic disorders, autoimmune disorders, and cancers etc. [8-11]. The histone deacetylases family comprises 18 members [12-13] which are further divided into four classes depending on their homological sequences with yeast HDACs [14-15]. The four classes are class 1 HDACs (HDAC 1,2,3 and 8) [14]; class II HDACs, which are further subclassified into class IIa (HDAC 4,5,7, and 9) and class IIb (HDAC 6 and 10) [4]; class III which resembles yeast SIR2 (SIRT 1-7) protein structures and Class IV that comprises only HDAC11 [16].

HDAC10 is an essential representative of class IIb HDACs and is configurationally indistinguishable from the HDAC6 isotype. It is localized in the cytoplasm and has poor lysine deacetylase activity. It can identify polyamines as substrate and thus, named as polyamine deacetylase [2-3, 17-18]. The HDAC10 gene is confined to chromosome number 22 [18], includes 20 exons with two spliced transcripts [19-20] and contains an N-terminal catalytic domain and a leucine-rich domain with a C-terminal end. The N-terminal catalytic domain of HDAC10 is analogous to the deacetylase domain of other class II HDACs, but the C-terminal catalytic domain does not contain any residues that are imperative for enzymatic actions. The presence of both catalytic domains in the HDAC10 structure may confer resistance to trapoxin B and sodium butyrate [18-19]. Christianson and his co-workers have found that Zn^{2+} binding groups are present in HDAC10 and have additional interactions with inhibitors when accommodating them [21]. They have also illustrated the structure of HDAC10 complexed with trifluoromethylketone (AAT) (PDB ID:1TD7), where canonical catalytic domain (PDAC domain) and smaller, non-catalytic (ΨDAC) domain assemble with unique butterfly-like architecture (Figure 1) [22].

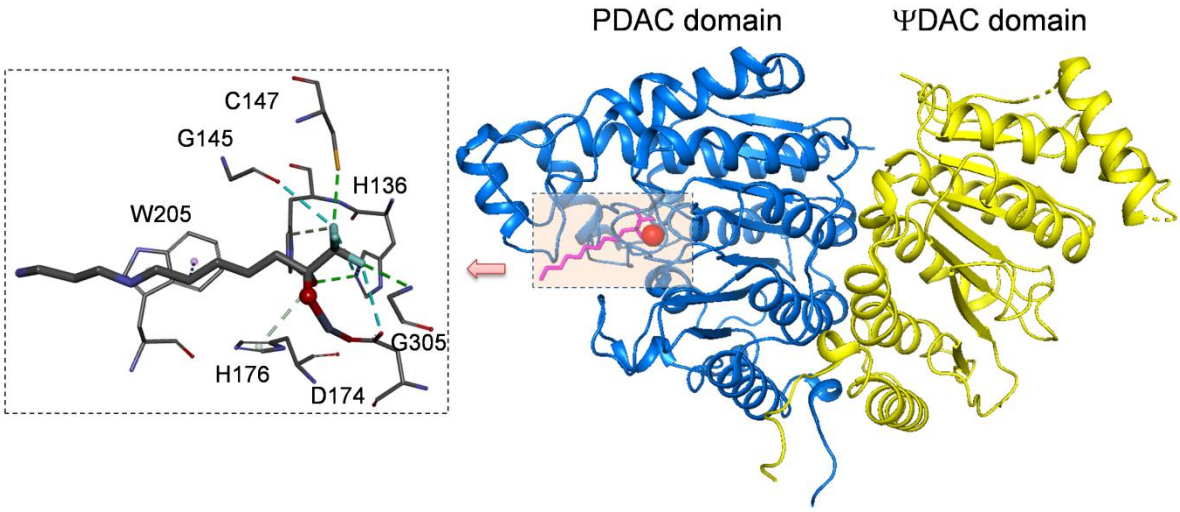


Figure 1. Structure of *Danio rerio* HDAC10 (PDB: 1TD7) showing the catalytic polyamine deacetylase (PDAC) domain (blue) and noncatalytic pseudo-deacetylase ((ΨDAC) domain (yellow). The important amino acids for the interactions with the trifluoromethylketone inhibitor are shown in the box.

Two well-known researchers, Hanahan and Weinberg established six and an additional four different markers of tumours, respectively [23,24]. These theories regarding different markers of tumours progressed the better view of the diverse types of tumours.

HDAC10 has a critical role in the regulation of growth in cancer cells. It has been observed from several pieces of literature that it guides cell proliferation in non-small-cell lung cancer (NSCLC) [25]. On the other hand, HDAC10 suppresses the growth of cancer cells in other malignant types of tumours, including lung adenocarcinoma (LA) [26] and renal cell carcinoma (RCC) [27]. The main driving molecule is Akt for the PI3K-AKT pathway, involved in survival, cell proliferation, and other functions [28]. In the same way, HDAC10 overexpression significantly facilitates NSCLC cell growth by regulating the Ser473 phosphorylation of AKT [25]. For LA, deletion in HDAC10 accelerates the KRAS-driven cancer progression both *in vivo* and *in vitro*. Mainly, through the transformation of the growth-factor β (TGF-β) pathway, deletion of HDAC10 promotes the expression of sex-determining region Y box protein 9 (SOX9), which significantly up-regulates the expression of SLUG, as well as CD44. These processes contribute to the growth of lung cancer spheres via SOX9-mediated stem-like properties, suggesting that HDAC10-TGF-β-SOX9-SLUG/CD44 pathway participates as an

essential modulator in lung adenocarcinoma [26]. AMP-activated protein kinase (AMPK) controls several biological activities in tumours that mediate through liver kinase B1 (LKB1), such as cell survival and transcription, via the mTOR pathway [29]. Inducing the LKB1–AMPK signalling, phosphorylated HDAC10 is transported from the nucleus to the cytoplasm and subsequently elevates the expression of glucose-6-phosphate dehydrogenase (G6PD) level. It also decreases the cellular reactive oxygen species (ROS) level and promotes LC cell proliferation [30]. The abovementioned scenarios suggest that the LKB1–AMPK-HDAC10-G6PD-ROS pathway may have a crucial role in tumour cell proliferation. However, in RCC cells, suppressed expression of HDAC10 significantly promotes the phosphorylation of β-catenin and thus plays a part in anti-proliferation [27]. A disquieted cell cycle is also a way for growth-regulation of cancer cells [31]. In previous, several studies proved that the inhibition of histone H3 deacetylation around the let-7f-2/miR-98 promoter, HDAC10 suppresses HMGA2 expression which is mostly directed towards cyclin A2 promoter, and additionally it inhibits the cyclin A2 transcription [26,32]. The signalling pathways: HDAC10-let-7f-2/miR-98-HMGA2-cyclin A2 hampers the G2/M transition and finally prevents LC cell proliferation. Both cell cycle inhibitors (such as P21 and P27) and promoters (such as cyclins E1 and D1) play a vital role in cancer progression [33-35]. HDAC10 plays an essential role by inhibiting the expression of P27, and P21 and enhancing that of cyclins D1 and E1 [25].

HDAC10 related to cell apoptosis

An insufficient amount of apoptosis will prevent cancer cells from dying [36]. A variety of proteins and signalling pathways are involved in the death of cells. It is known that overexpression in anti-apoptotic proteins (such as those in the Bcl-2 family), as well as down-regulated proteins (like Bid, BIK and BAK), might break the equilibrium between apoptosis and anti-apoptosis [37,38]. By targeting AKT, HDAC10 alters the expression of B-cell lymphoma-2 (BCL2) as well as BCL2 antagonist/killer (BAK), which induces apoptosis in LC [25]. In colorectal cancer, inhibited HDAC10 expression promotes cell apoptosis by depleting transcription factor 7 like 2 (TCF7L2), which impairs the Wnt pathway [39]. ROS produced as a result of mitochondrial damage or oxidative stress, may trigger caspases, and stimulate apoptosis [40,41]. According to Lee *et al*., a low level of HDAC10 in gastric cancer may activate proapoptotic molecules including caspase-3, caspase-9, and Bid through the thioredoxin interacting protein (TXNIP)-induced ROS signalling pathway [42]. Earlier

research has determined limited mechanisms in lung, colorectal and gastric tumours, including the HDAC10-AKT-BCL2-BAK pathway, the HDAC10-TCF7L2-Wnt pathway and the HDAC10-TXNIP-ROS-caspase-3/caspase-9/Bid pathway, although the existence of similar pathways in other cancers is yet unknown [25,39,42]. Deciphering HDAC10's possible significance in cancers will thus require advancements in research on the processes of this gene.

HDAC10 related to cell metastasis

The process of tumour cell metastasis is intricate and involves several processes such as cell adhesion, invasion, migration, and dissemination at various bodily organs [43,44]. Cancer cell lines primarily exhibit dysregulation of molecules involved in cellular invasion and migration, such as matrix metalloproteinases (MMPs) and S100A10 [45-47]. According to research by Zhao *et al*. [48], connexin 43 (Cx43) controls the expression of HDAC10, which is expressed at lower levels in lung giant cell carcinoma cells. Follistatin-like 1 (FSTL1) is produced more when it is overexpressed thanks to the interaction between the promoter of FSTL1 and HDAC10-mediated acetylation of H3 and H4. The Cx43-HDAC10-FSTL1 axis not only contributes to the low expression levels of S100A10, MMP-2 and laminin subunit α 4 (LAMA4), but also promotes MTSS I-BAR domain containing 1 (MTSS1) expression, which is crucial in the inhibition of both invasion and metastasis [48]. HDAC10 functions as a tumour suppressor in metastatic cervical squamous carcinoma cells by controlling MMPs. Basically, the pairs of HDAC10 and the promoters of MMP-2 (especially the AP1-binding site) and MMP-9 (especially the NF-κB- and sp1-binding sites), the decreased level of histone acetylation has a major function in the binding function of polymerase II, weakens the expression levels of MMP-2 and MMP-9 and restrains cancer cell invasion and migration [49]. Additionally, HDAC10 deacetylates acetyl-H3K9/14 and acetyl-C/EBP, which prevents their recruitment to the miR-223 promoter in sulfatide-treated hepatocellular carcinoma (HCC) cells. In addition to facilitating the production of integrin V subunit, reduced miR-223 expression also inhibits HCC cell movement and metastasis [50]. The β-catenin pathway is changed in a variety of malignancies and is crucial for the development of cancer. [51,52]. HDAC10 overexpression in renal cell carcinoma (RCC) limits RCC cell invasion by preventing the production of phosphorylated nuclear β-catenin. [27]. Even though these findings are fascinating, they primarily depended on overexpression experiments. Consequently, it is necessary to define the role of HDAC10 in cancer metastasis in more detail.

HDAC10 related to angiogenesis

In the early stages of cancer formation, angiogenesis is associated with essential activities [53]. Extracellular regulated protein kinases (ERKs), vascular endothelial growth factor (VEGF), TGF-β, and other pro-angiogenic proteins are involved [54-56]. According to Duan *et al*. [57], overexpressed HDAC10 deacetylates the H3 and H4 residues in the promoter of phosphatase and tensin homolog 22 (PTEN22), which prevents polymerase II from interacting with the PTEN22 promoter. In turn, the increased ERK1/2 activation caused by the PTEN22 expression being reduced eventually makes it easier to construct tubes both *in vivo* and *in vitro*. As a result, it was hypothesised that the HDAC10-PTEN22-ERK1/2 signalling pathway may be involved in the angiogenesis of tumours. Additionally, HDAC10 influences angiogenesis in gastric and colon cancer cells by reducing VEGFR. HDAC10 inhibits the interaction of heat shock protein (Hsp) 90 with VEGF receptor 1 (VEGFR1)/VEGFR2, although it favours the association of Hsp70 with VEGFR1/VEGFR2 [58]. VEGFRs are degraded by proteasomes as a result of this unbalanced binding, which may control angiogenesis.

Drug resistance and HDAC10

Cancer treatment frequently results in drug resistance. It can potentially result in treatment failure in addition to limiting the therapeutic impact [59]. The increased drug efflux rate, modified drug targets, and changed drug metabolism all lead to drug resistance [60,61]. In neuroblastoma, HDAC10 downregulation could improve doxorubicin sensitivity. As proposed by Ridinger *et al*. [62], reducing HDAC10 expression causes doxorubicin to accumulate intracellularly by preventing lysosomal exocytosis, which causes cell death with increased double-strand breaks (DSBs) and DNA damage. Additionally, therapeutic stress-activated autophagy contributes to drug resistance and multidrug resistance (MDR) [63]. Based on this phenomenon, Oehme *et al*. [64,65] discovered that HDAC10 deletion acetylated Hsp70/heat shock cognate 70 (Hsc70) impairs the autophagic flux by preventing the fusion of autophagosomes and lysosomes, allowing drug-resistant neuroblastoma cells to regain doxorubicin sensitivity. In a study that integrated bioinformatics studies and cell tests, it was discovered that decreased HDAC10 expression increases the cytotoxicity of cisplatin on ovarian cancer cells while also impeding DNA repair [66].

DNA double strands can be broken by a variety of cytotoxic processes, which causes cell death [67]. Several proteins, including the human single-stranded DNA binding protein 1 (hSSB1), will repair the damaged regions when they encounter DNA damage [68]. According to Wu *et*

al. [69], p300 stabilises DNA by acetylating hSSB1 Lys94 when it is localised at damaged sites. However, upregulated HDAC10 reverses this process and increases the chemosensitivity of cancer cells. Castration-resistant prostate cancer (CRPC), which develops as a result of androgen deprivation treatment (ADT) resistance, is a challenging condition [70]. According to earlier research, the androgen receptor (AR) signalling pathway can become worsened by the bromodomain and extra-terminal protein inhibitor (BETi). Furthermore, BETi-induced bromodomain-containing protein 4 (BRD4, a member of the BET family) expression is increased in prostate cancer cells, which provides resistance to these inhibitors [71,72]. Jin *et al*. [73] discovered that deubiquitinase ubiquitin-specific peptidase 17 like family member 2 (DUB3) is negatively correlated with low expression nuclear receptor co-repressor 2 (NCOR2)-HDAC10 complex, which may interact with BRD4 and further inhibit sensitivity to BETi in prostate cancer cells and *in vitro*. Experiments indicate that the NCOR2-HDAC10-DUB3-BRD4 signalling pathway may be beneficial as a regulator of drug resistance in cancers, even though the functions of HDAC10 in drug resistance have not yet been fully studied [73]. Silencing HDAC10 also lowers the transcriptional activity of AR-V7 and full-length AR (flAR) in prostate cells, making AR an important target in the treatment of CRPC [74], It indicates that HDAC10 may be useful for treating diseases.

HDAC10 with other biological functions

Cancer stem cells (CSCs) are cells that may promote cancer growth, progression, and resistance to therapies. They are always present in malignancies [75,76]. Numerous molecules, including as CD44, SLUG, and SOX9, may aid in determining the stem cell status [77,78]. Down-regulation of HDAC10 and activation of TGF-β in LA activates SOX9, SLUG, and CD44, which facilitates CSC-related cancer development [26]. HDAC10 influences immunological and inflammatory processes as well. HDAC10-deleted Foxp3+ Treg cells have immunosuppressive properties and may help treat immune-related colitis and extend the lifespan of cardiac allografts in animals with MHC mismatches [79]. Massive macrophages, particularly M2 macrophages, are likewise attracted by HDAC10 loss to the lung cancer microenvironment. Considering the possibility that M2 macrophages may release inflammatory inhibitors and stimulate cancer growth, HDAC10 may accelerate the development of tumours by escalating tumour inflammation [26,80]. Chaperone-mediated autophagy (CMA), a lysosomal-associated type of autophagy that occurs via selective cytosolic protein transfer, is another result of HDAC10 damage [81]. Lysosomal-associated protein 2 A

(LAMP2A) was shown to be more highly expressed in knockout HDAC10 cervical cancer cells, and the number of LAMP2A-positive lysosomes that accumulated close to the nucleus was also elevated [82], which would indicate that the activated CMA [83,84]. This quality may help in the treatment of cancer.

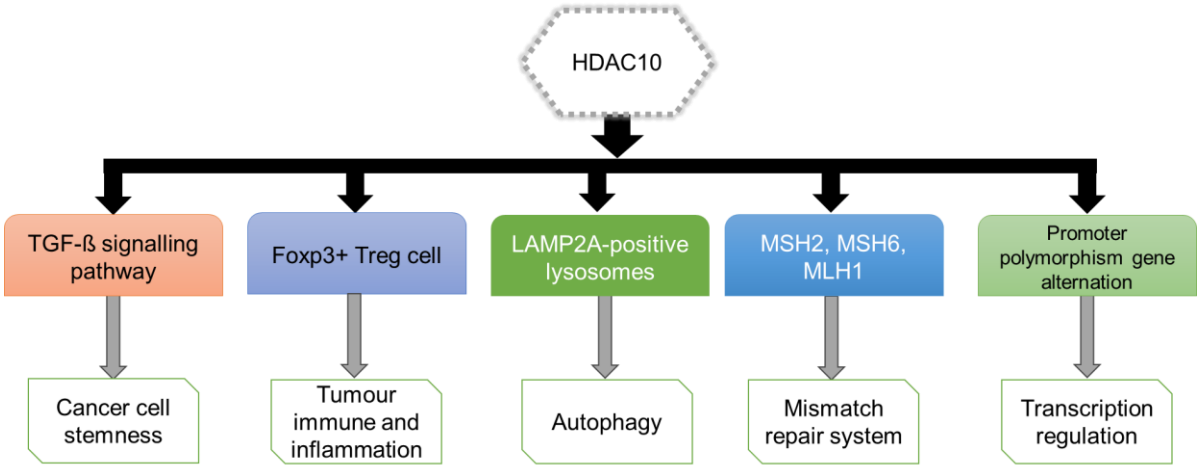


Figure 2. Role of HDAC10 in non-tumour diseases.

The mismatch repair mechanism (MMR) is crucial for fixing errors that arise during the DNA replication cycle [85]. It has been found that MMR frequently included mutL homolog 1 (MLH1), mutS homolog 2 (MSH2), and MSH6 [86]. In HeLa cells, MSH2 activity can be triggered by MSH2 deacetylation at Lys73 after better MMR activation [87]. Research comparing colon cancer cells to nearby cells also found that MSH2, MLH1, and MSH6 were negatively connected with low levels of HDAC expression in normal cells [88]. This might explain why cancer cells have a worse prognosis than normal ones. HDAC10 can carry out regulatory functions through altering other epigenetic factors, including as methylation, promoter polymorphism, and transcription [89-91]. HDAC10 is markedly methylated in adrenocortical carcinoma cells and adenomas, and its expression level sharply declines with hypermethylation [90]. HDAC10 was discovered to be down-regulated by Uzdensky *et al*. [91] as the melanoma progressed through transcription dysregulation. Due to the promoter polymorphism, HDAC10 also contributes to affecting the occurrence and age of HBV infection in HCC patients, notably HDAC10-589C > *T*. The 'T' allele in HDAC10-589C > *T* enhances transcription activity, and might spur HCC progression by increasing HDAC10 expression

[89]. Other research revealed that HDAC10 among anaplastic thyroid cancer specimens had genetic alterations such frameshift-type deletions [92]. Lourdusamy *et al.* [93], proposed that multiple genes, including *BRD1*, *HIRA* and *HDAC10* are expressed at beneath levels in spinal ependymoma. But further research is required to understand the precise processes driving the activation of HDAC10 in spinal ependymoma. These investigations might reveal a different mode of tumour control (Figure 2). Also, it can be helpful for multiple clinical application of HDAC10 targeted therapy. It is important to note that further research is required to determine the precise methods through which HDAC10 is involved in other processes.

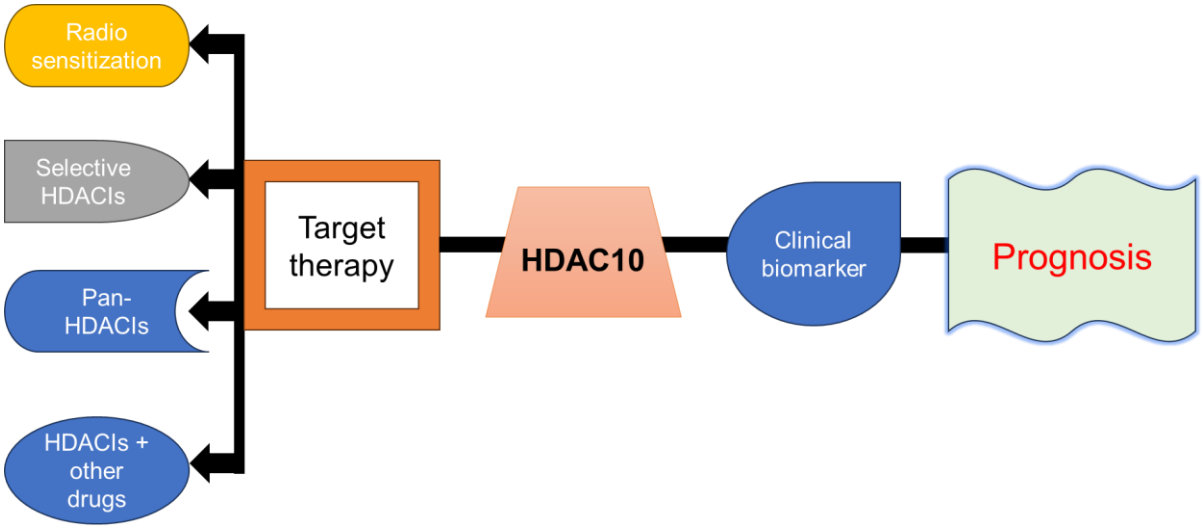


Figure 3. Clinical applications of HDAC10

Literature Review

Selective and non-selective HDAC10 inhibitors

Today, many HDACIs show strong broad-spectrum inhibition of HDACs (including HDAC10). Broad-spectrum HDACIs include suberoylanilide hydroxamic acid [94], CRA-024781 [95], CRA-026440 [96], AR42 [97], trichostatin A (TSA) [98], valproic acid (VPA) [99], and others. The novel hydroxamic acid-based HDACIs CRA-024781 and CRA-026440 exhibit potency against several HDACs, including HDAC1, 2, 3, 6, 8 and 10. In addition to reducing angiogenesis and tumour cell proliferation *in vitro*, CRA-024781 and CRA-026440 induce apoptosis and slow tumour growth *in vivo* [95,96]. Another group reported pan-HDACIs AR42 and sodium valproate, which likewise exhibit strong selectivity over HDAC1, 3, 8, and 10, and which pre-treatment with these HDACIs clearly showed *in vivo* and *in vitro* anti-tumour activities. Notably, PD-L1 and PD-L2 expression are reduced by AR42 and sodium valproate, which boosts the effectiveness of pazopanib by increasing a variety of immune cells, including M1 macrophages, neutrophils, NK cells, and T cells [9, 97]. TSA can also inhibit HDAC10 and 6; its N-terminal domain confers selectivity over HDAC10 for inhibition [99]. TSA attenuates the Wnt pathway by deleting the Wnt signalling pathway-associated protein TCF7L2 in a proteasome-dependent manner, inhibiting cell proliferation in vitro in colorectal cancer cells by inducing the inhibition of HDAC10 and HDAC6. A new approach was used to screen out the anti-inflammatory medication bufexamac, which is selective for HDAC10 and HDAC6 and whose cellular effectiveness may depend on interferon released by mononuclear cells [100].

Geraldy *et al*. [101] investigated HDAC10-specific inhibitors. However, they found that tubastatin A was tightly linked to HDAC10 in HeLa cells with its primary amine in the cap group due to some serious adverse effects of multitarget HDACIs. They discovered a hydrogen bond between a cap group nitrogen and a gatekeeper residue, Glu272, contributing to the tight binding based on an HDAC10 homology model. Unfortunately, because HDAC6 activity cannot be eliminated, discovering a highly selective HDAC10 inhibitor is still years away.

Design and synthesis of HDAC10 inhibitors

Azumamides A – E were synthesized by Villadsen *et al*. by altering the building block reaction's conditions [102]. They also looked at HDAC10's inhibitory characteristics at two distinct drug concentrations (5 and 50 M), and they found that both azumamides C and E significantly inhibit HDAC10. The N-benzylindole core of PCI-34051 and the tubastatin A are identical [101, 103]. Morgen *et al*. [104] synthesised di-hydroxamic acids by combining

these two inhibitors, drastically reducing neuroblastoma cells' survival, and serving as a potent inhibitor of HDAC10. To create M0017, the best model of human HDAC10.

Uba *et al*. [17] screened out ZINC19749069, the chemical with the highest rank in the ZINC database, due to its excellent stability when docked into an HDAC10 model with quisinostat. As a result, M0017 and ZINC19749069 have a lot of potential to be the best HDACI and scaffold .

Even though a growing number of studies may be relied upon to designate them as HDACIs, further research into the HDAC10 structure and biological characterisation is still necessary to understand how to choose and create site-specific HDACIs.

Rationale Behind the Study

Rationale behind selection of HDAC10 inhibitors

HDAC-10 has been indicated as a new target for novel compounds, possibly devoid of the severe side effects, which characterize the pan HDACs inhibitors presently available. In fact, HDAC-10 is recognized to promote cancer cell survival as a mediator of autophagy and to induce resistance to cytotoxic chemotherapeutic agents, particularly doxorubicin [62]. Herbst-Gervasoni *et al*. [21] based on their results on the X-ray crystal structures of a series of *N*8-acetylspermidine analogues containing different zinc-binding groups (ZBG) complexed with zebrafish HDAC-10 reported the importance of the electrostatic effects of E274 to confer specificity, besides the 3_{10} helix defined by the P(E, A) CE motif, that constrains the active site. The capability of the P(E, A)CE motif to accommodate inhibitors carrying bulky capping groups, provided their complementary to the P(E, A) CE motif helix.

Based on the reported activity of Tubastatin A towards HDAC10, several compounds structurally related have been designed. Among them, the hydroxamic acid derivative compound 36 displayed significant potency and selectivity (HDAC-10 pIC50 = 8.4 and pIC50 towards Class I HDACs ranging from 5.2 - 6.4 [105]. In addition, the affinity for HDAC-10 of several compounds extracted from *Psidium guajava* have been reported recently [106]. Given their high presence in many plants, they could represent an alternative area to discover selective HDAC-10 inhibitors.

Following a novel approach, based on the evidence that HDAC8 elevated levels correlate with poor prognosis in neuroblastoma treatment, dual HDAC8/10 inhibitors have been designed. Encouraging results are obtained by a series of hybrid di-hydroxamic acid derivatives [105]. Unexpectedly, the basic nitrogen presence in the cap group indicated as an essential condition in previous studies, was unnecessary for these series. Though most of these compounds showed poor cell permeability, the studies suggest that using prodrugs or substituting hydroxamic moiety with a different ZBG would improve their efficacy. The superimposition of the crystal structures of Compound 46a complexed with a ″humanized″ *Danio rerio* (zebrafish) HDAC10 and with *Danio rerio* HDAC-6, respectively, showed a different conformation of the capping group, thus providing additional information to generate HDAC10 selective compounds [21-22].

In this landscape, a quantitative structure-activity relationship (QSAR) study can be a good choice for the discovery of potential HDAC10 inhibitors that can be the prime solution for the treatment of HDAC10-related pathophysiological conditions.

Therefore, in this work, the machine learning-based QSAR study has been conducted on a diverse dataset of 484 compounds (obtained from BindingDB, website link: https://www.bindingdb.org/) with a wide range of HDAC10 inhibitory activity. It has depicted several crucial structural attributes from these unique 484 compounds which can assist in the designing and development of potent HDAC10 inhibitors for the treatment of HDAC10-related maladies.

Materials and Methods

Preparation of data set

In any type of QSAR-based molecular modelling study, literature study and dataset preparation are the primal parts. In this study, a web-accessible database called " The Binding Database" or BindingDB (https://www.bindingdb.org) [107] had been utilized and a series of 529 compounds were collected as .sdf format. The duplicate compounds were eliminated from the dataset with the help of the "*prepare ligands for QSAR*" module of Discovery Studio (DS) version 3.0 [108]. A total of 484 compounds was considered as a duplicate-free whole dataset. Then the dataset was classified into active (1) and inactive (0) classes based on mean biological activity [109]. Next, the biological sorting method was applied to split the whole dataset into a training set (N_{Train} = 363) and test set (N_{Test} = 121). The details of the training and test set has been provided in Table S1 and S2 (SMILES format) of Appendix, respectively.

Calculation of molecular descriptors for the development of Machine learning (ML) models

The calculation of molecular descriptors was done by using PaDEL descriptor calculator software [110-112]. The details of the configuration include a classifier fitness assessment and experimental results obtained on the training dataset by ten-fold cross-validation (10-*CV*). The trained classifiers were applied to the test dataset to observe their functionality. Along with that, several error measurement statistical and non-statistical metrics like: Mean Squared Error (*MSE*), Mean Average Error (*MAE*), *R-Squared error* and area under the receiver operating characteristics (*ROC*) Curve (*AUC*) score were applied to evaluate the difference between actual and predicted labels [109-113,114].

Configuration of the classical ML classifiers

Here, the Scikit-learn package was used to develop standard ML classifiers (Decision Tree, Random Forest, Gradient Boosting, and Linear Discriminant Analysis (LDA). These ML techniques are publicly available with detailed documentation and can be obtained via the sklearn library website [115]. The parameters for each classifier have been provided in Table 1, as well as the documentation URL mentioned in the respective column [116-119].

Table 1. Configuration details of the ML classifiers.

Classifier	Scikit-learn method	Parameters
Decision Tree	sklearn.tree.DecisionTreeClassifier	criterion='gini', min_samples_leaf=1, presort='deprecated', splitter='best'
Random Forest	sklearn.ensemble.RandomForestClassifier	n_estimators=100, min_samples_split=2, min_samples_leaf=1
Gradient Boosting Classifier	sklearn.ensemble.GradientBoostingClassifier	learning_rate=0.1, n_estimators=100, subsample=1.0, criterion='friedman_mse', min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_depth=3
LDA	sklearn.discriminant_analysis.LinearDiscriminantAnalysis	solver='svd', tol=0.0001

Bayesian Classification study

Bayesian classification is mainly based on the classical Naïve Bayes' theorem-based approach. The related Naïve Bayes theorem can be represented by the following equation,

$$P(wi|\mathrm{X}) = \frac{p(X|\mathrm{wi})P(wi)}{p(X)}$$

Here,

$P(wi|\mathrm{X})$ denotes posterior probability (probability of a hypothesis "w" being true given the observed data "X"

$P(wi)$ is prior probability of hypothesis,

$p(X|\mathrm{wi})$ denotes likelihood of seeing the evidence if the hypothesis is correct,

$p(X)$ is the likelihood of the evidence under any circumstances [120].

Here this method had been used as a ML-dependent approach to identify good and bad structural fingerprints/features for HDAC10 inhibitory action [120-122]. During the development of the Bayesian classification model, different descriptors such as extended connectivity fingerprints with a diameter of 6 (*ECFP_6*), number of rings, number of aromatic rings, number of hydrogen bond acceptor, number of hydrogen bond donor, *ALogP*, number of rotatable bonds were calculated. The *ECFPs* are intended to trace the explicit atom-type information and use specific atom types as an identifier. Model creation and validation were carried out using the training set **(**Appendix Table S1**)** and test set **(**Appendix Table S2**)** molecules, respectively. The cross-validation methods were employed to inspect the model. Analyses of distinctive true positives (*TP*), true negatives (*TN*), false positives (*FP*), false

negatives (*FN*) were made to evaluate the model's propensity for prediction [121]. The equations for determination of s *Sensitivity*, *Specificity*, *Concordance* are given below:

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$Concordance = \frac{TP + TN}{(TP + FN + TN + FP)}$$

Recursive Partitioning (RP) study

This is a multivariate strategy for classification-based 2D QSAR investigation. It is commonly used for critical prediction and classification in compounds with known biological activity [122, 124]. It employs a common method to construct trees that depict the succession of rules that must be followed to obtain a predicted value or class. In this analysis, a decision tree is formed in which each branch represents the value of some characteristics and each node is enriched with *active* or *inactive* compounds. The RP models were built using Discovery Studio (DS) Version 3.0 [108], and validated using cross-validation procedures. The best RP model for HDAC10 inhibitors was chosen based on its ability to discriminate between different types of inhibitors.

Molecular docking study

Molecular docking study was used to understand the binding interaction of the potential HDAC10 inhibitors with the HDAC10 enzyme. The docking study also validated the structural fingerprints identified by Bayesian classification as well as recursive partioning studies. Since the human HDAC10 X-ray crystallographic structure is not available, we collected a "humanized" form of HDAC10 (PDB ID: 6VNQ, source organism: *Danio rerio*) from the Protein Data Bank (PDB) [125]. Initially, the receptor preparation and optimization were performed. Different steps including the addition of missing hydrogens, state generation, Zn^{2+} metal binding site preparation and structural rectification of the protein were performed as

mentioned in the "*Protein Preparation Wizard* " of Schrodinger Maestro v12.1 [126]. As a force field option, OPLS_2005 was selected. The receptor grid was generated by using the "*Receptor Grid Generation*" module of Schrodinger Maestro v12.1. Similarly, ligands or the inhibitors were prepared by using the "*Ligprep*" module from Maestro v12.1 software [126]. The extra precision (XP) method of the Schrodinger Maestro v12.1 was used for the docking study.

Results and discussion

The whole modeling study has two main objectives - (A) The development of statistically validated ML models, (B) identification of structural fingerprints for HDAC10 inhibition. The different modeling techniques used in this study to fulfill these two objectives are shown in **Figure 4**.

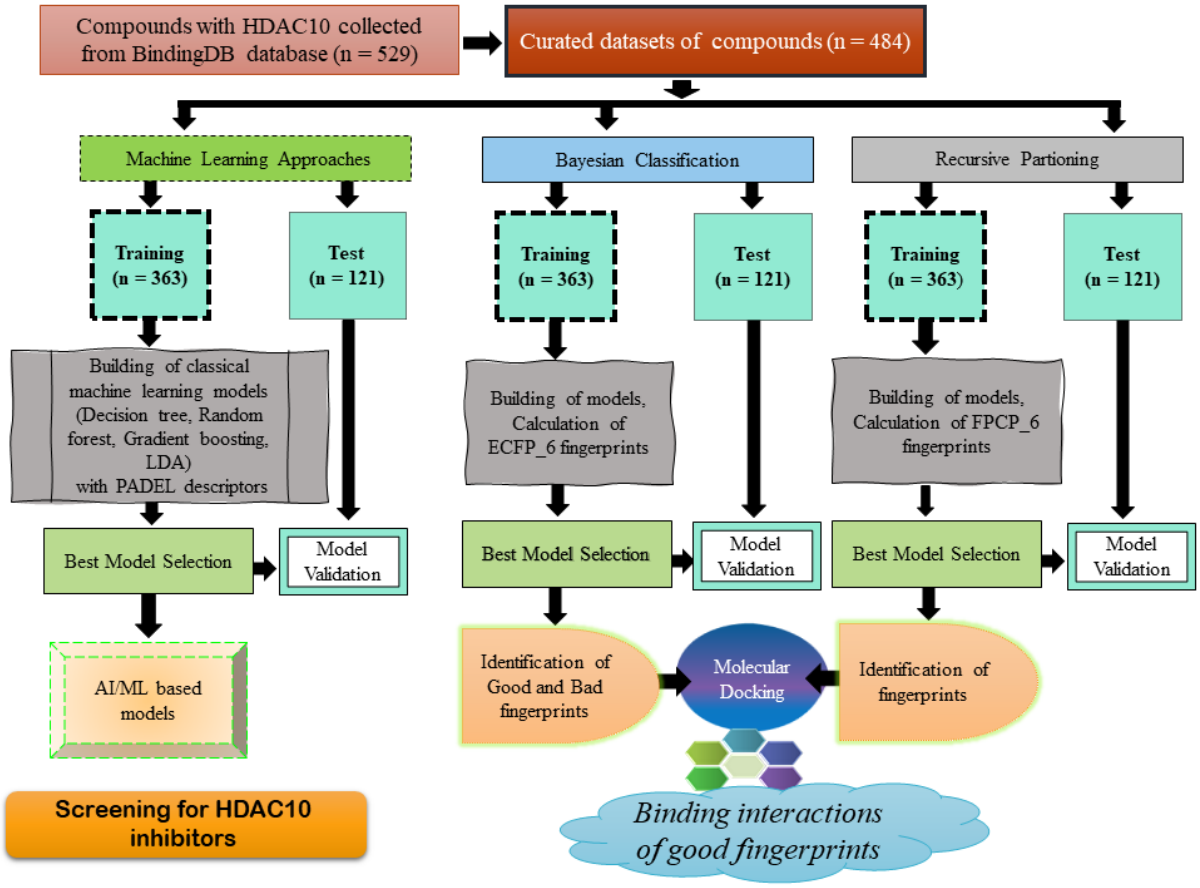


Figure 4. Overview of modeling strategies used in the present study

Development of ML models with PaDEL-descriptor for HDAC10 inhibition

Scikit learn package was used to develop different ML models. Here, four different classification algorithms namely Decision Tree, Random Forest, Gradient Boosting, and LDA were used. These algorithms are publicly accessible and can be imported from Scikit learn package [115-119]. The descriptors were calculated by using PaDEL-descriptor (http://www.yapcwsoft.com/dd/padeldescriptor/). During the validation phase of the ML models, the hyperparameters were adjusted. Additionally, the major goal of hyper-parameter tuning is to govern how the generated models behave through the optimized parameter values. The best performance of generated models may also achieve using these modifiable

parameters. To complete the process of tweaking the ML models' hyperparameters, we employed the looping approach. With this approach, a list of hyper-parameter values was specified before ML models were executed to determine the best values.

Fitness measurement of ML classifiers

We utilized the training set to assess the fitness of the ML classifiers. In **Table 2**, the fitness assessment property of the classifiers using evaluation matrices are shown. The training set fits for the Decision Tree, Random Forest, Gradient Boosting, and LDA classifiers with *Accuracy*, *Precision*, *Recall* and *F1-score* are around 94%. The suitability measurement has been graphically demonstrated using confusion matrices in **Figure 5**. Additionally, the heat map view of the experimental results for each classifier model is shown in **Figure 6**.

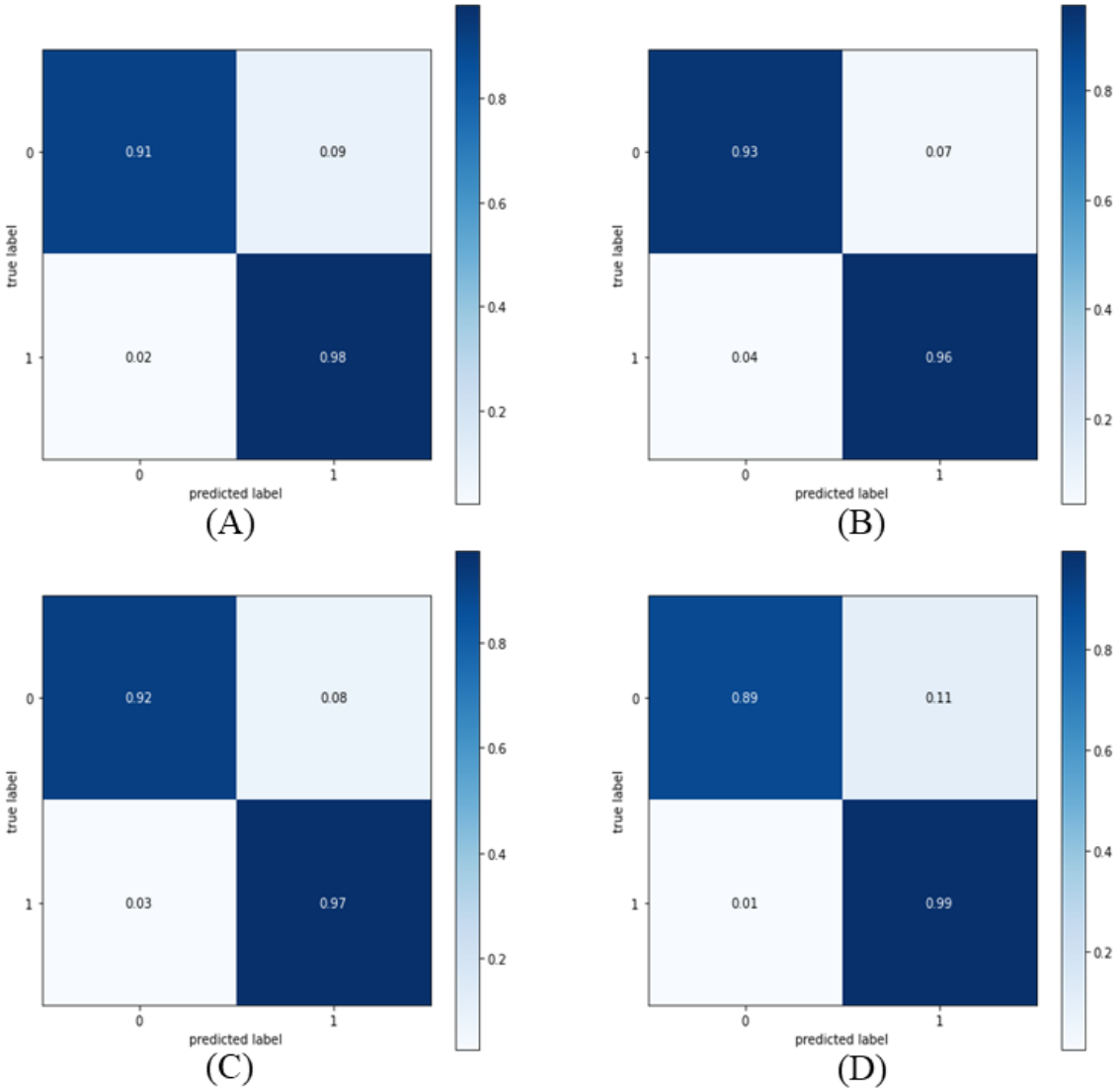


Figure 5. Confusion matrix of (A) Decision Tree, (B) Random Forest, (C) Gradient Boosting, and (D) LDA classifiers while fitness measurement

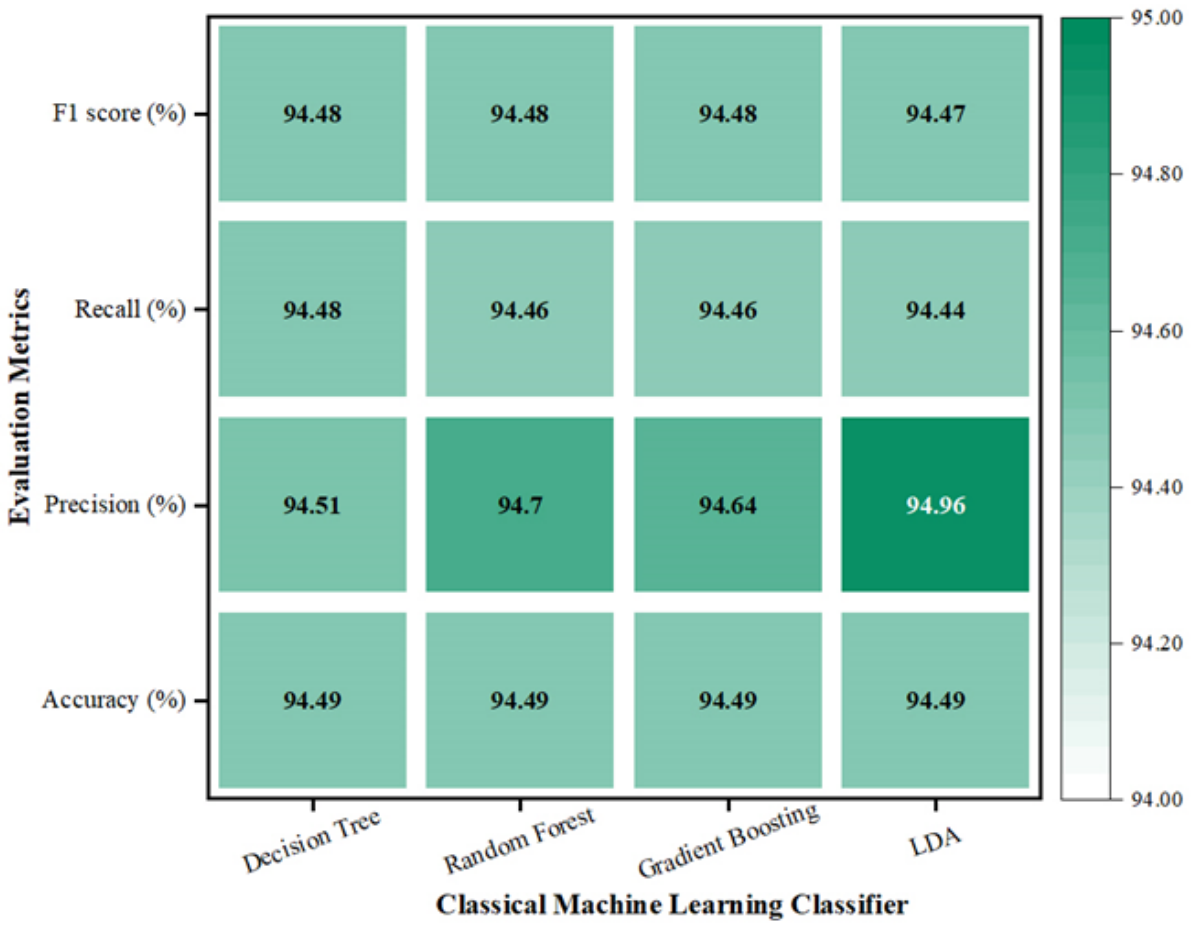


Figure 6. Heat map view of classical ML classifiers while fitness measurements.

Table 2. Fitness measurement results of classical ML classifiers

Classifier	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1-score* (%)
Decision Tree	94.49	94.51	94.48	94.48
Random Forest	94.49	94.70	94.46	94.48
Gradient Boosting	94.49	94.64	94.46	94.48
LDA	94.49	94.96	94.44	94.47

The *ROC* curve depicts the probability distribution that plots *TPR* (True Positivity Rates) against *FPR* (False Positivity Rates) [39]. The *ROC-AUC* score of 0.99 has been exceeded by all the classifiers (**Figure 7**).

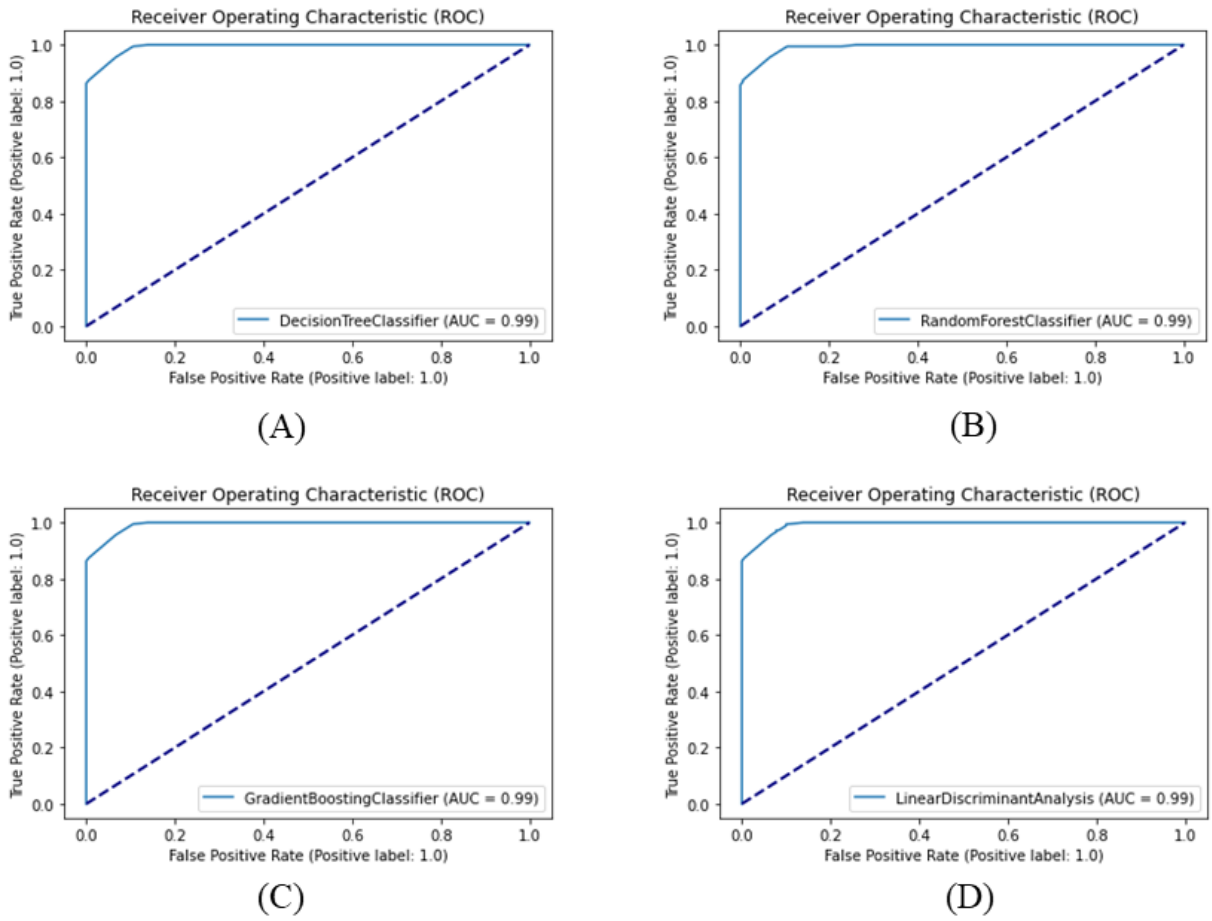


Figure 7. AUC-ROC score of classical ML classifiers while fitness measurement.

Cross-Validation Approach

The results of the ML models based on the 10 folds cross-validation approach are shown in **Table 3**. In this case, the overall good-performing classifier on the 10 folds cross-validation set is identified as the Gradient Boosting in terms of *Accuracy* rates (80.13%). Therefore, the remaining result has been found as a *Precision* score of 79.05%, *Recall* of 83.20%, and *F1-score* of 80.63%. However, the Decision Tree classifier has achieved the comparatively worst classification performance on the 10 folds cross-validation with an *Accuracy* of 74.33%, a *Precision* score of 74.92%, a *Recall* of 75.83%, and an *F1-score* of 74.41%, respectively. Additionally, the heat map view of the experimental result is shown in **Figure 8**.

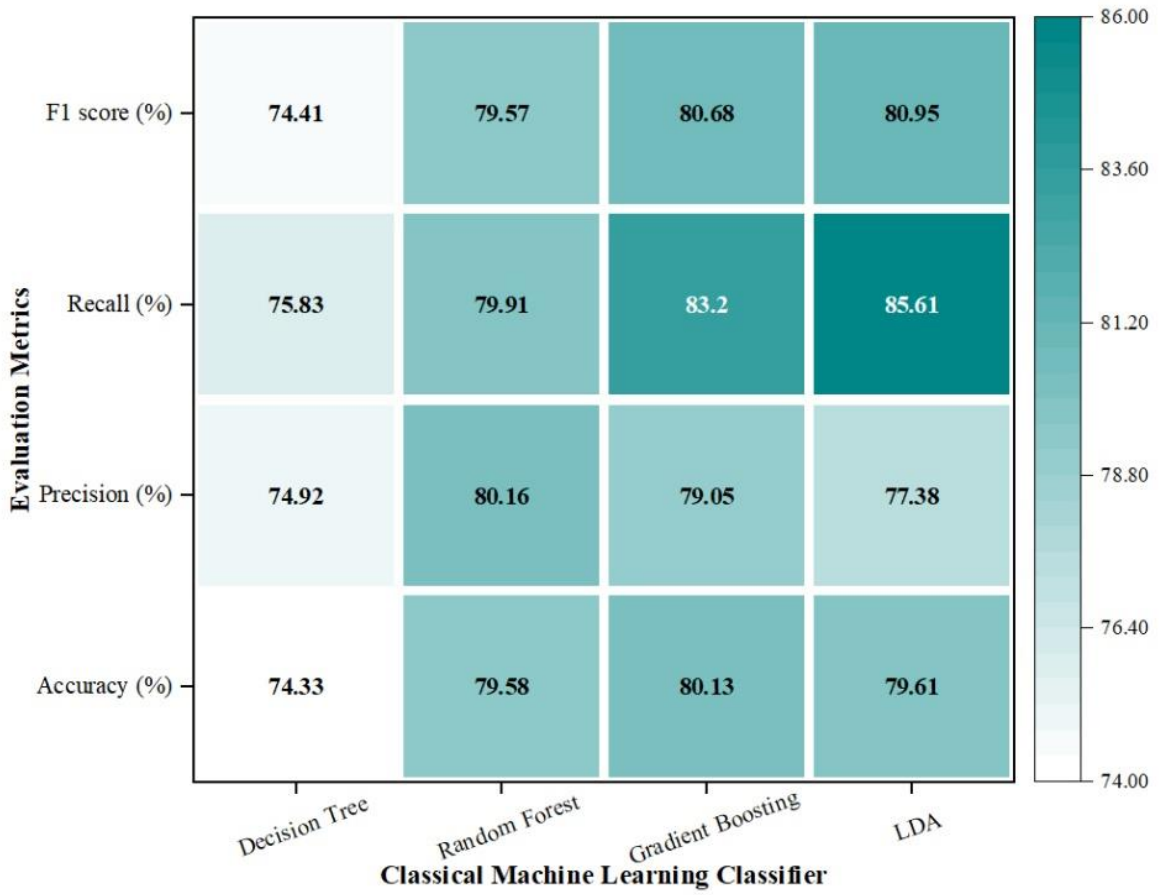


Figure 8. Heat map view of classical ML classifiers on 10 folds cross-validation approach while fitness measurement.

Table 3. Result of ML classifiers on 10 folds cross-validation approach while fitness measurement.

Classifier	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1-score* (%)
Decision Tree	74.33	74.92	75.83	74.41
Random Forest	79.58	80.16	79.91	79.57
Gradient Boosting	80.13	79.05	83.20	80.68
LDA	79.61	77.38	85.61	80.95

Results of the trained ML classifiers on the test set

The results of the ML classifiers on the test set are shown in **Table 4**. Here, the Random Forest has been identified as the good-performing classifier on the test set. The evaluation matrices *i.e.*, *Accuracy* of 80.16%, *Precision* score of 80.19%, *Recall* of 80.17%, and *F1-score* of 80.16% are found in case of Random Forest classifier.

Table 4. Result of ML classifiers on the test set.

Classifier	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1-score* (%)
Decision Tree	76.03	76.61	76.09	75.92
Random Forest	80.16	80.19	80.17	80.16
Gradient Boosting	79.33	79.40	79.35	79.33
LDA	76.85	77.61	76.78	76.66

Here also, the Decision Tree classifier has achieved the worst classification performance on the test set with an *Accuracy* of 76.03%, a *Precision* score of 76.61%, *Recall* of 76.09% and *F1-score* of 75.92%. Additionally, the heat map views of the experimental results and confusion matrix for each classifier are shown in **Figure 9** and **Figure 10**, respectively. The *ROC-AUC* plot of all the classifiers is depicted in **Figure 11**.

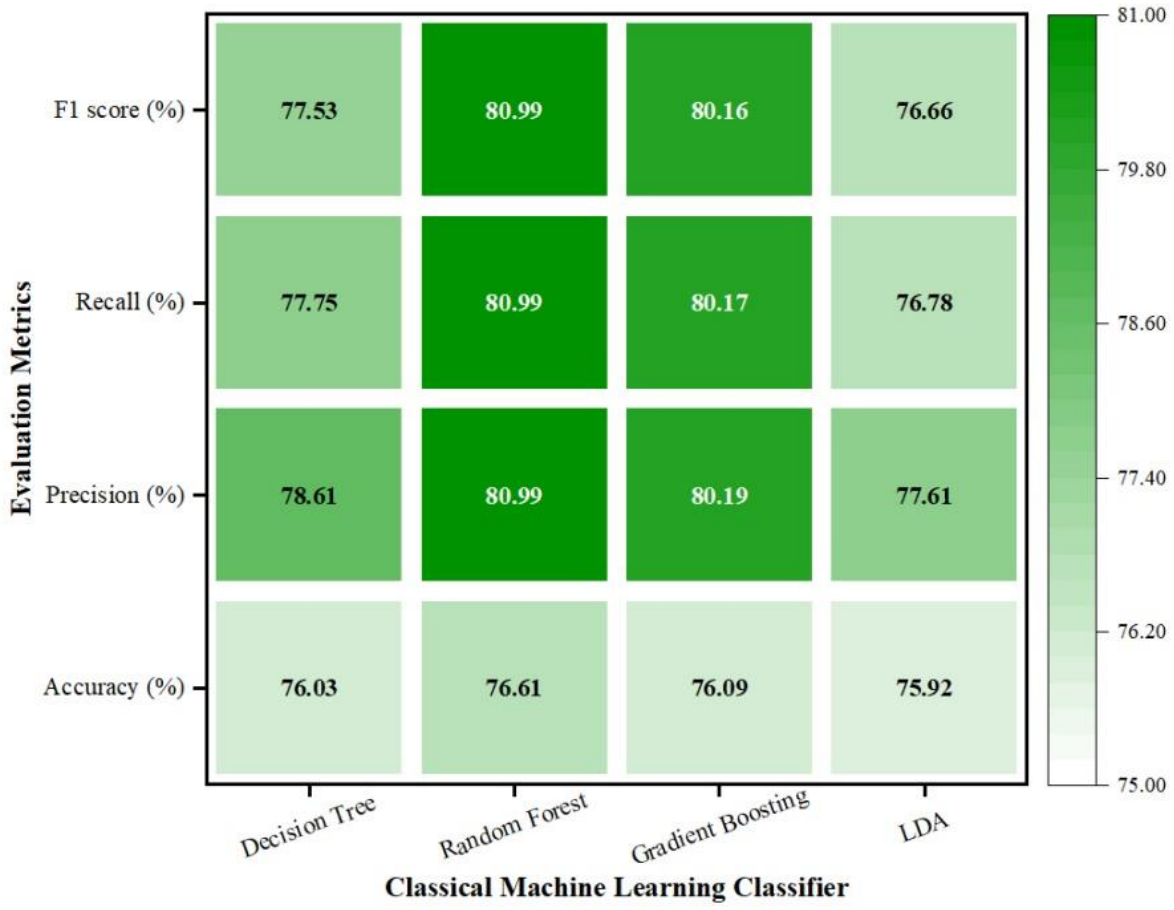


Figure 9. Heat map view of classical ML models on the testing dataset.

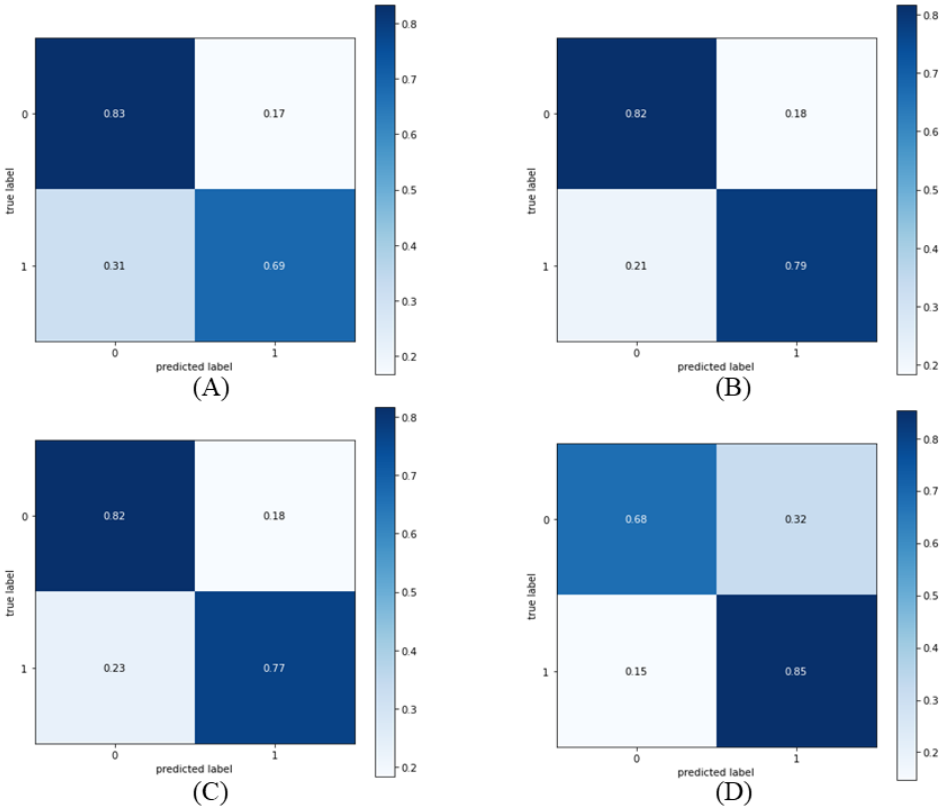


Figure 10. Confusion matrix of (A) Decision Tree, (B) Random Forest, (C) Gradient Boosting, and (D) LDA classifiers on the testing dataset.

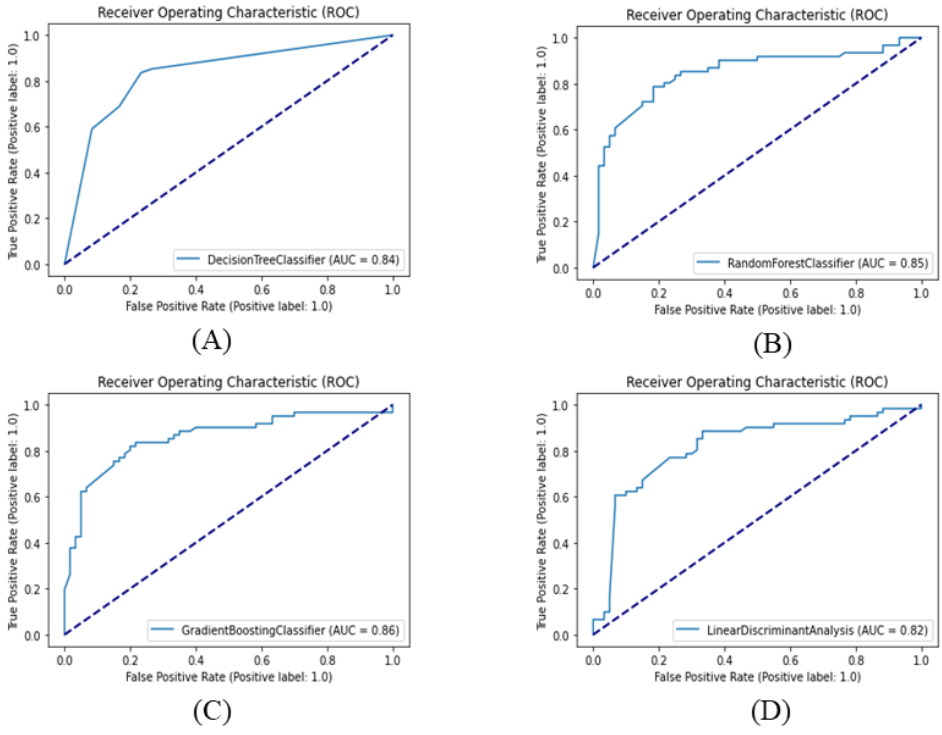


Figure 11. AUC-ROC score of classical ML classifiers on the testing dataset.

Performance measurement using *MAE*, *MSE*, *R-Squared error*, and Area under the *ROC* Curve (*AUC*) score

The loss function is primarily used to quantify the difference between the actual and predicted results. **Table 5** displays the computational loss value of ML classifiers. The LDA classifier was chosen as the best classifier with the lowest loss of the *MAE* floating values (*MAE* = 0.112), whereas the Decision Tree classifier had the highest loss value (*MAE* = 00.142). In terms of *MSE*, the LDA was recognized as the best classifier, with a minimal loss of 0.112 scores, while the Decision Tree classifier had the worst loss value when compared to the other classifiers (*MSE* = 00.131). The LDA classifier has the greatest goodness-of-fit meaning R^2 (score = 0.450), while the Gradient Boosting classifier has the poorest score (score = 00.536). Furthermore, each classifier has above the given acceptance criterion (*ROC-AUC* = 0.5). The LDA Tree, on the other hand, performs the poorest, with an *AUC* rate of 0.82. As a result, the Gradient Boosting classifier gets the highest *AUC*-score = 0.86. Overall, these four ML classification-based models are statistically significant. The models can be used for classification of unknown molecules into HDAC10 inhibitors based on their structure i.e., High throughput screening (HTS).

Table 5. Floating loss values of classical ML classifiers with ROC-AUC score.

Classifier		**Loss Values**		
	MAE	***MSE***	***R^2***	***ROC***
Decision Tree	0.142	0.131	0.478	0.84
Random Forest	0.128	0.125	0.534	0.85
Gradient Boosting	0.118	0.116	0.536	0.86
LDA	0.112	0.112	0.450	0.82

After satisfying our first objective of the study we have employed Bayesian classification and recursive partition studies by Discovery Studio version 3.0 [108] to identify the structural fingerprints for HDAC10 inhibition. Just developing a ML classification-based model will not give the medicinal chemists the complete guidance to design HDAC10 inhibitors. The identification of potential structural fingerprints modulating HDAC10 inhibition is necessary to guide the medicinal chemists to include/discard them in the development of HDAC10 inhibitors.

Bayesian classification study

The Bayesian classification model was constructed by using the "*Create Bayesian Model*" module of DS 3.0 [108]. The 5-fold cross-validated ROC_{5CV} value is found to be 0.871 and the externally cross-validated *ROC* score of the test set (ROC_{Test}) value is found to be 0.814. The *ROC* plots have been generated for training and test set as shown in **Figure 12** justifying the internal and external predictability of the model, respectively. Other statistical parameters like *Sensitivity*, *Specificity* and *Concordance* are also encouraging. Bayesian model shows good scores as reported in **Table 6** (*Sensitivity* = 0.902, *Specificity* = 0.867 and *Concordance* = 0.884 for training set) which justifies that the developed model is resilient and predictive.

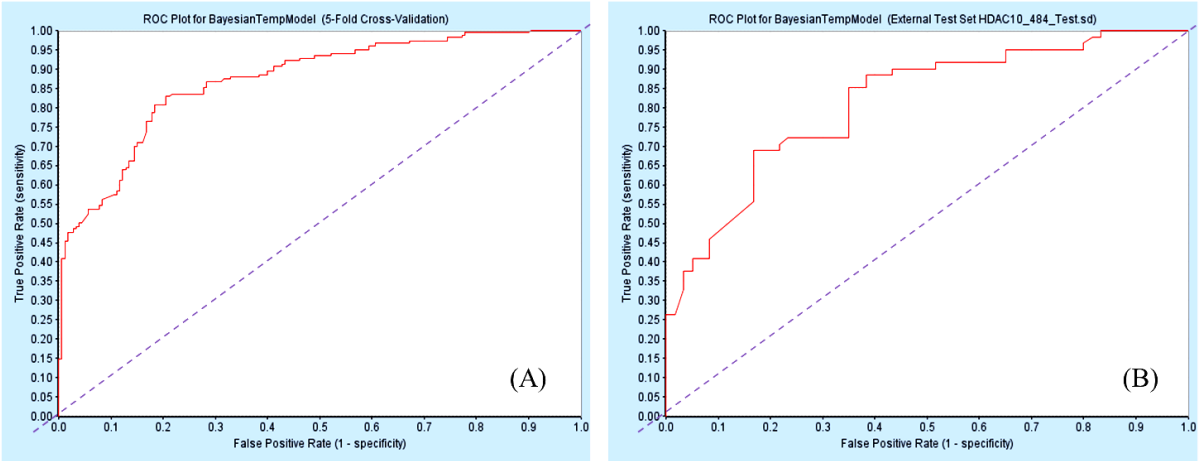


Figure 12. Bayesian ROC plots for **(A)** the training set and **(B)** the test set.

Table 6. Statistical parameters of the generated Bayesian model

Set	*ROC*	*ROC* rating	*TP*	*FN*	*FP*	*TN*	*Sensitivity*	*Specificity*	*Concordance*
Training[a]	0.871	Good	165	18	24	156	0.902	0.867	0.884
Test	0.814	Good	42	19	13	47	0.689	0.783	0.736

[a]5-fold cross-validation is performed for the training set to calculate the statistics; Where *TN*: true negative, *TP*: true positive, *FP*: false positive, *FN*: false negative

Depending on the Bayesian score, the top twenty good structural fingerprints (**G1-G20**) are identified and summarized in **Figure 13**. Compounds with favorable fingerprint show a good

sign for better HDAC10 inhibitory activities. The fragments G1, G2, G4 and G7 exhibit the importance of the carboxamido function present in a compound (Example: Compound **A053** with IC_{50} = 7 nM and **A061** with IC_{50} = 9 nM, **Figure 14**).

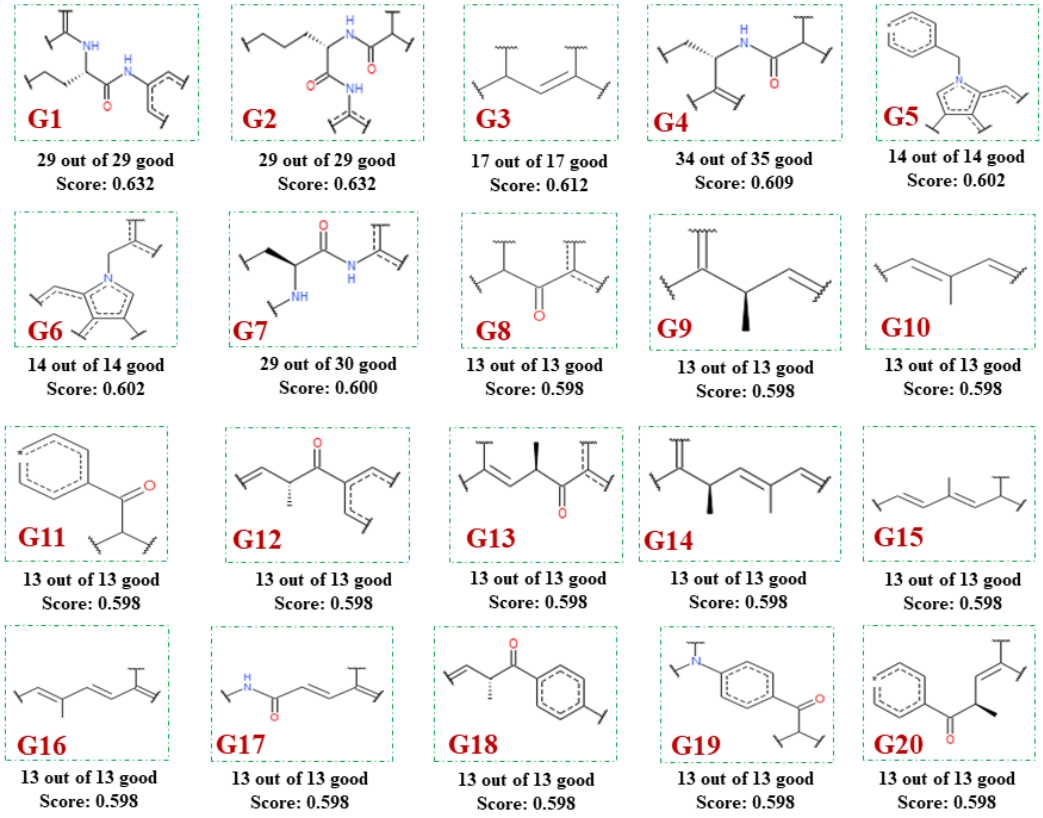


Figure 13: Top 20 good fingerprints identified from the Bayesian classification study

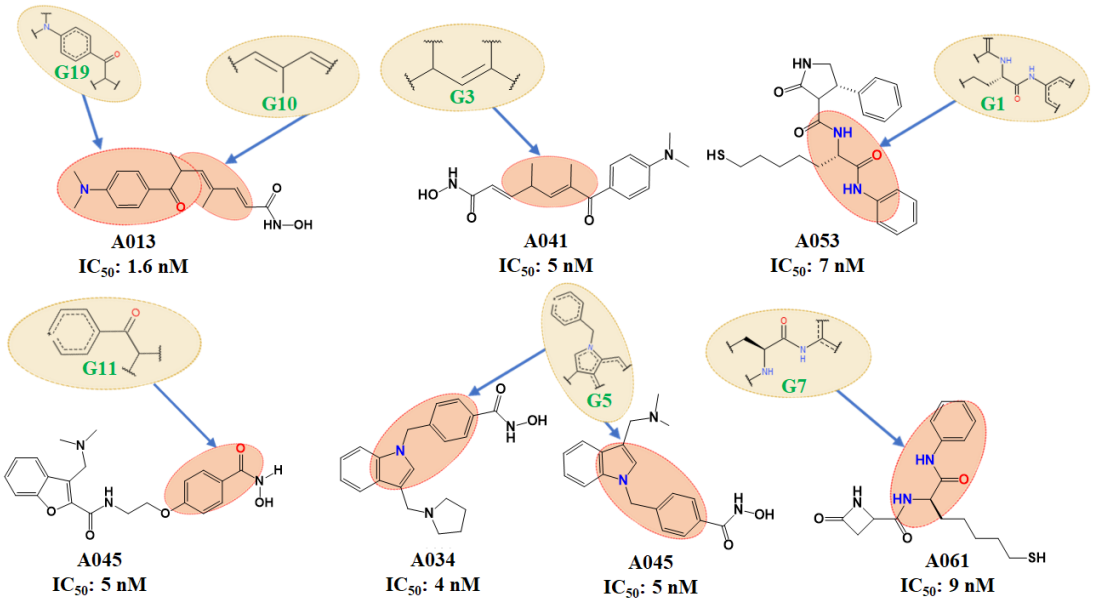


Figure 14. Structure of some active HDAC10 inhibitors with good Bayesian fingerprint.

Interestingly, compounds (Example: **A041**, IC_{50} = 5 nM) with fingerprint G3 (a unique 2, 4 dimethyl pent-2-ene structure) exhibits potent HDAC10 inhibition. The fingerprints like **G5** and **G6** containing 1, 3 substituted benzylindole moiety are well-attributed for imparting higher HDAC10 inhibitory activity (Example: **A034** with IC_{50} value: 4 nM and **A045** with IC_{50} = 5 nM). Fingerprints **G8-G20** suggest the importance of these fingerprints for HDAC10 inhibition (Example: **A013** with IC_{50} = 1.6 nM and A045 with IC_{50} = 5 nM).

Depending on the Bayesian score, the top twenty bad structural fingerprints (**B1-B20**) are also identified as shown in **Figure 15**. Poor HDAC10 inhibitory activity is indicated by the presence of the aniline moiety and secondary amine as a linker group in the compounds as shown by the bad fingerprints **B1, B3, B4, B6, B7, B8, B9**, and **B19** (IC_{50} values of compound **A395** and **A403** are 15,500 and 23,100 nM, respectively). Meanwhile, the 1,5 substituted indole analogues may lead towards the decreased HDAC10 inhibitory activity, according to fingerprints **B2** and **B5** (Compound **413** with IC_{50} = 59,800 nM, **Figure 16**).

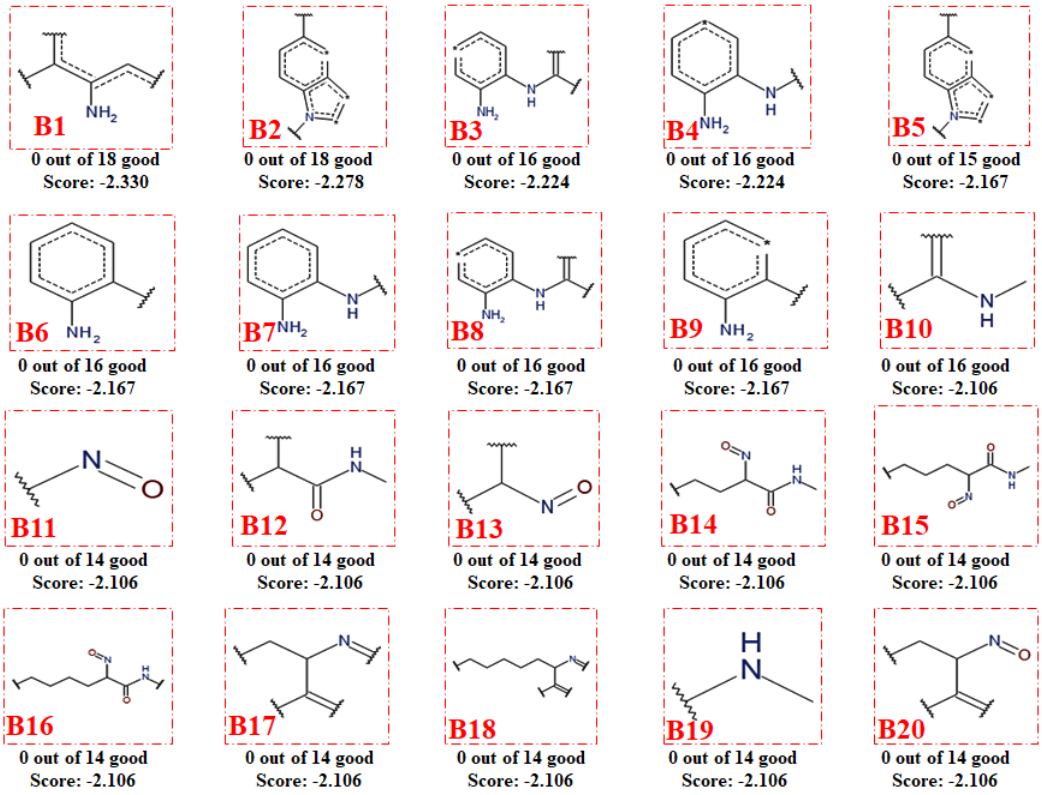


Figure 15. Top 20 bad fingerprints identified from the Bayesian classification study

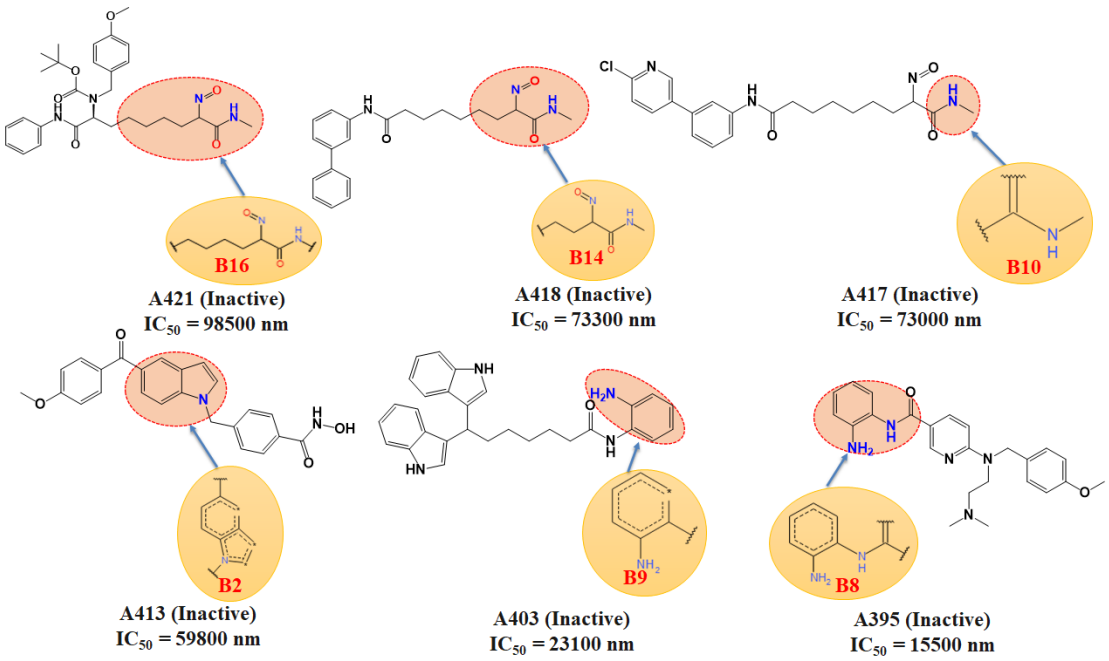


Figure 16. Structure of some inactive HDAC10 inhibitors with bad Bayesian fingerprints

The presence of the amide group and imine group (Fingerprint **B10**–**B18** and **B20**) attached to the chiral carbon as individually or both in compounds (Example: **A417, A418, A421**) results in detrimental to poor HDAC10 inhibition. Therefore, combining these results obtained from the Bayesian classification modeling study, it may be inferred that phenyl amine derivatives are less active but substitution with dimethyl group at N-atom of phenyl amine moiety along with an addition of keto group at *para* position may enhances the inhibitory activity. Furthermore, in hydroxamate derivatives, the benzyl indole containing compounds comes under both good and bad classes, only the substitution at 1, 3-position of indole group shows better inhibitory activity and distinguishes the active compounds from the less active/inactive.

Recursive partitioning study

In order to build the Recursive partitioning (RP) model, the same training and test sets (n_{Train} = 363 and n_{Test} = 121) were used. The RP model was generated using a combination of fingerprint feature (Feature-class fingerprint of diameter 6, *FPCP_6*) and different types of molecular properties like *AlogP*, molecular weight (*MW*), number of hydrogen bond donors (*nHBD*), Number of hydrogen bond acceptor (*nHBA*) by using Discovery Studio 3.0 [108]. A total of 8 tree(s) are constructed for the RP model. The details of the trees with the number of leaves,

statistics of confusion matrix, *ROC* value of training set, cross-validation results are given in **Table 3**.

Table 3. The results of the RP model as obtained from the training set of HDAC10 inhibitors.

Tree Information	**Confusion Matrix**			***ROC***	***ROC$_{CV}$***
Tree 1: 18 leaves Error Rate (training data): 49.045 Min alpha: 0		***1***	***0***	**0.940**	**0.8115**
	1	**161**	**22**		
	0	**27**	**153**		
Tree 2: 13 leaves Error Rate (training data): 50.136 Min alpha: 0.7799		*1*	*0*	0.9204	0.8117
	1	166	17		
	0	33	147		
Tree 3:12 leaves Error Rate (training data): 52.218 Min alpha: 2.34		*1*	*0*	0.9196	0.7933
	1	170	13		
	0	39	141		
Tree 4: 11 leaves Error Rate (training data): 54.896 Min alpha: 3.12		*1*	*0*	0.9142	0.7929
	1	149	34		
	0	21	159		
Tree 5: 9 leaves Error Rate (training data): 61.789 Min alpha: 3.899		*1*	*0*	0.8918	0.7381
	1	139	44		
	0	18	162		
Tree 6:5 leaves Error Rate (training data): 82.369 Min alpha: 5.459		*1*	*0*	0.8158	0.6985
	1	164	19		
	0	63	117		
Tree 7:4 leaves Error Rate (training data): 92.453 Min alpha: 12.48		*1*	*0*	0.7917	0.7015
	1	164	19		
	0	73	107		
Tree 8:2 leaves Error Rate (training data): 142.89 Min alpha: 25.74		*1*	*0*	0.6064	0.7015
	1	43	140		
	0	4	176		

The best RP model is shown in bold faces

Tree-1 is the minimally pruned tree with *ROC* score of 0.940. The error rate for Tree-1 is found to be 49.045 and min alpha is 0, which can be considered as a model with lesser error compared to other seven trees. A thorough examination revealed that Tree-1 is the best model with *ROC$_{CV}$* score of 0.8331. The test set result prediction for Tree-1 is similarly quite high (**Table 4**).

Table 4. The results of the RP model as obtained from the test set of HDAC10 inhibitors

Confusion Matrix			Tree ID	*ROC* Rating
Actual\Pred.	*1*	*0*	**1**	**Accuracy 0.831: Good**
1	**51**	**10**		
0	**17**	**43**		
Actual\Pred.	*1*	*0*	2	Accuracy 0.821: Good
1	52	9		
0	17	43		
Actual\Pred.	*1*	*0*	3	Accuracy 0.820: Good
1	54	7		
0	20	40		
Actual\Pred.	*1*	*0*	4	Accuracy 0.815: Good
1	45	16		
0	11	49		
Actual\Pred.	*1*	*0*	5	Accuracy 0.793: Fair
1	40	21		
0	9	51		
Actual\Pred.	*1*	*0*	6	Accuracy 0.702: Fair
1	52	9		
0	28	32		
Actual\Pred.	*1*	*0*	7	Accuracy 0.686: Poor
1	52	9		
0	30	30		
Actual\Pred.	*1*	*0*	8	Accuracy 0.: Poor
1	10	51		
0	6	54		

The best RP model is shown in bold faces

The decision tree with 18 leaves as shown in **Figure 17** exported six molecular properties and eleven structure fingerprints (**FP-1** to **FP-11**) based on the *FPCP_6* fingerprint. These eleven fingerprints play a crucial role to pick out the highly active and lower active/inactive HDAC10 inhibitors.

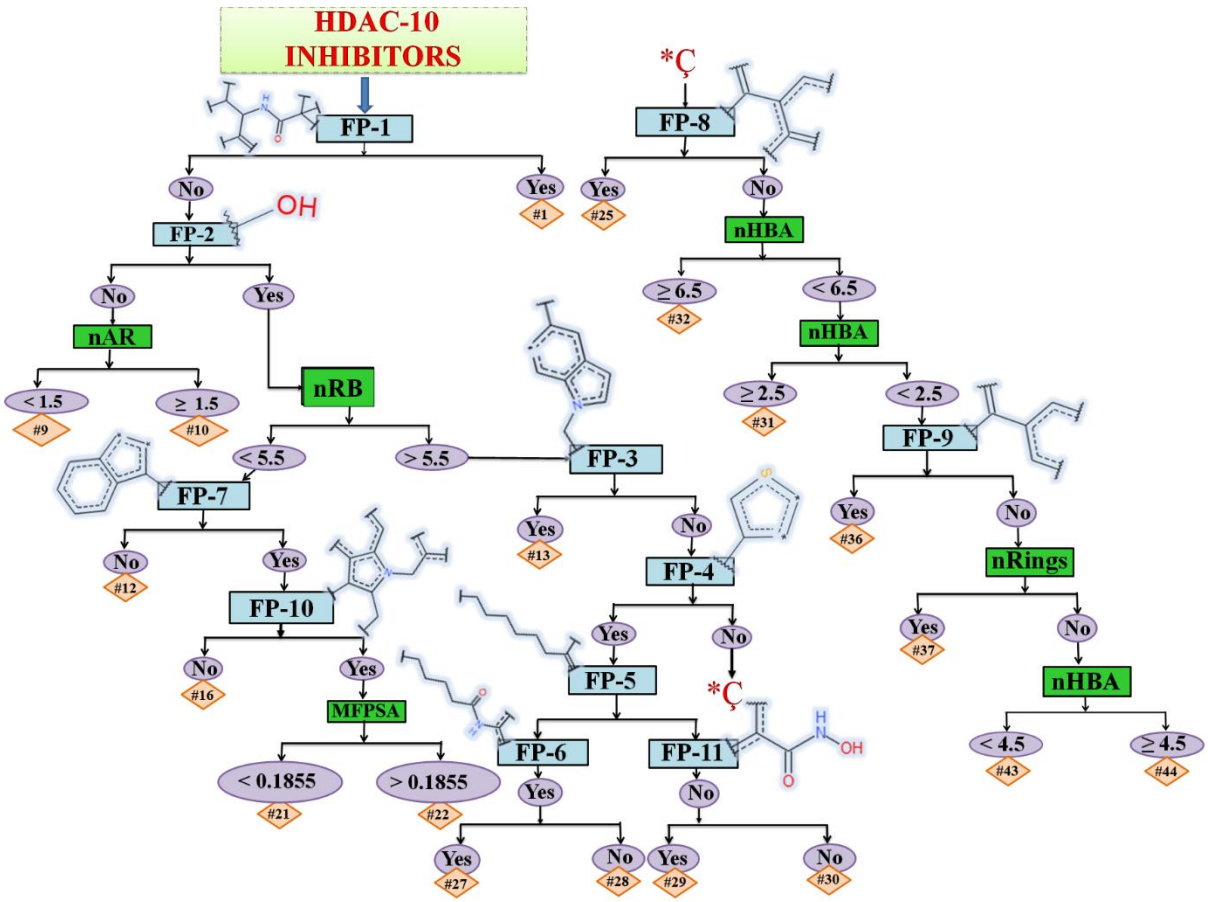


Figure 17. Classification of HDAC10 inhibitors through the decision tree into active and inactive classes by using RP (the Leaf ID is highlighted as "# number of the leaf" and *Ç is defined as continuation to FP-8).

Of the six molecular properties chosen by the decision tree, number of rotatable bonds (*nRB*), number of aromatic rings (*nAR*), molecular fractional polar surface area (*MFPSA*), number of hydrogen bond acceptor (*nHBA*), number of rings (*nRings*) are properties that determine hydrogen binding ability and bulkiness of the molecules. In the decision tree, the Leaf ID #1 comprises compounds with fingerprint-1 (FP-1) in their structure (Example: compound **A053** shown in **Figure 14** with **G1** fingerprint) but rest of the compounds do not contain the FP-1. Then Leaf ID #9 and Leaf ID #10 consist of compounds with *nAR* < 1.5 and *nAR* ≥ 1.5, not containing FP-2 (active hydroxy group) in their structure. Compounds containing FP-2 in their structure shows overall poor inhibitory activity. By analyzing the compounds **A421**, **A418**, **A417**, **A413**, **A403**, **A395** shown in the **Figure 16** with bad Bayesian fingerprints reveals the negative contribution of FP-2 towards HDAC10 inhibitory activity. Next, the Leaf ID #12 with compounds having *nRB* < 5.5 (Example: **A386**, **A391**, **A398**, **A427**, **A430**, **A434**) not

containing FP-7 (indole or indazole moiety) are more potent than the molecules containing FP-10 (indole fused with cyclic five or six membered ring). The molecules (Example: **A302**, **A339**, **A343**) having fingerprint FP-10 with leaf ID #16 are comparatively more active than the molecules (Example: **A381**, **A389**, **A407**) with leaf ID #22 where *MFPSA* < 0.1855 or leaf ID #21 with *MFPSA* > 0.1855. This *MFPSA* is an important parameter for drug transport property and oral bioavailability.

On other hand, compounds **A351**, **A354**, **A362**, **A366**, **A373** with *nRB* > 5.5 containing FP-3 (indole moiety substituted at 1,5 position) with leaf ID #13 has more inhibitory activity and the compounds with lower activity than these compounds are further divided on the basis of the presence of **FP-4** (thiophene or thiazole ring). Under this **FP-4**, the compounds having **FP-5** further subdivided into two parts with compounds containing **FP-6** and **FP-11**. The compounds containing **FP-6** with leaf ID #27 and #28 are comparatively more active than the compounds containing **FP-11** with leaf ID #29 and #30.

Further, the compounds under **FP-4** without **FP-5** in the structure are divided on the basis of **FP-8**, and compounds with leaf ID #25 is active and rest of relatively lower active compounds are again divided on the basis of *nHBA*. Under this class, compounds with *nHBA* ≥ 6.5 with leaf ID #32 are more active than *nHBA* < 6.5. Now, the compounds under *nHBA* < 6.5 are further subclassified on the basis of number of rings (*nRings*). Compounds consisting of *nRings* ≥ 2.5 with leaf ID #36 shows better activity than the compounds consisting of *nRings* < 2.5. So, the compounds with *nRings* < 2.5 are divided whether the **FP-9** is present or absent in their structure. Compounds with leaf ID #37 are mostly higher actives than the compounds not containing **FP-9**. Compounds without **FP-9** are finally subdivided on the *nHBA* and compounds with *nHBA* < 4.5 with leaf ID #43 and leaf ID #44 with *nHBA* < 4.5 are the lowest active compounds with negligible HDAC10 inhibitory activity (Example: **A253-A263**, **A275**, **A285**, **A402**).

In brief, the RP study suggests the importance of carboxyamido group for potent inhibitory action. Interestingly, the good fingerprints generated through Bayesian model (Fingerprint **G1, G2, G7**) also suggest the importance of the same fingerprint in structure. Additionally, less inhibitory action shown by the compounds with indole 1,5 substituted analogues which is also in agreement with the Bayesian bad fingerprint **B2** (**Figure 15**). Hence, the RP study creates a crucial link with Bayesian fingerprints which is significant and acceptable for HDAC10 drug design.

Binding interactions of some potential HDAC10 inhibitors at their good fingerprint part

Some promising HDAC10 inhibitors bearing the good fingerprints in their structure are selected for the binding interaction studies. The selected compounds are (i) **A013** with fingerprint **G10, G19**, (ii) **A030** containing **G11** fingerprint and (iii) **A034** and **A045** having **G5-G6** fingerprints, which we have previously described in **Figure 14**. Till now, the human HDAC10 X-ray crystallographic structure has not been solved. We have selected the "humanized" version of HDAC10 from *Danio rerio* (drHDAC10; PDB ID: 6VNQ; Resolution: 2.05 Å) where the substitution of A24E and D94A residues mimics the human HDAC10 [3]. Interestingly, in all four compounds the hydroxamate moiety interacts to the catalytic Zn^{2+} ion in the active site (**Figure 18**). 3D binding interactions of compounds [(A) **A013**, (B) **A030**, (C) **A034** and (D) **A045]** and the active site of amino acid residues of HDAC10 (PDB ID: 6VNQ) are given in **Figure 19.** For compound **A013**, the **G10** fingerprint (α,γ-methyl alkene group) undergoes π-alkyl interaction with residues A94, F146, I27 (**Figure 19**).

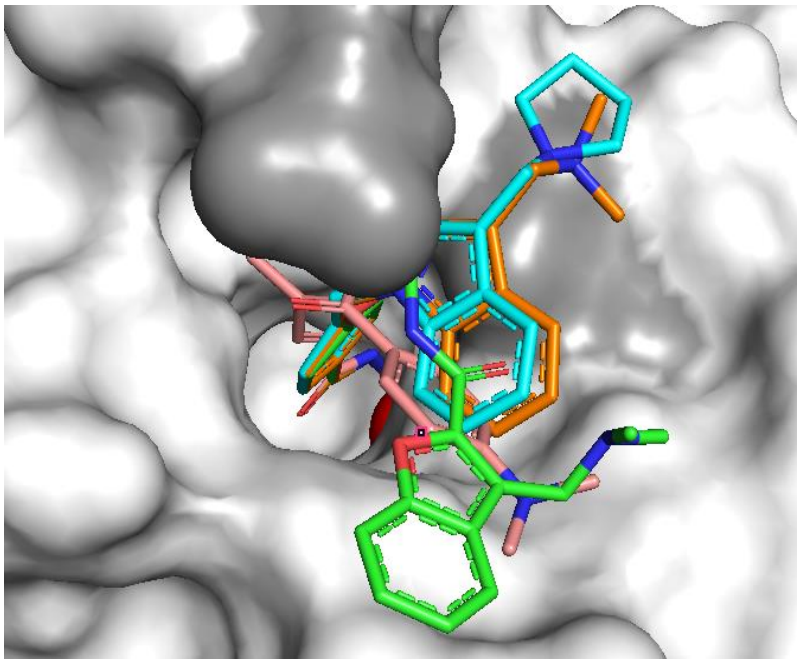


Figure 18. Binding orientations of HDAC10 inhibitors (A013, A030, A034, A045) inside the binding cavity of HDAC10.

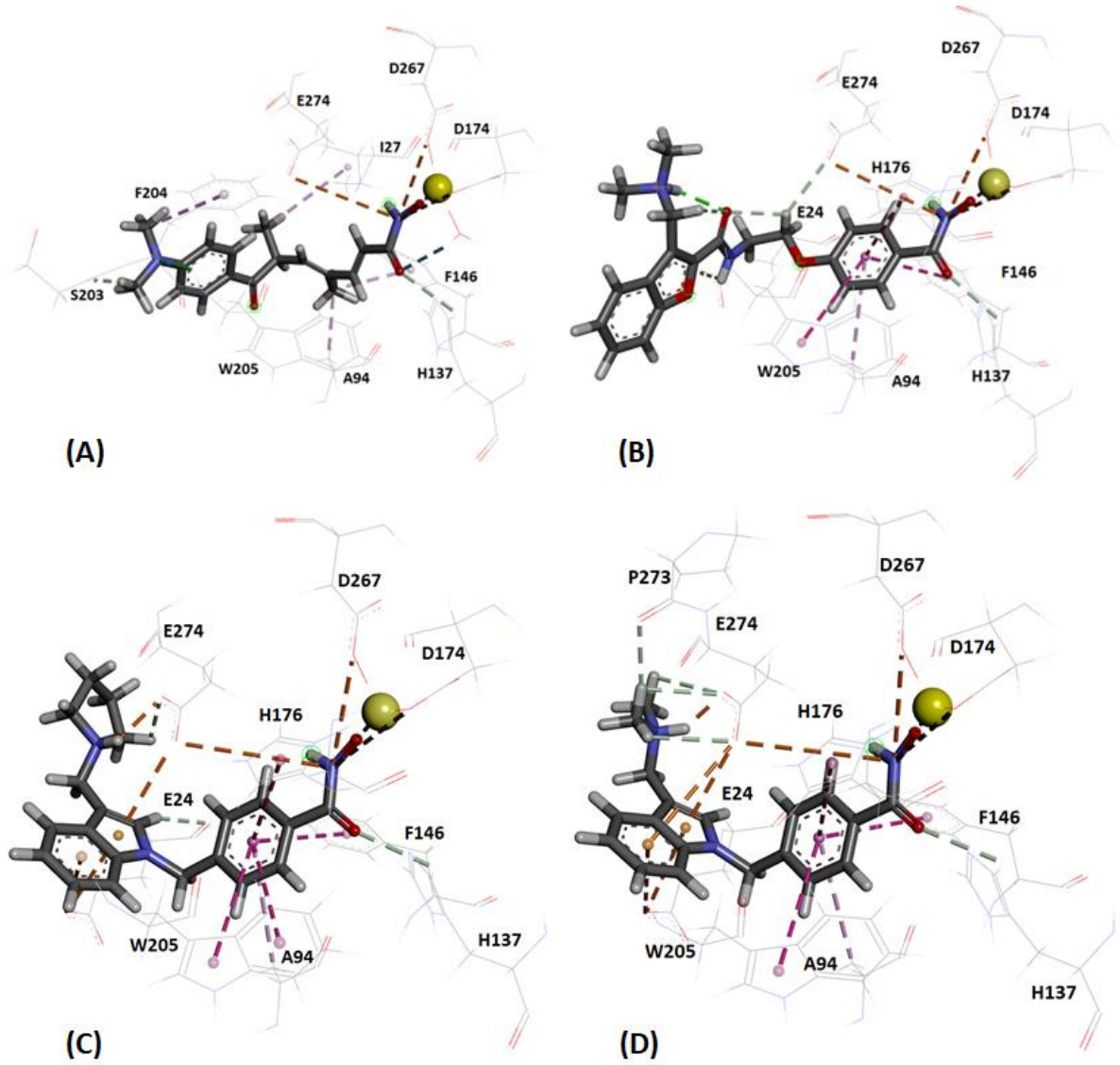


Figure 19. Binding interactions of (A) **A013**, (B) **A030**, (C) **A034** and (D) **A045** with the active site of amino acid residues of HDAC10 (PDB ID: 6VNQ). [Zinc ion is shown in yellow balls]

In **G19** fingerprint, the dimethyl amino group shows different types of interactions. It has interactions with S203 by carbon-hydrogen bond and also shows Van der Waals interaction with W205 and **π**-alkyl interaction with A94. Meanwhile, in compound A030, fingerprint **G11** (benzaldehyde moiety) shows π-π stacking interactionswith W205, F146, H176 and **π**-alkyl interaction with A94 as well as hydrogen bond interaction with H137. Compound **A034** and **A045** with **G5-G6** fingerprints shows π-anion interactions with E274 and E24 and π-π stacking interactions with W205, F146, H176 as well as parallel **π**-alkyl interaction with A94.

It is worth mentioning that the E274 residue conserved in HDAC10 enzyme as one of the "Gatekeeper" residues, and E24 is the "humanized" residue of tunnel like P(E,A)CE motif helix, which plays a key role for accommodating the ligands in the binding cavity. The hydroxamate moiety is the key scaffold to bind with these residues. It subsequently promotes the polyamine deacetylase (PDAC) activity and suppresses HDAC activity [21-22]. In a nutshell it can be stated that all the four compounds have followed almost same binding pattern inside the binding cavity (**Figure 18**). All the good fingerprints mentioned for the representative compounds has been participated in binding interaction with the binding pocket residues which reveals the significance of Bayesian fingerprints for HDAC10 inhibition. Water molecule has no significant effect in binding interaction. Overall, the fingerprints generated from our modeling studies can be future reference to design potential HDAC10 inhibitors.

Conclusion and future perspective

Different tumor and non-tumorous diseases have an evident pathophysiological role with HDAC10 in global health. Researchers have tried to develop different target based therapies, but most of them have failed due to the narrow range of safety and efficacy. Some recent researches revealed that out of all HDACs, HDAC10 is a crucial polyamine deacetylase which involves in several tumorous and non-tumorous diseases. Since the human HDAC10 crystal structure has not yet been elucidated, the ligand-based drug design is a viable alternative to accelerate the search for potential HDAC10 inhibitors. Though many HDAC10 inhibitors are undergoing various stages of clinical trials, unfortunately, most of them shows terrible adverse effects and failed at different phases clinical trials.

Here, the whole study is focused on the development of different ML models as well as crucial fingerprints regulating HDAC10 inhibitory properties. This study demonstrated statistically validated ML models to evaluate probable hits (HDAC10 inhibitors) from a pool of compounds. Out of all the classical ML models, Random Forest is the best classifier with the high performance and can be used as a model for screening of HDAC10 inhibitors. In addition, the statistically validated Bayesian classification and RP model successfully molecular fingerprints like carboxamido function, benzylindole and benzaldehyde are important for good inhibitory action. However, compounds containing bad fingerprints like indole 1,5 substituted analogues, aniline moiety with secondary amine as a linker group etc. are responsible for poor inhibition. Furthermore, the molecular docking study validated the probable binding patterns of the identified fingerprints at the HDAC10 active site. This current study also elucidated the importance of glutamate gatekeepers (E274 and E24) with binding mode analysis of selected inhibitors. Taken together, our *in-silico* approaches are capable to guide the chemist to design better active HDAC10 inhibitors. We hope that these will manifest a new outlook in the design of selective HDAC10 inhibitors in future. However, it is important to note that *in vitro* and *in vivo* studies are necessary in order to precisely conclude our findings.

References

References

[1] Amin, S. A., Adhikari, N., & Jha, T. (2017). Structure–activity relationships of hydroxamate-based histone deacetylase-8 inhibitors: reality behind anticancer drug discovery. *Future Medicinal Chemistry*, *9*(18), 2211-2237.

[2] Steimbach, R. R., Herbst-Gervasoni, C. J., Lechner, S., Stewart, T. M., Klinke, G., Ridinger, J., ... & Miller, A. K. (2022). Aza-SAHA derivatives are selective histone deacetylase 10 chemical probes that inhibit polyamine deacetylation and phenocopy HDAC10 knockout. *Journal of the American Chemical Society*, *144*(41), 18861-18875.

[3] Zeyen, P., Zeyn, Y., Herp, D., Mahmoudi, F., Yesiloglu, T. Z., Erdmann, F., ... & Sippl, W. (2022). Identification of histone deacetylase 10 (HDAC10) inhibitors that modulate autophagy in transformed cells. *European Journal of Medicinal Chemistry*, *234*, 114272.

[4] Cheng, F., Zheng, B., Wang, J., Zhao, G., Yao, Z., Niu, Z., & He, W. (2021). Histone deacetylase 10, a potential epigenetic target for therapy. *Bioscience Reports*, *41*(6), BSR20210462.

[5] Seto, E., & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harbor perspectives in biology*, *6*(4), a018713.

[6] Caron, C., Boyault, C., & Khochbin, S. (2005). Regulatory cross-talk between lysine acetylation and ubiquitination: role in the control of protein stability. *Bioessays*, *27*(4), 408-415. [7] F. Fischle, and Y. Wang, Histone and chromatin cross-talk, Curr. Opin. Cell Biol., 15 (2003), pp. 172–183. doi:10.1016/S0955-0674(03)00013-9

[8] Chakrabarti, A., Oehme, I., Witt, O., Oliveira, G., Sippl, W., Romier, C., ... & Jung, M. (2015). HDAC8: a multifaceted target for therapeutic interventions. *Trends in pharmacological sciences*, *36*(7), 481-492.

[9] Pojani, E., & Barlocco, D. (2022). Selective Inhibitors of Histone Deacetylase 10 (HDAC-10). *Current Medicinal Chemistry*, *29*(13), 2306-2321.

[10] Allfrey, V. G., Faulkner, R., & Mirsky, A. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences*, *51*(5), 786-794.

[11] Inoue, A., & Fujimoto, D. (1969). Enzymatic deacetylation of histone. *Biochemical and biophysical research communications*, *36*(1), 146-150.

[12] Grozinger, C. M., & Schreiber, S. L. (2002). Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. *Chemistry & biology*, *9*(1), 3-16..

[13] Ruijter, A. J. D., GENNIP, A. H. V., Caron, H. N., Kemp, S., & KUILENBURG, A. B. V. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochemical Journal*, *370*(3), 737-749.

[14] Dokmanovic, M., Clarke, C., & Marks, P. A. (2007). Histone deacetylase inhibitors: overview and perspectives. *Molecular cancer research*, *5*(10), 981-989.

[15] Bolden, J. E., Peart, M. J., & Johnstone, R. W. (2006). Anticancer activities of histone deacetylase inhibitors. *Nature reviews Drug discovery*, *5*(9), 769-784.

[16] Lakshmaiah, K. C., Jacob, L. A., Aparna, S., Lokanatha, D., & Saldanha, S. C. (2014). Epigenetic therapy of cancer with histone deacetylase inhibitors. *Journal of cancer research and therapeutics*, *10*(3), 469-478.

[17] Uba, A. I., & Yelekci, K. (2020). Crystallographic structure versus homology model: a case study of molecular dynamics simulation of human and zebrafish histone deacetylase 10. *Journal of Biomolecular Structure and Dynamics*, *38*(15), 4397-4406.

[18] Guardiola, A. R., & Yao, T. P. (2002). Molecular cloning and characterization of a novel histone deacetylase HDAC10. *Journal of Biological Chemistry*, *277*(5), 3350-3356.

[19] Tong, J. J., Liu, J., Bertos, N. R., & Yang, X. J. (2002). Identification of HDAC10, a novel class II human histone deacetylase containing a leucine-rich domain. *Nucleic acids research*, *30*(5), 1114-1123.

[20] Fischer, D. D., Cai, R., Bhatia, U., Asselbergs, F. A., Song, C., Terry, R., ... & Cohen, D. (2002). Isolation and characterization of a novel class II histone deacetylase, HDAC10. *Journal of Biological Chemistry*, *277*(8), 6656-6666.

[21] Herbst-Gervasoni, C. J., & Christianson, D. W. (2019). Binding of N 8-acetylspermidine analogues to histone deacetylase 10 reveals molecular strategies for blocking polyamine deacetylation. *Biochemistry*, *58*(49), 4957-4969.

[22] Hai, Y., Shinsky, S. A., Porter, N. J., & Christianson, D. W. (2017). Histone deacetylase 10 structure and molecular function as a polyamine deacetylase. *Nature communications*, *8*(1), 15368.

[23Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, *144*(5), 646-674.

[24] Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell*, *100*(1), 57-70.

[25] Yang, Y., Huang, Y., Wang, Z., Wang, H. T., Duan, B., Ye, D., ... & Kang, J. (2016). HDAC10 promotes lung cancer proliferation via AKT phosphorylation. *Oncotarget*, *7*(37), 59388.

[26] Li, Y., Zhang, X., Zhu, S., Dejene, E. A., Peng, W., Sepulveda, A., & Seto, E. (2020). HDAC10 regulates cancer stem-like cell properties in KRAS-driven lung adenocarcinoma. *Cancer research*, *80*(16), 3265-3278.

[27] Fan, W., Huang, J., & Xiao, H. (2015). Histone deacetylase 10 suppresses proliferation and invasion by inhibiting the phosphorylation of β-catenin and serves as an independent prognostic factor for human clear cell renal cell carcinoma. *International journal of clinical and experimental medicine*, *8*(3), 3734.

[28] Revathidevi, S., & Munirajan, A. K. (2019, December). Akt in cancer: Mediator and more. In *Seminars in cancer biology* (Vol. 59, pp. 80-91). Academic Press.

[29] Guertin, D. A., & Sabatini, D. M. (2007). Defining the role of mTOR in cancer. *Cancer cell*, *12*(1), 9-22.

[30] Shan, C., Lu, Z., Li, Z., Sheng, H., Fan, J., Qi, Q., ... & Zhang, S. (2019). 4-hydroxyphenylpyruvate dioxygenase promotes lung cancer growth via pentose phosphate pathway (PPP) flux mediated by LKB1-AMPK/HDAC10/G6PD axis. *Cell death & disease*, *10*(7), 525.

[31] Evan, G. I., & Vousden, K. H. (2001). Proliferation, cell cycle and apoptosis in cancer. *nature*, *411*(6835), 342-348.

[32] Li, Y., Peng, L., & Seto, E. (2015). Histone deacetylase 10 regulates the cell cycle g2/m phase transition via a novel let-7–HMGA2–cyclin a2 pathway. *Molecular and cellular biology*, *35*(20), 3547-3565.

[33] Coqueret, O. (2003). New roles for p21 and p27 cell-cycle inhibitors: a function for each cell compartment?. *Trends in cell biology*, *13*(2), 65-70.

[34] Qie, S., & Diehl, J. A. (2016). Cyclin D1, cancer progression, and opportunities in cancer treatment. *Journal of molecular medicine*, *94*, 1313-1326.35 Sutherland, R.L. and Musgrove, E.A. (2004) Cyclins and breast cancer. *J. Mamm. Gland Biol. Neoplasia* **9**, 95–104

[35] Sutherland, R. L., & Musgrove, E. A. (2004). Cyclins and breast cancer. *Journal of mammary gland biology and neoplasia*, *9*, 95-104.

[36] Wong, R. S. (2011). Apoptosis in cancer: from pathogenesis to treatment. *Journal of experimental & clinical cancer research*, *30*(1), 1-14.

[37] Gross, A., McDonnell, J. M., & Korsmeyer, S. J. (1999). BCL-2 family members and the mitochondria in apoptosis. *Genes & development*, *13*(15), 1899-1911.

[38] Reed, J. C. (1997, October). Bcl-2 family proteins: regulators of apoptosis and chemoresistance in hematologic malignancies. In *Seminars in hematology* (Vol. 34, No. 4 Suppl 5, pp. 9-19).

[39] Götze, S., Coersmeyer, M., Müller, O., & Sievers, S. (2014). Histone deacetylase inhibitors induce attenuation of Wnt signaling and TCF7L2 depletion in colorectal carcinoma cells. *International journal of oncology*, *45*(4), 1715-1723.

[40] Circu, M. L., & Aw, T. Y. (2010). Reactive oxygen species, cellular redox systems, and apoptosis. *Free radical biology and medicine*, *48*(6), 749-762.

[41] Zorov, D. B., Juhaszova, M., & Sollott, S. J. (2014). Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. *Physiological reviews*, *94*(3), 909-950.

[42] Lee, J. H., Jeong, E. G., Choi, M. C., Kim, S. H., Park, J. H., Song, S. H., ... & Kim, T. Y. (2010). Inhibition of histone deacetylase 10 induces thioredoxin-interacting protein and causes accumulation of reactive oxygen species in SNU-620 human gastric cancer cells. *Molecules and cells*, *30*, 107-112.

[43] Polacheck, W. J., Zervantonakis, I. K., & Kamm, R. D. (2013). Tumor cell migration in complex microenvironments. *Cellular and Molecular Life Sciences*, *70*, 1335-1356.

[44] Hamidi, H., & Ivaska, J. (2018). Every step of the way: integrins in cancer progression and metastasis. *Nature Reviews Cancer*, *18*(9), 533-548.

[45] Saiki, Y., & Horii, A. (2019). Multiple functions of S100A10, an important cancer promoter. *Pathology international*, *69*(11), 629-636.

[46] Castro-Castro, A., Marchesin, V., Monteiro, P., Lodillinsky, C., Rossé, C., & Chavrier, P. (2016). Cellular and molecular mechanisms of MT1-MMP-dependent cancer cell invasion. *Annual review of cell and developmental biology*, *32*, 555-576.

[47] Cui, N., Hu, M., & Khalil, R. A. (2017). Biochemical and biological attributes of matrix metalloproteinases. *Progress in molecular biology and translational science*, *147*, 1-73.

[48] Zhao, W., Han, H. B., & Zhang, Z. Q. (2011). Suppression of lung cancer cell invasion and metastasis by connexin43 involves the secretion of follistatin-like 1 mediated via histone acetylation. *The international journal of biochemistry & cell biology*, *43*(10), 1459-1468.

[49] Song, C., Zhu, S., Wu, C., & Kang, J. (2013). Histone deacetylase (HDAC) 10 suppresses cervical cancer metastasis through inhibition of matrix metalloproteinase (MMP) 2 and 9 expression. *Journal of Biological Chemistry*, *288*(39), 28021-28033.

[50] Dong, Y. W., Wang, R., Cai, Q. Q., Qi, B., Wu, W., Zhang, Y. H., & Wu, X. Z. (2014). Sulfatide epigenetically regulates miR-223 and promotes the migration of human hepatocellular carcinoma cells. *Journal of hepatology*, *60*(4), 792-801.

[51] Von Schulz-Hausmann, S. A., Schmeel, L. C., Schmeel, F. C., & Schmidt-Wolf, I. G. (2014). Targeting the Wnt/beta-catenin pathway in renal cell carcinoma. *Anticancer research*, *34*(8), 4101-4108.

[52] Shang, S., Hua, F., & Hu, Z. W. (2017). The regulation of β-catenin activity and function in cancer: therapeutic opportunities. *Oncotarget*, *8*(20), 33972.

[53] Kerbel, R., & Folkman, J. (2002). Clinical translation of angiogenesis inhibitors. *Nature Reviews Cancer*, *2*(10), 727-739.

[54] Arias, H. R., Richards, V. E., Ng, D., Ghafoori, M. E., Le, V., & Mousa, S. A. (2009). Role of non-neuronal nicotinic acetylcholine receptors in angiogenesis. *The international journal of biochemistry & cell biology*, *41*(7), 1441-1451.

[55] Rak, J., Joanne, L. Y., Klement, G., & Kerbel, R. S. (2000, December). Oncogenes and angiogenesis: signaling three-dimensional tumor growth. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 5, No. 1, pp. 24-33). Elsevier.

[56] Relf, M., LeJeune, S., Scott, P. A., Fox, S., Smith, K., Leek, R., ... & Harris, A. L. (1997). Expression of the angiogenic factors vascular endothelial cell growth factor, acidic and basic fibroblast growth factor, tumor growth factor β-1, platelet-derived endothelial cell growth factor, placenta growth factor, and pleiotrophin in human primary breast cancer and its relation to angiogenesis. *Cancer research*, *57*(5), 963-969.

[57] Duan, B., Ye, D., Zhu, S., Jia, W., Lu, C., Wang, G., ... & Kang, J. (2017). HDAC10 promotes angiogenesis in endothelial cells through the PTPN22/ERK axis. *Oncotarget*, *8*(37), 61338.

[58] Park, J. H., Kim, S. H., Choi, M. C., Lee, J., Oh, D. Y., Im, S. A., ... & Kim, T. Y. (2008). Class II histone deacetylases play pivotal roles in heat shock protein 90-mediated proteasomal degradation of vascular endothelial growth factor receptors. *Biochemical and biophysical research communications*, *368*(2), 318-322.

[59] Holohan, C., Van Schaeybroeck, S., Longley, D. B., & Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer*, *13*(10), 714-726.

[60] Longley, D. B., & Johnston, P. G. (2005). Molecular mechanisms of drug resistance. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, *205*(2), 275-292.

[61] Gottesman, M. M., Fojo, T., & Bates, S. E. (2002). Multidrug resistance in cancer: role of ATP–dependent transporters. *Nature reviews cancer*, *2*(1), 48-58.

[62] Ridinger, J., Koeneke, E., Kolbinger, F. R., Koerholz, K., Mahboobi, S., Hellweg, L., ... & Oehme, I. (2018). Dual role of HDAC10 in lysosomal exocytosis and DNA repair promotes neuroblastoma chemoresistance. *Scientific reports*, *8*(1), 10039.

[63] Li, Y.J., Lei, Y.H., Yao, N., Wang, C.R., Hu, N., Ye, W.C. et al. (2017) Autophagy and multidrug resistance in cancer. *Chin. J. Cancer* **36**, 52, https://doi.org/10.1186/s40880-017-0219-2

[64] Oehme, I., Linke, J. P., Böck, B. C., Milde, T., Lodrini, M., Hartenstein, B., ... & Witt, O. (2013). Histone deacetylase 10 promotes autophagy-mediated cell survival. *Proceedings of the National Academy of Sciences*, *110*(28), E2592-E2601.

[65] Oehme, I., Lodrini, M., Brady, N. R., & Witt, O. (2013). Histone deacetylase 10-promoted autophagy as a druggable point of interference to improve the treatment response of advanced neuroblastomas. *Autophagy*, *9*(12), 2163-2165.

[66] Islam, M. M., Banerjee, T., Packard, C. Z., Kotian, S., Selvendiran, K., Cohn, D. E., & Parvin, J. D. (2017). HDAC10 as a potential therapeutic target in ovarian cancer. *Gynecologic oncology*, *144*(3), 613-620.

[67] Bartkova, J., Hořejší, Z., Koed, K., Krämer, A., Tort, F., Zieger, K., ... & Bartek, J. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*, *434*(7035), 864-870.

[68] Richard, D. J., Savage, K., Bolderson, E., Cubeddu, L., So, S., Ghita, M., ... & Khanna, K. K. (2011). hSSB1 rapidly binds at the sites of DNA double-strand breaks and is required for the efficient recruitment of the MRN complex. *Nucleic acids research*, *39*(5), 1692-1702.

[69] Wu, Y., Chen, H., Lu, J., Zhang, M., Zhang, R., Duan, T., ... & Kang, T. (2015). Acetylation-dependent function of human single-stranded DNA binding protein 1. *Nucleic acids research*, *43*(16), 7878-7887.

[70] Shafi, A. A., Yen, A. E., & Weigel, N. L. (2013). Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacology & therapeutics*, *140*(3), 223-238.

[71] Welti, J., Sharp, A., Yuan, W., Dolling, D., Nava Rodrigues, D., Figueiredo, I., ... & de Bono, J. S. (2018). Targeting bromodomain and extra-terminal (BET) family proteins in castration-resistant prostate cancer (CRPC). *Clinical Cancer Research*, *24*(13), 3149-3162.

[72] Dai, X., Gan, W., Li, X., Wang, S., Zhang, W., Huang, L., ... & Wei, W. (2017). Prostate cancer–associated SPOP mutations confer resistance to BET inhibitors through stabilization of BRD4. *Nature medicine*, *23*(9), 1063-1071.

[73] Jin, X., Yan, Y., Wang, D., Ding, D., Ma, T., Ye, Z., ... & Huang, H. (2018). DUB3 promotes BET inhibitor resistance and cancer progression by deubiquitinating BRD4. *Molecular cell*, *71*(4), 592-605.

[74] Sun, H., Mediwala, S. N., Szafran, A. T., Mancini, M. A., & Marcelli, M. (2016). CUDC-101, a novel inhibitor of full-length androgen receptor (flAR) and androgen receptor variant 7

(AR-V7) activity: mechanism of action and in vivo efficacy. *Hormones and Cancer*, *7*, 196-210.

[75] Saygin, C., Matei, D., Majeti, R., Reizes, O., & Lathia, J. D. (2019). Targeting cancer stemness in the clinic: from hype to hope. *Cell stem cell*, *24*(1), 25-40.

[76] Batlle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nature medicine*, *23*(10), 1124-1134.

[77] Guo, W., Keckesova, Z., Donaher, J. L., Shibue, T., Tischler, V., Reinhardt, F., ... & Weinberg, R. A. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell*, *148*(5), 1015-1028.

[78] Ye, X., Tam, W.L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E. et al. (2015) Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260, https://doi.org/10.1038/nature14897

[79] Dahiya, S., Beier, U. H., Wang, L., Han, R., Jiao, J., Akimova, T., ... & Hancock, W. W. (2020). HDAC10 deletion promotes Foxp3+ T-regulatory cell function. *Scientific reports*, *10*(1), 424.

[80] Chen, Y., Song, Y., Du, W., Gong, L., Chang, H., & Zou, Z. (2019). Tumor-associated macrophages: an accomplice in solid tumor progression. *Journal of biomedical science*, *26*(1), 1-13.

[81] Kaushik, S., & Cuervo, A. M. (2018). The coming of age of chaperone-mediated autophagy. *Nature reviews Molecular cell biology*, *19*(6), 365-381.

[82] Obayashi, H., Nagano, Y., Takahashi, T., Seki, T., Tanaka, S., Sakai, N., ... & Maruyama, H. (2020). Histone deacetylase 10 knockout activates chaperone-mediated autophagy and accelerates the decomposition of its substrate. *Biochemical and biophysical research communications*, *523*(1), 246-252.

[83] Arias, E. (2017). Methods to study chaperone-mediated autophagy. *Methods in enzymology*, *588*, 283-305.

[84] Kaushik, S., & Cuervo, A. M. (2009). Methods to monitor chaperone-mediated autophagy. *Methods in enzymology*, *452*, 297-324.

[85] Baretti, M., & Le, D. T. (2018). DNA mismatch repair in cancer. *Pharmacology & therapeutics*, *189*, 45-62.

[86] Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell research*, *18*(1), 85-98.

[87] Radhakrishnan, R., Li, Y., Xiang, S., Yuan, F., Yuan, Z., Telles, E., ... & Seto, E. (2015). Histone deacetylase 10 regulates DNA mismatch repair and may involve the deacetylation of MutS homolog 2. *Journal of Biological Chemistry*, *290*(37), 22795-22804.

[88] Tao, X., Yan, Y., Lu, L., & Chen, B. (2017). HDAC10 expression is associated with DNA mismatch repair gene and is a predictor of good prognosis in colon carcinoma. *Oncology letters*, *14*(4), 4923-4929.

[89] Park, B. L., Kim, Y. J., Cheong, H. S., Lee, S. O., Han, C. S., Yoon, J. H., ... & Shin, H. D. (2007). HDAC10 promoter polymorphism associated with development of HCC among chronic HBV patients. *Biochemical and biophysical research communications*, *363*(3), 776-781.

[90] Fonseca, A. L., Kugelberg, J., Starker, L. F., Scholl, U., Choi, M., Hellman, P., ... & Carling, T. (2012). Comprehensive DNA methylation analysis of benign and malignant adrenocortical tumors. *Genes, Chromosomes and Cancer*, *51*(10), 949-960.

[91] Uzdensky, A., Demyanenko, S., Bibov, M., Sharifulina, S., Kit, O., Przhedetski, Y., & Pozdnyakova, V. (2014). Expression of proteins involved in epigenetic regulation in human cutaneous melanoma and peritumoral skin. *Tumor Biology*, *35*, 8225-8233.

[92] Kasaian, K., Wiseman, S. M., Walker, B. A., Schein, J. E., Zhao, Y., Hirst, M., ... & Jones, S. J. (2015). The genomic and transcriptomic landscape of anaplastic thyroid cancer: implications for therapy. *BMC cancer*, *15*, 1-11.

[93] Lourdusamy, A., Rahman, R., & Grundy, R. G. (2015). Expression alterations define unique molecular characteristics of spinal ependymomas. *Oncotarget*, *6*(23), 19780.

[94] Codd, R., Braich, N., Liu, J., Soe, C. Z., & Pakchung, A. A. (2009). Zn (II)-dependent histone deacetylase inhibitors: suberoylanilide hydroxamic acid and trichostatin A. *The international journal of biochemistry & cell biology*, *41*(4), 736-739.

[94] Buggy, J. J., Cao, Z. A., Bass, K. E., Verner, E., Balasubramanian, S., Liu, L., ... & Dalrymple, S. A. (2006). CRA-024781: a novel synthetic inhibitor of histone deacetylase enzymes with antitumor activity in vitro and in vivo. *Molecular cancer therapeutics*, *5*(5), 1309-1317.

[95] Cao, Z. A., Bass, K. E., Balasubramanian, S., Liu, L., Schultz, B., Verner, E., ... & Buggy, J. J. (2006). CRA-026440: a potent, broad-spectrum, hydroxamic histone deacetylase inhibitor with antiproliferative and antiangiogenic activity in vitro and in vivo. *Molecular cancer therapeutics*, *5*(7), 1693-1701.

[96] Booth, L., Roberts, J. L., Poklepovic, A., Kirkwood, J., & Dent, P. (2017). HDAC inhibitors enhance the immunotherapy response of melanoma cells. *Oncotarget*, *8*(47), 83155.

[97] De Souza, C., & P Chatterji, B. (2015). HDAC inhibitors as novel anti-cancer therapeutics. *Recent patents on anti-cancer drug discovery*, *10*(2), 145-162.

[98] Gurvich, N., Tsygankova, O. M., Meinkoth, J. L., & Klein, P. S. (2004). Histone Deacetylase Is a Target of Valproic Acid-Mediated Cellular Differentiation. In Cancer Research (Vol. 64, Issue 3, pp. 1079–1086). American Association for Cancer Research (AACR). https://doi.org/10.1158/0008-5472.can-03-0799

[99] Brush, M. H., Guardiola, A., Connor, J. H., Yao, T. P., & Shenolikar, S. (2004). Deactylase inhibitors disrupt cellular complexes containing protein phosphatases and deacetylases. *Journal of Biological Chemistry*, *279*(9), 7685-7691.

[100] Bantscheff, M., Hopf, C., Savitski, M. M., Dittmann, A., Grandi, P., Michon, A. M., ... & Drewes, G. (2011). Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nature biotechnology*, *29*(3), 255-265.

[101] Géraldy, M., Morgen, M., Sehr, P., Steimbach, R. R., Moi, D., Ridinger, J., ... & Miller, A. K. (2019). Selective inhibition of histone deacetylase 10: hydrogen bonding to the gatekeeper residue is implicated. *Journal of medicinal Chemistry*, *62*(9), 4426-4443.

[102] Villadsen, J. S., Stephansen, H. M., Maolanon, A. R., Harris, P., & Olsen, C. A. (2013). Total synthesis and full histone deacetylase inhibitory profiling of azumamides A–E as well as β2-epi-azumamide E and β3-epi-azumamide E. *Journal of medicinal chemistry*, *56*(16), 6512-6520.

[103] Balasubramanian, S., Ramos, J., Luo, W., Sirisawad, M., Verner, E., & Buggy, J. J. (2008). A novel histone deacetylase 8 (HDAC8)-specific inhibitor PCI-34051 induces apoptosis in T-cell lymphomas. *Leukemia*, *22*(5), 1026-1034.

[104] Morgen, M., Steimbach, R. R., Géraldy, M., Hellweg, L., Sehr, P., Ridinger, J., ... & Miller, A. K. (2020). Design and synthesis of dihydroxamic acids as HDAC6/8/10 inhibitors. *ChemMedChem*, *15*(13), 1163-1174.

[105] Pojani, E., & Barlocco, D. (2022). Selective Inhibitors of Histone Deacetylase 10 (HDAC-10). *Current Medicinal Chemistry*, *29*(13), 2306-2321.

[106] Adewole, K. E., & Ishola, A. A. (2021). A computational approach to investigate the HDAC6 and HDAC10 binding propensity of psidium guajava-derived compounds as potential anticancer agents. *Current drug discovery technologies*, *18*(3), 423-436.

[107] Liu, T., Lin, Y., Wen, X., Jorissen, R. N., & Gilson, M. K. (2007). BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic acids research*, *35*(suppl_1), D198-D201.

[108] Discovery Studio 3.0 (DS 3.0), Accelrys Inc., CA, USA, 2015. Available at www.accelrys.com

[109] Moinul, M., Amin, S. A., Kumar, P., Patil, U. K., Gajbhiye, A., Jha, T., & Gayen, S. (2022). Exploring sodium glucose cotransporter (SGLT2) inhibitors with Machine Learning Approach: A novel hope in anti-diabetes drug discovery. *Journal of Molecular Graphics and Modelling*, *111*, 108106.

[110] Bajorath, J., Kearnes, S., Walters, W. P., Meanwell, N. A., Georg, G. I., & Wang, S. (2020). Artificial intelligence in drug discovery: into the great wide open. *Journal of medicinal chemistry*, *63*(16), 8651-8652..

[111] Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, *32*(7), 1466-1474.

[112] Nandi, S., Kumar, P., Amin, S. A., Jha, T., & Gayen, S. (2021). First molecular modelling report on tri-substituted pyrazolines as phosphodiesterase 5 (PDE5) inhibitors through classical and machine learning based multi-QSAR analysis. *SAR and QSAR in Environmental Research*, *32*(11), 917-939.

[113] Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315-1316.

[114] de Moraes Batista, A. F., Miraglia, J. L., Rizzi Donato, T. H., & Porto Chiavegatto Filho, A. D. (2020). COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *MedRxiv*, 2020-04.

[115] "Supervised learning — scikit-learn 0.23.2 documentation." https://scikit-learn.org/stable/supervised_learning.html#supervised-learning (accessed Aug. 31, 2020).

[116] "sklearn.tree.DecisionTreeClassifier — scikit-learn 0.23.2 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html (accessed Aug. 31, 2020).

[117] "sklearn.ensemble.RandomForestClassifier — scikit-learn 0.23.2 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (accessed Aug. 31, 2020).

[118] "sklearn.ensemble.GradientBoostingClassifier — scikit-learn 0.24.1 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html (accessed Mar. 16, 2021).

[119] "sklearn.discriminant_analysis.LinearDiscriminantAnalysis — scikit-learn 0.23.2 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html (accessed Aug. 31, 2020).

[120] Theodoridis, S. (2015). *Machine learning: a Bayesian and optimization perspective*. Academic press.

[121] Amin, S. A., Adhikari, N., & Jha, T. (2020). Exploration of histone deacetylase 8 inhibitors through classification QSAR study: Part II. *Journal of Molecular Structure*, *1204*, 127529.

[122] Chen, L., Li, Y., Zhao, Q., Peng, H., & Hou, T. (2011). ADME evaluation in drug discovery. 10. Predictions of P-glycoprotein inhibitors using recursive partitioning and naive Bayesian classification techniques. *Molecular Pharmaceutics*, *8*(3), 889-900.

[123] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

[124] Amin, S. A., Nandi, S., Kashaw, S. K., Jha, T., & Gayen, S. (2022). A critical analysis of urea transporter B inhibitors: molecular fingerprints, pharmacophore features for the development of next-generation diuretics. *Molecular Diversity*, 1-11.

[125] https://www.rcsb.org/structure/6vnq (as accessed on 8th October 2022)
[126] Schrödinger Suite, 2018. Schrödinger. LLC, New York, USA. http://www.schrödinger.com/glide

Published articles

➢ Journal of Molecular Graphics and Modelling

Journal of Molecular Graphics and Modelling 123 (2023) 108510

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

Exploring structural requirements of HDAC10 inhibitors through comparative machine learning approaches

Arijit Bhattacharya [a, 1], Sk Abdul Amin [b, c, 1], Prabhat Kumar [d], Tarun Jha [c, **], Shovanlal Gayen [a, *]

[a] Laboratory of Drug Design and Discovery, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India
[b] Department of Pharmaceutical Technology, JIS University, 81, Nilgunj Road, Agarpara, Kolkata, West Bengal, India
[c] Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India
[d] Department of Computer Science, Institute of Science, Banaras Hindu University, Varanasi, 221005, India

ARTICLE INFO

Keywords:
Histone deacetylase 10 (HDAC10) inhibitors
Bayesian classification
Recursive partitioning
Decision tree
Machine learning
Molecular docking

ABSTRACT

Histone deacetylase (HDAC) inhibitors are in the limelight of anticancer drug development and research. HDAC10 is one of the class-IIb HDACs, responsible for cancer progression. The search for potent and effective HDAC10 selective inhibitors is going on. However, the absence of human HDAC10 crystal/NMR structure hampers the structure-based drug design of HDAC10 inhibitors. Different ligand-based modeling techniques are the only hope to speed up the inhibitor design. In this study, we applied different ligand-based modeling techniques on a diverse set of HDAC10 inhibitors (n = 484). Machine learning (ML) models were developed that could be used to screen unknown compounds as HDAC10 inhibitors from a large chemical database. Moreover, Bayesian classification and Recursive partitioning models were used to identify the structural fingerprints regulating the HDAC10 inhibitory activity. Additionally, a molecular docking study was performed to understand the binding pattern of the identified structural fingerprints towards the active site of HDAC10. Overall, the modeling insight might offer helpful information for medicinal chemists to design and develop efficient HDAC10 inhibitors.

1. Introduction

Histone deacetylases (HDACs) are called the "REMOVER" enzymes as these eliminate the -acetate moiety from acetylated ε-amino groups of histone lysine and other non-histone proteins [1–4]. They belong to the category of Zn^{2+} or nicotinamide adenine dinucleotide (NAD^+)-dependent proteolytic enzymes, involved in transcriptional subduing as well as chromatin condensation. HDACs execute post-translational changes, such as ubiquitination and methylation, and can exert effects on gene transcription by increasing the interactivity between DNA and histone [5–7]. The deacetylation nature of the HDACs takes part as a modulator or regulator in several types of bio-signaling pathways [8]. Thus, they have crucial pathophysiological role in several deadliest maladies including neurodegenerative dysfunctions, inflammation, metabolic disorders, autoimmune disorders, and cancers etc. [8–11]. The histone deacetylases family comprises 18 members [12,13] which are further divided into four classes depending on their homological sequences with yeast HDACs [14,15]. The four classes are class I HDACs (HDAC 1,2,3 and 8) [14]; class II HDACs, which are further subclassified into class IIa (HDAC 4,5,7, and 9) and class IIb (HDAC 6 and 10) [4]; class III which resembles yeast SIR2 (SIRT 1–7) protein structures and class IV that comprises only HDAC11 [16].

HDAC10 is an essential representative of class IIb HDACs and configurationally indistinguishable from the HDAC6 isotype. It is localized in cytoplasm and has poor lysine deacetylase activity. It can identify polyamines as substrate and thus, can be named as polyamine deacetylase [2,3,17,18]. The *HDAC10* gene is confined to chromosome number 22 [18], includes 20 exons with two spliced transcripts [19,20] and contains an N-terminal catalytic domain and a leucine-rich domain with a C-terminal end. The N-terminal catalytic domain of HDAC10 is analogous to the deacetylase domain of other class II HDACs, but the C-terminal catalytic domain does not contain any residues that are

* Corresponding author.
** Corresponding author.
E-mail addresses: tjupharm@yahoo.com (T. Jha), shovanlal.gayen@gmail.com (S. Gayen).
[1] Authors have equal contribution.

https://doi.org/10.1016/j.jmgm.2023.108510
Received 24 February 2023; Received in revised form 26 April 2023; Accepted 4 May 2023
Available online 16 May 2023

➢ Molecular Diversity

Molecular Diversity
https://doi.org/10.1007/s11030-023-10670-2

ORIGINAL ARTICLE

Exploring molecular fingerprints of different drugs having bile interaction: a stepping stone towards better drug delivery

Sourav Sardar[1] · Arijit Bhattacharya[1] · Sk. Abdul Amin[2] · Tarun Jha[3] · Shovanlal Gayen[1]

Received: 20 January 2023 / Accepted: 10 June 2023

Abstract
Bile acids are amphiphilic substances produced naturally in humans. In the context of drug delivery and dosage form design, it is critical to understand whether a drug interacts with bile inside the gastrointestinal (GI) tract or not. This study focuses on the identification of structural fingerprints/features important for bile interaction. Molecular modelling methods such as Bayesian classification and recursive partitioning (RP) studies are executed to find important fingerprints/features for the bile interaction. For the Bayesian classification study, the *ROC* score of 0.837 and 0.950 are found for the training set and the test set compounds, respectively. The fluorine-containing aliphatic/aromatic group, the branched chain of the alkyl group containing hydroxyl moiety and the phenothiazine ring etc. are identified as good fingerprints having a positive contribution towards bile interactions, whereas, the bad fingerprints such as free carboxylate group, purine, and pyrimidine ring etc. have a negative contribution towards bile interactions. The best tree (tree ID: 1) from the RP study classifies the bile interacting or non-interacting compounds with a *ROC* score of 0.941 for the training and 0.875 for the test set. Additionally, SARpy and QSAR-Co analyses are also been performed to classify compounds as bile interacting/non-interacting. Moreover, forty-six recently FDA-approved drugs have been screened by the developed SARpy and QSAR-Co models to assess their bile interaction properties. Overall, this attempt may facilitate the researchers to identify bile interacting/non-interacting molecules in a faster way and help in the design of formulations and target-specific drug development.

Sourav Sardar and Arijit Bhattacharya have contributed equally to this work.

✉ Tarun Jha
tjupharm@yahoo.com

✉ Shovanlal Gayen
shovanlal.gayen@gmail.com

[1] Laboratory of Drug Design and Discovery, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

[2] Department of Pharmaceutical Technology, JIS University, 81, Nilgunj Road, Agarpara, Kolkata, West Bengal, India

[3] Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

Appendix

Table S1. Training set compounds in SMILES format

Compound name	SMILES format	IC_{50} value (nM)	Binary
A001	[C@H]1(NC(=O)[C@]2(N=C(c3csc(CNC(=O)C[C@H](OC1=O)/C=C/CCSSCC/C=C/[C@H]1OC(=O)[C@@H](NC(=O)[C@@]4(CSC(=N4)c4csc(n4)CNC(=O)C1)C)C(C)C)n3)SC2)C)C(C)C	0.2	1
A002	s1c2nc(c1)C1=N[C@@](CS1)(C(=O)N[C@@H](C(=O)O[C@H](/C=C/CCS)CC(=O)NC2)C(C)C)C	0.5	1
A003	[C@@H]1(C(=O)O[C@@H](CC(=O)NCc2scc(n2)C2=N[C@](C(=O)N1)(CS2)C)/C=C/CCSSC[C@@H](C(=O)OC(C)(C)C)NC(=O)OC(C)(C)C)C(C)C	0.5	1
A005	O=C(Nc1ccc(c2nc(N3CCOCC3)c3c(n2)n(nc3)CCCCCCC(=O)NO)cc1)N1C[C@H](OCC1)C	0.58	1
A006	O=C(Nc1ccc(c2nc(N3CCOCC3)c3c(n2)n(nc3)CCCCCCC(=O)NO)cc1)N1C[C@H](OCC1)C	0.58	1
A007	n1(c(nc2c1nc(nc2N1CCOCC1)c1ccncc1)CN(c1ncc(cn1)C(=O)NO)C)C	0.86	1
A009	N1[C@@H](CCCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1cc(ccc1)C	1	1
A010	n1c(nc2c(c1N1CCOCC1)cnn2CCCCCCC(=O)NO)c1ccc(cc1)N	1.4	1
A011	n1c(nc2c(c1N1CCOCC1)cnn2CCCCCCC(=O)NO)c1ccc(cc1)N	1.4	1
A013	ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C	1.6	1
A014	N1([C@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1cc(ccc1)C)C	2	1
A015	n1c(sc(c1C(=O)N[C@H](c1ccc(cc1)C(=O)NO)C)C1CC1)c1ncc(s1)c1ccccc1	2	1
A017	N(C(=O)CCCCCC(=O)NO)c1nc(cs1)c1cccc(c1)NC(=O)OCC	2.1	1
A018	O=C(NO)/C=C/c1ccc(CNCCc2c3c([nH]c2C)cccc3)cc1	2.1	1
A019	n1(c(nc2c1nc(nc2N1CCOCC1)c1ccc(cc1)N)CN(c1ncc(cn1)C(=O)NO)C)C	2.8	1
A021	n1c(nc2c(c1N1CCOCC1)sc(c2)CN(c1ncc(cn1)C(=O)NO)C)c1cnc(cc1)OC	2.8	1
A022	n1c(nc2c(c1N1CCOCC1)sc(c2)CN(c1ncc(cn1)C(=O)NO)C)c1cnc(cc1)OC	2.8	1
A023	n1c(nc2c(c1N1CCOCC1)sc(c2)CN(c1ncc(cn1)C(=O)NO)C)c1cnc(cc1)OC	2.8	1
A025	ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C	3.5	1
A026	C1C(=O)NCc2scc(n2)C2=N[C@@](CS2)(C(=O)N[C@H](C(=O)O[C@@H]1/C=C/CCSCCS)C(C)C)C	3.6	1
A027	c1c(cccc1N)c1csc(n1)NC(=O)CCCCCC(=O)NO	3.7	1
A029	SCC/C=C/[C@@H]1CC(=O)N[C@@H](C(=O)N[C@H](C(=O)N/C(=C\C)/C(=O)N[C@H](C(=O)O1)C(C)C)CS)C(C)C	3.9	1
A030	C(=O)(NCCOc1ccc(cc1)C(=O)NO)c1c(c2c(cccc2)o1)CN(C)C	4	1

A031	**C(c1ccc(cc1)C(=O)NO)n1c2ccccc2c(c1)CN(C)C**	**4**	**1**
A033	**N1[C@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1cc(ccc1)C**	**4**	**1**
A034	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CN1CCCC1**	**4**	**1**
A035	**N1[C@@H](CCCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccc(cc1)C**	**4**	**1**
A037	**c1cc2c(cc1S(=O)(=O)N1CCN(CC1)c1ncc(cn1)C(=O)NO)cccc2**	**4.3**	**1**
A038	**O=C(NO)/C=C/c1ccc(CNCCc2c3c([nH]c2C)cccc3)cc1**	**4.5**	**1**
A039	**N1C(=O)[C@H]([C@@H](C1)c1ccccc1)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1**	**5**	**1**
A041	**c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C**	**5**	**1**
A042	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)C1N(CCC1)C**	**5**	**1**
A043	**c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C**	**5**	**1**
A045	**C(c1ccc(cc1)C(=O)NO)n1c2ccccc2c(c1)CN(C)C**	**5**	**1**
A046	**c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C**	**5**	**1**
A047	**N(C(=O)CCCCCC(=O)NO)c1nc(cs1)c1cccc(c1)NC(=O)OC(C)(C)C**	**5.7**	**1**
A049	**c1cccc(c1)NC(=O)[C@H](CCCCCS)NC(=O)[C@H]1N(C(=O)CCC1)C**	**6**	**1**
A050	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)n1nnc(c1)c1ccccc1**	**6.8**	**1**
A051	**n1(c(nc2c1nc(nc2N1CCOCC1)c1ccc(cc1)OC)CN(c1ncc(cn1)C(=O)NO)C)C**	**6.9**	**1**
A053	**N1C(=O)[C@@H]([C@H](C1)c1ccccc1)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1**	**7**	**1**
A054	**N1[C@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1cc(ccc1)C(F)(F)F**	**7**	**1**
A055	**n12c(c3c(C2)c(c2c(n3)cccc2)/C=N/OCCCCCC(=O)NCCSSCCNC(=O)/C(=N\O)/Cc2ccc(c(c2)Br)O)cc2c(c1=O)COC(=O)[C@]2(O)CC**	**7.6**	**1**
A057	**C(=O)(NCCOc1ccc(cc1)C(=O)NO)c1c(c2c(cccc2)o1)CN(C)C**	**7.9**	**1**
A058	**c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C**	**8**	**1**
A059	**c1c(cc2c(c1)c(cn2Cc1ccc(cc1)C(=O)NO)C(=O)c1cc(c(c(c1)OC)OC)OC)OC**	**8.2**	**1**
A061	**N1[C@@H](CC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1**	**9**	**1**
A062	**c1cccc(c1)NC(=O)[C@H](CCCCCS)NC(=O)[C@H]1C(=O)NCCC1**	**9**	**1**
A063	**n12c(c3c(C2)c(c2c(n3)ccc(c2)O)/C=N/OCCCCCC(=O)NCCSSCCNC(=O)/C(=N\O)/Cc2ccc(c(c2)Br)O)cc2c(c1=O)COC(=O)[C@]2(O)CC**	**9.1**	**1**
A065	**c1cc2n(cc(c2cc1)CNCC1CCN(CC1)c1ncc(cn1)C(=O)NO)C**	**10**	**1**
A066	**O=C(NO)/C=C/c1ccc(CNCCc2c3c([nH]c2C)cccc3)cc1**	**10**	**1**
A067	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)c1nnn(c1)c1ccc(cc1)I**	**10.7**	**1**
A069	**c1cc(cc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1)C(F)(F)F**	**11**	**1**

A070	**c1cc(cc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1)C(F)(F)F**	**11**	**1**
A071	**N1C(=O)/C(=C/C)/NC(=O)[C@H](NC(=O)[C@H](CC(=O)C[C@H](OC(=O)[C@@H]1C(C)C)/C=C/CCS)C(C)C)CS**	**11**	**1**
A073	**c1cccc(c1)NC(=O)[C@H](CCCCCS)NC(=O)[C@@H]1C(=O)NCCC1**	**11**	**1**
A074	**N1[C@@H](CCC1=O)C(=O)N[C@H](CCCCCS)C(=O)Nc1ccccc1**	**11**	**1**
A075	**c1cccc(c1)NC(=O)[C@H](CCCCCS)NC(=O)[C@@H]1NC(=O)CC1**	**11**	**1**
A077	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**11.6**	**1**
A078	**O=C(NO)/C=C/c1ccc(CNCCc2c3c([nH]c2C)cccc3)cc1**	**12.7**	**1**
A079	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**13**	**1**
A081	**c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)CC(O)c1cccc(c1)Br**	**13**	**1**
A082	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**14**	**1**
A083	**N1C(=O)[C@H]([C@H](C1)c1ccccc1)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1**	**14**	**1**
A085	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)c1nnn(c1)c1ccc(cc1)C(F)(F)F**	**14.5**	**1**
A086	**N1[C@@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1**	**15**	**1**
A087	**c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1CC(c1cccc(c1)Br)O**	**15.8**	**1**
A089	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)n1nnc(c1)C1CCCCC1**	**16.6**	**1**
A090	**c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1cccc(c1)N**	**17.5**	**1**
A091	**c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1Cc1ccc(cc1)F**	**17.6**	**1**
A093	**c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)Cc1ccccc1**	**18.5**	**1**
A094	**C(n1ncc(c1)c1c2c(ncn1)ncc2)CCCCCC(=O)NO**	**19**	**1**
A095	**c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1Cc1ccccc1**	**19.2**	**1**
A097	**c1(ccc(cc1)C(=O)NO)Cn1nc(c2c1cccc2)CN(C)C**	**20**	**1**
A098	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**20**	**1**
A099	**c1ccc2c(c1)n(cc2CCN(C)C)Cc1ccc(cc1)C(=O)NO**	**20**	**1**
A101	**CC(COC[C@@H]1[C@@H]2[C@H]([C@@H]([C@H](O1)O[C@@H]1[C@H](O[C@@H]([C@H]([C@@H]1O)O)O[C@@H]1[C@H](O[C@@H]([C@H]([C@@H]1O)O)O[C@@H]1[C@H](O[C@@H]([C@H]([C@@H]1O)O)O[C@@H]1[C@H](O[C@@H]([C@H]([C@@H]1O)O)O[C@@H]1[C@H](O[C@@H]([C@H]([C@@H]1O)O)O[C@@H]1[C@H](O[C@H](O2)[C@H]([C@@H]1O)O)COCC(C)O)COCC(C)O)COCC(C)O)COCC(C)O)COCC(C)O)COCC(C)O)O)O)O**	**20.3**	**1**
A102	**C1C(=O)NCc2scc(n2)C2=N[C@@](CS2)(C(=O)N[C@H](C(=O)O[C@@H]1/C=C/CCC(=O)NO)C(C)C)C**	**21**	**1**
A103	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**22.8**	**1**
A105	**c1ccc2c(c1)n(cc2CCN(C)C)Cc1ccc(cc1)C(=O)NO**	**25**	**1**

A106	**C(=O)(CCCCc1nn(cc1)Cc1ccc(cc1)Nc1ccccc1)NO**	**25**	**1**
A107	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)c1n(cnc1)C**	**25**	**1**
A109	**c1c(cccc1NC(=O)CCCCCCCC(=O)NO)c1cn(nn1)c1ccccc1**	**27.7**	**1**
A110	**c1c(cccc1NC(=O)CCCCCCCC(=O)NO)c1cn(nn1)c1ccccc1**	**27.7**	**1**
A111	**n1(cc(c(=O)c2c1ccc(c2)OC)C(=O)Nc1nc2c(s1)cccc2)CCCCCC(=O)NO**	**28**	**1**
A113	**c1c(ccc(c1)NC(=O)CCCCCCC(=O)NO)c1ccccc1NC(=O)[C@@H](N)Cc1c[nH]c2c1cccc2**	**28.3**	**1**
A114	**C(=O)(CCCCc1nn(cc1)Cc1ccc(cc1)c1ccccc1)NO**	**29**	**1**
A115	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**29.1**	**1**
A117	**c1cc(cc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1)C**	**30**	**1**
A118	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**30**	**1**
A119	**c1cc(cc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1)C**	**30**	**1**
A121	**C(=O)(CCCCCC(C)n1ncc(c1)c1c2c(ncn1)ncc2)NO**	**31**	**1**
A122	**c1cccc(c1)NC(=O)[C@H](CCCCCS)NC(=O)[C@@H]1NC(=O)CCC1**	**32**	**1**
A123	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)c1nnn(c1)c1ccc(cc1)F**	**32.8**	**1**
A125	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**35**	**1**
A126	**n1c(c(cc1c1ccc(cc1)O)c1cocc1)C(=O)NCc1ccc(cc1)C(=O)NO**	**35**	**1**
A127	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)c1ccccc1**	**36.1**	**1**
A129	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**38**	**1**
A130	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**39.6**	**1**
A131	**C(=O)(CCCCSc1nc(cc(=O)[nH]1)c1ccc(cc1)c1ccccc1)NO**	**40**	**1**
A133	**C(=O)([C@H](/C=C(/C=C/C(=O)NO)\C)C)c1ccc(N(C)C)cc1**	**40**	**1**
A134	**c1cc2n(cc(c2cc1)CNCC1CCN(CC1)c1ncc(cn1)C(=O)NO)C**	**40**	**1**
A135	**C(/C(=C/c1ccc(cc1)/C=C/C(=O)NO)/c1ccc(cc1)F)NC1CC1**	**42**	**1**
A137	**C(N1CCN(CC1)c1ncc(cn1)C(=O)NO)(/C=C/c1ccc(cc1)F)CO**	**43**	**1**
A138	**c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1ccc(cc1)N**	**44.9**	**1**
A139	**c1c(ccc(c1)NC(=O)CCCCCCC(=O)NO)c1ccccc1NC(=O)[C@@H](N)Cc1ccccc1**	**46.1**	**1**
A141	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)c1n(cnc1)C**	**50**	**1**
A142	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)CCC1**	**51**	**1**
A143	**n1cnc2c(c1N)c(nn2c1cccc(c1)c1cnnn1CCCCCC(=O)NO)c1ccc(cc1)Cl**	**51**	**1**
A145	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1c1ccccc1**	**55.7**	**1**
A146	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**58**	**1**
A147	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**61**	**1**
A149	**c1(cc2cc(c1)/C=C\CO[C@H](C(=O)Nc1c(OC2)cccc1)CCCCCC(=O)NO)OC**	**63**	**1**

A150	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**64**	**1**
A151	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1Cc1ccc(cc1)F**	**64.7**	**1**
A153	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1Cc1ccccc1**	**70.1**	**1**
A154	**c1(cc2cc(c1)/C=C\COC(C(=O)Nc1c(OC2)cccc1)CCCCCC(=O)NO)OC**	**71**	**1**
A155	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**71**	**1**
A157	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**72**	**1**
A158	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**72**	**1**
A159	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**72**	**1**
A161	**c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1ccc(cc1)NC(=O)C**	**73**	**1**
A162	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1CC(O)c1cccc(c1)Br**	**74.8**	**1**
A163	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**77**	**1**
A165	**OCCN(C(=O)Cc1ccc(C(=O)NO)cc1)c1ccccc1**	**79**	**1**
A166	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CCO**	**79**	**1**
A167	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**80**	**1**
A169	**c1cc(cc(c1)CCNC(=O)[C@H](CCCCCS)NC(=O)[C@H]1NC(=O)CCC1)C**	**81**	**1**
A170	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)CC(O)c1cccc(c1)Br**	**86.5**	**1**
A171	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)C1C(=O)NCC1c1ccccc1**	**88**	**1**
A173	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**92**	**1**
A174	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1**	**93**	**1**
A175	**C(=O)(CCCCc1nn(cc1)Cc1cc(ccc1)c1ccccc1)NO**	**94**	**1**
A177	**C(=O)(c1ccc(cc1)CN(C(=O)Nc1ccccc1)CCCC)NO**	**100**	**1**
A178	**C(=O)(NO)c1cnc(N(Cc2onc(n2)C2CCN(CC2)Cc2ccc(cc2)C)C)nc1**	**102**	**1**
A179	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)Cc1ccc(cc1)F**	**112**	**1**
A181	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)C1**	**113**	**1**
A182	**C1(=O)N(C(C(=O)N1c1ccc(cc1)Cl)CCC(=O)Nc1ccc(cc1)C(=O)NO)Cc1ccccc1**	**114**	**1**
A183	**N1([C@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1)C**	**116**	**1**
A185	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)C1**	**125**	**1**
A186	**c1(c(cc2c(c1)c(ncn2)Nc1cc(ccc1)C#C)OC)OCCCCCCCC(=O)NO**	**126**	**1**
A187	**c12ccc(nc1cccc2)N(CCCCCCC(=O)NO)c1ccccn1**	**126**	**1**
A189	**c1c(cnc(n1)N(c1ccccc1)c1ccccc1)C(=O)NCCCCCCCC(=O)NO**	**126**	**1**
A190	**C(=O)(CCCCc1nn(cc1)Cc1ccc(cc1)n1ccc2ccccc12)NO**	**136**	**1**

A191	**SCC/C=C/[C@@H]1CC(=O)N[C@@H](c2scc(C(=O)N/C(=C\C)/C(=O)N[C@H](C(=O)O1)C(C)C)n2)C(C)C**	**138**	**1**
A193	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)C1C(=O)NCC1**	**147**	**1**
A194	**[C@@H]1(Cc2cccc3c2cccc3)NC(=O)[C@@H](NC(=O)[C@H](C)NC(=O)C[C@H](CC(C)C)NC1=O)CCCCCC(=O)CC**	**150**	**1**
A195	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**150**	**1**
A197	**c1(onc(c1)COCCCCC(=O)NO)c1ccccc1**	**155**	**1**
A198	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CNC(=O)OC(C)(C)C**	**158**	**1**
A199	**c1ccc2c(c1)n(c1c2CCN(C1)C)Cc1ccc(cc1)C(=O)NO**	**158**	**1**
A201	**c1c(ccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1)C**	**163**	**1**
A202	**C1(=O)N(C(C(=O)N1c1ccc(cc1)Cl)CCC(=O)Nc1ccc(cc1)C(=O)NO)Cc1ccc(cc1)C**	**164**	**1**
A203	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**170**	**1**
A205	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**170**	**1**
A206	**C1C2CC3CC1CC(C2)(C3)NC(=O)NCCCCCC(=O)NO**	**173**	**1**
A207	**c1c(ccc(c1)NC(=O)NCCCCCCC(=O)NO)N(C)C**	**176**	**1**
A209	**C(=O)(NO)CCCCCC(C(=O)Nc1ccccc1)OCc1ccc(cc1)OC**	**179**	**1**
A210	**c1cc(ccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)Cc1ccccc1**	**188**	**1**
A211	**N1[C@@H](CCCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)NC1CCCC1**	**189**	**1**
A213	**c1c(cnc(n1)N(c1ccccc1)c1ccccc1)C(=O)NCCCCCCCC(=O)NO**	**194**	**1**
A214	**c1ccc2c(c1)n(c1c2CCN(C1)C)Cc1ccc(cc1)C(=O)NO**	**200**	**1**
A215	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CCOC**	**200**	**1**
A217	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CCO**	**200**	**1**
A218	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**200**	**1**
A219	**c1(cc2cc(c1)CCCO[C@@H](C(=O)Nc1c(OC2)cccc1)CCCCCC(=O)NO)OC**	**200**	**1**
A221	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**210**	**1**
A222	**N(C(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@@H]1NC(=O)CC1)CCc1ccccc1**	**210**	**1**
A223	**N(C(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)CCC1)CCc1ccccc1**	**227**	**1**
A225	**C(=O)(CCCCCC(c1c[nH]c2c1cccc2OC)c1c[nH]c2c1cccc2OC)NO**	**240**	**1**
A226	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)C1C(=O)NCCC1**	**244**	**1**
A227	**C(=O)(CCCCCC(c1c[nH]c2c1cc(cc2)[N+](=O)[O-])c1c[nH]c2c1cc(cc2)[N+](=O)[O-])NO**	**250**	**1**
A229	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CNC(=O)OC(C)(C)C**	**251**	**1**
A230	**C(=O)(Nc1ccccc1)CCCCCCCS**	**252**	**1**
A231	**c1(nc(cs1)c1ccccc1)NC(=O)CCCCc1onc(c1)C(=O)NO**	**254**	**1**

A233	**c1cccc(c1)NC(=O)CCCCCCS**	**270**	**1**
A234	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**278**	**1**
A235	**c1(n(c(cc1)/C=C/C(=O)NO)C)C(=O)c1ccccc1**	**280**	**1**
A237	**c1c(c(nc1c1ccc(cc1)OC)C(=O)NCc1ccc(cc1)C(=O)NO)c1ccoc1**	**294**	**1**
A238	**ONC(=O)CCCCCn1c(=O)c2cccc3cccc(c1=O)c23**	**300**	**1**
A239	**N1[C@@H](CCCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)NCc1ccc(cc1)C(F)(F)F**	**310**	**1**
A241	**c1ccc2c(c1)n(c(c2C)C)Cc1ccc(cc1)C(=O)NO**	**316**	**1**
A242	**c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1ccc(cc1)NC(=O)OC(C)(C)C**	**316**	**1**
A243	**c1(ccc(cc1)C(=O)NO)Cn1c2c(CSCC2)c2c1cccc2**	**316**	**1**
A245	**c1(ccc2c3c1cccc3c(=O)n(c2=O)CCCCCC(=O)NO)NCCC**	**320**	**0**
A246	**c1(cc2cc(c1)CCCOC(C(=O)Nc1c(OC2)cccc1)CCCCCC(=O)NO)OC**	**320**	**0**
A247	**c1ccc(cc1C(F)(F)F)NC(=O)CCCCCCC(=O)NO**	**374**	**0**
A249	**c1(ccc(cc1)C(=O)NO)Cn1c2c(CSCC2)c2c1cccc2**	**398**	**0**
A250	**c1(c(cc2c(c1)c(ncn2)Nc1cc(ccc1)C#C)OC)OCCCCCCCC(=O)NO**	**398**	**0**
A251	**c1(onc(c1)C(=O)NCCCCC(=O)NO)c1ccc(cc1)N**	**401**	**0**
A253	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**430**	**0**
A254	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**430**	**0**
A255	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**432**	**0**
A257	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**456**	**0**
A258	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**456**	**0**
A259	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**456**	**0**
A261	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**456**	**0**
A262	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**459**	**0**
A263	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**460**	**0**
A265	**c1(ccc(cc1)C(=O)NO)Cn1c2c(CN(CC2)C(=O)C)c2c1cccc2**	**501**	**0**
A266	**C(=O)(c1ccc(cc1)CN(C(=O)Nc1ccccc1)CCCC)NO**	**501**	**0**
A267	**c1cc(cc(c1)NC(=O)CCCCCCC(=O)NO)C**	**503**	**0**
A269	**c1ccc2c(c1)n(c(c2C)C)Cc1ccc(cc1)C(=O)NO**	**631**	**0**
A270	**c1(ccc(cc1)C(=O)NO)Cn1c2c(COCC2)c2c1cccc2**	**631**	**0**
A271	**c1c(ccc(c1)NC(=O)CCCCCCC(=O)NO)C**	**649**	**0**
A273	**c1(ccc2c3c1cccc3c(=O)n(c2=O)CCCCCC(=O)NO)NCc1ccccc1**	**660**	**0**
A274	**c1cccc2c1OCCCCCCOC(C(=O)N2)CCCCCC(=O)NO**	**683**	**0**
A275	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**686**	**0**
A277	**c1(onc(c1)C(=O)NCCCCC(=O)NO)c1ccc(cc1)NC(=O)OC(C)(C)C**	**690**	**0**
A278	**c1ccc(cc1)CC(=O)N(CCO)Cc1ccc(cc1)C(=O)NO**	**700**	**0**

A279	**c1c(c(nc1c1ccc(cc1)OC)C(=O)NCc1ccc(cc1)C(=O)NO)c1ccsc1**	**723**	**0**
A281	**C(=O)(CCCCc1nn(cc1)c1ccccc1)NO**	**774**	**0**
A282	**N(C(=O)c1cc(cc(c1)N=[N+]=[N-])CN=[N+]=[N-])c1ccc(c2onc(C(=O)NCCCCCCC(=O)NO)c2)cc1**	**776**	**0**
A283	**c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1ccc(cc1)NC(=O)OC(C)(C)C**	**794**	**0**
A285	**c1(onc(c1)C(=O)NCCCCC(=O)NO)c1ccccc1**	**852**	**0**
A286	**C(=O)(CCCCCC(c1c[nH]c2c1cccc2)c1c[nH]c2c1cccc2)NO**	**870**	**0**
A287	**C(=O)(c1c(cc(c(c1)C(C)C)O)O)N1CCc2onc(c2C1)C(=O)NCCCCCCC(=O)NO**	**886**	**0**
A289	**c1cc(ccc1NC(=O)c1ccc(cc1)C(C)(C)C)C(=O)NO**	**899**	**0**
A290	**n1c([nH]c(=O)c2c1sc1c2CCN(C1)C)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C**	**923**	**0**
A291	**c1(ccc(cc1N)F)NC(=O)c1ccc(cc1)CNC(=O)/C=C/1\NC(=O)c2c1cccc2**	**950**	**0**
A293	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**982**	**0**
A294	**c1(ccc(cc1)C(=O)NO)Cn1c2c(CCCC2)c2c1cccc2**	**1,000**	**0**
A295	**c1(ccc(cc1)C(=O)NO)Cn1c2c(COCC2)c2c1cccc2**	**1,000**	**0**
A297	**c1ccc2c(c1)c(nc(n2)C)N(c1cc(c(cc1)OC)OCCCC(=O)NO)C**	**1,176**	**0**
A298	**C(=O)(Nc1ccccc1)CCCCCS**	**1,220**	**0**
A299	**C(=O)(c1ccc(cc1)CNc1c2cccnc2ccc1)NO**	**1,256**	**0**
A301	**C(NC(=O)CCCCCCC(=O)NO)Cc1ccccc1**	**1,350**	**0**
A302	**n1(cccc(c1=O)/C=C/C(=O)NO)CCc1cc2c(cc1)cccc2**	**1,463**	**0**
A303	**c1ccc2c(c1)C(=O)c1c(C2=O)ccc(c1)/C=C/C(=O)NO**	**1,470**	**0**
A305	**c1(ccc(cc1)C(=O)NO)Cn1c2c(cccc2)c2c1[C@@H](N1[C@H](C2)C(=O)N(CC1=O)CCC)C**	**1,510**	**0**
A306	**c1cccc2c1OCCCCCOC(C(=O)N2)CCCCCC(=O)NO**	**1,530**	**0**
A307	**C(=O)(c1ccc(N(C)C)cc1)NCCCCCNC(=O)CS**	**1,540**	**0**
A309	**C(=O)(/C=C/c1cc2nc(n(c2cc1)CCN(CC)CC)CCCC)NO**	**1,585**	**0**
A310	**c1c(cnc(n1)N(c1ccccc1)c1ccccc1)C(=O)NCCCCCCC(=O)NO**	**1,585**	**0**
A311	**c1([C@@H]2CN(C[C@H]2C(=O)Nc2ccc(cc2)Cl)C)ccc(/C=C/C(=O)Nc2c(cccc2)N)cc1**	**1,600**	**0**
A313	**C(=O)(/C=C/c1ccc(cn1)NC(=O)C(Cc1ccccc1)c1ccccc1)NO**	**1,800**	**0**
A314	**C(=O)(c1ccc(cc1)CCn1c(=O)c2c3c(c1=O)ccc(c3ccc2)N1CCOCC1)NO**	**1,870**	**0**
A315	**S(CC/C=C/[C@@H]1CC(=O)NCc2scc(C(=O)N/C(=C\C)/C(=O)N[C@H](C(=O)O1)C(C)C)n2)C(=O)CCCCCCC**	**1,930**	**0**
A317	**c1(ccc(cc1)C(=O)NO)Cn1c2c(CN(CC2)C(=O)C)c2c1cccc2**	**1,995**	**0**
A318	**c1(ncnc2c1nc[nH]2)N[C@H](c1n(c(=O)c2c(n1)cccc2CCCCC(=O)NO)c1ccccc1)CC**	**2,073**	**0**
A319	**CCCCCCCC(=O)SCC/C=C/[C@@H]1CC(=O)NCc2scc(n2)C2=N[C@@](CS2)(C(=O)N[C@H](C(=O)O1)C(C)C)C**	**2,210**	**0**

A321	**c1c(c(cc(c1O)C(=O)N(c1ccc(cc1)NC(=O)CCCCCCC(=O)NO)C)C(C)C)O**	**2,670**	**0**
A322	**n1c(nc2c(c1NC1CCN(CC1)Cc1ccc(cc1)OC)cc(c(c2)OC)OC)Nc1ccc(cc1)NC(=O)CCCCCCC(=O)NO**	**2,690**	**0**
A323	**C(=O)(Nc1c(cccc1)N)CCCCCCC(=O)Nc1scc(n1)c1cccc(c1)N**	**2,910**	**0**
A325	**C(=O)(CCCCCCS)N1Cc2c(CC1)cccc2**	**2,930**	**0**
A326	**C(=O)(CCCCCCS)NCc1cc(ccc1)OCc1ccccc1**	**2,990**	**0**
A327	**OCCN(C(=O)Cc1ccc(C(=O)NO)cc1)c1ccccc1**	**3,000**	**0**
A329	**N(C(=O)CCCCCC(=O)Nc1ccccc1N)c1nc(cs1)c1cccc(c1)NC(=O)C**	**3,000**	**0**
A330	**C(=O)(/C=C/c1cc2nc(n(c2cc1)CCN(CC)CC)CCCC)NO**	**3,162**	**0**
A331	**c1c(O)c(cc(c1O)n1nnc(C(=O)NCCCCCCC(=O)NO)c1c1ccc(cc1)CN1CCOCC1)C(C)C**	**3,220**	**0**
A333	**c1([C@H]2CN(C[C@@H]2C(=O)Nc2ccc(cc2)Cl)C)ccc(/C=C/C(=O)Nc2c(cccc2)N)cc1**	**3,400**	**0**
A334	**C(=O)(c1ccc(N(C)C)cc1)NCCCCCCNC(=O)CS**	**3,430**	**0**
A335	**C(=O)(c1ccc(cc1)CNC(=O)c1cccn1C)NO**	**3,700**	**0**
A337	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**3,710**	**0**
A338	**C(=O)(CCCCCCS)NCCc1cc(ccc1)C**	**3,930**	**0**
A339	**c1(ccc(cc1)Cn1c2c(cccc2)c2c1cccc2)C(=O)NO**	**3,981**	**0**
A341	**ONC(=O)c1ccc(cc1)Cn1sc2ncccc2c1=O**	**4,260**	**0**
A342	**C12(C(=O)NC1)CCN(c1ccc(C(=O)Nc3cc(ccc3N)c3cccs3)cn1)CC2**	**4,400**	**0**
A343	**C(c1c[nH]c2c1cccc2)(c1c[nH]c2c1cccc2)c1ccc(cc1)/C=C/C(=O)NO**	**4,700**	**0**
A345	**c1ccc2c(c1)C(=O)c1c(C2=O)ccc(c1)C(=O)NO**	**4,997**	**0**
A346	**N(C(=O)CCCC(=O)NCCc1ccccc1)O**	**5,000**	**0**
A347	**CN1C[C@H](c2c(cc(/C=C/C(=O)Nc3c(cccc3)N)cc2)F)[C@H](C1)C(=O)Nc1ccc(cc1)Cl**	**5,000**	**0**
A349	**c1(ccc(cc1)C(=O)NO)Cn1c(cc2c1cccc2)C(C)(C)C**	**5,012**	**0**
A350	**[C@H](CCCCCS)(C(=O)NC1CCCC1)NC(=O)OC(C)(C)C**	**5,210**	**0**
A351	**O=S(=O)(n1c2ccc(C(=O)c3ccc(cc3)F)cc2cc1)c1cccc(c1)/C=C/C(=O)NO**	**5,240**	**0**
A353	**c1c(ccc(c1)/C=C/C(=O)NO)/C=N/OCc1ccc(cc1)C(=O)OC**	**5,430**	**0**
A354	**O=S(=O)(n1c2ccc(C(=O)c3ccc(cc3)OC)cc2cc1)c1cccc(c1)/C=C/C(=O)NO**	**5,890**	**0**
A355	**O=S1(=O)Cc2c(CC1)n(c1c2cc(cc1)F)Cc1ccc(cc1)C(=O)NO**	**5,900**	**0**
A357	**c1(cc(on1)CN1CCOCC1)c1ccc(cc1)/C=C/C(=O)NO**	**6,100**	**0**
A358	**C(=O)(CCCCCCS)NCc1ccc(cc1)C(F)(F)F**	**6,290**	**0**
A359	**C(=O)(/C=C/c1ccc(cc1)C(F)(F)F)NO**	**6,320**	**0**
A361	**N(C(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)CCC1)C1CCCC1**	**6,830**	**0**

A362	**C(n1c2ccc(C(=O)c3ccc(cc3)OC)cc2cc1)c1ccc(cc1)/C=C/C(=O)NO**	**7,050**	**0**
A363	**c1c(ccc(c1)/C=C/C(=O)NO)/C=N/OCc1c(c(c(c(c1F)F)F)F)F**	**7,360**	**0**
A365	**O=S1(=O)Cc2c(CC1)n(c1c2cccc1)Cc1ccc(cc1)C(=O)NO**	**7,700**	**0**
A366	**O=S(=O)(n1c2ccc(C(=O)c3cc(ccc3)OC)cc2cc1)c1cccc(c1)/C=C/C(=O)NO**	**7,740**	**0**
A367	**c1c(ccc(c1)/C=C/C(=O)NO)/C=N/OCCN1CCOCC1**	**7,840**	**0**
A369	**c1c(c(nc1c1ccc(cc1)O)C(=O)NCc1ccc(cc1)C(=O)Nc1c(cccc1)N)c1ccoc1**	**8,547**	**0**
A370	**c1c(c(ccc1)c1nc2c(n1)cc(cc2)C(=O)NO)C(F)(F)F**	**8,600**	**0**
A371	**C(=O)(c1ccc(N(C)C)cc1)NCCCNC(=O)CS**	**8,780**	**0**
A373	**O=S(=O)(n1c2ccc(C(=O)c3ccc(cc3)C)cc2cc1)c1cccc(c1)/C=C/C(=O)NO**	**9,100**	**0**
A374	**C1Cc2c(N(Cc3ccc(cc3)C(=O)NO)[C@H]3[C@H](C1)CCCC3)cc(cc2)C(F)(F)F**	**9,400**	**0**
A375	**C(=O)(c1ccc(cc1)CNc1cccc2cccnc12)NO**	**9,791**	**0**
A377	**O=S1(=O)C[C@@H]2[C@H](N(c3c1cccc3)Cc1ccc(cc1)C(=O)NO)CCCC2**	**10,000**	**0**
A378	**C1Cc2c(N(Cc3ccc(cc3)C(=O)NO)[C@H]3[C@H](C1)CCCC3)cccc2**	**10,000**	**0**
A379	**S(=O)(=O)(N1c2c(cccc2CC1)NCc1ccc(cc1)C(=O)Nc1ccccc1N)c1ccc(cc1)OC**	**10,000**	**0**
A381	**O=C(NO)c1ccc(Cn2c3CN4C(=O)N(C)C(=O)C(C4)c3c3c2cccc3)cc1**	**10,500**	**0**
A382	**c1(NC(=O)CCCCCNC(=O)CS)c2ncccc2ccc1**	**10,700**	**0**
A383	**c1(ccc(cc1)/C=C/C(=O)NO)c1ccc(cc1)O**	**11,000**	**0**
A385	**c1(c(cc2c(c1)c(ccn2)Oc1ccc(cc1F)NC(=O)C1(C(=O)Nc2ccc(cc2)F)CC1)OCCCCCC(=O)NO)OC**	**11,900**	**0**
A386	**c1c(cc2c(c1)ccn2Cc1ccc(cc1)OC)C(=O)NO**	**13,000**	**0**
A387	**c1c(cc2c(c1)ccn2Cc1ccc(cc1)OC)C(=O)NO**	**13,000**	**0**
A389	**O=S1(=O)c2c3c(n(Cc4ccc(cc4)C(=O)NO)c2CC1)cccc3**	**13,000**	**0**
A390	**c1(ccc(cc1)c1onc(c1)C(=O)NO)NC(=O)CNC(=O)c1cc(on1)c1ccc(cc1)NC(=O)OC(C)(C)C**	**13,300**	**0**
A391	**N1(Cc2ccc(cc2)C(=O)NO)c2c(CC[C@H]3[C@H]1CCCC3)cccc2**	**14,000**	**0**
A393	**c1cc(ccc1C(=O)NNCCCC)Br**	**15,300**	**0**
A394	**n1c(cc(o1)c1cccc(c1)N)C(=O)NCCCCCCC(C(=O)NC)N=O**	**15,400**	**0**
A395	**c1c(ccc(c1)CN(c1ncc(cc1)C(=O)Nc1c(cccc1)N)CCN(C)C)OC**	**15,500**	**0**
A397	**c1c(c(cc(c1)C(=O)c1ccc2c(c1)ccn2Cc1ccc(cc1)C(=O)NO)OC)OC**	**18,900**	**0**
A398	**c1c(c(cc(c1)C(=O)NO)F)Cn1c(nc2c1cccc2)C**	**20,400**	**0**
A399	**S1Cc2c(CC1)n(c1c2cc(cc1)F)Cc1ccc(cc1)C(=O)NO**	**21,000**	**0**
A401	**c1(cccc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)c1cnc(cc1)Cl**	**23,100**	**0**
A402	**C(=O)(Nc1ccccc1)CCCCC1CC(=NO1)C(=O)NO**	**23,900**	**0**

A403	**C(=O)(c1ccc(N(C)C)cc1)NCCCCNC(=O)CS**	**29,700**	**0**
A405	**c1(ncnc2c1nc[nH]2)N[C@H](c1nc(c2c(n1)cccc2)CCCCC(=O)NO)CC**	**33,730**	**0**
A406	**c1cccc(c1)NC(=O)[C@H](CCCCCC(C(=O)NC)N=O)NCc1ccc(cc1)OC**	**37,800**	**0**
A407	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**38,400**	**0**
A409	**c1c(cc2c(c1)ccn2Cc1ccc(cc1)OC)C(=O)NO**	**39,811**	**0**
A410	**C1N(Cc2n(C1)cc(c2)C(=O)NO)C(=O)c1cccn1C**	**41,000**	**0**
A411	**c1cc(cc(c1)C(=O)c1ccc2c(c1)ccn2Cc1ccc(cc1)C(=O)NO)OC**	**47,900**	**0**
A413	**c1c(ccc(c1)C(=O)c1ccc2c(c1)ccn2Cc1ccc(cc1)C(=O)NO)OC**	**59,800**	**0**
A414	**c1(cc(ccc1CC#N)c1c(cc(cc1)/C=C/C(=O)NO)Cl)C12CC3CC(C1)CC(C3)C2**	**64,000**	**0**
A415	**C(CCC(=O)Nc1ccccc1N)CCC(=O)Nc1ccc(cc1)C**	**66,200**	**0**
A417	**C(=O)(CCCCCC(c1c[nH]c2c1cccc2)c1c[nH]c2c1cccc2)Nc1ccccc1N**	**73,000**	**0**
A418	**c1cc(cc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)c1ccccc1**	**73,300**	**0**
A419	**c1c(cc2c(c1)cc(c(=O)o2)C(=O)/C=C/C=C/c1ccc(c(c1)OC)OC)OC**	**85,000**	**0**
A421	**c1cccc(c1)NC(=O)[C@H](CCCCCC(C(=O)NC)N=O)N(Cc1ccc(cc1)OC)C(=O)OC(C)(C)C**	**98,500**	**0**
A422	**c1c(cc2c(c1)ccn2Cc1ccc(cc1)OC)C(=O)NO**	**<10000**	**0**
A423	**c1(ccc(cc1)C(=O)NO)Cn1c(cc2c1cccc2)C(C)(C)C**	**<39811**	**0**
A425	**N(Cc1ccc(cc1)C(=O)Nc1ccccc1N)C(=O)OCc1cccnc1**	**>1000**	**0**
A426	**c1ccc(c(c1)CN1C[C@H](CC1)O)NC(=O)CCCCCCC(=O)NO**	**>10000**	**0**
A427	**O=C(NO)c1c2c(C(C)(C)N(C2)c2cnc(cn2)C(F)(F)F)ccc1**	**>10000**	**0**
A429	**c1(c(cc2c(c1)c(ccn2)Oc1ccc(cc1F)NC(=O)C1(C(=O)Nc2ccc(cc2)F)CC1)OCCCC(=O)NO)OC**	**>10000**	**0**
A430	**c1c(cc2c(c1)ccn2Cc1ccc(cc1)OC)C(=O)NO**	**>10000**	**0**
A431	**c1(cc(cc(c1)C(F)(F)F)C(=O)N(c1ccc(cc1)C(=O)NO)C(C)C)C(F)(F)F**	**>10000**	**0**
A433	**O=C(NO)c1c2c(C(C)(C)N(c3nc4cc(ccc4n3)C(F)(F)F)C2)ccc1**	**>10000**	**0**
A434	**c1(c(cc2c(c1)oc(n2)Nc1cc(ccc1)C(=O)NO)Cl)Cl**	**>10000**	**0**
A435	**C1(CCOCC1)(CNC(=O)c1cccc(c2nc(on2)C(F)(F)F)c1)c1nc(cs1)c1ccccc1**	**>10000**	**0**
A437	**n1c(cc(o1)c1cccc(c1)NC(=O)OC(C)(C)C)C(=O)NCCCCCCC(C(=O)NC)N=O**	**>100000**	**0**
A438	**n1c(cc(o1)c1ccc(cc1)NC(=O)OC(C)(C)C)C(=O)NCCCCCCC(C(=O)NC)N=O**	**>100000**	**0**
A439	**c1cc(cc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)c1cncc2c1ccc(c2)Cl**	**>100000**	**0**
A441	**c1cc(cc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)c1ccc(cc1)C(=O)O**	**>100000**	**0**
A442	**c1cc(cc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)Br**	**>100000**	**0**
A443	**c1c(ccc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)c1ccccc1**	**>100000**	**0**

A445	**C1(NC(=O)CCCCCCC(C(=O)NC)N=O)c2c(cccc2)c2c1cccc2**	**>100000**	**0**
A446	**n1cc(cc2c1cccc2)NC(=O)CCCCCCC(C(=O)NC)N=O**	**>100000**	**0**
A447	**c1cc(cc2c1cccc2)NC(=O)CCCCCCC(C(=O)NC)N=O**	**>100000**	**0**
A449	**N(Cc1ccc(cc1)C(=O)Nc1ccccc1N)C1=N[C@@H]([C@@H](O1)c1ccccc1)c1ccccc1**	**>100000**	**0**
A450	**c1(ccc2c(c1)C[C@@H](CC2)Nc1nccc(n1)c1cccnc1)C(=O)NO**	**>100000**	**0**
A451	**C(=O)(NO)/C=C/C=C/c1ccc(c2ccc(c(c2)C23CC4CC(C2)CC(C4)C3)OC)cc1**	**>100000**	**0**
A453	**c1(c(ccc(c1)c1ccc(cc1)/C=C/C(=O)NO)O)C12CC3CC(C1)CC(C3)C2**	**>100000**	**0**
A454	**O=S(=O)(n1c2ccc(C(=O)c3ccc(cc3)OC)cc2cc1)c1ccc(cc1)/C=C/C(=O)NO**	**>100000**	**0**
A455	**c1c(ccc(c1)C(=O)c1ccc2c(c1)ccn2Cc1ccc(cc1)C(=O)NO)Cl**	**>100000**	**0**
A457	**c1c(ccc(c1)C(=O)c1ccc2c(c1)ccn2Cc1ccc(cc1)C(=O)NO)C**	**>100000**	**0**
A458	**O=S(=O)(n1c2ccc(C(=O)c3cc(c(c(c3)OC)OC)OC)cc2cc1)c1cccc(c1)/C=C/C(=O)NO**	**>100000**	**0**
A459	**c1c(ccc(c1)C(=O)c1ccc2c(c1)ccn2Cc1ccc(cc1)C(=O)NO)F**	**>100000**	**0**
A461	**c1c(cc(c(c1)/C=C/C(=O)NO)c1ccc(cc1)c1ccccc1)OC**	**>20000**	**0**
A462	**c12cc(oc1cccc2)c1nc2c(cccc2)c(c1)C(=O)NC(c1ccccc1)c1ccccc1**	**>200000**	**0**
A463	**[O-][N+](=O)c1c(N)cc(N2CCN(CC2)C)cc1**	**>200000**	**0**
A465	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**>30000**	**0**
A466	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**>30000**	**0**
A467	**C1c2c(N(CC1)Cc1ccc(cc1)C(=O)NO)ccc(c2)Cl**	**>30000**	**0**
A469	**C1c2c(N(Cc3ccc(cc3)C(=O)NO)[C@H]3[C@H]1CCCCC3)cccc2**	**>30000**	**0**
A470	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)Nc1ccccc1N)c1cn(nn1)c1ccccc1**	**>30000**	**0**
A471	**C(=O)(Nc1c(cccc1)N)CCCCCCC(=O)Nc1ccc(cc1)c1nnn(c1)Cc1ccccc1**	**>30000**	**0**
A473	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**>30000**	**0**
A474	**c1cc(ccc1[C@@H]1C[C@H](CSc2nc(c(o2)c2ccccc2)c2ccccc2)O[C@@H](O1)c1ccc(cc1)NC(=O)CCCCCCC(=O)NO)CO**	**>30000**	**0**
A475	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**>30000**	**0**
A477	**c1(nc(cs1)c1ccccc1)NC(=O)CCCCc1onc(c1)CO**	**>30000**	**0**
A478	**c1(nc(cs1)c1ccccc1)NC(=O)CCCCc1onc(c1)C(=O)O**	**>30000**	**0**
A479	**c1(nc(cs1)c1ccccc1)NC(=O)CCCCc1onc(c1)C(=O)N**	**>30000**	**0**
A481	**c1ccc(c(c1)C(=O)N)Nc1cccc(c1)OCCc1ccccc1**	**>30000**	**0**
A482	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)Nc1ccccc1N)c1cn(nn1)c1ccccc1**	**>30000**	**0**
A484	**c1cc(cc2c1CN(CC2)C(=O)c1cccn1C)C(=O)NO**	**>50000**	**0**

Table S2. Test set compounds in SMILES format

Compound name	SMILES format	IC_{50} value (nM)	BINARY
A004	SCC/C=C/[C@@H]1CC(=O)NCc2scc(n2)C2=N[C@@](CS2)(C(=O)N[C@H](C(=O)O1)C(C)C)C	0.5	1
A008	N1C(=O)/C(=C/C)/NC(=O)[C@H](NC(=O)[C@H](CC(=O)C[C@H](OC(=O)[C@@H]1C(C)C)/C=C/CCS)C(C)C)CS	0.9	1
A012	n1(cc(c(=O)c2c1cccc2)C(=O)Nc1nc2c(s1)cccc2)CCCCCC(=O)NO	1.5	1
A016	c1cc2n(cc(c2cc1)CNCC1CCN(CC1)c1ncc(cn1)C(=O)NO)C	2	1
A020	n1(c(nc2c1nc(nc2N1CCOCC1)c1cnc(nc1)N)CN(c1ncc(cn1)C(=O)NO)C)C	2.8	1
A024	ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C	3.2	1
A028	ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C	3.8	1
A032	c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CN1CCCC1	4	1
A036	O=C(NO)/C=C/c1ccc(CNCCc2c3c([nH]c2C)cccc3)cc1	4	1
A040	N1[C@@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1cc(ccc1)C	5	1
A044	c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)C1N(CCC1)C	5	1
A048	c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)c1nnn(c1)c1cc(cc(c1)CO)CO	5.8	1
A052	N1[C@H](CCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)Nc1cc(ccc1)C(F)(F)F	7	1
A056	c1ccc(cc1)/C=C/c1scc(n1)CCCC(=O)NO	7.8	1
A060	c1cc(ccc1/C=C/C(=O)NO)CN(CCO)CCc1c[nH]c2c1cccc2	8.4	1
A064	c1c(cc(cc1)NC(=O)CCCCCCC(=O)NO)c1nnn(c1)C1CCCCC1	9.8	1
A068	c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C	11	1
A072	c1cccc(c1)NC(=O)[C@H](CCCCCS)NC(=O)[C@H]1NC(=O)CCC1	11	1
A076	c1c(cc(cc1)c1csc(n1)NC(=O)CCCCCCC(=O)NO)[N+](=O)[O-]	11.1	1
A080	c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO	13	1
A084	N1C(=O)[C@@H]([C@@H](C1)c1ccccc1)C(=O)N[C@@H](CCCCCS)C(=O)Nc1ccccc1	14	1
A088	c1(ccc(cc1)C(=O)NO)Cn1nc(c2c1cccc2)CN(C)C	16	1
A092	n1(c(nc2c1nc(nc2N1CCOCC1)c1cnc(cc1)OC)CN(c1ncc(cn1)C(=O)NO)C)C	18	1
A096	c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C	20	1
A100	c1cc(ccc1C(=O)C(=CC(C)C=CC(=O)NO)C)N(C)C	20	1
A104	c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cnnn1c1ccccc1	24.5	1
A108	c1(c(cc2c(c1)c(ncn2)Nc1cc(ccc1)C#C)OC)OCCCCCCC(=O)NO	26	1
A112	c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1cccc(c1)NC(=O)OC(C)(C)C	28.2	1

A116	**C[C@H](/C=C/C(=O)NO)/C=C(\C)/C(=O)c1ccc(cc1)N(C)C**	**29.19**	**1**
A120	**ONC(=O)/C=C/C(=C/[C@H](C(=O)c1ccc(N(C)C)cc1)C)/C**	**31**	**1**
A124	**c1c(cccc1NC(=O)CCCCCCC(=O)NO)c1cn(nn1)Cc1ccc(cc1)F**	**34.3**	**1**
A128	**SCC/C=C/[C@@H]1CC(=O)NCc2scc(C(=O)N/C(=C\C)/C(=O)N[C@H](C(=O)O1)C(C)C)n2**	**37**	**1**
A132	**N1[C@H](CCC1=O)C(=O)N[C@H](CCCCCS)C(=O)Nc1ccccc1**	**40**	**1**
A136	**c1c(ccc(c1)NC(=O)CCCCCCC(=O)NO)c1ccccc1NC(=O)[C@@H](N)Cc1ccc(cc1)O**	**42.1**	**1**
A140	**c1(n(cc(c1)/C=C/C(=O)NO)C)C(=O)c1ccccc1**	**50**	**1**
A144	**C(=O)(/C=C/c1c(=O)n(ccc1)CCCc1cc2c(cc1)cccc2)NO**	**52**	**1**
A148	**c1ccc2c(c1)n(cc2)Cc1ccc(cc1)C(=O)NO**	**63**	**1**
A152	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**68**	**1**
A156	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**72**	**1**
A160	**c1c(cc2c(c1)n(c(n2)CCCC(=O)O)C)N(CCCl)CCCl**	**72**	**1**
A164	**c1cc(ccc1CNC(=O)/C=C/c1cccnc1)C(=O)Nc1cc(ccc1N)F**	**78**	**1**
A168	**c1cc2cc(c1)COC/C=C/COCc1c(ccc(Nc3nccc2n3)c1)OCCCCCCC(=O)NO**	**80**	**1**
A172	**c1(onc(c1)C(=O)NCCCCCCC(=O)NO)c1ccc(cc1)NC(=O)OC(C)(C)C**	**90.7**	**1**
A176	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**97**	**1**
A180	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)CC1**	**113**	**1**
A184	**c1cccc(c1)NC(=O)[C@H](CCCCCC(=O)NO)NC(=O)[C@H]1NC(=O)CC1**	**125**	**1**
A188	**c1(ccc(cc1)C(=O)NO)Cn1cc(c2c1cccc2)CCOC**	**126**	**1**
A192	**C(CCCCCC(=O)NO)C(=O)Nc1cnn(c1)Cc1cc(cc(c1)N=[N+]=[N-])CN=[N+]=[N-]**	**140**	**1**
A196	**c1(cc2cc(c1)/C=C\CO[C@@H](C(=O)Nc1c(OC2)cccc1)CCCCCC(=O)NO)OC**	**152**	**1**
A200	**c1ccc2c(c1)n(cc2)Cc1ccc(cc1)C(=O)NO**	**158**	**1**
A204	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**170**	**1**
A208	**c1c(ccc(c1)CNC(=O)NCCC(=O)NO)N(C)C**	**178**	**1**
A212	**c1(cc2cc(c1)CCCO[C@H](C(=O)Nc1c(OC2)cccc1)CCCCCC(=O)NO)OC**	**190**	**1**
A216	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**200**	**1**
A220	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**208**	**1**
A224	**c1ncccc1/C=C/C(=O)NCc1ccc(cc1)C(=O)Nc1ccc(cc1N)F**	**235**	**1**
A228	**c1(nc(cs1)c1cc(ccc1)N)NC(=O)CCCCc1onc(c1)C(=O)NO**	**250**	**1**
A232	**C1(=O)N(C(C(=O)N1c1ccc(cc1)Cl)CCC(=O)Nc1ccc(cc1)C(=O)NO)Cc1ccc(cc1)Br**	**268**	**1**
A236	**c1c(c(nc1c1ccc(cc1)OC)C(=O)NCc1ccc(cc1)C(=O)NO)c1ccoc1**	**290**	**1**
A240	**OCCN(C(=O)Cc1ccc(C(=O)NO)cc1)c1ccccc1**	**316**	**1**

A244	**c1c(c(nc1c1ccc(cc1)O)C(=O)NCc1ccc(cc1)C(=O)NO)c1ccsc1**	**316**	**1**
A248	**S(CC/C=C/[C@@H]1CC(=O)N[C@@H](c2scc(C(=O)N/C(=C\C)/C(=O)N[C@H](C(=O)O1)C(C)C)n2)C(C)C)C(=O)CCCCCCC**	**395**	**0**
A252	**C[C@@H](/C=C/C(=O)NO)/C=C(\C)/C(=O)c1ccc(cc1)N(C)C**	**403.35**	**0**
A256	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**456**	**0**
A260	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**456**	**0**
A264	**c1ccc2c(c1)cc(cn2)NC(=O)CCCCc1onc(c1)C(=O)NO**	**471**	**0**
A268	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**604**	**0**
A272	**N1[C@@H](CCCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)NCc1cc(ccc1)OCc1ccccc1**	**658**	**0**
A276	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**686**	**0**
A280	**S(CC/C=C/[C@@H]1CC(=O)NCc2scc(C(=O)N/C(=C\C)/C(=O)N[C@H](C(=O)O1)C(C)C)n2)C(=O)C**	**751**	**0**
A284	**c1(onc(c1)C(=O)NCCCCC(=O)NO)c1cccc(c1)N**	**840**	**0**
A288	**c1(onc(c1)C(=O)NCCCCC(=O)NO)c1cccc(c1)NC(=O)OC(C)(C)C**	**891**	**0**
A292	**c1cccc2c1OCCCCCCCOC(C(=O)N2)CCCCCC(=O)NO**	**958**	**0**
A296	**C(=O)(CCCCCCS)NC1CCCC1**	**1,170**	**0**
A300	**c1ccccc1NC(=O)CCCCCCC(=O)NO**	**1,259**	**0**
A304	**c1(ncnc2c1nc[nH]2)N[C@H](c1n(c(=O)c2c(n1)cccc2CCCCCC(=O)NO)c1ccccc1)CC**	**1,502**	**0**
A308	**c1c(ccc(c1)/C=C/C(=O)NO)/C=N/OCc1ccc(cc1)[N+](=O)[O-]**	**1,580**	**0**
A312	**c1c(c(c2c(c1)cccc2)OCCCCCCCCC(=O)NO)C(=O)OC**	**1,600**	**0**
A316	**c1(ccc(cc1)Cn1c2c(cccc2)c2c1cccc2)C(=O)NO**	**1,995**	**0**
A320	**c1(ccc(cc1)C(=O)NO)Cn1c2c(CCCC2)c2c1cccc2**	**2,512**	**0**
A324	**N(C(=O)CCCCCC(=O)Nc1ccccc1N)c1nc(cs1)c1cccc(c1)N**	**2,910**	**0**
A328	**OCCN(C(=O)Cc1ccc(C(=O)NO)cc1)c1ccccc1**	**3,000**	**0**
A332	**c1(ccc(cc1)C(=O)Nc1cc(ccc1N)c1cccs1)NC(=O)C**	**3,400**	**0**
A336	**c1cc(ccc1[C@@H]1C[C@H](CSc2nc(c(o2)c2ccccc2)c2ccccc2)O[C@@H](O1)c1ccc(cc1)NC(=O)CCCCCCC(=O)NO)CO**	**3,710**	**0**
A340	**C(=O)(/C=C/c1cc(ccc1)S(=O)(=O)n1c2c(cccn2)cc1)NO**	**4,130**	**0**
A344	**C1(NC(=O)CCCCCCC(=O)NO)CCCC1**	**4,890**	**0**
A348	**c1ccc(CCNC(=O)CCCNC(=O)OCC)cc1**	**5,000**	**0**
A352	**N(C(=O)CCCCCC(=O)Nc1ccccc1N)c1nc(cs1)c1cccc(c1)NC(=O)OC(C)(C)C**	**5,300**	**0**
A356	**N1[C@@H](CCCC1=O)C(=O)N[C@@H](CCCCCS)C(=O)N1Cc2c(CC1)cccc2**	**6,000**	**0**
A360	**c1c(cc2c(c1)n(c1c2C[S+](CC1)[O-])Cc1ccc(cc1)C(=O)NO)F**	**6,400**	**0**
A364	**C(=O)(c1ccc(cc1)CN(C(=O)Nc1ccccc1)CCCC)NO**	**7,570**	**0**
A368	**O=S(=O)(n1c2ccc(C(=O)c3cc(c(cc3)OC)OC)cc2cc1)c1cccc(c1)/C=C/C(=O)NO**	**8,240**	**0**
A372	**C(=O)(Nc1ccccc1)CCCCc1cc(no1)C(=O)NO**	**9,050**	**0**

A376	**n1c(n(c(=O)c2c(cccc12)NCc1ncc(cc1)C(=O)NO)c1ccccc1)[C@H](CC)Nc1ncnc2[nH]cnc12**	**9,996**	**0**
A380	**c1ccc2c(c1)c(c([nH]2)C)/C=C/c1c(cccc1)c1ccc(cc1)/C=C/C(=O)NO**	**10,200**	**0**
A384	**N(Cc1ccc(cc1)C(=O)Nc1ccccc1N)C(=O)OCc1cccnc1**	**11,100**	**0**
A388	**O=S1(=O)C2c3c(N(Cc4ccc(cc4)C(=O)NO)C2CC1)cccc3**	**13,000**	**0**
A392	**n1c(n(c(=O)c2c(cccc12)NCc1ccc(cc1)C(=O)NO)c1ccccc1)[C@H](CC)Nc1ncnc2[nH]cnc12**	**14,700**	**0**
A396	**c1(ncnc2c1nc[nH]2)N[C@H](c1nc(c2c(n1)cccc2)CCCCCC(=O)NO)CC**	**16,900**	**0**
A400	**n1c(cc(o1)c1ccc(cc1)N)C(=O)NCCCCCCC(C(=O)NC)N=O**	**21,800**	**0**
A404	**c1c(ccc(c1)C(=O)NO)NCc1csc2c1cc(cc2)Br**	**33,000**	**0**
A408	**c1(ccc(cc1)C(=O)NO)Cn1c2c(cccc2)c2c1CN1C(C2)C(=O)N(CC1=O)CCC**	**39,500**	**0**
A412	**N(Cc1ccc(cc1)C(=O)Nc1ccccc1N)C(=O)OCc1cccnc1**	**50,100**	**0**
A416	**c1(ccc2c(c1)CN(CC2)Cc1ccc(o1)c1cccc(c1)[N+](=O)[O-])C(=O)NO**	**71,100**	**0**
A420	**c1(ccc(cc1)/C=C/C(=O)NO)c1ccc(cc1)OC**	**91,000**	**0**
A424	**N(Cc1ccc(cc1)C(=O)Nc1ccccc1N)C(=O)OCc1cccnc1**	**>1.0e+4**	**0**
A428	**c1c(c(c(cc1O)C)c1nc2c(n1)cc(cc2)C(=O)NO)C**	**>10000**	**0**
A432	**c1c(c(cc(c1O)C(=O)N(c1ccc(cc1)NC(=O)CCCCCCCC(=O)NO)C)C(C)C)O**	**>10000**	**0**
A436	**O=C(N)c1c2c(C(C)(C)N(C2)c2cnc(cn2)C(F)(F)F)ccc1**	**>10000**	**0**
A440	**c1(cccc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)c1cnccc1**	**>100000**	**0**
A444	**c1c(ccc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O)N(C)C**	**>100000**	**0**
A448	**c1cccc(c1)NC(=O)CCCCCCC(C(=O)NC)N=O**	**>100000**	**0**
A452	**c1c(ccc(c1)/C=C/C=C/C(=O)NO)c1ccc(cc1)OC**	**>100000**	**0**
A456	**O=S(=O)(n1c2ccc(C(=O)c3cc(c(c(c3)OC)OC)OC)cc2cc1)c1ccc(cc1)/C=C/C(=O)NO**	**>100000**	**0**
A460	**c1cccc(c1)NC(=O)[C@H](CCCCCC(C(=O)NC)N=O)N(Cc1ccccc1)Cc1ccccc1**	**>100000**	**0**
A464	**c1(C(=O)N[C@H](C(=O)NO)C(N)(C)C)ccc(cc1)OCC#CC**	**>30000**	**0**
A468	**C1[C@H]2[C@H](N(Cc3ccc(cc3)C(=O)NO)c3c1cccc3)CCCC2**	**>30000**	**0**
A472	**N(C(=O)CCCCCC(=O)Nc1ccccc1N)c1nc(cs1)c1cccc(c1)NC(=O)OCC**	**>30000**	**0**
A476	**C(=O)(Nc1ccccc1)NCCCCCCNC(=O)CS**	**>30000**	**0**
A480	**c1c(cc(cc1)NC(=O)CCCCCCC(=O)Nc1ccccc1N)n1nnc(c1)c1ccccc1**	**>30000**	**0**
A483	**c1ccc2c(c1)n(c1c2CN(CC1)C)Cc1ccc(cc1)C(=O)NO**	**>30000**	**0**