Application of Classification-based Molecular Modelling & non-linear Pattern Recognition Methodologies to pinpoint Crucial Structural Attributes of HDAC6 inhibitors

By

Sandeep Jana

EXAM ROLL NO: **M4PHB24005**

CLASS ROLL NO: **002211402009**

REG. NO: **163653** of **2022-2023**

Department of Pharmaceutical Technology

Jadavpur university

Session-2022-2024

Under The guidance of

Dr. Nilanjan Adhikari

Natural Science Laboratory

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

2024

Jadavpur University

Jadavpur, Kolkata-700032

CERTIFICATE OF APPROVAL

This is to certify that **Sandeep Jana** (Exam Roll No. **M4PHB24005**, Reg. No. **163653** of **2022-2023**) has sincerely carried out the research work on the subject entitled "***Application of Classification-based Molecular Modelling & non-linear Pattern Recognition Methodologies to pinpoint Crucial Structural Attributes of HDAC6 inhibitors***" under the supervision of **Dr. Nilanjan Adhikari**, Assistant professor, Natural Science Laboratory, 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 Chemistry) of Jadavpur University. He has carried out the research work independently and sincerely with proper care and attention to our entire satisfaction.

27/8/24.

Head of the Department
Department of Pharmaceutical Technology
Jadavpur University, Kolkata-700032

Head
Dept. of Pharmaceutical Technology
Jadavpur University
Kolkata - 700 032, West Bengal, India

Nilanjan Adhikari 27/08/2024

Dr. Nilanjan Adhikari
Natural Science Laboratory
Department of Pharmaceutical Technology
Jadavpur University, Kolkata-700032

DR. NILANJAN ADHIKARI
Assistant Professor
Dept. of Pharm. Tech.
Jadavpur University
Kolkata - 700 032

Dipak Laha 27.8.24

Dean
Faculty of Engineering and Technology
Jadavpur University, Kolkata-700032

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

27/08/2024 13:24

Acknowledgment

This journey from being a science student to becoming a contributor to science required a lot of guidance and assistance from countless people. And I can't miss this opportunity to thank them for their every bit of kindness.

I take this opportunity to express my heartfelt gratitude and indebtedness to my guide **Dr. Nilanjan Adhikari**, Assistant professor, Department of Pharmaceutical Technology, Jadavpur University, Kolkata for his excellent guidance, endless encouragement, thoughtful insights and freedom throughout the term paper till its successful completion.

I owe my deep respect to **Prof. Amalesh Shamanta**, Head of the Department and **Prof. Sanmoy Karmakar** former 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 AICTE and Jadavpur University for their financial and equipmental support for my M. Pharm course.

I am both extremely honoured and grateful to **Prof. Tarun Jha**, **Mr. Sandip Kr. Baidya, Mr. Suvankar Banerjee** and **Ms. Subha Mondal** for their priceless guidance and support which assisted me to gather knowledge about the different aspects of this work. I would express my sincere thanks to my laboratory seniors **Mr**. **Jigme Sangay Dorjay Tamang Mr. Rahul Jana, Mr. Chayenta Sen, Mr. Vishal Saha,** and my juniors **Ms**. **Barsha Saha** and **Ms**. **Ishita Biswas,** Natural Science Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata-700032

I would like to thank **Dr. Balaram Gosh**, BITS-Pilani, Hyderabad, India for the continuous encouragement, necessary help and support to perform my work.

I would like to express my thanks to **Tuhin, Subhamoy, Rudra and Subham** for always being there for me.

Last, but not the least, I would like to thanks my father, **Ananda Mohan Jana**, my mother, **Pratima Jana** and my eldest brother, **Sudip Kumar Jana** for all the love and support throughout my entire course of work. I wish to express my deepest sense of respect and love to my entire family members for their support, encouragement without which this work could be incomplete

Date:
Place:

Sandeep Jana

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis contains literature survey and original research work performed by me (Sandeep Jana) as a part of my Master of Pharmacy studies. All the information in this document have 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: Sandeep Jana
EXAM ROLL NO: **M4PHB24005**
CLASS ROLL NO: **002211402009**
REG. NO: **163653** of **2022-2023**
*Thesis Title: "**Application of Classification-based Molecular Modelling & non-linear Pattern Recognition Methodologies to pinpoint Crucial Structural Attributes of HDAC6 inhibitors"***

Sandeep Jana

Dedicated to

Baba, Maa, Dada

Teachers and Friends

Preface

Histone deacetylase (HDACs) play a key role in chromatin remodelling, serving as epigenetic regulators of gene expression. Since their discovery, HDAC inhibitors has become a prominent area of research, particularly for targeted epigenetic modulation related to cancer, neurodegenerative disorders, and inflammatory diseases. Histone deacetylase 6 (HDAC6), a class IIb member of this metalloenzyme family, has garnered significant interest due to its distinctive structure, cytoplasmic localization, and ability to deacetylate specific non-histone substrates in the cytoplasm, such as α-tubulin, Hsp90, cortactin, peroxiredoxin, and heat shock transcription factor-1 (HSF-1). Elevated levels of HDAC6 have been observed in several conditions, including various cancers, neurodegenerative diseases like Alzheimer's and Parkinson's, and rare disorders such as amyotrophic lateral sclerosis, Rett syndrome, and Charcot-Marie-Tooth disease. Over the past decade, researchers have focused on elucidating the full spectrum of HDAC6's physiological functions, particularly in cancer, due to its role in coordinating numerous cellular processes critical to cancer development. Despite such diverse roles of HDAC6, none of the potential selective inhibitors have been translated in clinics to date therefor surging the development of selective HDAC6 inhibitors as potential therapeutic agents for these conditions. As an aid to the quest here some computational techniques were implemented in a set of quinazoline-hydroxamate based HDAC6 inhibitors hoping to find critical structural alerts and non-linear functions ruling the biological activity of HDAC6 inhibitors.

Content

Chapter 1

Introduction and Motivations

1.1 Drug

In pharmacology, a drug[1] is any chemical substance that upon administration to a living organism generates a biological response. There are vast array of drugs capable of causing different physiological effects.[2]

Drugs are generally used to cure a disease and alleviate many symptoms of illnesses, though some are not used to specifically treat a particular disease but rather acts as psychoactive chemical substance by impacting the central nervous system. In pharmaceutical terms drugs are chemicals substances with known structures used to treat, cure, prevent or diagnose a disease or to promote well-being.

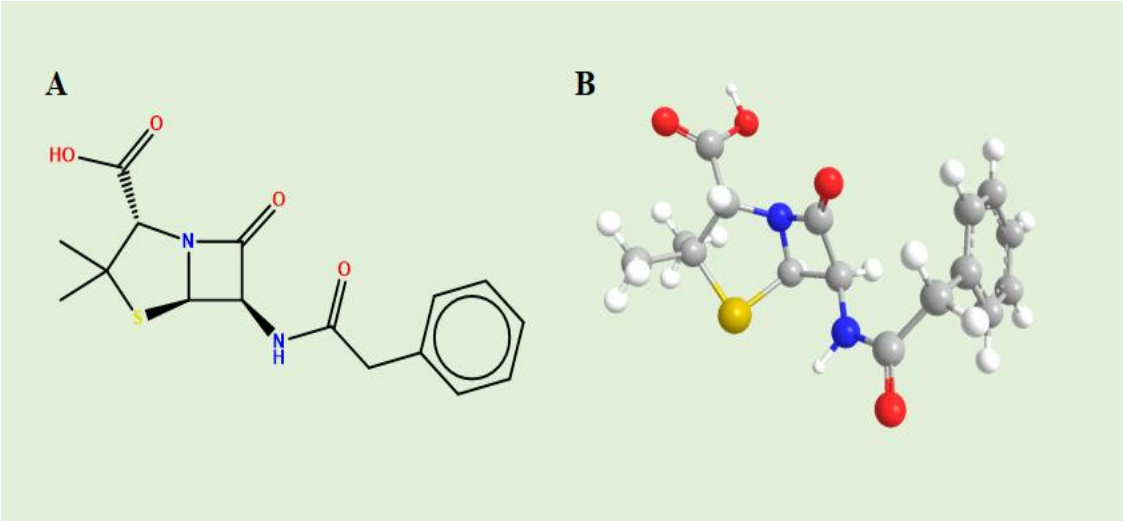


Figure 1.1: 2D (A) and 3D (B) molecular structure of penicillin G, a marvellous engineering by *Penicillium chrysogenum,* that saved millions of lives during World War II.

1.2 Drug Discovery

Discovering a drug from scratch is a long road with tremendous uncertainty. Most of the time investment of billions of dollars, man power and years of time leads to nowhere. As a consequence, number of potential drugs has become low whereas the need for newer drug has skyrocketed due to over use and drug tolerance in past few decades. Hence healthcare system continuously requires newer drugs to address the unmet medical needs across diverse therapeutic areas, and the pharmaceutical

industries primarily strive to deliver new drugs to market through complex activities of drug discovery.[3]

The art of drug discovery is just not only costly but rather goes through very time consuming and tedious process. To better shied light into this process Food and Drug Administration (FDA) has prescribed a timeline regarding the necessary steps to deliver a new drug that meets human requirements.

- The first stage in drug discovery and development process is to identify potential new drug candidates. Generally, researches dose this through new insights into the disease process that allows them to efficiently craft a product to stop or to optimize the effects of disease. Or else they can try to find new ways to evaluate small molecules to screen out beneficial effects against vast array of diseases or using cutting-edge technology to efficiently manipulate genetic material or target specific site within the body. The number of possible ways to identify a potential lead candidate is huge and, in this era, of artificial intelligence a lot of new ones are surfacing each and every day. In this stage thousands of compounds may be indicated as potential candidate for later stage of development to be medical treatment. However, after preliminary testing only a handful of compounds remains and are called for further development. Once a potential compound is identified for development, researchers move on to conduct experiments to gather information on the dosage, pharmacokinetic and pharmacodynamic properties of the concerned compound. This phase with all its exploratory phase takes around 6.5 years to complete.[4]
- The second stage is the preclinical research that concerns toxicity of the compounds. The number of tests performed is not so large but should provide sufficient information to decide whether the researcher should proceed to human trial. Typically, this phase can be done in few months.[4]
- Then comes the clinical stage where the drug is tested on humans to evaluate the effectiveness and possible side effects. Generally, this stage is sub-divided into 4 stages where a number of increasing volunteers participate in the test and correspondingly a number of decreasing drugs goes to the next stage. This stage requires around 5 years in order to complete all these clinical steps.[4]

- Finally, all the data from the clinical trials are submitted to FDA that decide if the study is sufficient and the drug is safe for commercialization or not. FDA, in case of an approval, will keep monitoring the new drug to have a complete picture, and in some case can change the indication of the drug or the dosage if necessary.

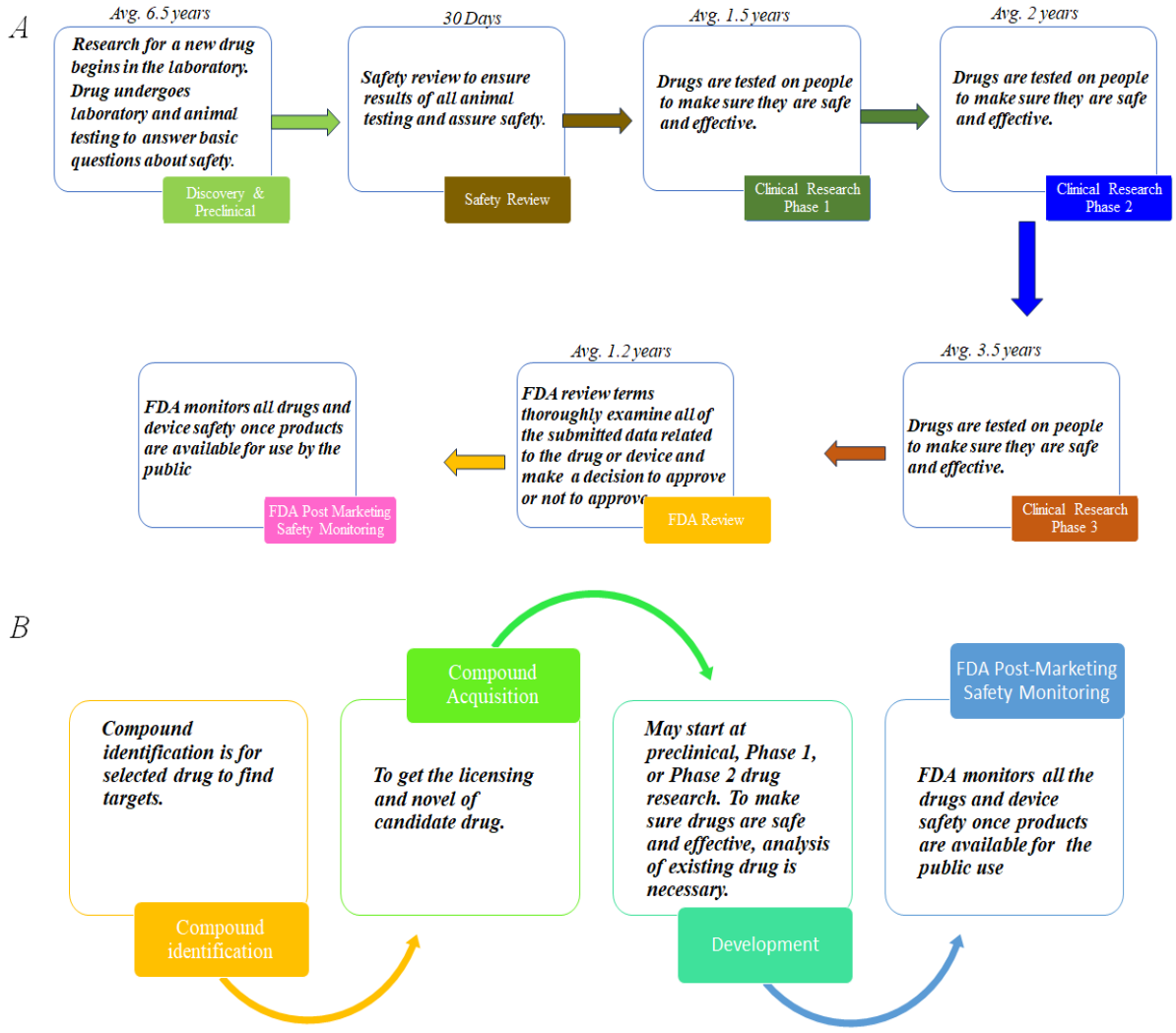


Figure 1.2: (A) Flowchart of the traditional drug development process. (B) Flowchart of drug repurposing

In the current scenario, all processes are time-consuming and expensive as well and the industry is under pressure owing to extremely stringent regulatory requirements, environmental concerns, and reduced incomes due to patent expiration. These issues have led to reduced R&D productivity in recent decade, hence innovative approaches

and increased collaboration between industry, academia, and governmental research institutions, with a common objective of constantly delivering quality medicines.[3]

1.3 Drug Repurposing

Drug repurposing also called as drug re-profiling, therapeutic switching or drug re-tasking is the identification of new therapeutic indications for known drugs. These drugs can either be commercially available and used daily in clinical setting or they can be drugs that have been "shelved" namely molecules that did not passed in clinical trials or for which projects have been discontinued for various reasons. In other words, drug repositioning can be defined *as renewing failed drugs and expanding successful ones*.[5]

As mentioned earlier, drug discovery is research and comes with no guarantee of success. Being so it is a high risk, slow, and expensive process.[3] The risk arises from the intense competition within the pharmaceutical sector, where developing a drug for an illness becomes a race against time to meet both health and economic demands.[6]

As of a report published by Eastern Research Group (ERG) only 2% of new molecule succeed in clinical trials and require around 10-15 years to reach the market for commercial use.[3] Alongside, PhRMA reports (Fig:1.3) shows that the yearly investment for drug discovery in USA is increasing year after year with respect to number of drugs approved by FDA.[7]

Certainly, the global trend is in constant rise with respect to money invested and decreasing number of approved drugs, making it harder to find a treatment for rare diseases that doesn't have a big share market. As the number of possible numbers of customer is low the volume of information on which researcher can rely becomes narrow making it too high-risk process both in health and economic front.

Drug repurposing seems to be a valid solution as this approach capitalizes on the fact that approved drug and many abandoned molecules have already been tested in humans and detailed information on their pharmacology, formulation, dose and toxicity is already available, making it advantageous in the economic front foot. Whereas, very list is known about the new molecules as these are in the preliminary stages of drug discovery. So, it is possible to reduce time, cost and risk of failure since the drug is already approved and declaring a new indication for a drug is much easier.[3]

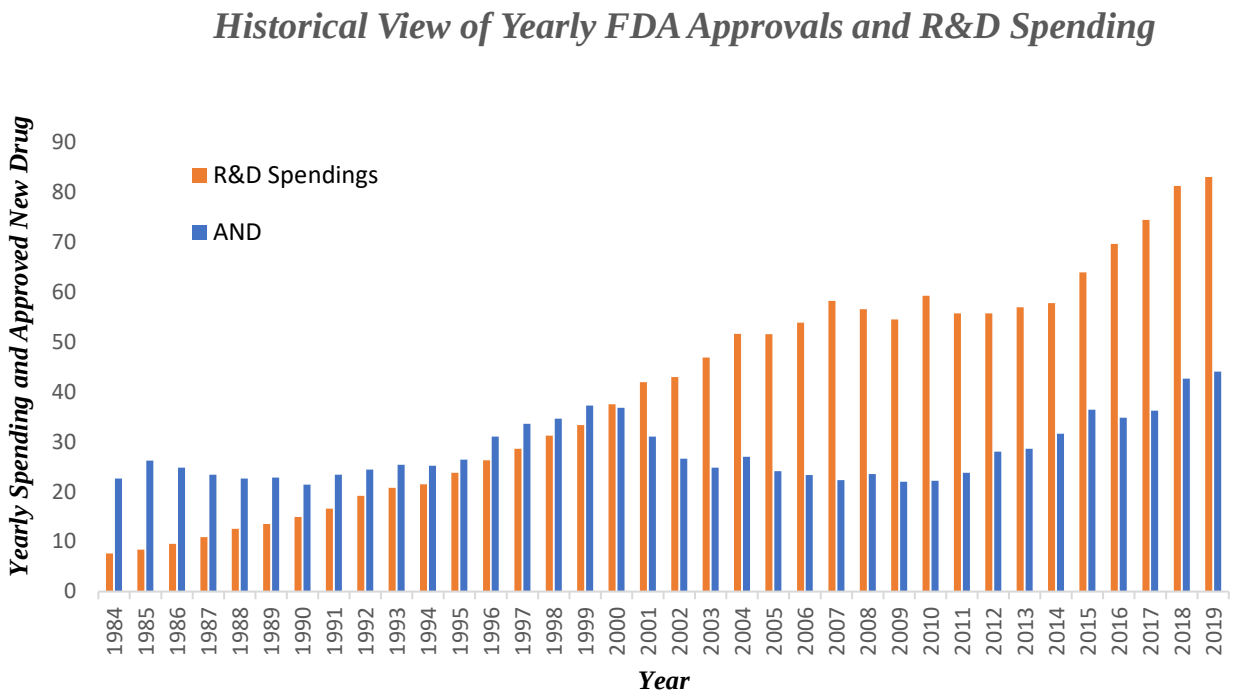


Figure 1.3: The amount of investment (in billion dollars) in drug development by Pharmaceutical Research and Manufacturers of America (PhRMA)[7] member companies and the number of approved drugs by FDA from 1984 to 2019.

1.3.1 Experimental approach towards drug discovery and repurposing

Historically, drug repurposing has often occurred as chance discoveries or unexpected findings. A drug becomes a potential candidate for repurposing when it's found to have off-target effects that could be beneficial for a different medical purpose or when new applications are identified. Successful instances of drug repurposing in the past have typically not followed a systematic approach but rather was serendipitous. Two classic example includes the repurposing of sildenafil citrate for treating erectile dysfunction that stemmed from retrospective clinical observations, and the repurposing of thalidomide for conditions like erythema nodosum leprosum and multiple myeloma which was an accidental discovery. These early successes have spurred efforts to develop a more systematic approach for identifying potential repurposable compounds, reducing reliance on chance discoveries. These methods have let to identification of numerous promising candidate drugs, some of which are undergoing advanced clinical trials. These repurposable drugs have the potential to treat both common and rare diseases, providing opportunities to bring valuable medications back to use that might otherwise have been overlooked.[8]

There are primarily two experimental approaches for the identification of a possible candidate for a drug repurposing. Proteomic methods like affinity chromatography and mass spectrometry are utilized to identify binding partners, focusing on structural compatibility. With the rise of chemical biology for target validation, analysing both target and off-targets of drugs and for repurposing has become standard practice. This method not only gathers valuable information but also aids in future research endeavours by providing a more comprehensive understanding of drug interaction and potential applications.[8] Phenotypic screening involves identification of compounds that exhibit effects relevant to a disease in model systems, regardless of prior knowledge of the target. In the realm of drug repurposing, if the screened compounds are either approved or under investigation, it can signal potential repurposing opportunities that can be readily explored further.[8]

1.3.2 Computational approach towards drug discovery and repurposing

The theory and clinical cases presented the reality of drug repurposing and I tried to briefly describe the fundamental reasons enabling new usages while emphasising the importance of chance discoveries in the process. Now it's a day dream of many medicinal chemists working in drug discovery field to be able to formally predict such repurposing scenarios and unveiling new pharmacology in an automated process. Numerous computational approaches have been developed since the beginning of cheminformatics in order to materialize this distant goal or at least get closer to it. A computational approach to the drug repurposing problem is not only cost-effective and time-efficient but can also adapt to different targets without requiring extensive efforts in majority of the cases.[3]

At the fundamental level, a repurposing initiative aims to connect a drug with a specific disease, essentially depicting a potential use or prescription probability for that molecule. To computationally generate new hypothesis for potential indications, biomedical concepts can be leveraged. Different biomedical concepts correspond to varying levels of abstraction within a biological system, ranging from biomedical details to broader perspectives, enabling computation or comparison. In general, a similarity metric is extracted from the specific property under investigation like chemical structure or gene expression levels, that serves as a descriptor to prioritize information and forecast potential new indications, resulting in the establishment of

connections between drugs, diseases, or molecular targets. These approaches can be categorized into groups based on the central property of analysis.

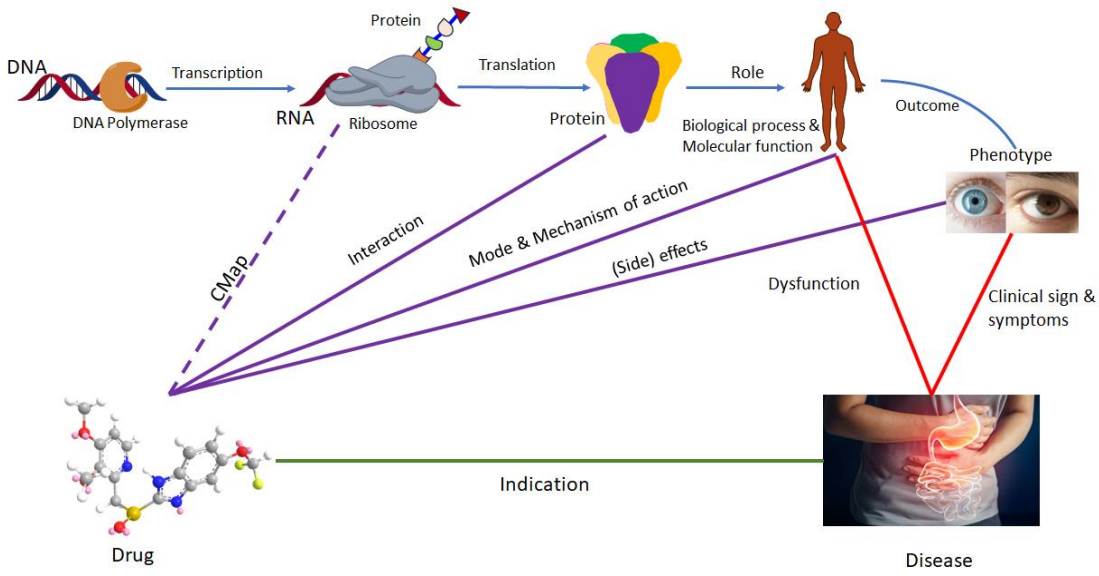


Figure 1.4: Conceptual Map of the relationship between the different biomedical concepts. Rational related to the drug and its action are in violate, disease in red and biomedical concepts are in blue. Computational drug repositioning methods are based on either one or a series of such concepts in order to forward new indication for a drug, ultimate goal (green edge).

1.3.2.1 Chemical structure-based approaches

Majority of the orally active drugs are small lipophilic molecules.[9] So, its logical to directly examine the chemical makeup when comparing drugs for similarity: if the structure are alike, they're expected to produce similar biological response. This rule of thumb is known as *similar property principle*[10] and it lies at the core of any quantitative structure-activity relationship study. Various techniques, such as fingerprints, clustering algorithms etc are employed to gauge structural alikeness between two chemicals.[11] These methodologies are instrumental in conducting ligand-based virtual screening, where active ligands are used as reference to identify structurally akin molecules within a target dataset, believed to possess similar biological activity.

When it comes to drug repurposing one can search only among approved compounds for instance. Noeske et al.[12] effectively utilized this strategy implementing an unsupervised machine learning algorithm (self-organising map), to group chemicals according to their structural characteristics. Molecular scaffolds were converted into vectors to be utilised during the clustering process. They were able to identify shared

activities among metabotropic glutamate receptor antagonists, observing their effects on additional protein targets such as dopamine D2, histamine H1, and muscarinic acetylcholine receptors. Though these findings were experimentally validated in vitro and shown to be active, yet were pharmacologically irrelevant due to weak binding. This new knowledge on off-target binding can pave the way for potential new usage for these drugs, by further modification and optimization of the molecular structure for instance.

Working with structural similarity for off-target identification, Keiser et al.[13] developed another intriguing method. For this project, known ligands were categorized according to their binding partners and chemical features. This method is known as *similarity ensemble approach and calculates whether a molecule will bind to a target based on the chemical features it shares with those of known ligands, using a statistical model to control for random similarity.*[14] In case of drug repurposing, researchers focused on testing only approved drugs. The findings from similarity analysis uncovered several off-target interactions. Subsequent retrospective analysis confirmed the approach's validity, and experimental validation of some predicted off-target bindings provided valuable insights into the pharmacological mechanisms of certain drugs. Notably, for fabahistin, the affinity for an off-target receptor (5-HT5A) was superior to its known primary receptor (H1), suggesting promising alternative therapeutic uses.

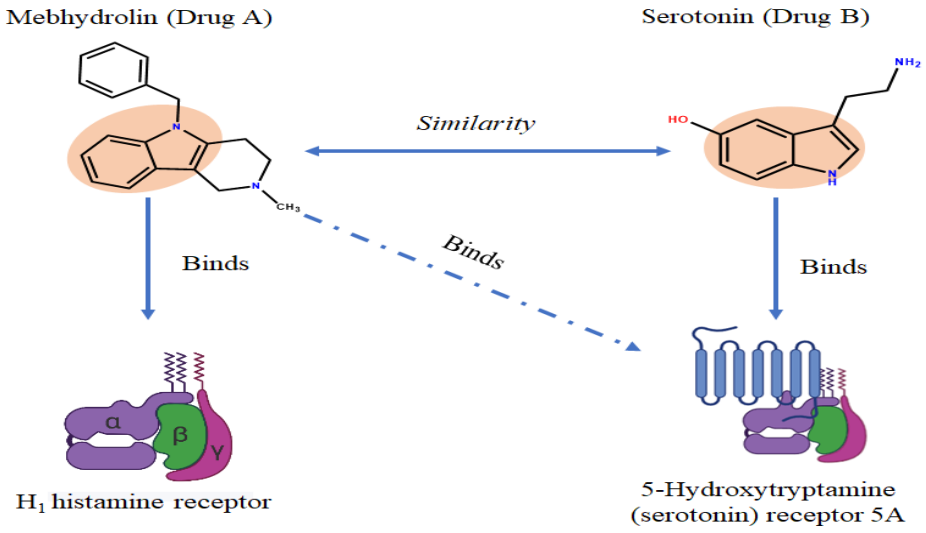


Figure 1.5: Drug repurposing using the chemical structure. Compounds with similar structures have similar biological activities (similarity principle). Drug A shares common scaffold with drug B. This observation leads to conclusion that drug A could be active on the canonical target of drug B.

This chemical structure-based approaches are intuitive and the foundation of it lies on the *similar property principle* (Fig 1.5). Although in reality, significantly small change such as alteration in bond order, an atom or even the nature of bond to a molecular structure can drastically change the biological activity. Furthermore, predictions from different methodologies show minimal agreement among them stressing the obscurity to pick the adequate one for the right scenario.[11] Other challenge lies in the fact that some compounds undergo chemical modification inside the body prior being pharmacologically active, therefore the structure as documented in databases may undermine the effectiveness of a predictive statistical model.

1.3.2.2 Gene expression and functional genomics-based approaches

Just like computer where each and every task is executed as a result of software execution, every living systems behaviour is dictated by its gene expression in a particular setting. Expression level of certain genes is very much dependent on the state of the system, and can be identified and quantised by the relative number of their messenger RNA (mRNA) molecules transcribed. Differently expressed genes can function as emissary to pervade a molecular effect, known as *gene expression signature*. This type of experimentation is generally conducted in microarray, containing probes for the genes of interest. The method offers a clear understanding of the condition under investigation and have been effectively employed (particularly the Connectivity Map)[15] to find indication for marketed drugs.

The axiom behind CMap stated that the action of a drug can be apprehended and nudged by looking at the gene expression level resulting from its administration into the biological system. Messenger RNA that function as gene expression signature, not only reflect the activity of drug but also can act as proxy of a disease state. Based on this assumption, Sirota et al.[16] conducted a set of experiments from Gene Expression Omnibus to capture disease signatures from gene expression profiles. Further integration of this data with similarity values between drugs, derived from the CMap, the researchers were able to identify negative correlation between a cluster of related diseases and the signatures of current treatment regime. This anti-correlation predicted cimetidine (anti-ulcer drug) to be a potential treatment for lung cancer, which was further validated by in vitro and in vivo experimentation on a mouse model. The research also found that topiramate, typically prescribed as an anticonvulsant, shows promise as a therapeutic agent for treating inflammatory bowel disease (IBD), a

condition without a current cure.[17] This new indication of topiramate was strongly supported by in vivo testing in a rodent model. This study underscores a valuable aspect of transcriptomics: even when limited molecular knowledge exists about the specific disease mechanisms, analysing gene expression patterns can effectively bypass detailed mechanistic understanding and accurately pinpoint potential treatment options.

Gene expression analysis and CMap has paved the way for numerous drug repurposing. This technique is very much handy as it does not require much prior knowledge about the action of drug or the pathology behind a phenotype; rather it majorly relies on creation of signatures directly from mRNA readouts to retrieve unknown drug-disease associations. Transcriptomics provides another valuable insight, suggesting that the functional role of drugs, rather than their chemical structures, is the primary determinant of success. In this approach, drugs are evaluated solely based on their impact and function within the biological system, as indicated by their gene expression signature. Consequently, the chemical composition of the drug becomes largely insignificant for this analysis.

Irrespective of such massive success, this technique suffers from significant drawbacks and requires improvement.[18] Firstly, the expression profile of the drug or disease must be available. The CMap provides a relatively small list of molecules which is far from being representative of all approved and experimental drugs, limiting the compounds that can be investigated. Secondly, gene expression profiles can potentially characterize disease conditions or drug actions. The CMap resource, however, lacks tissue-specific data as it was primarily build using response from cancer cells, limiting its relevance across all disease categories. Finally, transcriptomics data also present considerable challenges in terms of statistical analysis.[15]

1.3.2.3 Protein structure and molecular docking-based approaches

Majority of the small molecules if not all medicates their pharmacological effect by interacting with proteins that can be analysed with certain accuracy using computer software by modelling the three-dimensional (3D) structures of the target and the drug. This practice is called molecular docking, a commonly used method in drug discovery process, mainly used to identify and optimize binding affinities in the active site of the target in order to increase the potency of the drug developed.[19] Due to the widespread use of molecular docking, drug repositioning efforts using this method are quite common. Given that many compounds are known to interact with multiple proteins, the

objective is to identify these potential off-targets by screening them against the 3D structures of proteins in a specific database. If the predicted off-targets are relevant to a disease, the drug can be repositioned based on these findings.

In this context, several recent studies have concentrated on binding sites, comparing their relative similarities (Fig 1.6).[19] By examining only, the structures of protein active sites, researchers ensure they remain as close as possible to the biochemical and physical realities of the interaction. From an analysis of over 6000 binding site structures, De Franchi et al.[20] identified synapsin I, a protein involved in neurotransmitter release regulation, as a new target for the drug saturosporine, which is known to bind Pim-1 kinase. This finding was experimentally validated in vitro, though the pharmacological significance of this new target has yet to be demonstrated.

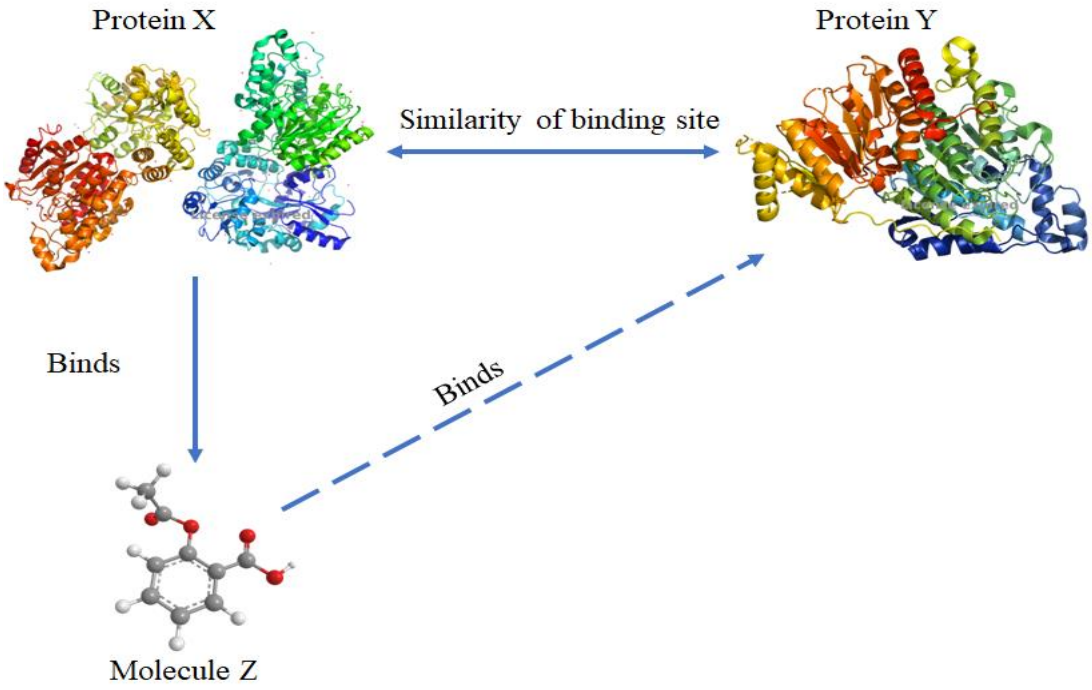


Figure 1.6: Drug repurposing using protein structure and binding site. The 3D structure of proteins and their respective binding sites can be compared using scoring function. On this basis, it is assumed that similar binding sites can bind same ligand as depicted in the picture.

Zahler et al.[21] conducted an inverse screening by docking a single compound across multiple binding sites to map the off-target binding landscape of kinase inhibitors. This class of drugs, widely used in cancer therapy, is known for its “promiscuous” behaviour. The virtual screening identified PDK1 as a new off-target of indirubin. This

prediction was validated in vitro using a phenotypic cell proliferation assay, demonstrating the effectiveness of the approach and offering insight into the side effects of kinase inhibitors.

Likely, Kinnings et al.[22] used molecular docking technique to address drug-resistant tuberculosis. It is true that occasionally the bacteria causing the illness will not respond to first-line treatment; in these situations, the pathogen must be eradicated. To optimize the problem the authors used a technique called '*selective optimization of side activity*' (SOSA), originally developed to progressively move away from the original indication and optimize a compound across protein families.[23] This methodology goes like this: first the binding site extraction from the 3D structure of protein sequence, second search for similar binding sites across the proteome using search algorithm, and finally manual docking analysis to make sure the physical interaction is possible. Using this method the researchers were able to predict the potential implication of two approved drugs entacapone and tolcapone in tuberculosis through binding of enoyl-acyl carrier protein reductase, an enzyme that facilitates synthesis of fatty acid in *Mycobacterium tuberculosis*. This hypothesis was validated experimentally using commercially available tablet, which not only showed effectiveness but also bypassed the drug resistance encountered in *Mycobacterium tuberculosis* making a valuable treatment alternative for affected patients.

Well, no method is absolute, so dose molecular docking, despite such success it also not immune to drawbacks. Firstly, 3D structure must be available, though databases such as Protein Data Bank (PDB) comes fourth as a saviour still they are very far from covering the whole proteome.[19] Secondly, automatically recognising a binding site possess a real challenge especially when the protein is crystalised without a ligand. Finally, as all methodologies are prone to generation of large number of false positives, experimental and manual validation become the only viable solution to judge the prediction. Furthermore, single amino acid difference can lead to a totally different pharmacology of binding site,[24] a major pitfall especially when the structures are analysed and aligned in an automated fashion.

Arguably, protein-based approaches are the closest methodology to the actual physical interaction between drug and protein target. Docking approaches offer an intricate, detailed view of the biochemical complex, but still face challenges in modelling. Additionally, identifying off-target proteins does not always led to repositioning

opportunities, and the results must always be interpreted within a broader biological context.

1.3.2.4 Phenotype and side-effect-based approaches

Phenotype is a set of characteristics or traits attributed to an organism, such as morphology, developmental, biochemical or physiological properties.[25] In biological sciences this concept is majorly used to indicate key observation when looking at a living organism. The phenotype is perhaps the most fundamental interaction between biomedical scientists and their subjects. For instance, as Darwin journeyed around the world, he gathered evidence for evolution by studying the phenotypes of barnacles.[26] Similarly, Gregor Mendel first described inheritance based on the traits observed in pea plants.[27] Though none of those scientists were aware of the molecular mechanism underling those observable patterns, yet their phenotypic observations were strong enough to forward valid conclusion. This practice is still in use in clinical settings, every time a doctor diagnoses a patient, he or she primarily relies on phenotypic signs present in the patient. This phenotypic-driven approach is also routinely used in drug discovery, interestingly some recent studies suggest that it's one of the best techniques to bring new medicine.[28] This high success rate of this method can be attributed to the fact that phenotypic observations more accurately reflect the underlying system, preserving the physiological context. This contrasts with target-based approach, where in vitro lead compound has higher likelihood of remaining active when progressing to animal models and, eventually clinical trials.[29]

In the context of drug repositioning, side effects can also be considered as phenotypes. The story of sildenafil is a classic example: regardless of a drugs potency in animal models or in vitro assay, its true pharmacological nature only emerges during clinical trials. Accurately characterizing these side effects can aid in repositioning a drug or uncovering new interaction patterns (Fig 1.7).

Drugs with similar target binding profiles tend to cause similar side effects.[30,31] Based on this justification, Campillos et al.[32] defined the adverse effect profiles for medicines that were approved and then utilized similarity among these to determine the drug's intended targets. Using text mining on package inserts, the side effects were first extracted in order to create a statistical model that would tell us how likely it was the two drugs would bind same target. The authors then focused on substances that, based on the model, had high likelihood of sharing a target but belonged to distinct therapeutic

categories. Twenty of these predictions were validated experimentally confirming thirteen of them of which eleven had inhibition constant less than ten micromolar. The method's novelty underscores how side effects are molecularly relevant, offering a pathway to identify off-target effects and repurpose therapeutic molecules for new indications. An intriguing aspect of this method is how it represents side effects. Like any phenotypic trait, words or terms based on observation remain the effective way to describe them. In this study the researchers utilized the Unified Medical Language System (UMLS),[33] a controlled vocabulary provided by the National Institute of Health (NIH). The experimental validation by the research team suggested that ontologies and controlled vocabularies can indeed produce reliable predictions.

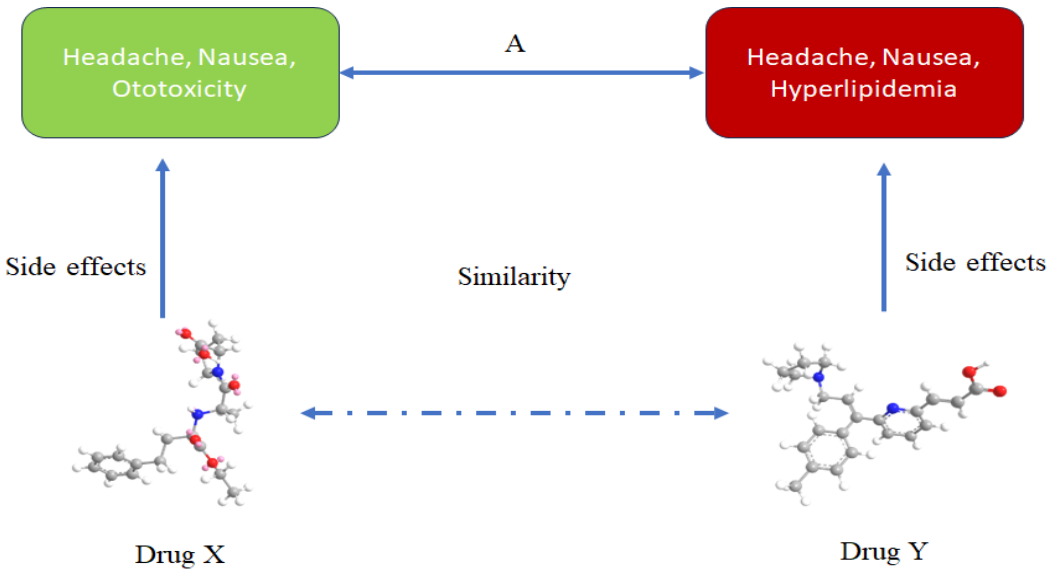


Figure 1.7: Drug repurposing using phenotype information. Knowledge about the phenotypic outcome triggered by a drug can be used in order to establish relative similarities. (A) The diagram illustrates a theoretical example using reported side-effects: the more side-effects are commonly shared by two drugs, the more similar these two drugs are. The similarity can be used to either derived potential off-target or new indication.

Another approach was laid down by Yang and Agarwal,[34] where they used side-effects from SIDER[35] database to link diseases in search of potential drug repositioning opportunities. Map was created connecting molecules to pathologies based on information available in pharmacogenomics knowledge base (PharmGKB).[36] This method values evidence indicating that drugs used to treat similar diseases tend to have

similar side effects. Considering this, it can be hypothesised that drugs having similar side-effects may have common mode of action, indicating that two drugs with significant number of similar side-effects can be used to treat the same pathology even if they belong to different chemical class.

Half a century ago, drug discovery centered around phenotypes. However, with the rise of molecular biology, the art of drug discovery took a sharp turn towards target-based approach.[29] Despite this shift, phenotype-based approach still remains valuable even today as it repots the effect of a given substance on entire organism, which is more relevant for clinical application.[37,38]

1.3.2.5 Genetic variation-based approaches

At the molecular level, genetic variations offer significant insights into drug repositioning opportunities. With the advent of high-throughput DNA sequencing techniques and advanced analysis pipelines, sequencing individuals and studying their genotypes has become more accessible, allowing researchers to pinpoint common mutations in DNA that are strongly associated with phenotypic traits. This approach, known as genome-wide association study (GWAS), is used to link single-nucleotide polymorphism (SNPs) to diseases. Information about SNPs and their diseases associations are publicly available in databases, such as one maintained by the National Human Genome Research Institute (http://www.genome.gov/gwastudies/). Sanseau et al.[39] utilised this resource to analysis and screen out potential new indications for protein targets identified through GWAS. The approach is based on the premise that an association between an SNP and a trait identified in a GWAS can be interpreted as a link between a gene and a disease (when the traits considered are disease, as depicted in Fig: 1.8). Knowing that a drug targets the product of a specific gene, one can infer that the drug's indication might correspond to the disease studied in the GWAS. For example, a SNP of the gene encoding 3-hydroxy-3-methylglutaryl-CoA was found significantly associated with the trait LDL cholesterol.[40] The statins, a class of drug known to target this gene product and are prescribes as cholesterol lowering agent (hypercholesterolemia). The authors were able to identify 97 cases where SNPs supported the current drug indication, providing greater confidence in the biological role of the protein. Conversely, they found 123 instances where the trait associated with the gene did not match the drug's current indication, suggesting these associations as opportunities for drug repositioning. For instance, denosumab, a monoclonal antibody

used to treat osteoporosis and bone cancer, targets the protein TNFSF11 (tumour necrosis factor superfamily, member 11).

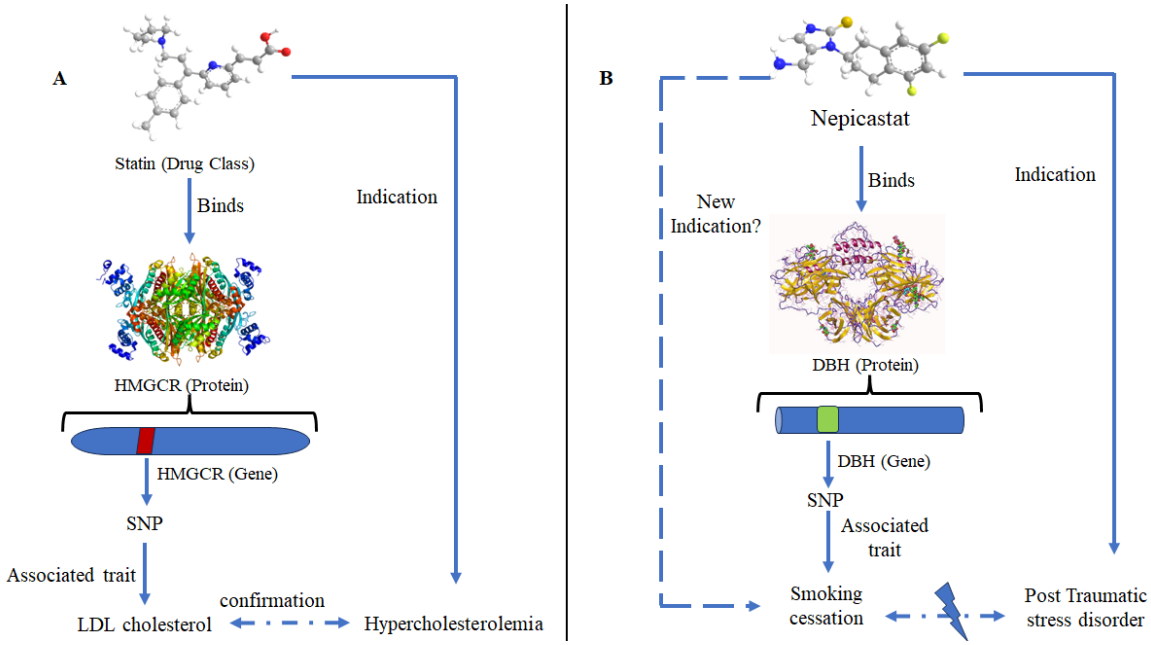


Figure 1.8: Drug repurposing using genetic information. (A) Single-nucleotide polymorphism (SNP) are associated with phenotypic trait, here LDL cholesterol. The gene where the SNP is found (HMGCR) encodes for a protein, targeted by statins. Statins are indicated as cholesterol lowering agents, which are confirmed by the trait associated with the SNP. (B) sometimes the trait associated with the SNP diverges from the indication of the drugs, as shown on the diagram (PTSD against smoking cessation). In such case, a repurposing hypothesis can be generated.

A SNP in this protein has been linked to Crohn's disease, indicating that denosumab could potentially be tested for treating Crohn's disease.[41] Another such example is nepicastat, a small molecule primarily indicated for treating cocaine addiction and post-traumatic stress disorder. Its target, dopamine beta-hydroxylase (DBH), has been associated with smoking cessation in a GWAS,[41] suggesting a potential new application of the drug for those who wants to quit smoking.

Just like any other methodology it also comes with pitfalls, for example we can take the instance of prediction made for NOS2 (nitric oxide synthase 2) inhibitor where GWAS predicted it be active against psoriasis though experiments showed no results. In practice, the relationship between genes and disease is complex, and additional information is often necessary to fully understand a drug's potential effects. Furthermore, GWAS does not indicate the direction of the pharmacological effect,

making it troublesome to determine whether the agonist or the antagonist will produce the desired results. Despite all these short comings, the remarkable advancements in genome sequencing suggest that this approach or a similar methodology, could become a major role player in drug repurposing in the coming years.

1.3.2.6 Disease network-based approaches

Traditionally, diseases have been classified based on the cause of the pathology (e.g., infection) or the observed biological dysfunction (e.g., uncontrolled cell growth). Since similar diseases are treated similarly, a more detailed understanding of the relationship between pathologies can lead to drug repositioning hypothesis. Here I will briefly outline some work done in this area, specifically on constructing a "diseasome" or a network of relationships between diseases.

Chiang and Butte[42] defined disease based on the available treatment regime and off-label indications. Although this approach is relatively simplistic, it is supported by successful examples and is commonly practiced in clinical settings. The authors proposed the use of medication that is only indicated for one of two similar diseases as a therapy for the other, a technique known as associative indication transfer. Using this approach on 700 diseases and 2000 medication they were able to generate over 150,000 new associations. Remarkably, the new indications aligned with the clinical trial data, with the predicted new uses frequently being reported by practitioner's (showing a 12-fold enrichment compared to random chances). For example, atorvastatin, a cholesterol-lowering agent, was predicted to be effective for asthma, Crohn's disease, and myocardial infarction; all these associations have been positively confirmed in clinical trials, validating the methodology. For the same drug some new association have no clinical knowledge, such as activity in breast cancer and osteosarcoma. Thus, it is feasible to explore the effects of a drug on these conditions. This work demonstrates on approach to relating diseases. Li and Agarwal[43] developed a similar methodology by constructing network based on shared pathway, where they created a map linking diseases using public resources such as Reactome, KEGG pathways, and text mining. Considering diseases with commonly deregulated pathways as similar. The properties of the resulting graph were analysed, and the authors showed how their work could shed light into disease relationships. Although no analysis for repositioning opportunities was conducted, the map can serve as a starting point for identifying such possibilities. Suthram et al.[44] used a similar approach except they constructed the

disease map based on gene expression profile and protein network rather than relying on shared pathway like Li and Agarwal. An analysis of such graph reviled 59 functional molecules shared by half of the disease studied. These molecules connect pathologies at the molecular level, aiding in the understanding of the system's internal dynamics. Similar to aforementioned methods once a disease network is established, it can be viewed as a stepping stone for generating drug repositioning hypothesis. In conclusion, although disease maps do not directly address drug repositioning, they can offer valuable insight into drug usage. These approaches also challenge the current system of classifying diseases by considering molecular information as signature of definition.

1.3.2.7 Machine learning and concepts combination approaches

Most of the aforementioned approaches heavily relies on concept of the map shown in (figure 1.4) and orients their analysis around it. It is completely possible to train a machine learning model using these biomedical descriptors to generate predictions. Two instances of drug repositioning study have come out from this perspective, where firstly a series of biomedical heuristics is defined to train a ML algorithm using known data and finally predictions are made using the trained model.

In the year 2011 Gottlieb et al.[45] presented a method called PREDICT based on this concept, where to train the ML algorithm they used drug-drug and disease-disease association separately. The drug-drug association were characterised using their chemical fingerprint followed by reported and predicted list of side effects. To further enrich model, they introduced information related to the targets of the drugs: such as the sequence similarity of the protein, distance in the protein-protein interaction network along with semantic similarity of their GO annotation. The disease-disease associations were defined in a pretty straightforward manner using semantic similarities derived from Human Phenotype ontology (HPO), using the annotation from the Online Mendelian Inheritance in Man (OMIM) database. The authors then trained a logistic regression classifier using this association maps to distinguish genuine association from the false ones. The model's performance was compared to predictions made by other methods, such as GUILT-by-association and CMap approaches,[15] presented earlier in this chapter. The evaluation revealed minimal overlap among the different mythologies, indicating difficulties in aligning the various datasets, as the diseases and drugs considered often differed. Subsequently, some drug repositioning predictions were made and addressed using clinical trials data. Approximately one-third of these

predictions had already been investigated, lending confidence to the methodology's outcomes. In the final step of the study, the authors replaced the disease-disease associations based on phenotypic similarity with the gene expression profiles. This step aimed to test the method for personalised medicine: given a patient's gene expression profile, could PREDICT identify the best drug for the individual? The results were promising, with the method achieving high recall and specificity (area under curve of 0.92 obtained from receiver operating characteristic curve), highlighting a solid proof-of-concept for the algorithm.

The second method presented by Napolitano et al.[46] is very similar to PREDICT only differing in the ML algorithm which was Support Vector Machine in their case. The algorithm was employed to predict therapeutic categories within the Anatomical Therapeutic Chemical (ATC) Classification system, with misclassifications being reinterpreted as drug repositioning hypothesis. The researchers also used structural similarity, protein-protein interaction network distance, and gene expression data as initial features to train the SVM. Following standard machine learning evaluation procedures, the authors generated repositioning predictions. The primary hypothesis suggested that anthelmintic compounds could be effective as antineoplastic agents and that antineoplastic drugs could be repurposed as systemic antibacterials.

Machine learning based approaches to drug repositioning offers a means to integrate various descriptors into a single statistical model, aiming to improve predictions accuracy. However, these techniques encounter significant challenges. One major issue is interpreting the repurposing hypothesis: the statistical model functions as a black box, obscuring the rationale behind selecting a compound. Many hypotheses turn out to be obvious cases that a biologist could easily explain by examining the chemical structure or known-off-target effects of the compounds. Thus, outcome may be due to overtraining the model. additionally, the biomedical significance of incorporating a large number of descriptors is questionable; given the complexity and uniqueness of diseases, excessive information may obscure critical biological mechanistic details, leading to less meaningful results.

1.4 Summary

A large number of approaches have been tried and tested in order to computationally repurpose drugs. This field of computational drug repurposing is still in its infancy, as revealed by two factors.

Firstly, it remains unclear which method yields the best results and why. The only definitive way to evaluate these predictions is when a drug, suggested by an in-silico hypothesis, becomes routinely used in the clinic. To my knowledge, no compelling example of this exists yet. This is not disappointing, as developing a new drug is lengthy process, often exceeding twelve years, and is fraught with lengthy legal and economic obstacles. Considering that the first study reported by PubMed for the keyword "computational drug repositioning" was published in 2006,[47] and this trend spiked in 2021 in the covid period while the world was going through a crisis, it seems reasonable not to have much clinical examples yet.

Secondly, each method tries to answer drug repurposing problem form different perspective or biomedical concept, adding multiple layers of complexity in the evaluation process. objectively aligning the results from various approaches is troublesome as the initial dataset pertain to different molecules and diseases, leading to different outcomes. It would be advantageous for the community to have a standardized dataset that includes both the legal indications and the known, confirmed alternative once. Computational methods could use such a resource to benchmark their performance, assess their predictive capabilities, and conduct error analysis. In this context, the immaturity of the field also fosters creativity, as evidenced by the numerous methods that have been developed.

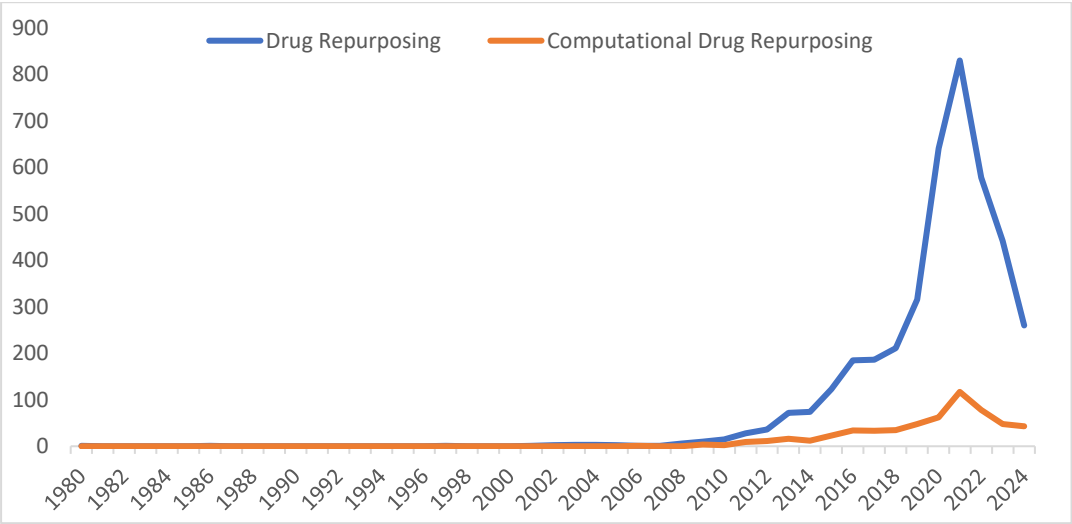


Figure 1.9: Evolution trend of the documents related to drug repositioning in PubMed database since 1980-2024.

In conclusion, computational drug repositioning is becoming an increasingly popular topic with in the scientific community, as illustrated in Figure 1.9. Numerous methods have been developed over the past decade which are summarised and discussed in this

chapter. Drug repositioning is a small puzzle of a bigger problem known as indication discovery and network biology.[48] This approach leverages our growing understanding of systemic behaviour to computationally design smart drug. Various levels of abstraction can be considered, as demonstrated by the range of biomedical concepts used as starting points.

Chapter 2

What makes HDAC6 a good target for drug?

2.1 Introduction

Before delving deep into the problem statement why HDAC6 is a good target, lets first discuss about what is it, where did it come from? To answer these questions, we have to delve into the realms of Epigenetics.

DNA the central molecule of life, holds the genetic information for all the eukaryotes. This DNA upon transcription forms RNA which ultimately forms protein through translation and genetic information is expressed in terms of phenotype. This whole process is called central dogma.[49] Epigenetics is a crucial part of it. Normally in cell DNA are supper coiled and is in dormant state. Epigenetic modulator such as HAT and HDAC plays the role of unwinding and rewinding of DNA from histone octamer, making it accessible to topoisomerase for central dogma to take place.[50] This HAT family of enzymes catalyses an acetylation reaction at specific lysine residues of Histone 2 protein of histone octamer complex.[51] This acetylation adds negative charge to histone tail, which ultimately repels negatively charged DNA molecule making it more accessible to topoisomerase and other enzymes associated with replication and transcription. HDAC family dose the simply opposite, it helps to splice of the extra acetate group which was added by the HAT enzyme, leading to shift in polarity and stable binding of DNA on histone octamer, hence genetic repression.[52]

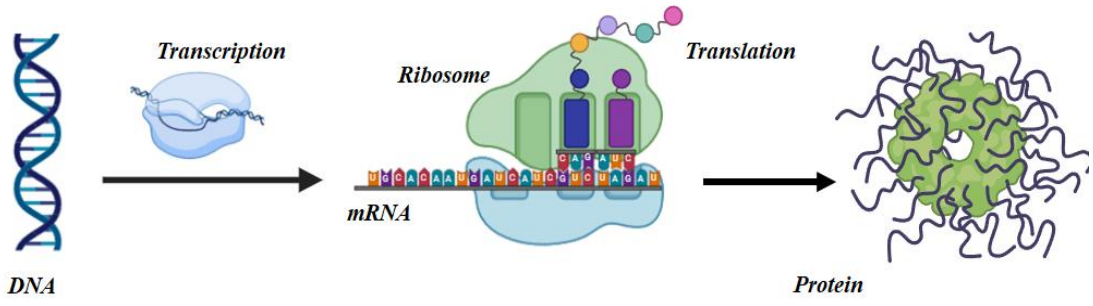


Figure 2.1: Central dogma of life.

Here I will briefly discuss about the HDACs and will focus majorly on HDAC6 as it is of our main concern. These Histone deacetylases or HDACs are the key regulator of cellular protein acetylation level.[51] The name may suggest that they are only specific to histone, well that is not the case, besides histones they also deacetylates proteins like p53, E2F, α-tubulin, heat shock protein 90 (HSP90), cortactin and Myo D. This wide

substrate specificity indicates that HDACs are a major role player in the cellular function and homeostasis.[53-55] Therefore, dysregulation of HDACs contributes to development of several diseases ranging from multiple forms of cancer, neurological disorders, inflammatory disease, autoimmune disease, to cardiac and pulmonary diseases, making this family of metalloenzyme crucial drug targets.[56-59]

To date, 18 HDAC isoforms have been discovered in humans that shows sequence homology to yeast protein ortholog hda1. These HDACs have been grouped into four different classes based on their co-factors and catalytic domain organisation, of which class I (HDAC1, HDAC2, HDAC3 and HDAC8) class IIa (HDAC4, HDAC5, HDAC7 and HDAC9), class IIb (HDAC6 and HDAC10), and class IV (HDAC11) uses a Zn^{2+} as a cofactor, whereas class III or sirtuins (SIRT1-7) exerts its catalytic functions using nicotinamide adenine dinucleotide (NAD^{+}) as co-enzyme.[60,61] Since identification of this class of enzymes, numerous inhibitors with varying efficiency (at nM to μM concentration) have been developed ranging from naturally derived trichostatin A and trapoxin to synthesised molecules like butyrate. The therapeutics of these inhibitors are judged on the basis of their capacity to acetylate different cellular proteins, upregulation of p21 as well as downregulation of tumour proliferation via apoptosis.[62] Currently on our arsenal we have six FDA approved HDAC inhibitors namely vorinostat (SAHA, 1), belinostat (2), Panobinostat (3), [FDA has been withdrawn in 2022], romidepsin (4) and pracinostat (5) and are primarily indicated for the treatment of refractory or relapsed cutaneous and peripheral T cell lymphomas as well as multiple myeloma.[58,63] Other than this five, chidamide (6) is another potent HDAC inhibitor approved in China for similar clinical condition. The major drawback these inhibitors suffer from is lack of isoform selectivity, leading to various adverse effects such as fatigue, nausea, vomiting, cardiotoxicity, and thrombocytopenia.[64,65]

With the advancements in the field of molecular biology and genetics several HDAC isoforms have been identified and studied, and of these isoforms class IIb member HDAC6 came to the center of attention due to its unique structure since its discovery in 1999.[66] Another unique aspect of HDAC6 is its cytoplasmic localization whereas other HDACs are nuclear enzyme. Being cytoplasmic it deacetylase specific cytosolic non-histone substrate such as α-tubulin, Hsp90, cortactin, peroxiredoxin, and heat shock transcription factor-1 (HSF-1).

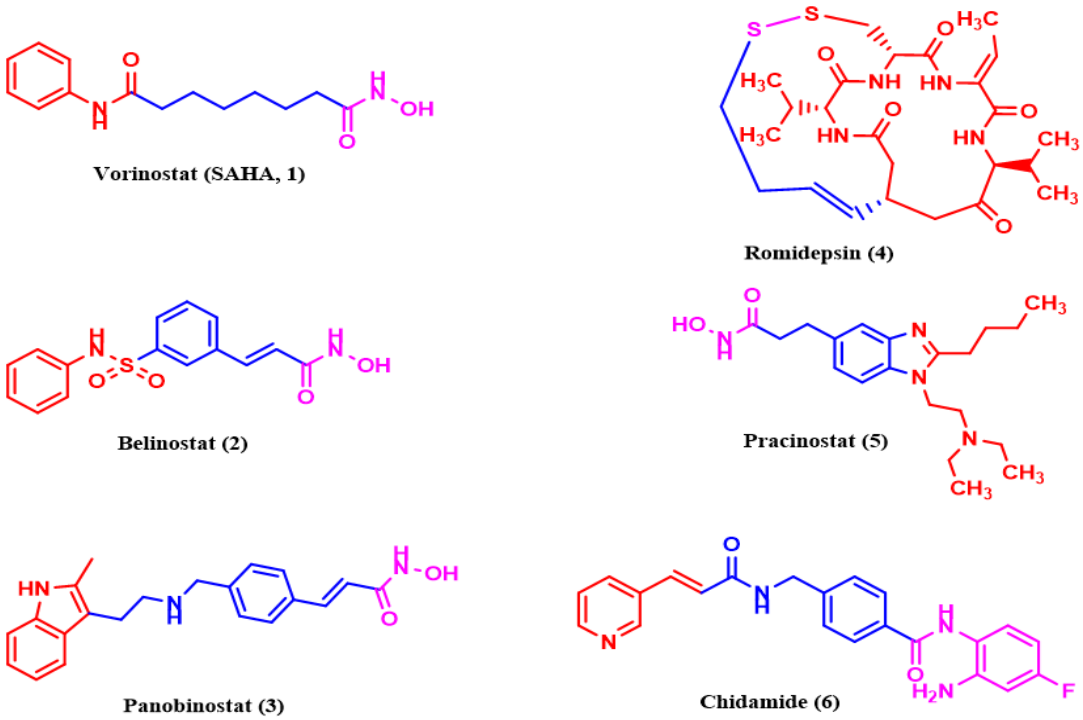


Figure 2.2: FDA approved HDAC inhibitors

The acetylation of lysine 40 on α-tubulin, regulated by HDAC6, is notably the first identified and most extensively researched physiological non-histone substrate.[63] Numerous studies have indicated that HDAC6 plays a role in tumorigenesis, development, and metastasis through diverse cellular pathway involving proteins like tubulin, protein ubiquitination, and Hsp90. Extensive research has also shown that selective inhibition of HDAC6 could be an effective treatment for different cancers such as malignant melanoma, lung cancer, and bladder cancer, as well as for neurodegenerative diseases like Alzheimer's disease (AD), Parkinson's disease (PD), and Huntington's disease.[63,67] Recent reports also demonstrated the potential of selective HDAC6 inhibitors in rare disease conditions such as amyotrophic lateral sclerosis (ALS), Rett syndrome, and Charcot-Marie-Tooth disease.[68] Later in this chapter, there is a detailed discussion on this disease and the role of HDAC6 in their development.

2.2 Classification of HDACs

As stated earlier, the 18 HDACs are categorised into class I, II, III and IV based on their similarity to yeast orthologues Rpd3, HdaI, and Sir2, respectively. Classes I, II and IV are referred to as classical HDACs, comprising 11 family members that are generally Zn^{2+} dependent metalloenzymes. In contrast, class III members are known as sirtuins

and require NAD+ as a vital cofactor.[60,69] Table 1.1 summarises all the HDAC isoforms, their chromosomal location, cellular localization and functions.

2.2.1 Class I HDACs

Class I HDAC family comprises of four members namely, HDAC1, HDAC2, HDAC3 and HDAC8, of these HDAC1, HDAC2, and HDAC3 functions together as subunits of multiprotein nuclear complex as a transcriptional repression factor.[69] HDAC1 and HDAC2 have very similar structure (sequence similarity index 82%) and mainly works as subunits of corset complex which upon activation represses the expression of neuronal gene in non-neuronal tissue. HDAC3 also functions as a transcription repressor. The last member of this family, i.e., HDAC8 though has about 34% sequence similarity with HDAC3, is not a component of any repressor complex so far.[60]

2.2.2 Class II HDACs

Class II HDAC family is further divided into two sub classes, Class IIa and Class IIb. Class IIa consists of four members namely HDAC4, HDAC5, HDAC7 and HDAC9 and are characterised by large N-terminal domain with a specific binding site for transcription factors like monocyte-specific enhancer factor-2 (MEF-2) and 14-3-3 protein that facilitates HDAC signalling.[60] These Class IIa features three of these 14-3-3 binding sites, and are majorly responsible for regulating cellular HDAC trafficking.[60,69] 14-3-3 upon activation stimulates the cytoplasmic retention or nuclear retention of class IIa HDACs using a phosphorylation-dependent kinase, like calcium/calmodulin-dependent protein kinase (CaMK) and protein kinase D (PKD), microtubule affinity-regulating kinase (MARK) and checkpoint kinase-1 (CHK-1).[70]

Class IIb members i.e., HDAC6 and HDAC10 are similar to each other with 55% sequence homology. HDAC6 is mainly cytoplasmic and deacetylase non-histone substrates such as α-tubulin, heat shock protein etc.[70] This particular isoform of HDAC contains two tandem deacetylase domain and a C-terminal zinc finger ubiquitin binding domain,[67] and is associated with cell mortality, adhesion and chaperone function. HDAC6 also regulates aggresome function, autophagy through its zinc finger ubiquitin binding domain.[69] Like other HDACs, HDAC10 also features an N-terminal catalytic domain and a C-terminal leucine-rich domain. The N-terminal catalytic domain of HDAC10 resembles the deacetylase domain found in other class II HDACs, whereas the C-terminal domain lacks the residues essential for enzymatic activity.[71]

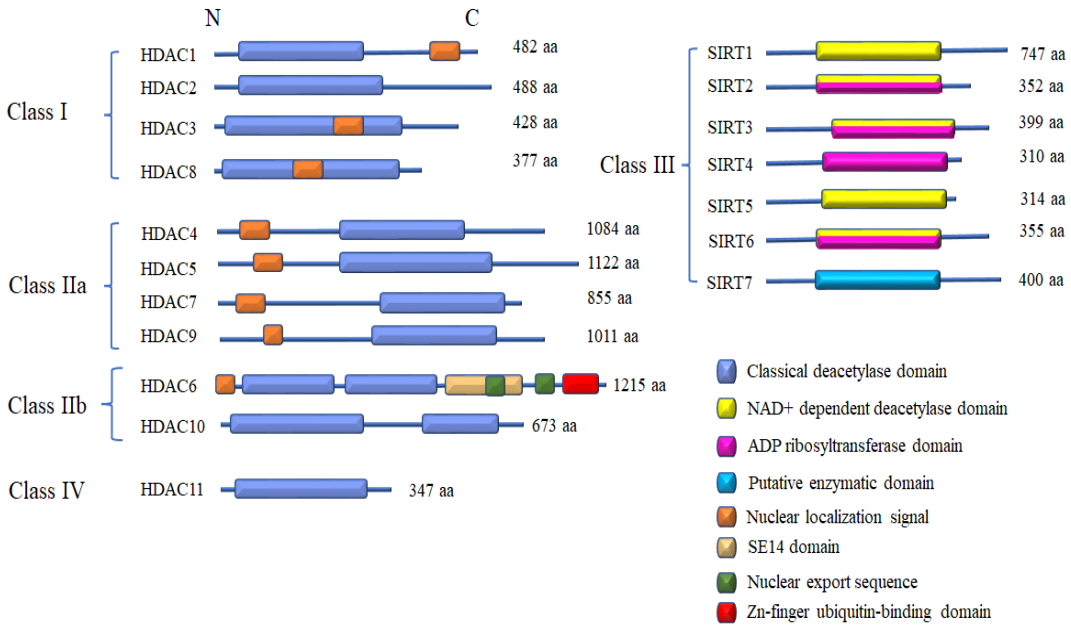


Figure 2.3: Domain organization of human HDACs. The total number of amino acid residues is shown on the right of each isoform. Enzymatic domains (or putative enzymatic domains) are shown in colour.

2.2.3 Class III HDACs

Recently discovered class III HDACs are adenine dinucleotide dependent enzymes, better known as sirtuins due to their sequence homology with yeast Sir2 silencing protein.[61] This class consists of seven members (namely SIRT1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6, SIRT7), and all have specific functions and are associated with various age-related neurodegenerative diseases like Alzheimer's and Parkinson's disease.[72] This class of HDACs are also involved in various physiological pathway such as mitochondrial dysfunction, stress response, oxidative stress, and inflammatory processes associated with neurodegeneration.[72]

2.2.4 Class IV HDACs

The sole member of this family, i.e., HDAC11 is the smallest known HDAC enzyme with a molecular weight of 39 kilodalton, characterised by an open reading frame responsible for encoding a 347-residue protein.[73] HDAC11 features a deacetylase domain, structurally similar to both class I and class II HDADCs.[73] HDAC11 is highly expressed is the brain, heart, kidney, muscles and testes, though least is known about its physiochemical function.[73,74]

2.3 Structural biology of HDAC6

In 1999, Verdel and Khochbin et al.[75] along with Grozinger et al.[76] identified HDAC6 following a GenBank database search for human HDACs that exhibited sequence similarity to yeast Hda1. The X-linked gene p11.22-23 composed of 21923 base pair encodes for HDAC6, that consists 1215 amino acids.[63] They observed that this tissue specific enzyme is primarily cytoplasmic and has highest expression in the heart, liver, kidney and pancreas.[76] HDAC6 exclusively contains two highly conserved catalytic domains. These domains are homologous to each other and functions independently, contributing to the overall activity of HDAC6. The helixes H17 and H18 of catalytic domain 1 (CD1) and helixes H36 and H37 of catalytic domain 2 connects these homologues domains, and forms a lar domain-domain interface by helices H13, H14, H15, and H18N of CD1 and H32, H33, and H34 of CD2.[77]

Structurally, HDAC6 contains five domain all total. From N to C terminal, the N terminal end is rich in arginine and lysine (A.a: 1-87) and functions as nuclear localization signal, followed by a nuclear export signal (A.a: 67-76) which is leucine enriched.[63,78] The catalytic domain 1 (A.a: 88-447) and catalytic domain 2 (A.a: 482-800) are the center of deacetylase activity and are followed by cytoplasmic retention signal (SE14; A.a: 884-1022) characterised by a tetra decapeptide serine glutamic acid repeat sequence.[63,79] At the C-terminus there is a unique zinc finger ubiquitin binding domain[80] that facilitates misfolded protein degradation through aggresome pathway.[81]

With the help of X-ray crystallography Miyake et al.[77] was able to image catalytic domains of HDAC6 exhibiting different substrate specificity. The catalytic domain 1 (CD1) comes with narrow substrate specificity primarily due to its active site being constricted by K330 making it inaccessible to wider substrate, whereas similar position in CD2 is occupied by relatively smaller amino acid L712 thus allowing it to interact with wider range of substrates.[58] Both of these catalytic domains are highly conserved and features a hydrophobic channel connecting the active sites opening and the zinc at the base of the pocket, made up of residues Pro83, Gly201, Phe202, and Trp261 in CD1, and Pro464, Gly582, Phe583, Phe643, and Leu712 in CD2. The two-charge relay system is made up of His192-Asp228 and His193-Gln235 dyads in CD1. In contrast CD2 has the classis dyad configuration with His573-Asp610 and His574-Asn617.[77] During the catalytic process, the zinc ion in the active site, coordinated by His573,

His574, and Asp612, activates a water molecule. This water molecule, supported by the proton relay system involving Tyr745 and Asp705, attacks the carbonyl carbon of the

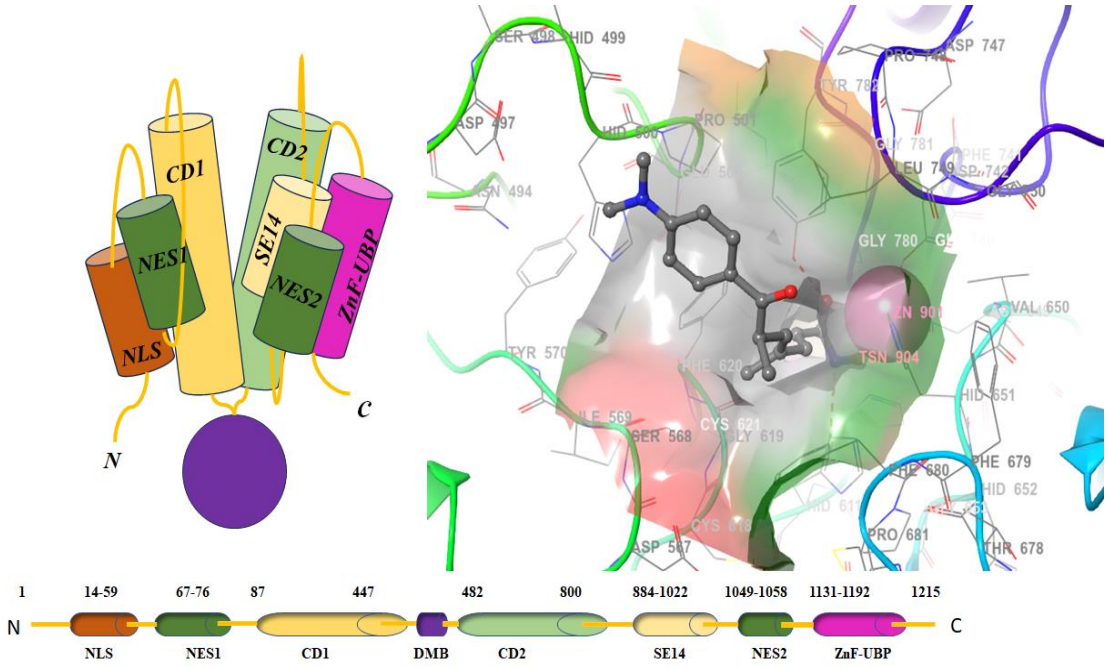


Figure 2.4: Mechanistic representation of HDAC6 (left). Catalytic domain 2 of HDAC6 with Trichostatin A (right).

acetyl group on the substrate's lysine residue, leading to the formation of a tetrahedral intermediate. The collapse of this intermediate results in the release of an acetate molecule and the regeneration of the free lysine residue. Additionally, Phe583 and Trp627 contribute to forming a hydrophobic pocket that stabilizes the substrate within the active site, ensuring proper orientation for the deacetylation reaction.[79] Understanding the role of these specific residues in the coordination and execution of the deacetylation process are essential for developing selective inhibitors for HDAC6.

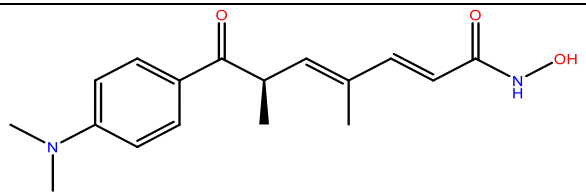
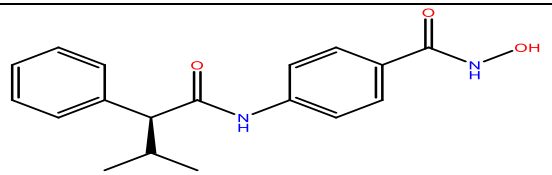
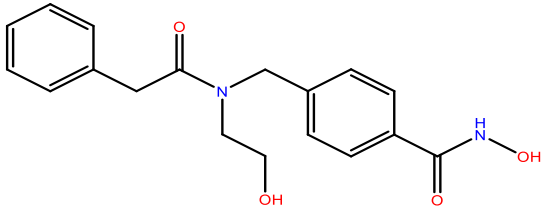
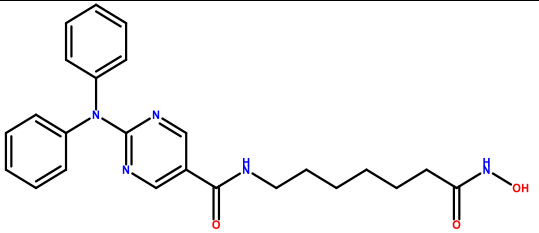
To date, researchers have identified a total of 92 X-ray crystal structures of HDAC6, derived from both Homo sapiens (humans) and Danio rerio (zebrafish). These structural studies have offered invaluable insights into the complex interactions between the ligand (inhibitor) and the receptor (HDAC6). The active sites in the catalytic domain 2 of HDAC6 in humans and zebrafish are nearly identical, with the exception of the N645M and N530D substitutions, respectively. This conservation suggests that the overall mechanism of action is preserved across species, making zebrafish a useful model for studying HDAC6 inhibitors. A comprehensive list of these HDAC6 crystal structures is available in the Protein Data Bank and is detailed in Table 2.2.

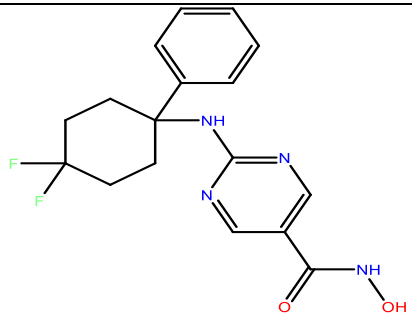
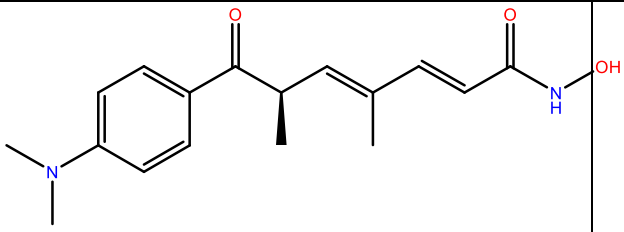
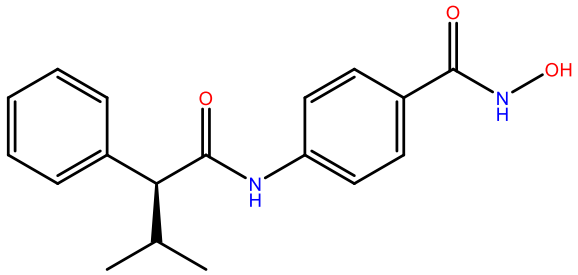
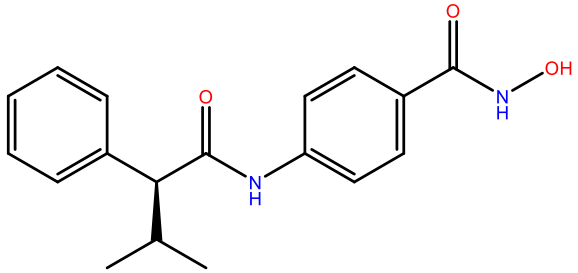
Table 2.1: Classification of HDAC isoforms, their cellular localization and functions

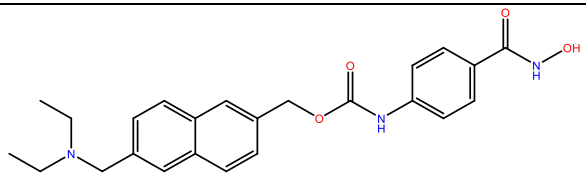
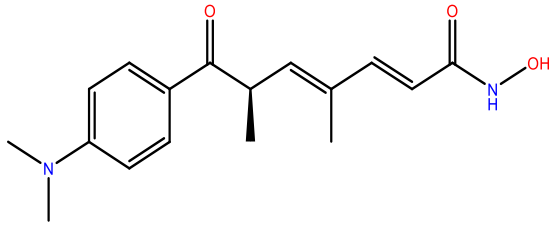
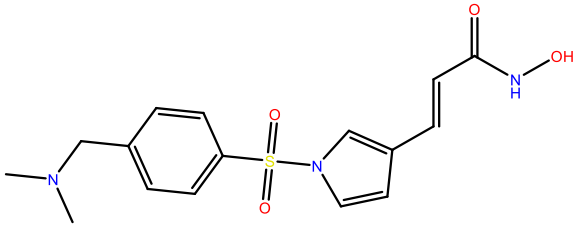
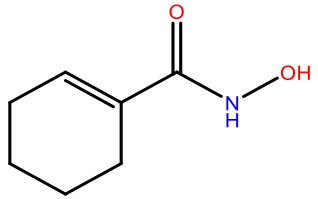
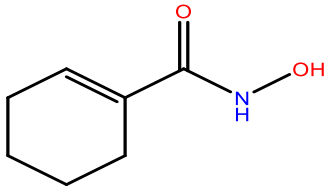
Super family	Family	Group	Class		HDAC isoform	Chromosomal location	Amino acids No.	Cellular localization	Physiological Role
Arginase/ deacetylase superfamily	Histone deacetylase family	Zn^{2+}	I		HDAC1	1p35 – p35.1	483	Nucleus	Proliferation control, apoptosis, transcription regulation, cell survival.
					HDAC2	6q21	488		Proliferation control and apoptosis, transcription repressor.
					HDAC3	5q31.3	428		Proliferation, differentiation, transcriptional repressor, Fox3 deacetylation.
					HDAC8	Xq13.1	377		Proliferation, differentiation, and cell survival.
			II	IIA	HDAC4	2q37.3	1084	Nucleus/ cytoplasmic	Differentiation, angiogenesis, cytoskeletal dynamics, and cell motility.
					HDAC5	17q21.31	1122		Differentiation, lymphocyte activation, endothelial cell function.
					HDAC7	12q13.11	912		Angiogenesis, Lymphocyte activation, thrombocyte differentiation.
					HDAC9	7p21	1069		Deacetylates FoxP3, Immunosuppressive activity.

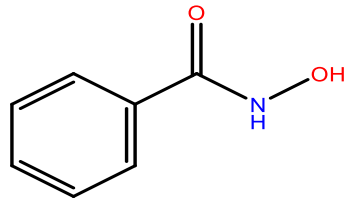
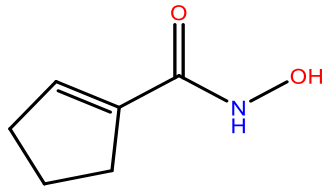
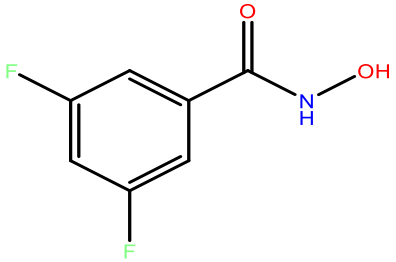
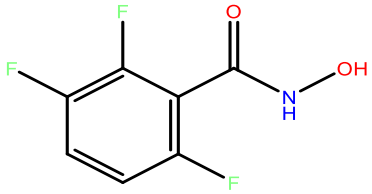
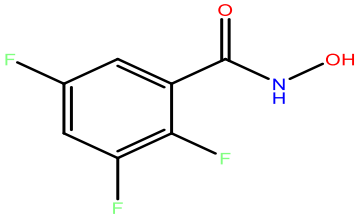
				IIB	HDAC6	Xp11.23	1215	Cytoplasmic	Regulation of protein degradation through aggresome pathway, Hsp90 chaperone activity, cytoskeletal dynamics, cell motility, and angiogenesis.
					HDAC10	2q13.33	669		Angiogenesis.
			IV		HDAC11	3p25.1	347	Nucleus	DNA replication and Immunomodulation by regulating the expression of IL-1.
Deoxyhypusine synthase like NAD/FAD-binding domain superfamily	Sir 2 regulator family	NAD^+	III		SIRT1	10q21.3	747	Nucleus/ cytoplasmic	DNA repair, cell survival, autoimmunity.
					SIRT2	19q13.2	389	Nucleus	DNA repair, cell survival, cell invasion.
					SIRT3	11p15.5	399	Mitochondrial	DNA repair, cell signaling apoptosis.
					SIRT4	12q24.31	314		Energy metabolism.
					SIRT5	6p23	310		Cell signaling pathways.
					SIRT6	19p13.3	355	Nucleus	DNA repair, metabolism regulation
					SIRT7	17q25.3	400		Apoptosis, cellular transformation.

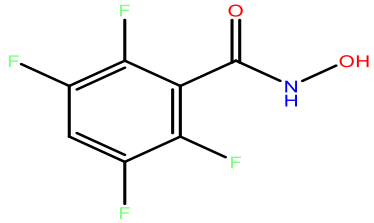
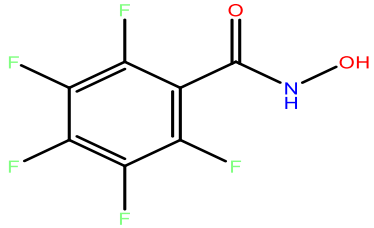
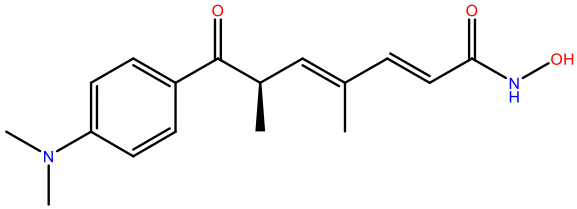
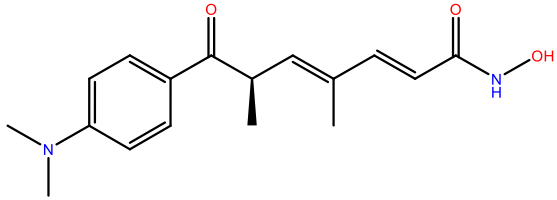
Table 2.2: List of reported crystal structures of HDAC6 as available from Protein Data Bank (as accessed on September, 2023)

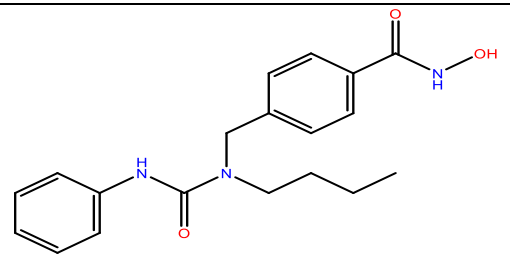
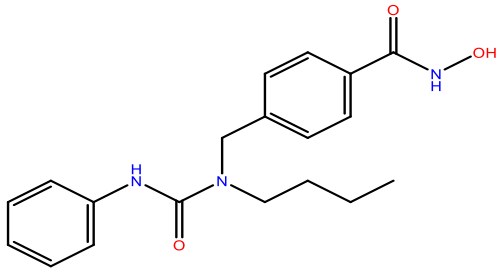
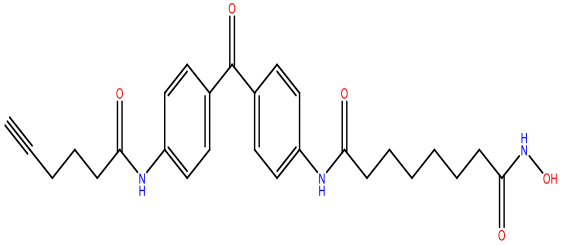
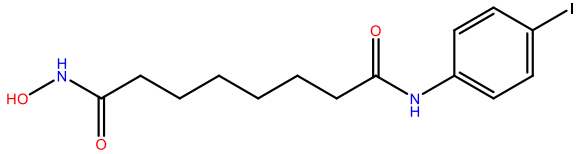
SL. No	PDB	Domain	Mutation	Co-crystalized with	Organism	Release	XRD Resolution	Ligand Structure	Ref
1	6UO2	CD1	-	Trichostatin A	*Danio rerio*	04-12-2019	1.65		82
2	6UO3	CD1	-	AR-42	*Danio rerio*	04-12-2019	1.09		82
3	5WGK	CD2	-	HPB	*Danio rerio*	06-12-2017	1.822		83
4	5WGL	CD2	-	ACY-1215	*Danio rerio*	06-12-2017	1.7		83

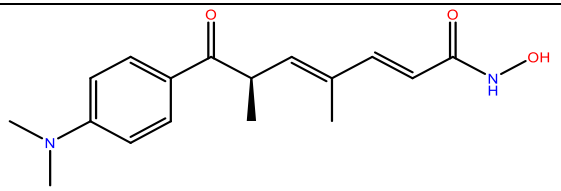
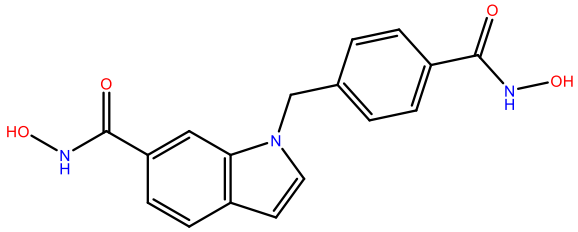
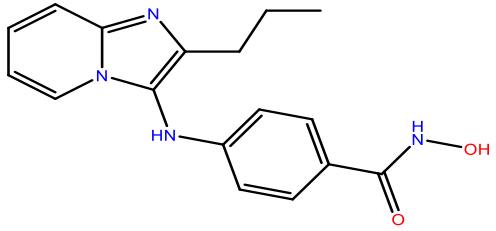
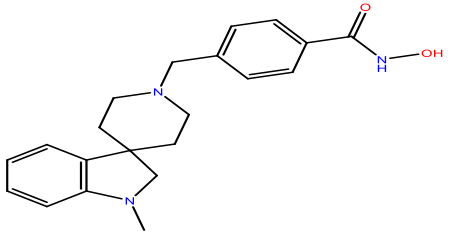
5	5WGM	CD2	-	ACY-1083	*Danio rerio*	06-12-2017	1.75		83
6	6UO4	CD1	Y363F	Trichostatin A	*Danio rerio*	04-12-2019	1.268		82
7	6UO5	CD1	Y363F	AR-42	*Danio rerio*	04-12-2019	1.439		82
8	6UO7	CD1	K330L	AR-42	*Danio rerio*	04-12-2019	1.395		82

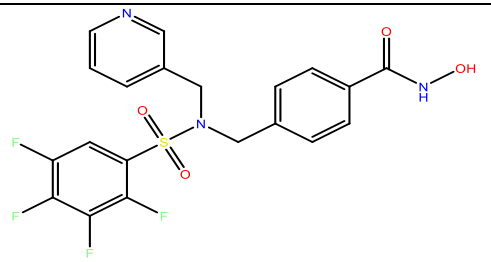
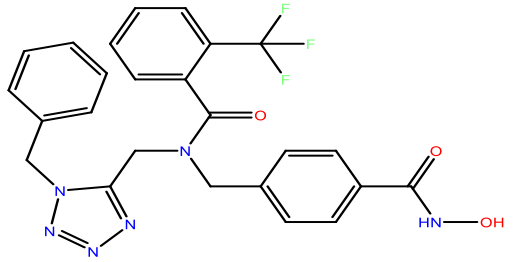
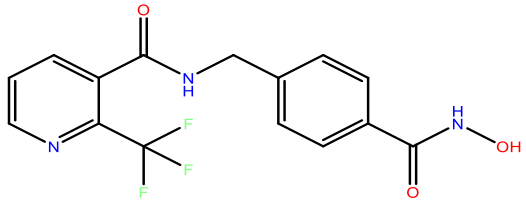
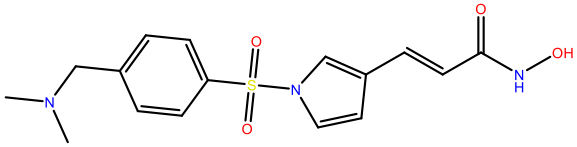
9	6UOC	CD1	K330L	Givinostat	*Danio rerio*	04-12-2019	1.4		82
10	5WGI	CD2	-	Trichostatin A	*Danio rerio*	06-12-2017	1.05		83
11	6UOB	CD1	K330L	Resminostat	*Danio rerio*	04-12-2019	1.58		82
12	6CSP	CD2	-	N-hydroxycyclohex-1-ene-1-carboxamide	*Danio rerio*	30-05-2018	1.237		84
13	6CSQ	CD2	-	N-hydroxycyclohex-1-ene-1-carboxamide	*Danio rerio*	30-05-2018	2.031		84

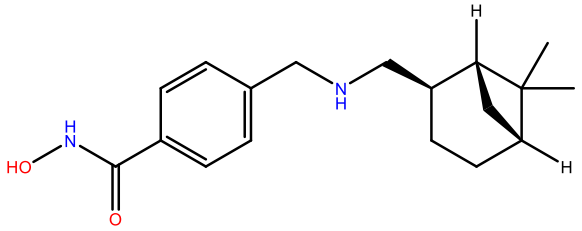
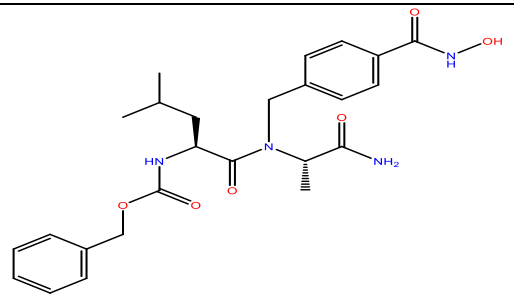
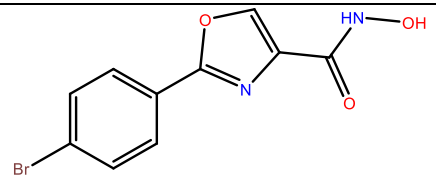
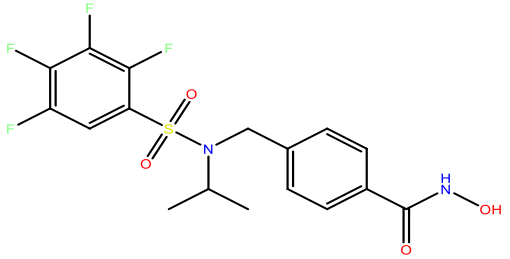
14	6CSR	CD2	-	Phenylhydroxamate	*Danio rerio*	30-05-2018	1.619		84
15	6CSS	CD2	-	Cyclopentenylhydroxamate	*Danio rerio*	30-05-2018	1.7		84
16	8D98	CD2	-	3,5-difluoro-N-hydroxybenzamide	*Danio rerio*	28-09-2022	1.66		85
17	8D99	CD2	-	2,3,6-trifluoro-N-hydroxybenzamide	*Danio rerio*	28-09-2022	1.79		85
18	8D9A	CD2	-	2,3,5-trifluoro-N-hydroxybenzamide	*Danio rerio*	28-09-2022	1.75		85

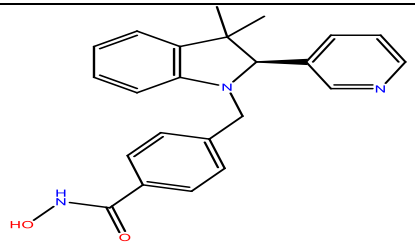
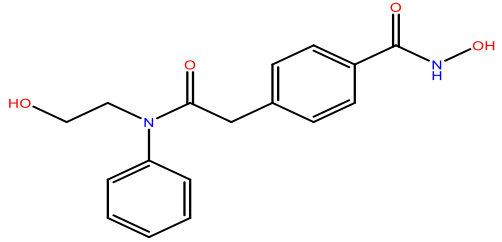
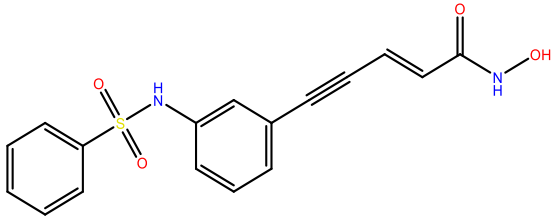
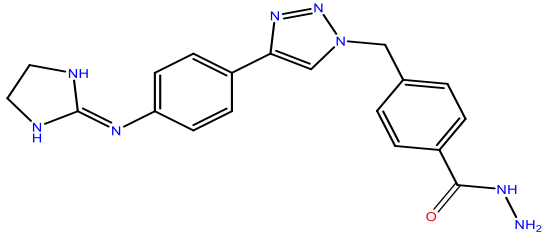
19	8D9B	CD2	-	2,3,5,6-tetrafluoro-N-hydroxybenzamide	*Danio rerio*	28-09-2022	1.63		85
20	8D9C	CD2	-	2,3,4,5,6-pentafluoro-N-hydroxybenzamide	*Danio rerio*	28-09-2022	1.82		85
21	8EQI	CD2	-	Cyclopeptide des4.2.0	*Danio rerio*	19-04-2023	2	-----	86
22	5G0G	CD1	-	Trichostatin A	*Danio rerio*	27-07-2016	1.499		77
23	5G0H	CD2	-	(S)-Trichostatin A	*Danio rerio*	27-07-2016	1.6		77

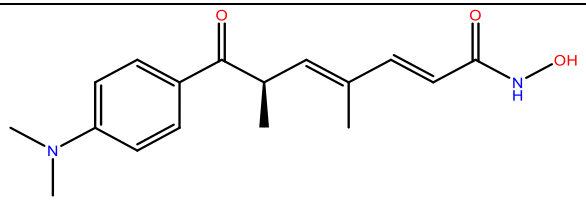
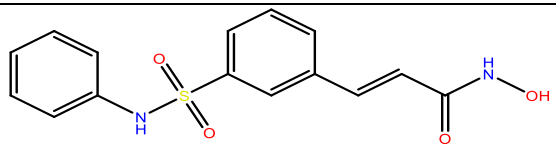
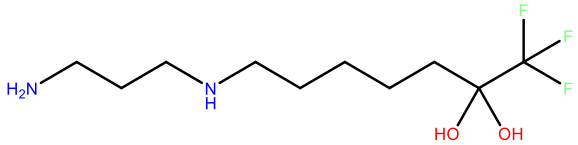
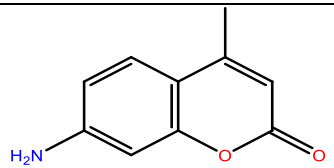
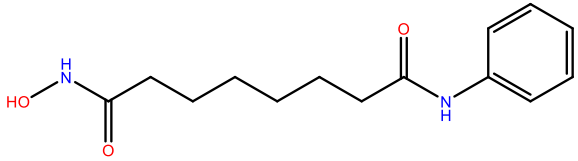
24	5G0I	CD1 & CD2 (Linker cleaved)	-	Nexturastat A	*Danio rerio*	27-07-2016	1.99		77
25	5G0J	CD1 & CD2 (Linker intact)	-	Nexturastat A	*Danio rerio*	27-07-2016	2.88		77
26	6WYP	CD1	K330L	SAHA-Bpyne	*Danio rerio*	02-09-2020	2.4		87
27	6WYQ	CD1	K330L	4-iodo-SAHA	*Danio rerio*	02-09-2020	1.9		87
28	5G0F	ZnF-UBP	-	na	*Danio rerio*	27-07-2016	1.9	------	77

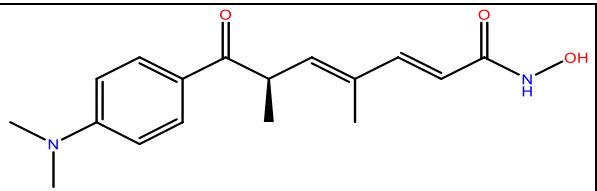
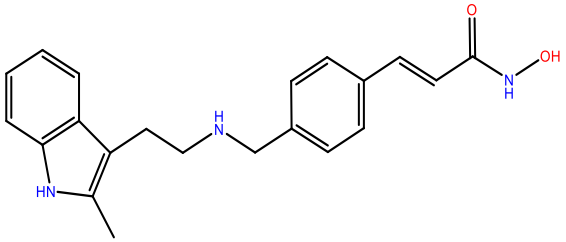
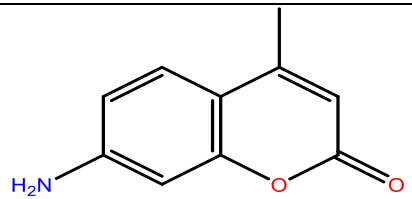
29	6WYO	CD1	H82F, F202Y	Trichostatin A	*Danio rerio*	02-09-2020	2.3		87
30	6VNR	CD2	-	N-hydroxy-1-{[4-(hydroxycarbamoyl)phenyl]methyl}-1H-indole-6-carboxamide	*Danio rerio*	13-05-2020	1.943		88
31	6CGP	CD2	-	MAIP-032	*Danio rerio*	13-06-2018	2.5		89
32	6V7A	CD2	-	NF2657	*Danio rerio*	02-12-2020	2.08		90

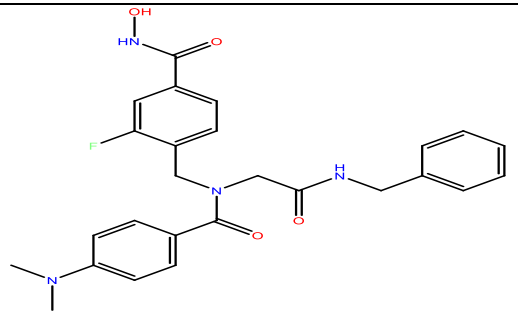
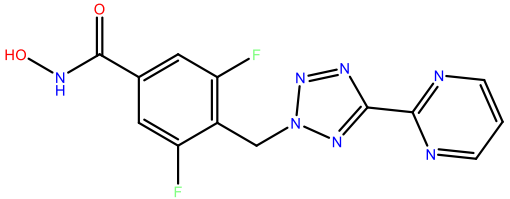
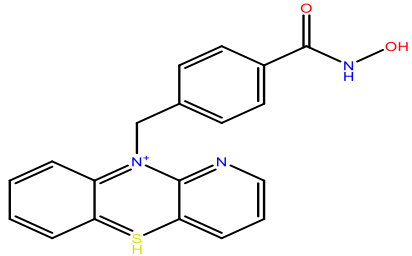
33	7JOM	CD2	-	TO-317	*Danio rerio*	16-06-2021	1.84		91
34	6PYE	CD2	-	NR160	*Danio rerio*	29-07-2020	1.48		92
35	7QNO	CD1 & CD2	-	na	*Danio rerio*	09-02-2022	2.38	-----	93
36	6PZO	CD2	-	YX-153	*Danio rerio*	02-05-2022	1.5		92
37	6PZR	CD2	-	Resminostat	*Danio rerio*	02-05-2022	2.3		92

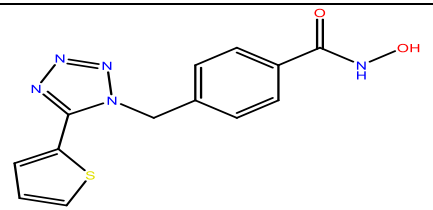
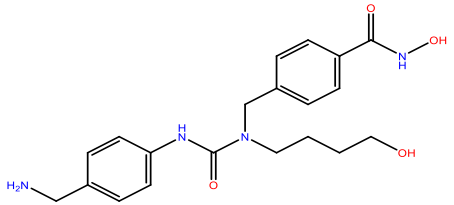
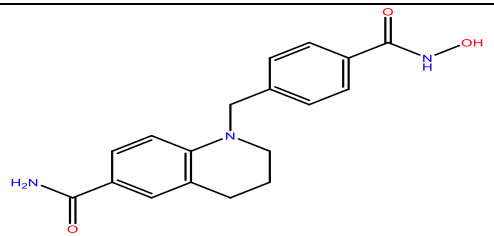
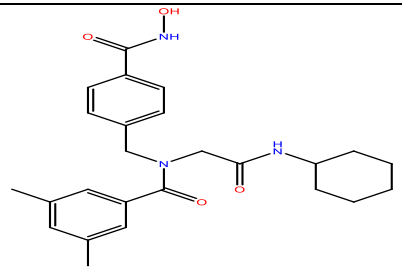
38	6PZS	CD2	-	JR005	*Danio rerio*	02-05-2022	1.79		92
39	6PZU	CD2	-	AP-1-62-A	*Danio rerio*	02-05-2022	1.74		92
40	6Q0Z	CD2	-	JS28	*Danio rerio*	02-05-2022	1.75		92
41	7UK2	CD2	-	NN-390	*Danio rerio*	02-11-2011	1.6		94

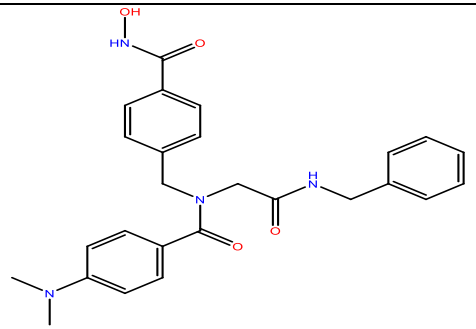
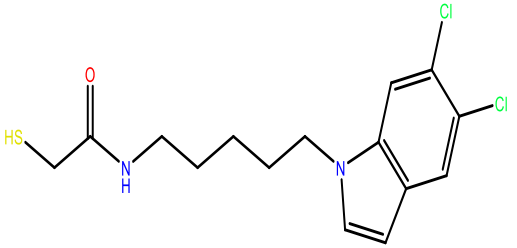
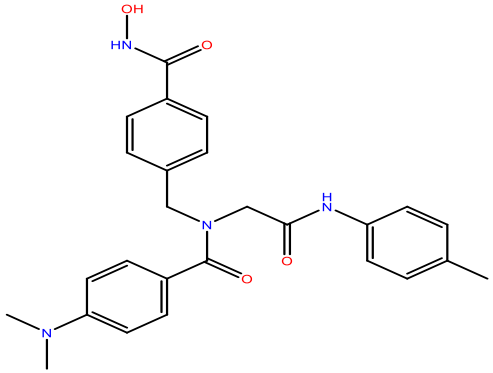
42	6V79	CD2	-	NF2376	*Danio rerio*	02-12-2020	2.039		95
43	5EF7	CD2	-	HPOB	*Danio rerio*	27-07-2016	1.9		96
44	5EFB	CD2	-	Oxamflatin	*Danio rerio*	27-07-2016	2.543		96
45	5EFG	CD2	-	na	*Danio rerio*	27-07-2016	2.25	------	96
46	8A8Z	CD2	-	ITF5924	*Danio rerio*	25-01-2023	1.6		97

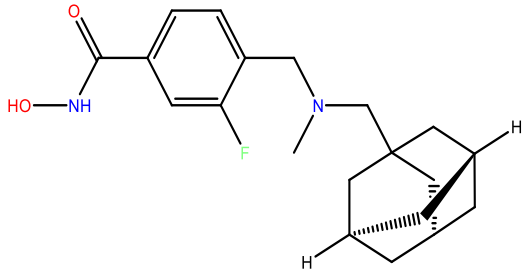
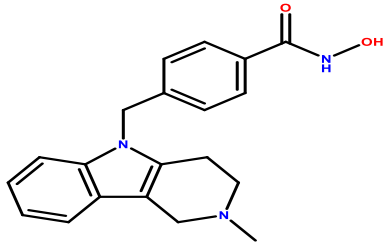
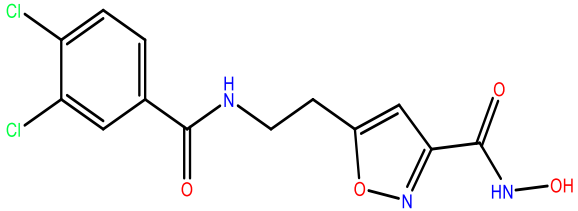
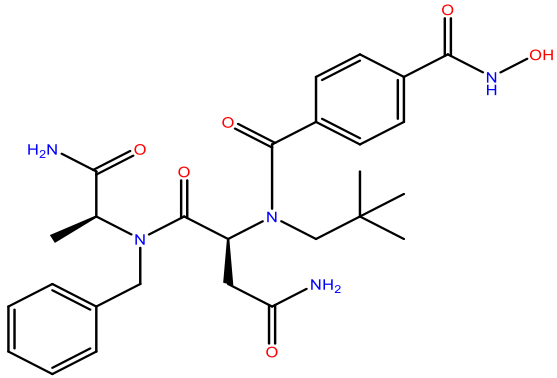
47	5EEF	CD1	-	Trichostatin A	*Danio rerio*	27-07-2016	2.151		96
48	5EEM	CD2	-	na	*Danio rerio*	27-07-2016	2	-----	96
49	5EEN	CD2	-	Belinostat	*Danio rerio*	27-07-2016	1.861		96
50	5EFH	CD2	-	7-[(3-aminopropyl)amino]-1,1,1-trifluoroheptane-2,2-diol	*Danio rerio*	27-07-2016	2.162		96
51	5EFN	CD2	H574A	7-amino-4-methyl-chromen-2-one	*Danio rerio, Homo sapiens*	27-07-2016	1.804		96
52	5EEI	CD2	-	SAHA	*Danio rerio*	27-07-2016	1.32		96

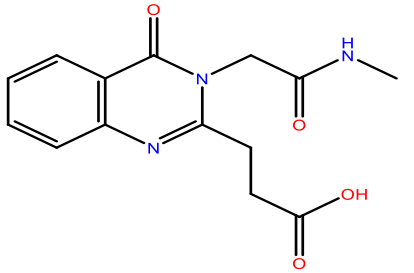
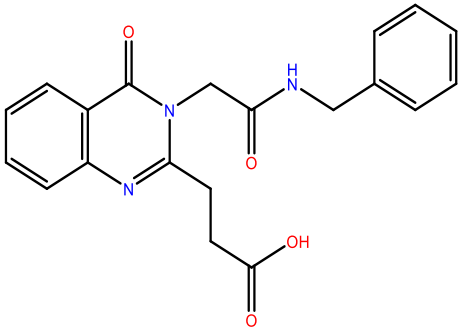
53	5EEK	CD2	-	Trichostatin A	*Danio rerio*	27-07-2016	1.59		96
54	5EF8	CD2	-	Panobinostat	*Danio rerio*	27-07-2016	2.6		96
55	5EFJ	CD2	-	-	*Danio rerio, Bipolaris zeicola*	27-07-2016	1.73	-----	96
56	5EFK	CD2	Y745F	7-amino-4-methyl-chromen-2-one	*Danio rerio*	27-07-2016	1.82		96

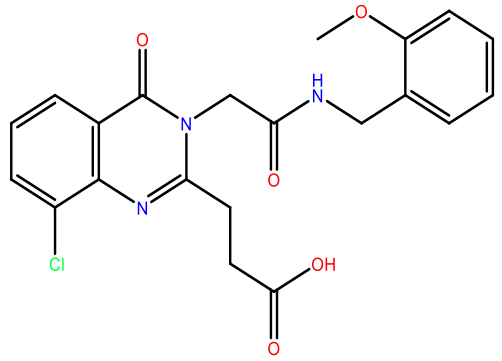
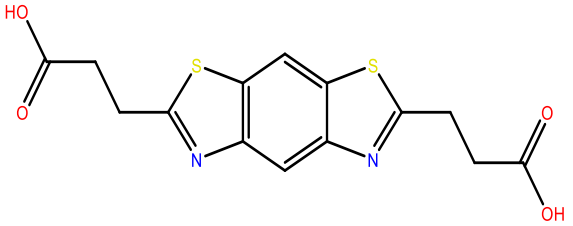
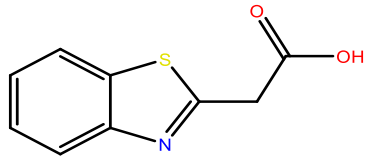
57	7U8Z	CD2	-	4-({N-[2-(benzylamino)-2-oxoethyl]-4-(dimethylamino) benzamido}methyl)-3-fluoro-N-hydroxybenzamide	*Danio rerio*	23-11-2022	1.85		98
58	7O2R	CD2	-	ITF3985	*Danio rerio*	27-10-2021	2.3		98
59	5W5K	CD2	-	KV70	*Danio rerio*	27-06-2018	2.7		99
60	6WSJ	CD2	-	Cyclopeptide des4.3.1	*Danio rerio*	28-04-2021	1.7	-------	100

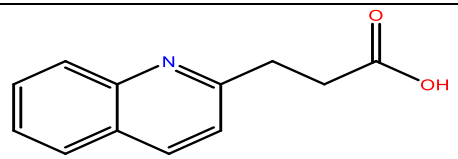
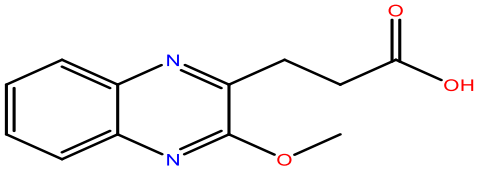
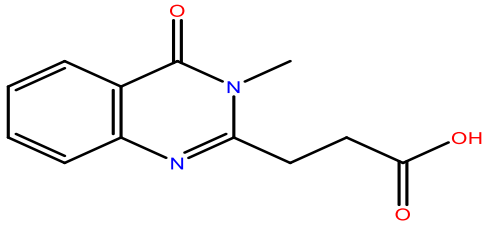
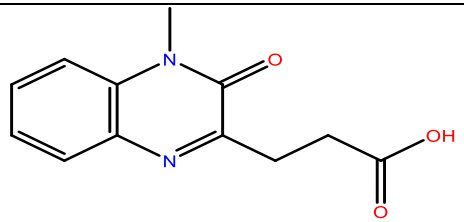
61	7O2P	CD2	-	ITF3756	*Danio rerio*	27-10-2021	1.9		98
62	6TCY	CD2	-	SS555	*Danio rerio*	04-11-2020	1.6		101
63	6ZW1	CD2	-	SW101	*Danio rerio*	04-08-2021	1.13		102
64	6DVL	CD2	-	DDK-115	*Danio rerio*	29-08-2018	2.1		103

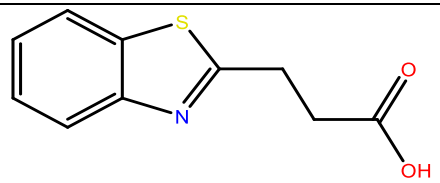
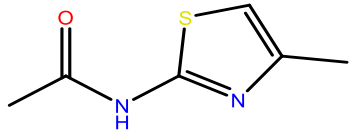
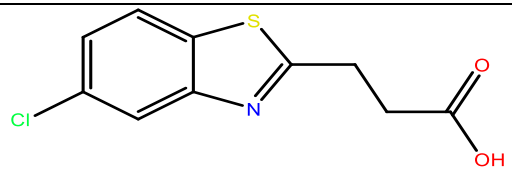
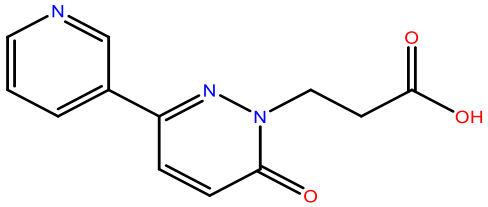
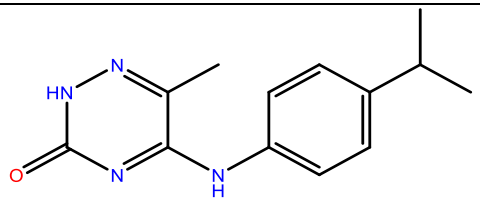
65	6DVN	CD2	-	DDK-137	*Danio rerio*	29-08-2018	2.2		103
66	6MR5	CD2	-	N-[5-(5,6-dichloro-1H-indol-1-yl)pentyl]-2-sulfanylacetamide	*Danio rerio*	05-12-2018	1.85		104
67	6DVM	CD2	-	DDK-122	*Danio rerio*	29-08-2018	1.47		103

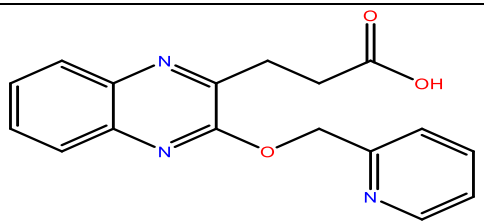
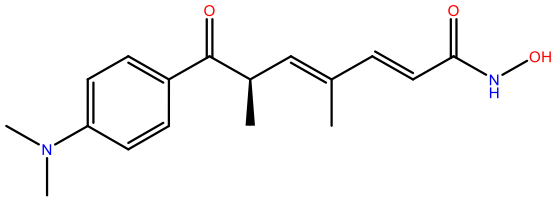
68	6DVO	CD2	-	Bavarostat	*Danio rerio*	29-08-2018	1.98		103
69	6THV	CD2	-	Tubastatin A	*Danio rerio*	15-07-2020	1.1		105
70	6R0K	CD2	-	SS208	*Danio rerio*	09-10-2019	1.15		106
71	6CW8	CD2	-	RTS-V5	*Danio rerio*	21-11-2018	1.9		107

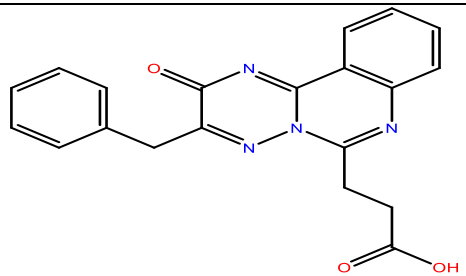
72	8G43	ZnF-UBP	-	3-(3-(2-(methylamino)2-oxoethyl)-4-oxo-3,4-dihydroquinazolin-2-yl)propanoic acid	*Homo sapiens*	03-05-2023	1.55		108
73	8G44	ZnF-UBP	-	3-(3-(2-(benzylamino)-2-oxoethyl)-4-oxo-3,4-dihydroquinazolin-2-yl)propanoic acid	*Homo sapiens*	03-05-2023	1.55		108

74	8G45	ZnF-UBP	-	3-[8-chloro-3-(2-{[(2-methoxyphenyl)methyl]amino}-2-oxoethyl)-4-oxo-3,4-dihydroquinazolin-2-yl]propanoic acid	*Homo sapiens*	03-05-2023	1.62		108
75	6CE6	ZnF-UBP	-	3,3'-(benzo[1,2-d:5,4-d']bis(thiazole)-2,6-diyl)dipropionic acid	*Homo sapiens*	28-02-2018	1.6		109
76	6CE8	ZnF-UBP	-	(1,3-benzothiazol-2-yl)acetic acid	*Homo sapiens*	28-02-2018	1.55		109

77	6CEA	ZnF-UBP	-	3-(quinolin-2-yl)propanoic acid	*Homo sapiens*	28-02-2018	1.6		109
78	6CEC	ZnF-UBP	-	3-(3-methoxy-2-quinoxalinyl)propanoic acid	*Homo sapiens*	28-02-2018	1.55		109
79	6CED	ZnF-UBP	-	3-(3-methyl-4-oxo-3,4-dihydroquinazolin-2-yl)propanoic acid	*Homo sapiens*	28-02-2018	1.7		109
80	6CEE	ZnF-UBP	-	3-(1-methyl-2-oxo-1,2-dihydroquinozalin-3-yl)propionic acid	*Homo sapiens*	28-02-2018	1.55		109

81	6CEF	ZnF-UBP	-	3-(1,3-benzothiazol-2-yl)propanoic acid	*Homo sapiens*	28-02-2018	1.8		109
82	5B8D	ZnF-UBP	-	N-(4-methyl-1,3-thiazol-2-yl)propanamide	*Homo sapiens*	27-07-2016	1.05		110
83	5KH3	ZnF-UBP	-	3-(5-chloro-1,3-benzothiazol-2-yl)propanoic acid	*Homo sapiens*	27-07-2016	1.6		110
84	5KH7	ZnF-UBP	-	3-[6-oxo-3-(3-pyridinyl)-1(6H)-pyridazinyl]propanoic acid	*Homo sapiens*	27-07-2016	1.7		110
85	5KH9	ZnF-UBP	-	5-[(4-isopropylphenyl)amino]-6-methyl-1,2,4-	*Homo sapiens*	27-07-2016	1.07		110

				triazin-3(2H)-one					
86	5WPB	ZnF-UBP	-	3-(3-(pyridin-2-ylmethoxy)quinoxalin-2-yl)propanoic acid	*Homo sapiens*	23-08-2017	1.55		110
87	3PHD	Complete structure	-	Ubiquitin	*Homo sapiens*	23-02-2011	3	-----	111
88	3CK5	ZnF-UBP	-	na	*Homo sapiens*	19-02-2008	1.55	-----	--
89	3GV4	ZnF-UBP	-	Ubiquitin C-terminal peptide RLRGG	*Homo sapiens*	28-04-2009	1.72	-----	--
90	5EDU	CD2	-	Trichostatin A	*Homo sapiens, Escherichia coli*	27-07-2016	2.79		96

91	5WBN	ZnF-UBP	-	3-(3-benzyl-2-oxo-2H-[1,2,4]triazino[2,3-c]quinazolin-6-yl)propanoic acid	*Homo sapiens*	02-08-2017	1.64		--
92	7ZYU	ZnF-UBP	-	DARPin (Designed Ankyrin repeat protein) F10	*Homo sapiens*	01-06-2022	2.43		112

2.4 Physiological functions of HDAC6

HDAC6 is a major regulator of cellular proliferation, apoptosis, cellular mortality, cellular oxidative stress pathway, misfolded protein degradation and heat shock response, in other word it's a major role player in maintaining cellular homeostasis. Histone acetylation and deacetylation are essential for genetic expression and are facilitated by histone acetyltransferase and histone deacetylase. As HDAC6 is cytoplasmic deacetylase, it primarily maintains acetylation balance of cytosolic non-histone proteins such as α-tubulin, cortactin, Hsp90, peroxiredoxins, surviving, Miro-1, ERK-1, HSF-1, Ku-70, etc.[113,114] various HDAC6 substrate their functions and its related disease conditions have been listed in table 2.3.

HDAC6 by affecting cytoskeleton (a cross-linking network consisting of microtubules, actine filaments etc, provides structural support to the cell) dynamics influences cell division, migration, angiogenesis and aggresome formation. α-tubulin is the first described and most studied substrate of HDAC6 and happens to be a building block of microtubules.[79] Hypoacetylation of tubulin promotes cell migration whereas hyperacetylation results in excessive accumulation of focal adhesion, thereby causing cellular adhesion and its acetylation state overseen by opposing action of α-tubulin acetyltransferase and HDAC6.

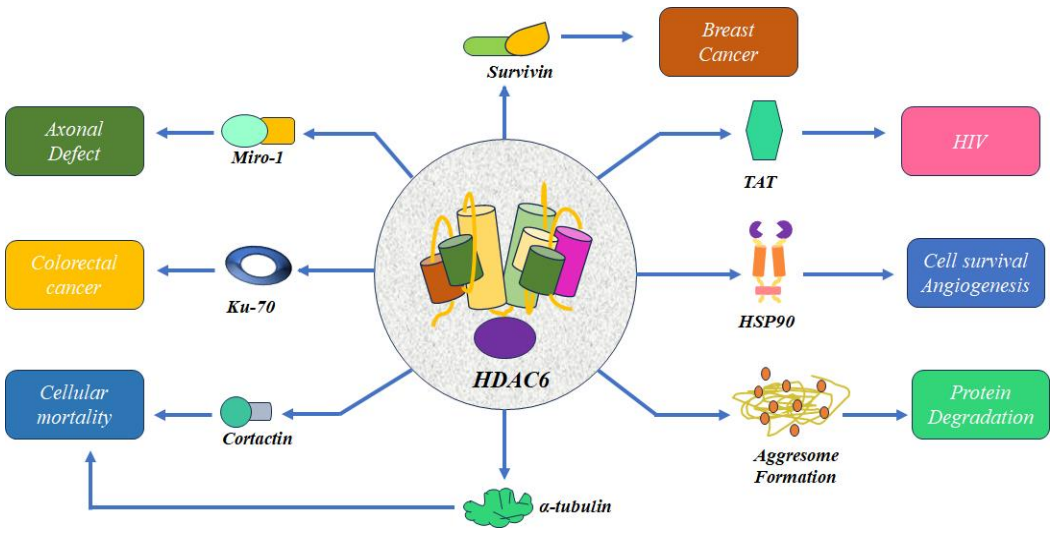


Figure 2.5: Physiological role of HDAC6

HDAC6 also regulates actin-dependent cell mortality by altering acetylation state of cortactin.[115] HDAC6 has also been found to influence PD-L1 expression via the STAT3 signalling pathway, suggesting its role in immunoregulation.[67]

Table 2.3: HDAC6 substrates, interacting proteins, and relevant biological functions

Substrate	Localization of substrate	Lysine residue deacetylated	Catalytic domain involved	HDAC6 function	Potential therapeutic target disease
α-tubulin	Cytoplasm	Lys 40	CD1 or CD2	Regulation of immune synapse formation, cell migration and chemotaxis, microtube dynamics.	Antigen presentation deficiencies, tumor cell metastasis (cancer), neurodegenerative disorders (Parkinson's disease, spinobulbar muscular atrophy, CMT disease)
Cortactin	Cytoplasm	Lys 87, 124, 161, 189, 198, 235, 272, 309, 319	CD1 + CD2	Regulation of cellular migration and F-actin binding	Cell migration and adhesion in cancer.
HSP90	Cytoplasm	Lys 294	CD1 + CD2 + BUZ	Misfolded protein degradation and clearance and regulation of glucocorticoid receptor and gene transcription activation.	Parkinson's disease, Alzheimer's disease, and cancer.
Miro-1	Mitochondria	Lys 105		Blocks mitochondrial transport and mediates axonal growth inhibition.	Axonal defect in CMT 2.

Peroxiredoxins	Cytoplasm and nucleus	Lys 196, 197	ND	Redox regulation	Neurodegenerative disorders and cancer.
Survivin	Nucleus	Lys 129	CD2	Anti-apoptotic function.	Breast cancer.
Ku-70	Cytoplasm	Lys 539, 542	ND	Suppression of apoptosis	Colorectal cancer.
Tat	Nucleus	Lys 28	CD2 + BUZ	Suppression of Tat-mediated transactivation of HIV	HIV
β-catenin	Cytoplasm	Lys 49	ND	Epidermal growth factor-induced β-catenin nuclear localization	Tumour cell (cancer)
ERK 1	Cytoplasm	Lys 72		Cell proliferation and growth, cell mobility and survival.	Cancer.
GSK3β	Cytoplasm and Nucleus	Ser 22	ND	GSK3β phosphorylates HDAC6 (Ser 22) to enhance tubulin deacetylase activity.	Neurodegenerative disorders
Aurora A	Centrosomes of interphase cell (cytoplasm)	----	ND	Aur A phosphorylates HDAC6 to activate tubulin deacetylase activity	Polycystic kidney disease and colorectal cancer.
CK2	Nucleus and cytoplasm	Ser 485	ND	CK2 phosphorylates HDAC6 (Ser458) to	Neurodegenerative disorders.

				increase tubulin deacetylase activity.	
Ubiquitin	Nucleus and cytoplasm	----	BUZ	Signal for cellular processes, such as protein degradation and endocytosis	Neurodegenerative disorders.
TRIM50	Cytoplasm	----	ND	E3 ubiquitin ligase in aggresome formation and protein degradation	Neurodegenerative diseases
Dynein	Cytoplasm	----	DMB	Aggresome formation and protein degradation	Neurodegenerative disease.
LooR	Nucleus	----	ND	Cofactor of nuclear receptor corepressor LCoR	ER-positive breast cancer
NFκB	cytoplasm	----	ND	Transcription factor in inflammation and cell growth control	Gene expression-related deficiencies.
BRMS1	Nucleus and cytoplasm	----	ND	Decrease metastasis suppressor activity	Cancer cell metastasis
Bax	Cytoplasm	----	ND	Pro-apoptotic protein block of apoptosis	Neuroblastoma

* *CD1= Catalytic domain 1, CD2= Catalytic domain 2, BUZ= ubiquitin binding domain, DMB= Dynein motor binding domain, ND = not determined*

Newley synthesised protein chain needs to be folded into tree-dimensional structure prior they can exert their biological function, and error in this folding leads to misfolded proteins. These misfolded proteins need to be degraded that they can be further recycled, and HDAC6 dose this job majorly in three ways.

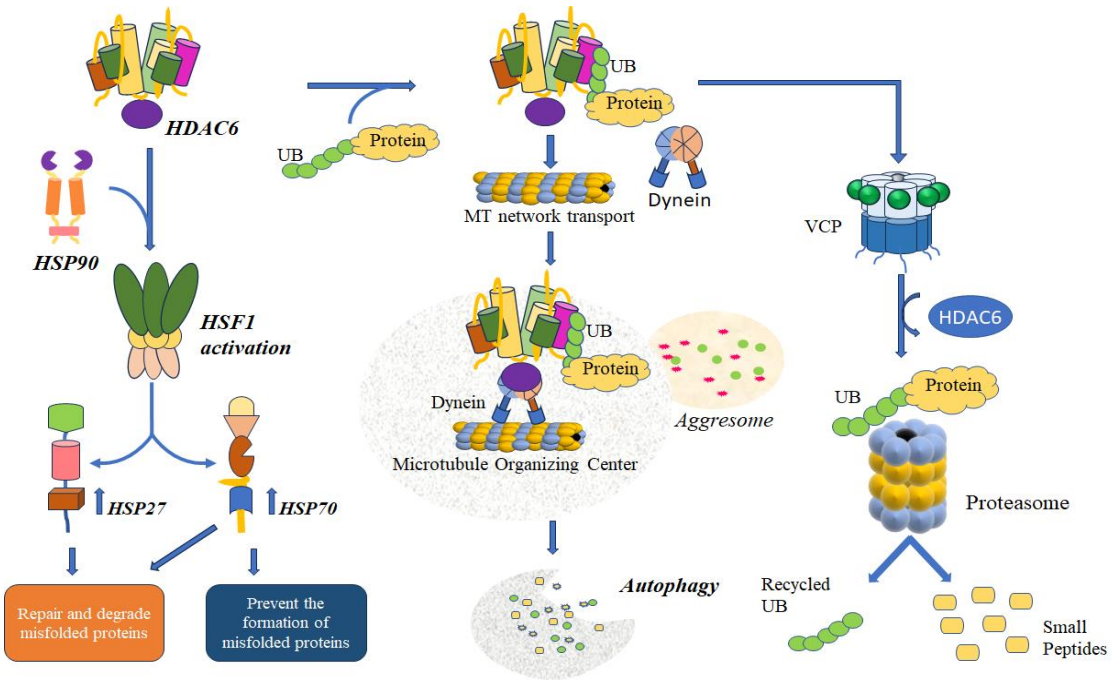


Figure 2.6: HDAC6 and misfolded protein degradation.

(a) Under non-stress conditions, to process misfolded proteins, HDAC6 binds with AAA-ATPase chaperone p97, a vasolin-containing protein known for its ability to disassemble polyubiquitinated proteins/HDAC6 complex.[81]

(b) If the proteosome-mediated pathway is inaccessible, then HDAC6 uses its ZBG to couple with ubiquitin tagged misfolded protein and carries it to microtubule organizing center (MTOC) using the microtubule network, where its cleared away via aggresome formation.[116]

(c) The third route HDAC6 uses to degrade misfolded protein is via heat shock protein 90 (Hsp90). HDAC6/Hsp90/HSF1 could promote heat shock transcription factor 1 (HSF1) which revamps expression of chaperone Hsp27 and Hsp70 down the line promoting repair and degradation of misfolded protein.[115]

2.5 The role of HDAC6 in diverse disease process

HDAC6 has a broad range of cellular substrate, getting it involved in pathophysiology of several disease conditions. Keeping this into mind, HDAC6 can be considered as a promising target for selective inhibitor design and development.

2.5.1 HDAC6 and cancer

HDAC6 overexpression is a characteristic of numerous cancers,[117] with studies indicating its necessity for effective oncogenic cell transmutation.[115] It is prominently involved in various signalling pathways, including the oncogenic Ras, AKT, and ERK1/2 pathways.[118] The above processes vitalize transformed cells to proliferate and survive independently of anchorage, bypassing anoikis -a specific type of programmed cell death triggered upon detachment of cells from the extracellular matrix and surrounding basement membrane.[114] Ongoing research on this field indicated that HDAC6 has a prominent role I n tumour development and the maintenance of a transformed phenotype. Further, many oncogenic proteins, depends on Hsp90-a substrate of HDAC6 for structural mutation and activation. Downregulation of HDAC6 contributes to hyperacetylation of Hsp90 and α-tubulin, thereby impairing the chaperone's function.[119]

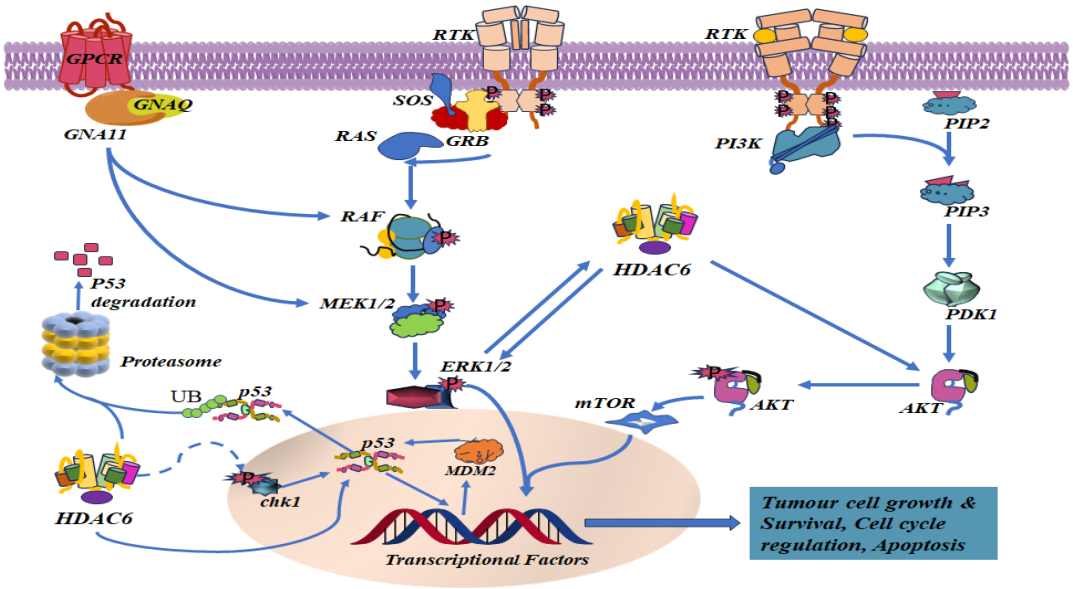


Figure 2.7 Involvement of HDAC6 in the cancer signalling pathway. Proposed model for HDAC6i mechanism of action targeting either MAPK/ERK or PI3K/AKT or p53 signalling pathway.

In hepatocellular carcinoma, the overexpression of HDAC6 induced by proinflammatory cytokines may enhance cell proliferation by supressing p53's transcriptional activity. HDAC6 inhibitors also function as tumour suppressor by reducing the activity of the Wnt/β-catenin signalling pathway in hepatocellular carcinoma.[63] HDAC6 stimulates cell proliferation, colony formation, cell migration, and invasion by directly interacting with the PTTPN1/ERK1/2 pathways, which target MMP-9.[120]

Recent studies on B16 murine melanoma cells with urea-derived HDAC6 inhibitor nexturastat A, have shown inhibition of cancer cell proliferation and induction of apoptosis, possibly by upregulation of cell cycle regulators like CYLD and acetylated microtubules, resulting in cell cycle disruption.[121]

2.5.2 HDAC6 and neurodegenerative diseases

Many neurodegenerative diseases such as Alzheimer's disease (AD), Huntington's disease (HD), Parkinson's disease (PD), and Charcot-Marie-Tooth disease are caused by the accumulation of protein aggregates, and HDAC6 due to the presence of ZBG or the ubiquitin binding domain is known for its ability to eliminate misfolded protein by augmenting autophagy.[63,122]

2.5.2.1 HDAC6 in Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder marked by the accumulation of extracellular β-amyloid peptide and intracellular neurofibrillary tangles (NFTs), which consists of cross β-fibrils formed by the misfolded protein tau (tubulin-associated unit).[123,124] Under normal condition, tau regulates microtubule dynamics, while HDAC6 controls tau phosphorylation and accumulation. In neuronal cells, tau hyperphosphorylation and aggregation lead to synaptic dysfunction, mitochondrial damage, and apoptotic cell death.[63,125] Studies have shown that HDAC6 regulates acetylated α-tubulin and Hsp90, playing a key role in mitochondrial axonal transport and protein aggregation/degradation by forming complexes with Hsp90, ubiquitin, and tau.[126] Further, proteasome inhibition results in the interaction between HDAC6 and tau, leading to increased co-localization of HDAC6 and tau in a perinuclear aggresome-like compartment. This process is independent of HDAC6's tubulin deacetylation activity.[127]

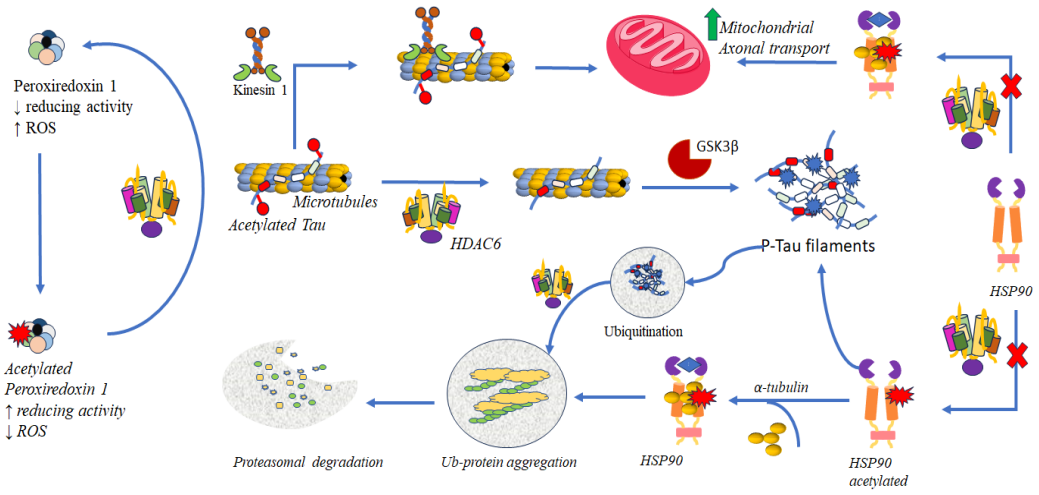


Figure 2.8 HDAC6 in Alzheimer disease pathogenesis.

Studies have shown that downregulation of HDAC6 in mouse model of Alzheimer's disease significantly improved learning, memory and α-tubulin acetylation.[63] It is known that HDAC6 directly influence acetylation level of peroxiredoxin 1 and 2 contributing in redox regulation (oxidative stress) which is a well-defined factor causing AD and aging.[128] Mitochondrial dysfunction and increased reactive oxygen species (ROS)[129] might be a possible mechanism in AD development. Further analysis using a HDAC6 knock-down mouse model suggested that the loss of HDAC6 activity might make neurones resistant to amyloid-β-induced mitochondrial trafficking deterioration, thereby restoring cognitive function.[130]

2.5.2.2 HDAC6 in Parkinson's disease

Parkinson's disease (PD) is a well-known progressive bradykinetic disorder[115] marked by gradual degeneration of the nigrostriatal dopaminergic pathway and the presence if Lewy bodies (insoluble cytoplasmic inclusion of α-synucleins).[131] The abrupt accumulation of α-synucleins in different regions of brain like substantia nigra, locus coeruleus and nucleus basalis of Meynert are the major contributing factor in the development of PD.[63,131] A mouse model of PD suggested that HDAC6 guards the dopaminergic neurons from cytotoxic α-synucleins aggregates by promoting aggresome formation by dissociating Hsp90 from Hsf1 complex.[132] Moreover, mutations in the DJ-1 gene, which is associated with early-onset of Parkinson's disease, lead to the misfolding and accumulation of this protein. This misfolded protein is subsequently targeted for elimination through binding with parkin and HDAC6. Parkin, an E3 ligase, forms a complex with the heterodimeric E2 enzyme UbcH13/Uev1a,

resulting in K63 linked polyubiquitination of the misfolded protein.[133] The polyubiquitinated proteins then bind to HDAC6 and DJ-1 aggregates, which are transported to aggresomes by the dynein motor complex.[134] Further studies have shown that parkin facilitates mitophagy (the removal of damage mitochondria) by recruiting HDAC6 and p62, which then form juxtanuclear mitochondria inclusion bodies resembling aggresomes.[131,134] In a recent study using a rat model of Parkinson's disease and Tubastatin A, it was observed that inhibiting HDAC6 increases the acetylation of α-synuclein, enhances the levels of Hsc70 and lamp2A (key components of chaperone-mediated autophagy), and reduces both α-synuclein expression and its toxicity.[135] This suggests that HDAC6 may be a promising therapeutic target for Parkinson's disease and other α-synucleinopathies.

2.5.2.3 HDAC6 in Huntington's disease

Huntington's disease (HD) is an autosomal inherited neurodegenerative disorder caused by a genetic alteration in the CAG triplet (Cytosine-Adenine-Guanosine), leading to the abnormal expansion of polyglutamine in proteins and resulting in the accumulation of huntingtin aggregates (HA).[131,136] In individuals without Huntington's disease, the CAG repeats range from 7 to 34 and may vary with age. However, CAG repeats exceeding 100 are associated with juvenile onset of the disease.[137] HD is characterised by uncontrolled excessive motor movements, cognitive impairment, and emotional deficits.[115] In Huntington's disease, alteration in cellular transport system, including microtubule-dependent transport and the intracellular transport of brain-derived neurotrophic factor (BDNF)-containing vesicles, are linked to the neuronal toxicity of huntingtin aggregates (HA).[138] Inhibition of HDAC6 is known to enhance microtubule-based transport by recruiting kinesin-1 and dynein/dynactin to more acetylated microtubules.[139] Conversely, studies using HDAC6 knockout mouse models shows increased tubulin acetylation but do not affect the progression of the disease.[140] Interestingly, another study suggested that HDAC6 is pivotal for the autophagic removal of aggregated huntingtin by recruiting the autophagic degradation machinery to the inclusion bodies.[141]

2.5.2.4 HDAC6 in Rett syndrome

Rett syndrome is a rare, progressive neurodevelopmental disorder that arises from a loss-of-function mutation in the X-linked MeCp2 gene.[142] MecP2 acts as a transcriptional repressor by interacting with DNA at CpG islands.[143] The loss of

functional MeCP2 mutations is linked to impaired BDNF trafficking and disrupted microtubule dynamics, highlighting the important role of HDAC6 in the neurobiology of Rett syndrome.[144] In neuronal cells, the acetylation level of microtubules controls the effectiveness of various crucial processes, including differentiation and migration, mitochondrial trafficking and the movement of BDNF-containing vesicles.[143,145] Additionally, inhibiting HDAC6 leads to accelerated axonal transport of BDNF and mitochondria in both anterograde and retrograde directions, thereby improving synaptic plasticity.[139,146]

2.5.2.5 HDAC6 in Charcot-Maire-Tooth disease

Charcot-Marie-Tooth (CMT) disease is the most prevalent inherited disorder affecting the peripheral nervous system, caused by mutations in the heat-shock protein genes (S135F orP182L). This leads to axonal CMT or distal hereditary motor neuropathy (distal HNM), and is characterised by muscle weakness, motor issues, and sensory loss. These symptoms have been linked to in vivo studies in transgenic mice.[147,148] Transgenic mice with mutations in this gene showed a reduction in the total number of mitochondria and defects in mitochondrial transport, which were improved by treatment with the HDAC6 inhibitor tubastatin A.[147] Aminoacyl transfer RNA (tRNA) synthetases, the largest gene/protein family associated with CMT, recently revealed HDAC6 as an intracellular factor interacting with glycyl tRNA synthetase (GlyRS or GARS) in CMT.[149] Administration of tubastatin A disrupted the interaction between GlyRS and HDAC6, leading to restored mitochondrial axonal transport, increased acetylation of α-tubulin, and improved muscle strength and motor performance.[147] This suggested that HDAC6 is highly promising target for treating CMT and potentially other peripheral neuropathies linked to axonal transport deficiencies.

2.5.2.6 HDAC6 in amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis is another progressive motor neurone disease caused by mutations in genes encoding superoxide dismutase 1 (SOD1), TAR DNA binding protein 43 (TDP-43), and fused in sarcoma (FUS).[147,150] It is marked by the selective degeneration of motor neurones in the motor cortex, brainstem, and spinal cord, resulting in progressive muscle weakness, paralysis, and eventually death, typically within 2-5 years after diagnosis.[147] Transgenic mice with the G93A mutation in superoxide dismutase 1 (SOD1) exhibited defects in axonal transport, but genetic deletion of HDAC6 significantly slowed disease progression and extended the survival

of these mutant SOD1^{G93A} mice.[147] Additionally, studies using induced pluripotent stem cells (iPSCs) derived from fibroblasts of ALS patients with various FUS mutations (R531H and P525L) demonstrated that treatment with HDAC6 inhibitors like tubastatin A or ACY-738 restored impaired axonal transport and increased the overlay of endoplasmic reticulum (ER) and mitochondria.[150,151] In a transgenic mouse model of ALS overexpressing wild-type FUS (Tg FUS+/+ mouse),[151] which exhibited histone hypoacetylation in the spinal cord and cortical tissue associated with progressive neurodegeneration, treatment with ACY-738 significantly extended the survival rate of the mutant mice. It also reduced neuromuscular denervation and muscle atrophy, thereby improving the ALS disease phenotype.[147]

2.5.3 HDAC6 in inflammation

Inflammation is a crucial biological response to infection by bacteria and viruses affecting the body surface and organs in mammals.[78] Recent research has highlighted the significant role of HDAC6 in the innate immune response to intracellular bacterial infections, particularly through Toll-like receptor-mediated signalling.[152] HDAC6 was also found to significantly influence the production of cytokines. This includes both pro-inflammatory cytokines (IL-6, IL-1β, TNFα, IL-17) and the anti-inflammatory cytokine IL-10. Inhibiting HDAC6 led to a reduction in the production of IL-6, IL-1β, and TNF-α in various mouse models of inflammatory diseases.[78] Additionally, disrupting HDAC6 leads to the recruitment of inflammatory antigen-presenting cells, which are essential for initiating T-cell activation and T-cell tolerance.[153] The suppressive activity of Foxp3+ regulatory T-cells is enhanced when HDAC6 is depleted in models of autoimmunity and inflammation.[154] Defects in the number or function of Foxp3+ regulatory T-cells, which are crucial for maintaining immune homeostasis, can result in autoimmunity. Research has also indicated that HDAC6 regulates HIV replication by modulating the deacetylation of Tat, thereby inhibiting viral transactivation.[155,156] HDAC6 has also been implicated in Sendai virus infection by deacetylating β-catenin, which serves as a co-activator of IRF-3 mediated transcription.[157] Due to its regulatory role in inflammation, inhibiting HDAC6 could be an effective treatment of various inflammatory disease, including rheumatoid arthritis, inflammatory bowel disease, and airway inflammation.

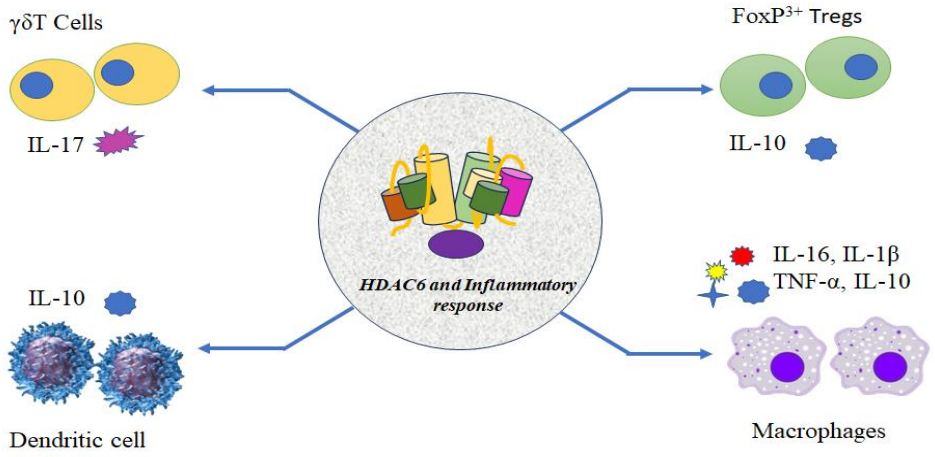


Figure 2.9: Role of HDAC6 in the regulation of inflammatory cells (macrophages, dendritic cells, γδT cells, and FoxP3^{+} tregs cells) and cytokines (IL-6, IL-1β, TNF-α, IL-10, and IL-17).

2.5.3.1 HDAC6 in rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease marked by inflammatory synovitis, proliferation and invasion of synovial tissue, resulting in the destruction of bone and cartilage.[158] HDAC6 has gained increasing attention in RA because it deacetylates not only histone but also non-histone protein such as α-tubulin and myeloid differentiation primary response 88 (MyD88). MyD88 is an essential adaptor molecule for the Toll-like receptor and IL-1 receptor in the NF-κB signalling pathway, and HDAC6 can deacetylate it.[159,160] Small molecule inhibitors of HDAC6 have been shown to reduce the production of pro-inflammatory cytokines IL-6, TNF-α, and IL-1β, leading to decrease in synovial inflammation and indicating their potential role in RA treatment. Tubastatin A, a selective HDAC6 inhibitor, has been demonstrated to effectively reduce synovial inflammation and protect against joint destruction in a collagen antibody-induced arthritis mouse model. similarly, CDK-l, another selective HDAC6 inhibitor, has been found to inhibit the expression of IL-6, TNF-α, and IL-1β while increasing IL-10 production. This results in a lower arthritis score and reduced proliferation of effector T cells in a collagen-induced arthritis mouse model.[161]

2.5.3.2 HDAC6 in inflammatory bowel disease

Inflammatory bowel disease (IBD) encompasses a group of chronic immune-mediated disorders affecting the gastrointestinal tract, including Crohn's disease and ulcerative colitis.[162] These conditions are characterised by recurrent inflammation and subsequent

damage to the gastrointestinal tract.[163] Current treatment for IBD, such as anti-inflammatory drugs, antibiotics, and biologics, often prove ineffective and can have adverse effects.[164] Research using knockdown mice and small molecule inhibitors has revealed that HDAC6 plays a crucial role in the progression of OBD. Selective inhibition of HDAC6 with molecules like BML-281 and LTB2 has been shown to effectively alleviate colitis induced by dextran sulfate in mouse model, BML-281 treatment protected against colonic inflammation and prevented the activation of inflammatory neutrophils. Similarly, LTB2 treatment was linked to reduced rectal bleeding and diarrhoea.[165,166] These findings suggest that HDAC6 inhibition may be potential in preventing colonic inflammation and treating IND in humans.

2.5.3.3 HDAC6 in airway inflammation

Airway inflammation is a key factor in many chromic respiratory diseases, including asthma chronic obstructive pulmonary disease (COPD). Asthma is characterised by persistent airway inflammation, increased airway responsiveness, and airway remodelling. In a mouse model of chronic allergic airway disease, the selective HDAC6 inhibitor Tubastatin A has been shown to effectively reduce airway inflammation, airway remodelling and airway hyperresponsiveness. These results indicate that HDAC6 may play a significant role in asthma treatment.[167]

Chronic obstructive pulmonary disease (COPD) is marked by epithelial cell dysfunction, ciliary shortening, impaired mucociliary clearance, and abnormal airway inflammation, primarily due to chronic cigarette smoking exposure. Current treatments for COPD are often inadequate and ineffective against exacerbation. However, HDAC6 inhibition with Tubastatin A has been shown to significantly reduce airway dysfunction induced by cigarette smoke, suggesting it could be a promising therapeutic approach for COPD.[168]

Research has also identified HDAC6 as a common factor in the development of dysregulated pro-inflammatory and fibrotic phenotypes in cystic fibrosis (CF), an inherited lung disease marked by extensive collagen deposition and tissue remodelling.[169] Fibrosis in CF is characterised by the excessive growth of tissue, an increase in myofibroblasts, and abnormal deposition of extracellular matrix components, a process known as epithelial-mesenchymal transition (EMT).[170]

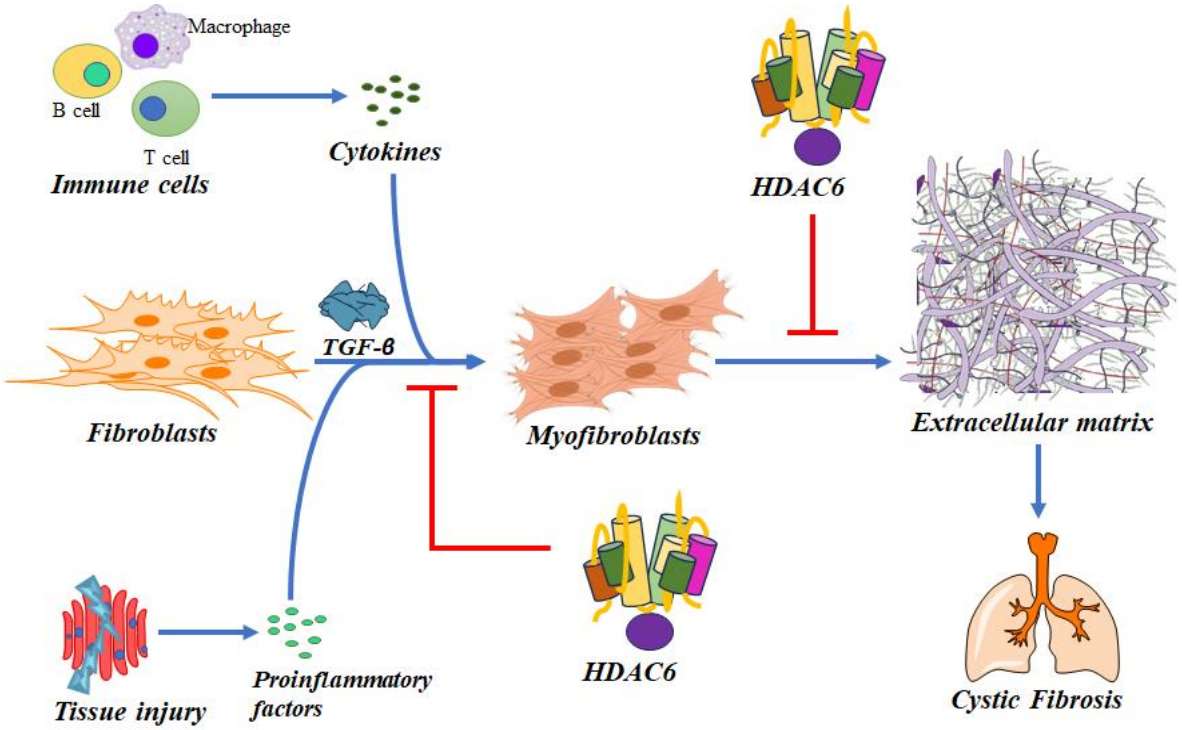


Figure 2.10: Role of HDAC6 in the regulation of the fibrotic process. HDAC6 inhibitors negatively regulate the fibrotic process by acting at different levels of the profibrotic cascade.

In cystic fibrosis (CF), EMT is a crucial process involving the loss of cell-cell junctions and cell surface molecule polarization, leading cells to adopt mesenchymal characteristics.[171] Recent studies have shown that tubacin, and HDAC6 blocker, reduces TGFβ1-induced EMT markers and inhibits SMAD3 activation in response to TGFβ1.[169] Since SMAD3 is a key component of TGFβ1 signalling, its inhibition disrupts HDAC6-dependent deacetylation of α-tubulin, highlighting the essential role of HDAC6 in EMT through the TGFβ1-SMAD3 signalling pathway.[172]

2.5.4 HDAC in acute kidney injury

Acute kidney injury (AKI), marked by a rapid decline in glomerular filtration rate, is a serious clinical issue associated with sever disease progression, high mortality rates, and an increased risk of developing chronic kidney disease (CKD).[173] Rhabdomyolysis accounts for 15% of AKI[174] cases and can be triggered by various factors, including metabolic disorders, trauma, infections, drugs and toxins.[175,176] While the exact mechanism are not fully understood, it is well-established that endoplasmic reticulum (ER) stress-induced apoptosis of tubular epithelial cells plays a crucial role in

rhabdomyolysis-related AKI.[177] Recent research has highlighted organelle-mediated stress, particularly ER stress, as a key pathophysiological factor in apoptosis. HDAC6 activation has been implicated in the development of rhabdomyolysis-induced AKI, contributing to renal tubular cell apoptosis, inflammatory responses, macrophage infiltration, and oxidative stress.[178,179]

Apoptosis or programmed cell death, is typically triggered by changes in the cell microenvironment.[180,181] It involves the activation of pro-apoptotic molecules and the deactivation of pro-survival ones.[182] In the tubular epithelium, apoptotic pathways can be activated through mechanisms such as caspase cascade activation, mitochondrial damage, and endoplasmic reticulum stress.[183] Apoptosis leads to the loss of renal epithelial cells, a hallmark of acute kidney injury.[184,185] Research indicates that caspase 3 activation is the primary mechanism driving renal tubular cell apoptosis in rhabdomyolysis-induced AKI.[177,185] Additionally, the Bcl-2 family plays a crucial role in regulating apoptosis.

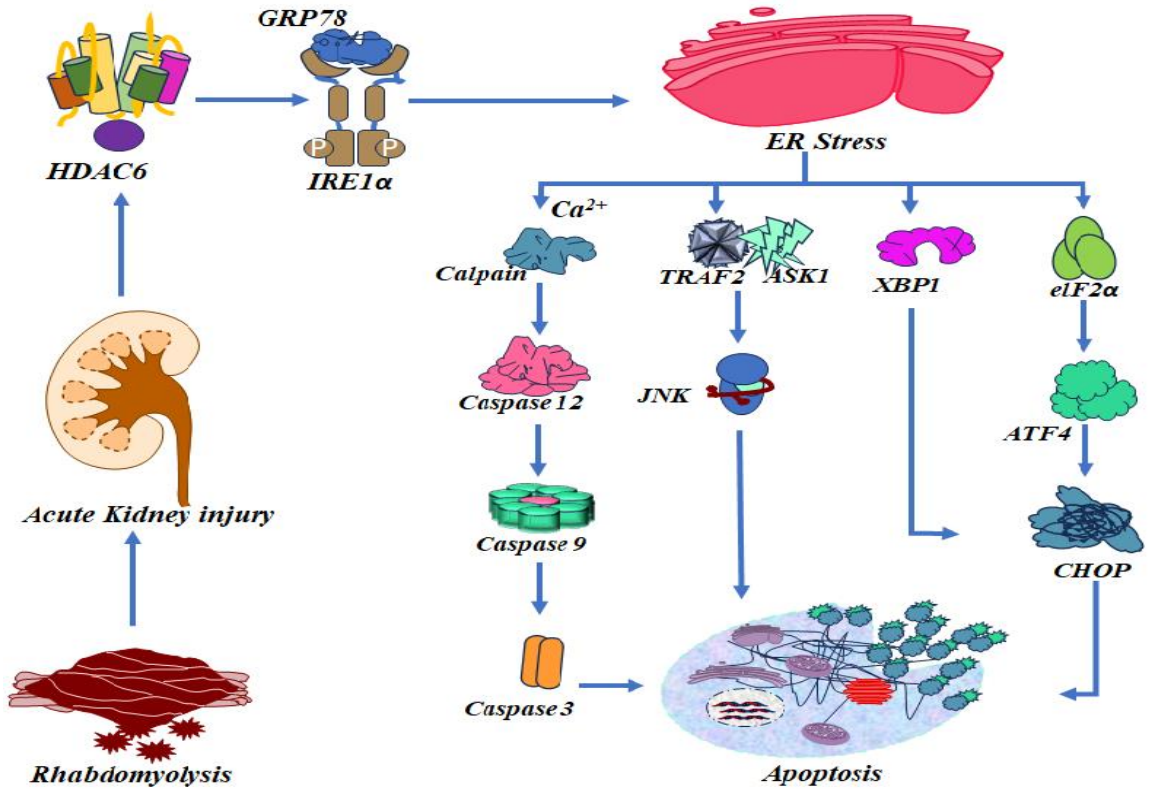


Figure 2.11: The involved mechanism of HDAC6 inhibitor against rhabdomyolysis induced-acute kidney injury.

Endoplasmic reticulum (ER) stress in acute kidney injury (AKI) can be triggered by various factors, including mutant protein aggregation, hypoxia, energy deprivation, and metabolic dysfunction.[186] A reduced capacity for protein folding in the ER results in the accumulation of misfolded proteins and initiates ER stress. Excessive ER stress leads to tubular cell apoptosis through three main signalling pathways: PERK-eIF2-ATF4, IRE1-VBP1, and ATF6.[187] HDAC6 primarily located in the cytoplasm, contributes to the acetylation of the ER-localized chaperone protein glucose-regulated protein (GRP78). Under normal conditions, GRP78 binds to signalling proteins PERK, ATF6 and IRE1, inhibiting their activation.[188] Research by Feng et al. demonstrated that inhibiting HDAC6 reduces ER stress, as indicated by lower GRP78 expression.[189] This suggests that HDAC6 inhibitors could be a promising treatment for rhabdomyolysis-induced acute kidney injury.

2.5.5 HDAC6 in myocardial dysfunction

Myocardial dysfunction is a major cause of early death after successful cardiopulmonary resuscitation (CPR) in patients with cardiac arrest (CA). this dysfunction may be driven by cell pyroptosis, a novel type of programmed cell death characterised by plasma membrane rupture and the release of inflammatory cytokines.[190] Pyroptosis is primarily mediated by NOD-like receptor protein 3 (NLRP3)-caspase-1 pathway. In this process, the NLRP3 inflammasome facilitates the conversion of pro-caspase-1 into active caspase-1, which then cleaves pro-inflammatory cytokines interleukin-1β and interleukin-18, as well as pyroptotic substrate gasdermin D. This results in cell death and excessive release of interleukin-1β and interleukin-18.[191] In a study by Jiefeng Xu et al. it was found that inhibiting HDAC6 with Tubastatin A blocked the activation of the NLRP3-caspase-1 pathway and reduced cell pyroptosis in an H9c2 cardiomyocyte hypoxia/reoxygenation (H/R) model.[190] additionally, Tubastatin A promoted the acetylation and nuclear translocation of transcription factor EB (TFEB), which is crucial pro regulating the autophagy-lysosome pathway and can inhibit autophagy induction.[202,203] The acetylation and nuclear translocation of TFEB resulted in the inhibition of NLRP3 inflammasome activation, leading to improved myocardial function and reduced cardiac injury following cardiac arrest and resuscitation by decreasing pro-inflammatory cytokines.[190]

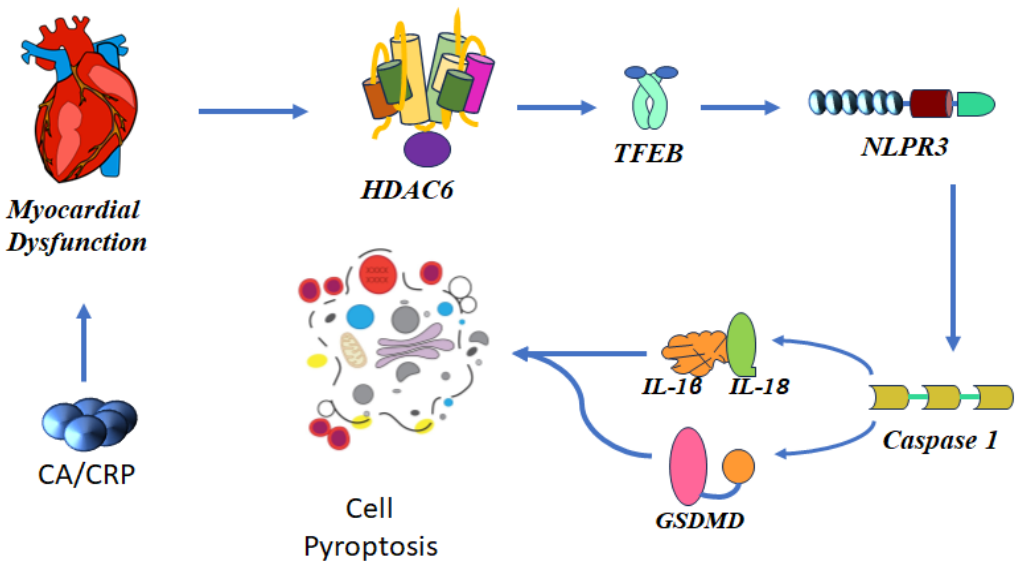


Figure 2.12: The proposed molecular mechanisms of the protective effect of HDAC6 inhibitors on myocardial dysfunction after cardiac arrest (CA) and resuscitation.

2.6 Inhibitors of HDAC6

A variety of HDAC inhibitors have been identified to date. Six of these- vorinostat, romidepsin, belinostat, Panobinostat (withdrawn), pracinostat (FDA-approved), and chidamide (approved by the China FDA)-are approved for the clinical treatment of refractory or relapsed cutaneous and/or peripheral T-cell lymphomas, or multiple myeloma.[58] Many other inhibitors are currently undergoing clinical trials. Most of the approved HDAC inhibitors are non-selective (pan-HDAC) or selective for class I HDACs,[194] which often leads to unwanted side effects.[115] Therefore, there is a need for highly isoform-specific HDAC inhibitors to better understand the biological roles of individual HDAC isoforms and to provide targeted therapies with minimal side effects. Most HDAC inhibitors share a common pharmacophore that includes a zinc bindi ng group (ZBG) or chelating group such as (hydroxamic acid, thiol, carboxylic acid, ketones or substituted aniline), a cap group (for surface recognition), and a linker connecting the ZBG and cap group.[63,194] Modification to any part of this pharmacophore can significantly affect potency, stability, and isoform selectivity. Notably, modifying the cap group is a promising strategy for enhancing isoform selectivity.[194]

2.6.1 Hydroxamic acid based HDAC6 inhibitors

The chelation of the Zn^{2+} ion is crucial for the inhibition of classical HDACs. Hydroxamic acid, as a zinc-binding group (ZBG), has been the most extensively studied

due to its strong ability to chelate Zn^{2+}. Since the discovery of the first hydroxamic acid-based HDAC inhibitor, Trichostatin A (TSA), by Yoshida et al.[195] in 1990, significant advancements have been made in this class of inhibitors. TSA, a faunistic antibiotic isolated from *Streptomyces hydroscopious* by Tsuji et al.[196] in 1976, paved the way for further development. Following TSA's discovery, several hydroxamic acid inhibitors have been approved, including Vorinostat (2006), Belinostat (2014), Pracinostat (2014), and Panobinostat (2015, later withdrawn), for the treatment of peripheral T-cell lymphomas and melanoma.[58,63] Currently, (R)-Trichostatin A is mainly used as a research tool. The availability of its crystal structure complexed with an HDAC enzyme has significantly contributed to the development of newer and more potent HDAC inhibitors.

***R*-Trichostatin A (7)**
HDAC Ki=3.4 nM;

***S*-Trichostatin A (7a)**
HDAC Ki= no inhibition

Figure 2.13: Trichostatin A

Trichostatin A features a structure comprising an N-methyl group, a linker with an alkenyl group, and a distinct hydroxamic acid tail. The hydroxamate group forms a bivalent coordination with the Zn^{2+} ion located at the base of the catalytic pocket, while the 4-dimethylaminobenzoyl moiety in the cap region interacts with residues around the channel rim leading to the active site. The linker region occupies the hydrophobic channel. [197] Various aryl substitution and structural modifications on the TSA scaffold have been explored to assess the structural-activity relationship and enhance interactions with the enzyme's catalytic domain. However, these derivatives generally exhibit lower potency compared to the natural TSA.

Common pharmacophore of HDAC inhibitors typically involves a scaffold composed of a six-carbon chain, a hydrophobic capping group, and a chelating group at the terminal end. A classic and extensively studied example of this design is suberonylanilide hydroxamic acid, known as SAHA (Vorinostat). This synthetic small molecule is renowned for its potent HDAC inhibitory activity.[198] SAHA was engineered to enhance lipophilicity by incorporating a hydrophobic phenyl ring at the

opposite end. Additionally, SAHA serves as a key example of an inverse amide analogue of trichostatin[199] and is notable for being the first HDAC inhibitor approved by US-FDA for the treatment of cutaneous T-cell lymphoma.

Stephen J. Haggarty and colleagues discovered tubacin (**8**) as the first selective HDAC inhibitor through a comprehensive chemical genetic screening of 7392 small molecules. Tubacin features a 2,3-dioxane structure and demonstrates up to a 317-fold selectivity for HDAC6 over HDAC1 and HDAC2.[200,201] Research indicates that the specific configuration of the dioxane ring, which interacts with the protein surface, contributing to this selectivity.[194] Tubacin was shown to inhibit α-tubulin deacetylation, suppress cell proliferation, and induce apoptosis without impacting histone acetylation, gene expression, or cell cycle progression in mammalian cells. Additionally, it does not exhibit toxicity towards normal haematological cells.[202] However, due to its high lipophilicity, attributed to its large cap group consisting of six lipophilic rings and its non-drug-like structure, tubacin is currently used primarily as a research tool.

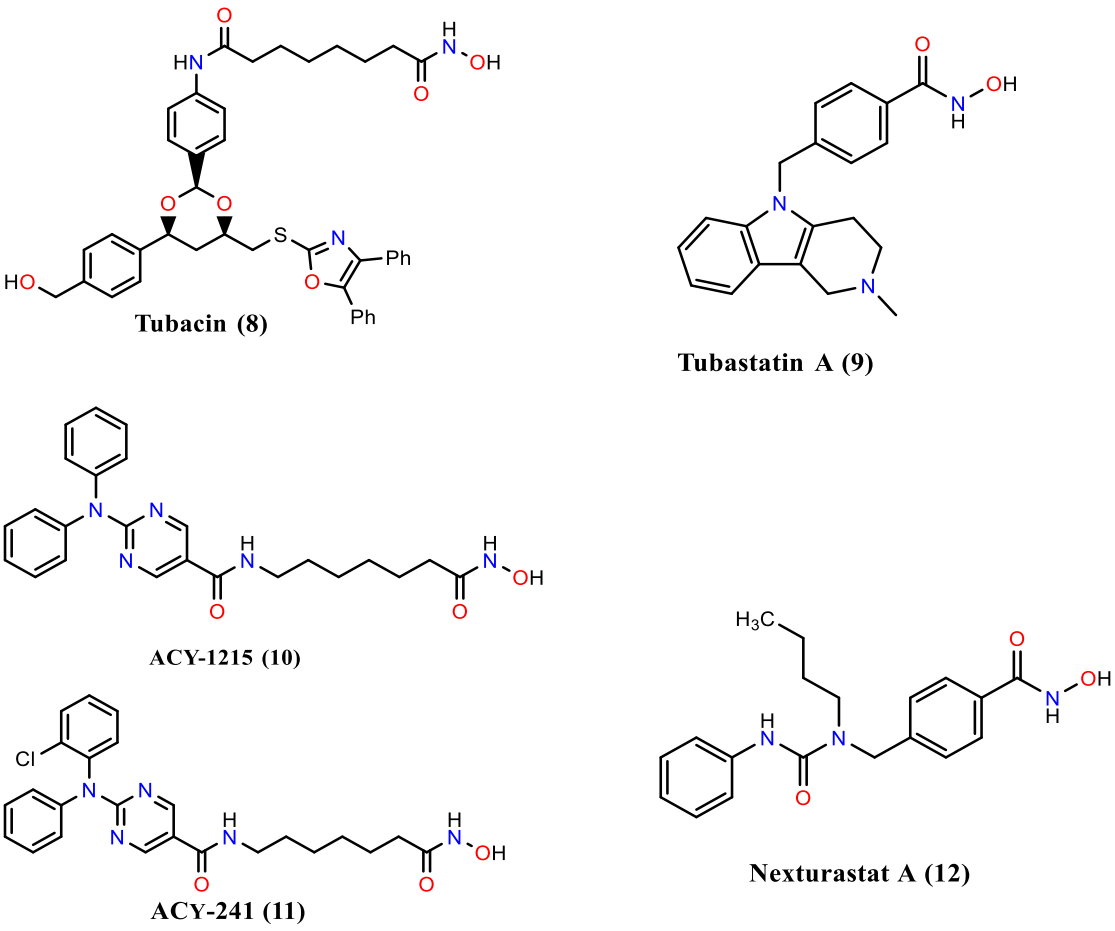


Figure 2.14: Some selective HDAC6 inhibitors.

Ricolinostat (ACY-1215) is the first selective oral HDAC6 inhibitor to advance to clinical trials for multiple myeloma (MM). it shares structural similarity with tubacin but offers improved drug-like properties.[203] As a micromolar dose (0.62μm), ricolinostat effectively increases α-tubulin acetylation without affecting histone acetylation. It is 12, 10 and 11 times more selective for HDAC6 compared to HDAC1, HDAC2, and HDAC3, respectively. Ricolinostat can be used alone or in combination with bortezomib, a proteasome inhibitor, for treating MM. The synergistic effect of combining HDAC inhibitors with bortezomib is not fully understood yet. Additionally, ricolinostat has shown a synergistic anti-MM effect when used with carfilzomib, enhancing carfilzomib-induced cell death by inhibiting aggresome formation.[204] Ricolinistat has been extensively studied in Phase I/II clinical trials, either alone or in combination with dexamethasone, bortezomib,[205] or lenalidomide[206] for relapsed or refractory multiple myeloma.[205] It has also been investigated as an anti-lymphoma agent in combination with bendamustine.[207] In mouse xenograft models of oral squamous cell carcinoma, ricolinostat has demonstrated potential in suppressing tumour growth and inducing apoptosis through various pathways involving miR-30d/PI3K/mTOR and ERK.[208] Although ricolinostat shares some structural characteristics with the pan-HDAC inhibitor SAHA, such as a long aliphatic linker and bivalent zinc binding, it is distinguished by its unique surface recognition domain that interacts with the cleft between the L1 and L7 loops of HDAC6, contributing to selectivity of HDAC6.[83]

As the successor to ACY-1215, an orally active second-generation analogue, ACY-241, was developed with an IC_{50} of 2.6 nM against HDAC6 and over 18-fold reduced potency towards Class I HDACs. Both ACY-1215 and ACY-241 features a 2-(diphenylamino) pyrimidine-5-carboxamide as the surface recognition group, which is likely responsible for their HDAC6 inhibitory potency and selectivity. ACY-241 is considered sightly superior to ACY-1215 due to the electron-withdrawing chlorine substitution on one of the phenyl rings. In Phase Ib clinical trials for multiple myeloma,[209] and in studies involving solid tumour, ACY-241, when combined with paclitaxel, enhanced anti-proliferative activity and increased cell death.[210] Additionally, in murine xenograft models of multiple myeloma, ACY-241 demonstrated a synergistic effect with pomalidomide, enhancing tumour growth inhibition, promoting apoptosis, and causing cell cycle arrest both in vitro and in vivo.[211] Currently, ACY-241 is being

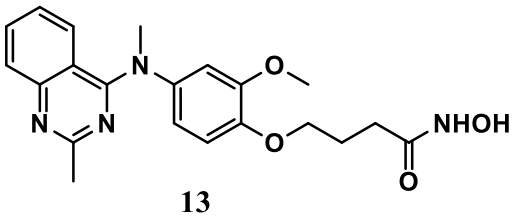
evaluated in phase I clinical trials, either alone or in combination, for treating various cancers.[212]

Compound **13**, a quinazoline hydroxamate, has been identified as a preferentially selective HDAC6 inhibitor with an IC50 of 17 nM and 200-fold selectivity over HDAC8.[213] It exhibits superior anti-proliferative effects against eleven different haematological and solid tumour cell lines compared to SAHA and ACY-1215. Additionally, Compound **13** has demonstrated 47.0% oral bioavailability in rats, indicating a favourable pharmacokinetic profile. Biological evaluations by Feng et al. revealed that Compound **13** can regulate endoplasmic reticulum stress and apoptosis, leading to the attenuation of rhabdomyolysis-induced acute kidney injury.[214]

2.6.2 N-hydroxy benzamide based HDAC6 inhibitors

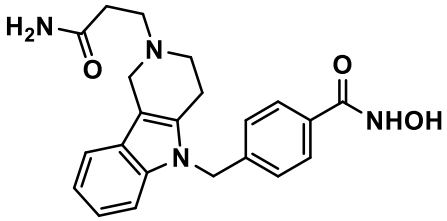
Unlike other classical HDAC isoforms, HDAC6 has a wider and shallower channel, suggesting that replacing traditional long-chain alkyl linkers with bulkier and shorter aromatic moieties can more efficiently achieve HDAC6 selectivity.[145] Additionally, a rigid and larger surface recognition group is better suited to occupy the rim region of HDAC6. To address this, Tubastatin A, a potent HDAC6 inhibitor with improved pharmacokinetic properties, was designed and synthesised. Tubastatin A is a tetrahydro-γ-carboline derivative with up to 1000-fold selectivity over all HDAC isoforms except HDAC8 (57-fold). Studies have shown that Tubastatin A can induce hyperacetylation of α-tubulin and slightly induce histone hyperacetylation when tested alone. In a homocysteic acid (HCA)-induced oxidative stress model, Tubastatin A demonstrated neuroprotective effects without causing neuronal cell death. It also exhibits various biological activities, including anti-inflammatory and anti-rheumatic effects,[215] suppression of hepatitis C virus proliferation,[216] and mitigation of stroke-induced brain inflammation and functional deficits.[217]

Following this, the structure of Tubastatin A (**9**) was modified to develop second-generation HDAC6 selective inhibitors. Some compounds in this series (compounds **14**, **15**, and **16**) showed sub-nanomolar inhibitory activity for HDAC6 with over 7000-fold selectivity over HDAC1. These compounds also enhanced the ability of Fox3+ regulatory T cells to suppress the mitotic division of effector T cells, indicating potential for further investigation into their use for treating autoimmune disorders.[218]



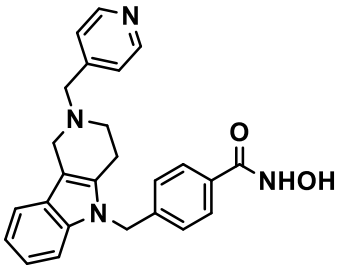
13

HDAC1 = 422 nM;
HDAC6 =17 nM;
HDAC8 = 3398 nM;
Other HDACs: Not reported



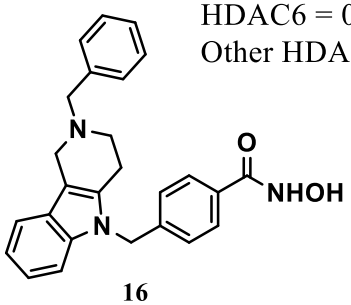
14

HDAC1 = 16100 nM;
HDAC6 = 0.459 nm;
Other HDACs: Not reported



15

HDAC1 = 2740 nM;
HDAC6 = 0.582 nm;
Other HDACs: Not reported



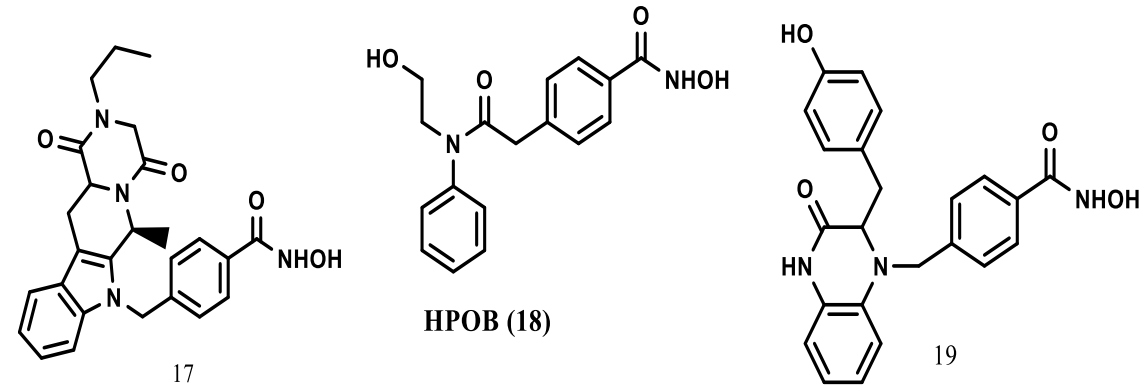
16

HDAC1 = 4320 nM;
HDAC6 = 0.872 nm;
Other HDACs: Not reported

Figure 2.15: Tubastatin A and Tubastatin A analogues with their respective IC_{50} values for different HDAC isoforms.

Leonhardt et al. designed and synthesized highly potent HDAC6 selective inhibitors featuring a large and rigid tetrahydro-β-carboline as the surface recognition group, tailored to fit the extensive active binding site of HDAC6. Among these, Compound **17** stands out as particularly potent, demonstrating superiority over Tubastatin A in cellular assays. Compound **17** induces rapid hyperacetylation of tubulin without affecting histone H3.[219] Additionally, it does not induce apoptosis or cell cycle arrest at sub-micromolar concentrations.

Nexturastat A (**12**), a potent HDAC6 inhibitor, has an IC50 value of 5.02 nM and is 601-fold more selective over HDAC8, exemplifying this (N-hydroxy benzamide) type of structure. Structure-activity relationship studies of Nexturastat A and its derivatives have shown that introducing a branching element, particularly to the nitrogen atom near the zinc-binding group, significantly enhances potency and selectivity for HDAC6.[91] Nexturastat A effectively increases the acetylation of α-tubulin and inhibits the growth of B16 melanoma cells.



17

HDAC1 = 1055 nM;
HDAC2 = 1870 nM;
HDAC4 = 1445 nM;
Other HDACs: Not reported

HDAC1 = 2900 nM;
HDAC2 = 4400 nM;
HDAC3 =1700 nM;
HDAC4 > 10000 nM;
HDAC5 > 10000 nM;
HDAC6 = 56 nM;

HDAC2 = 330 nM;
HDAC6 = 10 nM;
HDAC8 = 73 nM;
Other HDACs: Not reported

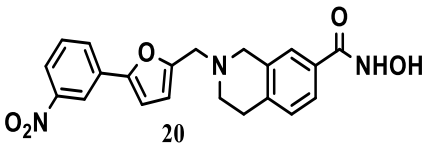
Figure 2.16: some *N*-hydroxy benzamide based HDAC6 inhibitor.

HPOB (**18**), a compound structurally similar to Nexturastat A, is also an HDAC6 selective inhibitor with an IC50 value of 56 nM. It effectively induces the acetylation of α-tubulin without impacting histone acetylation in both normal (HSF) and transformed (LNCAP, U87, and A549) cells. HPOB can be used alone or in combination to inhibit the growth of cancer cells. When used alone, HPOB inhibits the growth of both normal and transformed cells without causing cell death at concentrations ≤ 16 μM. However, when combined with other agents such as etoposide, doxorubicin, or SAHA, HPOB enhances the induced cell death in transformed cells.[220]

Smil et al. developed novel chiral 3,4-dihydroquinoxalin-2(1H)-one and piperazine-2,5-dione aryl hydroxamate (19) compounds, which exhibit high affinity and selectivity for HDAC6. They discovered that the selectivity and potency of these inhibitors are significantly influenced by the stereochemistry of the chiral moiety, while the substituent on the cap group has minimal impact on HDAC6 potency and selectivity. Additionally, the stereochemistry of the chiral moiety affects the acetylation levels of α-tubulin and histone H3.[221]

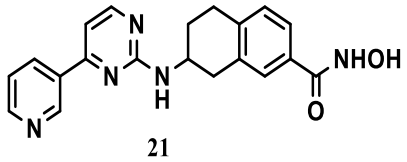
Guozhi Tang and colleagues developed a 2,7-disubstituted tetrahydroisoquinoline derivative (compound **20**) as a dual HDAC6/HDAC8 inhibitor, which exhibited up to 426-fold selectivity over HDAC1. However, compound **20** faced challenges such as high intrinsic clearance, N-C cleavage on the tetrahydroisoquinoline moiety, and poor solubility (<10 μg/ml).[222] To address these ADME (Absorption, Distribution,

Metabolism, and Excretion) issues, a second series of aminotetralin derivatives was designed and synthesized. Among these, aminotetralin-derived hydroxamates, particularly compound **21**, demonstrated good solubility (66 μg/ml) and metabolic stability while maintaining potency and selectivity for HDAC6 and HDAC8. Further structural optimization of compound **21** resulted in the development of molecule **22**, a tetrahydroquinoline-based HDAC6 selective inhibitor, which exhibited improved HDAC6 inhibitory activity (IC50 = 12 nM) and solubility (210 μg/ml).[223] Subsequently, scaffold hopping was employed to design a third series of amino-pyrrolidinone-based HDAC6 selective inhibitors, leading to the discovery of compound **23**, which displayed the highest affinity for HDAC6 in this series.[223]



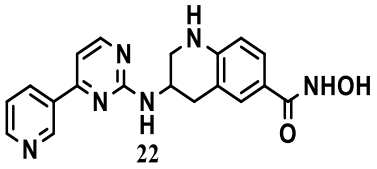
20

HDAC1 = 21300 nM; HDAC7 = 60700 nM;
HDAC2 = 109000 nM; HDAC8 = 30 nM;
HDAC3 = 6590 nM; HDAC9 = 65800 nM;
HDAC4 = 678000 nM; HDAC10 = 71100 nM;
HDAC5 = 36000 nM; HDAC11 =75300 nM;
HDAC6 = 50 nM;



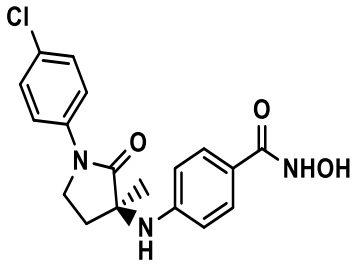
21

HDAC1 = 6310 nM; HDAC7 = 30800 nM;
HDAC2 > 100000 nM; HDAC8 = 8 nM;
HDAC3 > 100000 nM; HDAC9 = 35000 nM;
HDAC4 > 100000 nM; HDAC10 > 100000 nM;
HDAC5 > 100000 nM; HDAC11 > 100000 nM;
HDAC6 = 50 nM;



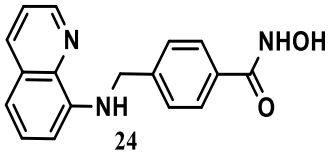
22

HDAC1 = 9960 nM;
HDAC6 = 12 nM;
HDAC8 = 470 nM;
Other HDACs: Not reported



23

HDAC1 = 74100 nM;
HDAC6 = 17 nM;
HDAC8 = 180 nM;
Other HDACs: Not reported



24

HDAC1 = 9550 nM; HDAC7 = 709 nM;
HDAC2 = 12500 nM; HDAC8 =1190 nM;
HDAC3 = 7750 nM; HDAC9 = 1530 nM;
HDAC4 = 4438 nM; HDAC10 =9791 nM;
HDAC5 = 3112 nM; HDAC11 = 376 nM;
HDAC6 = 0.29 nM;

Figure 2.17: Some *N*-hydroxy benzamide based HDAC6 inhibitors with their respective IC_{50} values against different HDAC isoforms

Dallavalle et al. discovered a series of arylamino/heteroarylamino hydroxamates as novel HDAC6 inhibitors.[224] Within this series, compound **24** demonstrated exceptional inhibitory activity against HDAC6 (IC50 = 0.29 nM) with up to 4000-fold selectivity over other HDAC isoforms. As a single agent, compound **24** showed significant anti-

proliferative activity against human multiple myeloma cell lines (RPMI-8226, U226, and NCI-H929), indicating its potential as a therapeutic candidate for multiple myeloma. Additionally, it exhibited good hepatocytic stability and high permeability.

2.6.3 N-hydroxycinnamamide based HDAC6 inhibitors

ST3595 (**25**), a *N*-hydroxycinnamamide based inhibitor capped with a biphenyl group, was initially studied as an HDAC2 inhibitor.[225] However, it was found to be more effective against HDAC6 (IC50 = 100 nM) compared to HDAC2 (IC50 = 1160 nM). When combined with paclitaxel, ST3595 exhibited a synergistic anti-tumour effect in wild-type p53 ovarian carcinoma cells[226] and also showed efficacy against pancreatic cancer cells.[227] Additionally, in studies involving non-small cell lung cancer (NSCLC) cell lines H460 and A549, as well as their cisplatin-resistant variants H460/Pt and A549/Pt, ST3595 demonstrated more significant antiproliferative activity in the H460/Pt cell subline compared to the other sublines.[228]

C1A (**26**), a weak HDAC6 inhibitor with an IC50 of 479 nM, was developed from the structure of the naturally occurring pan-HDAC inhibitor trichostatin A. C1A induces acetylation of α-tubulin and HSP-90 in a concentration-dependent manner.[229] Notably, despite containing a nitrogen mustard moiety, C1A does not cause non-specific DNA alkylation. In cellular assays, C1A inhibited the growth of one type of B-cell malignancy as well as eight different histological types of solid tumours. Additionally, C1A treatment led to an increase in the sub-G1 cell population and activation of caspase-3/7, indicating that its anti-proliferative effects may be mediated through an apoptotic mechanism.

A new series of quinazolin-4-one derivatives, featuring a quinazolin-4-one core structure and an N-hydroxycinnamamide moiety, has shown high affinity and selectivity for HDAC6. Among these, compound **27** stands out for its exceptional inhibitory potency against HDAC6 (IC50 = 8 nM). In in vitro biological evaluations, several potent compounds from this series induced neurite outgrowth and significantly increased acetylation of α-tubulin.[230] Following a range of biological assessments, compound **28** has emerged as a promising candidate for Alzheimer's disease (AD) treatment due to its lack of effect on human ether-a-go-go-related gene (HERG) ion channel activity (IC50 > 10 μM) and cytochrome P450 activity (IC50 > 6 μM).

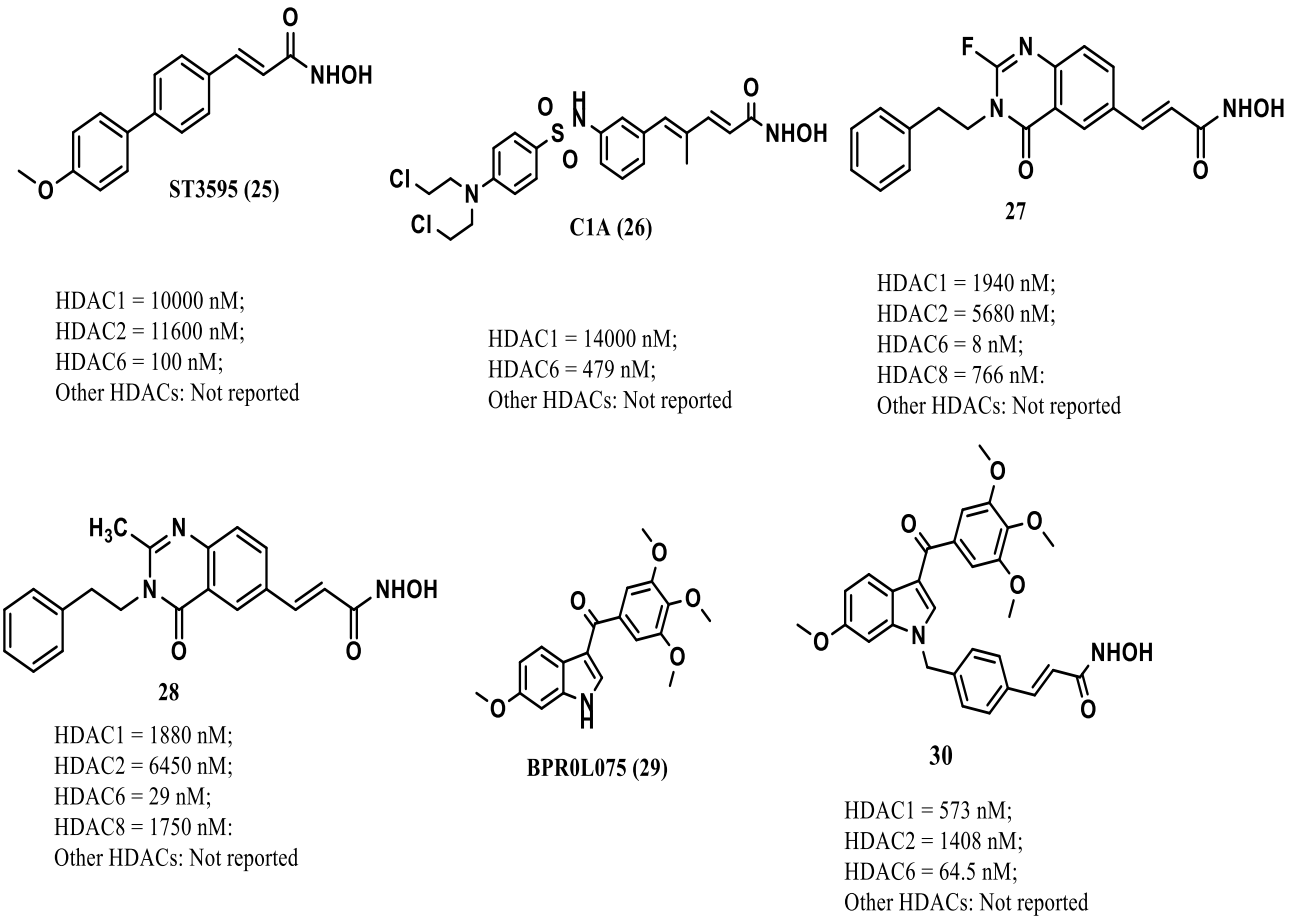


Figure 2.18: Some N-hydroxycinnamamide based HDAC6 inhibitors with their respective IC50 values.

BPR0L075 (**29**) is a novel synthetic compound initially discovered as part of research aimed at finding new microtubule inhibitors.[231] It binds to the colchicine-binding site on tubulin, inhibiting its polymerization, and shows significant anti-proliferative activity with IC50 values in the single-digit nanomolar range across various human cell lines. Structural modifications of BPR0L075, including the addition of an N-hydroxycinnamamide group at the N1 position, led to the development of 3-aroylindole hydroxamates as HDAC6 inhibitors. Among these, compound **30** emerged as the most potent, demonstrating selectivity for HDAC6 while also exhibiting tubulin inhibitory activity.[232] Additionally, compound **30** shows remarkable anti-proliferative effects in vitro and effectively inhibit the growth of multiple myeloma xenografts in vivo.[233]

2.6.4 HDAC6 inhibitors with novel ZBG

HDAC6 inhibitors are commonly utilized for treating central nervous system disorders. However, hydroxamates, a prevalent class of HDAC6 inhibitors, are linked with

genotoxicity. To address this issue, researchers have explored a range of novel zinc-binding groups (ZBGs) such as mercaptoacetamide, thiols, trifluoromethylketone, hydrazides, hydroxypyridones, hydroxylpyridine-thiones, and hydroxy ketones. These alternative ZBGs aim to develop more selective and potent HDAC6 inhibitors that are free from the toxicity associated with hydroxamates.

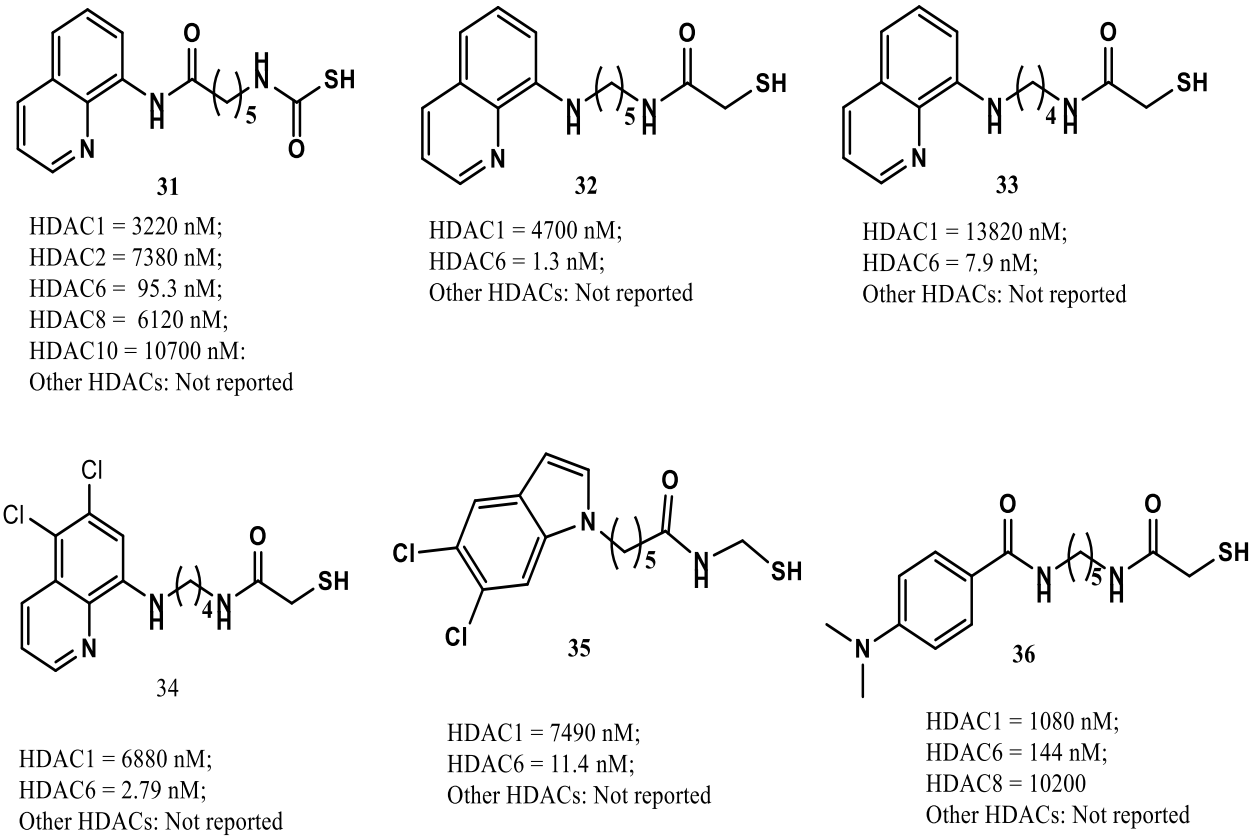
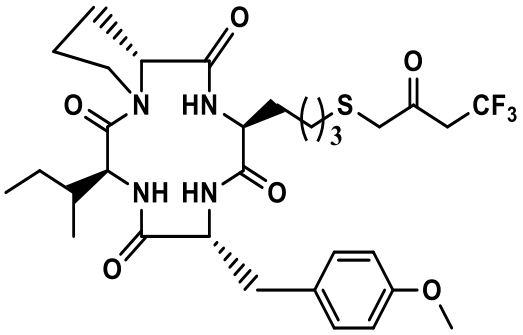


Figure 2.19: Some mercaptoacetamide based HDAC6 inhibitor with quinoline cap group.

Kozikowski et al. developed a series of HDAC inhibitors based on mercaptoacetamide, with quinoline emerging as an effective cap group. Compound **31** demonstrated significant affinity for HDAC6 (IC50 = 95.3 nM) and exhibited 33-fold and 64-fold greater selectivity over HDAC1 and HDAC8, respectively. However, substituting the quinoline moiety with other groups (compound **36**) resulted in decreased potency against HDAC6. Some of these mercaptoacetamide-based inhibitors also showed protective effects on cortical neurons subjected to oxidative stress, though they exhibited dose-dependent toxicity.[234] Additionally, compound **31** was found to influence the levels of the amyloid precursor protein (APP), Aβ synthase, and Aβ degradation enzymes.[235] Further structural modifications led to compound **32**, which

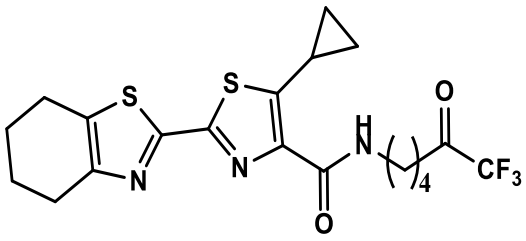
displayed enhanced potency against HDAC6 (IC50 = 1.3 nM). Compound **32** effectively induced acetylation of α-tubulin in a dose-dependent manner without affecting histone H3 acetylation levels.[236] Structure-activity relationship studies indicated that compounds with alkyl chain linkers of varying lengths had strong inhibitory potency against HDAC6, while benzyl linkers drastically reduced potency. To enhance lipophilicity and blood-brain barrier penetration, different halogen substitutions were made on the quinoline cap, resulting in compound **33**, a potent mercaptoacetamide-based HDAC6 selective inhibitor. Various analogues, including compounds **34** and **35**, exhibited high affinity and selectivity for HDAC6 over HDAC1.[237]

Trifluoromethyl ketone is another novel zinc-binding group (ZBG), but research has shown that it does not effectively achieve selectivity for HDAC6. For example, compound **37**, which uses a cyclic tetrapeptide as the cap group and trifluoromethyl ketone as the ZBG, demonstrated better inhibition of HDAC1 (IC50 = 47 nM) compared to HDAC6 (IC50 = 180 nM).[238] Similarly, compound **38**, which also employs trifluoromethyl ketone as the ZBG, exhibited comparable inhibition for both HDAC1 (IC50 = 19.38 nM) and HDAC6 (IC50 = 17.78 nM).[239]



37

HDAC1 = 47 nM; HDAC4 = 190 nM;
HDAC6 = 180 nM; HDAC8 = 230 nM;
Other HDACs: Not reported



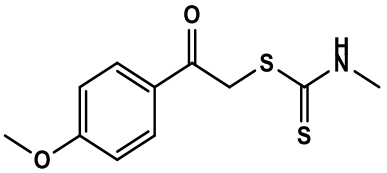
38

HDAC1 = 19.38 nM; HDAC3 = 20.79 nM;
HDAC4 = 23.42 nM; HDAC6 = 17.78 nM;
Other HDACs: Not reported

Figure 2.20: some Trifluoromethyl ketone based HDAC6 inhibitors.

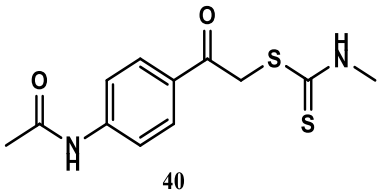
Dehmel et al. discovered several series of thiocarbonate analogues as highly effective and substrate-competitive HDAC6-selective inhibitors.[240] In their initial series, compounds with para-methoxy benzene as the cap group and various thiocarbonates as the zinc-binding group (ZBG) exhibited notable HDAC inhibitory activity. Among

these, compound **39**, a dithiocarbonate derivative, showed substantial affinity for HDAC6 (IC50 = 94 nM), though there was potential for further improvement. Substituting the phenylacetyl-bound sulfur in the headgroup with nitrogen (compound **40**), methylene (compound **41**), or oxygen (compound **42**) resulted in significant enhancements in HDAC6 inhibitory activity, with IC50 values of 18 nM, 23 nM, and 90 nM, respectively.



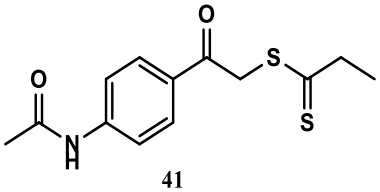
39

HDAC1 = 215 nM;
HDAC6 = 94 nM;
Other HDACs: Not reported



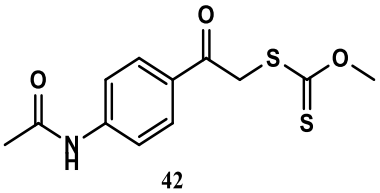
40

HDAC1 = 141 nM;
HDAC6 = 18 nM;
Other HDACs: Not reported



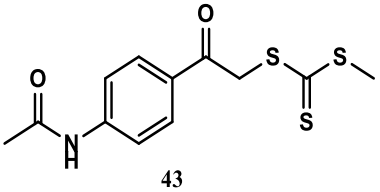
41

HDAC1 = 164 nM;
HDAC6 = 23 nM;
Other HDACs: Not reported



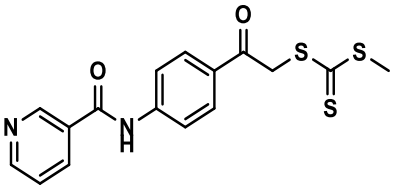
42

HDAC1 = 536 nM;
HDAC6 = 90 nM;
Other HDACs: Not reported



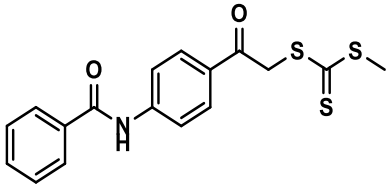
43

HDAC1 = 1680 nM;
HDAC6 = 180 nM;
Other HDACs: Not reported



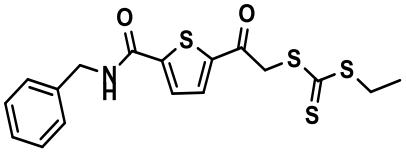
44

HDAC1 = 388 nM;
HDAC6 = 30 nM;
Other HDACs: Not reported



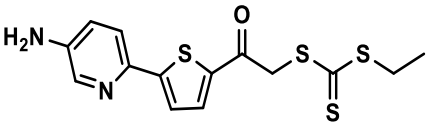
45

HDAC1 = 120 nM;
HDAC6 = 65 nM;
Other HDACs: Not reported



46

HDAC1 = 340 nM;
HDAC6 = 44 nM;
Other HDACs: Not reported



47

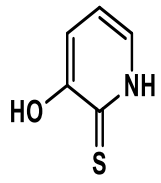
HDAC1 = 831 nM;
HDAC6 = 43 nM;
Other HDACs: Not reported

Figure 2.21: HDAC6 inhibitors with thiocarbonate as ZBG, and their respective IC_{50} values against different HDAC isoforms

However, the trithiocarbonate derivative (compound **43**) demonstrated a decrease in inhibitory potency against HDAC6. Subsequently, Dehmel et al.[240] investigated trithiocarbonates with phenylacetyl as the core structure, incorporating large amide

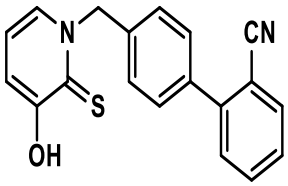
substituents to enhance HDAC inhibitory activity. In their third series of compounds, the replacement of the phenylacetyl moiety (compounds **44** and **45**) led to improved inhibitory potency against HDAC6 compared to compound **43**. Further modifications, substituting the benzene ring with a 2,5-disubstituted thiophen core in trithiocarbonates, produced compounds **46** and **47**. These compounds demonstrated strong HDAC6 inhibitory activity, with compound **47** in particular showing excellent selectivity for HDAC6 over HDAC1.

Patil et al. discovered 3-hydroxypyridin-2-thione (3-HPT, **48**) as a novel zinc-binding group (ZBG) that effectively inhibits HDAC6 with an IC50 of 681 nM, while showing minimal activity against HDAC1.[241] Further structural optimization led to compound **49**, a 3-HPT-derived HDAC inhibitor with enhanced inhibitory activity against HDAC6. These compounds also demonstrated significant growth inhibition of Jurkat cells and induced apoptosis in various cancer cell lines. Compared to 3-HPT, 1-hydroxypyridine-2-thione (1-HTP) proved to be more effective in achieving HDAC6 inhibitory potency and selectivity. Specifically, 1-HTP-6-carboxylic acid (compound **50**) showed HDAC6 inhibition with an IC50 value of 150 nM and 286-fold selectivity over HDAC1. The high potency, selectivity, and good metabolic stability of 1-HPT suggest that these molecules hold potential for further development in leukaemia treatment.[242]



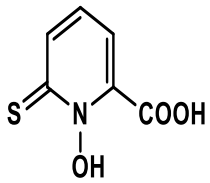
48

HDAC1: No inhibition
HDAC6 = 681 nM;
HDAC8 = 3675 nM;
Other HDACs: Not reported



49

HDAC1: No inhibition
HDAC6 = 372 nM;
HDAC8 = 1907 nM;
Other HDACs: Not reported



50

HDAC1 = 43000 nM; HDAC6 = 150 nM;
HDAC2 = 710000 nM; HDAC7 = 29000 nM;
HDAC4 = 71000 nM; HDAC8 = 3300 nM;
HDAC5 = 35000 nM; HDAC9 = 47000 nM;
Other HDACs: Not reported

Figure 2.22: HDAC6 inhibitors with other novel ZBG along with their respective IC_{50} values against different HDAC isoforms.

2.7 Anti-HDAC6 therapy in clinical trials

Clinically approved HDAC inhibitors like vorinostat (SAHA), romidepsin, belinostat, Panobinostat (which has been withdrawn), and chidamide are used for treating T cell lymphomas and multiple myeloma. However, their poor isoform selectivity can lead to significant toxicity and side effects, which limits their clinical use.

For example, during clinical trials of Panobinostat in high-risk MDS or AML patients following allogenic stem cell transplantation, 52% of participants experienced at least one severe (Grade 3 or 4) adverse event (AE) attributed to Panobinostat. The most common severe AEs were thrombocytopenia (24%) and neutropenia (19%).[243] These findings were corroborated by another Phase Ia/II clinical trial of Panobinostat, where Grade 3/4 adverse effects included thrombocytopenia (41.5%) and neutropenia (21%).[244] Similarly, the clinical use of vorinostat is also restricted due to its high rate of adverse effects;[210,245] for instance, when used in combination with bortezomib, 16% of patients experienced Grade 3/4 diarrhea, and 22% had Grade 4 thrombocytopenia.[245]

HDAC6 inhibitors (HDAC6is) are a focus of clinical research due to the unique structure of HDAC6. Currently, five HDAC6is are in clinical trials for various cancers, autoimmune diseases, and peripheral pain, but none have yet advanced for neurodegenerative diseases. Clinical trials conducted on various selective HDAC6 inhibitors are summarized in table 2.4.

Two notable HDAC6is, ACY-1215 (Ricolinostat) and ACY-241 (Citarinostat), developed by Acetylon Pharmaceuticals, are undergoing extensive clinical trials. Both share a similar structure with a long-chain hydroxamate scaffold and a large, rigid N,N-diphenyl 2-aminopyrimidine recognition group. They exhibit nanomolar potency against HDAC6 (ACY-1215: IC50 = 4.7 nM; ACY-241: IC50 = 2.6 nM) and demonstrate 12 to 13-fold selectivity over HDAC1.[211,246] In a Phase Ib clinical trial involving 38 patients with relapsed or refractory multiple myeloma, ACY-1215, administered at 160 mg daily for 21 days of a 28-day cycle, combined with lenalidomide and dexamethasone, achieved a 55% response rate with minimal adverse effects.[206] Additionally, ACY-1215 has been reported to mitigate nerve damage and alleviate pain, numbness, and muscle weakness caused by chemotherapy and chemoradiotherapy.[246-248]

KA2507, another HDAC6i developed by Karus Therapeutics, was under Phase II trials for biliary tract cancer but has since been withdrawn.[249] Chong Kun Dang Pharmaceutical Corp. (CKD) in Korea developed CKD-504 (structure not disclosed), which is currently in Phase I trials to evaluate its pharmacokinetics, pharmacodynamics, and toxicity in healthy adults, with potential applications for Huntington's disease. Additionally, CKD-506, a highly selective HDAC6i with at least 400-fold selectivity over other HDAC isoforms, has shown significant effects in mouse models of systemic lupus erythematosus (SLE) and chronic autoimmune diseases with minimal adverse effects.[250] This compound is undergoing Phase I studies in Europe and a Phase II trial for rheumatoid arthritis.[251]

JBI-802, a dual LSD1/HDAC6 inhibitor identified by Jubilant Therapeutics Inc., with IC50 values of 50 nM for LSD1 and 11 nM for HDAC6 and more than 100-fold selectivity over other HDAC isoforms, is currently in Phase II trials for solid tumours.[252]

Table 2.4: Clinical Trials conducted on various selective HDAC6 inhibitors

NCT Number	Agent	Other Agents	Inclusion	Phase	Start	End	Enrolment
NCT02400242	ACY-241	Pomalidomide Dexamethasone	Multiple Myeloma	I	07.05.2015	02.07.2023	85
NCT02935790	ACY-241	Nivolumab Ipilimumab	Malignant Melanoma	I	30.09.2016	07.04.2017	1
NCT02551185	ACY-241	---	Advanced Solid Tumors	I	22.12.1015	04.10.2019	20
NCT02635061	ACY-241	Nivolumab	Non-Small Cell Lung Cancer	I	25.08.2016	30.06.2023	16
NCT02091063	ACY-1215	---	Lymphoma Lymphoid Malignancies	I/II	02.04.2014	05.05.2019	24
NCT02632071	ACY-1215	Nab-paclitaxel	Metastatic Breast Cancer Breast Carcinoma	I	01.03.2016	30.09.2020	17
NCT01997840	ACY-1215	Pomalidomide Dexamethasone	Multiple Myeloma	I/II	01.03.2014	30.06.2023	103
NCT01323751	ACY-1215	---	Multiple Myeloma	I/II	7.2011	03.12.2016	120
NCT02189343	ACY-1215	Pomalidomide Dexamethasone	Multiple Myeloma	I	15.09.2014	30.04.2018	16
NCT02787369	ACY-1215	Ibrutinib Idelalisib	Recurrent Chronic Lymphoid leukemia	I	May-16	Apr-26	3
NCT01583283	ACY-1215	Lenalidomide Dexamethasone	Multiple Myeloma	I	12.07.2012	24.03.2021	38
NCT02088398	ACY-1215	---	Healthy	I	Mar-14	Apr-14	19

NCT03176472	ACY-1215	---	Painful Diabetic Peripheral Neuropathy	II	07.12.2020	15.02.2023	282
NCT02856568	ACY-1215	Cisplatin Gemcitabine Hydrochloride	Non-Resectable Cholangiocarcinoma Recurrent Cholangiocarcinoma Stage III Extrahepatic Bile Duct Cancer Stage III Intrahepatic Cholangiocarcinoma Stage IIIA Hilar cholangiocarcinoma Stage IIIB Hilar cholangiocarcinoma Stage IVA Extrahepatic Bile Duct Cancer Stage IVA Hilar cholangiocarcinoma Stage IVA Intrahepatic Cholangiocarcinoma Stage IVB Extrahepatic Bile Duct Cancer	I	01.05.2017	Oct-21	0
NCT05193851	ACY-1215	---	Peripheral Nervous System Disease	I	12.01.2022	06.01.2023	12
NCT05229042	ACY-1215	---	Peripheral Nervous System Disease	I	01.12.2022	28.04.2024	57
NCT02661815	ACY-1215	Paclitaxel Bevacizumab	Ovarian Cancer Fallopian Tube Cancer Primary Peritoneal Carcinoma	I	15.06.2016	28.07.2017	6
NCT03713892	CKD-504	---	Huntington Disease	I	23.05.2018	Dec-20	88

NCT04186156	KA2507	---	Biliary Tract Cancer	II	05.03.2020	Oct-23	0
NCT03008018	KA2507	---	Solid Tumor	I	07.08.2017	10.06.2020	20
NCT04204603	CKD-506	---	Rheumatoid Arthurites	II	30.11.2018	29.10.2019	122
NCT05268666	JBI-802	---	Locally Advanced Solid Tumor Metastatic Solid Tumor	I/II	08.04.2022	Aug-25	126

Chapter 3

Present work & the rationale Behind the work

Previously in the chapter 1 & 2 I have discussed about various computational techniques developed within the last decades for drug designing & repurposing and why HDAC6 is a valuable target for drug target. Despite of having such cutting-edge computational techniques and understanding of the crucial physiological role HDAC6 plays no inhibitor has made it to the market. This present work tries to bridge the gap between understanding and reality. To do that some of these computational techniques was applied to learn more deeply about the structural requirements of a nearly perfect HDAC6 inhibitor, both from the ligand perspective and the target perspective. To explore the ligand perspective a combination of machine learning algorithms was used to learn about non-linear patterns. And to understand how these ligands (inhibitors) and target (HDAC6) interact molecular docking and molecular dynamics simulation was performed. This knowledge might be helpful in designing new generation of HDAC6 inhibitors, or can be used as a screening tool in the vast chemical space.

Experimental

Chapter 4

Materials and Methods

4.1 Dataset preparation

To successfully test if any marketed drug has the potential to bind HDAC6, we need to define a hypothesis based on known ligands. In other words, firstly we need to learn about the structural attributes to the HDAC6 inhibitors. To do that a dataset of 142 quinazoline-containing, hydroxamate based HDAC6 inhibitors, exhibiting a wide range of HDAC6 inhibitory activities, was compiled from the literature (**Appendix Table S1**).[253-258] The structures were drawn and saved in an appropriate format using ChemDraw Ultra software (Cambridge soft corporation, U.S.A.). For structural refinement of these 142 hydroxamate-based HDAC6 inhibitors "*Prepare ligand for QSAR"* protocol in Discovery Studio 3.0[259] was used.

4.2 Descriptor generation and dataset division

Descriptors lies at the heart of QSAR, as they represent structural and physiochemical qualities or properties of molecules.[260,261] Having a set of informative descriptors are of outmost importance to build robust QSAR models. In this study, for machine learning-based regression analysis, we used 1,444 two-dimensional molecular descriptors and 12,775 fingerprint descriptors calculated by PaDEL descriptor software to represent the molecules. These descriptors include constitutional, physicochemical, thermodynamic, and topological indices, as well as PubChem Fingerprint (881 bits), Substructural Fingerprint (307 bits), Substructural Fingerprint Count (307 bits), Klekota-Roth Fingerprint (4,860 bits), Klekota-Roth Fingerprint Count (4,860 bits), Atom-Pairs 2D Fingerprint (780 bits), Atom-Pairs 2D Fingerprint Count (780 bits), and others.[262] The pool of 14,219 calculated descriptors was pre-processed using V-WSP 1.2 software to remove highly correlated features and features with no variance. A correlation cut-off value of 0.90 and a covariance threshold of 0.001 were applied, resulting in a subset of 892 descriptors.[263]

To conduct the classification-based QSAR study, the dataset molecules were classified as active or inactive using a threshold pIC50 of 7.0 (IC50 = 100 nM), based on the average pIC50 value of the total dataset molecules (Avg pIC50 = 7.231). Since dataset division is crucial for the development and validation of any QSAR model, several fundamental molecular properties were considered to achieve a balanced division. These properties included lipophilicity (AlogP), molecular weight (Mw), molecular

polar surface area (MPSA), number of hydrogen bond donors and acceptors (nHBA, nNBD), number of atoms (nAtoms), number of rings (nR), number of aromatic rings (nAr), and number of fragments (nFrag). The "generate training and test data" protocol of Discovery Studio 3.0[259] was used to split the dataset using the random per cluster (RPC) method, maintaining a 75:25 ratio for the training and test sets.[259] The same training and test sets were used to construct both the ML-based QSAR and classification-based QSAR models.

Further the marketed drugs data (3816 molecules) was collected from the CIMS database and were saved in appropriate format for future use.[264]

4.3 Feature selection and model development

In most machine learning techniques, feature selection is a crucial step as it reduces the risk of overfitting due to noisy, redundant descriptors and increases the interpretability and understanding of the resulting models.[265] To achieve this, having a good selection criterion that can measure the relevance of each descriptor to the desired output is essential. In this study, lasso regression was applied to the initial set of 892 pre-processed descriptors using "Lasso-CV_R_Feature_Selection_UI_v0.24.2.R," an R-based in-house software, to identify descriptors of high significance.[266] Finally, the 25 descriptors selected by lasso regression were subjected to the best subset selection method using "BestSubsetSelection_v2.1," with an r^2 value of >0.6 and q^2 value of >0.5, to correlate and extract the final set of features for the training set molecules.[267]

4.4 Development of QSAR models

In this study I have used four different machine learning algorithms[268,269] for regression analysis and three different calcification methods to learn about the crucial structural contributors of these HDAC6 inhibitors.

4.4.1 k-nearest neighbour (k-NN)

K-nearest neighbour (k-NN) is a non-linear, non-parametric, instance-based machine learning approach that predicts a ligand's bioactivity as the distance-weighted average of the bioactivity of its k nearest neighbours.[268-270] The k-NN method generalizes the 1-NN rule proposed by Cover and Hart in 1967, which is based on the principle that similar compounds exhibit similar properties.[271] The performance of k-NN depends on the chosen k value and the distance function used. In this study, similarity between compounds was determined using the Euclidean distance in a multidimensional

descriptor space. Since the Euclidean distances between a compound and its k nearest neighbours are not identical, the neighbour closest to the compound is given a higher weight in calculating the predicted activity. The equation used for prediction is:

$$\hat{y}i = \frac{\sum_{j=1}^{k} y_j w_{ij}}{\sum_{j=1}^{k} w_{ij}}$$

where y_i is the observed activity value for the nearest neighbour j, $\widehat{yi}$ is the predicted activity value for compound i, and w_{ij} are the weights defined as:

$$w_{ij} = \left(1 + \frac{d_{ij}^2}{\sum_{j'=1}^{k} d_{ij}^2}\right)^{-1}$$

where d_{ij} represents the Euclidean distance between compound i and its k nearest neighbour. The value of k (ranging from 1 to n−1, where n is the number of compounds in the dataset) is optimized during the model-building process to provide the best prediction for the training set. For test data prediction, the same equation and k value are used. The k-NN model was developed using WEKA 3.8 software.

4.4.2 Random Forest (RF)

Random Forest (RF) is a supervised, parallel, ensemble-based machine learning method developed by Breiman that uses decision trees as classifiers.[272,273] The fundamental concept of this algorithm is to build a collection of independent decision trees from random samples of the training data. When a new input sample is introduced, each tree provides a prediction, and the final output is determined by averaging the predictions of all the decision trees using the following equation:

$$\hat{y}(x) = \frac{1}{n_{tree}} \sum_{i=1}^{n_{tree}} \hat{f}\mathrm{i}\,(x)$$

where $\widehat{f_i}(x)$ represents the predicted output of the i^{th} tree, and $\widehat{f_i}(x)$ is the predicted response value by the RF model. The key advantage of the Random Forest algorithm over other machine learning techniques, such as artificial neural networks, support vector machines, or linear discriminant analysis, is its robustness against overfitting. The model tends to converge as the number of trees in the forest increases. During model building with WEKA 3.8, parameters such as the number of regression trees

(n_{tree}), the number of randomly selected descriptors at each node, and the number of training samples in each terminal node (node size) were optimized to achieve the best Random Forest model.

4.4.3 Artificial neural network (ANN)

Artificial Neural Networks (ANNs)[274,275] are supervised machine learning techniques inspired by the complex neuronal circuitry of the human brain, capable of modelling multidimensional, non-linear relationships between inputs and outputs. There are various types of ANNs, including feed-forward backpropagation networks (BP-NN), radial basis function networks, and probabilistic neural networks.[276,277] In this study, a feed-forward backpropagation network was employed, consisting of three layers: input, hidden, and output layers, which are connected unidirectionally through coefficients known as weights (artificial synapses). This architecture computes a numerical output for a given input vector. The input data (In) is processed in each neuron as follows:

$$Z_j = \sum_{i}^{n} W_{ij} A_i$$

where Z_j represents the value of the j^{th} hidden neuron, W_{ij} is the weight linking the i^{th} input neuron to the j^{th} hidden neuron, and A_i is the normalized value of the i^{th} input variable.

In the ANN algorithm, input and output values are rescaled to a range from -1 to +1 using the following formula:

$$A_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \times (r_{max} - Xr_{min}) + r_{min}$$

where X_i is the i^{th} real variable, A_i is the normalized value of X_i, X_{min} and X_{max} are the minimum and maximum values of X_i, and r_{min} and r_{max} are the target range limits for scaling.

During each iteration of the BP-NN algorithm, the weights are adjusted to minimize the difference between the actual and predicted outputs. The weight change can be described by:

$$\Delta W_{ij} + W_{ij} \rightarrow W_{ij}$$

$$\Delta W_{ij} = \eta(t - o) In_i$$

where *t* and *o* are the target and output values of the ANN for each sample, η is the learning rate that controls the amount of weight change at each iteration, and ΔW_{ij} denotes the change in weight W_{ij}.

4.4.4 Support vector machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm designed for classification and regression tasks, developed by Vapnik and colleagues based on statistical learning theory.[278] SVM works by constructing a hyperplane in a multidimensional feature space that maximizes the margin between two classes. For a sample data point (*xi*, *yi*), where $i=1,\ldots n$, $x \in R^d$, and $y \in (\pm 1)$, the equation for the optimal hyperplane is $w \cdot x + b = 0$. The margin, which is the distance between the hyperplane and the nearest data points, is given by $2/\|w\|$. Finding the optimal hyperplane can be framed as solving the following convex quadratic programming (QP) problem:

$$min \frac{1}{2} \| w \|^2,$$

Subject to: $y_i [w.x + b] \geq 1$

Support Vector Regression (SVR) extends SVM to regression problems by finding a hyperplane that minimizes the distance to all data points, as opposed to maximizing the margin in classification. For linear regression, the goal is to find an optimal hyperplane that estimates y with an ϵ-insensitive loss function, meaning the distance from the hyperplane to any data point is less than ϵ. This problem is framed as minimizing the model's complexity, equivalent to minimizing $\frac{1}{2} \| w \|^2$.. The corresponding quadratic programming problem is:

$$min \frac{1}{2} \| w \|^2$$

$$\text{Subject to: } \begin{cases} [\, y_i - w.x - b] \leq \varepsilon \\ [w.x + b - y_i] \leq \varepsilon \end{cases}$$

For nonlinear regression, SVM employs a kernel function to map the sample data into a higher-dimensional feature space.[279,280] The study used the Radial Basis Function (RBF) kernel to uncover nonlinear relationships in the data using WEKA 3.8 software. The RBF kernel is expressed as:

$$k\ (x, x_i) = exp(-\gamma \|x - x_i\|^2)$$

where k is the kernel function, x and x_i are vectors, and γ is a hyperparameter.

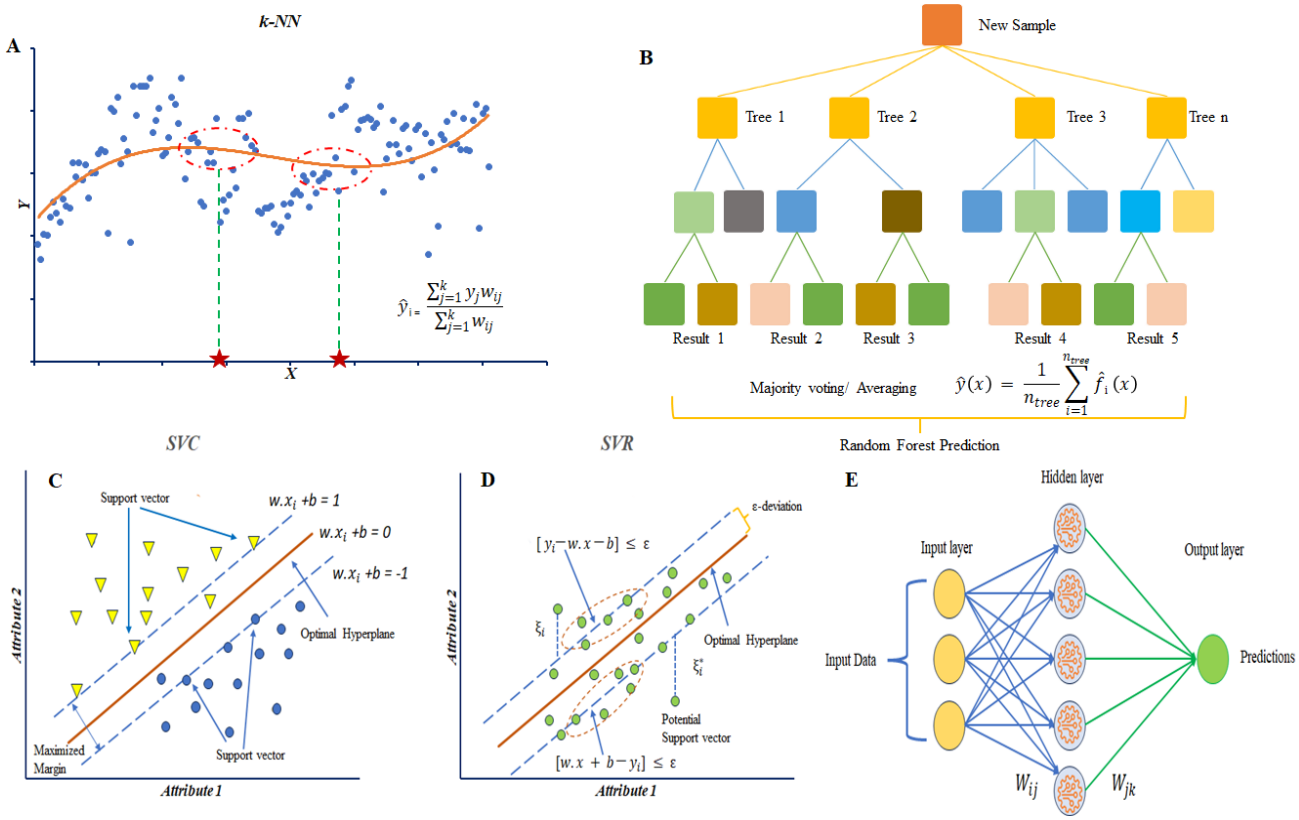


Figure 4.1: Graphical schematic representation of **(A)** *k*-Nearest neighbour (*k-NN*); (B) Random Forest (*RF*); **(C)** Support vector classifier (*SVC*); **(D)** Support vector regressor (SVR); **(E)** Artificial neural network (*ANN*).

4.4.5 Bayesian classification study

Bayesian classification employs Bayes' theorem,[281] formulated by Thomas Bayes in the 18th century, to calculate the probability of an event based on two factors: a prior probability and a likelihood function derived from a probability model of the observed data. Mathematically, it is expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Here, A represents the hypothesis or model, and B denotes the observed data. *P(A)* and *P(B)* are the prior belief and the evidence, respectively. *P(A/B)* is the posterior probability, and *P(B/A)* signifies the likelihood.

In this classification study, several fundamental molecular properties were calculated using Discovery Studio 3.0[259] and used as independent variables to construct the Bayesian classification model. These properties include lipophilicity (AlogP), molecular weight (MW), the number of hydrogen bond donors (nHBD) and acceptors (nHBA), the number of rotatable bonds (nRB), the number of rings (nR), and the number of aromatic rings (nAR), along with the molecular fractional polar surface area (MFPSA). Additionally, a topological descriptor, the atom-type extended connectivity fingerprint of diameter 6 (ECFP_6), was employed for fragment-based structural assessment in this Bayesian classification study.

4.4.6 Recursive partitioning study

Recursive partitioning,[281] also known as decision tree analysis, is a statistical learning algorithm used to predict an outcome based on a set of covariates (predictors). Introduced by Morgan and Sonquist in 1963 with the automatic interaction detection (AID) algorithm,[283] the main idea behind recursive partitioning is to group individuals with similar outcomes using the covariate set. Given an outcome and a covariate, the decision tree splits the data into nodes based on each unique value of each covariate, aiming to make the outcome homogeneous within each node and heterogeneous between nodes until a stopping criterion is met. This stopping criterion can depend on cross-validated fit indices (e.g., entropy, mean squared error), node size (e.g., the number of participants in a node), and minimum improvement in prediction accuracy.

In this study, the recursive partitioning (RP) model was developed using the *"Create Recursive Partitioning Model"* protocol in Discovery Studio 3.0.[259] The model incorporated the same fundamental molecular properties used in the Bayesian classification model, along with a topological descriptor called the functional class fingerprint of diameter 6 (FCFP_6), to classify the dataset molecules into different clusters. The decision tree was constructed using an entropy-based split method and a minimum sample per node of 10 as the stopping criterion.

4.4.7 SARpy analysis

SARpy, or Structure Activity Relationship in Python, is a QSAR approach that employs "String mining" to transparently discover relevant molecular fragments and derive rules directly from the data without prior knowledge.[282,284] The process of selecting an active ruleset involves three steps:

1. Fragmenting chemicals of arbitrary complexity to extract all possible substructures within a user-defined size range.

2. Validating the predictive power of each fragment by analysing the correlation between the occurrence of each molecular substructure and the experimental activity of the compounds containing the fragment.

3. Selecting the most predictive fragments and listing them as rules in the format "IF contains Structural alert THEN apply activity label."

The Likelihood Ratio (LR) is calculated as:

$$Likelihood\ Ratio\ (LR) = \frac{True\ positive}{False\ positive} \times \frac{Negatives}{Positives}$$

In this study, SARpy analysis was conducted with OPTIMAL precision using SARpy software developed by Gini and colleagues.[285] The analysis aimed to identify active structural alerts or rules for atoms numbering 2-18, with a minimum of 5 occurrences in the given dataset.

4.5 Evaluation of QSAR models

According to the OECD principle 2004,[286] validating any QSAR model is essential to confirm its reliability and predictivity. In this study, the performance and learning capability of machine learning models were assessed using the squared correlation coefficient (R2) and root mean square error (RMSE). Leave-one-out (LOO) cross-validation and mean absolute error (MAE) were also employed. For classification-based models, a Receiver Operating Characteristic (ROC)-based statistical evaluation was conducted to assess performance and predictivity. Additionally, to validate the performance of SARpy, Bayesian, and RP models, metrics such as sensitivity (Se), specificity (Sp), precision (Pr), and accuracy (Acc) were analysed. Statistical properties like Matthew's correlation coefficient (MCC), F1-measure, and balanced accuracy (AUCb) were also considered. Table 4.1 provides the mathematical descriptions and significance of these statistical validation parameters.

4.6 Molecular Docking and Molecular Dynamics (MD) simulation-based binding pattern analysis

Molecular docking is a computational technique that samples conformations of small molecules in protein binding site in order to predict the affinity and binding pattern of the small molecules through geometric and energy matching and recognition.

In the current sicario to conduct the molecular docking study, the X-ray diffraction solution structure of HDAC6 (PDB ID: 5EDU) was collected from the Protein Data Bank (PDB).[287] Initially, in order to prepare and optimize the protein molecule, as well as to add any missing hydrogens, state generation and structural refinement of the protein the "Protein Preparation Wizard" of Schrodinger Maestro v12.5[288] was utilised. During this process, the OPLS_2005 force field was chosen to perform restrain minimization of the protein structure. The "Receptor Grid Generation" wizard of Schrodinger Maestro v12.5 software was used to generate the receptor grid for molecular docking study. Subsequently, the "*Ligprep*" module present in maestro V12.5 was used to prepare the molecules before docking studies. Finally, the molecular docking study was conducted with *extra precision (XP)* method utilizing the *GLIDE* module of Schrodinger Maestro V 12.5.[269,289] Ten poses were generated for each ligand during docking process and the best poses were selected based on the docking score values and best orientation of the molecule in the active site of HDAC6.

For the MD simulation study, the best-docked poses of the compounds were selected based on their docking scores and binding site interactions with HDAC6 (PDB ID: 5EDU). Each HDAC6 (PDB ID: 5EDU)-docked compound complex underwent pre-processing and het-state generation using the Protein Preparation Wizard in Schrodinger Maestro v12.5 software.[288] During protein pre-processing, hydrogens were added, bond orders assigned, and co-crystallized water molecules removed. The Epik module of Schrodinger Maestro v12.5 software was employed for het-state generation at a pH of 7.0 (± 2.0).[288] Subsequently, the complex was optimized, and energy minimization was carried out using the OPLS_2005 force field, with the convergence of heavy atoms to an RMSD of 0.30 Å. The System Builder wizard in Schrodinger Maestro v12.5 software was then used to develop the simulation system. A cubic box with a 10 Å buffer distance between the system and the box boundary was chosen. The TIP3P solvent system was selected, and isotonic conditions were maintained by adding

Table 4.1: The mathematical description and significance of statistical validation parameters used for QSAR model development.

Parameter	*Description*	*Equation*	*Significance*
R^2	Squared correlation coefficient	$R^2 = 1 - \sum \frac{(y_{obs} - y_{pred})^2}{(y_{obs} - y_{mean})^2}$	Metric to check the goodness-of-fit of a regression model by measuring the variation of observed data with the predicted once
$\boldsymbol{Q^2_{Loo}}$	Leave-one-out cross-validation	$Q^2_{LOO} = 1 - \frac{\sum(y_{obs(training)} - y_{pred(training)})^2}{\sum(y_{obs(training)} - \bar{y}_{(training)})^2}$	Cross-validated $R^2(Q^2)$ is checked for internal validation
RMSE	Root means square error	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{n}}$	It gives a measure of model external validation. A lower value signifies good external predictivity.
MAE	Mean absolute error	$MAE = \frac{1}{n} \times \sum \lvert y_{obs} - y_{pred} \rvert$	Ensures that the trained model has no outlier prediction with huge errors
SE	Sensitivity	$SE = \frac{TP}{(TP + FN)}$	Indicates the accuracy of real prediction
SP	Specificity	$SP = \frac{TN}{(TN + FP)}$	Used to calculate the false positive rate

PR	Precision	$PR = \frac{TP}{(TP + FP)}$	Indicates the accuracy of predicted class
Q	Accuracy	$Q = \frac{(TP + TN)}{(TP + TN + FP + FN)}$	
MCC	Matthew's correlation coefficient	$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	Used to measure the quality of binary classification
F_1	F-measure value	$F1 = \frac{2TP}{(2TP + FP + FN)}$	Indicates the harmonic mean of recall
AUC_b	Balanced accuracy	$AUC_b = \frac{(SE + SP)}{2}$	

Na+ and Cl^- ions at a concentration of 0.15 M. Additionally, 5 Na+ ions were included to neutralize the system charges, with the OPLS_2005 force field applied.[269]

The Desmond module in Schrodinger Maestro v12.5[36] software was used to conduct 100 ns MD simulation studies on the docked most active and least active compounds, along with HDAC6-bound Trichostatin A (TSA, PDB ID: 5EDU), using the OPLS_2005 force field. Each 100 ns MD simulation study was carried out under NPT ensemble conditions at a temperature of 37°C (310.15K) and a pressure of 1.01325 bar. Additionally, a pre-simulation system relaxation was performed using a 2.0 fs RESPA integrator, a Nose-Hoover chain thermostat with a relaxation time of 1.0 ps, and a Martyna-Tobias-Klein barostat with a relaxation time of 2.0 ps.

Additionally, the stability of the complex was verified in terms of conventional root mean square deviation (RMSD), fluctuation (RMSF) of amino acid residues, the radius of gyration (Rg) of the main chain, *Prime*[290] Molecular Mechanics-Generalized Born Surface Area (MM-GBSA)[291], and principal component analysis (PCA)-based free energy landscape (FEL).[29]

Chapter 5

Result and Discussion

5.1 Feature Selection

The feature selection process was conducted using the training set compounds, and the training and test data distributions were validated through logP and molecular weight-based principal component clustering, as well as fundamental molecular property-based t-SNE distribution (Fig. 5.1A and 5.1B, respectively). The main objective of feature selection is to identify the most relevant and significant features while preventing data overfitting caused by noisy, redundant descriptors. In this study, a LOO-cross-validated lasso regression-based feature selection procedure was employed to identify highly important molecular descriptors (Fig. 5.1C and 5D) (Appendix Table S2). This was followed by best subset selection to filter out the most relevant descriptors for the dataset molecules (Appendix Table S3). The reliability and robustness of these selected features were further evaluated through the development of MLR models (Appendix Table S4).

5.2 Machine learning model optimization

Optimizing the learning parameters of machine learning (ML) models is crucial for their performance. In this study, the k-NN model was constructed by optimizing the number of neighbours, while the RF models were optimized by adjusting the number of trees using WEKA 3.8 software. For the SVM model, parameters such as kernel complexity (C), kernel width (γ), and the epsilon value for the radial basis function were optimized using AUTOWEKA 1.0 software. The same AUTOWEKA 1.0 software was also used to optimize the learning rate, learning momentum, number of hidden layer nodes, and epoch parameters for the ANN model. The final optimized models were selected based on their Q2, R2, and root mean square error (RMSE) values for the training set instances. Detailed depictions of the ML model optimization and their actual vs. predicted activity for each of the optimized models are provided in Fig. **5.1E** to **5.1P**.

5.3 Evaluation of Machine Learning (ML) model performance

All regression-based machine learning models were constructed using the selected ten features on the training set molecules ($N_{Training}$ = 104). The statistical quality of these optimized ML models is summarized in Table 5.1. Among the four regression-based ML models analysed, the Random Forest (RF) model with 90 trees and the Support

Vector Machine (SVM) model with an epsilon of 0.001, an optimal C of 524,288.0 (2^19), and an optimal γ value of 0.00000762939453125 (2^-17) were found to be the most capable of representing the activity variation of these HDAC6 inhibitors both internally and externally (Table 5.1).

Among the twenty k-NN models generated with the selected feature set and varying numbers of neighbours, the model with 9 neighbours was found to be the best. It exhibited an R^2 of 0.675 and a Q^2 of 0.549, with an R^2Pred value of 0.385 for the test set.

From the development of the Random Forest (RF) model, the RF model with 90 trees produced a Q^2 of 0.537 and an R^2Pred of 0.668 for the test set molecules, while maintaining an R^2 of 0.954 for the training set population.

Table 5.1: The statistical performance calculated for the optimized ML models

Model Type	R^2	***RMSE***	Q^2	$RMSE_{LOO\text{-}CV}$	R^2_{Pred}	$RMSE_{Pred}$
k-NN	0.657	0.404	0.549	0.462	0.385	0.560
RF	***0.954***	***0.173***	***0.537***	***0.463***	***0.668***	***0.419***
SVM	***0.738***	***0.348***	***0.661***	***0.397***	***0.564***	***0.483***
ANN	0.704	0.419	0.655	0.399	0.487	0.525

The parameter-optimized Artificial Neural Network (ANN) model, with an optimal configuration of 1 hidden layer, a learning rate of 0.1, a momentum of 0.3, and 10 epochs, exhibited R^2, Q^2, and R^2Pred values of 0.704, 0.655, and 0.487, respectively. Additionally, the optimized Support Vector Machine (SVM) model, with an epsilon of 0.001, an optimal C of 524,288.0 (2^19), and an optimal γ of 0.00000762939453125 (2^-17), produced an R^2 value of 0.738 and a Q^2 of 0.661, while showing an R^2Pred of 0.564 for the test set.

5.4 Interpretation of selected features used for machine learning

By analysing the model's descriptors, it is possible to gain valuable chemical insights into the activities and structural requirements that influence the inhibitory activity of these quinazoline-based HDAC6 inhibitors. Therefore, a critical examination of these selected features is essential not only for a deeper understanding of their inhibitory

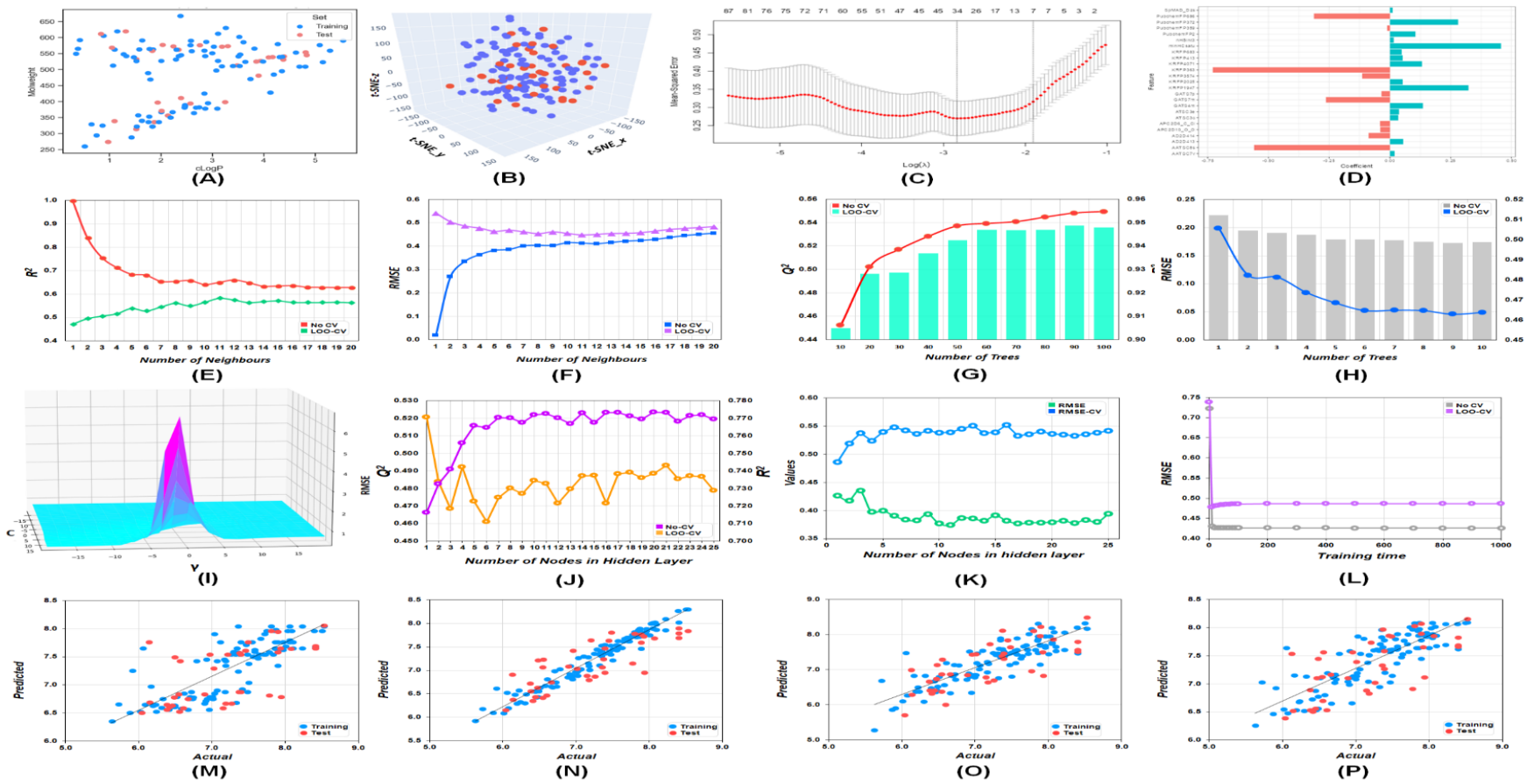


Fig. 5.1. (A) PCA-based chemical space distribution of training and test set compounds, **(B)** Molecular property-based *t*-SNE distribution of the training and test set molecules, **(C)** Mean squared error (*MSE*) vs log λ values for the LASSO-feature selection, **(D)** LASSO regression selected final set of features, **(E)** R^2 scores vs Number of neighbours for the *k*-NN model optimization, **(F)** *RMSE* scores *vs* Number of neighbours for the *k*-NN model optimization, **(G)** R^2 scores vs Number of trees for the random forest model optimization, **(H)** *RMSE* scores *vs* Number of trees for the random forest model optimization **(I)** optimization contour surface for the SVM model, **(J)** R^2 scores vs number of hidden layer nodes for the ANN model, **(K)** *RMSE* scores *vs* number of hidden layer nodes for the ANN model, **(L)** *RMSE* scores vs learning time for the ANN model, Observed versus predicted activity for the optimized **(M)** k-NN model, **(N)** random forest model, **(O)** SVM model, **(P)** ANN model,

mechanism but also for designing more potent molecules. Consequently, a comprehensive discussion of these selected features is provided.

KRFP363 is a Klekota-Roth fingerprint substructure that indicates the presence or absence of the N-methylbutan-1-amine function in a molecule. A detailed analysis of the dataset molecules revealed that the presence of this functionality in the linker region of compound 30 resulted in lower HDAC6 inhibitory efficacy (IC50 = 1,238 nM) (Fig. 4). Comparing this with other similar compounds such as compounds **27**, **28**, and **29**, it appears that the presence of the amine functionality in the linker region is unfavourable. A thorough examination of the crystal structure of HDAC6 (PDB ID: 5G0G & 5G0H) showed that both catalytic domains are highly conserved and feature narrow hydrophobic channels composed of residues Pro83, Phe202, Trp261, and Gly201 in catalytic domain 1, and Pro464, Phe583, Gly582, Leu712, and Phe643 in catalytic domain 2. This hydrophobic nature of the channel may be responsible for the unfavourable interaction with the amine functionality in the linker region, leading to decreased inhibitory potency.

ATSC8 represents the average Centered Broto-Moreau autocorrelation with a lag of 8, weighted by I-state. This 2D autocorrelation descriptor is calculated based on the Moreau-Broto autocorrelation, where lag 8 signifies the topological distances between two atoms in a molecule. It was observed that dataset molecules with higher negative values of this descriptor were effective HDAC6 inhibitors, including compounds **21**, **31**, **33**, **35**, **39**, **49**, **50**, **99**, and **137**.

Likewise, ***minHCsatu*** is a 2D Atom type Electrotopological state descriptor representing the minimum atom-type H E-State: H on C sp3 bonded to unsaturated C. It was observed that molecules with higher positive values of this feature were promising HDAC6 inhibitors, including compounds **24**-**26**, **31**, **40**-**44**, **62**, and **66**-**68**.

PubchemFP686 represents a 4-hydroxy ketone function (Fig. 5.2). It was observed that most molecules (compounds **70**-**77**, **109**, **113**, **117**, **120**) with this functionality in the linker were less effective HDAC6 inhibitors. In comparison to other compounds such as **78**-**95**, it appears that the 4-hydroxy ketone functionality may be detrimental to HDAC6 inhibition. However, replacing it with a 5-hydroxy ketone (compounds **78**-**86**) or a 6-hydroxy ketone (compounds **87**-**95**) leads to a significant improvement in HDAC6 inhibitory efficacy.

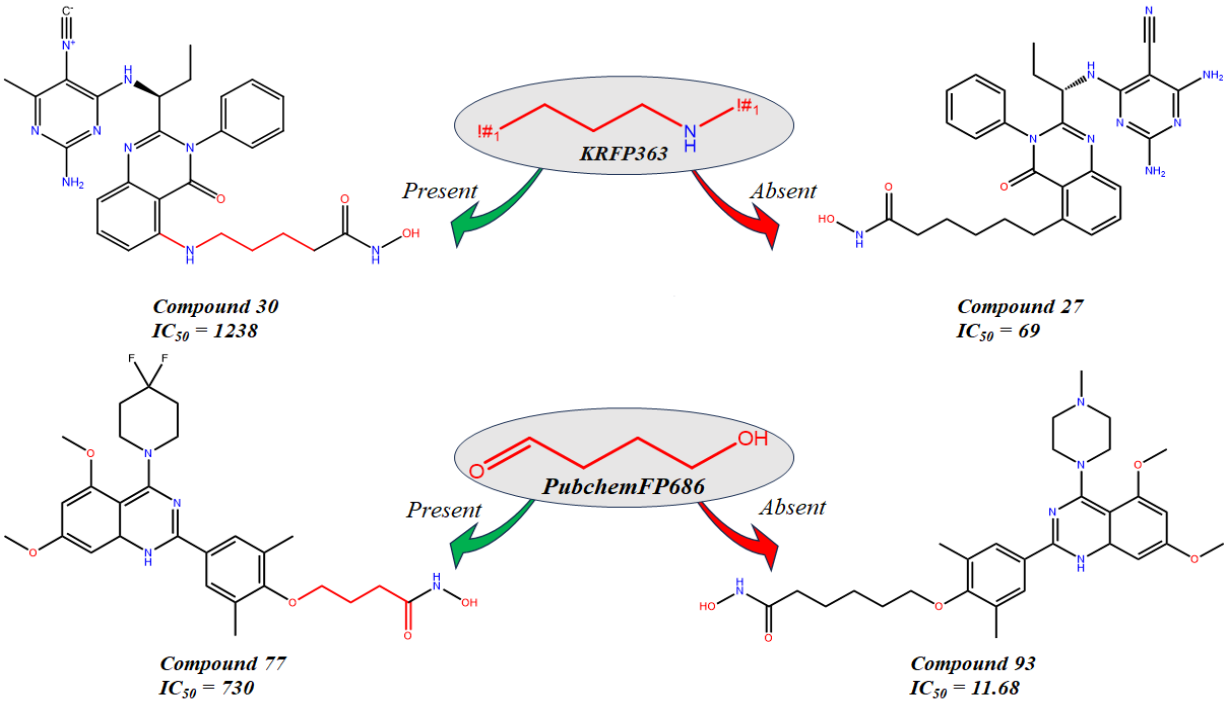


Fig. 5.2**.** Selected features with negative contributions from the ML model with their representative dataset compounds

PubchemFP372 represents the ethanamine function. This fragment was observed in the aryl moiety of the dataset compounds (Fig. 5.3). Detailed analysis revealed that the presence of this functionality, particularly in the linker region (as pyridine in compounds **24**, **26**, **36**, **45**, **46**, and as triazole in compounds **96**-**107**), led to more active HDAC6 inhibitors. It can be assumed that the lone pair on the nitrogen and the π electron cloud of the aromatic ring facilitate favourable electrostatic interactions at the HDAC6 active site, contributing to higher HDAC6 inhibition.

On the other hand, ***GATS4m*** is a 2D autocorrelation descriptor representing Geary autocorrelation-lag 4, weighted by mass. In this descriptor, the Geary coefficient, a distance-type function, can be any physicochemical property calculated for each atom in the molecule, such as electronegativity, atomic mass, dipole moment, polarizability, etc. Here, the property is atomic masses. It was observed that both increases and decreases of this parameter beyond a certain range led to a reduction in HDAC6 inhibition. This implies that a balanced topological distribution of atomic masses, along with the spatial molecular graph, is necessary for effective HDAC6 inhibition.

Additionally, the seventh selected descriptor, ***KRFP413***, represents the ethylbenzene function (Fig. 5.3). It was observed that compounds containing this substructural

feature in the 3rd position of the quinazoline ring are more effective HDAC6 inhibitors (compounds **128**, **130-134**, **137**, **138**, **140-142**), with the exception of compound **139** (IC50 = 747 nM). Further analysis revealed that replacing the ethylbenzene functionality with halogen-substituted methylbenzene (compounds **7-10**) led to a decrease in HDAC6 inhibition efficacy. This suggests that an appropriate degree of rotation and optimized lipophilicity for such substitution in the 3rd position of the quinazoline ring enhances inhibitory efficacy.

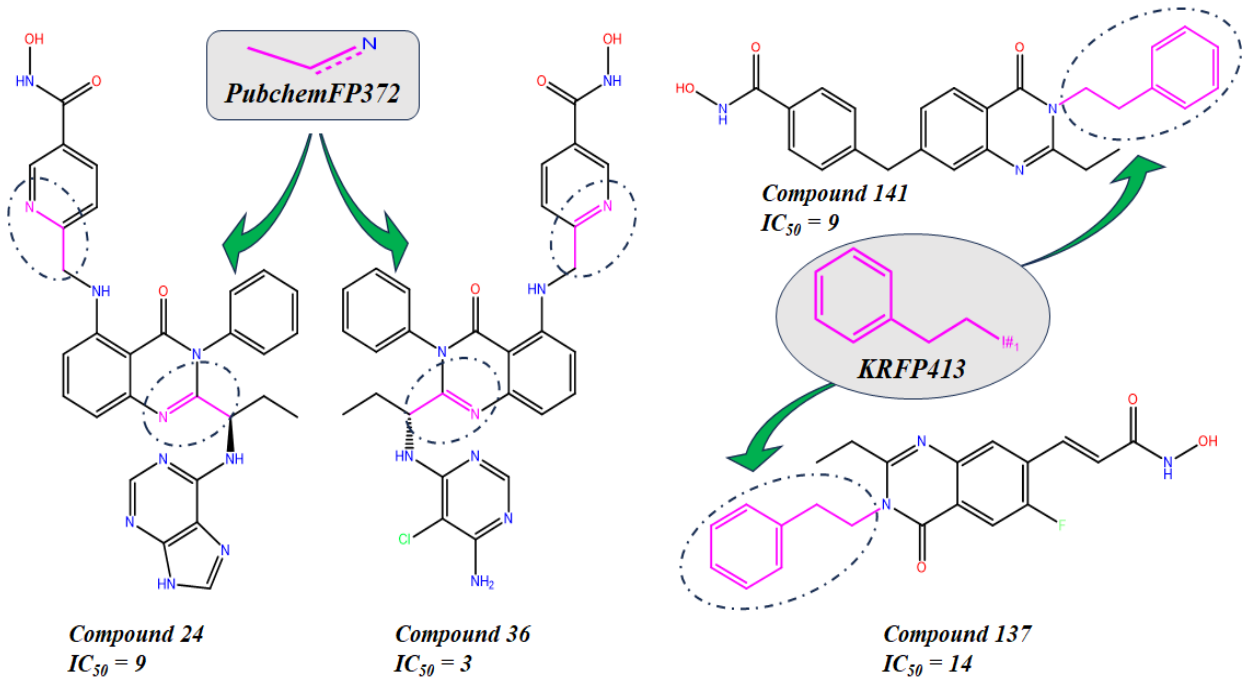


Figure 5.3: Selected features with positive contributions from the ML model with their representative dataset compounds

APC2D6_C_Cl is an atom pair count descriptor indicating the presence of a C-Cl group at a topological distance of 6. Structural analysis of the dataset molecules showed that both the frequency of the C-Cl feature and its relative position within the molecular structure significantly impact inhibitory efficacy. For instance, compound **2**, which has this feature attached to the 3rd position of the quinazoline ring, is a less effective HDAC6 inhibitor (IC50 = 2,385 nM).

GATS7p is a 2D descriptor representing Geary autocorrelation-lag 7, weighted by polarizabilities, which encodes the distribution of polarizability along the molecular topology. Polarizability refers to the ability of an atom or molecule to attract and distort the electron cloud of neighbouring species, a process known as polarization. It was

observed that in our dataset, compounds with values of this descriptor within a specific range demonstrated strong inhibitory efficacy. This suggests that an optimal distribution of polarizability throughout the molecule's topology can enhance HDAC6 inhibitory potency.

Lastly, the descriptor ***nHBint3*** represents the count of E-State descriptors indicating the strength of potential hydrogen bonds with a path length of 3. It was observed that an increase in the occurrence rate of this electro-topological state descriptor generally corresponds to lower IC50 values, indicating enhanced inhibitory activity. This suggests that as the potential for hydrogen bond formation increases, the molecules may achieve more thermodynamically stable interactions with the target protein (HDAC6). Therefore, for designing more potent HDAC6 inhibitors, it is important to consider an optimal number of hydrogen bond donor and acceptor groups.

5.5 Bayesian Classification model

In addition to our non-linear QSAR approach, the fragment-dependent Bayesian classification analysis proved highly effective in identifying key substructural features that influence the HDAC6 inhibitory activity of these hydroxamate derivatives. To assess the quality and reliability of the Bayesian classification model, several statistical metrics were calculated (Table 4.1), and the results demonstrated statistical reliability. The model showed a LOO cross-validation ROC score of 0.873, a ROC score of 0.870 from five-fold cross-validation, and a ROC score of 0.824 for the test set (Appendix Figure S1), highlighting its predictive performance and reliability. Additional statistical validation parameters are provided in Table 5.2 and Fig. 5.5(A).

Table 5.2 Calculated statistical parameters for the classification-based QSAR models

Model	*Dataset*	*ROC*	*ROC_{LOO-cv}*	*Se*	*Sp*	*Pr*	*ACC*	*MCC*	*F_1*	*AUC_b*
Bayesian	Training	0.870	0.873	0.912	0.833	0.911	0.885	0.745	0.911	0.872
Classification	Test	0.824	-	0.727	0.750	0.800	0.737	0.471	0.761	0.738
RP	Training	0.927	0.820	0.666	0.962	0.944	0.817	0.660	0.781	0.814
Decision tree	Test	0.767	--	0.545	0.750	0.750	0.631	0.295	0.631	0.647
1										
SARpy	Training	--	--	0.940	0.830	0.941	0.900	0.785	0.927	0.885
	Test	---	--	0.740	0.730	0.727	0.750	0.471	0.761	0.735

Additionally, this analysis identified a total of 40 ECFP_6 substructural fragments of the hydroxamate derivatives, which have either positive or negative effects on HDAC6 inhibition. The 20 beneficial (G1-G20) and 20 detrimental (B1-B20) molecular fingerprints are illustrated in table 5.3 and 5.4, respectively. Upon examination, the 20 beneficial molecular substructures (G1-G20) can be categorized into three groups: quinazoline, quinazoline-4-one, and pyrimidine moieties (G1-G4, G6, G8); alkane-substituted azole moieties (G10, G12-G13, G15-G20); and aliphatic amines (G5, G9, G11, G14).

Conversely, among the features negatively impacting HDAC6 inhibition, the pyrimidine moiety was the most frequently observed detrimental feature (B1-B2, B5-B6, B8, B10, B11-B12, and B14). Additionally, single nitrogen-containing saturated 6-membered heterocyclic rings (B4), benzyl ethers (B9, B17, B19-B20), and alkyl ethers (B13, B15) were also identified as negative regulators of HDAC6 inhibitory activity.

Further analysis of the substructures and molecules revealed that the most effective compounds, **36** and **46** (IC50 = 3 nM), not only feature quinazoline as a central scaffold but also include a pyrimidine connected to quinazoline via a secondary amine.

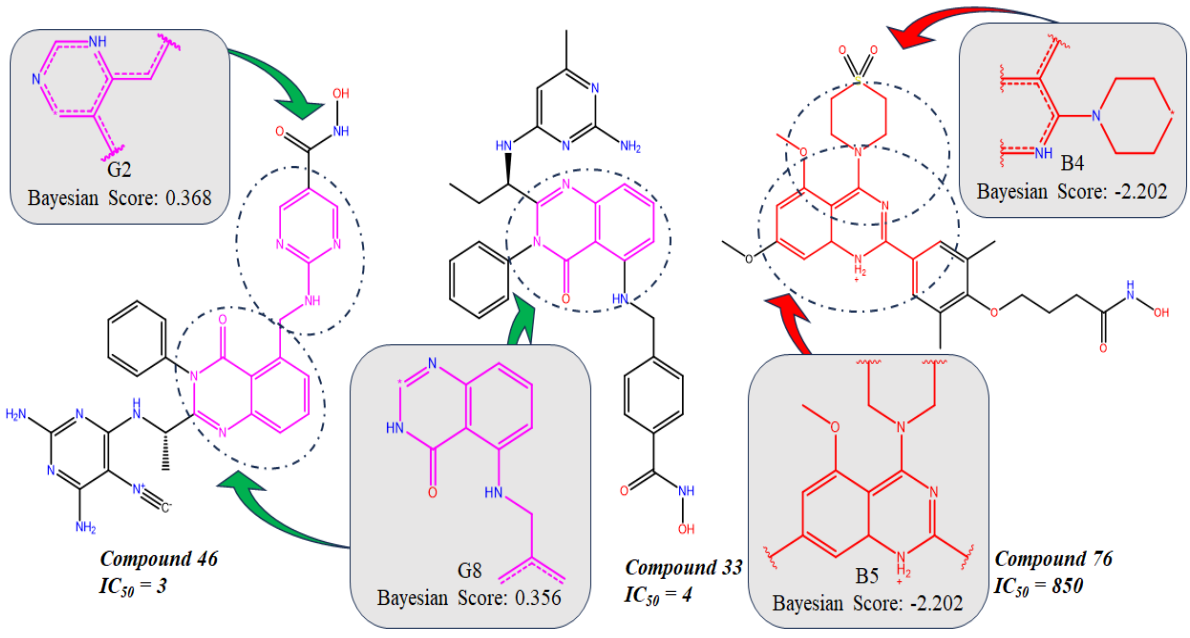


Figure 5.4: Representative dataset molecules with good and bad Bayesian fragments.

Similarly, compounds **31**-**35** and **37**-**45** also possess these structural features and were found to be potent HDAC6 inhibitors, with IC50 values ranging from 4 nM to 52 nM. This observation aligns with our regression analysis, which indicated that the strength

for potential hydrogen bonds of path length 3 (***nHBint3***) and the ethanamine function (***PubchemFP372***), present in pyridine, pyrimidine, and triazole rings, have a positive effect on HDAC6 inhibitory activity (Fig. 5.4).

Regarding the negative ECFP_6 fragments, it was noteworthy that the quinazoline scaffold with ether and tertiary amine functions (B2 and B5) was detrimental to HDAC6 inhibition. While quinazoline is a common core feature in many HDAC inhibitors, such as compounds **71**-**83**, which contain these unfavourable structural features, were found to be less effective HDAC6 inhibitors. These findings also suggest that the presence of an ether group may negatively impact the HDAC6 inhibitory activity of these hydroxamate analogues.

5.6 Recursive partitioning (RP) study

The CART-based recursive partitioning method generated four decision trees using fundamental molecular features and FCFP_6 to classify active and inactive compounds. Among these, the first decision tree, which utilized four substructural features, was chosen for further analysis due to its performance. This tree had an ROC of 0.927 and a cross-validated ROC (ROCcv) of 0.820 for the training data, and an ROC of 0.767 for the test set. For the training data, decision tree 1 demonstrated its predictive capability with 66.6% sensitivity (Se), 96.2% specificity (Sp), 94.4% precision (Pr), and 96.2% accuracy (ACC) (Fig. 5.5 A). During external validation, decision tree 1 also provided satisfactory predictions with 54.5% sensitivity, 75.0% specificity, 75.0% precision, and 63.1% accuracy (Fig. 5.5 A). The statistical validation parameters are detailed in Table 5.2.

Decision tree 1 (Fig. 5.5 B), which incorporates four substructural features, also uses nHBA (number of hydrogen bond acceptors), nHBD (number of hydrogen bond donors), and Mw (molecular weight) as key factors to classify the dataset molecules into seven distinct groups (Fig. 5.5B). This decision tree highlights the significance of butyl amine functionality and molecular weight in distinguishing between active and inactive compounds. It is noteworthy that while the butyl amine function was identified as a favourable feature (G14) in the Bayesian classification study, it appears to have a negative impact on HDAC6 inhibitory activity in the current recursive partitioning model.

Table 5.3: Good *ECFP_6* fragments obtained from the Bayesian classification study

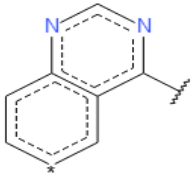
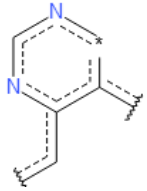
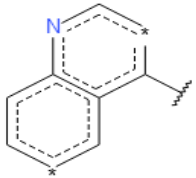
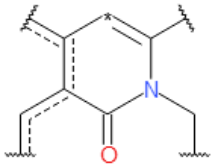
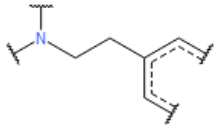
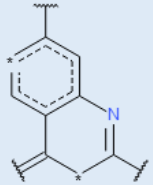
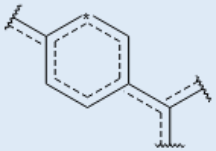
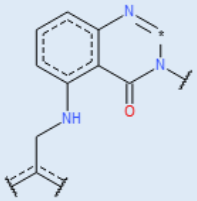
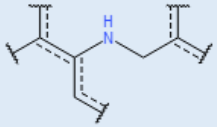
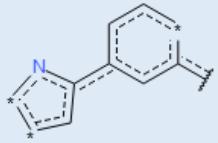
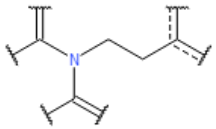
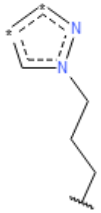
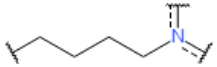
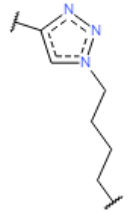
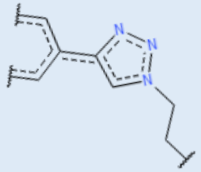
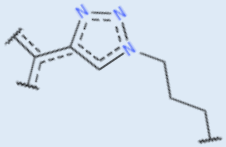
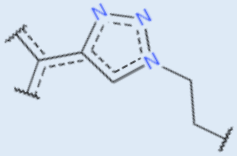
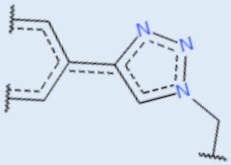
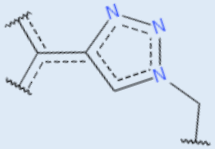
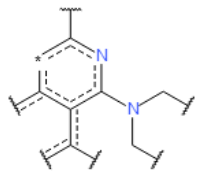
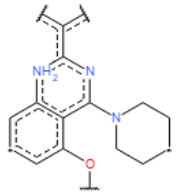
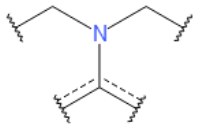
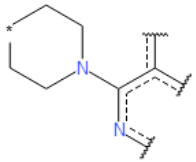
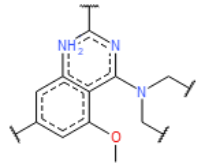
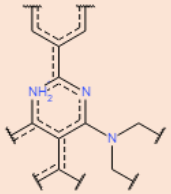
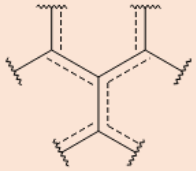
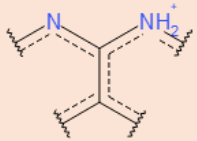
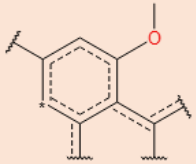
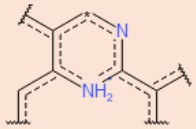
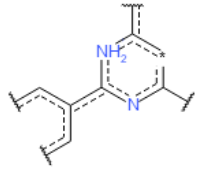
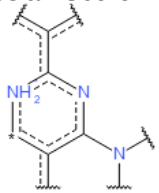
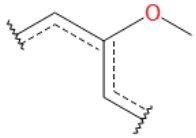
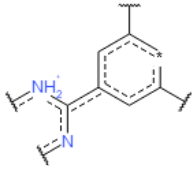
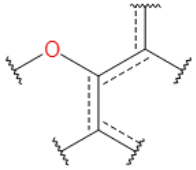
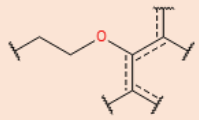
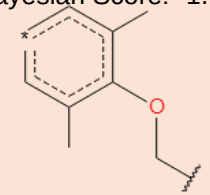
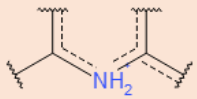
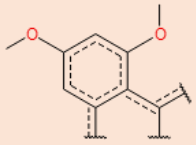
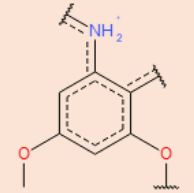
 G1: 726751110 14 out of 14 good Bayesian Score: 0.368	 G2: -1607747969 14 out of 14 good Bayesian Score: 0.368	 G3: 1569575246 14 out of 14 good Bayesian Score: 0.368	 G4: -1179243667 12 out of 12 good Bayesian Score: 0.363	 G5: -592856198 11 out of 11 good Bayesian Score: 0.360
 G6: 573062983 11 out of 11 good Bayesian Score: 0.360	 G7: -219423964 10 out of 10 good Bayesian Score: 0.356	 G8: -1002025377 10 out of 10 good Bayesian Score: 0.356	 G9: -4235950 10 out of 10 good Bayesian Score: 0.356	 G10: 1278310572 10 out of 10 good Bayesian Score: 0.356
 G11: -2119310481 10 out of 10 good Bayesian Score: 0.356	 G12: -1731122899 10 out of 10 good Bayesian Score: 0.356	 G13: 816396776 10 out of 10 good Bayesian Score: 0.356	 G14: -2118347459 10 out of 10 good Bayesian Score: 0.356	 G15: 1990926508 10 out of 10 good Bayesian Score: 0.356
 G16: 121650401 10 out of 10 good Bayesian Score: 0.356	 G17: 2129650112 10 out of 10 good Bayesian Score: 0.356	 G18: 311627079 10 out of 10 good Bayesian Score: 0.356	 G19: -1589534745 10 out of 10 good Bayesian Score: 0.356	 G20: -2041577508 10 out of 10 good Bayesian Score: 0.356

Table 5.4: Bad *ECFP_6* fragments obtained from the Bayesian classification study

 B1: 887355427 0 out of 12 good Bayesian Score: -2.202	 B2: -1796703767 0 out of 12 good Bayesian Score: -2.202	 B3: 1951894094 0 out of 12 good Bayesian Score: -2.202	 B4: -95545909 0 out of 12 good Bayesian Score: -2.202	 B5: 14006777 0 out of 12 good Bayesian Score: -2.202
 B6: 1478078755 0 out of 12 good Bayesian Score: -2.202	 B7: -1661653144 1 out of 19 good Bayesian Score: -1.927	 B8: -739716278 1 out of 16 good Bayesian Score: -1.768	 B9: 182451333 1 out of 16 good Bayesian Score: -1.768	 B10: 257408491 1 out of 16 good Bayesian Score: -1.768
 B11: -304302513 1 out of 16 good Bayesian Score: -1.768	 B12: -249806767 1 out of 16 good Bayesian Score: -1.768	 B13: 2055803015 1 out of 16 good Bayesian Score: -1.768	 B14: 1821498523 1 out of 16 good Bayesian Score: -1.768	 B15: -1531301414 1 out of 16 good Bayesian Score: -1.768
 B16: -742907450 1 out of 16 good Bayesian Score: -1.768	 B17: -1807893558 1 out of 16 good Bayesian Score: -1.768	 B18: -666326105 1 out of 16 good Bayesian Score: -1.768	 B19: 809943749 1 out of 16 good Bayesian Score: -1.768	 B20: -1393419274 1 out of 16 good Bayesian Score: -1.768

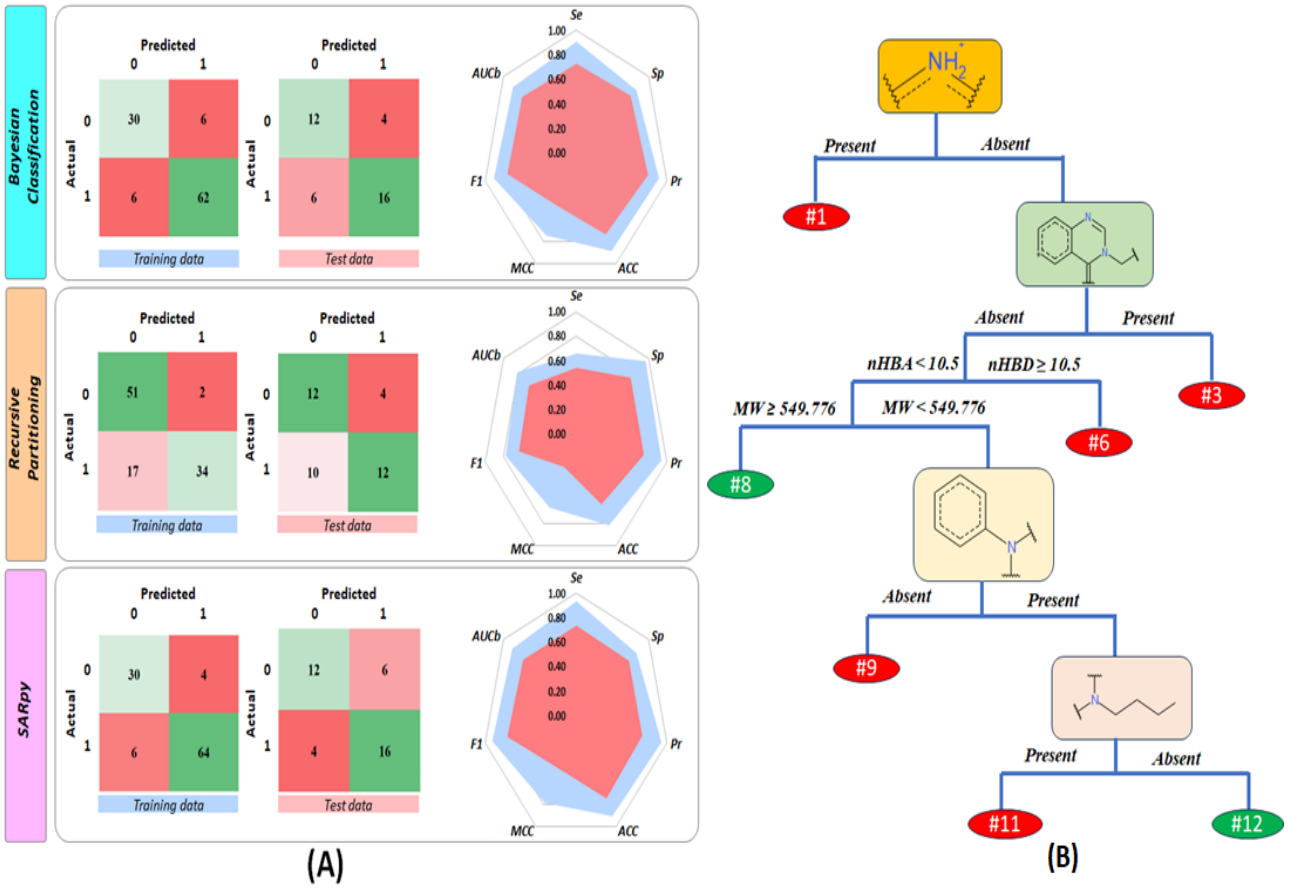


Figure 5.5: **(A)** Confusion matrix and calculated statistical parameters for the classification-based QSAR models (***blue:*** *Training data*, ***red:*** *Test data*), **(B)** Schematic representation of RP model generated decision tree 1.

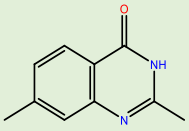
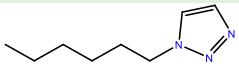
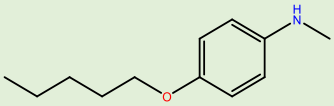
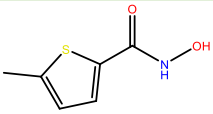
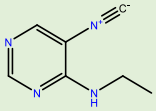
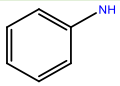
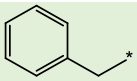
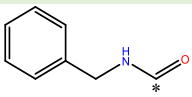
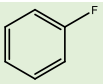
5.7 SARpy key structural attribute identification study

The smiles-mediated structural alert mining using SARpy analysis of hydroxamate-based inhibitors identified nine substructural features in smiles notation as part of the active rule set (Table 5.5). For the training set, SARpy analysis achieved 94.0% sensitivity (Se), 83.0% specificity (Sp), 94.1% precision (Pr), and 90.0% accuracy (ACC). When validated on the test set, the active ruleset showed 74.0% sensitivity, 73.0% specificity, 72.7% precision, and 75.0% accuracy. Additional statistical validation parameters related to SARpy analysis are detailed in Table 5.2.

Among the nine structural alerts identified by SARpy as part of the active ruleset (Table 5.5), several were noted for their positive contributions to HDAC6 inhibition. Specifically, the alert c12c(c(=O)n(c(n1)C))ccc(c2)C represents the quinazoline-4-one

functionality, and c1nnn(c1)CCCCCC denotes 1-hexyl triazole, both of which positively impact HDAC6 inhibition.

Table 5.5: SARpy active ruleset along with their corresponding structures and likelihood ratio

Active Ruleset	*Structure*	*Likelihood Ratio*
c12c(c(=O)n(c(n1)C))ccc(c2)C		Infinite
c1nnn(c1)CCCCCC		Infinite
N(C)c1ccc(cc1)OCCCCC		Infinite
Cc1ccc(s1)C(=O)NO		Infinite
C(Nc1c(cnc(n1))[N+]#[C-])C		5.29
Nc1ccc(cc1)		8.21
CCc1ccccc1		7.94
C(=O)NCc1ccc(cc1)		3.18
c1ccc(cc1)F		3.97

Similarly, Bayesian classification identified analogous features as beneficial fragments (G8 and G12-G13, G15-G20) with positive effects on HDAC6 efficacy. The alert N(C)c1ccc(cc1)OCCCCC highlights the positive influence of the N-methyl-4-(pentyloxy)aniline moiety on HDAC6 inhibition, while Cc1ccc(s1)C(=O)NO indicates that a 2-methylthiophene-containing hydroxamic acid moiety is favorable for activity. Additionally, C(Nc1c(cnc(n1))[N+]#[C-])C and Nc1ccc(cc1) denote N-ethyl-5-isocyanopyrimidin-4-amine and aniline, respectively, both contributing positively to HDAC6 inhibition and also identified as good fragments (G2 and G7) in our Bayesian

model. Finally, the structural alerts CCc1ccccc1, C(=O)NCc1ccc(cc1), and c1ccc(cc1)F, representing ethylbenzene, N-benzyl formamide, and fluorobenzene, respectively, were also positive contributors to HDAC6 inhibition. Notably, the common benzyl function among these three alerts was also recognized as a beneficial fragment (G7 and G10) in the Bayesian classification model.

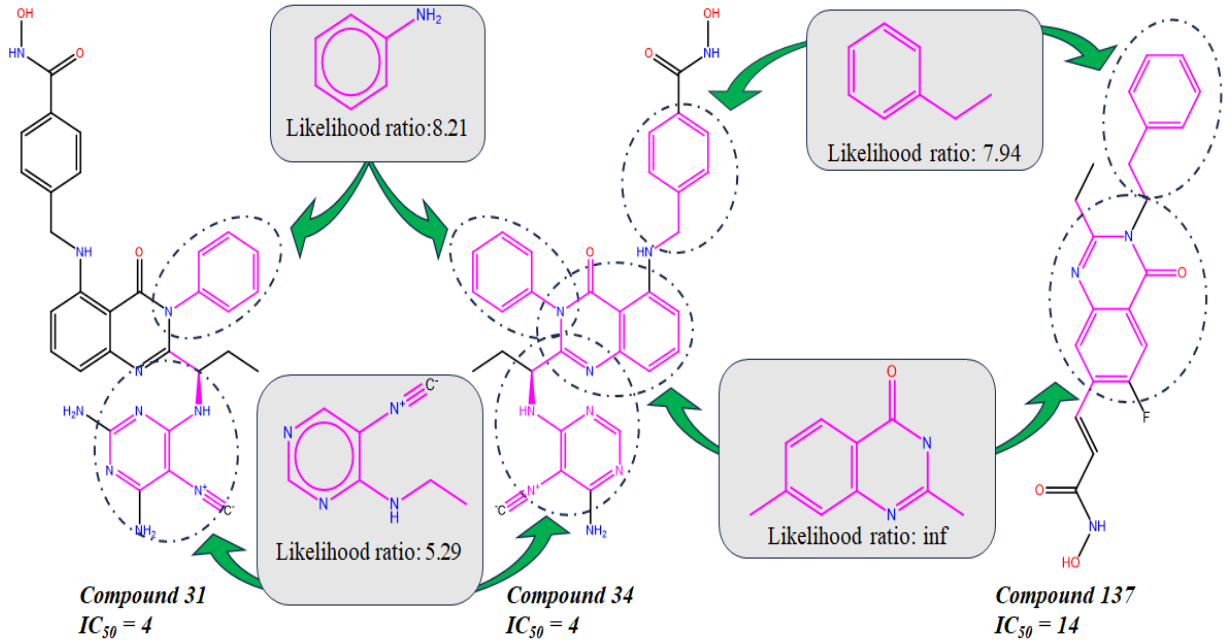


Figure 5.6: SARpy generated active fragments containing potent HFAC6 inhibitors.

Applying the active ruleset to the training set, 70 out of 104 structures matched, while 20 out of 38 test set molecules matched. Notably, compounds **31**, **34**, **38**, and **44**, which feature the active ruleset C(Nc1c(cnc(n1))[N+]#[C-])C and Nc1ccc(cc1) identified by SARpy structural alert mining, are effective HDAC6 inhibitors. Additionally, compounds **96-105** and **131-137**, which contain multiple structural alerts, were also found to be potent HDAC6 inhibitors (Fig. 5.6).

5.8 Molecular dynamics (MD) simulation study

Molecular dynamics (MD) simulations of the most effective (compound **36**) and least effective (compound **2**) quinoline-containing hydroxamate derivatives at the HDAC6 active site were conducted using the Desmond module of Schrodinger Maestro v12.5 software. Analysis of the RMSD values of the protein and ligand trajectories during the simulation (Fig. 5.7) revealed that the C-α chain of HDAC6 (PDB ID: 5EDU) exhibited minimal RMSD fluctuation (< 3Å) when bound to both compounds **36** and **2**. However,

the RMSD values for the atoms of compound **36** were noticeably higher compared to those for compound **2.** The RMSD for compound **2**, when complexed with HDAC6, exhibited similar fluctuations to the C-α chain of the protein, in contrast to the higher fluctuations observed with compound **36** (Fig. 5.7B vs 5.7A).

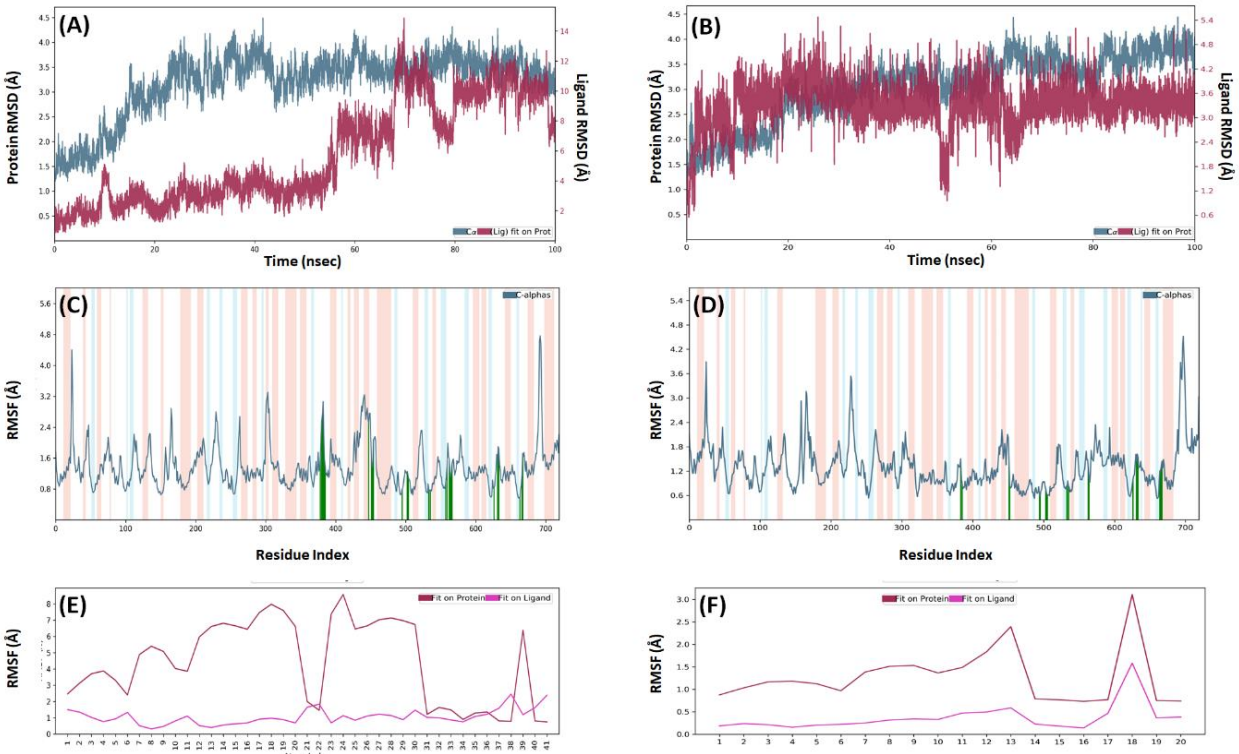


Figure 5.7: 100 ns trajectory *RMSD* plot for **(A)** compound **36**, and **(B)** compound **2**, *RMSF* plot for the C-α chain residues of HDAC6 (PDB ID: 5EDU) **(C)** compound **36**, and **(D)** compound **2**, *RMSF* of **(E)** compound **36**, and **(F)** compound **2** for 100 ns simulated period

Interestingly, this suggests that compound **2** binds more stably to the HDAC6 active site compared to compound **36**, despite being the least effective inhibitor in the series. The RMSD fluctuation of the C-α chain residues (Fig. 5.7C and 5.7D) showed similar fluctuations for most residues, except for Ser498, His499, Leu495, Trp496, Lys555, Ser563, Ser564, Asn565, Phe620, Gly619, Cys618, His610, and His611, which exhibited higher fluctuations when bound to compound **36** (Fig. 5.7C) compared to compound **2** (Fig. 5.7D). Conversely, residues such as Ser688, Ser689, and Pro708 showed higher fluctuations in the HDAC6-compound **2** complex compared to the HDAC6-compound **36** complex. Additionally, the heavy atoms of the cap group in compound **36** (atom numbers 1-20, 23-30, and 39, Fig. 5.7E) exhibited high fluctuations

(>2Å), whereas the smaller cap group of compound **2** (atom numbers 1-18, Fig. 5.7F) showed minimal fluctuation (< 2Å), except for the chlorine atom. This indicates that the high RMSD values observed for the most active compound **36** are due to significant fluctuations in the cap group heavy atoms, influenced by the highly flexible ethylene spacer connecting the cap to the linker phenyl ring.

In the analysis of protein-ligand interactions at the HDAC6 active site (Fig. 5.8), compound **36** interacted with a larger number of HDAC6 amino acid residues compared to compound **2** (Fig. 5.8A vs 5.8B). Specifically, compound **36** formed hydrogen bonds with Ser568, His611, Pro717, and His499, while establishing hydrophobic interactions with Phe620, His651, Tyr782, Pro501, Tyr570, Leu749, His500, His499, Phe679, Phe680, Met682, and Asn494. Additionally, ionic interactions were observed between compound **36** and the residues Asp649, His651, Asp742, and Tyr782 of HDAC6 (Fig. 5.8A).

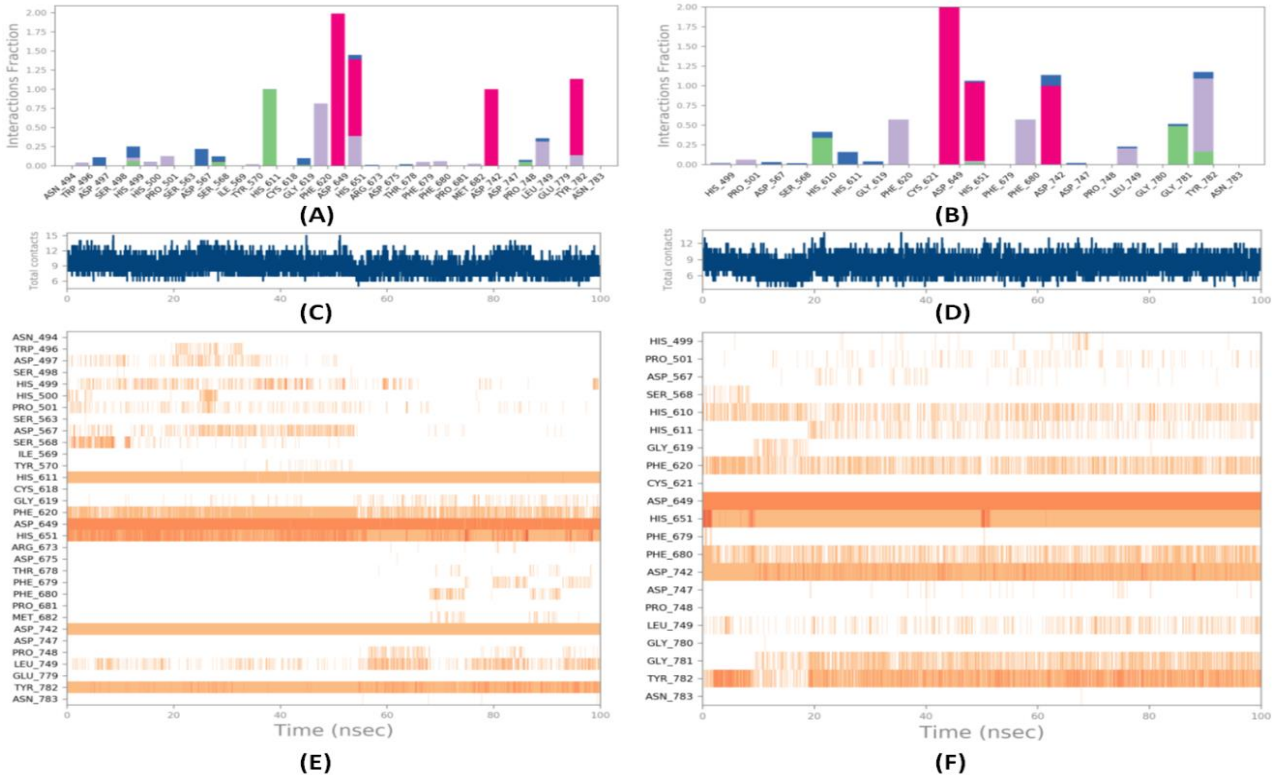


Figure 5.8: The interaction fraction recorded between **(A)** compound **36**, and **(B)** compound **2** and HDAC6 catalytic site (PDB ID: 5EDU), Overall contact frequency for **(C)** HDAC6-compound **36** complex, and **(D)** HDAC6-compound **2** complex, Contact frequency of **(E)** compound **36**, and **(F)** compound **2** with HDAC6 catalytic site residues

Although the least active compound **2** interacted with fewer HDAC6 amino acid residues, it established hydrogen bond interactions with only three residues: His610, Gly781, and Tyr782, as well as with His651. For hydrophobic interactions, compound **2** interacted with residues Phe620, Phe680, Leu749, Tyr782, Pro501, and His499. Additionally, ionic interactions were observed with Asp649, His651, and Asp782 (Fig. 5.8A).

During the 100 ns simulation period, it was observed that while compound **36** interacted with a larger number of HDAC6 catalytic site residues, many of these interactions were brief. The most significant and stable interactions were with residues Asp649, His651, His611, Phe620, Asp742, and Tyr782 (Fig. 5.8E). In contrast, compound **2** formed strong interactions primarily with residues Asp649, His651, Asp742, and Phe680 (Fig. 5.8F). However, compound **2** did not establish strong interactions with residues His610, His611, and Phe620 as effectively as compound **36**. Molecular generalized Born surface area (MM-GBSA) calculations for the ligand-protein binding energy were performed using the Prime module of Schrodinger Maestro software with a step of 10 (nsteps = 10) (Table 5.6). These calculations indicated that the most active compound demonstrated stable binding, whereas the least active compound formed a less stable complex with HDAC6. The calculated coulombic, van der Waals, and overall binding energies for these complexes during the simulation are detailed in Fig. 5.9A and 5.9B.

Table 5.6 Prime calculated MM-GBSA binding free energies for the simulated compounds.

Complex	***Energy (kcal/mol)***		
	Avg. $\Delta G_{Binding}$	***Avg. $\Delta G_{Binding\text{-}Coulomb}$***	***Avg. $\Delta G_{Binding\text{-}van\ der\ Waals}$***
HDAC6- Compound 36	- 8.575	- 31.378	- 37.976
HDAC6- Compound 2	+ 3.992	-14.673	-31.406

In addition to the basic MD analyses, the trajectory principal component analysis (Trajectory PCA) was used to calculate the free energy landscape (FEL) with the g_sham script from GROMACS 2020.6. The FEL contour maps reveal that both compounds **36** and **2**, despite having some unstable high-energy conformations (red to reddish brown areas), also exhibited similar stable low-energy conformations (violet to bluish violet regions) (Fig. 5.9A and 5.9B). However, although both compounds

showed stable binding, the conformation of the compound **2**-HDAC6 complex had a higher energy (cyan cleft, Fig. 5.9B) compared to the compound 36-HDAC6 complex. This indicates that, despite greater fluctuation in the cap group of compound **36**, it adopted multiple stable low-energy conformations during the simulation, which may contribute to its superior HDAC6 inhibitory activity compared to compound **2**.

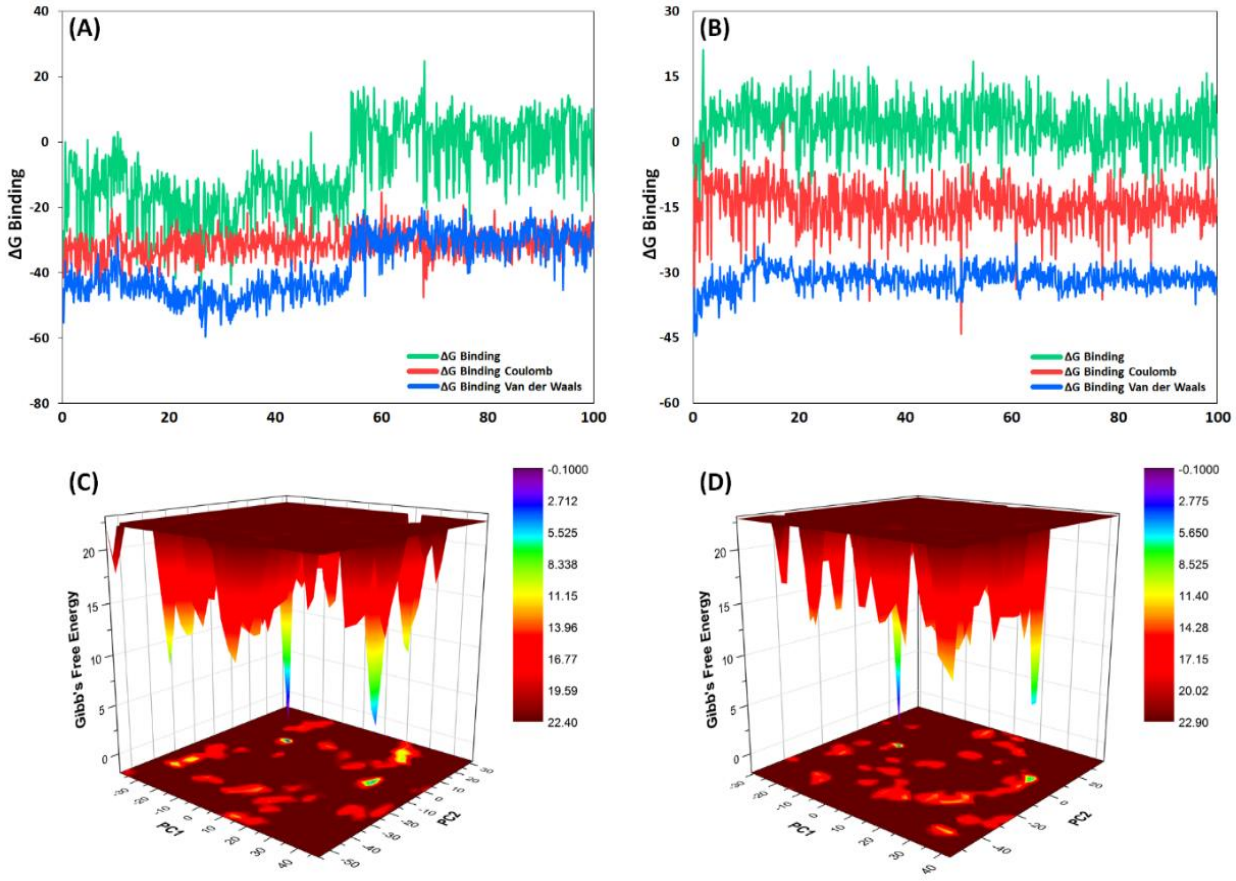


Figure 5.9: The calculated Gibb's free energy for the simulated **(A)** HDAC6-compound **36** complex, and **(B)** HDAC6 compound **2** complex, the trajectory PCA-based free energy landscape plot for **(C)** HDAC6-compound **36** complex, and **(D)** HDAC6 compound **2** complex.

From Fig. 5.9A and 5.9B, it is evident that the compound **36**-HDAC6 complex reached its most stable conformation at 26.6 ns, with the least stable conformation occurring at 68.2 ns (Fig. 5.9A). Conversely, for the compound **2**-HDAC6 complex, the initial conformation at 0.2 ns was the most stable, while the conformation at 0.3 ns was the least stable throughout the 100 ns simulation (Fig. 5.9B). The binding mode analysis of both the most active and least active compounds at various time points (Fig. 5.10) uncovered several notable observations regarding their interaction with HDAC6 CD1.

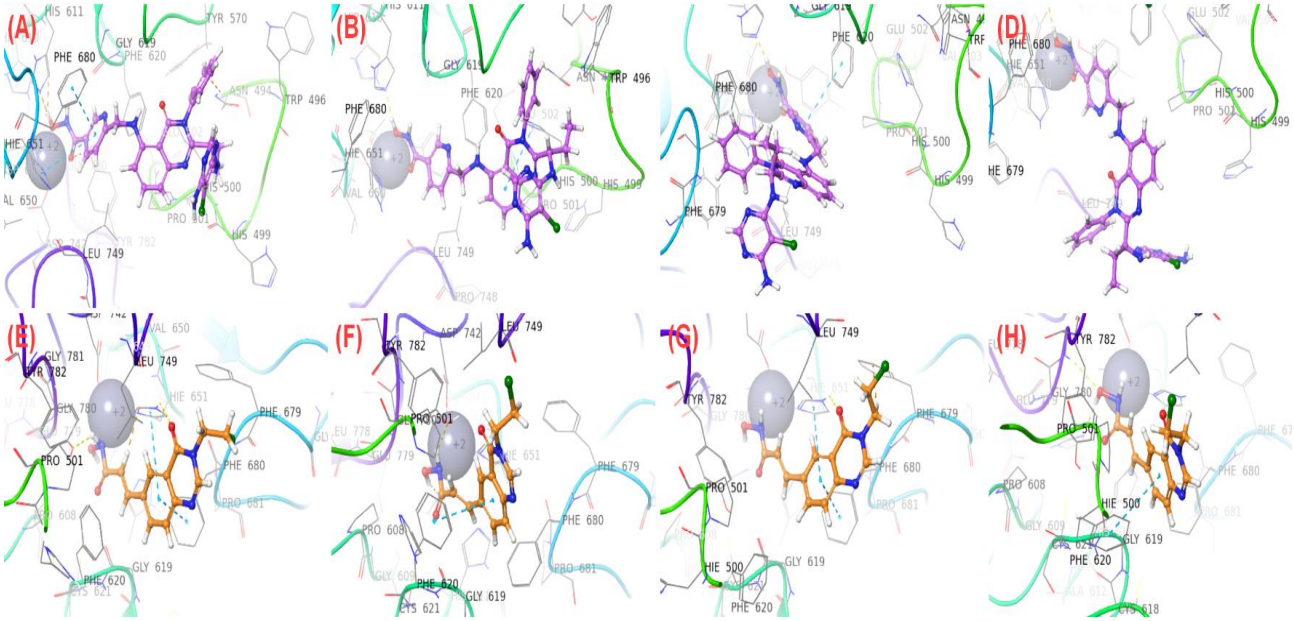


Figure 5.10: The binding mode of compound **36** at HDAC6 active site (PDB ID: 5EDU) at **(A)** 0ns, **(B)** 26.6 ns, **(C)** 68.2 ns, **(D)** 100 ns, the binding mode of compound **2** at HDAC6 active site (PDB ID: 5EDU) at **(E)** 0.03 ns, **(F)** 0.2 ns, **(G)** 50 ns, **(H)** 100 ns

For compound **36**, with its large and bulky quinazoline cap group, the cap moiety exhibited greater fluctuation compared to the smaller cap group in compound **2** (Fig. 5.10A-D vs 5.10E-H). This increased fluctuation in the cap group is reflected in the RMSD plot of the ligand versus the protein (Fig. 5.7). Notably, in the most stable conformation at 26.6 ns (Fig. 5.10B), the fused phenyl ring of the quinazoline cap in compound **36** was observed to form a π-π interaction with His500 in the proximal loop of HDAC6 CD1. However, this interaction was absent in the least stable conformation at 68.2 ns (Fig. 5.10C). For the less active compound **2**, the ethylene chloride substitution in the quinazoline cap was positioned near the carbonyl oxygen between Phe679 and Phe780 in the most stable conformation (0.03 ns, Fig. 5.10E), while it was significantly displaced in the least stable conformation (0.2 ns, Fig. 5.10F). These observations were also supported by the ML models (APC2D6_C_Cl), which suggested that such fragments contribute to HDAC6 binding and influence HDAC6 inhibitory activity and selectivity.

Chapter 6

Conclusion

The success of any research lies in the outcomes it produces and the conclusion it draws, which can uncover new or unexplored scientific insights. These findings can enhance understanding and deepen knowledge in the specific field of study. In response to the economic pressures on the drug market to be cost-effective and time-efficient, computational chemistry, including techniques like computer aided drug design and molecular modelling, along with virtual screening, are now recognised as the most effective and rapid methods for introducing new chemical entities into the market. These computational methods may reduce the lead generation time but are unable to bypass the clinical trials, as its next to impossible to computationally model a whole human body with its ecstatic complexity to understand the effect of human's internal environment on the lead molecule. In other words what a lead molecule and human body dose to each other cannot be comprehended by computational study. To reduce time in this area of drug discovery, drug repurposing becomes the most obvious option, as the knowledge of the drug is already available, the job is to just find a new application for it. This computational study on HDAC6 and its inhibitors was quite helpful to widen the understanding of HDAC6, its pathophysiology and its inhibitors. The machine learning (ML) and fragment based structural analysis study was able to identify key structural features ruling the biological activity. In this study substituted quinazoline moiety, ethylene chloride function, piperazine ring, heterocyclic nitrogen, and sulfur-containing 6-membered saturated ring, number of hydrogen bond donor-acceptor groups, and molecular weight of these compounds came forth as a leading factor for HDAC6 activity variation. Finally, the key finding of this study is that the quinazoline cap group in HDAC6 inhibitors can lead to strong and selective inhibition of HDAC6. Despite its larger size and grater mobility after binding, the bulky quinazoline cap can establish multiple interactions with the residues of HDAC6's pocket-forming loops. These interactions between HDAC6 loop residues and a thoughtfully designed cap moiety in inhibitors could pave the way for the future development of HDAC6-specific inhibitors.

References

References

1. T I Oprea, J. E., Cristian, G., Bologa, T., Buranda, A., Bruce, S., Edwards, J. W., Jarvik, H. D., Mark, K., & Haynes, B. (2011). Drug repurposing from an academic perspective. *Drug Discovery Today: Therapeutic Strategies*, *8*(3–4), 61–69.
2. Chan, J., Wang, X., Turner, J. A., Baldwin, N. E., & Gu, J. (2019). Breaking the paradigm: Dr Insight empowers signature-free, enhanced drug repurposing. *Bioinformatics (Oxford, England)*, *35*(16), 2818–2826. https://doi.org/10.1093/bioinformatics/btz006
3. Xue, H., Li, J., Xie, H., & Wang, Y. (2018). Review of drug repositioning approaches and resources. *International Journal of Biological Sciences*, *14*(10), 1232–1244. https://doi.org/10.7150/ijbs.24612
4. Sinha, S., & Vohora, D. (2018). Drug discovery and development: An overview. *Pharmaceutical medicine and translational clinical research*, 19-32.
5. Barratt, M. J., & Frail, D. E. (2012). *Drug repositioning: Bringing new life to shelved assets and existing drugs*. Wiley. com.
6. Talevi, A., & Bellera, C. L. (2020). «Challenges and opportunities with drug repurposing: finding strategies to find alternative uses of therapeutics. *Expert Opinion on Drug Discovery*, *15*, 397–401. https://doi.org/10.1080/17460441.2020.1704729
7. www.cbo.gov/publication/57025#data.
8. Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2019). Drug repurposing: progress, challenges and recommendations. *Nature Reviews. Drug Discovery*, *18*(1), 41–58. https://doi.org/10.1038/nrd.2018.168
9. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *23*(1–3), 3–25. https://doi.org/10.1016/s0169-409x(96)00423-1
10. Johnson, M. A., & Maggiora, G. M. (1990). *Concepts and applications of molecular similarity*. Wiley.

11. Eckert, H., & Bajorath, J. (2007). Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discovery Today*, *12*(5–6), 225–233. https://doi.org/10.1016/j.drudis.2007.01.011
12. Noeske, T., Sasse, B. C., Stark, H., Parsons, C. G., Weil, T., & Schneider, G. (2006). Predicting compound selectivity by self-organizing maps: Cross activities of metabotropic glutamate receptor antagonists. *ChemMedChem, 1*(10), 1066–1068.
13. Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., ... & Roth, B. L. (2009). Predicting new molecular targets for known drugs. *Nature*, *462*(7270), 175-181.
14. Lounkine, E., Keiser, M. J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J. L., ... & Urban, L. (2012). Large-scale prediction and testing of drug activity on side-effect targets. *Nature*, *486*(7403), 361-367.
15. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., ... & Golub, T. R. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *science*, *313*(5795), 1929-1935.
16. Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., ... & Butte, A. J. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine*, *3*(96), 96ra77-96ra77.
17. Dudley, J. T., Sirota, M., Shenoy, M., Pai, R. K., Roedder, S., Chiang, A. P., ... & Butte, A. J. (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Science translational medicine*, *3*(96), 96ra76-96ra76.
18. LINCS (January 2014). Connectivity map. http://lincscloud.org/.
19. Haupt, V. J., & Schroeder, M. (2011). Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Briefings in bioinformatics*, *12*(4), 312-326.
20. De Franchi, E., Schalon, C., Messa, M., Onofri, F., Benfenati, F., & Rognan, D. (2010). Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. *PLoS One*, *5*(8), e12214.
21. Zahler, S., Tietze, S., Totzke, F., Kubbutat, M., Meijer, L., Vollmar, A. M., & Apostolakis, J. (2007). Inverse in silico screening for identification of kinase inhibitor targets. *Chemistry & biology*, *14*(11), 1207-1214.

22. Kinnings, S. L., Liu, N., Buchmeier, N., Tonge, P. J., Xie, L., & Bourne, P. E. (2009). Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS computational biology*, *5*(7), e1000423.
23. Wermuth, C. G. (2006). Selective optimization of side activities: the SOSA approach. *Drug discovery today*, *11*(3-4), 160-164.
24. Kruger, F. A., Rostom, R., & Overington, J. P. (2012, December). Mapping small molecule binding data to structural domains. In *BMC bioinformatics* (Vol. 13, pp. 1-13). BioMed Central.
25. Wikipedia. Phenotype. http://en.wikipedia.org/wiki/Phenotype.
26. Darwin, C. (1891). *The Origin of Species by Means of Natural Selection: Or the Preservation of Favoured Races in the Struggle for Life*. John Murray, Albemarle Street..
27. Mendel, G. (1866). Versuche über Pflanzenhybriden, Verhandlungen des naturforschenden Vereins. *Brünn: Georg Gastl's Buchdruckerei*.
28. Swinney, D. C., & Anthony, J. (2011). How were new medicines discovered?. *Nature reviews Drug discovery*, *10*(7), 507-519.
29. Duran-Frigola, M., & Aloy, P. (2012). Recycling side-effects into clinical markers for drug repositioning. *Genome medicine*, *4*, 1-4.
30. Fliri, A. F., Loging, W. T., Thadeio, P. F., & Volkmann, R. A. (2005). Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nature chemical biology*, *1*(7), 389-397.
31. Fliri, A. F., Loging, W. T., & Volkmann, R. A. (2007). Analysis of system structure–function relationships. *ChemMedChem: Chemistry Enabling Drug Discovery*, *2*(12), 1774-1782.
32. Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., & Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, *321*(5886), 263-266.
33. Bodenreider, O. (2004). The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, *32*(suppl_1), D267-D270.
34. Yang, L., & Agarwal, P. (2011). Systematic drug repositioning based on clinical side-effects. *PloS one*, *6*(12), e28025.

35. Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., & Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, *6*(1), 343.
36. Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., ... & Klein, T. E. (2012). Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics*, *92*(4), 414-417.
37. Hoehndorf, R., Oellrich, A., Dumontier, M., Kelso, J., Rebholz-Schuhmann, D., & Herre, H. (2010). Relations as patterns: bridging the gap between OBO and OWL. *BMC bioinformatics*, *11*, 1-8.
38. Hoehndorf, R., Schofield, P. N., & Gkoutos, G. V. (2011). PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic acids research*, *39*(18), e119-e119.
39. Sanseau, P., Agarwal, P., Barnes, M. R., Pastinen, T., Richards, J. B., Cardon, L. R., & Mooser, V. (2012). Use of genome-wide association studies for drug repositioning. *Nature biotechnology*, *30*(4), 317-320.
40. Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N. P., Rieder, M. J., ... & Orho-Melander, M. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature genetics*, *40*(2), 189-197.
41. Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., ... & Parkes, M. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics*, *42*(12), 1118-1125.
42. Chiang, A. P., & Butte, A. J. (2009). Systematic evaluation of drug–disease relationships to identify leads for novel drug uses. *Clinical Pharmacology & Therapeutics*, *86*(5), 507-510.
43. Li, Y., & Agarwal, P. (2009). A pathway-based view of human diseases and disease relationships. *PloS one*, *4*(2), e4346.
44. Suthram, S., Dudley, J. T., Chiang, A. P., Chen, R., Hastie, T. J., & Butte, A. J. (2010). Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS computational biology*, *6*(2), e1000662.

45. Gottlieb, A., Stein, G. Y., Ruppin, E., & Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular systems biology*, *7*(1), 496.
46. Napolitano, F., Zhao, Y., Moreira, V. M., Tagliaferri, R., Kere, J., D'Amato, M., & Greco, D. (2013). Drug repositioning: a machine-learning approach through data integration. *Journal of cheminformatics*, *5*, 1-9.
47. Li, Y. Y., An, J., & Jones, S. J. (2006). A large-scale computational approach to drug repositioning. *Genome Informatics*, *17*(2), 239-247.
48. Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nature chemical biology*, *4*(11), 682-690.
49. Chang, H. Y., & Qi, L. S. (2023). Reversing the central dogma: RNA-guided control of DNA in epigenetics and genome editing. *Molecular cell*, *83*(3), 442-451.
50. AD, G. (2009). Epigenetics: a landscape takes shape. *Cell*, *128*, 635-638.
51. Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., ... & Guan, K. L. (2010). Regulation of cellular metabolism by protein lysine acetylation. *Science*, *327*(5968), 1000-1004.
52. Yoshida, M., Furumai, R., Nishiyama, M., Komatsu, Y., Nishino, N., & Horinouchi, S. (2001). Histone deacetylase as a new target for cancer chemotherapy. *Cancer chemotherapy and pharmacology*, *48*, S20-S26.
53. Glozak, M. A., Sengupta, N., Zhang, X., & Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. *gene*, *363*, 15-23.
54. Das, C., & Kundu, T. (2005). Transcriptional Regulation by the Acetylation of Nonhistone Proteins in Humans-A New Target for Therapeutics. *IUBMB life*, *57*(3), 137-149.
55. Juan, L. J., Shia, W. J., Chen, M. H., Yang, W. M., Seto, E., Lin, Y. S., & Wu, C. W. (2000). Histone deacetylases specifically down-regulate p53-dependent gene activation. *Journal of Biological Chemistry*, *275*(27), 20436-20443.
56. Shakespear, M. R., Halili, M. A., Irvine, K. M., Fairlie, D. P., & Sweet, M. J. (2011). Histone deacetylases as regulators of inflammation and immunity. *Trends in immunology*, *32*(7), 335-343.
57. Bagchi, R. A., & Weeks, K. L. (2019). Histone deacetylases in cardiovascular and metabolic diseases. *Journal of Molecular and Cellular Cardiology*, *130*, 151-159.
58. Zhang, X. H., Qin-Ma, Wu, H. P., Khamis, M. Y., Li, Y. H., Ma, L. Y., & Liu, H. M. (2021). A review of progress in histone deacetylase 6 inhibitors research: structural

specificity and functional diversity. *Journal of medicinal chemistry*, *64*(3), 1362-1391.

59. Sarkar, R., Banerjee, S., Amin, S. A., Adhikari, N., & Jha, T. (2020). Histone deacetylase 3 (HDAC3) inhibitors as anticancer agents: A review. *European journal of medicinal chemistry*, *192*, 112171.

60. Ruijter, A. J. M. de, van Gennip, A. H., Caron, H. N., Kemp, S., & van Kuilenburg, A. B. P. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. *The Biochemical Journal*, *370*(3), 737–749. https://doi.org/10.1042/bj20021321

61. Yuan, H., & Marmorstein, R. (2012). Structural basis for sirtuin activity and inhibition. *Journal of Biological Chemistry*, *287*(51), 42428-42435.

62. Lin, H. Y., Chen, C. S., Lin, S. P., Weng, J. R., & Chen, C. S. (2006). Targeting histone deacetylase in cancer therapy. *Medicinal research reviews*, *26*(4), 397-413.

63. Pulya, S., Amin, S. A., Adhikari, N., Biswas, S., Jha, T., & Ghosh, B. (2021). HDAC6 as privileged target in drug discovery: A perspective. *Pharmacological research*, *163*, 105274.

64. Duvic, M., Talpur, R., Ni, X., Zhang, C., Hazarika, P., Kelly, C., ... & Frankel, S. R. (2007). Phase 2 trial of oral vorinostat (suberoylanilide hydroxamic acid, SAHA) for refractory cutaneous T-cell lymphoma (CTCL). *Blood*, *109*(1), 31-39.

65. DeAngelo, D. J., Mesa, R. A., Fiskus, W., Tefferi, A., Paley, C., Wadleigh, M., ... & Bhalla, K. N. (2013). Phase II trial of panobinostat, an oral pan-deacetylase inhibitor in patients with primary myelofibrosis, post–essential thrombocythaemia, and post–polycythaemia vera myelofibrosis. *British journal of haematology*, *162*(3), 326-335.

66. Grozinger, C. M., Hassig, C. A., & Schreiber, S. L. (1999). Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proceedings of the National Academy of Sciences*, *96*(9), 4868-4873.

67. Li, T., Zhang, C., Hassan, S., Liu, X., Song, F., Chen, K., ... & Yang, J. (2018). Histone deacetylase 6 in cancer. *Journal of hematology & oncology*, *11*, 1-10.

68. Brindisi, M., Saraswati, A. P., Brogi, S., Gemma, S., Butini, S., & Campiani, G. (2019). Old but gold: tracking the new guise of histone deacetylase 6 (HDAC6) enzyme as a biomarker and therapeutic target in rare diseases. *Journal of medicinal chemistry*, *63*(1), 23-39.

69. Witt, O., Deubzer, H. E., Milde, T., & Oehme, I. (2009). HDAC family: What are the cancer relevant targets?. *Cancer letters*, *277*(1), 8-21.
70. Seto, E., & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harbor perspectives in biology*, *6*(4), a018713.
71. Shinsky, S. A., & Christianson, D. W. (2018). Polyamine deacetylase structure and catalysis: prokaryotic acetylpolyamine amidohydrolase and eukaryotic HDAC10. *Biochemistry*, *57*(22), 3105-3114.
72. Yalcin, G. (2018). Sirtuins and neurodegeneration. *Journal of Neurology & Neuromedicine*, *3*(1).
73. Liu, S. S., Wu, F., Jin, Y. M., Chang, W. Q., & Xu, T. M. (2020). HDAC11: a rising star in epigenetics. *Biomedicine & pharmacotherapy*, *131*, 110607.
74. Haberland, M., Johnson, A., Mokalled, M. H., Montgomery, R. L., & Olson, E. N. (2009). Genetic dissection of histone deacetylase requirement in tumor cells. *Proceedings of the National Academy of Sciences*, *106*(19), 7751-7755.
75. Verdel, A., & Khochbin, S. (1999). Identification of a new family of higher eukaryotic histone deacetylases: coordinate expression of differentiation-dependent chromatin modifiers. *Journal of Biological Chemistry*, *274*(4), 2440-2445.
76. Grozinger, C. M., Hassig, C. A., & Schreiber, S. L. (1999). Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proceedings of the National Academy of Sciences*, *96*(9), 4868-4873.
77. Miyake, Y., Keusch, J. J., Wang, L., Saito, M., Hess, D., Wang, X., ... & Matthias, P. (2016). Structural insights into HDAC6 tubulin deacetylation and its selective inhibition. *Nature chemical biology*, *12*(9), 748-754.
78. Ran, J., & Zhou, J. (2019). Targeted inhibition of histone deacetylase 6 in inflammatory diseases. *Thoracic Cancer*, *10*(3), 405-412.
79. Hubbert, C., Guardiola, A., Shao, R., Kawaguchi, Y., Ito, A., Nixon, A., ... & Yao, T. P. (2002). HDAC6 is a microtubule-associated deacetylase. *Nature*, *417*(6887), 455-458.
80. Hard, R. L., Liu, J., Shen, J., Zhou, P., & Pei, D. (2010). HDAC6 and Ubp-M BUZ domains recognize specific C-terminal sequences of proteins. *Biochemistry*, *49*(50), 10737-10746.
81. Boyault, C., Sadoul, K., Pabion, M., & Khochbin, S. (2007). HDAC6, at the crossroads between cytoskeleton and cell signaling by acetylation and ubiquitination. *Oncogene*, *26*(37), 5468-5476.

82. Osko, J. D., & Christianson, D. W. (2019). Structural basis of catalysis and inhibition of HDAC6 CD1, the enigmatic catalytic domain of histone deacetylase 6. *Biochemistry*, *58*(49), 4912-4924.
83. Porter, N. J., Mahendran, A., Breslow, R., & Christianson, D. W. (2017). Unusual zinc-binding mode of HDAC6-selective hydroxamate inhibitors. *Proceedings of the National Academy of Sciences*, *114*(51), 13459-13464.
84. Porter, N. J., Wagner, F. F., & Christianson, D. W. (2018). Entropy as a driver of selectivity for inhibitor binding to histone deacetylase 6. *Biochemistry*, *57*(26), 3916-3924.
85. Watson, P. R., Bai, P., Wang, C., Cragin, A. D., Hooker, J. M., & Christianson, D. W. (2022). Aromatic ring fluorination patterns modulate inhibitory potency of fluorophenylhydroxamates complexed with histone deacetylase 6. *Biochemistry*, *61*(18), 1945-1954.
86. Watson, P. R., Gupta, S., Hosseinzadeh, P., Brown, B. P., Baker, D., & Christianson, D. W. (2023). Macrocyclic octapeptide binding and inferences on protein substrate binding to histone deacetylase 6. *ACS chemical biology*, *18*(4), 959-968.
87. Osko, J. D., & Christianson, D. W. (2020). Binding of inhibitors to active-site mutants of CD1, the enigmatic catalytic domain of histone deacetylase 6. *Acta Crystallographica Section F: Structural Biology Communications*, *76*(9), 428-437.
88. 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.
89. Mackwitz, M. K., Hamacher, A., Osko, J. D., Held, J., Schöler, A., Christianson, D. W., ... & Hansen, F. K. (2018). Multicomponent synthesis and binding mode of imidazo [1, 2-a] pyridine-capped selective HDAC6 inhibitors. *Organic letters*, *20*(11), 3255-3258.
90. Saraswati, A. P., Relitti, N., Brindisi, M., Osko, J. D., Chemi, G., Federico, S., ... & Campiani, G. (2020). Spiroindoline-capped selective HDAC6 inhibitors: design, synthesis, structural analysis, and biological evaluation. *ACS Medicinal Chemistry Letters*, *11*(11), 2268-2276.
91. Olaoye, O. O., Watson, P. R., Nawar, N., Geletu, M., Sedighi, A., Bukhari, S., ... & Gunning, P. T. (2021). Unique molecular interaction with the histone deacetylase 6 catalytic tunnel: crystallographic and biological characterization of a model chemotype. *Journal of medicinal chemistry*, *64*(5), 2691-2704.

92. Reßing, N., Sönnichsen, M., Osko, J. D., Schöler, A., Schliehe-Diecks, J., Skerhut, A., ... & Hansen, F. K. (2020). Multicomponent synthesis, binding mode, and structure–activity relationship of selective histone deacetylase 6 (HDAC6) inhibitors with bifurcated capping groups. *Journal of medicinal chemistry*, *63*(18), 10339-10351.
93. Langousis, G., Sanchez, J., Kempf, G., & Matthias, P. (2022). Expression and Crystallization of HDAC6 Tandem Catalytic Domains. In *HDAC/HAT Function Assessment and Inhibitor Development: Methods and Protocols* (pp. 467-480). New York, NY: Springer US.
94. Garcha, H. K., Nawar, N., Sorger, H., Erdogan, F., Aung, M. M. K., Sedighi, A., ... & Gunning, P. T. (2022). High Efficacy and Drug Synergy of HDAC6-Selective Inhibitor NN-429 in Natural Killer (NK)/T-Cell Lym. *Pharmaceutics*, *15*(11)).
95. Campiani, G., Cavella, C., Osko, J. D., Brindisi, M., Relitti, N., Brogi, S., ... & Prasse, A. (2021). Harnessing the role of HDAC6 in idiopathic pulmonary fibrosis: design, synthesis, structural analysis, and biological evaluation of potent inhibitors. *Journal of medicinal chemistry*, *64*(14), 9960-9988.
96. Hai, Y., & Christianson, D. W. (2016). Histone deacetylase 6 structure and molecular basis of catalysis and inhibition. *Nature chemical biology*, *12*(9), 741-747.
97. Cellupica, E., Caprini, G., Cordella, P., Cukier, C., Fossati, G., Marchini, M., ... & Steinkühler, C. (2023). Difluoromethyl-1, 3, 4-oxadiazoles are slow-binding substrate analog inhibitors of histone deacetylase 6 with unprecedented isotype selectivity. *Journal of Biological Chemistry*, *299*(1).
98. Reßing, N., Schliehe-Diecks, J., Watson, P. R., Sönnichsen, M., Cragin, A. D., Schöler, A., ... & Hansen, F. K. (2022). Development of fluorinated peptoid-based histone deacetylase (HDAC) inhibitors for therapy-resistant acute leukemia. *Journal of medicinal chemistry*, *65*(22), 15457-15472.
99. Vögerl, K., Ong, N., Senger, J., Herp, D., Schmidtkunz, K., Marek, M., ... & Bracher, F. (2019). Synthesis and biological investigation of phenothiazine-based benzhydroxamic acids as selective histone deacetylase 6 inhibitors. *Journal of medicinal chemistry*, *62*(3), 1138-1166.
100. Hosseinzadeh, P., Watson, P. R., Craven, T. W., Li, X., Rettie, S., Pardo-Avila, F., ... & Baker, D. (2021). Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites. *Nature Communications*, *12*(1), 3384.

101. Noonepalle, S., Shen, S., Ptáček, J., Tavares, M. T., Zhang, G., Stránský, J., ... & Villagra, A. (2020). Rational design of suprastat: a novel selective histone deacetylase 6 inhibitor with the ability to potentiate immunotherapy in melanoma models. *Journal of Medicinal Chemistry*, *63*(18), 10246-10262.
102. Shen, S., Picci, C., Ustinova, K., Benoy, V., Kutil, Z., Zhang, G., ... & Kozikowski, A. P. (2021). Tetrahydroquinoline-capped histone deacetylase 6 inhibitor SW-101 ameliorates pathological phenotypes in a Charcot–Marie–Tooth type 2A mouse model. *Journal of Medicinal Chemistry*, *64*(8), 4810-4840.
103. Porter, N. J., Osko, J. D., Diedrich, D., Kurz, T., Hooker, J. M., Hansen, F. K., & Christianson, D. W. (2018). Histone deacetylase 6-selective inhibitors and the influence of capping groups on hydroxamate-zinc denticity. *Journal of medicinal chemistry*, *61*(17), 8054-8060.
104. Porter, N. J., Shen, S., Barinka, C., Kozikowski, A. P., & Christianson, D. W. (2018). Molecular basis for the selective inhibition of histone deacetylase 6 by a mercaptoacetamide inhibitor. *ACS Medicinal Chemistry Letters*, *9*(12), 1301-1305.
105. Shen, S., Svoboda, M., Zhang, G., Cavasin, M. A., Motlova, L., McKinsey, T. A., ... & Kozikowski, A. P. (2020). Structural and in vivo characterization of Tubastatin A, a widely used histone deacetylase 6 inhibitor. *ACS medicinal chemistry letters*, *11*(5), 706-712.
106. Shen, S., Hadley, M., Ustinova, K., Pavlicek, J., Knox, T., Noonepalle, S., ... & Villagra, A. (2019). Discovery of a new isoxazole-3-hydroxamate-based histone deacetylase 6 inhibitor SS-208 with antitumor activity in syngeneic melanoma mouse models. *Journal of medicinal chemistry*, *62*(18), 8557-8577.
107. Bhatia, S., Krieger, V., Groll, M., Osko, J. D., Reßing, N., Ahlert, H., ... & Hansen, F. K. (2018). Discovery of the first-in-class dual histone deacetylase–proteasome inhibitor. *Journal of medicinal chemistry*, *61*(22), 10299-10309.
108. Harding, R. J., Franzoni, I., Mann, M. K., Szewczyk, M. M., Mirabi, B., Ferreira de Freitas, R., ... & Arrowsmith, C. H. (2023). Discovery and characterization of a chemical probe targeting the zinc-finger ubiquitin-binding domain of HDAC6. *Journal of Medicinal Chemistry*, *66*(15), 10273-10288.
109. Ferreira de Freitas, R., Harding, R. J., Franzoni, I., Ravichandran, M., Mann, M. K., Ouyang, H., ... & Schapira, M. (2018). Identification and structure–activity relationship of HDAC6 zinc-finger ubiquitin binding domain inhibitors. *Journal of Medicinal Chemistry*, *61*(10), 4517-4527.

110. Harding, R. J., Ferreira de Freitas, R., Collins, P., Franzoni, I., Ravichandran, M., Ouyang, H., ... & Arrowsmith, C. H. (2017). Small molecule antagonists of the interaction between the histone deacetylase 6 zinc-finger domain and ubiquitin. *Journal of Medicinal Chemistry*, *60*(21), 9090-9096.
111. Ouyang, H., Ali, Y. O., Ravichandran, M., Dong, A., Qiu, W., MacKenzie, F., ... & Zhai, R. G. (2012). Protein aggregates are recruited to aggresome by histone deacetylase 6 via unanchored ubiquitin C termini. *Journal of Biological Chemistry*, *287*(4), 2317-2327.
112. Wang, L., Moreira, E. A., Kempf, G., Miyake, Y., Esteves, B. I. O., Fahmi, A., ... & Matthias, P. (2022). Disrupting the HDAC6-ubiquitin interaction impairs infection by influenza and Zika virus and cellular stress pathways. *Cell reports*, *39*(4).
113. Li, Y., Shin, D., & Kwon, S. H. (2013). Histone deacetylase 6 plays a role as a distinct regulator of diverse cellular processes. *The FEBS journal*, *280*(3), 775-793.
114. Li, T., Zhang, C., Hassan, S., Liu, X., Song, F., Chen, K., ... & Yang, J. (2018). Histone deacetylase 6 in cancer. *Journal of hematology & oncology*, *11*, 1-10.
115. Liang, T., & Fang, H. (2018). Structure, functions and selective inhibitors of HDAC6. *Current Topics in Medicinal Chemistry*, *18*(28), 2429-2447.
116. Y. Kawaguchi, J.J. Kovacs, A. McLaurin, J.M. Vance, A. Ito, T.P. Yao, The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress, Cell 115 (2003) 727–738.
117. Haakenson, J., & Zhang, X. (2013). HDAC6 and ovarian cancer. *International journal of molecular sciences*, *14*(5), 9514-9535.
118. Aldana-Masangkay, G. I., & Sakamoto, K. M. (2011). The role of HDAC6 in cancer. *BioMed Research International*, *2011*(1), 875824.
119. Yang, P., Zhang, L., Zhang, Y., Zhang, J., & Xu, W. (2013). HDAC6: physiological function and its selective inhibitors for cancer treatment. *Drug discoveries & therapeutics*, *7*(6), 233-242.
120. Liu, J., Luan, W., Zhang, Y., Gu, J., Shi, Y., Yang, Y., ... & Qi, F. (2018). HDAC6 interacts with PTPN1 to enhance melanoma cells progression. *Biochemical and biophysical research communications*, *495*(4), 2630-2636.
121. Wickström, S. A., Masoumi, K. C., Khochbin, S., Fässler, R., & Massoumi, R. (2010). CYLD negatively regulates cell-cycle progression by inactivating HDAC6 and increasing the levels of acetylated tubulin. *The EMBO Journal*, *29*(1), 131-144.

122. Pandey, U. B., Nie, Z., Batlevi, Y., McCray, B. A., Ritson, G. P., Nedelsky, N. B., ... & Taylor, J. P. (2007). HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. *Nature*, *447*(7146), 860-864.
123. Zhang, L., Sheng, S., & Qin, C. (2013). The role of HDAC6 in Alzheimer's disease. *Journal of Alzheimer's Disease*, *33*(2), 283-295.
124. Qureshi, T., & Chinnathambi, S. (2022). Histone deacetylase-6 modulates Tau function in Alzheimer's disease. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1869*(8), 119275.
125. Choi, H., Kim, H. J., Yang, J., Chae, S., Lee, W., Chung, S., ... & Mook-Jung, I. (2020). Acetylation changes tau interactome to degrade tau in Alzheimer's disease animal and organoid models. *Aging Cell*, *19*(1), e13081.
126. Shen, S., & Kozikowski, A. P. (2020). A patent review of histone deacetylase 6 inhibitors in neurodegenerative diseases (2014-2019). *Expert opinion on therapeutic patents*, *30*(2), 121-136.
127. Ding, H., Dolan, P. J., & Johnson, G. V. (2008). Histone deacetylase 6 interacts with the microtubule-associated protein tau. *Journal of neurochemistry*, *106*(5), 2119-2130.
128. Parmigiani, R. B., Xu, W. S., Venta-Perez, G., Erdjument-Bromage, H., Yaneva, M., Tempst, P., & Marks, P. A. (2008). HDAC6 is a specific deacetylase of peroxiredoxins and is involved in redox regulation. *Proceedings of the National Academy of Sciences*, *105*(28), 9633-9638.
129. Johri, A., & Beal, M. F. (2012). Mitochondrial dysfunction in neurodegenerative diseases. *Journal of Pharmacology and Experimental Therapeutics*, *342*(3), 619-630.
130. Shukla, S., & Tekwani, B. L. (2020). Histone deacetylases inhibitors in neurodegenerative diseases, neuroprotection and neuronal differentiation. *Frontiers in pharmacology*, *11*, 537.
131. Simões-Pires, C., Zwick, V., Nurisso, A., Schenker, E., Carrupt, P. A., & Cuendet, M. (2013). HDAC6 as a target for neurodegenerative diseases: what makes it different from the other HDACs?. *Molecular neurodegeneration*, *8*, 1-16.
132. Du, Y., Wang, F., Zou, J., Le, W., Dong, Q., Wang, Z., ... & Li, Y. (2014). Histone deacetylase 6 regulates cytotoxic α-synuclein accumulation through induction of the heat shock response. *Neurobiology of aging*, *35*(10), 2316-2328.

133. Ding, W. X., & Yin, X. M. (2008). Sorting, recognition and activation of the misfolded protein degradation pathways through macroautophagy and the proteasome. *Autophagy*, *4*(2), 141-150.
134. Olzmann, J. A., Li, L., Chudaev, M. V., Chen, J., Perez, F. A., Palmiter, R. D., & Chin, L. S. (2007). Parkin-mediated K63-linked polyubiquitination targets misfolded DJ-1 to aggresomes via binding to HDAC6. *The Journal of cell biology*, *178*(6), 1025-1038.
135. Francelle, L., Outeiro, T. F., & Rappold, G. A. (2020). Inhibition of HDAC6 activity protects dopaminergic neurons from alpha-synuclein toxicity. *Scientific reports*, *10*(1), 6064.
136. Lin, C. H., Tallaksen-Greene, S., Chien, W. M., Cearley, J. A., Jackson, W. S., Crouse, A. B., ... & Detloff, P. J. (2001). Neurological abnormalities in a knock-in mouse model of Huntington's disease. *Human molecular genetics*, *10*(2), 137-144.
137. Kumar, V., Kundu, S., Singh, A., & Singh, S. (2022). Understanding the role of histone deacetylase and their inhibitors in neurodegenerative disorders: current targets and future perspective. *Current neuropharmacology*, *20*(1), 158.
138. Pineda, J. R., Pardo, R., Zala, D., Yu, H., Humbert, S., & Saudou, F. (2009). Genetic and pharmacological inhibition of calcineurin corrects the BDNF transport defect in Huntington's disease. *Molecular brain*, *2*, 1-11.
139. Dompierre, J. P., Godin, J. D., Charrin, B. C., Cordelieres, F. P., King, S. J., Humbert, S., & Saudou, F. (2007). Histone deacetylase 6 inhibition compensates for the transport deficit in Huntington's disease by increasing tubulin acetylation. *Journal of Neuroscience*, *27*(13), 3571-3583.
140. Bobrowska, A., Paganetti, P., Matthias, P., & Bates, G. P. (2011). Hdac6 knock-out increases tubulin acetylation but does not modify disease progression in the R6/2 mouse model of Huntington's disease. *PloS one*, *6*(6), e20696.
141. Iwata, A., Riley, B. E., Johnston, J. A., & Kopito, R. R. (2005). HDAC6 and microtubules are required for autophagic degradation of aggregated huntingtin. *Journal of Biological Chemistry*, *280*(48), 40282-40292.
142. Chahrour, M., & Zoghbi, H. Y. (2007). The story of Rett syndrome: from clinic to neurobiology. *Neuron*, *56*(3), 422-437.
143. Brindisi, M., Saraswati, A. P., Brogi, S., Gemma, S., Butini, S., & Campiani, G. (2019). Old but gold: tracking the new guise of histone deacetylase 6 (HDAC6)

enzyme as a biomarker and therapeutic target in rare diseases. *Journal of medicinal chemistry*, *63*(1), 23-39.

144. Gold, W. A., Lacina, T. A., Cantrill, L. C., & Christodoulou, J. (2015). MeCP2 deficiency is associated with reduced levels of tubulin acetylation and can be restored using HDAC6 inhibitors. *Journal of Molecular Medicine*, *93*, 63-72.
145. Butler, K. V., Kalin, J., Brochier, C., Vistoli, G., Langley, B., & Kozikowski, A. P. (2010). Rational design and simple chemistry yield a superior, neuroprotective HDAC6 inhibitor, tubastatin A. *Journal of the American Chemical Society*, *132*(31), 10842-10846.
146. Chen, S., Owens, G. C., Makarenkova, H., & Edelman, D. B. (2010). HDAC6 regulates mitochondrial transport in hippocampal neurons. *PloS one*, *5*(5), e10848.
147. Shen, S., & Kozikowski, A. P. (2020). A patent review of histone deacetylase 6 inhibitors in neurodegenerative diseases (2014-2019). *Expert opinion on therapeutic patents*, *30*(2), 121-136.
148. d'Ydewalle, C., Krishnan, J., Chiheb, D. M., Van Damme, P., Irobi, J., Kozikowski, A. P., ... & Van Den Bosch, L. (2011). HDAC6 inhibitors reverse axonal loss in a mouse model of mutant HSPB1–induced Charcot-Marie-Tooth disease. *Nature medicine*, *17*(8), 968-974.
149. Mo, Z., Zhao, X., Liu, H., Hu, Q., Chen, X. Q., Pham, J., ... & Yang, X. L. (2018). Aberrant GlyRS-HDAC6 interaction linked to axonal transport deficits in Charcot-Marie-Tooth neuropathy. *Nature communications*, *9*(1), 1007.
150. Guo, W., Naujock, M., Fumagalli, L., Vandoorne, T., Baatsen, P., Boon, R., ... & Van Den Bosch, L. (2017). HDAC6 inhibition reverses axonal transport defects in motor neurons derived from FUS-ALS patients. *Nature communications*, *8*(1), 861.
151. Rossaert, E., Pollari, E., Jaspers, T., Van Helleputte, L., Jarpe, M., Van Damme, P., ... & Van Den Bosch, L. (2019). Restoration of histone acetylation ameliorates disease and metabolic abnormalities in a FUS mouse model. *Acta neuropathologica communications*, *7*, 1-19.
152. Moreno-Gonzalo, O., Ramírez-Huesca, M., Blas-Rus, N., Cibrián, D., Saiz, M. L., Jorge, I., ... & Sánchez-Madrid, F. (2017). HDAC6 controls innate immune and autophagy responses to TLR-mediated signalling by the intracellular bacteria Listeria monocytogenes. *PLoS pathogens*, *13*(12), e1006799.
153. Cheng, F., Lienlaf, M., Perez-Villarroel, P., Wang, H. W., Lee, C., Woan, K., ... & Villagra, A. (2014). Divergent roles of histone deacetylase 6 (HDAC6) and histone

deacetylase 11 (HDAC11) on the transcriptional regulation of IL10 in antigen presenting cells. *Molecular immunology*, *60*(1), 44-53.

154. de Zoeten, E. F., Wang, L., Butler, K., Beier, U. H., Akimova, T., Sai, H., ... & Hancock, W. W. (2011). Histone deacetylase 6 and heat shock protein 90 control the functions of Foxp3+ T-regulatory cells. *Molecular and cellular biology*.
155. Huo, L., Li, D., Sun, X., Shi, X., Karna, P., Yang, W., ... & Zhou, J. (2011). Regulation of Tat acetylation and transactivation activity by the microtubule-associated deacetylase HDAC6. *Journal of Biological Chemistry*, *286*(11), 9280-9286.
156. Valenzuela-Fernández, A., Alvarez, S., Gordon-Alonso, M., Barrero, M., Ursa, A., Cabrero, J. R., ... & Sánchez-Madrid, F. (2005). Histone deacetylase 6 regulates human immunodeficiency virus type 1 infection. *Molecular biology of the cell*, *16*(11), 5445-5454.
157. Zhu, J., Coyne, C. B., & Sarkar, S. N. (2011). PKC alpha regulates Sendai virus-mediated interferon induction through HDAC6 and β-catenin. *The EMBO journal*, *30*(23), 4838-4849.
158. Firestein, G. S. (2003). Evolving concepts of rheumatoid arthritis. *Nature*, *423*(6937), 356-361.
159. Li, G., Jiang, H., Chang, M., Xie, H., & Hu, L. (2011). HDAC6 α-tubulin deacetylase: a potential therapeutic target in neurodegenerative diseases. *Journal of the neurological sciences*, *304*(1-2), 1-8.
160. Li, M., Hu, W., Wang, R., Li, Z., Yu, Y., Zhuo, Y., ... & Zhu, Y. (2022). Sp1 S-sulfhydration induced by hydrogen sulfide inhibits inflammation via HDAC6/MyD88/NF-κB signaling pathway in adjuvant-induced arthritis. *Antioxidants*, *11*(4), 732.
161. Oh, B. R., Suh, D. H., Bae, D., Ha, N., Choi, Y. I., Yoo, H. J., ... & Song, Y. W. (2017). Therapeutic effect of a novel histone deacetylase 6 inhibitor, CKD-L, on collagen-induced arthritis in vivo and regulatory T cells in rheumatoid arthritis in vitro. *Arthritis research & therapy*, *19*, 1-16.
162. Lee, J. Y., Ma, H. W., Kim, J. H., Park, I. S., Son, M., Ryu, K. H., ... & Cheon, J. H. (2023). Novel histone deacetylase 6 inhibitor confers anti-inflammatory effects and enhances gut barrier function. *Gut and liver*, *17*(5), 766.
163. Lee, J. W., Lee, S. M., Chun, J., Im, J. P., Seo, S. K., Ha, N., ... & Kim, J. S. (2020). Novel histone deacetylase 6 inhibitor CKD-506 inhibits NF-κB signaling in

intestinal epithelial cells and macrophages and ameliorates acute and chronic murine colitis. *Inflammatory bowel diseases*, *26*(6), 852-862.

164. Nielsen, O. H., Coskun, M., Steenholdt, C., & Rogler, G. (2015). The role and advances of immunomodulator therapy for inflammatory bowel disease. *Expert review of gastroenterology & hepatology*, *9*(2), 177-189.
165. Do, A., Reid, R. C., Lohman, R. J., Sweet, M. J., Fairlie, D. P., & Iyer, A. (2017). An HDAC6 inhibitor confers protection and selectively inhibits B-cell infiltration in DSS-induced colitis in mice. *Journal of Pharmacology and Experimental Therapeutics*, *360*(1), 140-151.
166. Liu, T., Wang, R., Xu, H., Song, Y., & Qi, Y. (2017). A highly potent and selective histone deacetylase 6 inhibitor prevents DSS-induced colitis in mice. *Biological and Pharmaceutical Bulletin*, *40*(6), 936-940.
167. Ren, Y., Su, X., Kong, L., Li, M., Zhao, X., Yu, N., & Kang, J. (2016). Therapeutic effects of histone deacetylase inhibitors in a murine asthma model. *Inflammation Research*, *65*, 995-1008.
168. Lam, H. C., Cloonan, S. M., Bhashyam, A. R., Haspel, J. A., Singh, A., Sathirapongsasuti, J. F., ... & Choi, A. M. (2013). Histone deacetylase 6–mediated selective autophagy regulates COPD-associated cilia dysfunction. *The Journal of clinical investigation*, *123*(12), 5212-5230.
169. Barone, S., Cassese, E., Alfano, A. I., Brindisi, M., & Summa, V. (2022). Chasing a breath of fresh air in cystic fibrosis (CF): therapeutic potential of selective HDAC6 inhibitors to tackle multiple pathways in CF pathophysiology. *Journal of Medicinal Chemistry*, *65*(4), 3080-3097.
170. Huaux, F., Noel, S., Dhooghe, B., Panin, N., Lo Re, S., Lison, D., ... & Leal, T. (2013). Dysregulated proinflammatory and fibrogenic phenotype of fibroblasts in cystic fibrosis. *PloS one*, *8*(5), e64341.
171. Hilliard, T. N., Regamey, N., Shute, J. K., Nicholson, A. G., Alton, E. W., Bush, A., & Davies, J. C. (2007). Airway remodelling in children with cystic fibrosis. *Thorax*, *62*(12), 1074-1080.
172. Shan, B., Yao, T. P., Nguyen, H. T., Zhuo, Y., Levy, D. R., Klingsberg, R. C., ... & Lasky, J. A. (2008). Requirement of HDAC6 for transforming growth factor-β1-induced epithelial-mesenchymal transition. *Journal of Biological Chemistry*, *283*(30), 21065-21073.

173. Venkatachalam, M. A., Weinberg, J. M., Kriz, W., & Bidani, A. K. (2015). Failed tubule recovery, AKI-CKD transition, and kidney disease progression. *Journal of the American Society of Nephrology*, *26*(8), 1765-1776.
174. Chatzizisis, Y. S., Misirli, G., Hatzitolios, A. I., & Giannoglou, G. D. (2008). The syndrome of rhabdomyolysis: complications and treatment. *European journal of internal medicine*, *19*(8), 568-574.
175. Sever, M. S., Vanholder, R., & Lameire, N. (2006). Management of crush-related injuries after disasters. *New England Journal of Medicine*, *354*(10), 1052-1063.
176. Bosch, X., Poch, E., & Grau, J. M. (2009). Rhabdomyolysis and acute kidney injury. *The New England Journal of Medicine, 361*(1), 62–72.
177. Feng, Y., Ma, L., Liu, L., Hong, H. G., Zhang, X., Guo, F., ... & Fu, P. (2016). Rhabdomyolysis induced AKI via the regulation of endoplasmic reticulum stress and oxidative stress in PTECs. *RSC advances*, *6*(111), 109639-109648.
178. Li, Z., Zhuang, M., Zhang, L., Zheng, X., Yang, P., & Li, Z. (2016). Acetylation modification regulates GRP78 secretion in colon cancer cells. *Scientific Reports*, *6*(1), 30406.
179. Shi, Y., Xu, L., Tang, J., Fang, L., Ma, S., Ma, X., ... & Liu, N. (2017). Inhibition of HDAC6 protects against rhabdomyolysis-induced acute kidney injury. *American Journal of Physiology-Renal Physiology*, *312*(3), F502-F515.
180. Dang, S., Yu, Z. M., Zhang, C. Y., Zheng, J., Li, K. L., Wu, Y., ... & Wang, R. X. (2015). Autophagy promotes apoptosis of mesenchymal stem cells under inflammatory microenvironment. *Stem cell research & therapy*, *6*, 1-9.
181. Su, Z., Yang, Z., Xu, Y., Chen, Y., & Yu, Q. (2015). Apoptosis, autophagy, necroptosis, and cancer metastasis. *Molecular cancer*, *14*, 1-14.
182. Youle, R. J., & Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nature reviews Molecular cell biology*, *9*(1), 47-59.
183. Savitskaya, M. A., & Onishchenko, G. E. (2015). Mechanisms of apoptosis. *Biochemistry (Moscow)*, *80*, 1393-1405.
184. Havasi, A., & Borkan, S. C. (2011). Apoptosis and acute kidney injury. *Kidney international*, *80*(1), 29-40.
185. Linkermann, A., Chen, G., Dong, G., Kunzendorf, U., Krautwald, S., & Dong, Z. (2014). Regulated cell death in AKI. *Journal of the American Society of Nephrology*, *25*(12), 2689-2701.

186. Taniguchi, M., & Yoshida, H. (2015). Endoplasmic reticulum stress in kidney function and disease. *Current opinion in nephrology and hypertension*, *24*(4), 345-350.
187. Gardner, B. M., & Walter, P. (2011). Unfolded proteins are Ire1-activating ligands that directly induce the unfolded protein response. *Science*, *333*(6051), 1891-1894.
188. Walter, P., & Ron, D. (2011). The unfolded protein response: from stress pathway to homeostatic regulation. *science*, *334*(6059), 1081-1086.
189. Feng, Y., Huang, R., Guo, F., Liang, Y., Xiang, J., Lei, S., ... & Fu, P. (2018). Selective histone deacetylase 6 inhibitor 23BB alleviated rhabdomyolysis-induced acute kidney injury by regulating endoplasmic reticulum stress and apoptosis. *Frontiers in Pharmacology*, *9*, 274.
190. Xu, J., Zhao, X., Jiang, X., He, L., Wu, X., Wang, J., ... & Zhang, M. (2022). Tubastatin A Improves Post-Resuscitation Myocardial Dysfunction by Inhibiting NLRP3-Mediated Pyroptosis Through Enhancing Transcription Factor EB Signaling. *Journal of the American Heart Association*, *11*(7), e024205.
191. Seungwha, P., Kim, J. K., Prashanta, S., & Eun-Kyeong, J. (2021). An update on the regulatory mechanisms of NLRP3 inflammasome activation. *Cellular & Molecular Immunology*, *18*(5), 1141-1160.
192. Settembre, C., Di Malta, C., Polito, V. A., Arencibia, M. G., Vetrini, F., Erdin, S., ... & Ballabio, A. (2011). TFEB links autophagy to lysosomal biogenesis. *science*, *332*(6036), 1429-1433.
193. Kim, S. H., Kim, G., Han, D. H., Lee, M., Kim, I., Kim, B., ... & Lee, M. S. (2017). Ezetimibe ameliorates steatohepatitis via AMP activated protein kinase-TFEB-mediated activation of autophagy and NLRP3 inflammasome inhibition. *Autophagy*, *13*(10), 1767-1781.
194. Dallavalle, S., Pisano, C., & Zunino, F. (2012). Development and therapeutic impact of HDAC6-selective inhibitors. *Biochemical pharmacology*, *84*(6), 756-765.
195. Yoshida, M., Kijima, M., Akita, M., & Beppu, T. (1990). Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *Journal of Biological Chemistry*, *265*(28), 17174-17179.
196. Tsuji, N., Kobayashi, M., Nagashima, K., Wakisaka, Y., & Koizumi, K. (1976). A new antifungal antibiotic, trichostatin. *The Journal of antibiotics*, *29*(1), 1-6.

197. Tan, S., & Liu, Z. P. (2015). Natural Products as Zinc-Dependent Histone Deacetylase Inhibitors. *ChemMedChem*, *10*(3), 441-450.
198. Butler, L. M., Agus, D. B., Scher, H. I., Higgins, B., Rose, A., Cordon-Cardo, C., ... & Richon, V. M. (2000). Suberoylanilide hydroxamic acid, an inhibitor of histone deacetylase, suppresses the growth of prostate cancer cells in vitro and in vivo. *Cancer research*, *60*(18), 5165-5170.
199. Rodriquez, M., Aquino, M., Bruno, I., Martino, G. D., Taddei, M., & Gomez-Paloma, L. (2006). Chemistry and biology of chromatin remodeling agents: state of art and future perspectives of HDAC inhibitors. *Current medicinal chemistry*, *13*(10), 1119-1139.
200. Haggarty, S. J., Koeller, K. M., Wong, J. C., Grozinger, C. M., & Schreiber, S. L. (2003). Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. *Proceedings of the National Academy of Sciences*, *100*(8), 4389-4394.
201. Haggarty, S. J., Koeller, K. M., Wong, J. C., Butcher, R. A., & Schreiber, S. L. (2003). Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chemistry & biology*, *10*(5), 383-396.
202. Aldana-Masangkay, G. I., Rodriguez-Gonzalez, A., Lin, T., Ikeda, A. K., Hsieh, Y. T., Kim, Y. M., ... & Sakamoto, K. M. (2011). Tubacin suppresses proliferation and induces apoptosis of acute lymphoblastic leukemia cells. *Leukemia & lymphoma*, *52*(8), 1544-1555.
203. Vogl, D. T., Raje, N., Hari, P., Jones, S. S., Supko, J. G., Leone, G., ... & Jagannath, S. (2014). Phase 1B results of ricolinostat (ACY-1215) combination therapy with bortezomib and dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma (MM). *Blood*, *124*(21), 4764.
204. Amengual, J. E., Johannet, P., Lombardo, M., Zullo, K., Hoehn, D., Bhagat, G., ... & O'Connor, O. A. (2015). Dual targeting of protein degradation pathways with the selective HDAC6 inhibitor ACY-1215 and bortezomib is synergistic in lymphoma. *Clinical Cancer Research*, *21*(20), 4663-4675.
205. Vogl, D. T., Raje, N., Jagannath, S., Richardson, P., Hari, P., Orlowski, R., ... & Lonial, S. (2017). Ricolinostat, the first selective histone deacetylase 6 inhibitor, in combination with bortezomib and dexamethasone for relapsed or refractory multiple myeloma. *Clinical Cancer Research*, *23*(13), 3307-3315.

206. Yee, A. J., Bensinger, W. I., Supko, J. G., Voorhees, P. M., Berdeja, J. G., Richardson, P. G., ... & Raje, N. S. (2016). Ricolinostat plus lenalidomide, and dexamethasone in relapsed or refractory multiple myeloma: a multicentre phase 1b trial. *The Lancet Oncology*, *17*(11), 1569-1578.
207. Cosenza, M., Civallero, M., Marcheselli, L., Sacchi, S., & Pozzi, S. (2017). Ricolinostat, a selective HDAC6 inhibitor, shows anti-lymphoma cell activity alone and in combination with bendamustine. *Apoptosis*, *22*, 827-840.
208. Cao, J., Lv, W., Wang, L., Xu, J., Yuan, P., Huang, S., ... & Hu, J. (2018). Ricolinostat (ACY-1215) suppresses proliferation and promotes apoptosis in esophageal squamous cell carcinoma via miR-30d/PI3K/AKT/mTOR and ERK pathways. *Cell death & disease*, *9*(8), 817.
209. Niesvizky, R., Richardson, P. G., Gabrail, N. Y., Madan, S., Yee, A. J., Quayle, S. N., ... & Raje, N. S. (2015). ACY-241, a novel, HDAC6 selective inhibitor: synergy with immunomodulatory (IMiD®) drugs in multiple myeloma (MM) cells and early clinical results (ACE-MM-200 Study).
210. Huang, P., Almeciga-Pinto, I., Jarpe, M., van Duzer, J. H., Mazitschek, R., Yang, M., ... & Quayle, S. N. (2017). Selective HDAC inhibition by ACY-241 enhances the activity of paclitaxel in solid tumor models. *Oncotarget*, *8*(2), 2694.
211. North, B. J., Almeciga-Pinto, I., Tamang, D., Yang, M., Jones, S. S., & Quayle, S. N. (2017). Enhancement of pomalidomide anti-tumor response with ACY-241, a selective HDAC6 inhibitor. *PLoS One*, *12*(3), e0173507.
212. Niesvizky, R., Richardson, P. G., Yee, A. J., Nooka, A. K., Raab, M. S., Shain, K. H., ... & Bensinger, W. I. (2016). Selective HDAC6 inhibitor ACY-241, an oral tablet, combined with pomalidomide and dexamethasone: safety and efficacy of escalation and expansion cohorts in patients with relapsed or relapsed-and-refractory multiple myeloma (ACE-MM-200 Study). *Blood*, *128*(22), 3307.
213. Yang, Z., Wang, T., Wang, F., Niu, T., Liu, Z., Chen, X., ... & Chen, L. (2016). Discovery of selective histone deacetylase 6 inhibitors using the quinazoline as the cap for the treatment of cancer. *Journal of medicinal chemistry*, *59*(4), 1455-1470.
214. Feng, Y., Huang, R., Guo, F., Liang, Y., Xiang, J., Lei, S., ... & Fu, P. (2018). Selective histone deacetylase 6 inhibitor 23BB alleviated rhabdomyolysis-induced acute kidney injury by regulating endoplasmic reticulum stress and apoptosis. *Frontiers in Pharmacology*, *9*, 274.

215. Vishwakarma, S., Iyer, L. R., Muley, M., Singh, P. K., Shastry, A., Saxena, A., ... & Narayanan, S. (2013). Tubastatin, a selective histone deacetylase 6 inhibitor shows anti-inflammatory and anti-rheumatic effects. *International immunopharmacology*, *16*(1), 72-78.
216. Kozlov, M. V., Kleymenova, A. A., Konduktorov, K. A., Malikova, A. Z., & Kochetkov, S. N. (2014). Selective inhibitor of histone deacetylase 6 (tubastatin A) suppresses proliferation of hepatitis C virus replicon in culture of human hepatocytes. *Biochemistry (Moscow)*, *79*, 637-642.
217. Wang, Z., Leng, Y., Wang, J., Liao, H. M., Bergman, J., Leeds, P., ... & Chuang, D. M. (2016). Tubastatin A, an HDAC6 inhibitor, alleviates stroke-induced brain infarction and functional deficits: potential roles of α-tubulin acetylation and FGF-21 up-regulation. *Scientific reports*, *6*(1), 19626.
218. Kalin, J. H., Butler, K. V., Akimova, T., Hancock, W. W., & Kozikowski, A. P. (2012). Second-generation histone deacetylase 6 inhibitors enhance the immunosuppressive effects of Foxp3+ T-regulatory cells. *Journal of medicinal chemistry*, *55*(2), 639-651.
219. Leonhardt, M., Sellmer, A., Krämer, O. H., Dove, S., Elz, S., Kraus, B., ... & Mahboobi, S. (2018). Design and biological evaluation of tetrahydro-β-carboline derivatives as highly potent histone deacetylase 6 (HDAC6) inhibitors. *European journal of medicinal chemistry*, *152*, 329-357.
220. Lee, J. H., Mahendran, A., Yao, Y., Ngo, L., Venta-Perez, G., Choy, M. L., ... & Marks, P. A. (2013). Development of a histone deacetylase 6 inhibitor and its biological effects. *Proceedings of the National Academy of Sciences*, *110*(39), 15704-15709.
221. Smil, D. V., Manku, S., Chantigny, Y. A., Leit, S., Wahhab, A., Yan, T. P., ... & Déziel, R. (2009). Novel HDAC6 isoform selective chiral small molecule histone deacetylase inhibitors. *Bioorganic & medicinal chemistry letters*, *19*(3), 688-692.
222. Tang, G., Wong, J. C., Zhang, W., Wang, Z., Zhang, N., Peng, Z., ... & Chen, L. (2014). Identification of a novel aminotetralin class of HDAC6 and HDAC8 selective inhibitors. *Journal of medicinal chemistry*, *57*(19), 8026-8034.
223. Lin, X., Chen, W., Qiu, Z., Guo, L., Zhu, W., Li, W., ... & Tang, G. (2015). Design and synthesis of orally bioavailable aminopyrrolidinone histone deacetylase 6 inhibitors. *Journal of medicinal chemistry*, *58*(6), 2809-2820.

224. Lee, H. Y., Nepali, K., Huang, F. I., Chang, C. Y., Lai, M. J., Li, Y. H., ... & Liou, J. P. (2018). (N-Hydroxycarbonylbenylamino) quinolines as selective histone deacetylase 6 inhibitors suppress growth of multiple myeloma in vitro and in vivo. *Journal of medicinal chemistry*, *61*(3), 905-917.
225. Dallavalle, S., Cincinelli, R., Nannei, R., Merlini, L., Morini, G., Penco, S., ... & Zunino, F. (2009). Design, synthesis, and evaluation of biphenyl-4-yl-acrylohydroxamic acid derivatives as histone deacetylase (HDAC) inhibitors. *European journal of medicinal chemistry*, *44*(5), 1900-1912.
226. Zuco, V., De Cesare, M., Cincinelli, R., Nannei, R., Pisano, C., Zaffaroni, N., & Zunino, F. (2011). Synergistic antitumor effects of novel HDAC inhibitors and paclitaxel in vitro and in vivo. *PLoS One*, *6*(12), e29085.
227. Minjie, S., Defei, H., Zhimin, H., Weiding, W., & Yuhua, Z. (2015). Targeting pancreatic cancer cells by a novel hydroxamate-based histone deacetylase (HDAC) inhibitor ST-3595. *Tumor Biology*, *36*, 9015-9022.
228. Zuco, V., Cassinelli, G., Cossa, G., Gatti, L., Favini, E., Tortoreto, M., ... & Perego, P. (2015). Targeting the invasive phenotype of cisplatin-resistant non-small cell lung cancer cells by a novel histone deacetylase inhibitor. *Biochemical pharmacology*, *94*(2), 79-90.
229. Kaliszczak, M., Trousil, S., Åberg, O., Perumal, M., Nguyen, Q. D., & Aboagye, E. O. (2013). A novel small molecule hydroxamate preferentially inhibits HDAC6 activity and tumour growth. *British journal of cancer*, *108*(2), 342-350.
230. Yu, C. W., Chang, P. T., Hsin, L. W., & Chern, J. W. (2013). Quinazolin-4-one derivatives as selective histone deacetylase-6 inhibitors for the treatment of Alzheimer's disease. *Journal of medicinal chemistry*, *56*(17), 6775-6791.
231. Kuo, C. C., Hsieh, H. P., Pan, W. Y., Chen, C. P., Liou, J. P., Lee, S. J., ... & Chang, J. Y. (2004). BPR0L075, a novel synthetic indole compound with antimitotic activity in human cancer cells, exerts effective antitumoral activity in vivo. *Cancer research*, *64*(13), 4621-4628.
232. Lee, H. Y., Lee, J. F., Kumar, S., Wu, Y. W., HuangFu, W. C., Lai, M. J., ... & Liou, J. P. (2017). 3-Aroylindoles display antitumor activity in vitro and in vivo: Effects of N1-substituents on biological activity. *European journal of medicinal chemistry*, *125*, 1268-1278.
233. Wu, Y. W., Hsu, K. C., Lee, H. Y., Huang, T. C., Lin, T. E., Chen, Y. L., ... & HuangFu, W. C. (2018). A novel dual HDAC6 and tubulin inhibitor, MPT0B451,

displays anti-tumor ability in human cancer cells in vitro and in vivo. *Frontiers in pharmacology*, *9*, 205.

234. Kozikowski, A. P., Chen, Y., Gaysin, A., Chen, B., D'Annibale, M. A., Suto, C. M., & Langley, B. C. (2007). Functional differences in epigenetic modulators superiority of mercaptoacetamide-based histone deacetylase inhibitors relative to hydroxamates in cortical neuron neuroprotection studies. *Journal of medicinal chemistry*, *50*(13), 3054-3061.
235. Sung, Y. M., Lee, T., Yoon, H., DiBattista, A. M., Song, J. M., Sohn, Y., ... & Hoe, H. S. (2013). Mercaptoacetamide-based class II HDAC inhibitor lowers Aβ levels and improves learning and memory in a mouse model of Alzheimer's disease. *Experimental neurology*, *239*, 192-201.
236. Segretti, M. C., Vallerini, G. P., Brochier, C., Langley, B., Wang, L., Hancock, W. W., & Kozikowski, A. P. (2015). Thiol-based potent and selective HDAC6 inhibitors promote tubulin acetylation and T-regulatory cell suppressive function. *ACS medicinal chemistry letters*, *6*(11), 1156-1161.
237. Lv, W., Zhang, G., Barinka, C., Eubanks, J. H., & Kozikowski, A. P. (2017). Design and synthesis of mercaptoacetamides as potent, selective, and brain permeable histone deacetylase 6 inhibitors. *ACS Medicinal Chemistry Letters*, *8*(5), 510-515.
238. Jose, B., Oniki, Y., Kato, T., Nishino, N., Sumida, Y., & Yoshida, M. (2004). Novel histone deacetylase inhibitors: cyclic tetrapeptide with trifluoromethyl and pentafluoroethyl ketones. *Bioorganic & medicinal chemistry letters*, *14*(21), 5343-5346.
239. Gong, C. J., Gao, A. H., Zhang, Y. M., Su, M. B., Chen, F., Sheng, L., ... & Nan, F. J. (2016). Design, synthesis and biological evaluation of bisthiazole-based trifluoromethyl ketone derivatives as potent HDAC inhibitors with improved cellular efficacy. *European Journal of Medicinal Chemistry*, *112*, 81-90.
240. Dehmel, F., Weinbrenner, S., Julius, H., Ciossek, T., Maier, T., Stengel, T., ... & Beckers, T. (2008). Trithiocarbonates as a novel class of HDAC inhibitors: SAR studies, isoenzyme selectivity, and pharmacological profiles. *Journal of medicinal chemistry*, *51*(13), 3985-4001.
241. Patil, V., Sodji, Q. H., Kornacki, J. R., Mrksich, M., & Oyelere, A. K. (2013). 3-Hydroxypyridin-2-thione as novel zinc binding group for selective histone deacetylase inhibition. *Journal of medicinal chemistry*, *56*(9), 3492-3506.

242. Muthyala, R., Shin, W. S., Xie, J., & Sham, Y. Y. (2015). Discovery of 1-hydroxypyridine-2-thiones as selective histone deacetylase inhibitors and their potential application for treating leukemia. *Bioorganic & medicinal chemistry letters*, *25*(19), 4320-4324.
243. Bug, G., Burchert, A., Wagner, E. M., Kröger, N., Berg, T., Güller, S., ... & Ottmann, O. G. (2017). Phase I/II study of the deacetylase inhibitor panobinostat after allogeneic stem cell transplantation in patients with high-risk MDS or AML (PANOBEST trial). *Leukemia*, *31*(11), 2523-2525.
244. DeAngelo, D. J., Spencer, A., Bhalla, K. N., Prince, H. M., Fischer, T., Kindler, T., ... & Ottmann, O. G. (2013). Phase Ia/II, two-arm, open-label, dose-escalation study of oral panobinostat administered via two dosing schedules in patients with advanced hematologic malignancies. *Leukemia*, *27*(8), 1628-1636.
245. Dimopoulos, M., Siegel, D. S., Lonial, S., Qi, J., Hajek, R., Facon, T., ... & Anderson, K. C. (2013). Vorinostat or placebo in combination with bortezomib in patients with multiple myeloma (VANTAGE 088): a multicentre, randomised, double-blind study. *The lancet oncology*, *14*(11), 1129-1140.
246. Krukowski, K., Ma, J., Golonzhka, O., Laumet, G. O., Gutti, T., Van Duzer, J. H., ... & Kavelaars, A. (2017). HDAC6 inhibition effectively reverses chemotherapy-induced peripheral neuropathy. *Pain*, *158*(6), 1126-1137.
247. Benoy, V., Vanden Berghe, P., Jarpe, M., Van Damme, P., Robberecht, W., & Van Den Bosch, L. (2017). Development of improved HDAC6 inhibitors as pharmacological therapy for axonal Charcot–Marie–Tooth disease. *Neurotherapeutics*, *14*, 417-428.
248. Ma, J., Trinh, R. T., Mahant, I. D., Peng, B., Matthias, P., Heijnen, C. J., & Kavelaars, A. (2019). Cell-specific role of histone deacetylase 6 in chemotherapy-induced mechanical allodynia and loss of intraepidermal nerve fibers. *Pain*, *160*(12), 2877-2890.
249. Tsimberidou, A. M., Beer, P. A., Cartwright, C. A., Haymaker, C., Vo, H. H., Kiany, S., ... & Wistuba, I. I. (2021). Preclinical development and first-in-human study of KA2507, a selective and potent inhibitor of histone deacetylase 6, for patients with refractory solid tumors. *Clinical Cancer Research*, *27*(13), 3584-3594.
250. Choi, E. W., Song, J. W., Ha, N., Choi, Y. I., & Kim, S. (2018). CKD-506, a novel HDAC6-selective inhibitor, improves renal outcomes and survival in a mouse model of systemic lupus erythematosus. *Scientific reports*, *8*(1), 17297.

251. Clinical trial status of CDK-506: https://clinicaltrials.gov/ct2/show/NCT04204603
252. Noce, B., Di Bello, E., Fioravanti, R., & Mai, A. (2023). LSD1 inhibitors for cancer treatment: Focus on multi-target agents and compounds in clinical trials. *Frontiers in pharmacology*, *14*, 1120911.
253. Zhao, L., Fu, L., Li, G., Yu, Y., Wang, J., Liang, H., ... & Wang, Y. (2023). Three-dimensional quantitative structural-activity relationship and molecular dynamics study of multivariate substituted 4-oxyquinazoline HDAC6 inhibitors. *Molecular Diversity*, *27*(3), 1123-1140.
254. Li, Z., Zhao, C., He, G., Wang, Y., Wang, Y., & Ma, X. (2022). Identification of PI3K/HDAC Dual-targeted inhibitors with subtype selectivity as potential therapeutic agents against solid Tumors: Building HDAC6 potency in a Quinazolinone-based PI3Kδ-selective template. *Bioorganic & medicinal chemistry*, *73*, 117028.
255. Yao, D., Li, C., Jiang, J., Huang, J., Wang, J., He, Z., & Zhang, J. (2020). Design, synthesis and biological evaluation of novel HDAC inhibitors with improved pharmacokinetic profile in breast cancer. *European Journal of Medicinal Chemistry*, *205*, 112648.
256. Ding, C., Chen, S., Zhang, C., Hu, G., Zhang, W., Li, L., ... & Jiang, Y. (2017). Synthesis and investigation of novel 6-(1, 2, 3-triazol-4-yl)-4-aminoquinazolin derivatives possessing hydroxamic acid moiety for cancer therapy. *Bioorganic & medicinal chemistry*, *25*(1), 27-37.
257. Yang, Z., Wang, T., Wang, F., Niu, T., Liu, Z., Chen, X., ... & Chen, L. (2016). Discovery of selective histone deacetylase 6 inhibitors using the quinazoline as the cap for the treatment of cancer. *Journal of medicinal chemistry*, *59*(4), 1455-1470.
258. Yu, C. W., Chang, P. T., Hsin, L. W., & Chern, J. W. (2013). Quinazolin-4-one derivatives as selective histone deacetylase-6 inhibitors for the treatment of Alzheimer's disease. *Journal of medicinal chemistry*, *56*(17), 6775-6791.
259. Discovery Studio 3.0., Accelrys Inc., San Diego, USA, 2011.
260. Mauri, A., Consonni, V., & Todeschini, R. (2017). Molecular descriptors. In *Handbook of computational chemistry* (pp. 2065-2093). Springer International Publishing.
261. Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, *32*(7), 1466-1474.

262. Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, *32*(7), 1466-1474.
263. The simple, user-friendly and reliable online standalone tools, softwares. http://dtclab.webs.com/software-tools
264. https://www.mims.com/india
265. Shahlaei, M. (2013). Descriptor selection methods in quantitative structure–activity relationship studies: a review study. *Chemical reviews*, *113*(10), 8093-8103.
266. https://glmnet.stanford.edu/articles/glmnet.html
267. Engebretsen, S., & Bohlin, J. (2019). Statistical predictions with glmnet. *Clinical epigenetics*, *11*(1), 123.
268. Sun, L., Yang, H., Li, J., Wang, T., Li, W., Liu, G., & Tang, Y. (2018). In silico prediction of compounds binding to human plasma proteins by QSAR models. *ChemMedChem*, *13*(6), 572-581.
269. Banerjee, S., Baidya, S. K., Ghosh, B., Nandi, S., Mandal, M., Jha, T., & Adhikari, N. (2023). Quantitative structural assessments of potential meprin β inhibitors by non-linear QSAR approaches and validation by binding mode of interaction analysis. *New Journal of Chemistry*, *47*(15), 7051-7069.
270. Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning*, *6*, 37-66.
271. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.
272. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
273. Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences*, *43*(6), 1947-1958.
274. Nantasenamat, C., Naenna, T., Ayudhya, C. I. N., & Prachayasittikul, V. (2005). Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network. *Journal of computer-aided molecular design*, *19*, 509-524.
275. Nantasenamat, C., Worachartcheewan, A., Jamsak, S., Preeyanon, L., Shoombuatong, W., Simeon, S., ... & Prachayasittikul, V. (2015). AutoWeka:

toward an automated data mining software for QSAR and QSPR studies. *Artificial neural networks*, 119-147.

276. Wesolowski, M., & Suchacz, B. (2012). Artificial neural networks: theoretical background and pharmaceutical applications: a review. *journal of aoac international*, *95*(3), 652-668.
277. Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd..
278. Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, *10*(5), 988-999.
279. Amin, S. A., Adhikari, N., Gayen, S., & Jha, T. (2017). First report on the structural exploration and prediction of new BPTES analogs as glutaminase inhibitors. *Journal of Molecular Structure*, *1143*, 49-64.
280. Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, *36*(5), 355-362.
281. Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, *340*(6137), 1177-1178.
282. Banerjee, S., Baidya, S. K., Ghosh, B., Jha, T., & Adhikari, N. (2023). Exploration of structural alerts and fingerprints for novel anticancer therapeutics: a robust classification-QSAR dependent structural analysis of drug-like MMP-9 inhibitors. *SAR and QSAR in Environmental Research*, *34*(4), 299-319.
283. Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, *58*(302), 415-434.
284. Ferrari, T., Cattaneo, D., Gini, G., Golbamaki Bakhtyari, N., Manganaro, A., & Benfenati, E. (2013). Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. *SAR and QSAR in Environmental Research*, *24*(5), 365-383.
285. Gini, G., Ferrari, T., Lombardo, A., Cassano, A., & Benfenati, E. (2019). A new QSAR model for acute fish toxicity based on mined structural alerts. *J Toxicol Risk Assess*, *5*(1), 016.
286. Oecd, O. (2004). The OECD principles of corporate governance. *Contaduría y Administración*, (216).
287. RCSB Protein Data Bank, https://www.rcsb.org/. As accessed in August 2022
288. Schrodinger Suite. Schrodinger LLC, New York, USA, 2019.

289. Banerjee, S., Baidya, S. K., Ghosh, B., Adhikari, N., & Jha, T. (2022). The first report on predictive comparative ligand-based multi-QSAR modeling analysis of 4-pyrimidinone and 2-pyridinone based APJ inhibitors. *New Journal of Chemistry*, *46*(24), 11591-11607.
290. https://gohom.win/ManualHom/Schrodinger/Schrodinger_2015-2_docs/maestro/help_Maestro/prime/prime_mmgbsa_panel.html
291. Ylilauri, M., & Pentikäinen, O. T. (2013). MMGBSA as a tool to understand the binding affinities of filamin–peptide interactions. *Journal of chemical information and modeling*, *53*(10), 2626-2633.
292. Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., & De Gioia, L. (2009). Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *Journal of molecular graphics and modelling*, *27*(8), 889-899.

Appendix

Appendix Table S1. Dataset compounds with their smiles notation, HDAC6 inhibitory activity, and binary categories

Cpd No	*Smiles*	*IC_{50}*	*pIC_{50}*	*Binary*	*Class*	*Set*
1	c1(ccc2c(c1)c(=O)n(cn2)CC)/C=C/C(=O)NO	1354	5.868	0	Inactive	Training
2	c1(ccc2c(c1)c(=O)n(cn2)CCCl)/C=C/C(=O)NO	2385	5.623	0	Inactive	Training
3	c1(ccc2c(c1)c(=O)n(cn2)CCC)/C=C/C(=O)NO	925	6.034	0	Inactive	Test
4	c1(ccc2c(c1)c(=O)n(cn2)CCCC)/C=C/C(=O)NO	974	6.011	0	Inactive	Training
5	c1(ccc2c(c1)c(=O)n(cn2)CC1CCCC1)/C=C/C(=O)NO	504	6.298	0	Inactive	Training
6	c1(ccc2c(c1)c(=O)n(cn2)Cc1ccccc1)/C=C/C(=O)NO	285	6.545	0	Inactive	Test
7	c1(ccc2c(c1)c(=O)n(cn2)Cc1ccccc1F)/C=C/C(=O)NO	424	6.373	0	Inactive	Test
8	c1(ccc2c(c1)c(=O)n(cn2)Cc1cccc(c1)F)/C=C/C(=O)NO	588	6.231	0	Inactive	Training
9	c1(ccc2c(c1)c(=O)n(cn2)Cc1ccc(cc1)F)/C=C/C(=O)NO	253	6.597	0	Inactive	Test
10	c1(ccc2c(c1)c(=O)n(cn2)Cc1ccc(cc1)Cl)/C=C/C(=O)NO	329	6.483	0	Inactive	Training
11	c1(ccc2c(c1)c(=O)n(cn2)Cc1ccc(cc1)C)/C=C/C(=O)NO	360	6.444	0	Inactive	Training
12	c1c(cc2c(c1)c(=O)n(cn2)CC)/C=C/C(=O)NO	68	7.167	1	Active	Training
13	c1c(cc2c(c1)c(=O)n(cn2)CCCl)/C=C/C(=O)NO	113	6.947	0	Inactive	Training
14	c1c(cc2c(c1)c(=O)n(cn2)CCC)/C=C/C(=O)NO	67	7.174	1	Active	Test
15	c1c(cc2c(c1)c(=O)n(cn2)CCCC)/C=C/C(=O)NO	124	6.907	0	Inactive	Training
16	c1c(cc2c(c1)c(=O)n(cn2)CC1CCCC1)/C=C/C(=O)NO	72	7.143	1	Active	Test
17	c1c(cc2c(c1)c(=O)n(cn2)Cc1ccccc1)/C=C/C(=O)NO	247	6.607	0	Inactive	Training
18	c1c(cc2c(c1)c(=O)n(cn2)Cc1ccccc1F)/C=C/C(=O)NO	98	7.009	1	Active	Test
19	c1c(cc2c(c1)c(=O)n(cn2)Cc1cccc(c1)F)/C=C/C(=O)NO	96	7.018	1	Active	Training
20	c1c(cc2c(c1)c(=O)n(cn2)Cc1ccc(cc1)F)/C=C/C(=O)NO	48	7.319	1	Active	Training
21	c1c(cc2c(c1)c(=O)n(cn2)Cc1ccc(cc1)Cl)/C=C/C(=O)NO	41	7.387	1	Active	Test
22	c1c(cc2c(c1)c(=O)n(cn2)Cc1ccc(cc1)C)/C=C/C(=O)NO	44	7.357	1	Active	Test

23	c1ccc2c(c1Cc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1ncnc2c1nc[nH]2)CC)c1ccccc1	889	6.051	0	Inactive	Training
24	c1ccc2c(c1NCc1ncc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1ncnc2c1nc[nH]2)CC)c1ccccc1	9	8.046	1	Active	Training
25	c1ccc2c(c1CNC(=O)c1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1ncnc2c1nc[nH]2)CC)c1ccccc1	10	8.000	1	Active	Training
26	c1ccc2c(c1CNC(=O)c1ccc(cn1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1ncnc2c1nc[nH]2)CC)c1ccccc1	6	8.222	1	Active	Training
27	c1ccc2c(c1CCCCCC(=O)NO)c(=O)n(c(n2)[C@@H](Nc1nc(nc(c1C#N)N)N)CC)c1ccccc1	69	7.161	1	Active	Training
28	c1ccc2c(c1CCCCCC(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C)[N+]#[C-])CC)c1ccccc1	28	7.553	1	Active	Training
29	c1ccc2c(c1CCCCCC(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(cnc(n1)N)[N+]#[C-])CC)c1ccccc1	45	7.347	1	Active	Training
30	c1ccc2c(c1NCCCCC(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C)[N+]#[C-])CC)c1ccccc1	1238	5.907	0	Inactive	Training
31	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)N)[N+]#[C-])CC)c1ccccc1	4	8.398	1	Active	Test
32	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C)[N+]#[C-])CC)c1ccccc1	12	7.921	1	Active	Test
33	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1cc(nc(n1)N)C)CC)c1ccccc1	4	8.398	1	Active	Training
34	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(ncn1)N)[N+]#[C-])CC)c1ccccc1	4	8.398	1	Active	Test
35	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(ncn1)N)Cl)CC)c1ccccc1	4	8.398	1	Active	Test
36	c1ccc2c(c1NCc1ncc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(ncn1)N)Cl)CC)c1ccccc1	3	8.523	1	Active	Test
37	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C)Cl)CC)c1ccccc1	22	7.658	1	Active	Training

38	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)C)N)[N+]#[C-])CC)c1ccccc1	14	7.854	1	Active	Training
39	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)N)Cl)CC)c1ccccc1	9	8.046	1	Active	Training
40	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C(F)(F)F)[N+]#[C-])CC)c1ccccc1	52	7.284	1	Active	Training
41	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C1CC1)[N+]#[C-])CC)c1ccccc1	47	7.328	1	Active	Training
42	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C(F)F)[N+]#[C-])CC)c1ccccc1	27	7.569	1	Active	Training
43	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)N)[N+]#[C-])C)c1ccccc1	6	8.222	1	Active	Training
44	c1ccc2c(c1NCc1ccc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)C)[N+]#[C-])C)c1ccccc1	8	8.097	1	Active	Training
45	c1ccc2c(c1CNC(=O)c1ncc(cc1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)N)[N+]#[C-])C)c1ccccc1	16	7.796	1	Active	Training
46	c1ccc2c(c1CNc1ncc(cn1)C(=O)NO)c(=O)n(c(n2)[C@@H](Nc1c(c(nc(n1)N)N)[N+]#[C-])C)c1ccccc1	3	8.523	1	Active	Training
47	c1ccc2c(c1F)c(=O)n(c(n2)[C@H](CC)Nc1c2c(ncn1)[nH]cn2)CCCCCC(=O)NO	121	6.917	0	Inactive	Training
48	c1ccc2c(c1F)c(=O)n(c(n2)[C@H](CC)Nc1c2c(ncn1)[nH]cn2)Cc1ccc(cc1)C(=O)NO	45	7.347	1	Active	Training
49	c1ccc2c(c1F)c(=O)n(c(n2)[C@H]1N(c2ncnc3[nH]cnc23)CC2(C1)CC2)Cc1ccc(cc1)C(=O)NO	28	7.553	1	Active	Training
50	c1ccc2c(c1F)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)Cc1ccc(cc1)C(=O)NO	27	7.569	1	Active	Training
51	c1ccc2c(c1F)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)Cc1ccc(s1)C(=O)NO	32	7.495	1	Active	Training
52	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CCC1)Cc1ccc(cc1)C(=O)NO	280	6.553	0	Inactive	Test

53	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)Cc1ccc(cc1)C(=O)NO	331	6.480	0	Inactive	Test
54	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@@H](Nc1nc(nc(N)c1C#N)N)C)Cc1ccc(s1)C(=O)NO	66	7.180	1	Active	Training
55	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@@H](Nc1nc(nc(N)c1C#N)N)CC)Cc1ccc(s1)C(=O)NO	45	7.347	1	Active	Training
56	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CCC1)Cc1ccc(s1)C(=O)NO	67	7.174	1	Active	Training
57	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)Cc1ccc(s1)C(=O)NO	13	7.886	1	Active	Training
58	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)Cc1onc(c1)C(=O)NO	604	6.219	0	Inactive	Training
59	c1ccc2c(c1Cl)c(nc(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CCC1)NCc1ccc(cc1)C(=O)NO	267	6.573	0	Inactive	Training
60	c1ccc2c(c1Cl)c(nc(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)NCc1ccc(cc1)C(=O)NO	398	6.400	0	Inactive	Training
61	c1ccc2c(c1Cl)c(nc(n2)[C@H]1N(c2nc(nc(N)c2C#N)N)CC2(C1)CC2)NCCc1ccc(cc1)C(=O)NO	219	6.660	0	Inactive	Training
62	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OC(=O)NCc1ccc(cc1)C(=O)NO)c1nc(nc(c1C#N)N)N)c1ccccc1	87	7.060	1	Active	Training
63	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@H](C1)OC(=O)NCc1ccc(cc1)C(=O)NO)c1nc(nc(c1C#N)N)N)c1ccccc1	168	6.775	0	Inactive	Training
64	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OCc1ccc(cc1)C(=O)NO)c1nc(nc(c1C#N)N)N)C	13	7.886	1	Active	Training
65	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OCc1ccc(s1)C(=O)NO)c1nc(nc(c1C#N)N)N)C	13	7.886	1	Active	Test
66	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OC(=O)NCc1ccc(cc1)C(=O)NO)c1nc(nc(c1C#N)N)N)C	27	7.569	1	Active	Training

67	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OC(=O)N(Cc1ccc(cc1)C(=O)NO)C)c1nc(nc(c1C#N)N)N)C	11	7.959	1	Active	Training
68	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OC(=O)NCCc1ccc(cc1)C(=O)NO)c1nc(nc(c1C#N)N)N)C	36	7.444	1	Active	Test
69	c1ccc2c(c1Cl)c(=O)n(c(n2)[C@H]1N(C[C@@H](C1)OC(=O)NCc1ccc(s1)C(=O)NO)c1nc(nc(c1C#N)N)N)C	43	7.367	1	Active	Test
70	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1CCCC1)OC	390	6.409	0	Inactive	Test
71	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1CCSCC1)OC	420	6.377	0	Inactive	Test
72	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1CCOCC1)OC	350	6.456	0	Inactive	Training
73	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1[C@H](COCC1)C)OC	360	6.444	0	Inactive	Test
74	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1[C@@H](COCC1)C)OC	330	6.481	0	Inactive	Training
75	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1CCN(CC1)C)OC	650	6.187	0	Inactive	Training
76	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1CCS(=O)(=O)CC1)OC	850	6.071	0	Inactive	Training
77	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCC(=O)NO)C)N1CCC(CC1)(F)F)OC	730	6.137	0	Inactive	Test
78	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCCC1)OC	310	6.509	0	Inactive	Training
79	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCCCC1)OC	480	6.319	0	Inactive	Training
80	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCSCC1)OC	290	6.538	0	Inactive	Test
81	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCOCC1)OC	210	6.678	0	Inactive	Training

82	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1[C@H](COCC1)C)OC	190	6.721	0	Inactive	Training
83	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1[C@@H](COCC1)C)OC	220	6.658	0	Inactive	Test
84	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCN(CC1)C)OC	110	6.959	0	Inactive	Training
85	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCS(=O)(=O)CC1)OC	130	6.886	0	Inactive	Training
86	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCC(=O)NO)C)N1CCC(CC1)(F)F)OC	340	6.469	0	Inactive	Test
87	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCCC1)OC	101.4	6.994	0	Inactive	Training
88	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCCCC1)OC	203.5	6.691	0	Inactive	Training
89	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCSCC1)OC	129.2	6.889	0	Inactive	Test
90	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCOCC1)OC	119.2	6.924	0	Inactive	Training
91	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1[C@H](COCC1)C)OC	98.5	7.007	1	Active	Training
92	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1[C@@H](COCC1)C)OC	101.7	6.993	0	Inactive	Training
93	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCN(CC1)C)OC	11.68	7.933	1	Active	Test
94	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCS(=O)(=O)CC1)OC	55.3	7.257	1	Active	Test
95	c1c(cc2c(c1OC)c(nc(n2)c1cc(c(c(c1)C)OCCCCCC(=O)NO)C)N1CCC(CC1)(F)F)OC	186.1	6.730	0	Inactive	Training
96	c12c(ncnc1Nc1cc(c(cc1)F)Cl)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	9.5	8.022	1	Active	Training

97	c12c(ncnc1Nc1cc(c(cc1)F)Cl)ccc(c2)c1nnn(c1)CCCCCCC(=O)NO	8.4	8.076	1	Active	Training
98	c12c(ncnc1Nc1c(c(ccc1)Cl)F)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	4.1	8.387	1	Active	Training
99	c12c(ncnc1Nc1c(c(ccc1)Cl)F)ccc(c2)c1nnn(c1)CCCCCCC(=O)NO	3.2	8.495	1	Active	Training
100	c12c(ncnc1Nc1ccc(c(c1)Cl)OCc1cc(ccc1)F)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	94.2	7.026	1	Active	Training
101	c12c(ncnc1Nc1ccc(c(c1)Cl)OCc1cc(ccc1)F)ccc(c2)c1nnn(c1)CCCCCCC(=O)NO	19.5	7.710	1	Active	Training
102	c12c(ncnc1Nc1ccc(c(c1)Cl)OCc1nccs1)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	14.3	7.845	1	Active	Training
103	c12c(ncnc1Nc1ccc(c(c1)Cl)OCc1nccs1)ccc(c2)c1nnn(c1)CCCCCCC(=O)NO	13.6	7.866	1	Active	Training
104	c12c(ncnc1Nc1ccc(c(c1)C)Oc1ccc(nc1)C)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	12.9	7.889	1	Active	Test
105	c12c(ncnc1Nc1ccc(c(c1)C)Oc1ccc(nc1)C)ccc(c2)c1nnn(c1)CCCCCCC(=O)NO	13.8	7.860	1	Active	Training
106	c12c(ncnc1Nc1ccc(c(c1)OC)Oc1ccccc1)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	18.5	7.733	1	Active	Training
107	c12c(ncnc1Nc1ccc(c(c1)OC)Oc1ccccc1)ccc(c2)c1nnn(c1)CCCCCC(=O)NO	16	7.796	1	Active	Test
108	c1ccc2c(c1)c(nc(n2)C)N(C)c1ccc(cc1)OCC(=O)NO	8.6	8.066	1	Active	Test
109	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCCCC(=O)NO)cccc2	196	6.708	0	Inactive	Training
110	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCCCCC(=O)NO)cccc2	57	7.244	1	Active	Training
111	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCCCCCCC(=O)NO)cccc2	34	7.469	1	Active	Training
112	c1ccc2c(c1)c(ncn2)N(C)c1ccc(cc1)OCC(=O)NO	14	7.854	1	Active	Training
113	c12c(c(ncn1)N(C)c1ccc(cc1)OCCCC(=O)NO)cccc2	34	7.469	1	Active	Training
114	c12c(c(ncn1)N(C)c1ccc(cc1)OCCCCC(=O)NO)cccc2	49	7.310	1	Active	Training

115	c12c(c(ncn1)N(C)c1ccc(cc1)OCCCCCC(=O)NO)cccc2	19	7.721	1	Active	Training
116	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCC(=O)NO)CCC2	17	7.770	1	Active	Training
117	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCCCC(=O)NO)CCC2	40	7.398	1	Active	Training
118	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCCCCC(=O)NO)CCC2	63	7.201	1	Active	Test
119	c12c(c(nc(n1)C)N(C)c1ccc(cc1)OCCCCCC(=O)NO)CCC2	20	7.699	1	Active	Training
120	c12c(c(nc(n1)C)N(C)c1cc(c(cc1)OC)OCCCC(=O)NO)cccc2	17	7.770	1	Active	Test
121	c12c(c(nc(n1)C)N(C)c1cc(c(cc1)OC)OCCCCC(=O)NO)cccc2	111	6.955	0	Inactive	Test
122	c12c(c(nc(n1)C)N(C)c1cc(c(cc1)OC)OCCCCCC(=O)NO)cccc2	23	7.638	1	Active	Training
123	c12c(c(=O)n(c(n1)C)c1ccccc1)c(ccc2)/C=C/C(=O)NO	1920	5.717	0	Inactive	Training
124	c12c(c(=O)n(c(n1)C)c1ccccc1)cc(cc2)/C=C/C(=O)NO	32	7.495	1	Active	Training
125	c12c(c(=O)n(c(n1)C)c1ccccc1)ccc(c2)/C=C/C(=O)NO	88	7.056	1	Active	Training
126	c12c(c(=O)n(c(n1)C)c1ccccc1)cccc2/C=C\C(=O)NO	690	6.161	0	Inactive	Training
127	c12c(c(=O)n(c(n1)C)Cc1ccccc1)ccc(c2)/C=C/C(=O)NO	24	7.620	1	Active	Training
128	c12c(c(=O)n(c(n1)C)CCc1ccccc1)ccc(c2)/C=C/C(=O)NO	29	7.538	1	Active	Training
129	c12c(c(=O)n(c(n1)C)CCC1c3c(N=C1)cccc3)ccc(c2)/C=C/C(=O)NO	15	7.824	1	Active	Training
130	c12c(c(=O)n(cn1)CCc1ccccc1)ccc(c2)/C=C/C(=O)NO	35	7.456	1	Active	Training
131	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)ccc(c2)/C=C/C(=O)NO	11	7.959	1	Active	Training
132	c12c(c(=O)n(c(n1)CC)CCCc1ccccc1)ccc(c2)/C=C/C(=O)NO	33	7.481	1	Active	Test
133	c12c(c(=O)n(c(n1)CCC)CCc1ccccc1)ccc(c2)/C=C/C(=O)NO	41	7.387	1	Active	Training
134	c12c(c(=O)n(c(n1)C(C)C)CCc1ccccc1)ccc(c2)/C=C/C(=O)NO	13	7.886	1	Active	Training

135	c12c(c(=O)n(c(n1)CC)CCc1ccc(cc1)OC)ccc(c2)/C=C/C(=O)NO	41	7.387	1	Active	Test
136	c12c(c(=O)n(c(n1)CC)CCc1ccc(cc1)F)ccc(c2)/C=C/C(=O)NO	43	7.367	1	Active	Training
137	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)cc(c(c2)/C=C/C(=O)NO)F	14	7.854	1	Active	Training
138	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)cc(c(c2)F)/C=C/C(=O)NO	8	8.097	1	Active	Training
139	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)cc(c(c2)Cl)/C=C/C(=O)NO	747	6.127	0	Inactive	Test
140	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)cc(cc2)Cc1ccc(cc1)C(=O)NO	11	7.959	1	Active	Training
141	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)ccc(c2)Cc1ccc(cc1)C(=O)NO	9	8.046	1	Active	Training
142	c12c(c(=O)n(c(n1)CC)CCc1ccccc1)cccc2Cc1ccc(cc1)C(=O)NO	79	7.102	1	Active	Training

Appendix Table S2. LASSO coefficients of features obtained from LASSO regression

Selected feature	***LASSO coefficient***
KRFP363	-0.72921
AATSC8s	-0.55877
minHCsatu	0.456812
KRFP1947	0.322965
PubchemFP686	-0.3124
PubchemFP372	0.280104
GATS7m	-0.2635
GATS4m	0.135438
KRFP4071	0.131816
KRFP3574	-0.11414
PubchemFP2	0.104557
AD2D414	-0.08778
AD2D413	0.054307
KRFP2025	0.052232
KRFP413	0.052191
KRFP683	0.048442
APC2D6_C_Cl	-0.04093
APC2D10_O_O	-0.04004
ATSC3e	0.037471
GATS7p	-0.03476
ATSC3c	0.033796
AATSC7v	0.019275
nHBint3	0.015602
PubchemFP359	-0.01144
SpMAD_Dzs	0.010697

Appendix Table S3: Final selected features set use for ML model development

Cpd No	*KRFP 363*	*AATSC8 s*	*minHCsat u*	*Pubche mFP68 6*	*PubchemF P372*	*GATS4m*	*KRFP41 3*	*APC2 D6_C_ Cl*	*GATS7p*	*nHBint 3*	*pIC$_{50}$*
Training set											
1	0	0.975846	0	0	0	1.155545	0	0	1.211365	1	5.868
2	0	0.99185	0	0	0	0.842496	0	2	1.117043	1	5.623
4	0	0.755645	0	0	0	1.023775	0	0	1.092702	1	6.011
5	0	0.653826	0	0	0	0.901953	0	0	1.058371	1	6.298
8	0	0.574075	0	0	0	1.062331	0	0	1.04934	1	6.231
10	0	0.623105	0	0	0	1.021685	0	0	0.926777	1	6.483
11	0	0.622224	0	0	0	1.06159	0	0	0.874368	1	6.444
12	0	0.050251	0	0	0	1.146667	0	0	1.039849	1	7.167
13	0	0.025843	0	0	0	0.836867	0	2	0.963803	1	6.947
15	0	0.039727	0	0	0	1.016773	0	0	0.9707	1	6.907
17	0	-0.14251	0	0	0	1.070641	0	0	0.933116	1	6.607
19	0	-0.17533	0	0	0	1.055516	0	0	0.933137	1	7.018
20	0	0.448495	0	0	0	1.116937	0	0	0.935496	1	7.319
23	0	0.014383	0.917529	0	0	0.954958	0	0	1.096149	2	6.051
24	0	0.01387	0.927201	0	1	0.985302	0	0	1.109223	2	8.046
25	0	0.139627	0.931132	0	0	1.025688	0	0	1.112707	2	8
26	0	0.1705	0.935265	0	1	1.00901	0	0	1.129503	2	8.222
27	0	-0.09497	0.621586	0	0	0.908116	0	0	1.127141	2	7.161
28	0	-0.05475	0.619768	0	0	0.94371	0	0	1.152103	2	7.553
29	0	-0.03194	0.619768	0	1	0.942216	0	0	1.128084	2	7.347
30	1	-0.06516	0.639768	0	0	0.947368	0	0	1.132857	2	5.907
33	0	-0.10355	0.885257	0	0	0.954654	0	0	1.138427	2	8.398
37	0	-0.09214	0.884269	0	0	1.093569	0	1	1.127915	3	7.658

38	0	-0.07842	0.93609	0	0	0.971959	0	0	1.153737	2	7.854
39	0	-0.11295	0.898158	0	0	1.016502	0	1	1.12231	4	8.046
40	0	-0.08141	1.027927	0	0	0.849215	0	0	1.137008	2	7.284
41	0	-0.05591	0.93609	0	0	0.968614	0	0	1.082268	2	7.328
42	0	-0.06877	0.997315	0	0	0.887841	0	0	1.140788	2	7.569
43	0	-0.01154	0.949979	0	0	0.971898	0	0	1.088356	2	8.222
44	0	-0.00827	0.93609	0	0	0.99129	0	0	1.11038	2	8.097
45	0	0.18702	0.963042	0	1	1.004232	0	0	1.101754	2	7.796
46	0	0.039105	0.957588	0	1	0.999584	0	0	1.129193	2	8.523
47	0	-0.1403	0.641426	0	0	0.986321	0	0	1.052455	2	6.917
48	0	-0.12967	0.926037	0	0	1.023033	0	0	1.017348	2	7.347
49	0	-0.15948	0.926037	0	0	0.891206	0	0	1.045996	1	7.553
50	0	-0.32757	0.950759	0	0	0.866254	0	0	1.053048	1	7.569
51	0	-0.12996	0.942453	0	0	0.799092	0	0	1.054302	2	7.495
54	0	0.330075	0.911337	0	0	0.933102	0	2	1.017915	3	7.18
55	0	0.111355	0.911337	0	0	0.91141	0	2	0.999939	3	7.347
56	0	0.340541	0.911337	0	0	0.855866	0	2	0.857524	2	7.174
57	0	0.010251	0.911337	0	0	0.798135	0	2	0.983796	2	7.886
58	0	-0.0646	0.953434	1	0	0.816648	0	2	1.073414	1	6.219
59	0	0.048403	0	0	0	0.911722	0	2	0.888335	1	6.573
60	0	-0.17982	0	0	0	0.856494	0	2	0.992759	1	6.4
61	0	-0.16984	0	0	0	0.853008	0	2	1.059219	1	6.66
62	0	-0.01149	1.062858	0	0	0.910302	0	3	0.951027	1	7.06
63	0	-0.01149	1.062858	0	0	0.910302	0	3	0.951027	1	6.775
64	0	0.057103	0.967247	0	0	0.943701	0	1	0.799375	1	7.886
66	0	-0.04367	1.008242	0	0	0.978295	0	1	0.769754	1	7.569
67	0	0.018763	1.008242	0	0	0.980611	0	1	0.719523	1	7.959

72	0	0.047316	0.667989	1	0	1.006226	0	0	1.29176	1	6.456
74	0	0.03631	0.667989	1	0	1.019132	0	0	1.279273	1	6.481
75	0	0.045138	0.665767	1	0	0.985661	0	0	1.268346	1	6.187
76	0	0.2464	0.671664	1	0	0.90409	0	0	1.190748	1	6.071
78	0	-0.00754	0.624329	0	0	0.96318	0	0	1.257313	1	6.509
79	0	-0.03355	0.624329	0	0	0.986711	0	0	1.265385	1	6.319
81	0	-0.0028	0.628235	0	0	0.996058	0	0	1.26353	1	6.678
82	0	-0.01065	0.628235	0	0	1.009945	0	0	1.253595	1	6.721
84	0	-0.00462	0.626282	0	0	0.97888	0	0	1.242254	1	6.959
85	0	0.201026	0.631533	0	0	0.896721	0	0	1.17075	1	6.886
87	0	0.001456	0.600364	0	0	0.956446	0	0	1.247487	1	6.994
88	0	-0.02342	0.600364	0	0	0.980132	0	0	1.255923	1	6.691
90	0	0.005911	0.603824	0	0	0.987454	0	0	1.253688	1	6.924
91	0	-0.00207	0.603824	0	0	1.00207	0	0	1.244826	1	7.007
92	0	-0.00207	0.603824	0	0	1.00207	0	0	1.244826	1	6.993
95	0	0.23684	0.609624	0	0	1.061651	0	0	1.255052	1	6.73
96	0	0.105926	0.609027	0	1	0.930579	0	1	0.848142	1	8.022
97	0	-0.09482	0.591642	0	1	0.911888	0	1	0.847545	1	8.076
98	0	0.033808	0.610257	0	1	0.950846	0	1	0.853124	2	8.387
99	0	-0.16147	0.592677	0	1	0.930948	0	1	0.852134	2	8.495
100	0	0.060734	0.613584	0	1	0.967556	0	3	0.961334	1	7.026
101	0	-0.08723	0.595841	0	1	0.952131	0	3	0.953606	1	7.71
102	0	0.072794	0.610507	0	1	0.862626	0	1	0.8772	1	7.845
103	0	-0.08855	0.593013	0	1	0.849061	0	1	0.871311	1	7.866
105	0	-0.07997	0.594631	0	1	1.060955	0	0	0.909816	1	7.86
106	0	0.045527	0.614623	0	1	1.037828	0	0	1.003923	1	7.733
109	0	0.027389	0.650761	1	0	1.160863	0	0	0.990905	1	6.708

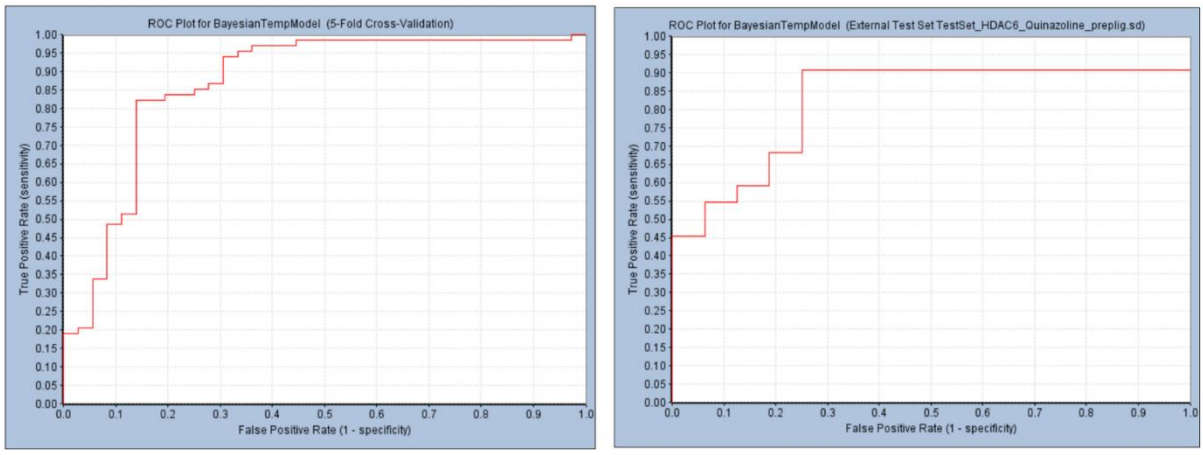
110	0	0.100747	0.613429	0	0	1.132432	0	0	0.963061	1	7.244
111	0	0.067849	0.590954	0	0	1.109396	0	0	0.963093	1	7.469
112	0	0.025671	0.901439	0	0	1.143778	0	0	1.007358	2	7.854
113	0	0.011644	0.650761	1	0	1.183985	0	0	1.012545	1	7.469
114	0	0.113012	0.613429	0	0	1.150321	0	0	0.975791	1	7.31
115	0	0.075919	0.590954	0	0	1.123426	0	0	0.974936	1	7.721
116	0	0.025135	0.893071	0	0	1.133556	0	0	0.847534	2	7.77
117	0	0.024424	0.644791	1	0	1.184072	0	0	0.873721	1	7.398
119	0	0.054876	0.586479	0	0	1.136483	0	0	0.863157	1	7.699
122	0	0.083624	0.607092	0	0	1.083198	0	0	0.980505	1	7.638
123	0	0.311838	0.614105	0	0	0.921993	0	0	1.151695	1	5.717
124	0	0.68001	0.607146	0	0	0.991864	0	0	0.974233	1	7.495
125	0	0.101273	0.614105	0	0	0.984517	0	0	1.107827	1	7.056
126	0	0.510553	0.623803	0	0	0.980294	0	0	1.034816	1	6.161
127	0	-0.03269	0.597488	0	0	1.074918	0	0	1.054966	1	7.62
128	0	-0.05339	0.587536	0	0	1.106777	1	0	1.057567	1	7.538
129	0	0.00346	0.597897	0	0	1.15542	0	0	0.971428	1	7.824
130	0	-0.13122	0	0	0	1.10262	1	0	0.897733	1	7.456
131	0	-0.06206	0.587536	0	0	1.128684	1	0	0.998319	1	7.959
133	0	0.002248	0.587536	0	0	1.049409	1	0	1.008555	1	7.387
134	0	-0.07659	0.587536	0	0	1.068878	1	0	0.956312	1	7.886
136	0	-0.08629	0.602536	0	0	1.145674	0	0	0.990436	1	7.367
137	0	-0.3722	0.610974	0	0	1.128217	1	0	0.998525	1	7.854
138	0	-0.02932	0.611189	0	0	1.124847	1	0	1.008272	1	8.097
140	0	-0.04993	0.578484	0	0	1.107456	1	0	1.003572	1	7.959
141	0	2.99E-04	0.584016	0	0	1.103022	1	0	0.926795	1	8.046
142	0	-0.02903	0.591401	0	0	1.117529	1	0	1.054083	1	7.102

					Test set						
3	0	0.863275	0	0	0	1.064066	0	0	1.280938	1	6.034
6	0	0.593243	0	0	0	1.07774	0	0	1.050959	1	6.545
7	0	0.439351	0	0	0	1.201827	0	0	1.051699	1	6.373
9	0	1.132576	0	0	0	1.123751	0	0	1.051699	1	6.597
14	0	0.03914	0	0	0	1.056253	0	0	1.130139	1	7.174
16	0	0.029094	0	0	0	0.896199	0	0	0.954204	1	7.143
18	0	-0.2708	0	0	0	1.195013	0	0	0.935496	1	7.009
21	0	-0.10928	0	0	0	1.016946	0	0	0.817663	1	7.387
22	0	-0.03437	0	0	0	1.055386	0	0	0.771571	1	7.357
31	0	-0.10198	0.949979	0	0	0.960971	0	0	1.143976	2	8.398
32	0	-0.07842	0.93609	0	0	0.980531	0	0	1.148597	2	7.921
34	0	-0.05936	0.93609	0	0	0.987637	0	0	1.114491	2	8.398
35	0	-0.07202	0.884269	0	0	1.037223	0	1	1.105148	4	8.398
36	0	-0.04999	0.889269	0	1	1.050337	0	1	1.100107	4	8.523
52	0	0.070647	0.919643	0	0	0.892129	0	2	0.855113	1	6.553
53	0	-0.18379	0.919643	0	0	0.840533	0	2	1.007333	1	6.48
65	0	-0.03544	0.960979	0	0	0.874094	0	1	0.735034	2	7.886
68	0	-0.09733	1.002711	0	0	0.967477	0	1	0.814077	1	7.444
69	0	0.263724	1.004203	0	0	0.964423	0	1	0.70477	3	7.367
70	0	0.046627	0.663545	1	0	0.971268	0	0	1.285873	1	6.409
71	0	0.043544	0.662886	1	0	0.85352	0	0	1.161716	1	6.377
73	0	0.03631	0.667989	1	0	1.019132	0	0	1.279273	1	6.444
77	0	0.278811	0.675264	1	0	1.088861	0	0	1.290695	1	6.137
80	0	-0.01028	0.62375	0	0	0.848806	0	0	1.142193	1	6.538
83	0	-0.01065	0.628235	0	0	1.009945	0	0	1.253595	1	6.658
86	0	0.23606	0.63471	0	0	1.074263	0	0	1.264137	1	6.469

89	0	-0.00139	0.599852	0	0	0.845091	0	0	1.13796	1	6.889
93	0	0.004463	0.602094	0	0	0.973113	0	0	1.233587	1	7.933
94	0	0.202813	0.606799	0	0	0.890587	0	0	1.165915	1	7.257
104	0	0.061746	0.612296	0	1	1.077487	0	0	0.912311	1	7.889
107	0	-0.09027	0.596684	0	1	1.021809	0	0	0.995978	1	7.796
108	0	0.045832	0.901439	0	0	1.113224	0	0	0.984191	2	8.066
118	0	0.085678	0.608288	0	0	1.15797	0	0	0.856377	1	7.201
120	0	0.185117	0.677609	1	0	1.12661	0	0	1.004574	1	7.77
121	0	0.092919	0.633934	0	0	1.102767	0	0	0.981087	1	6.955
132	0	0.001403	0.581076	0	0	1.11313	1	0	0.984877	1	7.481
135	0	-0.04021	0.597536	0	0	1.165557	0	0	1.047996	1	7.387
139	0	0.361198	0.580073	0	0	1.106298	1	1	1.073357	1	6.127

Appendix Table S4. Statistical validation parameters for MLR selected features used for the ML model development.

MLR Model	***Validation Parameters***	***Values***
pIC_{50} = 7.37902(+/-0.70571) -1.38773(+/-0.36842) ***KRFP363*** -0.73307(+/-0.16993) ***AATSC8s*** +0.67696(+/-0.14479) ***minHCsatu*** -0.44968(+/-0.14359) ***PubchemFP686*** +0.59108(+/-0.10738) ***PubchemFP372*** +1.12091(+/-0.49139) ***GATS4m*** +0.30805(+/-0.13704) ***KRFP413*** -0.1472(+/-0.05717) ***APC2D6***_C_Cl -1.8882(+/-0.33024) GATS7p +0.256(+/-0.07326) ***nHBint3***	N_{Train}	104
	SEE	0.3608
	R^2	0.74726
	R^2_A	0.72008
	PRESS	12.10657
	F	22.49677 (DF:10 ,93)
	Q^2	0.66458
	Avg. rm^2_{LOO}	0.543
	N_{Test}	38
	r^2	0.567
	r_0^2	0.53683
	$RMSE_p$	0.48286
	R^2_{pred} (Q^2f_1)	0.54148
	Q^2f_2	0.53679
	Avg. rm^2_{test}	0.44624



(A) **(B)**

Appendix Fig. S1. *ROC* plot obtained for **(A)** training set, and **(B)** test set for the Bayesian classification model.

Preprints

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2023.2298394

A combined ligand-based and structure-based *in silico* molecular modeling approach to pinpoint the key structural attributes of hydroxamate derivatives as promising meprin β inhibitors

Sandeep Jana[a], Suvankar Banerjee[a], Sandip Kumar Baidya[a], Balaram Ghosh[b], Tarun Jha[a] and Nilanjan Adhikari[a]

[a]Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India; [b]Epigenetic Research Laboratory, Department of Pharmacy, Birla Institute of Technology and Science-Pilani, Hyderabad Campus, Shamirpet, Hyderabad, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

Human meprin β is a Zn^{2+}-containing multidomain metalloprotease enzyme that belongs to the astacin family of the metzincin endopeptidase superfamily. Meprin β, with its diverse tissue expression pattern and wide substrate specificity, plays a significant role in various biological processes, including regulation of IL-6R pathways, lung fibrosis, collagen deposition, cellular migration, neurotoxic amyloid β levels, and inflammation. Again, meprin β is involved in Alzheimer's disease, hyperkeratosis, glomerulonephritis, diabetic kidney injury, inflammatory bowel disease, and cancer. Despite a crucial role in diverse disease processes, no such promising inhibitors of meprin β are marketed to date. Thus, it is an unmet requirement to find novel promising meprin β inhibitors that hold promise as potential therapeutics. In this study, a series of arylsulfonamide and tertiary amine-based hydroxamate derivatives as meprin β inhibitors has been analyzed through ligand-based and structure-based *in silico* approaches to pinpoint their structural and physiochemical requirements crucial for exerting higher inhibitory potential. This study identified different crucial structural features such as arylcarboxylic acid, sulfonamide, and arylsulfonamide moieties, as well as hydrogen bond donor and hydrophobicity, inevitable for exerting higher meprin β inhibition, providing valuable insight for their further future development.

ARTICLE HISTORY
Received 15 July 2023
Accepted 16 October 2023

KEYWORDS
Meprin β inhibitor; QSAR; Bayesian model; HQSAR; recursive partitioning; MD simulation

1. Introduction

Every single cell of a living organism goes through various reversible or irreversible modifications throughout its life cycle. Proteolysis by proteases is an irreversible post-translational modulation among them. During the last decade, enzymes involved in proteolysis have been pointed out to be key components not only in the modulation of the immune system, both the development of neurons as well as neurodegeneration but also in taking part crucial roles in apoptosis and cancer progression by extracellular matrix (ECM) remodeling (Broder & Becker-Pauly, 2013). Sterchi and Bond in the early 1980s noticed unexpected proteolysis in the intestine of patients after pancreatic surgery (Sterchi et al., 1982) as well as in mouse kidneys (Beynon et al., 1981), respectively. This observation led to the discovery of novel multidomain zinc-dependent metalloproteases later named meprin (metalloprotease from renal tissue). Meprins are among the astacin family of the metzincin endopeptidases. Shortly thereafter, two individual genes (namely *MEP1A* and *MEP1B*) from different chromosomes where a 41% similarity in sequence was noticed for the meprin α and meprin β proteases (Bond & Beynon, 1995). Mature meprin β has a very characteristic structure with an *N*-terminal signal peptide accompanied by a pro-peptide domain, as well as an astacin-like protease domain with a Zn^{2+}-dependent catalytic site, a MAM domain, a TRAF domain, an EGF-like domain, and also a transmembrane domain along with a cytosolic tail at the *C*-terminal region (Figure 1) (Banerjee et al., 2023; Broder & Becker-Pauly, 2013). The only differentiating factor between meprin α and meprin β is the presence of an inserted domain in meprin α between the TRAF and the EGF domains.

The meprins are majorly expressed in the enterocyte in the small intestine and colon, brush border membrane of the kidney proximal tubule, and to a small extent in the epidermis, blood vessels, lungs, brain, and immune cells (Peters & Becker-Pauly, 2019; Sterchi et al., 1982). Due to this diverse expression, meprin β regulates a number of substrates such as procollagen I, collagen IV, TREM 2, IL-6R, CD109, CD99, MUC-2, amyloid precursor protein (APP), E-cadherin, and IL-1β (Banerjee et al., 2023; Broder & Becker-Pauly, 2013). This diverse tissue expression pattern and wide substrate specificity give meprin β the privilege of influencing various

CONTACT Tarun Jha tjupharm@yahoo.com; Nilanjan Adhikari nilanjan_juphar@rediffmail.com Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India.

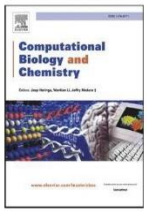
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2023.2298394.

Computational Biology and Chemistry 110 (2024) 108051

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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



An assessment of crucial structural contributors of HDAC6 inhibitors through fragment-based non-linear pattern recognition and molecular dynamics simulation approaches

Suvankar Banerjee [a], Sandeep Jana [a], Tarun Jha [a], Balaram Ghosh [b], Nilanjan Adhikari [a,*]

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

[b] Epigenetic Research Laboratory, Department of Pharmacy, Birla Institute of Technology and Science-Pilani, Hyderabad Campus, Shamirpet, Hyderabad 500078, India

ARTICLE INFO

Keywords:
HDAC6
HDAC6 inhibitor
Classification-QSAR
Non-linear pattern recognition
Molecular docking
Molecular dynamics simulation

ABSTRACT

Amidst the Zn^{2+}-dependant isoforms of the HDAC family, HDAC6 has emerged as a potential target associated with an array of diseases, especially cancer and neuronal disorders like Rett's Syndrome, Alzheimer's disease, Huntington's disease, etc. Also, despite the availability of a handful of HDAC inhibitors in the market, their non-selective nature has restricted their use in different disease conditions. In this situation, the development of selective and potent HDAC6 inhibitors will provide efficacious therapeutic agents to treat different diseases. In this context, this study has been carried out to evaluate the potential structural contributors of quinazoline-cap-containing HDAC6 inhibitors *via* machine learning (ML), conventional classification-dependant QSAR, and MD simulation-based binding mode of interaction analysis toward HDAC6 binding. This combined conventional and modern molecular modeling study has revealed the significance of the quinazoline moiety, substitutions present at the quinazoline cap group, as well as the importance of molecular property, number of hydrogen bond donor-acceptor functions, carbon-chlorine distance that significantly affects the HDAC6 binding of these inhibitors, subsequently affecting their potency . Interestingly, the study also revealed that the substitutions such as the chloroethyl group, and bulky quinazolinyl cap group can affect the binding of the cap function with the amino acid residues present in the loops proximal to the catalytic site of HDAC6. Such contributions of cap groups can lead to both stabilization and destabilization of the cap function after occupying the hydrophobic catalytic site by the aryl hydroxamate linker-ZBG functions.

1. Introduction

Epigenetics, a heritable change in genetic regulation, functioning through the phenotypic change in chromatin structure without altering DNA sequence, influences fundamental cellular and molecular regulation processes necessary for life to exist (Jones, 2007). It is actively influenced by post-translational modifications, such as DNA methylation (Law and Jacobsen, 2009), histone modification (Kouzarides, 2007), chromatin structure remodeling (Goldberg et al., 2007), and noncoding RNA regulation (Hirota et al., 2008). Among these post-translational modifications, the dynamically reversible acetylation of α-amino termini of the lysine residue in histone, catalyzed by histone acetyltransferase (HAT) and histone deacetylase (HDAC) is probably the best understood and widely occurring process (Zhao et al., 2010). The dynamics of these enzymes regulate the equilibrium between the acetylation levels of nuclear and cytoplasmic proteins, which is pivotal for maintaining cellular homeostasis (Yoshida et al., 2001). The shift in equilibrium towards HAT contributes to hyperacetylation of the target gene resulting in continuous genetic expression whereas the shift towards HDAC results in continuous genetic repression (Park and Kim, 2020). Therefore, dysregulation of HDAC leads to the development of several diseases, including several forms of cancer, neurological abnormalities, autoimmune disease, as well as cardiac and pulmonary disease (Shakespear et al., 2011; Bagchi and Weeks, 2019; Zhang et al., 2021; Sarkar et al., 2020). Not only histone but also several other non-histone substrates of HDACs are heat shock protein 90 (Hsp90), α-tubulin, p53, E2F, cortactin, and Myo-D, leading to much more diverse and complicated functions in cellular processes (Glozak et al., 2005;

* Corresponding author.
E-mail address: nilanjan_juphar@rediffmail.com (N. Adhikari).

https://doi.org/10.1016/j.compbiolchem.2024.108051
Received 31 December 2023; Received in revised form 28 February 2024; Accepted 8 March 2024
Available online 11 March 2024