

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Науково-навчальний інститут високих технологій

Допущено до захисту
Кафедра молекулярної біотехнології та біоінформатики
Зав. кафедри: к.б.н, доц. О.Ю. Нипорко
Протокол №____ засідання кафедри
від “____” _____ 2024 р.

**Дослідження ліганд-білкових взаємодій ефекторів з встановленою та
потенційною дією на гістондеацетилази 2го класу**

Випускна кваліфікаційна робота магістра
студента напряму підготовки 091 біологія
Дзьобака Андрія Володимировича

Науковий керівник від кафедри
Доцент кафедри молекулярної біотехнології
та біоінформатики к.б.н., **Самофалова
Дарія Олексіївна**

)

Робота виконана на базі ТОВ «Хімічна компанія «Сполука» під керівництвом
завідуючої лабораторії комп'ютерної хімії Самофалової Д.О.

Оцінка захисту роботи

Київ – 2024 р.

АНОТАЦІЯ

Дзьобак А.В. Дослідження ліганд-білкових взаємодій ефекторів з встановленою та потенційною дією на гістондеацетилази 2го класу. Випускна кваліфікаційна робота магістра за спеціальністю 091 Біологія ОП «Біоінформатика і структурна біологія».

У роботі проведено реконструкцію та аналіз структури гістондеацетилаз класу II – HDAC4, HDAC6, HDAC7, HDAC10, проведено молекулярний докінг з зібраною бібліотекою ефекторів з встановленою інгібуючою дією до родини гістондеацетилаз та бібліотекою для високопропускового віртуального скринінгу, ідентифіковано потенційні ліганди з селективною та неселективною дією до білків та клас-специфічні ліганди, обчислено дескрипторні характеристики, що використовуються для передбачення фармакологічної дії в правилах Ліпінського та визначено клас-специфічні відмінності сполук. Побудовано з використанням методу зменшення вимірності UMAP хімічний простір ефекторів для кожного з білків-об'єктів дослідження. Визначено потенційні клас-специфічні інгібітори гістондеацетилаз.

Методичні підходи та результати дослідження було апробовано на 2х конференціях: 28th National Meeting on Medicinal Chemistry (NMMC28) Balabon O, Stykhylias M, Samoilenko V, Dzobak A. Docking-based design of novel inhibitors targeting mrp1 in cancer drug discovery. 17th-20th September 2023. Chieti, а також на IV конференції молодих учених «Біологія рослин та біотехнологія». 16-18.06.2024 за результатами якої було видано збірник тез, що включає публікацію «Порівняльний аналіз ліганд-білкових взаємодій гістондеацетилаз класу IIb різного еволюційного походження», Стихилиас М. М., Дзьобак А. В., Раєвський О. В., Блюм Я.

Ключові слова: HDAC4, HDAC6, HDAC7, HDAC10, хімічний простір, молекулярний докінг.

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

HDAC	– гістондеацетилаза
ДНК	– дезоксирибонуклеїнова кислота
НАТ	– гістонацетилтрансфераза

ЗМІСТ

АНОТАЦІЯ	2
ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	3
ЗМІСТ	4
ВСТУП	6
РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ	8
1.1 Загальна характеристика гістондеацетилаз	8
1.2 Класифікація гістондеацетилаз.....	10
1.2.1 Надродина аргіназ/деацетилаз	11
1.2.2 Родина сиртуїнів.....	14
1.3 Функції HDAC в контексті здоров'я та патологічних станів людини	14
1.4 Інгібітори гістондеацетилаз.	16
1.4.1 Структурно-орієнтований спосіб драг-дизайну.....	16
1.4.2 Похідні бістіазолу.....	17
1.4.3 Похідні бензаміду.....	17
1.4.4 Малі кільцеві молекули	18
1.4.5 Селективні інгібітори.	18
РОЗДІЛ 2. МАТЕРІАЛИ ТА МЕТОДИ	21
2.1 Відбір датасетів	21
2.2 Побудова структурних моделей білка HDAC10.....	21
2.3 Оптимізація структури білку.....	23
2.4 Визначення рецепторів на поверхні білку.	24
2.5 Побудова карти зв'язування сайту	24
2.6 Підготовка референтного сету та HTS Stock Compounds до докінгу	25
2.7 Докінг лігандів	25
2.8 Визначення фізично-хімічних параметрів та побудова дескрипторів для аналізу характеристик молекул-хітів.....	26
2.9 Побудова хімічного простору для інгібіторів HDAC10.....	27
2.10 Статистичний аналіз значень медіан розподілів дескрипторних характеристик для об'єктів дослідження та підкласів	28

РОЗДІЛ 3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ	31
3.1 Визначення доменної організації та структурної класифікації об'єктів дослідження.	31
3.2 Реконструкція моделей протеїну HDAC10	32
3.3 Оптимізація просторової структури білків	35
3.4 Визначення поверхонь сайтів зв'язування з білком та побудова карт зв'язування.....	37
3.5 Підготовка бібліотек низькомолекулярних сполук до молекулярного докінгу.....	40
3.6 Молекулярний докінг референтного сету ефекторів	41
3.8 Статистичний аналіз з метою пошуку підклас-специфічних характеристик лігандів.	46
3.9 Побудова хімічного простору.	48
ВИСНОВКИ	54
СПИСОК ДЖЕРЕЛ	55
<i>Додаток 1</i>	60
<i>Додаток 2</i>	62

ВСТУП

Дослідження взаємодії родини гістондеацетилаз продовжується вже майже 40 років, фокусуючись на епігенетичному аспекті регуляції експресії генів внаслідок деацетилювання лізинових залишків гістонів, на їх негістонних партнерах, що залучені до різноманітних аспектів клітинної регуляції, зокрема поділу клітин, аутофагії та апоптозу. А також важливій ролі HDAC в патогенезі різноманітних захворювань – онкологіях, хворобі Паркінсона, обструктивному захворюванні легень, тощо.

Великий спектр функцій та білків з якими взаємодіють представники родини HDAC вимагають точного та ефективного інгібування не просто родини білків, а конкретного класу або навіть конкретного ферменту. Разом з цим тема селективного інгібування частини або одного з представників родини HDAC є доволі малодослідженою, але надзвичайно актуальною.

Відповідно до потреби розуміння принципу селективних взаємодій було обрано об'єкти дослідження: «гістондеацетилази класу II - HDAC4,6,7,10». А **метою роботи** було порівняння властивостей ефекторів гістондеацетилаз підкласів Па та Пб для встановлення білок-специфічних характеристик лігандів з селективною встановленою дією та потенційних інгібіторів HDAC.

Відповідно до мети було поставлено наступні **завдання**:

1. На основі структурних даних та класифікації родини ферментів зокрема 4,6,7, реконструювати та верифікувати просторову структуру гістонової деацетилази типу 10.

2. Проаналізувати потенційні сайти зв'язування на поверхні обраних гістондеацетилаз та виконати пошук інгібіторів для об'єктів дослідження методом високопродуктивного віртуального скринінгу.

3. Визначивши селективні інгібітори кожного типу та проаналізувати дескрипторні характеристики обраних молекул.

4. Побудувати та описати хімічний простір інгібіторів родини HDAC з врахуванням дослідження підклас-специфічних характеристик лігандів.

РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ

1.1 Загальна характеристика гістондеацетилаз

Гістондеацетилази (HDAC) є протеїнами, що каталізують відщеплення ацетильної функціональної групи від залишку лізину. Їхня ензиматична активність діє як на гістонні так і негістонні білки. В людей охарактеризовано 18 представників родини гістондеацетилаз, що використовують або цинк- або NAD^+ - залежний механізм деацетилювання ацетиллізинового субстрату. Ця функціональна властивість напряду впливає на структуру хроматину – при деацетилюванні гістони мають змогу більш тісно з'єднуватися з ДНК, що в свою чергу призводить до збільшення ступеню конденсації хроматину.

Фундаментальною одиницею хроматину є октамер чотирьох гістонів – тетрамер H3/H4 та два H2A/H2B димери, що оточені 146 парами нуклеїнових основ (Рис.1.1). При цьому структурна модифікація гістонів H3 і H4 значно краще охарактеризована, порівняно з H2A і H2B [1].

Під час активації транскрипції генів доступ рибосом до ДНК забезпечується модифікацією нуклеосоми. Ацетилювання у всіх гістонів відбувається на N-кінці еволюційно консервативного ϵ -амінокислотного лізинового залишку за участю гістонлізин ацетилтрансферазних ферментів (HATs/KATs).

Деацетилювання ж є процесом, що відповідає за баланс експресії генів та є оберненим до ацетилювання процесом, що розпочинається з залучення гістондеацетилазних комплексів шляхом зв'язування з метил-СpG-зв'язуючими, або іншими СpG-асоційованими білками, що рекрутуються метильними групами так званих СpG-острівців безпосередньо ДНК. При цьому результати досліджень вказують на те, що HDAC залучені не лише до

регуляції транскрипції, а до епігенетичної регуляції в цілому, контролюючи, запобігаючи або видаляючи інші лізин-залежні пострнансляційні модифікації – метилювання, убіквітинації та сумоїлювання [2]. Таким чином гістондеацетилази є залученими до фундаментальних клітинних процесів – транскрипції та епігенетичної регуляції.

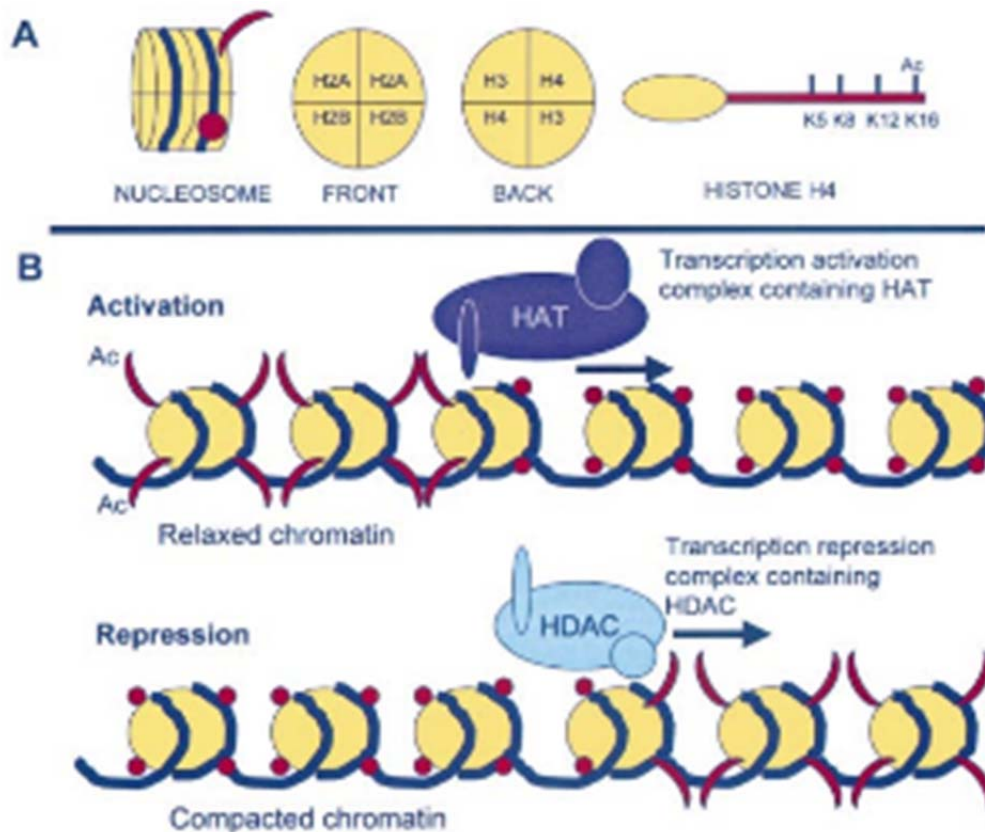


Рис. 1.1 Структура хроматину. Принцип дії гістондеацетилаз [3].

Разом з цим гістондеацетилази залучені також до регуляції цілої низки клітинних процесів [4], будучи, зокрема залученими до деацетилювання протеїнів p53m E2F, α -тубуліну та MyoD [5].

Баланс між гістон-ацетилтрансферазами (HAT), які приєднують ацетильні групи до гістонів, і HDAC, які їх видаляють, має вирішальне значення для регуляції експресії генів і є основним об'єктом досліджень в епігенетиці та розробці терапевтичних агентів.

Таким чином, HDAC є ключовими регуляторами динаміки хроматину та експресії генів, що має широке значення для клітинної функції, розвитку та захворювань. Їх різноманітні ролі та механізми дії роблять їх важливими мішенями для терапевтичного втручання при різних станах, включаючи рак, нейродегенеративні захворювання та запальні розлади.

1.2 Класифікація гістондеацетилаз

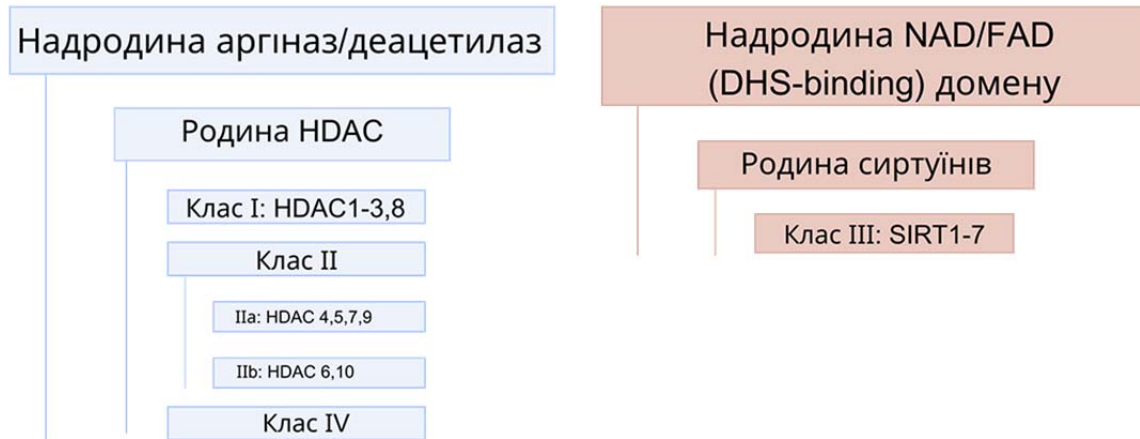
Гістондеацетилази розділені на дві сім'ї протеїнів – безпосередньо гістондеацетилаз та Sir2 регуляторну родину. В протеомі людини гістондеацетилази також традиційно розділені за класами відповідно до схожості послідовностей.

До першого класу (Class I) входять протеїни HDAC1, HDAC2, HDAC3 і HDAC8. Вони є структурно схожими до дріжджової деацетилази Rpd3 і розташовані переважно в ядрі. Вони беруть участь у регуляції експресії генів і є критично важливими для прогресії та диференціації клітинного циклу.

Гістондеацетилази класу II (Class II) поділяються на клас IIa (HDAC4, HDAC5, HDAC7 і HDAC9) і клас IIb (HDAC6 і HDAC10). Вони є структурно схожими до дріжджового білку Hda1 та до білків Hos. HDAC класу IIa переміщуються між ядром і цитоплазмою і беруть участь у тканинно-специфічній експресії генів і процесах розвитку. HDAC класу IIb мають додаткові функціональні домени, що включають їх роль у цитоплазматичних шляхах передачі сигналу.

Представники класу III гістондеацетилаз (Class III) також відомі як сиртуїни (SIRT1-7). Вони гомологічні дріжджовій деацетилазі Sir2 і потребують NAD⁺, як кофактор для своєї ферментативної активності. Сиртуїни регулюють різні фізіологічні процеси, включаючи старіння, транскрипцію, апоптоз і стійкість до стресу, шляхом деацетилювання гістонових і негістонових білків

HDAC класу IV (Class IV) - цей клас має характеристики, як клас I, так і клас II і HDAC11 є єдиним представником свого класу. Його функції менш вивчені, але вважається, що він бере участь у регуляції експресії генів та



імунних реакцій.

Рис. 1.2 Схематичне зображення класифікації родини гістондеацетилаз.

1.2.1 Надродина аргіназ/деацетилаз

Надродина аргіназ/деацетилаз або класична сім'я HDAC складається з трьох класів протеїнів – I, II та IV. Протеїни в класах мають спільного «попередника», схожі структурно та мають високу гомологію послідовностей, разом з цим мають схожі функції. Надродина містить аргіназоподібні імідогідролази, гістонові деацетилази. Основною гіпотезою походження родини є еволюційні зміни прокаріотичного ферменту, що мав функцію зворотнього ацетилювання та деацетилювання аміноалкільної групи ДНК-зв'язуючої молекули, досягаючи таким чином регуляторного ефекту[6]. При цьому цікавим є те, що HDAC надродина аргіназ/деацетилаз еволюційно виникли раніше за гістони, що може вказувати на те, що вихідним субстратом родини є не гістоновий білок.

1.2.1.1 Клас I

Гістонові деацетилази класу I рекрутуються та активуються спорідненими білками-корекспресорами, що спрямовують активність HDAC на конкретні локуси хроматину. Це в свою чергу спричиняє просторовий та часовий контроль експресії генів.

Найближчу структурну схожість представники класу мають до регулятора транскрипції дріжджів (*S. cerevisiae*) - Rpd3.

Клас I асоціюється з чотирма великими корекспресорними білковими комплексами – SMRT/NCOR для HDAC3 та Sin3A, CoREST і NuRD для HDAC1 і HDAC2 [7].

Взаємодія гістондеацетилаз з гістонами в контексті функціональних та конформаційних зв'язків з великими корекспресорними білками є об'єктом пильного дослідження. При цьому було доведено, що HDAC1 та HDAC2 також мають здатність до зв'язування безпосередньо з ДНК-зв'язуючими протеїнами, зокрема YY1 [8].

Структурно протеїни HDAC1 і HDAC2 мають велику ступінь спорідненості, з N-кінцевим каталітичним доменом, що складає більшу частину молекули.

Іншою цікавою функцією гістондеацетилаз класу I є їхня участь в епігенетичному регулюванні транскрипції внаслідок деацетилювання гістон лізин ϵ -N-L-лактилювання [K(L-la)]. Важливим є розуміння того, що саме HDAC I класу беруть участь в забезпеченні сигнальної функції кінцевого метаболіту – L-лактату [9].

1.2.1.2 Клас II

Представники класу II найбільш близькі за гомологією послідовностей доменів до HDA1 – деацетилази описаної в дріжджів та мають додаткові доменні послідовності, які не зустрічаються в класі I і, власне, є причиною відокремлення HDAC4, HDAC5 та HDAC6 в окрему таксономічну групу.

Структурно найбільш спорідненими є HDAC4 та HDAC5, ферменти класу II містять каталітичний домен на С-термінальному кінці молекули.

На відміну від HDAC класу I, представники класу II експресуються в меншій кількості типів клітин і, відповідно, тканин, таким чином даючи нам можливість зробити висновок, що вони можуть бути залучені до клітинної диференціації і процесів розвитку [10]. Наприклад був доведений вплив HDAC 4,5,7,9 на диференціацію м'язових клітин шляхом репресії міогенного фактору транскрипції MEF2.

Клітинна їх локалізація є предметом дослідження, але активність ферментів класу спостерігалася в цитоплазмі, що може бути клас-специфічною особливістю.

HDAC 6 є особливим членом класу, оскільки містить дуплікацію – два деацетилазних каталітичних домени, що можуть функціонувати незалежно.

HDAC 10 є білком, що найбільш гомологічно схожий до HDAC6, адже також містить дуплікацію каталітичного деацетилазного домену. Ця особливість дозволяє виокремити HDAC6 та HDAC10 в окремий клас – IIb.

1.2.1.3 Клас IV

Єдиним членом класу IV є HDAC 11, що має гомологічну послідовність до доменів як класу I так і класу II. Ця гістондеацетилаза була виявлена внаслідок BLAST-пошуку GenBank, де Hos3 протеїн був цільовою послідовністю до HDAC11. Згідно філогенетичного аналізу можна зробити висновок, що HDAC найбільш близько споріднений з HDAC3 та HDAC8, що в свою чергу може вказувати на вищу ступінь спорідненості з класом I, порівняно з класом II [3].

Каталітичний домен ензиму локалізований на N-кінці, та його активність не пов'язана з наведеними вище білковими комплексами. Дослідження Glozak et. al. [11] вказують на можливий вплив HDAC11 на стабільність

фактору реплікації CDT1. Разом з цим, HDAC11 залишається білком, що потребує подальшого детального вивчення.

1.2.2 Родина сиртуїнів

Надродина NAD/FAD дезоксигіпузинсинтазозв'язуючого (DHS-binding) домену включає в себе серед іншого білки мовчазного інформаційного регулятора 2 (Sir2). Білок-засновник Sir2 був ідентифікований в *S. cerevisiae* внаслідок генетичного скринінгу, ціллю якого було визначення генів, залучених до сайленсингу транскрипції.

Сиртуїни виділені в окрему родину, що містить клас III – нікотинамід аденіннуклеотид-залежних гістондеацетилаз. Членами родини є білки SIRT1-7, вони є філогенетично консервативними, тобто їх структура схожа як в прокариот, так і в еукаріотичних організмів, архей.

1.2.2.1 Клас III

Першими нову родину гістондеацетилаз описали Лорейн Піллус та Джеф Боке [12]. Структурно сиртуїни зберігають 22-50% схожості амінокислотного ланцюга та 27-88.

Клас III особливий в контексті функцій – його представники мають дві функції моно-АДФ-рибохилтрансферази та гістондеацетилази. Їх клітинна локалізація варіюється: SIRT1 і SIRT2 присутні як в ядрі, так і в цитоплазмі; SIRT3 - в ядрі і мітохондріях; SIRT4 і SIRT5 - в мітохондріях; SIRT6 - в ядрі; а SIRT7 - в ядерці. Подібно до HDAC класів I, II і IV, сиртуїни націлені на негістонові субстрати в еукаріотів.

1.3 Функції HDAC в контексті здоров'я та патологічних станів людини

Потреба в вивченні механізмів функціонування HDAC простягається далі ніж лише регуляція епігенетичної експресії генів, розуміння модифікування гістонів, організації хроматину. HDAC, як родина білків є критично важливою для організмів, адже з одинадцяти протеїнів лише три при нокауті відповідного гена не призводять до важких патофізіологічних станів або летального випадку при спостереженні за нокаут фенотипом [13].

HDAC є ключовою точкою інтересу для дослідників, що займаються раковими захворюваннями. Відмінна від нормальної експресія HDAC в тому числі пов'язана з підвищеним ризиком виявлення недоброякісних пухлин, більшій летальності захворювань [2]. Разом з тим Dovey et al. спостерігали значний внесок гістондеацетилаз до стабільності геному, при вивченні розвитку T-клітин мишей [14]. Схожі результати отримали також Santoro et al. в контексті виявлення пухлиноsupресуючих властивостей HDAC1 [15].

Пропоноване пояснення цих спостережень полягає в тому, що втрата HDAC їх функцій може сприяти втраті репресії певних генів, в свою чергу викликаючи злякисну трансформацію клітини.

Окрім ракових захворювань, гістондеацетилази пов'язують з патофізіологічними станами серця. Так, зокрема надмірну експресію гістондеацетилаз класу I пов'язують з виникненням ішемічної хвороби серця внаслідок дослідження Ca^{2+} /кальмодулін-залежної моделі серцевої недостатності, спричиненої активацією кінази II [16].

Так Song et al. в дослідження виявили, що надмірна експресія гістондеацетилаз класу IIa призводить до синергічного інгібування MEF2-залежної транскрипції, тобто інгібує гіпертрофію кардіоміцитів, запобігаючи розширенню шлуночків та прояву патологічного фенотипу серця [17].

Іншою дослідженою роллю HDAC є запальні процеси. Зокрема в пацієнтів з хронічним обструктивним захворюванням легень, при дослідженні біоптату та клітин крові було виявлено суттєве зниження, як експресії так і активності HDAC2. Зниження активності корелювало з тяжкістю захворювання та силою запального процесу. HDAC3 асоційована з

активацією значної частки запальних клітинних каскадів під впливом ліпополісахаридної стимуляції в макрофагах.

З огляду на варіативність патофізіологічних станів, що асоційовані з HDAC, логічним є пошук інгібіторів, що матимуть змогу інгібувати дію HDAC та таким чином мати регуляторний вплив на їх функції.

1.4 Інгібітори гістондеацетилаз

1.4.1 Структурно-орієнтований спосіб драг-дизайну

Логічним підходом до пошуку інгібіторів HDAC є дизайн, що повторює проміжний продукт тетраедричного оксианіону ензиму. Відповідно молекула має утворювати координаційні зв'язки з катіоном цинку в сайті зв'язування та матиме тонкий гідрофобний лінкер, який зможе поміститися в канал субстрату. Також важливим є “якір”(кеп) з іншої сторони лінкеру, який може вступати в взаємодію з краєм ензиму та посилювати зв'язування інгібітора з гістондеацетилазою. Такий структурно-орієнтований раціональний спосіб побудови інгібітора був запропонований Jung et al. в 1997 році.

Головною молекулою, похідні якої стали цілями для інгібуючої дії є гідроксамова кислота ($R-C(=O)-N(-OH)-R'$), завдяки її властивості хелатувати йони цинку в активному центрі.

Першим з таких похідних є Vorinostat (Рис. 1.3), що відповідає всім умовам фармакофорної гіпотези сформованої вище. Цей інгібітор HDAC затвердили в 2006 році з цілю лікування Т-клітинної лімфоми шкіри. В 2016 році Chen et al. запропонував використовувати морфолінопурин як кап, в контексті дизайну нової серії похідних для інгібування гістондеацетилаз [18]. Результати досліджень вказують на меншу цитотоксичність та більшу оральну біодоступність, порівняно з інгібіторами гістондеацетилаз, що використовуються в якості лікарських препаратів.

Всього затверджено п'ять інгібіторів гістондеацетилаз, а ще близько 30 беруть участь в клінічних дослідженнях.



Рис. 1.3 Структура Vorinostat [18].

Наразі, окрім Vorinostat (Рис. 1.3) затверджені FDA для лікування: Romidepsin, Belinostat, Panobinostat, Tucidinostat.

1.4.2 Похідні бістіазолу

Це клас структурно спрощених бістіазол-похідних молекул на основі природньої сполуки лагразолу з тіазол-тіазоліновим кепом. Вперше запропонований Chen et al. в 2014 році [19].

Декілька з синтезованих сполук продемонстрували мікромольні або наномольні значення інгібування IC₅₀, одна сполука (8f) проявила ефективність у полегшенні клінічних симптомів енцефаломієліту у мишей при пероральному введенні.

Також похідні бістіазолу – бістіаностат в поєднанні з bortezomib показують значне інгібування росту пухлин на моделі дослідження численної мієломи мишей, при оральному споживанні [20].

1.4.3 Похідні бензаміду

Іншим типовим класом інгібіторів гістондеацетилаз є бензаміди. Зокрема нещодавно Tucidinstat було запропоновано для лікування пацієнтів з периферальною Т-клітинною лімфомою, де він показав об'єктивний відсоток відповідей 46% [21]. Зокрема ще 5 інших інгібіторів класу знаходяться на різних етапах клінічних випробувань.

1.4.4 Малі кільцеві молекули

Підхід до дизайну різних кеп-груп, що складаються з циклопептидів де амінокислоти є гідрофобними, а лінкер – алкільний ланцюг, виокремив інший клас інгібіторів гістондеацетилаз. Одним з перших запропонованих інгібіторів є природній циклічний тетрапептид Apicidin, що вирізнявся кетонною групою для хелатування цинку. Іншими представниками класу, що вивчаються на клінічних дослідженнях є лагразол, LB100, M-344 [22].

Загалом, основною проблемою інгібіторів гістондеацетилаз є висока цитотоксичність, що робить їх обмежено придатними для клінічного використання проти твердих пухлин а також пацієнтів з лейкемією. Відповідною є потреба сформувані клас-специфічні, або навіть ізоформ-специфічні інгібітори гістондеацетилаз.

1.4.5 Селективні інгібітори.

Селективні інгібітори класу I є різними за типом інгібіторами – 2020 року була описана серія похідних арилкетонів, де арилкетон виступав як цинк-зв'язуюча група, дві похідних сполуки продемонстрували високу селективність [23].

Іншим цікавим селективним інгібітором є запропонований Li et al. молекула бензамідного похідного, зокрема проявивши і антипухлинні властивості в U937 та HCT116 ксенографтових моделях. Цей дизайн ліг в

основу ще одного, додаткового похідного о-амінобензаміду, що показало сильне зв'язування та селективність до класу.

Декілька досліджень спостерігали ізоформ-селективні інгібітори для HDAC3 [24].

Дослідження тетразаміщених циклопропанових похідних гідроксамової кислоти показало їх високу афінність саме до класу IIa HDAC [25]. Разом з цим частина сполук була розроблена з інгібуванням конкретних ізоформ. Для HDAC10 та HDAC6 – похідні на базі Тубастатину А – Marbosta 100.

Для класу III та IV таргетні інгібітори є суттєво менш дослідженими, зокрема з огляду на малий об'єм інформації, що є відомим про ензиматичну активність відповідних представників класів. Наприклад лише нещодавно було з'ясовано, що HDAC11 працює як дефаттіацилаза (деацетилаза жирних кислот), а не як класична деацетилаза гістонів [26].

Цитотоксичність, обмежена ефективність, резистентність цілі до ліків змушує дослідників не лише робити ліки високоспеціалізованими до класу, або ізоформи, а й також більше звертатися до інгібіторів подвійної дії. Зокрема рак, захворювання центральної нервової системи потребують комплексного підходу, що міг би впливати на одразу декілька цілей в організмі пацієнта. Наприклад інгібування різних регуляторних каскадів може дати більший синергетичний ефект, наряду з вищою ефективністю лікування та меншою шкодою для пацієнта [27].

Щодо інгібіторів гістондеацетилаз подвійної дії виділяють також такі додаткові мішені: фосфодіестерази (PDE), фосфатидинітозол 3-кінази (PI3K), янус кінази (JAK1), протеїни з бромодоменом (BRD4), нікотинамідфосфорибозитолтрансферази, індоламін2,3-діоксигенази, EZH2, та білок б-клітинної лімфоми 2 (Bcl 2).

З розуміння важливості гістондеацетилаз, як основних регуляторів експресії генів, клітинних процесів, а також усвідомлення нагальності вивчення додаткових функцій HDAC, виростає розуміння необхідності мати можливість таргетно та з високою ефективністю зв'язати конкретну

ізоформу. Великим бонусом буде можливість визначення того, чи мають високоселективні інгібітори потенційну можливість таргетингу інших клітинних процесів, з метою лікування патофізіологічних процесів.

РОЗДІЛ 2. МАТЕРІАЛИ ТА МЕТОДИ

2.1 Відбір датасетів

Оскільки об'єктами дослідження є фізико-хімічні параметри та структурна взаємодія малих органічних молекул з протеїнами HDAC10 та HDAC6 класу ІІв для пошуку потенційних лігандів було використано референсний датасет структур інгібіторів родини гістондеацетилаз отриманий з баз даних ChEMBL, PubMed, GooglePatents, ChemPub. Сет було агрегровано з використанням всіх доступних баз даних та відфільтровано, відкинувши неактивні сполуки, сполуки з низькою активністю (IC_{50} , K_i та ін., більше ніж 10 000 нМ, інгібування < 50%), а також сполуки з низькою масою (<120) та реактивні сполуки.

Структури сполук в форматі SMILES були опрацьовані програмним забезпеченням SYBYL-X (Tripos, USA) позбавлені солей та інших домішок, для структур було згенеровано двовимірні координати та збережено в форматі SD з видаленням дублікатів.

Тестовим датасетом було обрано HTS Stock Compounds – базу сполук компанії LifeChemicals, що містить понад 500 000 оригінальних малих органічних сполук для високопродуктивного скрінінгу.

2.2 Побудова структурних моделей білка HDAC10

Інформацію про послідовність та функції білка HDAC10, а також множинне структурне вирівнювання та інформацію про ідентичність послідовностей було отримано з веб-сервісу UniProtKB.

Для побудови та структурних моделей білка HDAC10 використовувалися два сервіси з моделювання.

I-TASSER (Iterative Threading ASSEmbly Refinement) – веб-сервіс що використовує ієрархічний підхід до передбачення структури протеїнів та структурної анотації, відносно їх функцій. Моделювання поєднує моделі глибокого навчання з моделюванням за гомологією з метою передбачення структур та функцій протеїнів, з використанням протоколів парсингу за послідовністю, згортанням доменів по черзі, інтердоменною збіркою доменів та функціональною анотацією, що базується на відомих структурних властивостях білку [28-30]. Оцінка згенерованої моделі відбувалася з використанням метрик сервісу :

- C-score (confidence score) – метрики передбачення якості отриманої моделі I-TASSER, що обчислюється виходячи з значення вирівнювання шаблонів протягування та параметрів збіжності моделювання збірки. Нормальним значенням цього параметру вважається -5.4, де вищі значення означають більшу оцінку якості збірки алгоритмом.
- RMSD (Root Mean Square Deviation) – в біоінформатиці середньоквадратичне відхилення атомних позицій – міра середньої відстані між атомами накладених молекул. Для вивчення конформацій білкових структур середньоквадратичне відхилення використовують при визначенні схожості трьохвимірної структури молекули за допомогою СКВ атомних координат $C\alpha$ після оптимального накладання молекули-зразка.
- TM-score – показник вимірювання структурної подібності між двома структурами, що зважає малу відстань між атомом згенерованої та оригінальної структури більше ніж велику, що дозволяє цьому показнику бути нечутливим до локальних помилок (неправильно орієнтація хвоста). Значення >0.5 вказує на правильну топологію моделі, а значення < 0.17 означає випадкову схожість.

Ще одним сервісом, що використовувався для моделювання структури HDAC10 був Robetta сервер з використанням технології глибинного навчання RoseTTAfold.

Методом оцінки точності в сервісі є GDT – метрика що описує відхилення між реальною моделлю, отриманою з даних експерименту та згенеровану з використанням серверу. Найкраще значення 1.00, найгірше – 0.

2.3 Оптимізація структури білку

Оптимізація структури білку проводилася в програмному забезпеченні Shrodinger Maestro з використанням академічної ліцензії.

Для використання функціоналу Structure refinement необхідно використати повноатомну структуру що містить всі гідрогени. Для цього було використано Protein Preparation.

Protein Refinement - це функціонал, що використовує методи молекулярної динаміки з метою мінімізації енергії структури або їх частин з врахуванням змін координат. Для мінімізації використовувалося поле OPLS3e та система розчинника VSGB.

З огляду на обмежені обчислювальні потужності та великий розмір білку, для мінімізації енергії системи було обрано Conjugate Gradient Method (CGM).

Суть методу полягає в ітеративному вирішенні лінійного рівняння форми $Ax = b$, де A – відома квадратна симетрична позитивно-невизначена матриця, b – відомий вектор, а x – невідомий. В контексті мінімізації енергії вектор x – позначає сукупність атомних координат n атомів системи. Зі зміною координат відбувається зміна енергії системи V . З використанням градієнтного спуску визначається такий вектор x що задовольнятиме мінімум V . Градієнт функції завжди буде вказувати в такому випадку в напрямку найбільшої зміни енергії [31].

Для визначення якості оптимізації структури було використано інструмент Protein Structure Quality Report, що дає інформацію про вигляд графіку Рамачандрана, відмінність в довжинах зв'язків, кутів, значення – G- та B- факторів, планарності, відсутніх атомів та стеричних зіткненнях.

Основними мірами є кількість кутів, що не задовольняють умови Рамачандрана, RMSD для зв'язків та кутів, кількість амінокислот в яких остов або бічні ланцюги не задовольняють стандартним умовам, відхилення планарності пептидів, бічних ланцюгів а також торсійних кутів.

2.4 Визначення сайтів зв'язування на поверхні білку

Для визначення сайтів зв'язування на поверхні білку були використані інструменти PDBSum та SiteMap. Вони дають інформацію про потенційне положення «пор» або «сайтів» що здатні до зв'язування лігандів, дають уявлення про потенційну ефективність зв'язування та можливість алостеричної взаємодії. Метриками для SiteMap є SiteScore, розмір сайту, сила зв'язування ліганду з сайтом, гідрофобні або гідрофільні взаємодії в сайті, кількість донорів та акцепторів водню.

2.5 Побудова карти зв'язування сайту

Після визначення активного сайту необхідно реконструювати карту зв'язування – модель, що репрезентує фізичні та хімічні характеристики об'єму рецептора. Для цього було використано GridGeneration з вказуванням відповідного об'єму відведеного для потенційного ліганда та загального розміру області де може знаходитися ліганд.

2.6 Підготовка референтного сету та HTS Stock Compounds до докінгу

Сполуки, отримані внаслідок фільтрації референтного сету були збережені в форматі *.sdf та використані для підготовки бази для докінгу сполук з використанням інструменту LigPrep. В основі лежить генерація тривимірних структур для великої кількості молекул, що використовує вихідні дво- чи тривимірні дані. Це дозволяє нам не лише генерувати ліганди з мінімізованою енергією і правильною хіральністю для кожної вихідної молекули, а генерувати різні стани іонізації, таутомери, стереоізомери і конформації циклів, паралельно з цим накладаючи фільтри, що допомагають відкидати сполуки за параметрами, зокрема молекулярною вагою, наявністю певних груп.

Параметри генерації включали:

- Стандартне силове поле;
- Генерація іонізації молекул при рН від 5.0 до 9.0;
- Збереження вказаної стереоізомерії при варіації всіх інших стереоцентрів.

Відповідно структури було імпортовано і для кожної було додано неявні атоми водню, де це необхідно, відкинуто небажані молекули (солі, домішки), нейтралізовано заряджені групи, згенеровано іонізаційні стани, таутомери. згенеровано різні варіанти хіральностей та оптимізовано геометрію.

2.7 Докінг лігандів

Для побудови моделі взаємодії згенерованої молекули та рецептора використовували інструмент Glide Docking. Цей інструмент, розроблений для

високопродуктивного віртуального скринінгу потенційних лігандів, залежно від афінності та типу зв'язування.

Методологія роботи інструменту полягає в використанні серії ієрархічних фільтрів з метою пошуку можливих положень ліганду в сайті зв'язування рецептору.

Було згенеровано та додано карту зв'язування – віртуальну область простору біля рецептора, що описує параметри рецептора та безпосередньо буде наповнюватися лігандами для оцінки їх потенційної взаємодії з рецептором, та безпосередньо характером цих взаємодій.

Карта зв'язування є набором полів, що прогресивно відсіюють сполуки, таким чином генеруючи точніші та точніші передбачення характеру взаємодії сполук та взагалі можливості взаємодії.

Таким чином після взаємодії з картою зв'язування генерується набір конформацій (скруток) лігандів в просторі, положення яких доопрацьовується відносно рецептора використовуючи OPLS3 та OPLS2005 моделі. Після цього найкращі результати положення для ліганда мінімізуються в полі рецептора з врахуванням обертальних зв'язків ліганда, виконуючи таким чином пост-докінгову оптимізацію.

Надалі було обрано файл, що містить ліганди для докування. Стандартний потенціал для неполярних частин ліганда (scaling factor) було обрано в значенні 0.8.

Точність для оцінки взаємодій було обрано стандартну, 10-15% отриманих результатів було відібрано для проведення екстра точної оцінки взаємодій.

Інші налаштування було залишено стандартними, включно з ініціацією пост-докінгової мінімізації, що дозволить збільшити якість відображуваних даних.

2.8 Визначення фізично-хімічних параметрів та побудова дескрипторів для аналізу характеристик молекул-хітів

Для визначення фізико-хімічних параметрів сполук були використано хемоінформатичний програмний продукт RDKit, що є прекомпільованою на мові C++, але використовує мову програмування Python, як мову програмування другого рівня для команд. Також використовувалися бібліотеки NumPy, Matplotlib та Pandas. Код програми для обчислення дескрипторів та побудови гістограм:

```
def calculate_descriptors(smiles_list):
    descriptors = {'H_donors': [], 'H_acceptors': [], 'MolWt': [], 'cLogP': [],
                  'TPSA': [], 'FSP3': []}

    for smiles in smiles_list:
        mol = Chem.MolFromSmiles(smiles)
        if mol: # Check if the molecule could be parsed
            descriptors['H_donors'].append(Lipinski.NumHDonors(mol))
            descriptors['H_acceptors'].append(Lipinski.NumHAcceptors(mol))
            descriptors['MolWt'].append(Descriptors.MolWt(mol))
            descriptors['cLogP'].append(Descriptors.MolLogP(mol))
            descriptors['TPSA'].append(Descriptors.TPSA(mol))
            descriptors['FSP3'].append(rdMolDescriptors.CalcFractionCSP3(mol))

    return descriptors

def plot_histograms(descriptors):
    fig, axs = plt.subplots(3, 2, figsize=(12, 12))
    axs = axs.flatten()

    for i, (key, values) in enumerate(descriptors.items()):
        axs[i].hist(values, bins=10, color='turquoise', edgecolor='black')
        axs[i].set_title(f'Histogram of {key}')
        axs[i].set_xlabel(key)
        axs[i].set_ylabel('Frequency')

    plt.tight_layout()
    plt.show()
```

2.9 Побудова хімічного простору для інгібіторів HDAC10

Хімічний простір – хемоінформатичний концепт, що описує простір властивостей молекули та простягається вздовж заданого набору дескрипторів та граничних умов.

Найпростіший хімічний простір є відрізком, наразі будуються багатовимірні простори, кожна сполука в яких описується набором векторів, кількість векторів визначає вимірність просторів.

Зрозумілою є необхідність в побудові не лише комп'ютерної моделі, а й візуалізації результатів дослідження. Візуалізація, наприклад, п'ятидесятивимірної моделі є непростою задачею, тому використовуються методи зниження вимірності.

В дослідженні було використано Python-бібліотеку ChemPlot, що є іноваційною в контексті використання в процесі кластеризації і розміщення сполук в хімічному просторі додаткового набору даних наданого користувачем, для визначення схожості між сполуками, детальніше метод було описано в дослідженні Basak et al [32]. Дана метрика викликається при побудові класу бібліотеки ключовим словом «tailored».

Код для побудови графіку:

```
from pandas import read_csv
from chemplot import Plotter
import matplotlib.pyplot as plt

data = read_csv(r"***filename**.smi", sep = '\t')
cp = Plotter.from_smiles(data["SMILES"])
cp.umap(random_state=228)
cp.cluster()
cp.interactive_plot(show_plot=True)
cp.visualize_plot(kind="hex")

plt.show()
```

Для зменшення вимірності було обрано метод UMAP (uniform manifold approximation and projection), що характеризується великою ефективністю обрахунку для великих масивів даних.

2.10 Статистичний аналіз значень медіан розподілів дескрипторних характеристик для об'єктів дослідження та підкласів

Статистичний аналіз на нормальність розподілу вибірки було проведено використовуючи тест Шапіро-Вілка, гомогенність розподілу варіантів було проаналізовано використовуючи тест Левена. Оскільки дані були

непараметричними, було використано попарне порівняння характеристик тестом Крускала-Волліса з корекцією Бонфероні.

Скрипт для проведення статистичної оцінки даних був написаний з використанням мови програмування Python v3.10 та модулю `scipy.stats`:

```
def check_assumptions(dataframes, properties):
    results = {}
    for target, df in zip(targets, dataframes):
        print(f"\nChecking assumptions for {target}")
        results[target] = {}
        for prop in properties:
            print(f"Analyzing {prop}:")
            data = df[prop].dropna() # Ensure no NaN values

            # Normality Test: Shapiro-Wilk
            shapiro_stat, shapiro_p = stats.shapiro(data)
            normality = shapiro_p > 0.05
            print(f" Normality (Shapiro-Wilk test): {'Pass' if normality else 'Fail'}, p-
value = {shapiro_p:.4f}")

            # Store the results
            results[target][prop] = {
                'Normality': 'Parametric' if normality else 'Nonparametric'
            }

        # Homogeneity of variances (Levene's test) across targets for each property
        for prop in properties:
            combined_data = [df[prop].dropna() for df in dataframes]
            levene_stat, levene_p = stats.levene(*combined_data)
            homogeneity = levene_p > 0.05
            print(
                f"\nHomogeneity of variances (Levene's test) for {prop}: {'Pass' if homogeneity
else 'Fail'}, p-value = {levene_p:.4f}")
            for target in targets:
                results[target][prop]['Homogeneity'] = 'Parametric' if homogeneity else
'Nonparametric'

    return results

# Run the checks
assumption_results = check_assumptions(dataframes, properties)

# Function to perform Kruskal-Wallis and pairwise comparisons
def nonparametric_tests(dataframes, properties):
    results = {}
    for prop in properties:
        print(f"\nAnalyzing {prop} across all targets:")
        # Collect data for this property from all dataframes
        data_by_target = [df[prop].dropna() for df in dataframes]
        # Kruskal-Wallis test
        kw_stat, kw_p = stats.kruskal(*data_by_target)
        print(f"Kruskal-Wallis test: H-statistic = {kw_stat}, p-value = {kw_p}")

        if kw_p < 0.05:
            # Perform pairwise comparisons if the Kruskal-Wallis test is significant
            print("Performing pairwise comparisons with Bonferroni correction:")
            num_comparisons = len(dataframes) * (len(dataframes) - 1) / 2
            alpha = 0.05 / num_comparisons # Bonferroni correction

            comparisons = []
            for i in range(len(dataframes)):
                for j in range(i + 1, len(dataframes)):
                    mw_stat, mw_p = stats.mannwhitneyu(dataframes[i][prop],
dataframes[j][prop],
```

```
        alternative='two-sided')
    comparisons.append((targets[i], targets[j], mw_stat, mw_p))
    print(
        f" {targets[i]} vs {targets[j]}: U-statistic = {mw_stat}, p-value =
{mw_p:.4f}, {'Significant' if mw_p < alpha else 'Not significant'}")

    results[prop] = {'Kruskal-Wallis': {'H-statistic': kw_stat, 'p-value': kw_p},
                    'Pairwise Comparisons': comparisons if kw_p < 0.05 else []}
return results
```

РОЗДІЛ 3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

3.1 Визначення доменної організації та структурної класифікації об'єктів дослідження

Доменну організацію представників родини HDAC для виду *Homo sapiens* було проаналізовано з метою встановлення клас-специфічних особливостей організації послідовностей для протеїнів. Аналіз було проведено з використанням профільних баз даних (GeneNames, InterPro) та їх он-лайн інструментів. Це дозволило визначити особливості класу II, зокрема, на відміну від класу I, представники якого переважно локалізовані в ядрі, клас II експресується як в ядрі клітини так і в цитоплазмі та залучений до сигналіngu.

Представники класу IIa – HDAC 4,5,7,9 характеризуються N-термінальним доменом, що містить мотив, що дозволяє MEF2 протеїнам зв'язуватися з білком. Ці протеїни є задіяними у декількох шляхах передачі сигналів, тому цей домен дозволяє HDAC класу IIa брати участь в сигнал-залежній регуляції ензиматичної активності. Відмінності в доменній організації між підкласом IIa та IIb можна спостерігати на рис. 3.1.

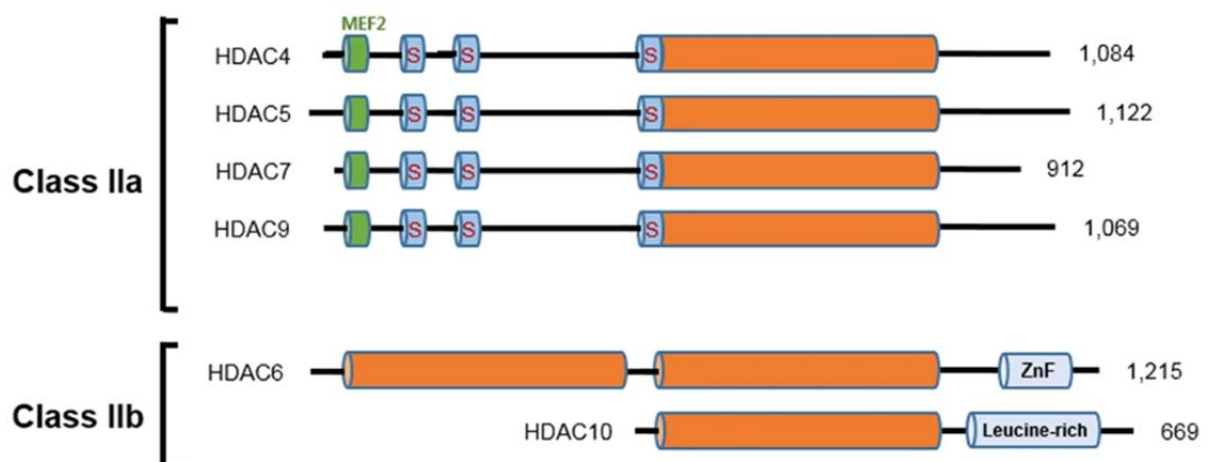


Рис. 3.1 Частина зображення з дослідження Suk-Yol et al., що показує структурні відмінності між класом IIa та IIb родини гістондеацетилаз [33].

Представники класу IIb – характеризуються двома каталітичними доменами, при цьому в HDAC10 домен лейцин-збагачений і каталітично неактивний, не взаємодіють з MEF2, маючи ширший спектр субстратів і негістонних цілей, зокрема тубулінів.

Оскільки послідовність каталітичного домену HDAC є висококонсервативною, важливим є враховувати не лише взаємодії ліганду з доменом, а й з загальною структурою білка.

Внаслідок проведеного аналізу структур та дослідження особливостей доменної архітектури для класу II родини HDAC було встановлено за необхідне використовувати загальну структуру для кожного з об'єктів дослідження. З огляду відсутності експериментально отриманої структури білку HDAC10, представнику підкласу IIb, було встановлено необхідність реконструкції просторової структури даного білку.

3.2 Реконструкція моделей протеїну HDAC10

Внаслідок аналізу наявних баз даних тривимірних структур протеїнів було визначено, що єдина модель отримана методом рентгенструктурного аналізу належить *Danio rerio*.

З використанням сервісу UniProtKB було проведено множинне вирівнювання послідовностей, завдяки результатам якого було встановлено ступінь ідентичності послідовностей. Було встановлено, що амінокислотна послідовність білку *H. sapiens* є ідентичною HDA10 *D.rerio* на 46%, з вищим ступенем ідентичності послідовностей (62%) в домені HDAC.

Також проведено аналіз ідентичності структур HDAC10 *Mus musculus* та HDAC10 *Rattus norvegicus*, що між собою показали найбільшу ідентичність в 88.1%. Найбільш ідентичною до людської була послідовність *M. musculus*, що є вказаним на рис. 3.2.

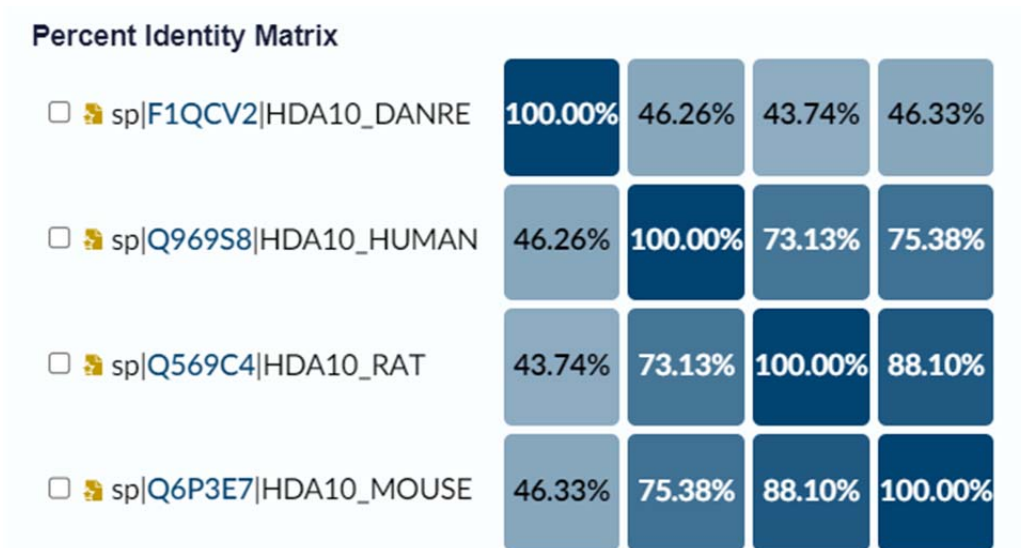


Рис 3.2 Матриця ідентичності послідовностей для гомологів HDAC10.

30% є достатнім порогом, для того, щоб стверджувати про те, що білки є еволюційно схожими та мають схожу просторову орієнтацію, але для впевненого моделювання за гомологією, для коректного представлення взаємодії лігандів з білком бажаною є ідентичність послідовностей принаймні в 75% [34].

З огляду на відсутність експериментально отриманих моделей просторової організації амінокислотного ланцюга протеїну HDAC10 людини та відсутніх достатньо близьких моделей інших видів була необхідність відтворити та побудувати теоретичну модель з використанням сервісів для моделювання Robetta та I-TASSER, візуалізовані моделі зображено на рис 3.3.

Послідовність, отриману з бази UniProt було завантажено в сервіси для моделювання, отримані моделі було оцінено за внутрішніми структурними параметрами сервісів.

Значення C-рейтингу для методу RoseTTAFold становили 0.83 (1.0 – найвище значення, 0.0 найнижче, вище значення – більша впевненість).

Для методу I-TASSER значення C-score для найкращої моделі становило -0.43 (-5.2 середнє значення, вищі значення – більша впевненість).

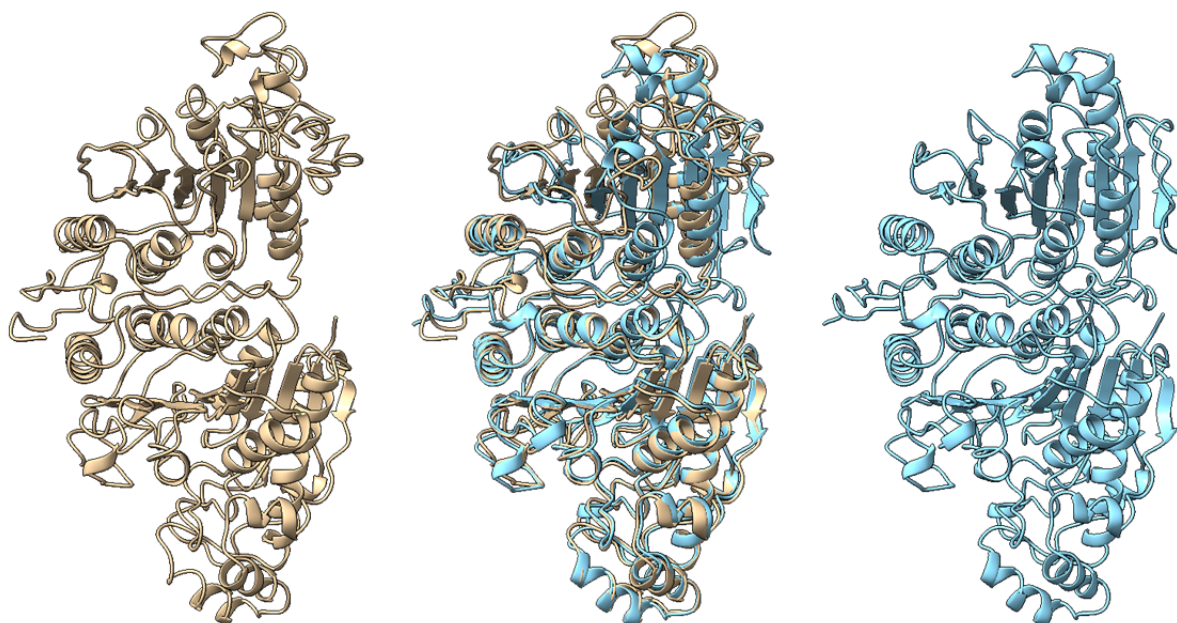


Рис 3.3 Результати молекулярного фітінгу отриманих структур I-TASSER (зліва) і Robetta (справа)

Відповідно структури було проаналізовано за критеріями порушення енергетично дозволених регіонів для ϕ та ψ дігедральних кутів вуглецевого скелета, що також є зображеними на карті Рамачандрана, а також середньоквадратичним відхиленням для кутів та довжин зв'язків (Табл.3.1).

Таблиця 3.1

Показники якості отриманих внаслідок моделювання структур з використанням сервісів Robetta та I-TASSER

Модель	Ramachandran Violations	RMS Bond Dev.	RMS Angle Dev.	Backbone	Sidechain
HDAC10_ITASSER_1	45	0.14	2.419	62	53
HDAC10_Robetta_2	72	0.19	4.064	90	81

Досліджувана модель отримана з використанням сервісу I-TASSER мала більш точне передбачення просторової орієнтації білку та мала на 27

порушень менше в метриці, що вимірює порушення дозволених регіонів двогранних кутів – 45 проти 72 в Robetta. Схожа картина спостерігалася для інших показників – середньоквадратичне відхилення в моделі I-TASSER було на 0.05 меншим для довжин зв'язків, а для кутів на 1.645, детальніше результати оцінки якості отриманих структур є представленими в таблиці 3.1.

Помітно, що I-TASSER є більш точним в передбаченні коректної структури HDAC10, зокрема. При цьому обидва методи дають доволі точні передбачення структури для висококонсервативного домену HDAC (до 300 амінокислот в послідовності), з послідуочим спаданням впевненості в передбаченні структури, як видно з рисунку 3.4.

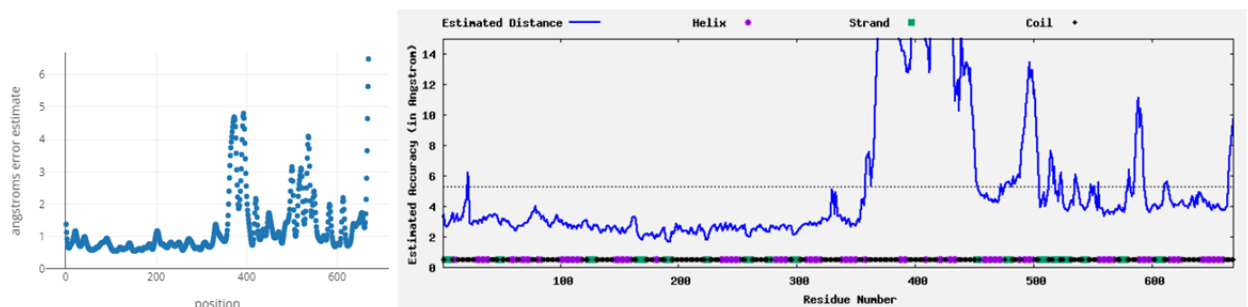


Рис. 3.4 Графіки точності передбачення правильної конформації, залежно від амінокислотного залишка для Robetta та I-TASSER відповідно.

Було встановлено, що промодель, отримана з використанням сервісу I-TASSER має більшу точність в контексті наданих вище показників, та зокрема метод показує високу впевненість в конформації структури активного центру, що міститься в межах 1-323 амінокислот з меншими значеннями RMSD для кутів та довжини зв'язків.

3.3 Оптимізація просторової структури білків

Окрім отриманої реконструкцією з використанням сервісів I-TASSER і Robetta структури HDAC10 для дослідження було також використано наступні структури протеїнів з бази даних PDB:

Назва протеїну	Код структури в базі PDB
HDAC4	2VQM
HDAC7	3C0Z
HDAC6	5EDU

Всі структури було оптимізовано за наступним алгоритмом:

1. Видалення оригінальних гідрогенів та додання нових воднів з використанням внутрішнього алгоритму;
2. Встановлення зв'язків нульового порядку з металами;
3. Встановлення дисульфідних зв'язків;
4. Кепування термінальних груп;
5. Генерація het станів;
6. Оптимізація водневих зв'язків (поле OPLS3e);
7. Мінімізація енергії протеїна з RMSD для важких атомів 0.3 Å та силовим полем OPLS3e;
8. Додатково використовувалася розширена оптимізація методом convergence gradient, молекулярно динамічне моделювання, зокрема з використанням алгоритму Монте-Карло.

Згідно з результатами, спостерігається висока якість структури для HDAC6, та не виявлено порушень за жодним з пунктів оцінювання.

Структура HDAC10 отримана теоретично, з методами моделювання протягуванням, за гомологією та з використанням машинного навчання. Додання йону цинку збільшило кількість порушень енергетично дозволених регіонів до 73. RMSD для довжин зв'язків HDAC10 знаходиться в діапазоні схожому до аналогічних значень білків, структури яких отримані експериментально (HDAC4,7), всі структури мають відхилення в межах 0.15-0.19, а RMSD для квадратичних кутів є вдвічі більшим для HDAC10 (4.065), порівняно з структурами HDAC4 (1.956) та HDAC7 (1.949), що можна спостерігати з даних наданих в таблиці 3.2.

За наявності більших обчислювальних потужностей було б гарно провести молекулярне моделювання взаємодії структури HDAC10 з розчинником для отримання стабільної нативної конформації. Проте в рамках даного дослідження це не мало ключової необхідності, з огляду на високу впевненість сервісів моделювання в побудові саме домену HDAC, точності отриманої моделі достатньо для передбачення взаємодій лігандів з активним центром молекули.

Таблиця 3.2

Визначення показників якості структури протеїна для об'єктів дослідження

Модель	Ramachandran Violations	RMS Bond Dev.	RMS Angle Dev.	Backbone	Sidechain
HDAC10_ITASSER_Zn	73	0.19	4.065	90	81
HDAC6	0	0	0	0	0
HDAC4_2VQM	15	0.15	1.956	9	8
HDAC7_3COZ	13	0.14	1.949	6	10

В результаті аналізу якості структурної оптимізації було отримано моделі структур білків HDAC4,6,7,10, що були використані в подальшому для побудови карт зв'язування з ліганду з поверхнею білка.

3.4 Визначення поверхонь сайтів зв'язування з білком та побудова карт сайтів зв'язування

Після отримання стабільних конформацій протеїнів необхідно було визначити поверхню зв'язування ліганда з білком. Це необхідно для побудови гріду - віртуальної області простору біля сайту зв'язування, що описує параметри рецептора та безпосередньо буде наповнюватися лігандами для оцінки їх потенційної взаємодії з рецептором та подальшим проведенням молекулярного докінгу в побудований грід.

Визначення активних центрів було проведено з використанням сервісу аналізу доменної архітектури SMART та інструментом SiteMap.

Важливою вимогою до сайту є можливість взаємодії лігандів з йоном цинку. Використаний інструмент дозволяє візуально визначати можливість такої взаємодії, оскільки обчислює поле, що описує зв'язок метал-ліганд та на рис. 3.5 зображене рожевим кольором.



Рис 3.5 Побудова сайтмепу на прикладі структури HDAC10.

Саме взаємодія з йоном цинку є типовою для білків, що взаємодіють з доменом HDAC та стоїть в принципі розробки інгібіторів родини. Відповідно до поставлених критеріїв було обрано сайти зв'язування для кожного з досліджуваних об'єктів обчислювані характеристики яких наведено в таблиці 3.3.

Таблиця 3.3

Параметри сайтів зв'язування для об'єктів дослідження

Блок	SiteScore	Size	Dscore	Volume
HDAC10	0.939	96	0.958	243.1
HDAC6	0.730	37	0.634	102.2
HDAC4	0.989	149	1.014	416
HDAC7	1.021	105	0.956	216

SiteScore в межах 0.8 вважається оптимальним для визначення різниці між селективними та неселективними сполуками, більше одиниці вважається гарним. В випадку HDAC6 значення є дещо меншим оптимального, але це пояснюється виходячи з рівняння, що описує SiteScore [35].

$$\text{SiteScore} = 0.0733 \sqrt{n} + 0.6688 e - 0.20 p$$

Де n – кількість точок сайту (Size), e – міра відкритості сайту до розчинника та p – міра гідрофільності сайту. Відповідно для меншого сайту значення SiteScore також буде менше. При цьому сайт HDAC6 незначно відхиляється від оптимального значення та продемонстрував можливість утворення зв'язку з йоном цинку.

D-score обраховується подібно до SiteScore, але вплив гідрофільності не обмежений максимальним значенням 1.0, що дає більш точне уявлення про можливість взаємодії лігандів з цією сполукою.

Показник Volume описує внутрішню поверхню сайту та об'єм простору, де ліганд може взаємодіяти з поверхнею білку. Спостерігається, що HDAC10 та HDAC7 мають подібні значення показників для всіх критеріїв оцінювання сайту зв'язування, зокрема SiteScore (0.939 та 1,01 відповідно), що може вказувати на доволі вдало побудовану модель структури HDAC10, що демонструє збіжність даних сайту зв'язування з

експериментально отриманою структурою HDAC7. Структура HDAC6 має меншу кількість точок де ліганд може взаємодіяти з активним центром – 37 проти 96,105 та 149 для інших об'єктів дослідження. Це пояснює менший показник SiteScore (0.73) та Dscore (0.634) та може бути асоційовано з отриманою більш закритою конформацією.

Структура HDAC4 має більший розмір (149) ідентифікованого сайту та відповідно високий показник Dscore (1.014). Це також може бути асоційовано з отриманою відповідно відкритою конформацією домену HDAC. Детально інформація стосовно характеристик сайтів зв'язування побудованих в рамках дослідження описана в таблиці 3.3.

Відповідно до результатів побудови та аналізу характеристик сайтів зв'язування було побудовано карти сайту зв'язування які в подальшому використовувалися для молекулярного докінгу лігандів до структур білків-об'єктів дослідження.

3.5 Підготовка бібліотек низькомолекулярних сполук до молекулярного докінгу

При молекулярному докінгу важливим є докування одразу декількох можливих форм молекули, адже різні конформації можуть мати різну енергію зв'язування та афінність до структури активного центру. Іншим важливим аспектом вимог до структур в випадку молекулярного докінгу є коректно згенеровані тривімірні координати та стабільна конформація молекули з мінімальною енергією.

Відповідно, також може виникати потреба в виправленні структури молекули, фільтрації за масою та іншими дескрипторними характеристиками. Для підготовки референсного сету ефекторів з встановленою дією на родину HDAC та сету з сполук для високопродуктивного віртуального скринінгу було використано інструмент LigPrep.

Сполуки отримані з баз даних було стандартизовано та мінімізовано їх енергію з використанням поля OPLS3e. Для іонізованих сполук було згенеровано можливі стани при $\text{pH } 7.0 \pm 2.0$. Також для сполук було видалено солі, згенеровано таутомери, вказані хіральності було збережено, інші центри стереоізомерії було використано для генерації до 32 комбінацій на ліганд.

Для референсного сету ефекторів з доведеною дією було згенеровано 37 155 сполук.

Для сету сполук для високопродуктивного скринінгу було згенеровано 1 176 261 молекул.

Внаслідок проведеного аналізу вибірок та їх підготовки до докінгу з використанням методу LigPrep було отримано дві бібліотеки – референсний сет ефекторів з доведеною активністю та HTS Stock Compounds.

3.6 Молекулярний докінг референтного сету ефекторів

Для об'єктів дослідження було проведено молекулярний докінг, де лігандами виступали структури з референсного сету з доведеною взаємодією з доменом HDAC та отримано дані про амінокислоти, які беруть участь в зв'язуванні лігандів в кожному з об'єктів дослідження.

Для HDAC6 (*H. sapiens*) Asp497(H-bond acceptor), His500(π - π -стекингова взаємодія), Ser568(H-bond acceptor), His651(π - π -стекингова взаємодія), Trp782(π - π -стекингова взаємодія).

В структурі білку HDAC10 (*H. sapiens*) Gly143 (H-bond acceptor), Phe144(π - π -стекингова взаємодія), Trp203(π - π -стекингова взаємодія, π -катіонна взаємодія), а також Glu272 (сольовий місток) були залучені до утворення зв'язків з інгібітором.

Амінокислоти HDAC4 (*H. sapiens*) Hid158 (H-bond acceptor), Phe168 (π - π -стекингова взаємодія), Gly167(H-bond acceptor, H-bond donor), Phe227(H-bond acceptor, H-bond donor, π - π -стекингова взаємодія) було ідентифіковано, як ті що утворюють зв'язки з лігандами.

Внаслідок дослідження амінокислот, що взаємодіють з лігандами в білку HDAC7 (*H. sapiens*) було ідентифіковано Asp626(H-bond acceptor), His669(H-bond donor), His670(H-bond donor), Gly678(H-bond acceptor), Phe679(π - π -стекингова взаємодія), His709(π - π -стекингова взаємодія, π -катіон), Pro809(H-bond acceptor).

Помітно, що амінокислоти Gly, Phe є висококонсервативними в сайті та спостерігаються в HDAC4,7,10.

Результати докінгу показують високу кількість лігандів з низьким показником рейтингу докінгу (Docking Score)

Таблиця 3.4

Результати молекулярного докінгу представників II класу з референтним сетом ефекторів з доведеною взаємодією

Білок	DockingScore max.	Ebind. max	Кількість сполук з DS > 7	Ligand_efficiency
HDAC10	-9.37	49.52	688	-0.27
HDAC6	-10.19	43.96	4822	-0.22
HDAC4	-10.75	74.76	8414	-0.15
HDAC7	-9.93	34.84	4665	-0.50

Відповідно до результатів докінгу, спостерігається висока афінність структур до великої кількості проаналізованих лігандів. Згідно з Halgren et al. для гідрофобного типу взаємодії значення -8 є дуже хорошим показником [36], відповідно при значеннях -9-(-10) спостерігається доволі висока афінність лігандів до протеїнів, що підтверджується також тим, що цей сет було зібрано як селективний до домену HDAC. Найнижчий показник DockingScore було встановлено в HDAC4 -10.75, як і найбільшу кількість сполук, що проявили афінність - 8414. Найменша кількість сполук була встановлена для HDAC10 – 688, для HDAC6 – 4822, для HDAC7 – 4665.

Разом з цим для всіх з досліджених ферментів було виявлено головний механізм інгібування – зв’язування хелатуючої групи на гідрофобному лінкері з йоном цинку.

На рисунку 3.6 зображено візуалізовані зв’язки з амінокислотами активного центру та йоном цинку для сполуки CHEMBL4793145, що проявила найбільшу афінність до білку HDAC10. Зв’язки було передбачено та побудовано з допомогою ChimeraX. Також побудовано поверхню білка, з прогнозованою ліпофільністю за шкалою Кайла-Дулітла [37].

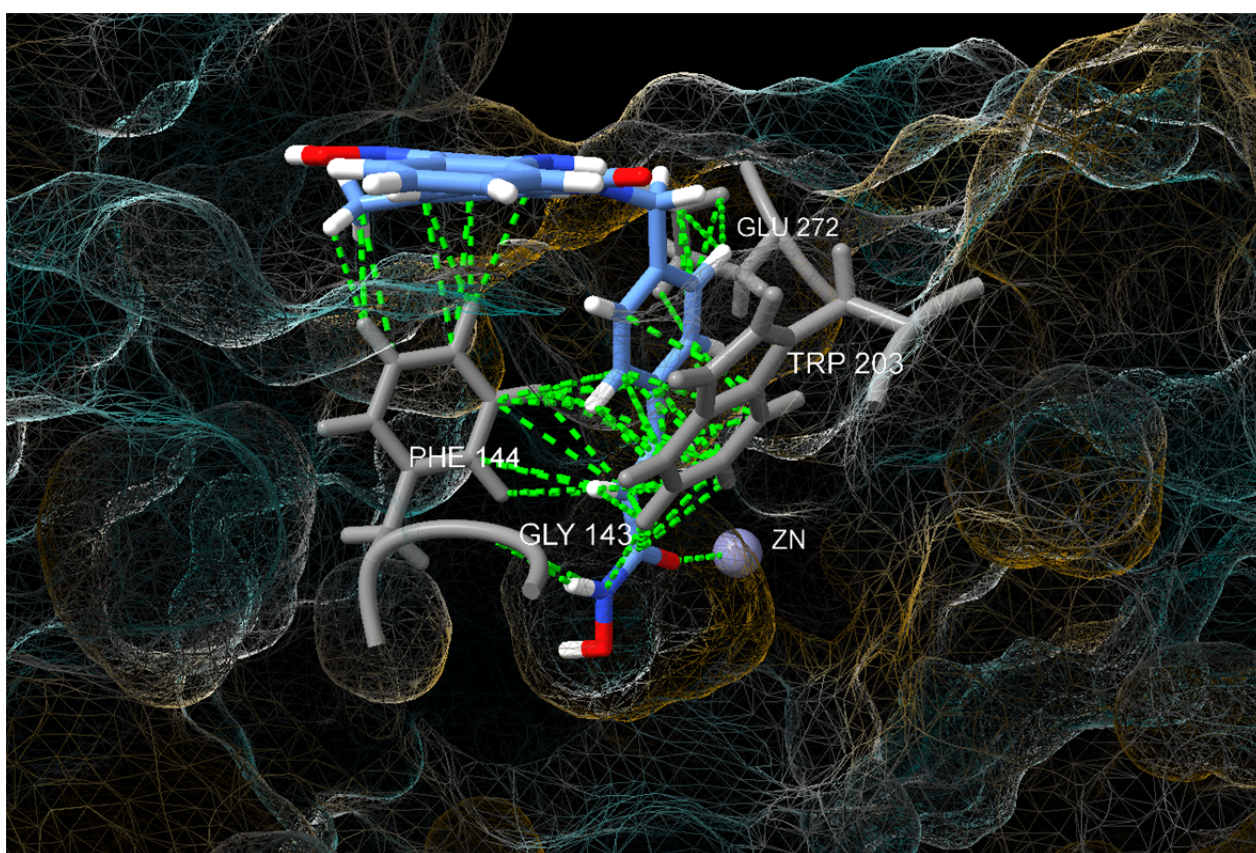


Рис. 3.6 Механізм взаємодії ефектора CHEMBL4793145 в активному центр білка HDAC10.

За допомогою цієї візуалізації можна простежити структуру гідрофобного каналу та взаємодію хелатуючої групи гідроксамової кислоти з йоном цинку. Таким чином виконавши докінг цільових структур з

референтним сетом було проведено апробацію змодельованої структури HDAC10 та реконструйованих методом PCA структур HDAC4, 6, 7.

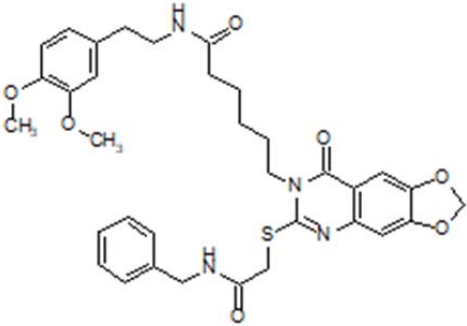
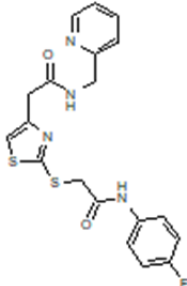
3.7 Пошук потенційних ефекторів за допомогою молекулярного докінгу загальної хімічної бібліотеки

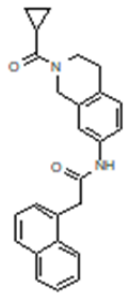
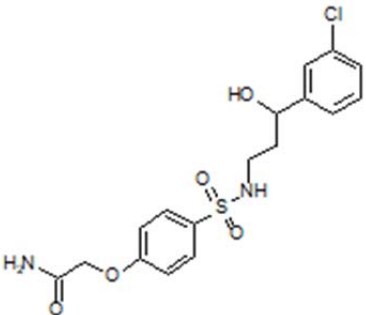
Використовуючи результати дослідження референтного сету ефекторів з встановленою інгібуючою дією, було проведено пошук потенційних селективних до підкласу або до конкретної мішені лігандів.

Виконавши молекулярний докінг та проаналізувавши його результати було отримано список сполук, що проявляють високу селективність (< -8 для обраного білку та > -8 для інших білків-об'єктів дослідження)

Таблиця 3.5

Сполуки з найвищим DockingScore та селективністю до вказаного білку
(продовження таблиці на стор. 42)

Білок	ID сполуки	DockingScore	Зоображення
HDAC4	F3168-2340	-10.470	
HDAC7	F2335-0642	-10.302	

HDAC6	F2416-0766	-9.752	
HDAC10	F6416-9301	-8.329	

Відповідно до результатів докінгу для білків HDAC4 та HDAC7 спостерігалася більша кількість сполук, що проявляють високу афінність – 27337 та 10465 відповідно. Це пояснюється, зокрема, більшим розміром сайту, що дозволяє більшій кількості реактивних груп взаємодіяти не лише з структурою активного центру, а й з поверхнею білка навколо неї. Таким чином спостерігалася не лише взаємодія з йоном цинку, а й з амінокислотами що утворюють гідрофобний тунель в активному центрі. При цьому гарний DockingScore вказує на найнижчу енергію зв'язування, та, відповідно, високу афінність ліганду.

Разом з цим, як зазначено в таблиці 3.6, лише невелика частина сполук проявила селективність в зв'язуванні до представників класу IIb - HDAC6 та HDAC10 - 1996 та 9 відповідно. Всі об'єкти дослідження мають гарні значення DockingScore – нижче -8.

За результатами докінгу було відібрано 6 сполук з найнижчим значенням DockingScore для кожного з об'єктів дослідження та підкласу. Першу з них наведено в таблиці 3.6, а детальні результати дослідження знаходяться в додатку 1.

Таблиця 3.6

DockingScore та кількість сполук з високою афінністю за результатами докінгу сету HTS Stock Compounds

Білок	DockingScore max.	Кількість сполук з DS > 8
HDAC10	-8.572	9
HDAC6	-9.752	1996
HDAC4	-10.75	27337
HDAC7	-10.78	10465

3.8 Статистичний аналіз з метою пошуку підклас-специфічних характеристик лігандів

Було проаналізовано 6 дескрипторних характеристик, які вважаються ключовими в контексті пошуку нових лікарських сполук та використовуються в емпіричних правилах Lipinski's RO5 та RO3 (Табл.3.6).

Ними є молекулярна вага, кількість донорів й акцепторів водню, кількість sp³-гібридизованих атомів вуглецю (fsp³), площа топологічної полярної поверхні (TPSA) та логарифм коефіцієнту розчинності сполуки в октанолі в співвідношенні до розчинності в воді (clogP). Головним завданням було проаналізувавши статистичні дані, зробити висновки про спільність або відмінність дескрипторних характеристик між класами.

Сполуки з найнижчим DockingScore відібрані для кожного з об'єктів дослідження та підкласів.

Для донорів водневого зв'язку статистично значна відмінність була знайдена для всіх об'єктів дослідження, окрім попарного тесту Манна-Уїтні для HDAC4 і HDAC6 ($p = 0.407$).

Для акцепторів водневого зв'язку значення для HDAC6 також не відрізнялася від значень представників класу IIa ($p = 0.499$, 0.017 для HDAC4

та HDAC7 відповідно), при цьому статистична відмінність була знайдена для HDAC10.

Таблиця 3.6

IDNUMBER	docking_score	Target	SMILES
F8881-8238	-8.37493	HDAC10	<chem>CC(N)Cc1ccc(O)cc1O</chem>
F3168-2340	-10.4698	HDAC4	<chem>COc1ccc(CCNC(=O)CCCCn2c(SCC(=O)NCc3ccccc3)nc3cc4c(cc3c2=O)OCO4)cc1OC</chem>
F2416-0766	-9.75213	HDAC6	<chem>O=C(Cc1cccc2ccccc12)Nc1ccc2c(c1)CN(C(=O)C1CC1)CC2</chem>
F2335-0642	-10.3024	HDAC7	<chem>O=C(Cc1esc(SCC(=O)Nc2ccc(F)cc2)n1)NCc1ccccc1</chem>
F6524-1480	-9.88	IIa	<chem>O=C(Cn1ccc2ccccc21)NCc1nc(-c2ccc[nH]c2=O)no1</chem>
F6506-1999	-8.19	IIb	<chem>NC(=O)COc1ccc(S(=O)(=O)NC(O)(c2ccoc2)c2cccs2)cc1</chem>

Для молекулярної ваги та cLogP та TPSA вибірки всіх протеїнів мали статистично достовірну різницю. ($p < 0.05$)

Найважливішим спостереженням є відмінності в розподілі FSP3, де HDAC4,6,10 не мали статистично значимої відмінності між собою, на противагу HDAC7, що мала достовірну відмінність від наведених вище білків ($p = 0.000$).

Внаслідок статистичного аналізу було встановлено, що об'єкти дослідження мають статистично значущу різницю в медіанах значень дескрипторів для всіх характеристик, окрім fsp3. При цьому в декількох тестах медіани значень для HDAC4 були схожими з HDAC6.

В контексті міжкласових відмінностей аналіз показав статистично достовірні ($p < 0.05$) менші значення медіани для дескрипторних характеристик донорів водневого зв'язку, FSP3 та cLogP для HDAC підкласу

Пb порівняно з підкласом Па. Тобто сполуки селективні до підкласу Пb в середньому мають більшу TPSA та більшу молекулярну вагу, з більшою кількістю акцепторів водню.

Детальні зображення розподілів параметрів та результати статистичного аналізу вказані в додатку 2.

В результаті аналізу отриманих статистичних даних про відмінності в медіанах розподілу дескрипторних характеристик лігандів, ідентифікованих при проведенні молекулярного докінгу HTS Stock Compounds було встановлено достовірно менші медіанні значення донорів водневого зв'язку, FSP3, cLogP для представників гістондеацетилаз класу Пb порівняно з представниками Па.

3.9 Побудова хімічного простору

Для побудови хімічного простору використовували ліганди з високою афінністю. Відсічною точкою були значення >7 або, у випадку коли ефекторів з таким значенням було багато, 10 000 сполук (1% від загальної вибірки).

Для збільшення якості візуалізації даних було обрано значення $\text{min_dist} = 0.9$ при зменшенні розмірності. Значення ключа що використовувався методом зменшення розмірності було встановлено на 322.

В результаті було отримано зображення хімічного простору для всіх об'єктів дослідження.

Аналізуючи хімічний простір класу Па було зроблено висновок, що головна маса сполук, що утворюють великий кластер в центрі хімічного простору для HDAC7 в більшості складається з сполук, які утворюють два цикли або гетероцикли, з'єднані між собою амідним зв'язком. Для HDAC4 більшість основного кластеру є збіжною з HDAC7, але при цьому є більшою частка сполук, що містять три цикли (Рис.3.7).

Зрозумілим є те, що при аналізі масивних кластерів важко отримати єдину характеристику дизайну селективного інгібітора, але з використанням аналізу хімічного простору є можливість визначити тенденції, яким необхідно слідувати в дизайні тих чи інших інгібіторів. Разом з цим використання аналізу хімічного простору дозволяє виявити скаффолди - сполуки на основі яких утворюються віддалені кластери і таким чином виявити потенційні хімічні структури, що можуть виступати як ефектори з високою афінністю і унікальними фізико-хімічними параметрами.

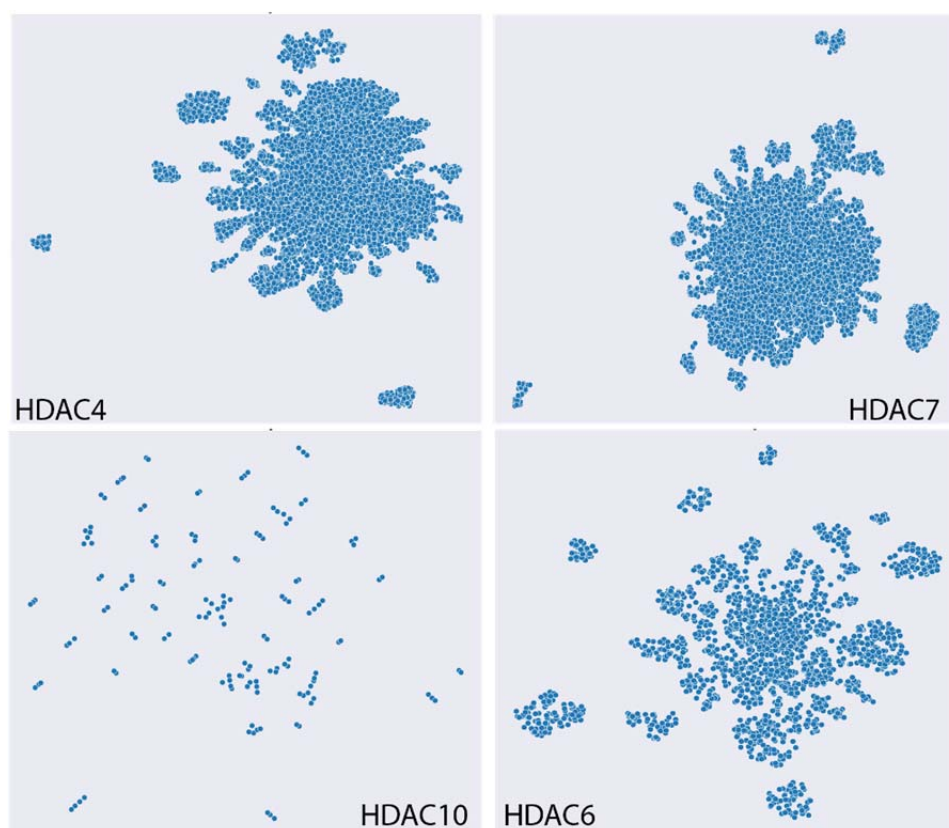


Рис 3.7 Результати візуалізації хімічного простору лігандів для HDAC4-10.

Таким чином на прикладі хімічного простору HDAC4 що містить 3 віддалених кластери, було ідентифіковано потенційні скаффолди для ефекторів, зображені на рис 3.8.

За результатами генерування хімічного простору було утворено інтерактивні моделі в форматі *.html які можна використовувати для аналізу інгібіторів HDAC II класу. Один з прикладів хімічного простору з репрезентативними сполуками для кожного кластеру наведено на рис. 3.9.

При побудові хімічного простору з бібліотекою ChemPlot ми також можемо додати ще один вимір для побудови простору, в цьому випадку враховуються не лише структурні особливості сполук та дескрипторні характеристики, але й дані з масиву який передається спільно з структурами. В нашому випадку це DockingScore.

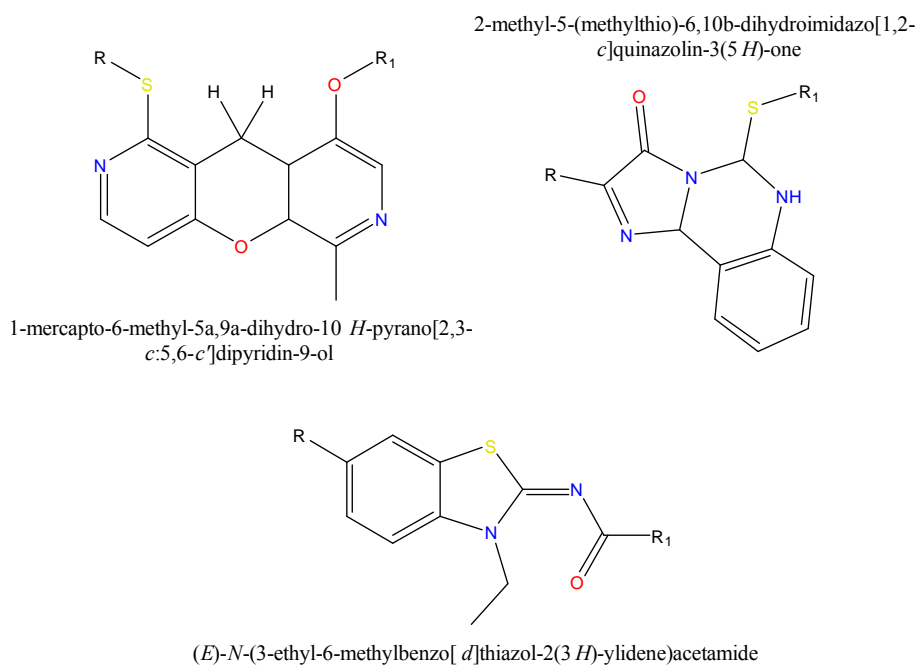


Рис 3.8 Скафолди сполук з віддалених кластерів, отримані внаслідок аналізу хімічного простору HDAC4

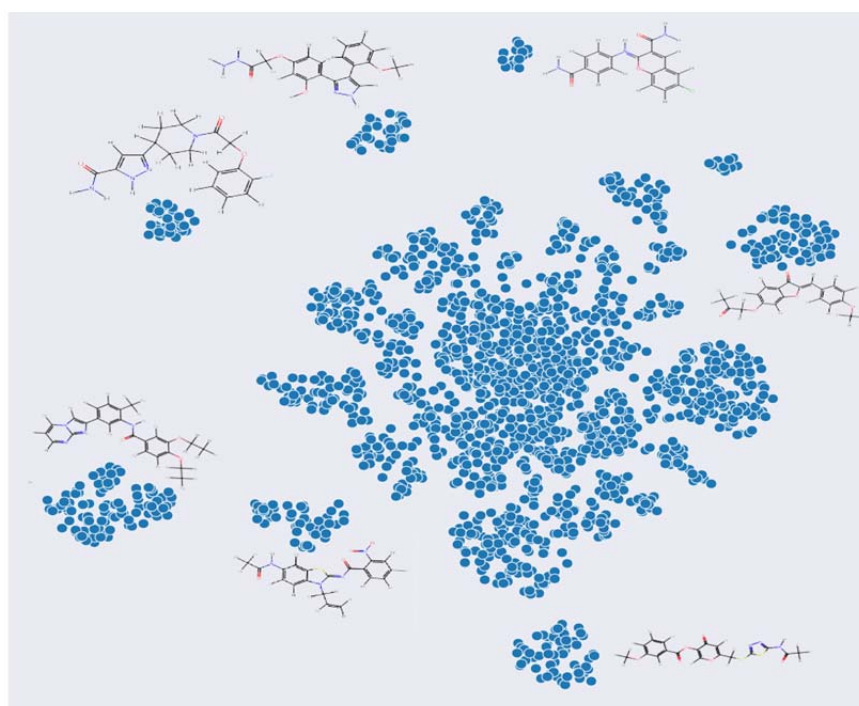


Рис. 3.9 Візуалізація хімічного простору ефекторів HDAC6 з представленням репрезентативних сполук для кожного з кластерів.

Хімічний простір з використанням DockingScore було побудовано використовуючи бібліотеку ChemPlot з використанням методу target, до якого було передано масив даних про афінність лігандів. За результатами побудови можна ідентифікувати кластери сполук, що мають високу афінність та є структурно схожими. На прикладі хімічного простору ефекторів HDAC4 (рис. 3.10) розглянемо кластери інгібіторів з високою активністю, зображенні темно-фіолетовим та чорним кольорами.

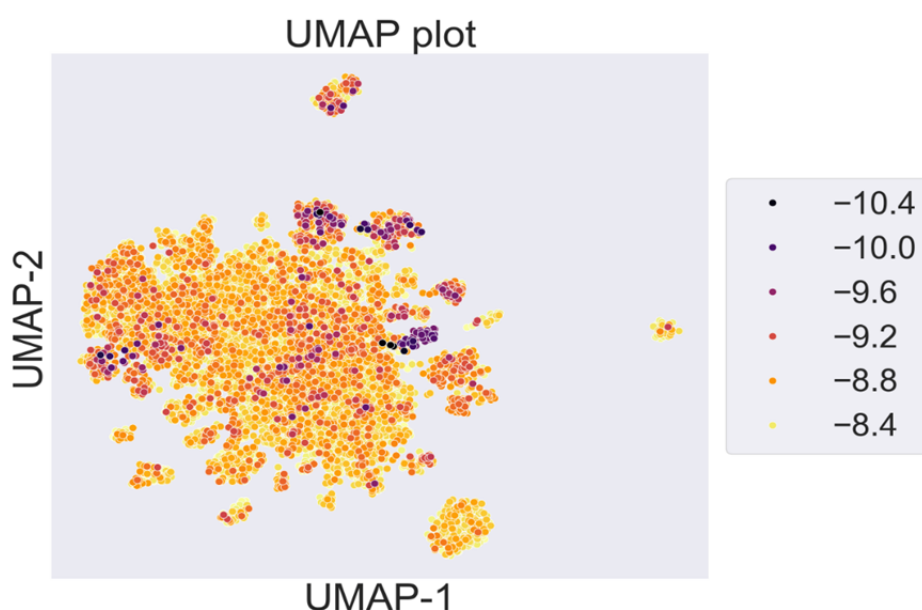


Рис. 3.10 Візуалізація хімічного простору ефекторів HDAC4 з врахуванням показника DockingScore.

Як ми бачимо структури утворюють 4 чітких кластери, але наявні і сполуки що розташовані по більшій площі хімічного простору. Проаналізуємо кластери.

За результатами дослідження кластерів було обрано репрезентативні сполуки з рисунку 3.11.

Вони відповідають спільній фармакофорній гіпотезі, оскільки мають групу що взаємодіятиме з цинком, лінкер, та циклічні молекули що

виступають в якості кепа та утворюють зв'язок з зовнішньою частиною активного центра, росташованій на поверхні білка.

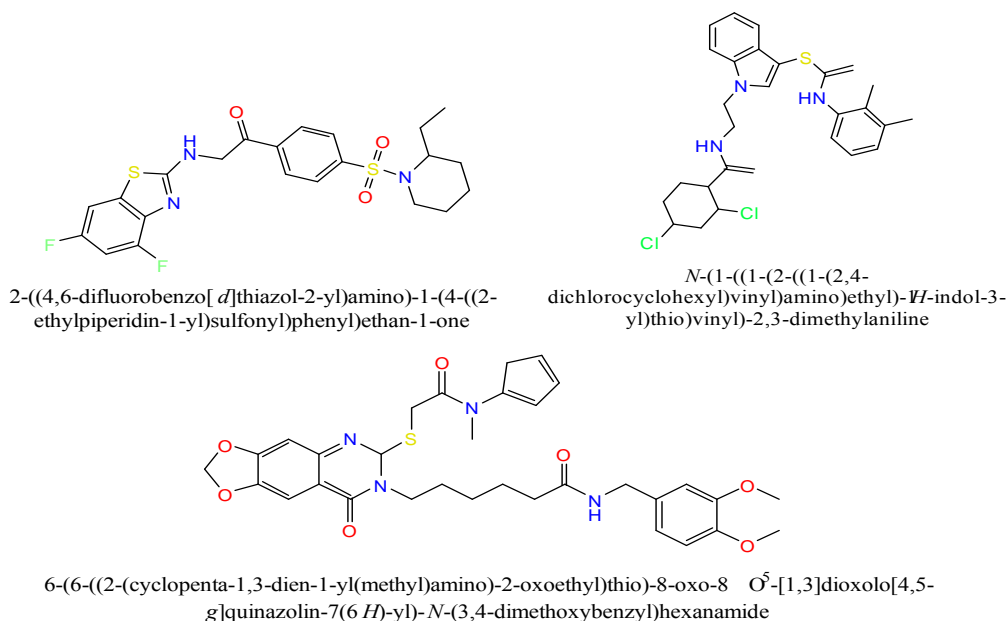


Рис. 3.11 Репрезентативні сполуки отримані внаслідок аналізу кластерів ефекторів з високою афінністю до HDAC4.

Відповідно до цього дослідження важливим аспектом є проведення генерації хімічного простору з ідентифікацією кластерів що мають спільний скафолд та мають унікальні фізико-хімічні властивості або є структурно схожими і утворюють кластери на основі даних про величину афінності ліганда до цільової структури активного центру протеїну. Отримані внаслідок цього дослідження результати можуть бути використані як набір тренувальних даних та для аналізу при побудові моделей машинного навчання, що зможуть утворювати вибірки хімічно схожих сполук та досліджувати їх на предмет селективного інгібування обраної мішені. Таким чином ми маємо змогу швидко та з великою впевненістю ідентифікувати високоселективні терапевтичні таргети, які будуть мати різноманітні фізико-хімічні параметри, що є важливим при клінічних випробуваннях.

В результаті побудови хімічного простору для ідентифікованих потенційних ефекторів білків-об'єктів дослідження, було отримано карти хімічного простору, які потенційно можуть використовуватися для

ідентифікації терапевтичних агентів та генерації хімічного простору на основі кластерів сполук з високою афінністю, методами машинного навчання.

ВИСНОВКИ

1. Реконструйовано та верифіковано просторову структуру гістондеацетилази типу 10, включно з реконструкцією будови каталітичного домена на основі комбінації методів молекулярного моделювання.

2. Охарактеризовано механізм взаємодії ефекторів в активних сайтах гістондеацетилаз HDAC10, HDAC6, HDAC4, HDAC7 та встановлено консервативний склад сайтів зв'язування з лігандами.

3. Проведено пошук ефекторів для представників класу II родини HDAC використовуючи метод молекулярного докінгу та встановлено сполуки з високою афінністю до гістондеацетилаз HDAC10, HDAC6, HDAC4, HDAC7. Зокрема, ідентифіковано 36 сполук з високою селективністю до об'єктів дослідження.

4. Обчислено визначені дескрипторні характеристики молекул та використовуючи методи статистичного аналізу визначено статистично достовірні різниці медіан з метою пошуку клас-специфічних характеристик, що свідчать про достовірно менші медіанні значення донорів водневого зв'язку, FSP3, cLogP для представників гістондеацетилаз класу IIb порівняно з представниками IIa.

5. Використовуючи інструмент ChemPlot, було побудовано та методом зменшення розмірності UMAP візуалізовано інтерактивні моделі хімічного простору ефекторів для гістондеацетилаз HDAC10, HDAC6, HDAC4, HDAC7, проаналізовано клас-специфічні характеристики та ідентифіковано характерні скеффолди, що запропоновані для подальшої лід-оптимізації при хімічному синтезі кінцевих сполук.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] P. A. Wade, “Transcriptional control at regulatory checkpoints by histone deacetylases: Molecular connections between cancer and chromatin,” *Human Molecular Genetics*, vol. 10, no. 7. 2001, doi: 10.1093/hmg/10.7.693.
- [2] E. Seto and M. Yoshida, “Erasers of histone acetylation: The histone deacetylase enzymes,” *Cold Spring Harb. Perspect. Biol.*, vol. 6, no. 4, 2014, doi: 10.1101/cshperspect.a018713.
- [3] A. J. M. De Ruijter, A. H. Van Gennip, H. N. Caron, S. Kemp, and A. B. P. Van Kuilenburg, “Histone deacetylases (HDACs): Characterization of the classical HDAC family,” *Biochemical Journal*, vol. 370, no. 3. 2003, doi: 10.1042/BJ20021321.
- [4] C. Choudhary *et al.*, “Lysine acetylation targets protein complexes and co-regulates major cellular functions,” *Science (80-.)*, vol. 325, no. 5942, 2009, doi: 10.1126/science.1175371.
- [5] L. J. Juan *et al.*, “Histone deacetylases specifically down-regulate p53-dependent gene activation,” *J. Biol. Chem.*, vol. 275, no. 27, 2000, doi: 10.1074/jbc.M000202200.
- [6] D. D. Leipe and D. Landsman, “Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily,” *Nucleic Acids Res.*, vol. 25, no. 18, 1997, doi: 10.1093/nar/25.18.3693.
- [7] C. J. Millard *et al.*, “Class I HDACs share a common mechanism of regulation by inositol phosphates,” *Mol. Cell*, vol. 51, no. 1, 2013, doi: 10.1016/j.molcel.2013.05.020.

[8] Y.-L. Yao, W.-M. Yang, and E. Seto, "Regulation of Transcription Factor YY1 by Acetylation and Deacetylation," *Mol. Cell. Biol.*, vol. 21, no. 17, 2001, doi: 10.1128/mcb.21.17.5979-5991.2001.

[9] C. Moreno-Yruela *et al.*, "Class I histone deacetylases (HDAC1-3) are histone lysine deacetylases," *Sci. Adv.*, vol. 8, no. 3, 2022, doi: 10.1126/sciadv.abi6696.

[10] S. C. Galasinski, K. A. Resing, J. A. Goodrich, and N. G. Ahn, "Phosphatase inhibition leads to histone deacetylases 1 and 2 phosphorylation and disruption of corepressor interactions," *J. Biol. Chem.*, vol. 277, no. 22, 2002, doi: 10.1074/jbc.M201174200.

[11] M. A. Glozak and E. Seto, "Acetylation/deacetylation modulates the stability of DNA replication licensing factor Cdt1," *J. Biol. Chem.*, vol. 284, no. 17, 2009, doi: 10.1074/jbc.M809394200.

[12] C. B. Brachmann, J. M. Sherman, S. E. Devine, E. E. Cameron, L. Pillus, and J. D. Boeke, "The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability," *Genes Dev.*, vol. 9, no. 23, 1995, doi: 10.1101/gad.9.23.2888.

[13] A. D. Bondarev, M. M. Attwood, J. Jonsson, V. N. Chubarev, V. V. Tarasov, and H. B. Schiöth, "Recent developments of HDAC inhibitors: Emerging indications and novel molecules," *British Journal of Clinical Pharmacology*, vol. 87, no. 12, 2021, doi: 10.1111/bcp.14889.

[14] O. M. Dovey *et al.*, "Histone deacetylase 1 and 2 are essential for normal T-cell development and genomic stability in mice," *Blood*, vol. 121, no. 8, 2013, doi: 10.1182/blood-2012-07-441949.

[15] F. Santoro *et al.*, "A dual role for Hdac1: oncosuppressor in tumorigenesis, oncogene in tumor maintenance," *Blood*, vol. 121, no. 17, 2013, doi: 10.1182/blood-2012-10-461988.

- [16] G. Jin, K. Wang, Y. Zhao, S. Yuan, Z. He, and J. Zhang, "Targeting histone deacetylases for heart diseases," *Bioorganic Chemistry*, vol. 138. 2023, doi: 10.1016/j.bioorg.2023.106601.
- [17] C. L. Zhang, T. A. McKinsey, S. Chang, C. L. Antos, J. A. Hill, and E. N. Olson, "Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy," *Cell*, vol. 110, no. 4, 2002, doi: 10.1016/S0092-8674(02)00861-9.
- [18] Y. Chen *et al.*, "Development of Purine-Based Hydroxamic Acid Derivatives: Potent Histone Deacetylase Inhibitors with Marked in Vitro and in Vivo Antitumor Activities," *J. Med. Chem.*, vol. 59, no. 11, 2016, doi: 10.1021/acs.jmedchem.6b00579.
- [19] F. Chen *et al.*, "Potent and orally efficacious bithiazole-based histone deacetylase inhibitors," *ACS Med. Chem. Lett.*, vol. 5, no. 6, 2014, doi: 10.1021/ml400470s.
- [20] Y. bo Zhou *et al.*, "Pharmacodynamic, pharmacokinetic, and phase 1a study of bithianostat, a novel histone deacetylase inhibitor, for the treatment of relapsed or refractory multiple myeloma," *Acta Pharmacol. Sin.*, vol. 43, no. 4, 2022, doi: 10.1038/s41401-021-00728-y.
- [21] S. Rai *et al.*, "Oral HDAC inhibitor tucidinostat in patients with relapsed or refractory peripheral T-cell lymphoma: phase IIb results," *Haematologica*, vol. 108, no. 3, 2023, doi: 10.3324/haematol.2022.280996.
- [22] J. Chang *et al.*, "High-throughput screening identifies two novel small molecule enhancers of recombinant protein expression," *Molecules*, vol. 25, no. 2, 2020, doi: 10.3390/molecules25020353.
- [23] J. Liu *et al.*, "Selective Class i HDAC Inhibitors Based on Aryl Ketone Zinc Binding Induce HIV-1 Protein for Clearance," *ACS Med. Chem. Lett.*, vol. 11, no. 7, 2020, doi: 10.1021/acsmchemlett.0c00302.

[24] F. Cao, M. R. H. Zwinderman, and F. J. Dekker, “The process and strategy for developing selective histone deacetylase 3 inhibitors,” *Molecules*, vol. 23, no. 3. 2018, doi: 10.3390/molecules23030551.

[25] C. A. Luckhurst *et al.*, “Potent, Selective, and CNS-Penetrant Tetrasubstituted Cyclopropane Class Iia Histone Deacetylase (HDAC) Inhibitors,” *ACS Med. Chem. Lett.*, vol. 7, no. 1, 2016, doi: 10.1021/acsmchemlett.5b00302.

[26] Z. Kutil, Z. Novakova, M. Meleshin, J. Mikesova, M. Schutkowski, and C. Barinka, “Histone Deacetylase 11 Is a Fatty-Acid Deacylase,” *ACS Chem. Biol.*, vol. 13, no. 3, 2018, doi: 10.1021/acscchembio.7b00942.

[27] R. B. Mokhtari *et al.*, “Combination therapy in combating cancer,” *Oncotarget*, vol. 8, no. 23. 2017, doi: 10.18632/oncotarget.16723.

[28] Y. Zhang, “I-TASSER server for protein 3D structure prediction,” *BMC Bioinformatics*, vol. 9, 2008, doi: 10.1186/1471-2105-9-40.

[29] A. Roy, A. Kucukural, and Y. Zhang, “I-TASSER: A unified platform for automated protein structure and function prediction,” *Nat. Protoc.*, vol. 5, no. 4, 2010, doi: 10.1038/nprot.2010.5.

[30] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, “The I-TASSER suite: Protein structure and function prediction,” *Nature Methods*, vol. 12, no. 1. 2014, doi: 10.1038/nmeth.3213.

[31] J. R. Shewchuk, “An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. {E}dition $\frac{1}{4}$ (<http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf>),” *Science (80-.)*, vol. 0, no. 2, 1994.

[32] S. C. Basak and G. D. Grunwald, “Predicting mutagenicity of chemicals using topological and quantum chemical parameters: A similarity based study,” *Chemosphere*, vol. 31, no. 1, 1995, doi: 10.1016/0045-6535(95)00122-O.

[33] S. Y. Park and J. S. Kim, “A short guide to histone deacetylases including recent progress on class II enzymes,” *Experimental and Molecular Medicine*, vol. 52, no. 2, 2020, doi: 10.1038/s12276-020-0382-4.

[34] E. Lindahl and A. Elofsson, “Identification of related proteins on family, superfamily and fold level,” *J. Mol. Biol.*, vol. 295, no. 3, 2000, doi: 10.1006/jmbi.1999.3377.

[35] T. A. Halgren, “Identifying and characterizing binding sites and assessing druggability,” *J. Chem. Inf. Model.*, vol. 49, no. 2, 2009, doi: 10.1021/ci800324m.

[36] T. A. Halgren *et al.*, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening,” *J. Med. Chem.*, vol. 47, no. 7, 2004, doi: 10.1021/jm030644s.

[37] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydropathic character of a protein,” *J. Mol. Biol.*, vol. 157, no. 1, 1982, doi: 10.1016/0022-2836(82)90515-0.

ДОДАТКИ

Додаток 1

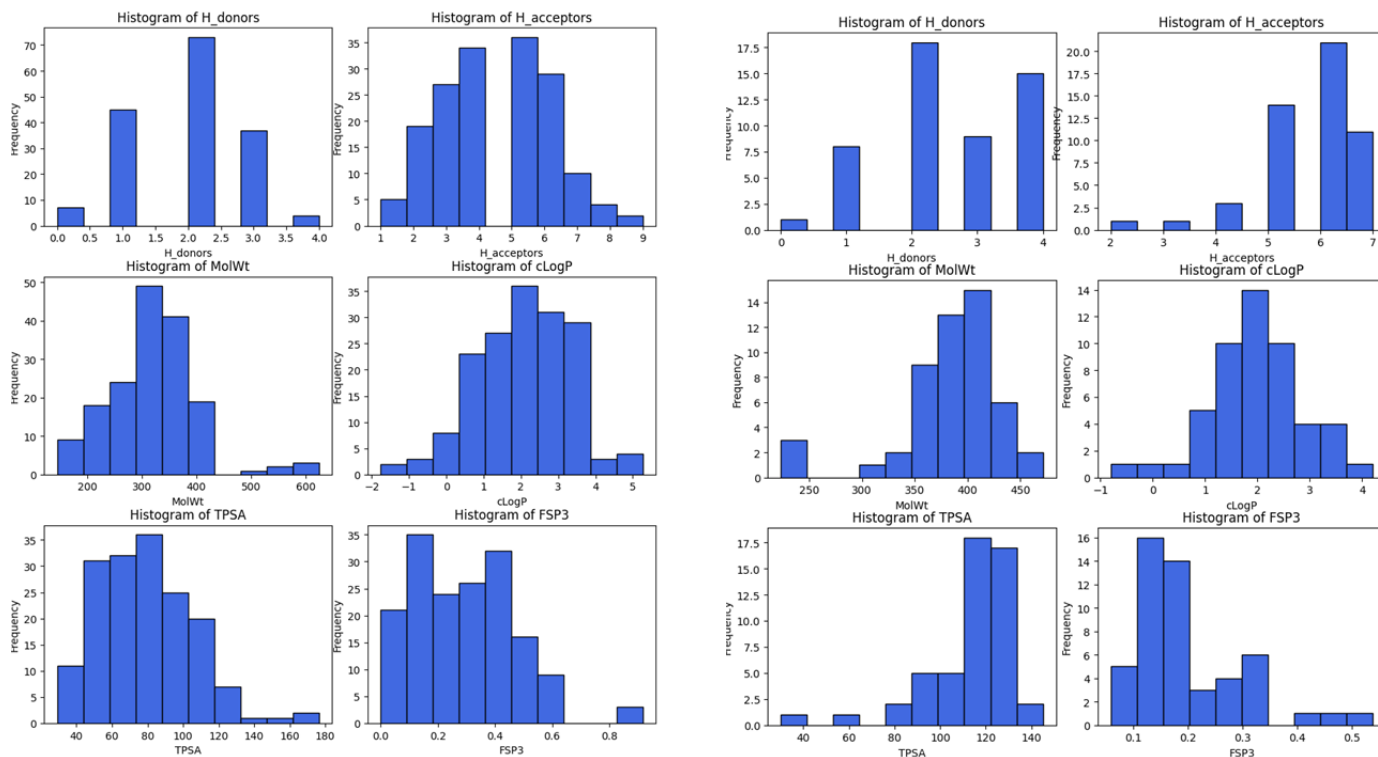
Відібрані внаслідок молекулярного докінгу бібліотеки HTS Stock
Compounds ліганди з високою селективністю

IDNUMBER	docking_score	Target	SMILES
F8881-8238	-8.37493	HDAC10	<chem>CC(N)Cc1ccc(O)cc1O</chem>
F6416-9301	-8.32913	HDAC10	<chem>NC(=O)COc1ccc(S(=O)(=O)NCCC(O)c2cccc(Cl)c2)cc1</chem>
F3406-6059	-8.13747	HDAC10	<chem>NC(=O)CSc1enn(CC(=O)NCc2ccc3c(c2)OCO3)c(=O)c1Cl</chem>
F2199-0237	-8.02451	HDAC10	<chem>CC(=O)N1CCN(S(=O)(=O)c2ccc(OCC(N)=O)cc2)CC1</chem>
F5331-0054	-7.81284	HDAC10	<chem>NC(=O)CCc1csc(NC(=O)Nc2ccccc2)n1</chem>
F6513-5982	-7.63343	HDAC10	<chem>NC(=O)COc1ccc(S(=O)(=O)NCc2cnn3c2CCCC3)cc1</chem>
F3168-2340	-10.4698	HDAC4	<chem>COc1ccc(CCNC(=O)CCCCN2c(SCC(=O)NCc3ccccc3)nc3cc4c(cc3c2=O)OCO4)</chem>
F0642-3464	-10.4423	HDAC4	<chem>Cc1cc(C)c2nc(NNC(=O)c3nc4ccccc4s3)sc2c1</chem>
F0642-3364	-10.3453	HDAC4	<chem>Cc1ccc2sc(NNC(=O)c3coc4ccccc4c3=O)nc2c1C</chem>
F0617-0133	-10.2524	HDAC4	<chem>O=C(CSc1cn(CCNC(=O)c2ccc(F)cc2)c2ccccc12)NCC1CCCO1</chem>
F3168-2328	-10.1357	HDAC4	<chem>COc1ccc(CCNC(=O)CCCCN2c(SCc3cc(=O)n4cc(Cl)ccc4n3)nc3cc4c(cc3c2=O)O</chem>
F0617-0419	-10.1244	HDAC4	<chem>Cc1ccc(C(=O)NCCn2cc(SCC(=O)NCC3CCCO3)c3ccccc32)cc1</chem>
F2416-0766	-9.75213	HDAC6	<chem>O=C(Cc1cccc2ccccc12)Nc1ccc2c(c1)CN(C(=O)C1CC1)CC2</chem>
F5225-0361	-9.68389	HDAC6	<chem>CCC(=O)Nc1ccc(SCC(=O)Nc2cc(OC)ccc2OC)nn1</chem>
F6252-4761	-9.40785	HDAC6	<chem>O=C(NCc1ccc(N2CCCC2=O)cc1)c1cc2c(=O)[nH]c(-c3ccsc3)cn2n1</chem>
F2416-0095	-9.40021	HDAC6	<chem>CC(=O)N1CCc2ccc(NC(=O)Cc3cccc4ccccc34)cc2C1</chem>
F6195-0303	-9.37785	HDAC6	<chem>CC(=O)Nc1nc2c(s1)CN(CC(=O)Nc1ccc3c(c1)OCCO3)CC2</chem>
F5834-2458	-9.35931	HDAC6	<chem>COCCNC(=O)Nc1ccc2c(c1)CN(C(C)=O)CC2</chem>

F2335-0642	-10.3024	HDAC7	<chem>O=C(Cc1csc(SCC(=O)Nc2ccc(F)cc2)n1)NCc1ccccn1</chem>
F2335-0484	-10.2101	HDAC7	<chem>COc1ccc(NC(=O)CSc2nc(CC(=O)NCc3ccccc3)cs2)cc1</chem>
F2335-0928	-10.1077	HDAC7	<chem>Cc1ccc(NC(=O)CSc2nc(CC(=O)NCc3ccccc3)cs2)cc1</chem>
F6524-1565	-10.0677	HDAC7	<chem>O=C(NCc1nc(-c2ccc[nH]c2=O)no1)OCc1ccccc1</chem>
F6436-3065	-9.99704	HDAC7	<chem>O=C1NCCc2ccc(NC(=O)N3CCNC3=O)cc21</chem>
F6125-2085	-9.90808	HDAC7	<chem>CS(=O)(=O)Nc1cccc(NC(=O)C2CN(C(=O)COc3ccccc3F)C2)c1</chem>
F6524-1480	-9.88	IIa	<chem>O=C(Cn1ccc2ccccc21)NCc1nc(-c2ccc[nH]c2=O)no1</chem>
F6524-1527	-9.62	IIa	<chem>O=C(/C=C\Cc1ccccc1)NCc1nc(-c2ccc[nH]c2=O)no1</chem>
F0325-0328	-9.53	IIa	<chem>CC(=O)Nc1ccc(Cl)cc1C(=O)c1ccccc1Cl</chem>
F5462-0023	-9.47	IIa	<chem>O=C(CCc1csc(NC(=O)c2ccc[nH]c2=O)n1)NCc1cccs1</chem>
F1361-0034	-9.46	IIa	<chem>CC(=O)NC1C(Oc2ccccc2Cl)OC(CO)C(O)C1O</chem>
F6524-1583	-9.45	IIa	<chem>O=C(/C=C\Cc1ccccc1F)NCc1nc(-c2ccc[nH]c2=O)no1</chem>
F6506-0883	-8.47	IIb	<chem>NC(=O)COc1ccc(S(=O)(=O)NCC(O)(c2ccsc2)c2ccco2)cc1</chem>
F6506-1999	-8.19	IIb	<chem>NC(=O)COc1ccc(S(=O)(=O)NCC(O)(c2ccoc2)c2cccs2)cc1</chem>
F6506-1399	-8.13	IIb	<chem>NC(=O)COc1ccc(S(=O)(=O)NCC(O)(c2ccco2)c2cccs2)cc1</chem>
F6506-0367	-8.11	IIb	<chem>NC(=O)COc1ccc(S(=O)(=O)NCC(O)(c2ccoc2)c2ccco2)cc1</chem>
F6571-5982	-8.05	IIb	<chem>CC(Cc1csc2ccccc12)NS(=O)(=O)c1ccc(OCC(N)=O)cc1</chem>
F6553-7335	-7.88	IIb	<chem>NC(=O)COc1ccc(S(=O)(=O)N2CCc3sc(Cl)cc3C2)cc1</chem>

Додаток 2

Розподіл параметрів обраних дескрипторних характеристик та статистичний аналіз для селективних інгібіторів класу IIa та класу IIb



Checking assumptions for class_IIa

Analyzing H_donors:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Analyzing H_acceptors:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0002

Analyzing MolWt:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Analyzing cLogP:

Normality (Shapiro-Wilk test): Pass, p-value = 0.6305

Analyzing TPSA:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0039

Analyzing FSP3:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Checking assumptions for class_IIb

Analyzing H_donors:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0001

Analyzing H_acceptors:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Analyzing MolWt:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Analyzing cLogP:

Normality (Shapiro-Wilk test): Pass, p-value = 0.6083

Analyzing TPSA:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Analyzing FSP3:

Normality (Shapiro-Wilk test): Fail, p-value = 0.0000

Homogeneity of variances (Levene's test) for H_donors: Fail, p-value = 0.0019

Homogeneity of variances (Levene's test) for H_acceptors: Fail, p-value = 0.0001

Homogeneity of variances (Levene's test) for MolWt: Fail, p-value = 0.0020

Homogeneity of variances (Levene's test) for cLogP: Fail, p-value = 0.0048

Homogeneity of variances (Levene's test) for TPSA: Fail, p-value = 0.0039

Homogeneity of variances (Levene's test) for FSP3: Fail, p-value = 0.0000

Analyzing H_donors across all targets:

Kruskal-Wallis test: H-statistic = 13.164759605043777, p-value = 0.0002852633814123597

Performing pairwise comparisons with Bonferroni correction:

class_Ia vs class_Ib: U-statistic = 2882.0, p-value = 0.0003, Significant

Analyzing H_acceptors across all targets:

Kruskal-Wallis test: H-statistic = 27.321343953419397, p-value = 1.7229898823870673e-07

Performing pairwise comparisons with Bonferroni correction:

class_Ia vs class_Ib: U-statistic = 2217.5, p-value = 0.0000, Significant

Analyzing MolWt across all targets:

Kruskal-Wallis test: H-statistic = 45.28265675417785, p-value = 1.7055278301989074e-11

Performing pairwise comparisons with Bonferroni correction:

class_Ia vs class_Ib: U-statistic = 1594.0, p-value = 0.0000, Significant

Analyzing cLogP across all targets:

Kruskal-Wallis test: H-statistic = 0.8240414059257598, p-value = 0.36400119729307523

Analyzing TPSA across all targets:

Kruskal-Wallis test: H-statistic = 60.155625192996794, p-value = 8.764605246604452e-15

Performing pairwise comparisons with Bonferroni correction:

class_Ia vs class_Ib: U-statistic = 1192.0, p-value = 0.0000, Significant

Analyzing FSP3 across all targets:

Kruskal-Wallis test: H-statistic = 18.9543500834936, p-value = 1.3388367125158608e-05

Performing pairwise comparisons with Bonferroni correction:

class_Ia vs class_Ib: U-statistic = 5939.5, p-value = 0.0000, Significant

