

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ВИСОКИХ ТЕХНОЛОГІЙ

Завідувач кафедри молекулярної біології та
біоінформатики Олексій Юрійович Нипорко
Протокол № ____ засідання
кафедри від “ ____ ” _____ 20__ р.

**Приорітизація низькомолекулярних сполук як потенційних
хімічних зондів**

Випускна кваліфікаційна робота магістра
студентки спеціальності 091 Біологія
ОП «Біоінформатика та структурна біологія»
Маханькової Олександри Вікторівни

Науковий керівник
професор кафедри
супрамолекулярної хімії, д.х.н,
проф. **Волочнюк Дмитро Михайлович**

Оцінка захисту роботи

Київ – 2024 р.

Анотація

Маханькова О.В. Приорітизація малих молекул як потенційних хімічних зондів. Випускна кваліфікаційна робота магістра за спеціальністю 091 Біологія, ОП «Біоінформатика та структурна біологія». У цій роботі представлені результати розробки та аналізу бази даних білків та їх класифікації з використанням методу BLAST для ідентифікації потенційних мішеней малих молекул. Окрема увага в роботі приділена розробці інструменту для полегшення виявлення малих молекул, що є потенційними хімічними зондами. Використання комплексного підходу, поєднуючи біоінформатику та хемоінформатику, відкриває нові можливості для виявлення та приорітизації малих молекул як потенційних хімічних зондів з високою селективністю та активністю.

Ключові слова: малі молекули, хімічні зонди, біоінформатика, BLAST, класифікація білків, хемоінформатика, білкові мішені, приорітизація.

Annotation

Makhankova O.V. Prioritization of Small Molecules as Potential Chemical Probes. Master's thesis for the specialty 091 Biology, EP "Bioinformatics and Structural Biology". This work presents the results of the development and analysis of a protein database and their classification using the BLAST method for identifying potential targets of small molecules. Special attention in the work is given to the development of a tool to facilitate the detection of small molecules that are potential chemical probes. The use of a comprehensive approach, combining bioinformatics and chemoinformatics, opens new opportunities for the detection and prioritization of small molecules as potential chemical probes with high selectivity and activity.

Keywords: small molecules, chemical probes, bioinformatics, BLAST, protein classification, chemoinformatics, protein targets, prioritization.

Зміст

Вступ.....	4
Розділ 1 Хімічні зонди та їх роль в біохімічних дослідженнях.....	6
1.1 Визначення та значення хімічних зондів у дослідженнях.....	6
1.2 Огляд ролі хімічних зондів в академічних дослідженнях, валідації мішеней та ідентифікації захворювань.....	7
1.3 Характеристики ефективних хімічних зондів.....	9
1.4 Приклади застосування хімічних зондів у дослідженнях хвороб.....	11
Розділ 2 Хімічні зонди у розробці терапії.....	14
2.1 Від ідентифікації хімічних зондів до ризиків та інновацій у поліфармакології.....	14
2.2 Важливість хімічних зондів для досліджень геному в контексті пошуку нових мішеней для терапії.....	15
2.3 Передослідженість та супутні з нею проблеми.....	18
2.4 Перспективи дослідження механізмів захворювань з використанням хімічних зондів.....	20
Розділ 3 Експериментальна частина.....	23
3.1 Класифіковані та некласифіковані мішені.....	23
3.2 Розробка інструменту для класифікації мішеней.....	25
3.3 Відбір потенційних хімічних зондів.....	30
Розділ 4 Результати та обговорення.....	34
4.1 Огляд методології та основні результати.....	34
4.2 Аналіз результатів.....	35
Висновки.....	36
Список використаних джерел.....	38

Вступ

У сучасному світі наукових відкриттів та технологічного прогресу, галузі біоінформатики та хемоінформатики відіграють ключову роль у розумінні складних біологічних процесів та розробці нових лікарських препаратів. Ці дисципліни забезпечують аналіз великих обсягів даних, що отримуються з експериментів, дозволяючи вченим розшифровувати біологічні механізми на молекулярному рівні, ідентифікувати потенційні мішені для лікарських засобів та досліджувати ці мішені.

У цьому контексті, хімічні зонди є незамінними інструментами, які відкривають нові можливості для дослідження живих організмів та розробки лікувальних стратегій. Хімічні зонди — це ретельно вивчені молекули, що мають біоактивність і селективність щодо цільових білків. Термін "хімічний зонд" відділяє такі сполуки від ліків, що використовуються у клінічних умовах, але не є доступними чи не мають дослідженого механізму дії, а також від таких біоактивних інгібіторів, лігандів, агоністів чи антагоністів, які спрямовані на певний білок, але недостатньо охарактеризовані чи селективні, і від "зондів", які є лабораторними реагентами для біофізичних досліджень та візуалізації. Хімічні зонди є інструментами для визначення функцій окремих білкових мішеней і відіграють важливу роль у дослідженнях[1].

Мета роботи: Створити інструмент для зручного пошуку потенційних хімічних зондів

Для вирішення поставленої мети потрібно виконати такі **завдання**:

1. Отримати актуальні дані з бази класифікованих протеїнів Enamine Bioactive Compounds станом на січень 2024 року.
2. Створити на основі цієї локальної бази даних інструмент класифікації протеїнів за допомогою методу BLAST та класифікувати нові надходження білків у базу.

3. Створити інструмент для зручного виявлення потенційних хімічних зондів.

4. Проаналізувати отримані результати.

Об'єктом дослідження є класифіковані та уніфіковані білки в локальній базі даних, а також низькомолекулярні сполуки, що є потенційними хімічними зондами.

Предмет дослідження: методи класифікації протеїнів, зокрема за допомогою BLAST, а також методи створення інструменту для фільтрації сполук за визначеними критеріями для хімічних зондів.

Методи дослідження: розробка програмного інструменту на основі локальної бази даних для класифікації протеїнів з використанням методу BLAST. Розробка інструменту для відбору потенційних хімічних зондів за фільтрами значень біоактивності та селективності.

Структура роботи. Розділ 1-2. Огляд літератури. Розділ 3. Експериментальна частина. Розділ 4. Результати та їх обговорення. Висновки. Список використаних джерел.

Розділ 1.

Хімічні зонди та їх роль в біохімічних дослідженнях

1.1 Визначення та значення хімічних зондів у дослідженнях

Розробка нових ліків часто починається з ідентифікації нової мішені за допомогою скринінгів. На фазі валідації використовуються різні стратегії для оцінки, чи відіграє ідентифікована мішень ключову роль у процесі захворювання та чи може модуляція її активності привести до терапевтичної ефективності[2]. На цьому етапі хімічні зонди є надзвичайно корисними. На відміну від молекулярних зондів, які викликають нокдаун або нокаут мішені протягом доволі тривалого періоду часу, хімічні зонди швидко пригнічують активність цільового білка. Крім того, у поєднанні з молекулярними зондами хімічні зонди можуть розрізняти ефекти, що виникають через присутність мішені в клітині та через інгібування активності каталітичної або білок-білкової взаємодії. Необхідність хімічних зондів у дослідженнях призвела до постійного зростання їх кількості. Кожен хімічний зонд повинен відповідати мінімальним критеріям, а саме: потужності, селективності та активності у клітинах. Ці властивості можуть змінюватися залежно від природи цільового білка, але в цілому хімічні зонди повинні мати активність *in vitro* не більша ніж 100 нМ (не менше за 7 рХ), селективність до цільового білка є принаймні в 30 разів більшою порівняно з білками з того самого сімейства, і активність у клітинах в концентраціях нижче за 1 мкМ. Подібним чином, рекомендації щодо використання хімічних зондів у експериментах в клітинах можуть відрізнятися, тим не менш, головною рекомендацією є використання хімічних зондів у концентраціях, наближених до підтверджених під час експериментів *in vitro*. Навіть найбільш селективний хімічний зонд стане неселективним, якщо використовувати його у високій концентрації. По-друге, рекомендується використовувати хімічний зонд разом з неактивним щодо цільового білку,

але структурно подібним аналогом, який слугував би негативним контролем. По-третє, кожен білок має мати хоча б два ортогональні хімічні зонди. Таким чином, для будь-якого білка ідеальний експеримент у клітинах мав би проводитись з використанням принаймні двох ортогональних хімічних зондів в адекватних концентраціях і супроводжуватися, якщо це можливо, структурно подібними, але неактивними сполуками[3].

Важливість хімічних зондів у сучасній науці не можна переоцінити. Вони знайшли широке застосування у різних сферах, включаючи медичну діагностику, фармацевтику, клітинну біологію, екологію. За допомогою хімічних зондів можна досліджувати сигнальні шляхи, вивчати механізми захворювань, відстежувати ефективність лікарських препаратів, а також розробляти нові. Це значно підвищує шанси на успішне лікування.

Усвідомлення важливості хімічних зондів в сучасних дослідженнях підкреслює не тільки їхню роль у відкриттях та інноваціях, але й акцентує на необхідності продовжувати розвиток інструментів, що відкриватимуть нові перспективи для вчених у пошуку відповідей на найскладніші питання, поставлені наукою сьогодні.

1.2 Огляд ролі хімічних зондів в академічних дослідженнях, валідації мішеней та ідентифікації захворювань

У дослідженні хвороб і розробці ліків, хімічні зонди використовуються в комбінації з генетичними технологіями, такими як RNAi та CRISPR, для підтвердження функцій потенційних біологічних мішеней. Ці зонди є незамінними для визначення ролі біомолекул у розвитку захворювань. Щоб покращити процес валідації таргетів при розробці нових медикаментів, рекомендується звертати увагу на чотири ключові аспекти: забезпечення достатнього рівня присутності зонду в клітині; підтвердження того, що зонд точно взаємодіє з цільовим таргетом; демонстрація здатності зонду впливати

на активність цілі; підтвердження здатності зонду впливати на відповідні фенотипічні зміни (рис 1.1) [4].



Рисунок 1.1 – Чотири ключові аспекти для покращення валідації таргетів

Автори підкреслюють, що перевірка взаємодії зонду з таргетом є важливою для зв'язування експозиції зонду з його функціональною фармакологією та здатністю викликати фенотипічні зміни.

Значимість хімічних зондів у академічних дослідженнях білкових функцій і їх потенціал як таргетів для лікувальних препаратів стимулювала створення співробітництва між промисловими та академічними дослідниками з метою розробки зондів для аналізу всього протеома. До 2035 року дослідники прагнуть перенести успіхи геноміки в клінічні дослідження, особливо у розробку нових низькомолекулярних сполук для лікування захворювань[5]. (рис 1.2)



Рисунок 1.2 – Часова шкала дослідження хімічних зондів та майбутніх перспектив

Однією з важливих переваг хімічних зондів є їх здатність емулювати фармакологію лікарських засобів. Незважаючи на те, що для хімічних зондів немає вимог до характеристик, що необхідні для ліків, таких як властивості абсорбції, розподілу, метаболізму, екскреції та токсичності (ADMET), все ж існує упевненість у тому, що зонди можуть служити вихідною точкою для прискорення процесу розробки нових медикаментів.

1.3 Характеристики ефективних хімічних зондів

Але все ж необхідно вилучати з досліджень сполуки, які проявляють себе як особливо проблемні через їхню неконкретну реакційну здатність по відношенню до білків, утворення колоїдних агрегатів, які можуть неспецифічно зв'язуватися з білками та інактивувати їх, а також викликати токсичні ефекти в клітинах, наприклад, шкодячи мембранам через ефект, відомий як фосфоліпоз. Ці сполуки також можуть впливати на показники біологічних есеїв через їх спектральні або флуоресцентні властивості. Зазвичай такі сполуки класифікують як з'єднання, які створюють перешкоди в біоесеях (PAINS) [6]. Для виявлення такої хімічної проміскупності застосовують спеціалізоване програмне забезпечення та алгоритми, а також проводять базові лабораторні аналізи для перевірки наявності згаданих проблемних властивостей. Такі сполуки повинні бути виключені з подальшого розгляду та використання як хімічні зонди.

Також для відбору хімічних зондів найвищої якості критично важливо застосовувати строгі критерії. Це включає використання так званих «факторів придатності» [7] і рекомендацій, розроблених Глобальним Консорціумом Структурної Геноміки (SGC) та міжнародною групою фахівців. Необхідно, щоб активність зонду проти цільового білка була високою, з показниками IC₅₀ або K_d не вище 30-100 нМ. Важливо також спостерігати активність зонду в клітинних системах при концентраціях не

більше 1-5 мкМ, ідеально — на субмікромолярному рівні. Вища активність сполуки зменшує ймовірність небажаних взаємодій з іншими білками.

Не менш важливою є глибока характеристика хімічного зонда, особливо з точки зору його селективності. Важливо, щоб зонд був у багато разів більш селективним (мінімум 1.5 ΔрХ) для цільового білка порівняно з іншими білками цієї ж родини. Широкий профіль взаємодій, що оцінюється проти фармакологічних панелей, які використовуються у промисловості для оцінки безпеки та мінімізації побічних ефектів потенційних лікарських засобів, є надзвичайно корисним. На наступному рисунку (рис 1.3) наводяться основні відмінності між хімічним зондом та ліками:



Рисунок 1.3 – Основні відмінності між хімічними зондами та ліками

Оцінка і вибір хімічних зондів для досліджень на клітинах та використання на тваринних моделях охоплює велику кількість чинників, що може бути значним ускладненням. Існують ризики надмірно дидактичного підходу, який може створити враження нав'язливості або евангелізму, а

також ускладнити процес публікації нових хімічних зондів або вибір серед відомих. Проте, як відзначили деякі дослідники [1], строгі вимоги можуть становити виклик для невеликих дослідницьких груп, зокрема для науковців на початку кар'єри, які прагнуть відповідати високим стандартам і публікувати свої перші роботи з новими хімічними зондами. Тому слід застосовувати практичний підхід, який відповідає конкретним потребам. Настанови щодо кращих практик є корисними, однак вони не є жорсткими правилами. На ранніх етапах розвитку зондів для певної мети, менш оптимальні реагенти можуть бути корисними як "першопрохідники", які покладають шлях до більш кваліфікованих хімічних зондів, що з'являються згодом. Ці першопрохідні зонди можуть бути не такими активними, як би хотілося, мати деякі побічні ефекти, або не підходити для використання в дослідках на тваринах, але вони все ще зберігають свою корисність. Для таких зондів існує місце доки ми усвідомлюємо їх неідеальність, чітко визначаємо їх властивості та обмеження, і не перебільшуємо результати в публікаціях. Також варто згадати, що відкриття та оцінювання якісних хімічних зондів може відбуватися швидко і ефективно завдяки багатодисциплінарним співпрацям та взаємодіям між хіміками, біологами, фармакологами, структурними біологами та експертами з хвороб.

1.4 Приклади застосування хімічних зондів у дослідженнях хвороб

Хімічні зонди стали незамінним інструментом у розробці ліків, зокрема у визначенні важливості білкових мішеней у патогенезі захворювань. Прикладом є дослідження ферментів, які беруть участь у синтезі холестеролу. Використання статинів, як хімічних зондів, дозволило вченим глибше зрозуміти механізм дії HMG-CoA редуктази, ключового ферменту у біосинтезі холестеролу[8]. Це вивчення не лише збільшило наше розуміння метаболізму холестеролу, але й сприяло розробці нових ліпідних

знижувальних терапій, що сьогодні є стандартом лікування гіперхолестеролемії.

Іншим ключовим застосуванням хімічних зондів є їх використання у дослідженні взаємодій між білками та їх маленькими молекулярними інгібіторами в контексті онкології. Наприклад, інгібітори кіназ, такі як иматиніб, специфічно ціляться на абнормальні білкові мішені, характерні для деяких форм раку, як BCR-ABL у хронічному мієлоїдному лейкозі[9]. Цей підхід дозволяє не тільки інгібувати ріст та поділ ракових клітин, але й значно зменшує побічні ефекти порівняно з традиційною хіміотерапією.

Зонди також відіграють важливу роль у вивченні нейродегенеративних захворювань. Інгібітори, що ціляться на ферменти, які каталізують фосфорилування тау-білка, використовуються для дослідження їхньої ролі у формуванні нейрофібрилярних клубочків, які є патогномонічними для хвороби Альцгеймера[10]. Ці хімічні зонди допомагають у визначенні потенціалу цільових молекул для подальшої розробки ліків, призначених для лікування або затримки прогресування захворювання.

У галузі імунології хімічні зонди використовуються для вивчення взаємодій, що відбуваються між різними клітинними рецепторами та їх лігандами. Наприклад, модулятори TLR (Toll-like receptors) допомагають вивчати молекулярні шляхи активації імунної відповіді, що може призвести до розробки нових вакцин або імуномодуляторів, здатних ефективно боротися з інфекційними агентами або аутоімунними розладами[11].

Крім того, хімічні зонди стають незамінними у розробці стратегій лікування метаболічних розладів. Зонди, що інгібують ключові ферменти гліколізу або жирнокислотного обміну, дозволяють детально вивчати метаболічні шляхи, що є важливими для розуміння та лікування цукрового діабету та ожиріння[12].

Важливість хімічних зондів також очевидна у вивченні антибактеріальної резистентності, де вони використовуються для ідентифікації і валідації нових мішеней в боротьбі з мультирезистентними

штамами бактерій. Наприклад, зонди, які блокують механізми резистентності, такі як ефлюксні помпи, можуть підвищити ефективність існуючих антибіотиків, дозволяючи знову використовувати препарати, які раніше вважалися неефективними.

Розділ 2.

Хімічні зонди у розробці терапії

2.1 Від ідентифікації хімічних зондів до ризиків та інновацій у поліфармакології

Вивчення та розробка низькомолекулярних сполук для взаємодії з макромолекулами є цікавою, оскільки вони завжди використовувались як метод, що не потребує інвазивних втручань, для вивчення функцій білків у біології та медицині. Цей метод доповнює більш радикальні підходи, такі як втручання з допомогою РНК (RNAi), сприяючи розвитку біологічних наук та новітніх методів лікування. В сучасних наукових дослідженнях наразі відбувається активний пошук нових хімічних зондів для анотації малодосліджених сегментів людського геному і подальшої перевірки молекулярних мішеней. Це є ключем для прогресу в хімічній біології та відкритті нових медичних препаратів.

Проте використання хімічних зондів не позбавлене ризиків, а розробка якісних зондів та контрольних об'єктів продовжує викликати жваві обговорення. Критичним аспектом є селективність до основної білкової мішені, оскільки нерозуміння механізму дії зонду та ігнорування потенційних непомічених побічних ефектів може призвести до хибного тлумачення результатів, що, в свою чергу, може спричинити наукові плутанини та збільшення невиправданих витрат.

Для розв'язання проблем селективності хімічні зонди зазвичай перевіряються *in vitro* на виборці білків, філогенетично пов'язаних з їх основною мішенню. Багато зондів також проходять широкий профіль тестувань проти розмаїття інших мішеней. Втім, обмежені можливості лабораторних випробувань *in vitro* не завжди дозволяють виявити всі потенційні поліфармакологічні взаємодії, що може призвести до невиявлених віддалених поліфармакологічних ефектів[13].

In silico профілювання мішеней ефективно використовується для ідентифікації нових мішеней лікарських препаратів і відбору кандидатів з фенотипічного скринінгу. Ці технології дозволяють оцінити поліфармакологію малих молекул на значно більшій частині протеому, ніж це можливо зі стандартними *in vitro* панелями. Також це дозволяє забезпечити більш повне розуміння можливих взаємодій між хімічними зондами та мішенями. Зокрема, *in silico* методи, такі як молекулярне моделювання та комп'ютерне прогнозування, стають невід'ємною частиною процесу відкриття нових медичних препаратів. Ці технології допомагають визначити потенційні непомічені побічні ефекти до того, як зонди будуть використані в клінічних випробуваннях, знижуючи таким чином ризики та витрати, пов'язані з розробкою лікарських засобів. Окрім того, стрімкий розвиток технологій в області штучного інтелекту відкриває нові можливості для оптимізації процесу відбору та перевірки хімічних зондів. Алгоритми машинного навчання можуть обробляти величезні масиви даних про взаємодії між молекулами та білками, що значно прискорює процес виявлення хімічних зондів з найкращими властивостями. Це дозволяє ефективніше знаходити потенційні ліки та інструменти для біомедичних досліджень, оптимізуючи їхню активність і специфічність. І також це сприяє більш точній настройці хімічних зондів, підвищуючи їх ефективність і зменшуючи ймовірність взаємодій з нецільовими мішенями, що важливо для уникнення можливих токсичних ефектів.

Ці напрямки досліджень забезпечують не тільки посилення безпеки і ефективності хімічних зондів, але й відкривають шлях до більш успішної та інноваційної терапії в майбутньому.

2.2 Важливість хімічних зондів для досліджень геному в контексті пошуку нових мішеней для терапії

У сучасному світі фармацевтичних розробок існує значний розрив між великою кількістю кодуючих білок генів людини та обмеженою кількістю мішеней, що активно використовуються у медичних препаратах. Наразі мішеней сучасних лікарських засобів у 100 разів менше, ніж генів, що кодують білки. Вважають, що потенційних таргетів для ліків може бути в десять разів більше, аніж наразі відомо, і вони все ще очікують на своє відкриття. Існує безліч білків у сигнальних шляхах, що мають сильний зв'язок зі захворюваннями, які не можуть бути ефективно вивчені за допомогою традиційних підходів.

У цьому контексті величезне значення має наявність хімічних зондів для кожного потенційно "лікувального" місця зв'язування в людському протеомі. Втім, досягнення такої амбіційної мети потребує значних зусиль. Тим часом, інгібітори з вузькими, але добре анотованими профілями селективності можуть слугувати як хемогеномний набір для валідації мішеней. Збірка таких наборів інгібіторів особливо здійсненна для основних сімейств мішеней ліків, таких як білкові кінази, GPCR, протеази та інші. Ці великі сімейства білків дозволяють розробку проектів зондів для великої кількості поки що не вивчених членів сім'ї. Класичні сім'ї білків охоплюють більше ніж 2000 потенційних нових мішеней для ліків. Одним з відомих прикладів використання хімічних зондів є розробка інгібіторів кіназ, які використовуються для лікування раку. Наприклад, застосування вузькопрофільного інгібітора для розробки нових терапевтичних принципів – кіназний інгібітор кризотиніб, розроблений для цілі рецепторної тирозинкінази c-MET[14]. Після відкриття, що споріднена кіназа ALK є ключовим фактором у частині випадків недрібноклітинного раку легень (NSCLC), кризотиніб був швидко адаптований для цього застосування (рис. 2.1), що призвело до його схвалення для лікування ALK-залежного NSCLC[15].

На жаль, для більшості членів класичних родин білків, включно з кіназами, не розроблено селективних інгібіторів. У деяких випадках спосіб

зв'язування або місце зв'язування інгібітора також є релевантними, збільшуючи кількість можливих стратегій націлювання в добре вивченому пулі мішеней для ліків. Кінази, наприклад, виконують як структурну, так і ферментативну функції, і обидва активні центри можуть бути ціллю для малих молекул[16]. Алостеричні інгібітори для кіназ ERK1/2 та АКТ тому викликають різні фармакологічні ефекти і показують різні результати в моделях пухлин порівняно з інгібіторами, що діють лише на каталітичну активність.

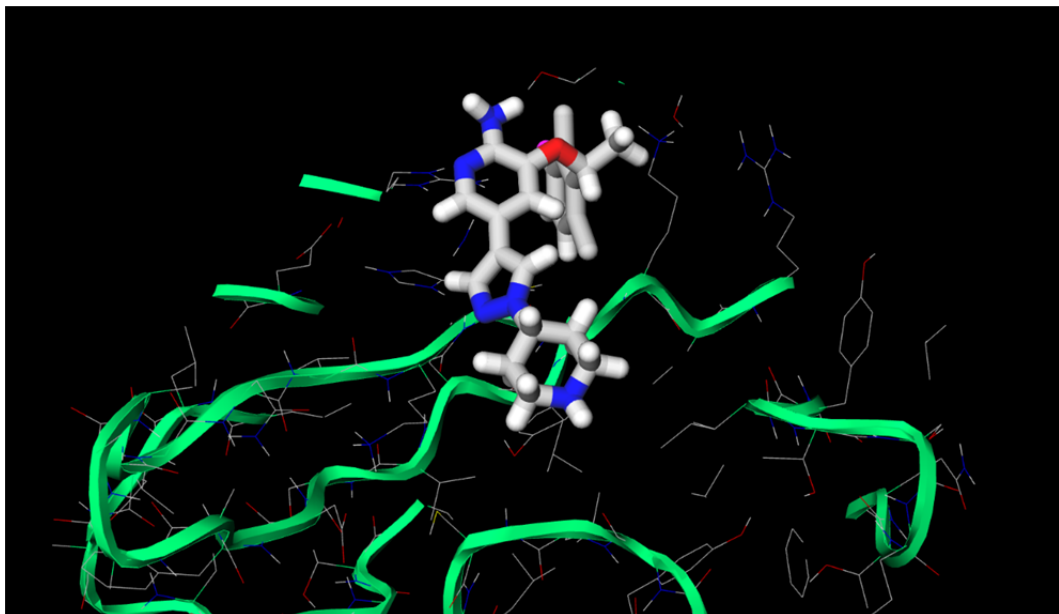


Рисунок 2.1 – Кризотиніб в активному центрі ALK

Розвиток початкових хімічних хітів для сімейств білків, які раніше вважалися непридатними для таргетування, також може розширити простір цілей. Така група важкодоступних мішеней включає, наприклад, білки з великою кількістю доменів взаємодії. Наприклад, ацетил-лізинові зв'язувальні модулі сімейства бромодоменів не вважалися потенційними мішенями для розробки ліків до тих пір, поки не були розроблені перші потужні інгібітори[17]. Інгібітори BET були винайдені завдяки дослідженням, проведеним за допомогою хімічних зондів. Вони наразі проходять клінічні випробування як потенційні ліки проти раку. Ці інгібітори

можуть стати основою для нових протипухлинних терапій, використовуючи отримані знання для боротьби з різними видами раку, оскільки ці зонди показали високу ефективність у моделях раку, особливо в лікуванні дифузної великоклітинної В-клітинної лімфоми та інших форм злоякісних новоутворень.

Отож, інновації в технологіях протеоміки значно підсилили можливості хімічних зондів. Завдяки поступам у мас-спектрометрії та інших аналітичних методах, вчені можуть зараз ідентифікувати взаємодії між маленькими молекулами та білками з високою точністю та великою пропускнуою здатністю. І це дозволяє точно вимірювати афінітет зв'язування і вибірковість хімічних зондів, покращуючи здатність прогнозування їхнього впливу на живий організм. Перспективи розвитку хімічних зондів в майбутньому здаються обнадійливими, з огляду на їхній потенціал для глибокого розуміння біологічних процесів і розвитку нових терапевтичних стратегій. Розширення наборів зондів для включення більш широкої варіативності біологічних мішеней і зв'язувальних сайтів відкриває можливості для інноваційних досліджень у багатьох галузях медицини. Це може включати розвиток нових інгібіторів для таких складних захворювань, як хвороба Альцгеймера, а також удосконалення існуючих терапій шляхом більш точного націлювання на молекулярні мішені, що пов'язані з хворобою. Це буде вимагати продовження тісної співпраці між академічними, урядовими та промисловими секторами для вирішення ключових наукових викликів, пов'язаних з визначенням та валідацією нових біологічних мішеней. Використання хімічних зондів для розуміння молекулярної етіології захворювань і механізму дії препаратів надасть значні переваги у лікуванні багатьох станів, підвищуючи ефективність медичних втручань і зменшуючи частоту та важкість побічних ефектів.

2.3 Передослідженість та супутні з нею проблеми

Як було зазначено вище, хімічні зонди є ключовими інструментами в біологічних дослідженнях, зокрема у вивченні протеїнів, які ще не досліджені або мало відомі. Ці молекули дозволяють вченим зрозуміти функції цільових білків у клітинних процесах, виявляючи потенційні терапевтичні мішені. Це особливо актуально в контексті диспропорції в увазі до деяких білків у порівнянні з іншими (рис 2.2). Наприклад, якщо зосередити свою увагу на такій родині білків, як BCL-2, які, до речі, є одними з дуже перспективних мішеней для лікування раку[18], то можна побачити, що значна увага зосереджена лише на декількох членах, в той час як інші, потенційно важливі, майже не вивчаються.

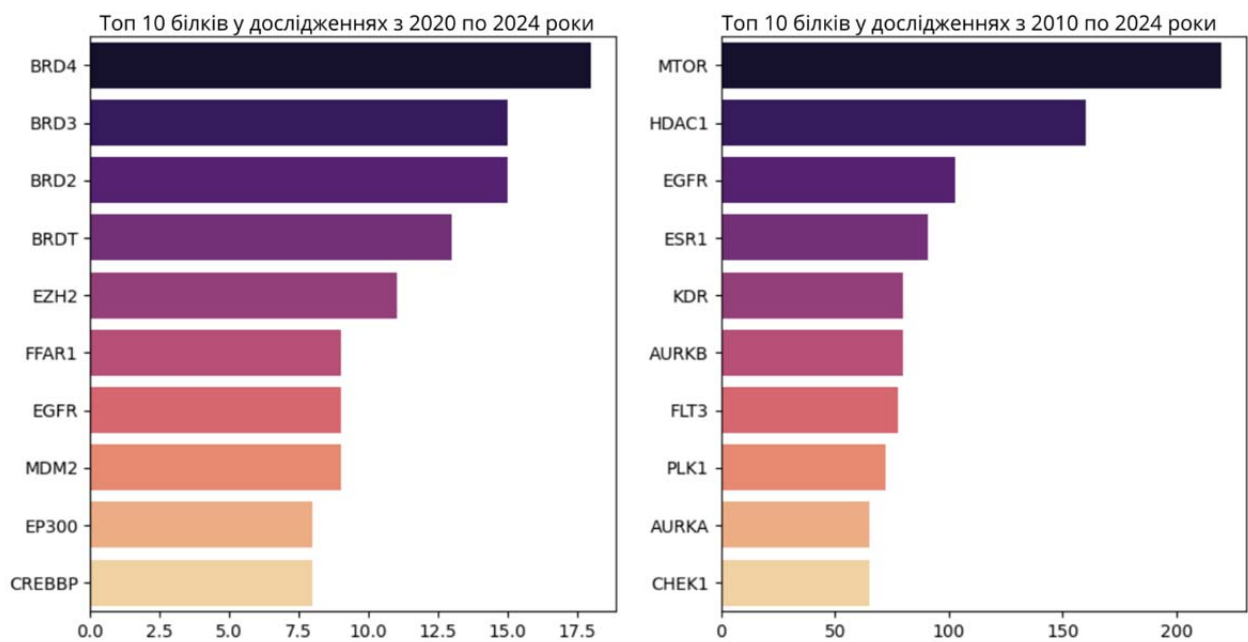


Рисунок 2.2 – Найбільш досліджувані білки за останні 4 і 14 років

Однією з причин цього є те, що вчені мають схильність занурюватися глибше в уже знайомі області. Крім того, системи фінансування часто уникають ризиків, віддаючи перевагу дослідженням, що базуються на вивчених білках, для яких легше обґрунтувати актуальність і значущість. Застосування хімічних зондів може допомогти подолати цю перешкоду, оскільки вони дозволяють швидко оцінювати біологічну активність білків і

забезпечують засоби для їхнього вивчення в реальних клітинних процесах. Наприклад, використання інгібіторів або активаторів дозволяє дослідникам пропонувати та перевіряти гіпотези про функції недосліджених білків, встановлюючи зв'язки між білковою активністю та фенотипами захворювань. Такий підхід також може прискорити відкриття лікарських засобів шляхом визначення нових мішеней або перевірки ефективності потенційних препаратів.

Проте, навіть з такими ініціативами, розвиток хімічних зондів і їх широке впровадження залишаються викликом через високі витрати і час, необхідний для їх розробки і характеристики. Наявність наборів хімічних зондів для широкого спектру білків могла б радикально змінити баланс у біологічних дослідженнях, дозволяючи швидше і ефективніше вивчати недосліджені молекулярні шляхи і механізми захворювань.

2.4 Перспективи дослідження механізмів захворювань з використанням хімічних зондів

Хімічні зонди, що впливають на активність конкретних білків, можуть бути використані для вивчення патогенезу різних захворювань на клітинному та молекулярному рівнях. Дослідження, які використовують ці інструменти, можуть допомогти визначити критичні мішені для потенційної терапії і зрозуміти, як зміни в експресії або функціях цих білків впливають на розвиток хвороб.

Наприклад, біполярний розлад (BD) є складним психіатричним захворюванням з високою ймовірністю спадковості та генетичної гетерогенності. Попри існування численних досліджень[19, 20], що зосереджуються на ідентифікації генетичних факторів, що сприяють розвитку BD, точні молекулярні механізми, що лежать в основі цього розладу, залишаються недостатньо зрозумілими. Одним з потужних інструментів для глибшого розуміння цих механізмів є застосування

хімічних зондів. Вони можуть бути використані для модуляції активності конкретних кіназ, які, як виявлено в різних дослідженнях[21], грають важливу роль у сигнальних шляхах, що пов'язані з BD. Наприклад, дослідження ферменту диацилгліцерол кінази ета (DGKH), який показав значущу асоціацію з BD у геномних асоціаційних дослідженнях (GWAS), може бути розширене з використанням специфічних інгібіторів, розроблених як хімічні зонди. Це дозволить вченим більш детально вивчити роль DGKH у регуляції маніакальних та депресивних станів.

Хімічні зонди також можуть стати у нагоді для вивчення молекулярних механізмів, які можуть спричинити психіатричні прояви BD. Використовуючи ці інструменти, дослідники можуть визначити, як зміни у функціонуванні ANK3[22] та інших генів, впливають на клітинні функції і патологію. Точкове інгібування або активація цих білків може допомогти розкрити їх ролі в сигнальних шляхах, пов'язаних з BD, і визначити нові терапевтичні цілі.

Діабет 2-го типу (T2D) також є одним з поширених захворювань і частою причиною смерті та інвалідності у світі. Ідентифікація білків, пов'язаних із T2D, має велике значення для розуміння молекулярних механізмів захворювання та розробки нових терапевтичних стратегій. У цьому контексті, хімічні зонди можуть відіграти ключову роль у дослідженні цих генетичних варіантів на клітинному та молекулярному рівнях.

Значення SLC16A11 та SLC16A13 як генетичних маркерів для T2D було ідентифіковано в результаті аналізу 9,2 мільйонів однонуклеотидних поліморфізмів (SNPs) у дослідженні A. Williams et al. [23]. Хімічні зонди, що специфічно націлені на ці білки, можуть допомогти в дослідженні впливу цих генів на метаболізм триацилгліцеролу, що є ключовим фактором у патогенезі T2D. Це може включати розробку інгібіторів або активаторів, які модулюють активність окремих білків, тим самим вивчаючи їх роль у гомеостазі глюкози та ліпідів у клітинних моделях.

Також хімічні зонди можуть бути використані для валідації функціональної ролі ідентифікованих генетичних мішеней у розвитку T2D. За допомогою цих молекул, дослідники можуть експериментально модулювати активність відповідних білків у клітинах або на тваринних моделях, спостерігаючи за змінами у метаболічних шляхах, які ведуть до розвитку діабету. Це допоможе підтвердити генетичні асоціації, виявлені у GWAS, і забезпечити доказову базу для розробки нових лікарських препаратів, заснованих на цих мішенях.

Розділ 3.

Експериментальна частина

Отже, метою даної роботи було створення інструменту для зручного пошуку потенційних хімічних зондів, а також інструменту для пришвидшення процесу класифікації та уніфікації білків. Ця робота поділялася на 4 частини:

1. Отримати актуальні дані з бази класифікованих протеїнів EBC [24] станом на січень 2024 року.
2. Створити на основі цієї локальної бази даних інструмент класифікації протеїнів за допомогою методу BLAST та класифікувати нові надходження unclassified proteins у базу.
3. Створити інструмент, що буде допомагати виявляти потенційні хімічні зонди.
4. Проаналізувати отримані результати.

3.1 Класифіковані та некласифіковані мішені

Класифікація білків відіграє ключову роль у розумінні біологічних процесів, що відбуваються в живих організмах. Однак, це залишається складним завданням для наукової спільноти з кількох причин, які охоплюють біоінформатичні, біологічні та технічні аспекти. Нижче наведено декілька ключових аспектів, які роблять класифікацію білків значним науковим викликом:

1. Білків дуже багато і майже в кожного з них є декілька синонімів. В різних літературних джерелах один і той самий білок може мати різні назви через що задача уніфікації є доволі складною.
2. Білки відіграють безліч ролей у клітині, включаючи каталіз реакцій, сигнальні процеси, структурні функції, транспорт молекул, імунну відповідь тощо. Через цю різноманітність, один білок може виконувати

кілька функцій або взаємодіяти з різними молекулами, що ускладнює його класифікацію.

3. Навіть незначні зміни у амінокислотному складі білка можуть суттєво змінити його функцію або взаємодію з іншими білками. Ці варіації ускладнюють класифікацію, оскільки білки з високим ступенем схожості за послідовністю можуть мати різні функції.
4. З розвитком геноміки та протеоміки, кількість відомих білків зростає експоненційно. Зі збільшенням кількості даних зростає й складність їхньої обробки та аналізу. Великі дані вимагають потужних обчислювальних ресурсів і складних алгоритмів для ефективної класифікації.
5. Досі існує багато білків, які не були повністю досліджені, і чия функція залишається невідомою. Це створює проблему для класифікації, оскільки необхідно інтегрувати нові дані з вже існуючими базами знань.

У базі даних Enamine Bioactive Compounds був значний арсенал класифікованих білкових мішеней, а також тих, що потребували класифікації та уніфікації (рис 3.1, 3.2).

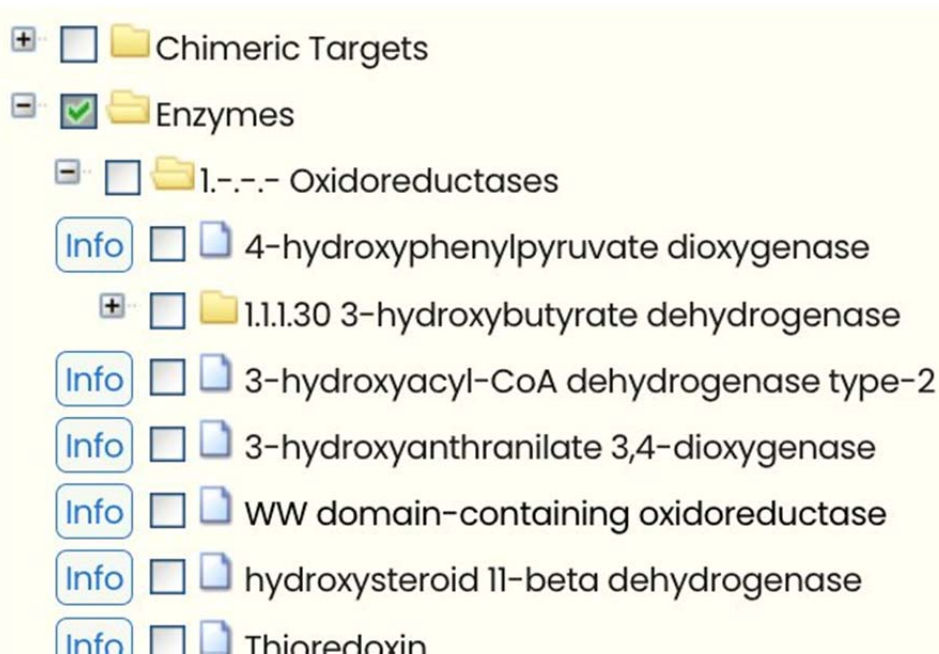


Рисунок 3.1 – Частина дерева класифікованих мішеней Enamine Bioactive Compounds

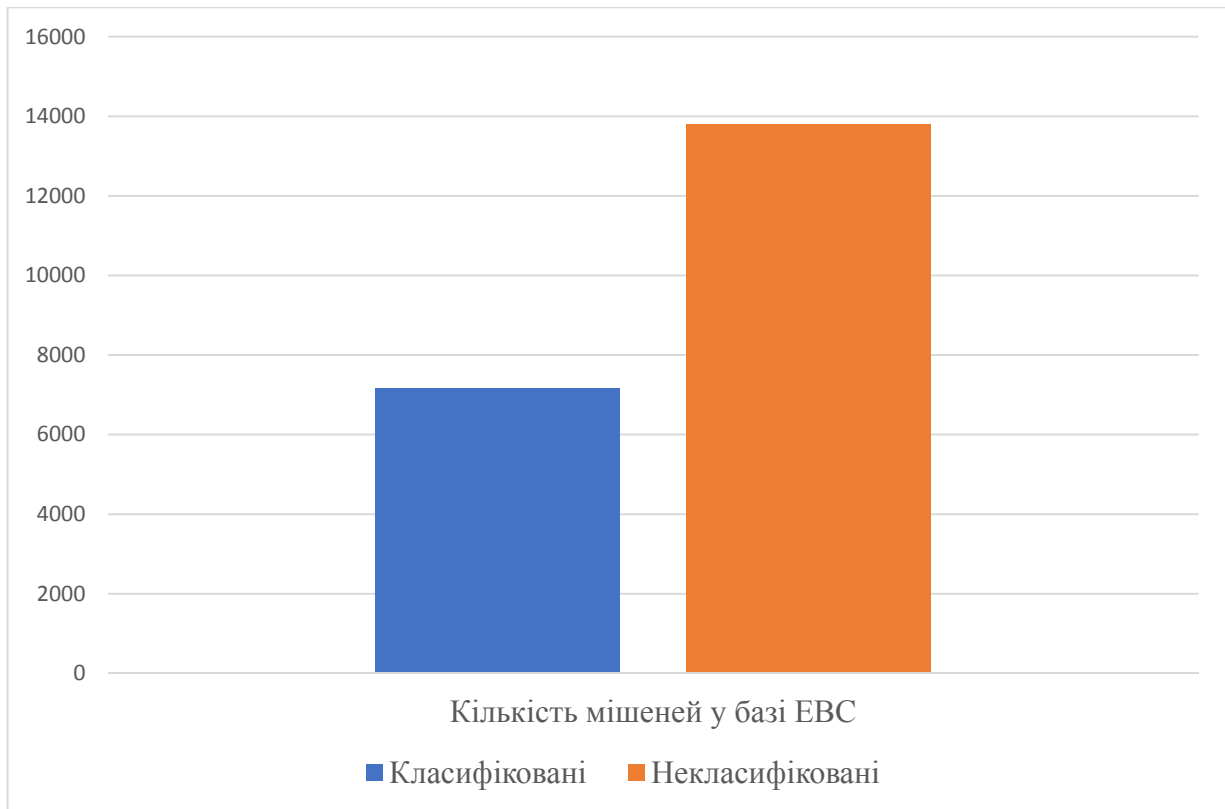


Рисунок 3.2 – Співвідношення класифікованих і некласифікованих мішеней у базі даних EVC

За раніше використовуваною методологією, процес класифікації та уніфікації проводився вручну, що забирало значну кількість робочого часу і ресурсів. Однак, з постійним зростанням обсягу даних і накопиченням нових мішеней, стала очевидною необхідність у вдосконаленні цього процесу. Тому було прийнято рішення розробити інструмент для автоматизованої класифікації та уніфікації білків.

3.2 Розробка інструменту для класифікації мішеней

Отож, в умовах постійного збільшення обсягу даних та необхідності їх ефективного аналізу було обрано метод BLAST [25] (Basic Local Alignment Search Tool), що є одним із найбільш використовуваних інструментів в біоінформатиці для швидкого порівняння амінокислотних або нуклеотидних послідовностей. Інструмент дозволяє здійснювати пошук подібностей між

білковими або нуклеотидними послідовностями, порівнюючи їх із послідовностями з великих баз даних. Методика BLAST базується на локальному вирівнюванні, що дозволяє ідентифікувати найбільш схожі регіони між двома послідовностями без необхідності вирівнювання їх повністю.

BLAST оптимізований для швидкого сканування великих баз даних, що дозволяє отримувати результати в короткі терміни, що було критично важливим для роботи із великою кількістю даних. Метод підтримує різні типи пошуку, включаючи пошук білкових послідовностей у білкових базах даних (BLASTP), нуклеотидних послідовностей у нуклеотидних базах (BLASTN), а також перехресні пошуки, які включають трансляцію нуклеотидних послідовностей в білкові перед порівнянням (BLASTX). BLAST є широко доступним і підтримується численними науковими спільнотами, що гарантує постійне оновлення та підтримку інструменту. Завдяки сучасним алгоритмам і використанню складних матриць, можна ефективно ідентифікувати найбільш значущі регіони схожості між послідовностями, що забезпечує високу точність результатів.

Отже, вибір BLAST як основного інструменту для класифікації таргетів у роботі було зумовлено його високою ефективністю та надійністю у визначенні схожості білкових послідовностей.

Було створено локальну базу даних, яка містить амінокислотні послідовності та UniProtID всіх класифікованих білкових мішеней з дерева EBC (рис 3.3).

Для отримання необхідних сиквенсів було використано API бази даних UniProt[26], що дозволило автоматизувати процес завантаження актуальних даних. Цей підхід забезпечив доступ до оновлених версій сиквенсів, які були необхідні для точного і ефективного BLAST аналізу. Використання API також дозволило фільтрувати сиквенси за специфічними параметрами, такими як тип організму. Завдяки цьому база була доповнена сиквенсами білків мишей і пацюків для збільшення варіабельності та репрезентативності

даних. Таким само чином було викачано амінокислотні послідовності некласифікованих протеїнів.

```

unknown.fasta
>A0A0M5JEZ5
MLSIEEPWEKISSSFDEIFDINVKSYFSGISAALPELKKTNGSVVMTASVSSHAVGGGSCYIASKHAVLGMVKALAYELAPEVRVNAVSPGGAVT
SLCGPASAGFDKMMKMPGIDDIMKGLTPLGFAAKPEDVVAPYLLASRKQKGFITGTVISIDGGMALGRK
>A0A2Z5ENT9
LAGKVAVVTGSGRGIGKAMAIELAKRGAKVAVNYANAVEGAEQVVKIKALGNGSDAAAFKANVGNVEESEKLMDDVVAHFGLDICCNSGVVVF
GHFKDVTPEEFDRVFTINTRGQFFVAKAAYKRMEMGGRIILMGSITGQAKGVPKHAVYSGSKGAIETFRCTMAI
>A0A140DCH8
MLKNIFLKDEIRKVSVPVFIAGAGPSGLTTAALLAKYGVKVLAVSQHSGTANSPRAHITNTRAMEVFRDLGIEEKVYDVASVEEMENVWCTRLAG
QELARLSSWGYGDDRRSEYLTASPSKMCNAAQHVLPEVLLDVARESGAEVLFNTEVLEIEQSSEKYYALVRNRETGEEYIVESDYAVGADGARSIV
ARSLGFEFEGKMGEGHAVNCWFDADLSKYCADRKAVLFPRTLQPGNDWIGGGTFICVKPWNDWILLFMYNPDNADAETNEEAVLERIYTLIGDRDV
DVNIKATISKWTVNMFASKMVKERIVLMGDAHRHPPAGGLGTNTCVQDAFNLSWKLAYIINGYAGQDLLQSYSDERQPVAKQMVNRAITAMVNMQ
AVPQALGFFPGQSEEDGDALVFSPTTEEGKTRRKKMREALELQNWQFNGIGGEAGHHYLSHAIVSDGTTAEFNSSENSQVHIHPTTTTPGATLPH
AWLQRNGVNVSSIDLVGNGKFTIITGLGGDAWLRSAKVSESLGIDVNLVSVSDGDI TDHYGAWADVSEIEDDGCLLVRPDRFVAYRSKGLVSDT
DAAISQAFRAILCL
>P54632
MTREENVYMAKLAEQAERYEEMVEAMKVAELDVELTVEERNLLSVAYKNVIGARRASWRIISSIEQKEESKGNENHVKKIKEYKCKVEKELTDIC
NDILEVLESHLIVSSASGESKVFYKMKGDYFRYLAEFATGNPRKTSAESL IAYKAASDIIVTELPPHTPIRLGLALNFSVFYIEILNSPDACN
LAKTAFDDAIAELDTLSEDSYKSTLIMQLLRDNLTLWTSVHNMEKNQDGGDDQNEPGM
>W0FUS4
MDFSGKNVWVTGAGKIGYATALAFVEAGAKVTGFDQAFTEQEQYPFATEVMDVADAQVAQVCQRLLAETERLDALVNAAGILRMGATDQLSKEDW
QQTFAVNVGGAFNLFQQTMMQFRRQGGAI VTVASDAAHTPRIGMSAYGASKAALKSLALSVGLELAGSGVRCNVVSPGSDTDTMQRTLWVSDDAE
EQRIRGFGEQFKLGIPLGKIARPEIANTILFLASDLASHITLQDIVVDGGSTLGA
>A0A559L2V3
MAALECLRDIGLDRDVEQGDYETASPCQPADIPQTLLEPILIRHATHNKFQVRFDTVLSSTFKDEKTGRIIATLKDKFSNLEYKVGTKYLFADGA
QSQVVKQLDPLERKPGQGLAINVLVKADLSHLVKHRKGNLHWVMPDKEHPEFGWGMGIVRMVWPWDEWMIFFPNRDCDPNIQVSKEEYLQRVRD
FIGDDTQAEILNISKWFVNEIVAERYSDGSIHCLGDAVHRHPPMNGLSNTCIQDAFNLAWKVAYVEKGIASPALLDTYSTERQPVGKSIITRAN
DAFRDHFHVWDAMGALPDKVEDRRKIVDELGSASLQGEARRRRFRDVIQTSHFHLGLIEMGQRYSGPGIYDADEPAPYSHVGAADDPILYYDA
STYPGCRLPHVWLNTSIPSEPVSTIDLGHGKFLFTGIGGEAWKTAAKDVSEN LGVPLDVHVI GYRQQWEDFYFDWQRVGVGVEDSGAVLVRPDRP
PNIVPLFITRGDHMTMECQSPSWREKRTV VTRNLNPKSLDEKHFVQETEAVILMNRLDDPSNFYSYRSLYASSVASILAWGFRAITLDSFWYKD
VSAMIEKWL EATEPGATPPVDLIPWLWYIPGKWKSRVYKMRDHMDKWSOARAMVDDRRARGDRRECMIDMKLDEYKNGWPMSONHAFNNLEGELM

```

Рисунок 3.3 – Файл .fasta, що є локальною базою даних для BLAST

Під час пошуку виникли деякі проблеми, такі як неканонічні назви білків, відсутність деяких сиквенсів у базі даних UniProt або відмінності у геномах менш поширених організмів (рис 3.4). Для вирішення цих проблем ми використовували додаткові джерела даних та алгоритми фільтрації.

Unknown targets	UniprotID	organism_simpl	Sequence
1,2-dihydroxy-1,2-dihydronaphthalene dehydroge...	A0A0M5JEZ5	bacteria	MLSIEEPWEKISSSFDEIFDINVKSYFSGISAALPELKKTNGSVVM...
1,3,8-Trihydroxynaphthalene Reductase	A0A2Z5ENT9	fungi	LAGKVAVVTGSGRGIGKAMAIELAKRGAKVAVNYANAVEGAEQVVK...
1-naphthol 2-hydroxylase	A0A140DCH8	bacteria	MLKNIFLKDEIRKVSVPVFIAGAGPSGLTTAALLAKYGVKVLAVSQ...
14-3-3-like protein	P54632	eukaryota	MTREENVYMAKLAEQAERYEEMVEAMKVAELDVELTVEERNLLSV...
2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	W0FUS4	bacteria	MDFSGKNVWVTGAGKIGYATALAFVEAGAKVTGFDQAFTEQEQYPF...

Рисунок 3.4 – Білки, що не пройшли алгоритм фільтрації за організмами

Для підрахунку оцінки BLAST використовувалася матриця BLOSUM (Blocks Substitution Matrix) 62 (рис 3.5). Ця матриця дозволяє нараховувати бали за співпадіння амінокислот та віднімати їх за неспівпадіння.

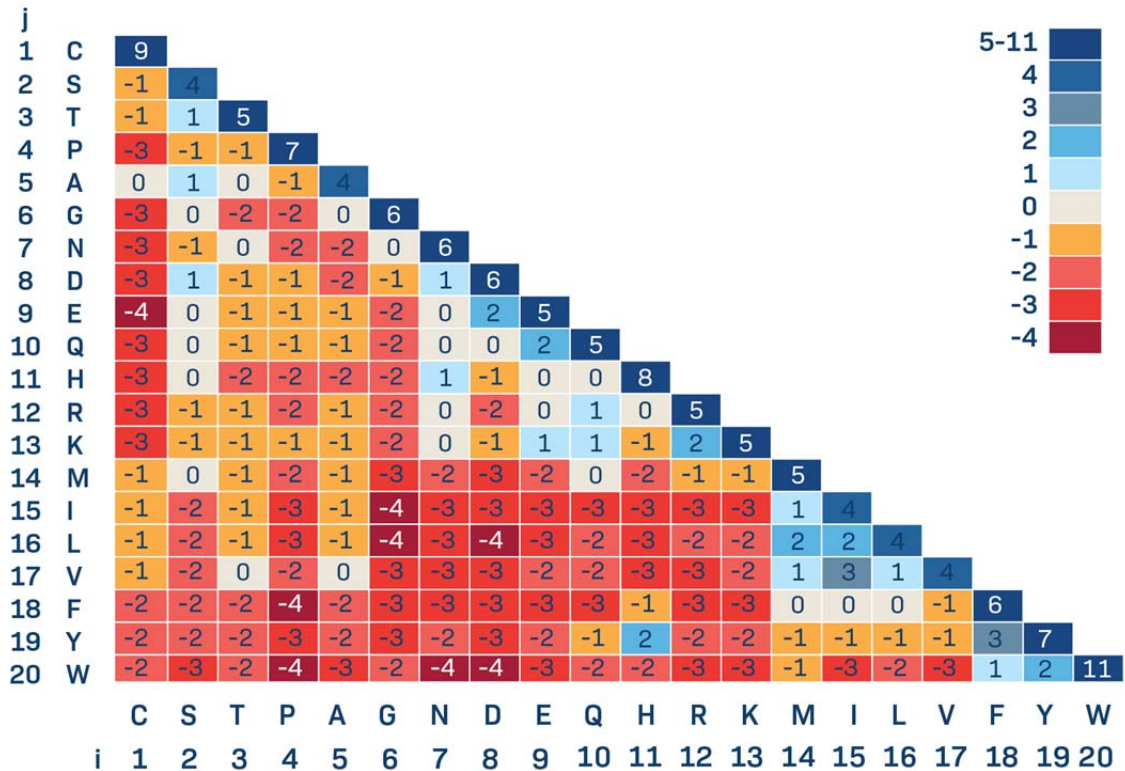


Рисунок 3.5 – Розподіл балів за допомогою матриці BLOSUM 62

Матриця BLOSUM є ключовим інструментом у біоінформатиці для оцінки вирівнювань білкових послідовностей. Назва "BLOSUM62" відноситься до конкретної матриці у сімействі BLOSUM, де "62" означає, що матриця була створена на основі порівнянь сегментів послідовностей, що мають принаймні 62% схожості.

Матриці BLOSUM створюються з використанням блоків послідовностей білків, які представляють вирівняні, консервативні області білків, що не містять вставок чи пропусків. Ці блоки отримуються з багатьох білків із баз даних для формування статистичної основи матриці. Для кожної матриці BLOSUM визначається поріг схожості для блоків, які включаються у аналіз.

Як було зазначено вище, матриця BLOSUM62 містить значення, які представляють "штрафи" або "бонуси" за заміну однієї амінокислоти на іншу

під час процесу вирівнювання амінокислотних послідовностей. Кожен елемент матриці вказує на штраф за заміну однієї амінокислоти на іншу. Якщо амінокислота замінюється на іншу, яка часто зустрічається у подібних білкових послідовностях, ця заміна отримує позитивний бал (бонус). І навпаки якщо заміна відбувається рідко, заміна отримує негативний бал (штраф).

Вибір матриці BLOSUM62 базувався на природі аналізу білків. Для аналізу, де очікувався середній рівень схожості між білками, BLOSUM62 виявилася оптимальною, оскільки матриця дозволяє ефективно оцінювати штрафи та бонуси за заміни, забезпечуючи баланс між чутливістю і специфічністю вирівнювання.

Було успішно здійснено автоматизовану класифікацію та уніфікацію білкових мішеней (рис. 3.6), що значно збільшило ефективність та швидкість аналізу даних.

	in_uniprot	to_uniprot	bitscore	family_name_pred	family_name_true
240353	A0A087WYF4	P41183	139.0	BTB (POZ) domain containing TFs	Zinc finger protein
240358	A0A087WYF4	A0A5K4F1D0	78.2	Zinc finger protein	Zinc finger protein
228684	A0A0A1PA92	A0A0A1PLU1	233.0	3.5.2.6 Beta-lactamase	3.5.2.6 Beta-lactamase
178637	A0A0B2UZH1	A0A1U7R6F4	113.0	Homeobox protein	Homeobox protein
178638	A0A0B2UZH1	A0A6I8VNX9	95.9	Homeobox protein	Homeobox protein
...
70140	X5KCU7	A0A4Y2WSZ8	45.4	CYP3 family	CYP2 family
70142	X5KCU7	Q9UNU6	45.4	CYP5, CYP7 and CYP8 families	CYP2 family
70141	X5KCU7	A0A4Y2WLM5	44.7	CYP3 family	CYP2 family
70143	X5KCU7	A0A8T9B1Q5	44.7	CYP2 family	CYP2 family
70144	X5KCU7	G3V7X8	43.1	CYP24, CYP26 and CYP27 families	CYP2 family

Рисунок 3.6 – Результати передбачень родин білків за допомогою інструменту, що використовує технологію BLAST

Переваги інструменту для класифікації білків:

1. Автоматизований процес значно прискорив класифікацію та уніфікацію мішеней, зменшивши витрати часу та ресурсів.

2. Доповнення локальної бази даних амінокислотними послідовностями не лише людей, а й мишей і щурів, дозволило досягти високого рівня точності у класифікації і знизило ймовірність помилок.
3. Інструмент розроблено з урахуванням можливості масштабування, що дозволяє легко впроваджувати його для обробки навіть найбільших обсягів даних.
4. Також є можливість налаштування під конкретні потреби та специфіку даних, що забезпечує гнучкість та адаптивність інструменту.

Завдяки цьому інструменту була значно покращена ефективність та якість процесів класифікації та уніфікації таргетів. Це дозволяє не лише збільшити продуктивність команди, а й швидше реагувати на новітні відкриття та вимоги ринку, що є ключовим для успіху у сучасному світі інформаційних технологій.

3.3 Відбір потенційних хімічних зондів

Відкриття нових хімічних зондів є критичним кроком у медичних дослідженнях та розробці лікарських засобів. Хімічні зонди, які ефективно взаємодіють з біологічними молекулами, дозволяють глибше зрозуміти біологічні процеси на молекулярному рівні та ідентифікувати потенційні терапевтичні мішені. В цьому контексті, вибір ефективних і специфічних хімічних зондів стає надзвичайно важливим.

У межах даної роботи було зосереджено увагу на відборі хімічних зондів з бази даних "Enamine Bioactive Compounds", що містить понад 10000 біоактивних анотованих сполук. Ці сполуки були ретельно відібрані та проаналізовані за допомогою автоматизованих інструментів для оцінки їхньої біоактивності та селективності. Дані про біоактивність цих сполук були зібрані з різних джерел, що дозволило нам провести всебічний аналіз їхнього потенціалу як хімічних зондів.

Мета цього дослідження полягала у відборі обмеженої кількості сполук, які не тільки проявляють високу біоактивність, але й відрізняються високою селективністю до специфічних білкових таргетів. Встановлені високі критерії для відбору (біоактивність вище за 7 pX та селективність вище за 1.5 pX) дозволяють нам зосередитися на тих сполуках, що мають найбільший потенціал в ролі хімічних зондів.

Як показано на рис 3.7, у базі даних Enamine Bioactive Compounds медіанне значення біоактивності (pX) сполук до всіх доступних мішеней знаходиться між 7 і 8. Також було проведено аналіз біоактивності, що базувався на одному з порогових значень саме для хімічних зондів — 7 pX (рис. 3.8).

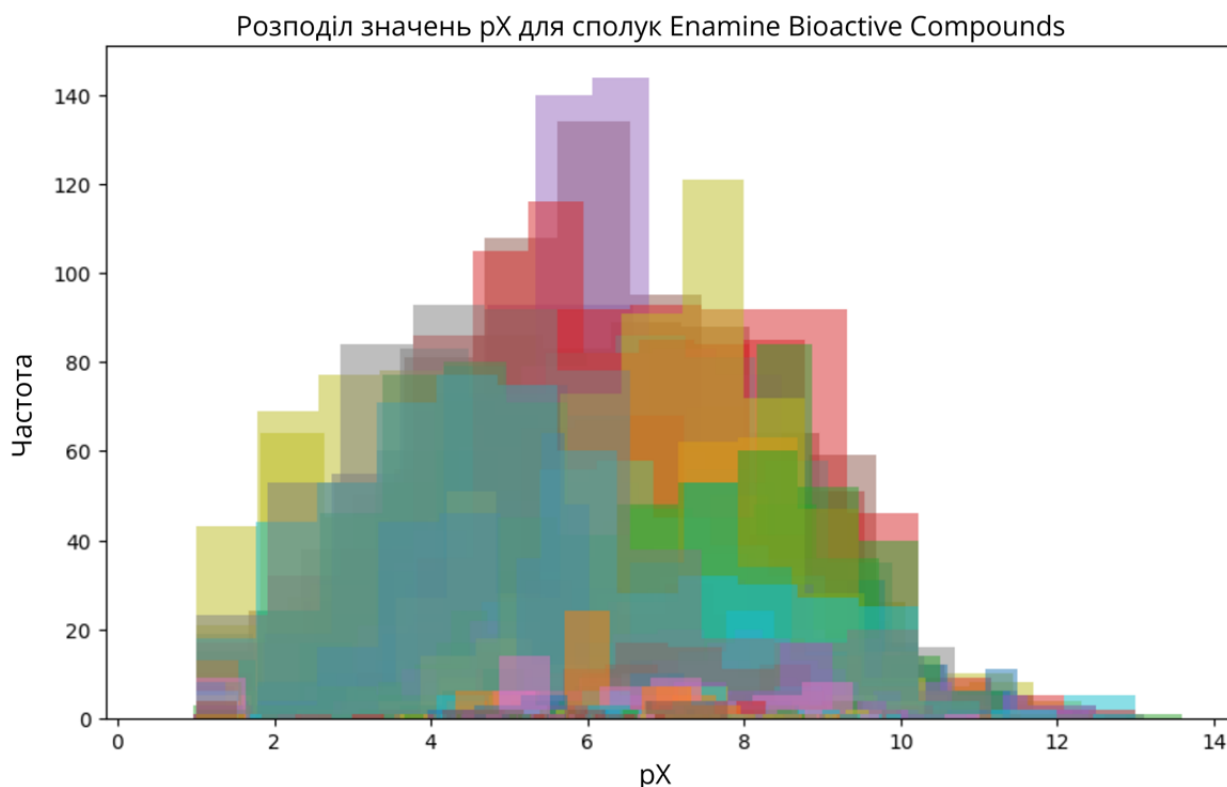


Рисунок 3.7 – Розподіл значень pX для всіх сполук в базі даних Enamine Bioactive Compounds

Можна побачити, що більшість сполук є достатньо біоактивною хоча б до однієї з білкових мішеней, однак не було виявлено жодної сполуки, біоактивність якої до всіх мішеней перевищувала б порогове значення. Після проведення даного аналізу було відсіяно всі сполуки з низьким

значенням біоактивності. Для цього було написано скрипт мовою програмування python, який розділяє модулятори на категорії «активні» та «неактивні». Таким чином було отримано датафрейм, що містив для кожної сполуки: унікальний ідентифікаційний номер, всі можливі білкові мішені, родини білкових мішеней та значення біоактивності (pX) по відношенню до кожної мішені.

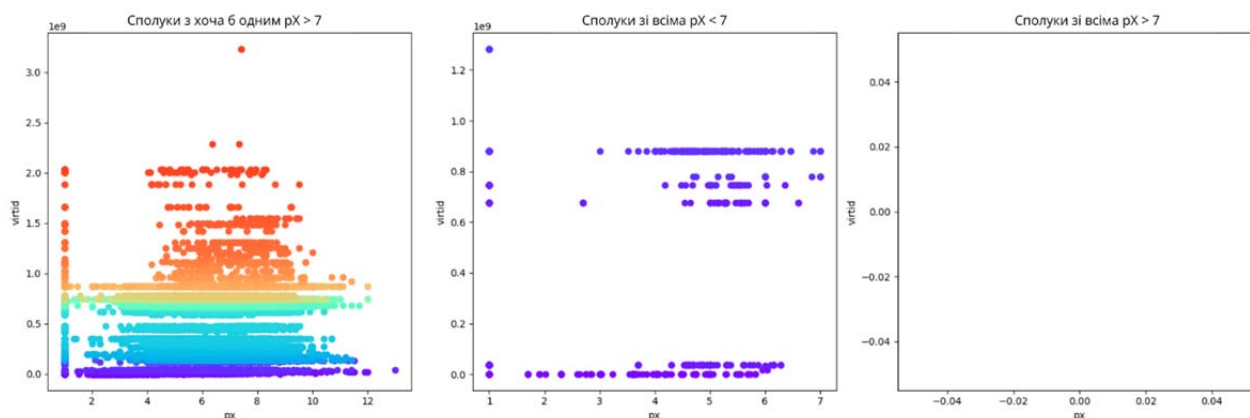


Рисунок 3.8 – Розподіл значень біоактивності (pX) сполук з бази даних Enamine Bioactive Compounds з використанням порогового значення для потенційних хімічних зондів

На наступному кроці необхідно було перевірити чи є сполуки достатньо селективними. Для цього було створено інструмент, що зчитує датафрейм, сортує вже класифіковані мішені за родинями, знаходить всі сполуки, що є біоактивними по відношенню до мішені з конкретної родини білків та залишає у базі потенційних хімічних зондів сполуки, що є селективними до лише одного білка з родини з мінімальним кроком у 1.5 pX. Якщо під час дії алгоритму фільтрації для мішені було відібрано декілька біоактивних та селективних модуляторів, то обирався найбільш активний, а інші відсіювались.

У результаті цього відбору було ідентифіковано 128 потенційних хімічних зондів, які будуть відправлені на додатковий скринінг, щоб підтвердити їх біоактивність і селективність, забезпечуючи таким чином основу для наступних етапів розробки і досліджень.

Розділ 4.

Результати та обговорення

4.1 Огляд методології та основні результати

Для досягнення мети в даній роботі була розроблена і впроваджена комплексна методологія відбору сполук. З бази даних «Enamine Bioactive Compounds», яка містить понад 10,000 біоактивних сполук, було відібрано початковий набір сполук для подальшого аналізу. Сполуки мали бути перш за все добре анотованими. Не менш важливим було визначення строгих критеріїв для відбору сполук, які включали мінімальну біоактивність вище за 7 рХ та селективність вище за 1.5 рХ. Ці критерії забезпечили зосередження на найбільш активних і селективних сполуках, що мають потенціал для застосування в біомедичних дослідженнях.

Методологія відбору сполук, що була застосована в даному дослідженні, показала високу ефективність. Завдяки автоматизації процесу аналізу і сортування стало можливим швидко і точно відфільтрувати сполуки, які не відповідали заданим параметрам, що дозволило зосередитися лише на найбільш перспективних кандидатах. Використання автоматизованих інструментів також забезпечило швидку класифікацію і уніфікацію білкових мішеней, що сприяло ідентифікації некласифікованих білків та поліпшенню точності аналізу.

У наступному підрозділі буде розглянуто аналіз отриманих результатів, включаючи оцінку біоактивності та селективності відібраних сполук, а також обмеження проведеного дослідження.

4.2 Аналіз результатів

У рамках даного дослідження було вжито великих зусиль для ідентифікації і відбору потенційних хімічних зондів, які можуть бути використані в біомедичних дослідженнях та посприяти розробці нових лікарських засобів. Використання бази даних "Enamine Bioactive Compounds" з понад 10000 анотованих біоактивних сполук дозволило ефективно відібрати ті зонди, які відповідають жорстким критеріям біоактивності та селективності, що були встановлені для даної роботи. В результаті було відібрано 128 потенційних хімічних зондів, що задовільняють вищезазначені критерії.

Одним із ключових аспектів роботи було збереження вищих значень біоактивності для сполук, по яких існувала інформація з декількох джерел. Цей підхід забезпечив, що аналіз враховував найбільш переконливі дані, що збільшує ймовірність того, що відібрані сполуки будуть ефективними у біологічних моделях.

Для аналізу отриманої інформації було побудовано графіки розподілу біоактивності та селективності відібраних хімічних зондів (розділ 3, рис 3.8). Графіки показали, що більшість сполук мають значну біоактивність, з середнім значенням > 7 рХ. Розподіл селективності також показав, що більшість відібраних сполук мають селективність вище за 1.5 рХ, що підтверджує їх потенціал для використання в біомедичних дослідженнях.

Проте, робота також мала певні обмеження. Всі дані, які були використані в аналізі, базувались лише на *in vitro* дослідженнях. Це може обмежувати можливість екстраполяції отриманих результатів на *in vivo* умови. Додатково, хоча база даних Enamine надає широкий спектр сполук, вона не обов'язково включає всі можливі біоактивні молекули, які можуть бути важливими для певних білкових мішеней. Також для подальшого підтвердження потенціалу відібраних хімічних зондів необхідно провести їх скринінг біоактивності та дослідження в *in vivo* моделях.

Висновки

1. У рамках даної роботи було успішно відібрано 128 потенційних хімічних зондів з бази даних «Enamine Bioactive Compounds». Відбір проводився на основі строгих критеріїв, які включали мінімальну біоактивність сполуки вище за 7 pX та селективність вище за 1.5 pX. Ці критерії дозволили сконцентруватися на сполуках із значною активністю та специфічністю, що оптимізує їх потенційне використання в подальших біомедичних дослідженнях.
2. Застосування автоматизованого інструменту для класифікації та уніфікації білкових мішеней забезпечило точне та швидке ідентифікування некласифікованих білків. Це також значно підвищило ефективність процесу, дозволяючи сфокусуватися на найбільш перспективних кандидатах.
3. Сфокусований підхід до відбору зменшив кількість сполук, що потребують подальшого тестування, тим самим оптимізуючи використання ресурсів і часу. Приорітизація сполук з найвищою зареєстрованою біоактивністю, що була отримана з різних джерел, допомогла забезпечити високу надійність і точність аналітичних даних.
4. Отримані результати мають важливі наслідки для подальшого розвитку в галузі хімічної біології та розробки лікарських засобів. Відібрані хімічні зонди можуть бути використані для глибших біомедичних досліджень, що потенційно призведе до відкриття нових терапевтичних засобів.
5. Перспективи подальшого розвитку включають проведення додаткових *in vivo* експериментів для підтвердження біоактивності та селективності відібраних сполук. Також необхідно розширити аналіз на інші бази даних біоактивних сполук для підвищення різноманітності та повноти відібраних хімічних зондів. Це сприятиме подальшому удосконаленню

методології та підвищенню ефективності відбору хімічних зондів для біомедичних досліджень.

Список використаних джерел

1. Arrowsmith, C. H., Audia, J. E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P. E., Brown, P. J., Bunnage, M. E., Buser-Doepner, C., Campbell, R. M., Carter, A. J., Cohen, P., Copeland, R. A., Cravatt, B., Dahlin, J. L., Dhanak, D., Edwards, A. M., Frederiksen, M., Frye, S. V., Gray, N., Grimshaw, C. E., Hepworth, D., Howe, T., Huber, K. V., Jin, J., Knapp, S., Kotz, J. D., Kruger, R. G., Lowe, D., Mader, M. M., Marsden, B., Mueller-Fahrnow, A., Müller, S., O'Hagan, R. C., Overington, J. P., Owen, D. R., Rosenberg, S. H., Ross, R., Roth, B., Schapira, M., Schreiber, S. L., Shoichet, B., Sundström, M., Superti-Furga, G., Taunton, J., Toledo-Sherman, L., Walpole, C., Walters, M. A., Willson, T. M., Workman, P., Young, R. N., and Zuercher, W. J. (2015) The promise and peril of chemical probes. *Nature Chemical Biology* 11, 536–541.
2. Rowe, C. C., Ellis, K. A., Rimajova, M., Bourgeat, P., Pike, K. E., Jones, G., Frupp, J., Tochon-Danguy, H., Morandau, L., O'Keefe, G., Price, R., Raniga, P., Robins, P., Acosta, O., Lenzo, N., Szoeker, C., Salvado, O., Head, R., Martins, R., Masters, C. L., Ames, D., and Villemagne, V. L. (2010) Amyloid imaging results from the Australian imaging, biomarkers and lifestyle (AIBL) study of aging. *Neurobiology of Aging* 31, 1275–1283.
3. Hartung, I. V., Rudolph, J., Mader, M. M., Mulder, M. P., and Workman, P. (2023) Expanding chemical probe space: Quality criteria for covalent and Degradable probes. *Journal of Medicinal Chemistry* 66, 9297–9312.
4. Bunnage, M. E., Chekler, E. L., and Jones, L. H. (2013) Target validation using chemical probes. *Nature Chemical Biology* 9, 195–199.
5. Müller, S., Ackloo, S., Al Chawaf, A., Al-Lazikani, B., Antolin, A., Baell, J. B., Beck, H., Beedie, S., Betz, U. A., Bezerra, G. A., Brennan, P. E., Brown, D., Brown, P. J., Bullock, A. N., Carter, A. J., Chaikuad, A., Chaineau, M., Ciulli, A., Collins, I., Dreher, J., Drewry, D., Edfeldt, K.,

- Edwards, A. M., Egner, U., Frye, S. V., Fuchs, S. M., Hall, M. D., Hartung, I. V., Hillisch, A., Hitchcock, S. H., Homan, E., Kannan, N., Kiefer, J. R., Knapp, S., Kostic, M., Kubicek, S., Leach, A. R., Lindemann, S., Marsden, B. D., Matsui, H., Meier, J. L., Merk, D., Michel, M., Morgan, M. R., Mueller-Fahrnow, A., Owen, D. R., Perry, B. G., Rosenberg, S. H., Saikatendu, K. S., Schapira, M., Scholten, C., Sharma, S., Simeonov, A., Sundström, M., Superti-Furga, G., Todd, M. H., Tredup, C., Vedadi, M., von Delft, F., Willson, T. M., Winter, G. E., Workman, P., and Arrowsmith, C. H. (2022) Target 2035 – update on the quest for a probe for every protein. *RSC Medicinal Chemistry* 13, 13–21.
6. Baell, J. B., and Nissink, J. W. (2017) Seven Year itch: Pan-assay interference compounds (pains) in 2017—utility and limitations. *ACS Chemical Biology* 13, 36–44.
 7. Workman, P., and Collins, I. (2010) Probing the probes: Fitness factors for small molecule tools. *Chemistry & Biology* 17, 561–577.
 8. Efimova, E. V., Ricco, N., Labay, E., Mauceri, H. J., Flor, A. C., Ramamurthy, A., Sutton, H. G., Weichselbaum, R. R., and Kron, S. J. (2018) HMG-COA reductase inhibition delays DNA repair and promotes senescence after tumor irradiation. *Molecular Cancer Therapeutics* 17, 407–418.
 9. Zhang, J., Ma, C., Yu, Y., Liu, C., Fang, L., and Rao, H. (2023) Single amino acid-based PROTACs trigger degradation of the oncogenic kinase BCR–ABL in chronic myeloid leukemia (CML). *Journal of Biological Chemistry* 299, 104994.
 10. Watanabe, H., Ono, M., and Saji, H. (2015) Novel PET/SPECT probes for imaging of tau in Alzheimer's disease. *The Scientific World Journal* 2015, 1–6.
 11. Franco, A. R., Artusa, V., and Peri, F. (2023) Use of fluorescent chemical probes in the study of toll-like receptors (TLRs) trafficking. *Methods in Molecular Biology* 57–74.

12. Park, J. M., Wu, M., Datta, K., Liu, S.-C., Castillo, A., Lough, H., Spielman, D. M., and Billingsley, K. L. (2017) Hyperpolarized sodium [1-¹³C]-glycerate as a probe for assessing glycolysis in vivo. *Journal of the American Chemical Society* 139, 6629–6634.
13. Antolin, A. A., Ameratunga, M., Banerji, U., Clarke, P. A., Workman, P., and Al-Lazikani, B. (2020) The kinase polypharmacology landscape of clinical PARP inhibitors. *Scientific Reports* 10.
14. Tacar, S. Y., Yilmaz, M., Oz, B., and Tural, D. (2021) Crizotinib for *c-met*-amplified advanced NSCLC: A single-center experience. *Tumori Journal* 108, 258–262.
15. Brugières, L., Cozic, N., Houot, R., Rigaud, C., Sibon, D., Arfi-Rouche, J., Bories, P., Cottreau, A. S., Delmer, A., Ducassou, S., Garnier, N., Lamant, L., Leruste, A., Millot, F., Moalla, S., Morschhauser, F., Nolla, M., Pagnier, A., Reguerre, Y., Renaud, L., Schmitt, A., Simonin, M., Verschuur, A., Hoog Labouret, N., Mahier Ait Oukhatar, C., and Vassal, G. (2023) Efficacy and safety of crizotinib in Alk-positive systemic anaplastic large-cell lymphoma in children, adolescents, and adult patients: Results of the French ACSÉ-crizotinib trial. *European Journal of Cancer* 191, 112984.
16. Arter, C., Trask, L., Ward, S., Yeoh, S., and Bayliss, R. (2022) Structural features of the protein kinase domain and targeted binding by small-molecule inhibitors. *Journal of Biological Chemistry* 298, 102247.
17. To, K. K., Xing, E., Larue, R. C., and Li, P.-K. (2023) BET bromodomain inhibitors: Novel design strategies and therapeutic applications. *Molecules* 28, 3043.
18. Campbell, K. J., and Tait, S. W. (2018) Targeting bcl-2 regulated apoptosis in cancer. *Open Biology* 8, 180002.
19. Gordovez, F. J., and McMahon, F. J. (2020) The genetics of bipolar disorder. *Molecular Psychiatry* 25, 544–559.
20. Young, A. H., and Juruena, M. F. (2020) The neurobiology of bipolar disorder. *Bipolar Disorder: From Neuroscience to Treatment* 1–20.

21. Kalkman, H. O. (2012) Potential opposite roles of the extracellular signal-regulated kinase (ERK) pathway in autism spectrum and bipolar disorders. *Neuroscience & Biobehavioral Reviews* 36, 2206–2213.
22. Lett, T. A., Zai, C. C., Tiwari, A. K., Shaikh, S. A., Likhodi, O., Kennedy, J. L., and Müller, D. J. (2011) *ANK3*, *cacna1c* and *znfx804* gene variants in bipolar disorders and psychosis subphenotype. *The World Journal of Biological Psychiatry* 12, 392–397.
23. Márquez, L. C., Churchhouse C, García, O. H., Gómez, V. M., Burt NP, Aguilar, S. C., González, V. C., Florez JC, Orozco L, Haiman CA, Tusié, L. T., and Altshuler D. (2013) Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico. *Nature* 506, 97–101.
24. Bioactive libraries. *Enamine Bioactive Compounds*.
25. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.L., (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421.
26. Coudert, E., Gehant, S., de Castro, E., Pozzato, M., Baratin, D., Neto, T., Sigrist, C. J., Redaschi, N., Bridge, A., Bridge, A. J., Aimo, L., Argoud-Puy, G., Auchincloss, A. H., Axelsen, K. B., Bansal, P., Baratin, D., Neto, T. M., Blatter, M.-C., Bolleman, J. T., Boutet, E., Breuza, L., Gil, B. C., Casals-Casas, C., Echioukh, K. C., Coudert, E., Cuche, B., de Castro, E., Estreicher, A., Famiglietti, M. L., Feuermann, M., Gasteiger, E., Gaudet, P., Gehant, S., Gerritsen, V., Gos, A., Gruaz, N., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Kerhornou, A., Le Mercier, P., Lieberherr, D., Masson, P., Morgat, A., Muthukrishnan, V., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Poux, S., Pozzato, M., Pruess, M., Redaschi, N., Rivoire, C., Sigrist, C. J., Sonesson, K., Sundaram, S., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., A-Jee, H. B., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Joshi, V., Jyothi, D., Kandasamy,

S., Lock, A., Luciani, A., Lugaric, M., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Mishra, A., Moulang, K., Nightingale, A., Pundir, S., Qi, G., Raj, S., Raposo, P., Rice, D. L., Saidi, R., Santos, R., Speretta, E., Stephenson, J., Tootoo, P., Turner, E., Tyagi, N., Vasudev, P., Warner, K., Watkins, X., Zaru, R., Zellner, H., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., and Wang, Y. (2022) Annotation of biologically relevant ligands in uniprotkb using chebi. *Bioinformatics* 39.