

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

Завідувач кафедри молекулярної біотехнології та біоінформатики
доцент Нипорко О.Ю.

Протокол № засідання кафедри

від “__” _____ 2024 р.

ДОСЛІДЖЕННЯ НОВИХ ПОТЕНЦІЙНИХ ЛІГАНДІВ CRBN
МЕТОДОМ ВІРТУАЛЬНОГО СКРІНІНГУ

Випускна кваліфікаційна робота магістра
студентки спеціальності 091

Біологія

ОП «Біоінформатика і структурна біологія»

Бештинарської Анни Олександрівни

Науковий керівник від кафедри

доцент кафедри молекулярної

біотехнології та біоінформатики

к.б.н Нипорко Олексій Юрійович

Робота виконана у НВП «Снамін»

під керівництвом директора к.б.н. Бориско П.О.

Оцінка захисту роботи

Київ – 2024р.

АНОТАЦІЯ

Субстратний рецептор E3-лігазного комплексу CRBN має великий потенціал як інструмент таргетного протеасомного деградування білків. Лише нещодавно відкривши та зрозумівши принцип, за яким CRBN впливає на метаболічні процеси в клітинах, дослідники активно почали розробляти інструменти для модуляції його роботи. Наша робота концентрується на пошуку нових лігандів CRBN, які можуть бути використані для розробки молекулярного клею чи молекул PROTAC. За результатами роботи було знайдено 5 лігандів, які показали високу оцінку стикування та інші фізико-хімічні параметри.

The substrate receptor of the E3-ligase complex CRBN has great potential as a tool for targeted proteasomal degradation of proteins. Recently discovered and understood the principle by which CRBN affects metabolic processes in cells, researchers have actively begun to develop tools to modulate its work. Our work is focused on finding new CRBN ligands that can be used to develop a molecular glue or PROTAC molecule. According to the results of the work, 5 ligands were found, which showed a high assessment of docking and other physicochemical parameters.

Ключові слова: Cereblon, CRBN, E3-лігаза, PROTAC, віртуальний скринінг, молекулярний докінг.

ЗМІСТ

| | |
|---|----|
| ВСТУП | 5 |
| РОЗДІЛ 1. Огляд літератури..... | 6 |
| 1.1 Cereblon, як субстратний рецептор E3-лігазного комплексу..... | 6 |
| 1.2 Молекулярна структура CRBN..... | 9 |
| 1.3 Огляд існуючих лігандів CRBN..... | 13 |
| 1.4 Використання лігандів CRBN для створення молекул PROTAC.... | 14 |
| 1.5 Дослідження лігандів CRBN методом віртуального скринінгу..... | 15 |
| РОЗДІЛ 2. Експериментальна частина | 17 |
| 2.1. Пошук кристалічних структур CRBN в біоінформатичних базах даних..... | 17 |
| 2.2 Огляд кристалічних структур білка CRBN..... | 19 |
| 2.3 Підбір речовин для контрольної вибірки..... | 21 |
| 2.4 Молекулярний докінг контрольної вибірки сполук..... | 23 |
| 2.5 Побудова та аналіз ROC кривих..... | 24 |
| 2.6 Порівняння повторного стикування лігандів..... | 26 |
| 2.7 Кореляційний аналіз оцінки стикування та інгібування сполук.... | 28 |
| 2.8 Віртуальний скринінг бібліотеки..... | 29 |
| РОЗДІЛ 3. Результати досліджень та їхнє обговорення..... | 32 |
| ВИСНОВКИ | 40 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ | 41 |

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

CRBN – Cereblon.

CRL – Cullin-RING ligase.

PROTAC – Proteolysis-targeting chimera.

IMiD – Immunomodulatory drug.

TDB – Thalidomide binding domain.

RCSB PDB – Research Collaboratory for Structural Bioinformatics Protein Data Bank.

IC50 – 50% inhibitory concentration.

ROC – receiver operating characteristic.

AUC – area under curve.

RMSD – Root mean square deviation.

ВСТУП

Пошук нових лікарських препаратів є одним з постійних завдань біології та хімії. Поєднуючись у таких науках, як біохімія та біоінформатика, вони надають нам нові інструменти для пошуку та розробки лікарських засобів, які будуть більш селективним та афінними або узагалі вражати абсолютно нові мішені.

Одним з нещодавно відкритих способів впливати на метаболічні процеси є використання білка CRBN. Ця мішень бере участь у регуляції багатьох метаболічних процесах. CRBN є субстратним рецептором E3-лігазного комплексу, який є частиною системи протеасомного деградування білків. Модулюючи роботу субстратного рецептору можна не лише пригнічувати роботу комплексу, а і впливати на його субстратну специфічність, тим самим змінюючи мішені для убіквітинування.

Це створює нові можливості для лікування не лише захворювань, що безпосередньо пов'язані з функціонуванням CRBN, а також широкого спектру інших захворювань, пов'язаних з надмірною або небажаною продукцією білків, завдяки розробці таких інструментів, як молекулярний клей або молекули PROTAC. Дослідження лігандів CRBN може сприяти розробці засобів для лікування різноманітних нейродегенеративних, генетичних та онкологічних захворювань, для яких на даний момент не існує лікування.

РОЗДІЛ 1

ОГЛЯД ЛІТЕРАТУРИ

Cereblon, або CRBN - субстратний рецептор у складі E3-лігазного комплексу, що відповідає за протеасомну деградацію білків за участі 26S протеасом [1], [2]. CRBN відіграє ключову роль у багатьох метаболічних процесах, що опосередковує його імуномодельючі або тератогенні властивості [1], [3], [4]. Це дозволяє, впливаючи на CRBN за допомогою різноманітних лігандів, знаходити нові підходи у лікуванні різноманітних захворювань. Окрім цього, CRBN можна використовувати для мобілізації E3-лігазного комплексу та цільового протеасомного деградування білків, що не є його первинними субстратами. Одним з таких підходів є використання молекул PROTAC, які зв'язують з CRBN на одному кінці молекули та з таргетним білком з іншого. Схожим підходом є розробка молекулярного клею – молекул, які змінюють поверхню зв'язування CRBN з мішенями таким чином, щоб викликати зв'язування неосубстрату [5], [6]. За рахунок цього E2-лігаза здійснює убіквітинування таргетного білка, що призводить до протеасомної деградації білка-мішені [7]. Усе це робить пошуки різноманітних лігандів CRBN актуальною задачею та перспективною областю досліджень, що в результаті може змінити наші підходи до лікування багатьох захворювань.

1.1 Cereblon, як компонент E3-лігазного комплексу

E3-лігазний комплекс – білкова система, що складається з лігаз, які здійснюють убіквітинування білка-мішені. Зазвичай, самою E3-лігазою називають групу білків, як зв'язують таргетний білок на одному кінці та E2 лігазу на іншому [8]. За посередництва E2-E3-лігазного комплексу убіквітин переноситься від E1-лігази на субстрат, яким виступає білок-мішень [8], [9].

Cullin-RING убiквітиновi лiгази (CRL) представляють собою найбільшу родину E3-лiгаз, що бере участь у сигнальній трансдукції, регуляції клітинного циклу, канцерогенезі та ембріональному розвитку [1], [10]. E3-лiгазний комплекс, у складі якого присутній CRBN, має назву CRL4^{CRBN}.

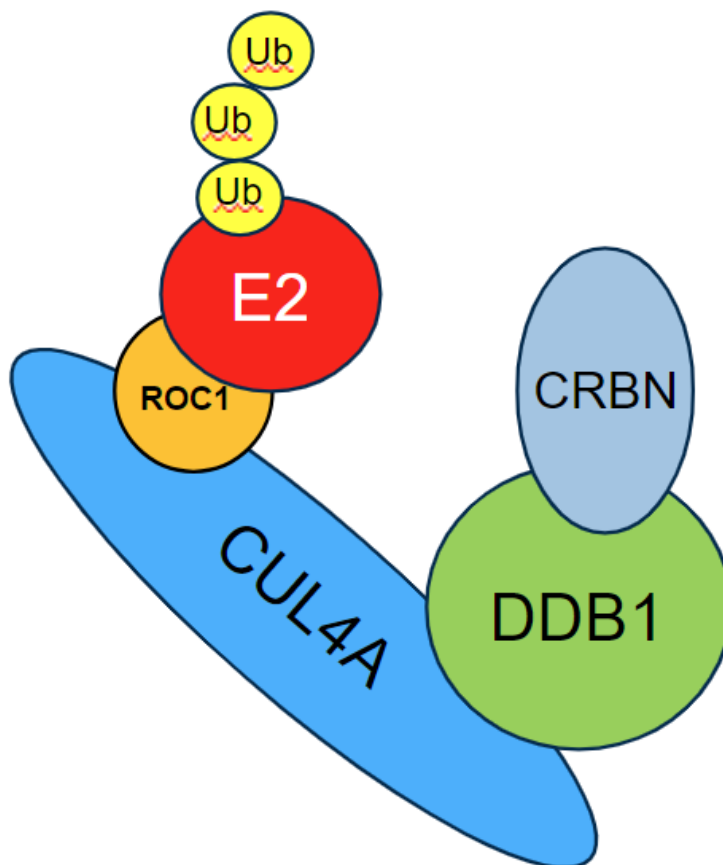


Рисунок 1.1 – схематичне зображення структури E3-лiгазного комплексу (CRBN, DDB1, CUL4A та ROC1 (RBX1)) разом з E2-лiгазою та убiквітином.

До складу комплексу, структуру якого представлено на рисунку 1.1, входять CRBN, DDB1, CUL4A та RBX1. RBX1, або ROC1 – білок, що забезпечує зв'язування E2-лiгази з комплексом за допомогою RING домену, що включає структуру «цинковий палець», а також виступає регулятором cullins1 [2], [9], [11]. Cullin4 (CUL4A/CUL4B) відіграє роль скефолд-білка, до якого байндяться інші структурні компоненти комплексу. Саме з ним зв'язується RBX1 та DDB1 (DNA binding protein), який містить в своїй структурі сайт зв'язування DCAF (DDB1–cullin4–associated factor). Подальші

дослідження показали, що одним з DCAF є CRBN [12]. DDB1 виступає свого роду корецептором, хоча і сам може виступати субстратним рецептором E3-лігази. Така властивість DDB1 може бути пов'язана із аутоубіквітуванням CRBN комплексом [13]. Кристалічну структуру комплексу продемонстровано на малюнку 1.2.

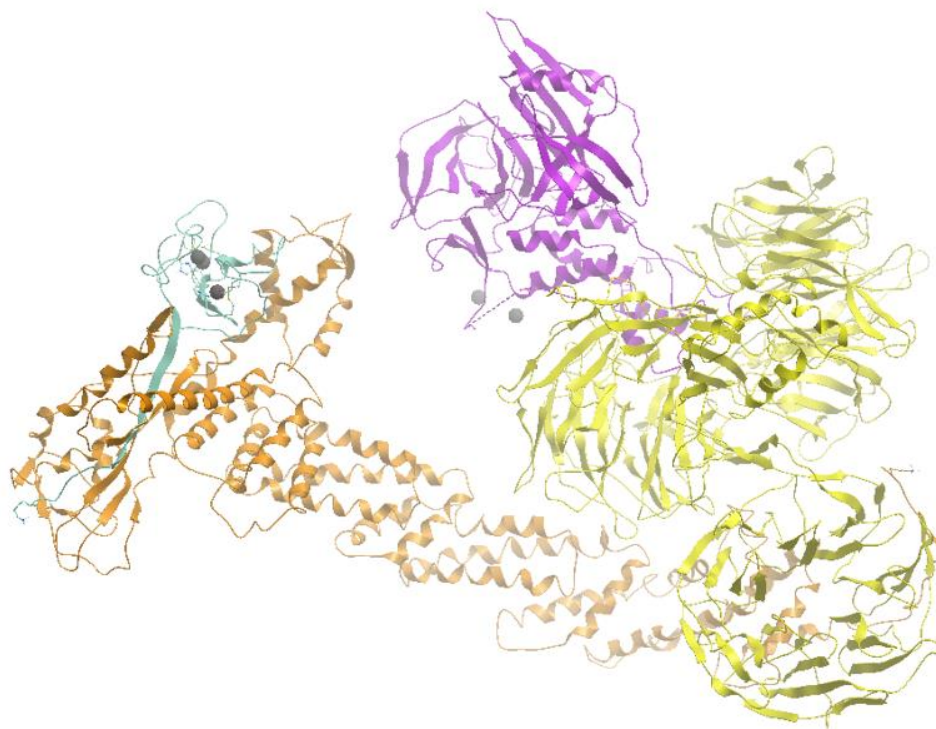


Рисунок 1.2 – кристалічна структура E3-лігазного комплексу. Зображення складено з 2 кристалів – 8OIZ та 2HYE [14], [15], що були об'єднані за допомогою суперпозиції DDB1. Голубим позначено RBX1, оранжевим – CUL4A, жовтим – DDB1, фіолетовим – CRBN.

Cereblon у свою чергу виступає субстратним рецептором комплексу, зв'язуючись з різноманітними субстратами, в ролі яких виступають білки-мішені. Він має цитоплазматичну, мембранну та ядерну локалізацію [1], [11], що пояснює його вплив на різні білки цитоплазми та транскрипційні фактори. До прикладу він опосередковує протеасомну деградацію MEIS2, що виступає транскрипційним фактором, який регулює ембріональний розвиток, зокрема ріст кінцівок та зумовлює тератогенний вплив CRBN [1]. Проте використовуючи різноманітні ліганди ми можемо впливати на субстратну

специфічність CRBN та зв'язувати з лігазним комплексом неосубстрати. Як приклад можна привести зв'язування білка з IMiDs, такими як помалідомід, що призводить до зміни білка-мішені на IKZF1 – білок пов'язаний з розвитком множинної мієломи [7], [16]. З використанням молекул PROTAC з'являється ще більше можливостей для цільового убіквітування білків, оскільки, завдяки своїй варіативності, PROTAC можна використовувати для великої кількості білків, у тому числі таких, що не піддаються інгібуванню у інший спосіб.

1.2 Молекулярна структура CRBN

Cereblon складається з 442 амінокислот та має молекулярну масу ~51kDa. Білок містить 2 домени – Lon-подібний протеазний домен розташований на N-кінці та CUL7 домен, що розміщений на C-кінці.

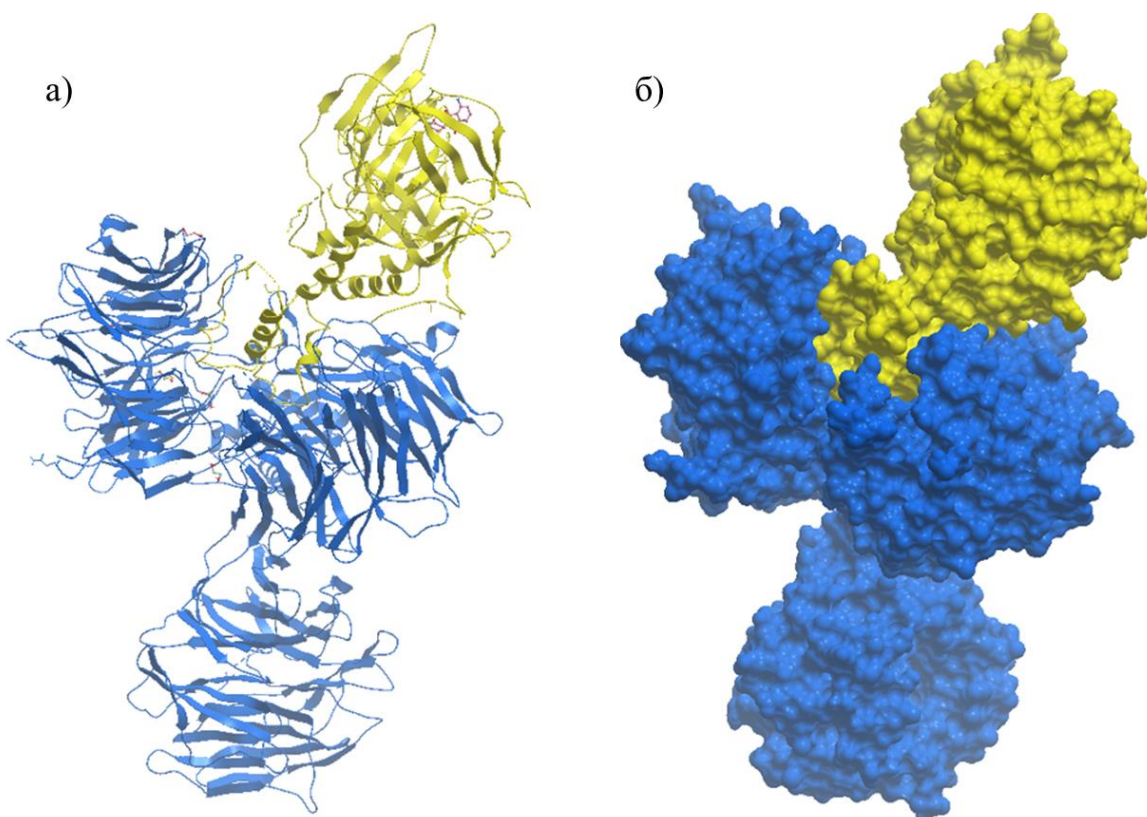


Рисунок 1.3 – зв'язування CRBN позначено жовтим, та DDB1 – синім. а) структура у вигляді стрічки; б) структура в вигляді поверхні зв'язування.

Lon домен розміщений в межах 76-318 амінокислоти та має протеазоподібну структуру з петлею, що опосередковує зв'язування CRBN з DDB1 [12]. Разом з цим варто відзначити, що CRBN не містить структури WD-повтору, характерну для інших представників групи DCAF. Безпосередньо зв'язується із DDB1 серія спіралей, що утворюються амінокислотами зі 188 по 248, а саме структура спіраль-поворот-спіраль (амінокислоти 221-248), ланцюг (амінокислоти 191-197) та спіраль (амінокислоти 198-209) [12], що можна побачити на рисунку 1.3.

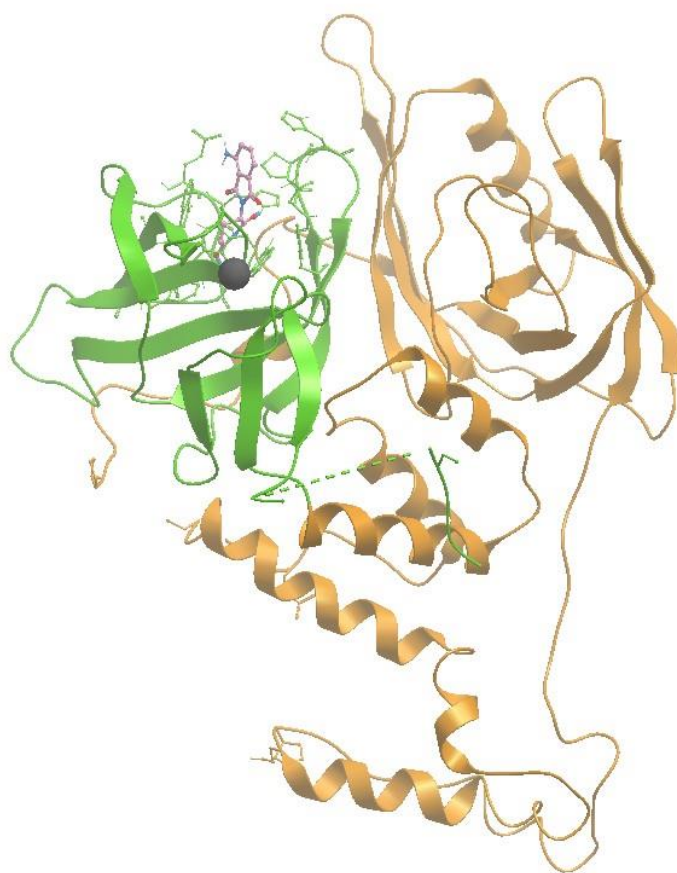


Рисунок 1.4 – Кристалічна структура CRBN. Lon домен позначено оранжевим, CULT – зеленим.

CULT домен, або Талідомід зв'язуючий домен (TBD) включає в себе амінокислоти 318-426. Саме тут розташований сайт зв'язування IMiDs та клітинних ліганди [1]. Домен розміщується дистально від сайту зв'язування DDB1 та займає меншу частину білкової молекули, що можна побачити на рисунку 1.4. Домен має структурну гомологію до

метіонінсульфоксидредуктази та до двониткових РНК доменів [12].

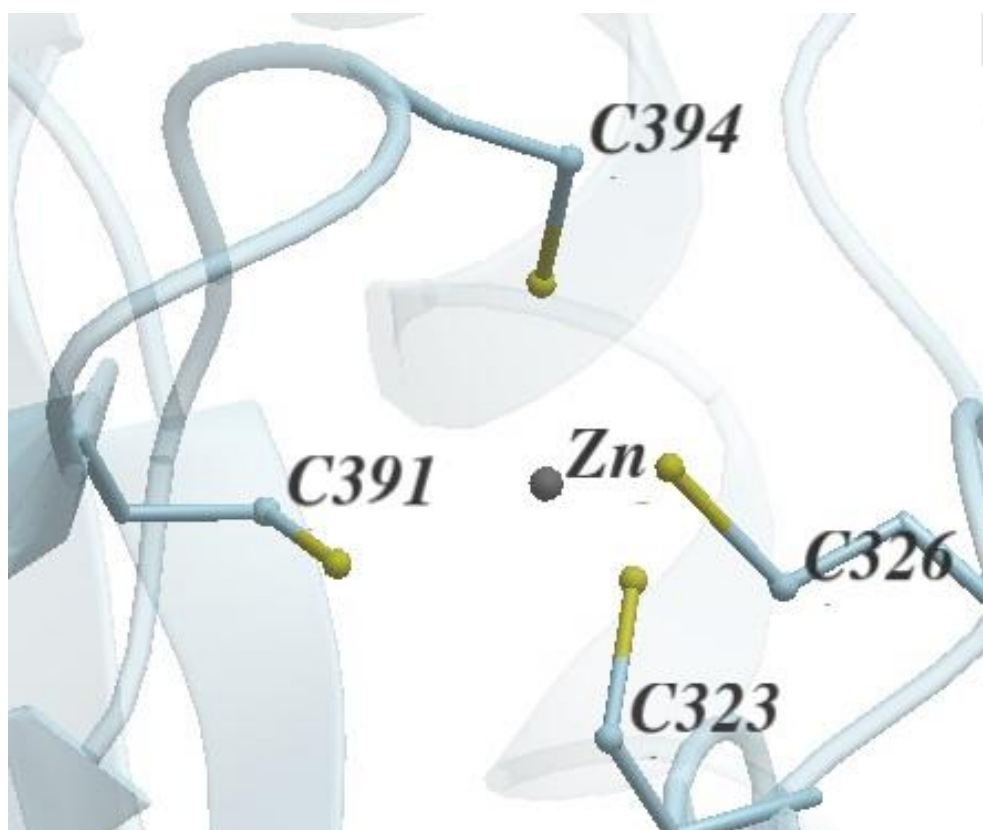


Рисунок 1.5 – амінокислоти, що координують атом цинку у структурі CRBN.

Домен на С кінці складається з 6 антипаралельних β -складчастих структур ($\beta 3$ – $\beta 4$ – $\beta 8$ – $\beta 7$ – $\beta 6$ – $\beta 5$), що утворюють ядро, яке включає Zn. Як показано на рисунку 1.5, іон координується 4 цистеїнами – Cys323, Cys326, Cys391 та Cys394 та знаходиться на відстані ~ 18 Å від сайту зв'язування IMiDs [17].

Невеликий гідрофобний карман зв'язує талідомід та його похідні водневими зв'язками з трьома триптофанами – Trp380, Trp386 та Trp400 та Phe402. В межах кишені глютаримідне кільце талідоміду утворює 2 водневі зв'язки з основами His378 та Trp380 та третій водневий зв'язок з бічним ланцюгом His378, як показано на рисунку 1.6. Також, один з оксигенів у структурі ізоіндолінового кільця орієнтований в сторону His357, проте відстань між ними занадто велика для утворення водневого зв'язку.

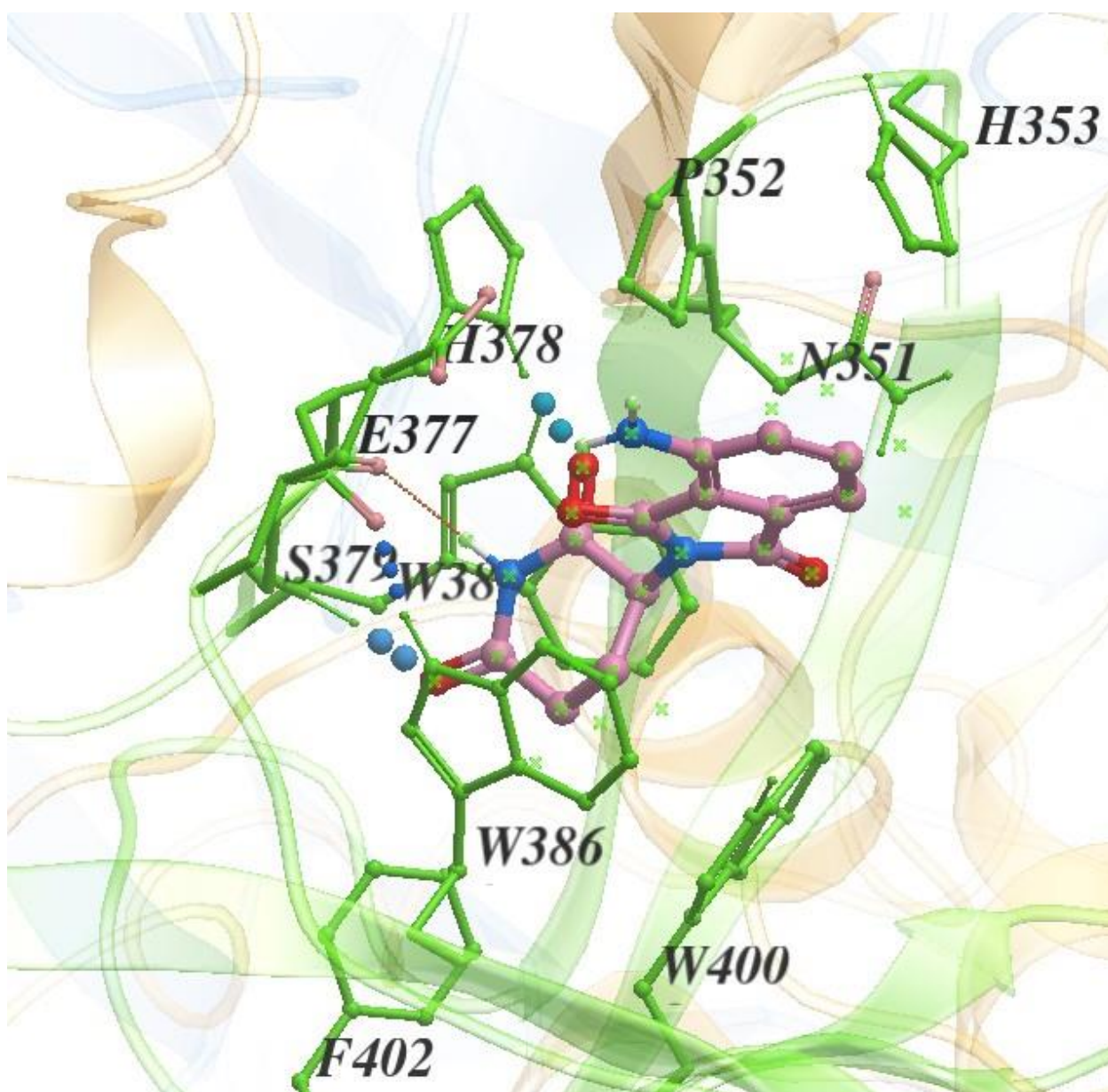


Рисунок 1.6 – структура сайту зв’язування IMiDs. Рожевим позначено помалідомід. Пунктирною лінією позначено водневі зв’язки.

TBD є дуже консервативною ділянкою білка, яка зберігає свою структуру у багатьох видів. Мутації у цій ділянці призводять до зниження зв’язування IMiDs та CRBN. Це спостерігається у гризунів, що, маючи 4 мутації у цій ділянці, не проявляють чутливості до IMiDs [3], [12], [18].

Саме завдяки зв’язуванню TBD доменом різних лігандів, змінюється поверхня зв’язування субстратів E3-лігазного комплексу, що за підбору правильних лігандів дає можливість регулювати субстратну специфічність та спрямовувати роботу комплексу на убіквітинування обраних мішеней.

1.3 Огляд існуючих лігандів CRBN

Першим відкритим лігандом CRBN був талідомід. Молекула, що складається з двох функціональних частин – глутаримідного кільця та фталімідного кільця, як показано на рисунку 1.7. Саме глутаримідне кільце відповідає за зв'язування молекули з гідрофобною кишенню.

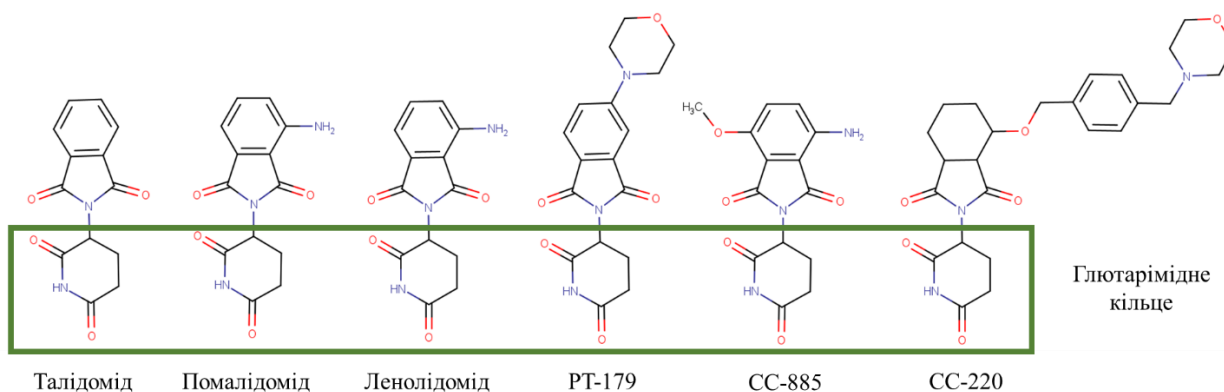


Рисунок 1.7 – структура талідоміду та приклади його аналогів.

Більшість інших лігандів CRBN на даний момент є похідними талідоміду і відрізняються структурами, що з'єднані з глутаримідним кільцем. Ми можемо бачити, що найвідоміші похідні талідоміду – помалідомід та ленолідомід, відрізняються від нього лише наявністю додаткової NH_2 групи та зміненою кількістю зв'язаних оксигенів. В свою чергу глутаримідне кільце зазвичай не піддається змінам.

У пошуках нових лігандів CRBN дослідники зазвичай використовують талідомід як основу для створення власних похідних на його основі [8], [19], [20], [21], [22]. Використання глутаримідного кільця як основи обумовлене розташуванням донорних та акцепторних груп таким чином, що вони утворюють три водневі зв'язки з амінокислотами гідрофобної кишені. His378 утворює два водневі зв'язки – один з киснем та один з NH групою. Третій водневий зв'язок утворюється між киснем глутаримідного кільця та Trp380 [21]. Це визначає стабільність утвореного між молекулами зв'язку та зумовлює популярність талідоміду, як основи для створення нових талідомідо-подібних лігандів.

На даний момент представлено дуже мало досліджень про не-талідомідо-подібні ліганди CRBN [23], [24]. Проте ті що є, показують, що напрямок має потенціал для розвитку. У своїх дослідженнях автори наводять результати, що відтворюються у тестах *in vitro* та мають хорошу активність на низьких концентрація. Така мала кількість досліджень цього питання приводить нас до необхідності пошуку нових можливих лігандів CRBN, що в подальшому може привести до відкриття нових сполук з кращими властивостями та розробки нових підходів до створення афекторів CRBN.

1.4 Використання лігандів CRBN для створення молекул PROTAC

При створенні молекул PROTAC – proteolysis-targeting chimera, за основу береться ідея таргетного деградування певних субстратів, якими найчастіше виступають білки. PROTAC є гетерофункціональною малою молекулою, що складається з 2 лігандів білків та лінкера, що їх з'єднує [5]. Один з лігандів зв'язує білок, який підлягає протеасомному деградуванню, у той час як інший зв'язується з субстратним рецептором E3-лігази, яким, в нашому, випадку виступає CRBN.

Перший PROTAC було розроблено у 2001 році. Він зв'язував таргетний білок з лігазним комплексом SCF, що призводило до убіквітинування мішені та його подальшої деградації [5]. Подальше встановлення механізму взаємодії IMiDs та CRBN дозволило використовувати цей клас молекул як основу для створення молекул PROTAC, що рекрутують CRL4^{CRBN} E3-лігазний комплекс.

На даний момент розроблені PROTAC для зв'язування CRBN з такими білками як естрогенові рецептори, різноманітні кінази, трансдуктори сигналу та транскрипційні фактори. Усі ці молекули знаходяться на різних етапах клінічних досліджень. Створення молекул з різноманітними лігандами для певних мішеней дозволяє значно розширити наші підходи до розробки лікарських препаратів та є перспективним напрямком дослідження.

1.5 Дослідження лігандів CRBN методом віртуального скринінгу

Віртуальний скринінг є дуже популярним методом біологічних досліджень, оскільки він дозволяє за витрати менших ресурсів отримати результати, що можуть порівнюватися з результатами отриманими у тестах *in vitro*. Якість отриманих результатів може залежати від обраних підходів, кристалічних структур та масиву даних попередніх досліджень.

Дослідження, що використовують віртуальний скринінг для пошуку лігандів CRBN, можна поділити на ті, що шукають більш афінні та менш токсичні аналоги талідоміду, та ті, що шукають принципово нові ліганди.

До першої категорії можна віднести Jian Yan та Zhibing Zheng і їхнє дослідження [25]. Автори провели дослідження лігандів, що були створені за допомогою білдинг блоків, використовуючи молекулярний докінг та молекулярну динаміку. Як зазначають автори, три відібрані після віртуального скринінгу ліганди показали здатність інгібувати 50% зв'язування CRBN з талідомідом на концентрації 200 нМ, що є у 13 разів меншим ніж значення IC50 для леналідоміда. У іншому своєму дослідженні автори проводили аналіз вибірки зі 580 000 сполук, які попередньо були проаналізовані в ADME скринінгу та відібрано не токсичні сполуки, що обмежені молекулярною масою 220-350 г/моль. Наступним етапом було проведення молекулярного докінгу в програмі GOLD з використанням кристалічної структури 4TZ4. В результаті докінгу було відібрано топ 500 сполук, які перейшли у наступні етапи дослідження. За результатами аналізу водневих зв'язків та RMSD було обрано 5 речовин, які були рекомендовані для проведення подальших досліджень у *in vitro* та *in vivo* тестах [26].

Інший підхід до пошуку лігандів обрали Feng Xiong at al [24]. В своєму дослідженні вони аналізували речовини, що не мають в основі глутарімідного кільця, за допомогою молекулярного докінгу, ADME скринінгу, біохімічних та клітинних тестів. У результаті вони отримали 15 сполук, які потенційно можуть стати новими успішними лігандами CRBN. В іншому своєму

дослідженні використовуючи молекулярного докінгу, ADME скринінгу, та фармакофорний пошук було отримано 11 сполук, які в подальших клітинних дослідженнях показали здатність інгібувати зв'язування CRBN з іншими лігандами на рівні мікромолярних та навіть наномолярних концентрацій [23].

CRBN, як субстратний рецептор E3-лігазного комплексу CRL4^{CRBN}, є основною мішенню дослідження модуляції роботи комплексу. Залучаючи такі білки, як CRBN, DDB1, CUL4A та RBX1 комплекс зв'язує на одному кінці E2-лігазу та субстрат, в ролі якого виступає білок мішень, на іншому. Це призводить до убіквітинування субстрату та його подальшій протеасомній деградації. Субстратна специфічність CRBN може змінюватися через його зв'язування з різними лігандами. CRBN зв'язує ліганди у гідروفобній кишені, що розміщується в TBD домені білка. В кишені розташовані три триптофани, що і утворюють гідروفобну поверхню, а саме Trp380, Trp386 та Trp400. Основними існуючими на даний момент лігандами CRBN є талідомід та його похідні. В своїй структурі вони усі містять глютаримідне кільце та різноманітні варіанти зв'язаних з кільцем структур, що і зумовлює різноманіття IMiDs. Проте зараз розвивається і пошук принципово інших лігандів CRBN і результати досліджень показують, що цей напрямок має великий потенціал. Відкриття нових лігандів сприяє розробці таких структур як PROTAC. Ці молекули включають у себе ліганди CRBN, щоб зв'язувати лігазу з неосубстратами, які зв'язуються з другим лігандом у складі молекули. Розробка нових молекул PROTAC має великий потенціал до зміни наших підходів дослідження та лікування багатьох різноманітних захворювань. У свою чергу, використання методу віртуального скринінгу дозволяє нам дослідити нові ліганди CRBN, що не містять в собі глютаримідного кільця та обрати ті сполуки, які зможуть потенційно показати хороші результати у подальших етапах дослідження чим просунуть нас далі у дослідження таких складних тем, як лікування раку та різноманітних генетичних захворювань.

РОЗДІЛ 2

ЕКСПЕРИМЕНТАЛЬНА ЧАСТИНА

Для проведення дослідження використовувалося програмне забезпечення ICM Pro, оскільки воно має високий показник успішності та точності молекулярного докінгу у порівнянні з доступними аналогами [27]. За допомогою ICM Pro було проведено попередній молекулярний докінг вибірки, що складалася з відомих лігандів CRBN та декоїв. Також було проведено порівняння розміщення ліганду в кристалі, отриманого кристалографічним методом або методом криоелектронної мікроскопії, та розміщення ліганду після його молекулярного докінгу у той самий кристал. Після аналізу результатів попереднього докінгу було обрано кристал 7BQV [17]. В подальшому, було проведено молекулярний докінг вибірки із 150 400 різноманітних сполук в результаті чого було отримано оцінку для кожної речовини та список із 30 000 найкращих сполук, що в подальшому, можуть перейти в наступні етапи досліджень.

2.1 Пошук кристалічних структур CRBN в біоінформатичних базах даних

Проведення досліджень в галузі біоінформатики вимагає роботи з великою кількістю різноманітних баз даних, які містять інформацію про структуру та взаємодію різноманітних білків, малих молекул та нуклеотидних послідовностей. Зазвичай робота з будь-яким білком починається з пошуку у базі даних «UniProt». «UniProt» – це база, що надає безкоштовний доступ до білкових послідовностей та анотацій до них. Наразі в базі представлено майже 250 мільйонів різноманітних протеїнів [28]. За допомогою інструменту пошуку за запитом «CRBN» знаходимо нашу мішень. За результатом пошуку отримуємо 3235 варіантів, серед яких обираємо той, що має назву білка:

«Protein cereblon», та організм: «Human». На сторінці протеїну ми бачимо короткі описи функцій білка, його розміщення в клітині, таксономію та іншу інформацію, яка може бути корисною при проведенні будь-яких досліджень пов'язаних з білковими молекулами. Головне, що ми отримуємо це унікальний UniProtID людського CRBN – Q96SW2. В подальшому цей ID використовується для пошуку білка мішені у інших базах даних.

Для пошуку лігандів білка методом віртуального скринінгу наступним кроком необхідно обрати кристал, що буде використано для докінгу. «The Research Collaboratory for Structural Bioinformatics Protein Data Bank» (RCSB PDB) – це база даних, що містить велику кількість даних про структуру білків та нуклеотидних послідовностей. База агрегує дані про дослідження кристалічної структури різноманітних білкових молекул, що були отримані за допомогою таких методів як рентгеноструктурний аналіз, кріоелектрона мікроскопія та інші. Ресурс також надає можливість переглянути такі параметри кристалу, як роздільна здатність, що є важливим при виборі структури для подальшого докінгу, оскільки різне програмне забезпечення має вищу точність результатів для кристалів з різною роздільною здатністю [27].

За запитом «Q96SW2» було попередньо відібрано кристали, що містять у своєму складі повнорозмірний CRBN або частину його амінокислотної послідовності. Серед них було відсіяно ті, у яких був відсутній сайт зв'язування IMiDs – TDB. Також було відкинуто кристали, роздільна здатність яких становила більше 3 Å, оскільки точність отриманих даних з використанням таких кристалів може бути значно нижчою, у порівнянні з аналогами із вищою роздільною здатністю [27].

Також додатково було проведено пошук у базі «PDBe». База представляє собою ресурс Європейського біоінформатичного інституту, що збирає, систематизує та поширює дані про дослідження структури біологічних макромолекул. Для пошуку повторюємо процедуру із попереднього абзацу – пошук структур за запитом, сортування та відбір підходящих кристалів.

В результаті було отримано 9 кристалічних структур, які відповідають

критеріям відбору, тобто мають роздільну здатність менше 3 Å та містять у повну послідовність сайту, який ми досліджуємо. У вибірку кристалів для подальшого дослідження були включені не лише повнорозмірні білки, а й частини послідовності, оскільки кристали мали високу роздільну здатність.

2.2 Огляд кристалічних структур білка CRBN

Після первинного відбору кристалічних структур, необхідно детальніше оглянути обрані кристали для вибору найкращого для проведення подальшого докінгу. При виборі структури необхідно зважати не лише на роздільну здатність, а і на інші параметри, що впливають на якість молекулярного докінгу, а саме метод, за допомогою якого було отримано сполуку, та наявність ліганда в кристалі. Також слід звернути увагу на наявність мутацій в амінокислотній послідовності [29].

Таблиця 2.1 – перелік кристалічних структур та їхні основні характеристики.

РА – рентгеноструктурний аналіз; ЕМ – електронна мікроскопія.

| PDB ID | Роздільна здатність | Метод | Наявність мутацій | Ліганд |
|--------|---------------------|-------|-------------------|-------------------------|
| 7BQV | 1.8 Å | РА | 366 C>S | (S)-5-гідрокситалідомід |
| 7BQU | 1.9 Å | РА | 366 C>S | (S)-талідомід |
| 8TNQ | 2.41 Å | ЕМ | Відсутні | PT-179 |
| 5FQD | 2.45 Å | РА | Відсутні | Леноледомід |
| 8TNR | 2.5 Å | ЕМ | Відсутні | PT-179 |
| 8OIZ | 2.5 Å | РА | Відсутні | Помалідомід |
| 8OJH | 2.72 Å | РА | Відсутні | СС-885 |
| 8U16 | 2.9 Å | РА | Відсутні | Помалідомід |
| 8U15 | 2.95 Å | РА | Відсутні | СС-220 |

За допомогою пошуку в біоінформатичних базах даних було знайдено

9 кристалічних структур, наведених у таблиці 2.1, які потенційно підходять для проведення молекулярного докінгу.

Серед кристалічних структур слід виділити окремі деталі кожного кристалу. 7BQV має найвищу роздільну здатність (1.8 Å) серед знайдених структур, а також отриманий за допомогою рентгеноструктурного аналізу. В його амінокислотній послідовності є заміна цистеїну на серин в 366 позиції. Автори зазначають, що мутацію було зроблено для кристалізації, проте вона не впливає на сайт зв'язування IMiD [30]. 366 амінокислота розміщена латерально відносно сайту зв'язування в неструктурованій частині ланцюга і тому не впливає на здатність білку байндити ліганд. Така ж мутація міститься і в структурі 7BQU. Проте, кристал має нижчу роздільну здатність – 1.9 Å. Окрім самого CRBN в структурі кристалах міститься білок SALL4, а також ліганд, що обумовлює зв'язування з SALL4. У кристалі 7BQV лігандом виступає (S)-5-гідрокситалідомід, а у кристалі 7BQU – талідомід.

Кристали 8TNQ та 8TNR отримані за допомогою електронної мікроскопії, що може означати дещо меншу якість даних. Роздільна здатність цієї пари 2.41 та 2.5 Å відповідно, що є значно нищим у порівнянні з попередньою парою. Проте ці кристали у своїй структурі містять повнорозмірний CRBN, а також DDB1, який зв'язується із Lon доменом білка. Також обидва кристали містять ліганд, який є похідним талідоміду – PT-179 [31].

5FQD має роздільну здатність 2.45 Å та отриманий за допомогою рентгеноструктурного аналізу. Окрім самого CRBN, кристал містить DDB1 та SK1 α , яка стає субстратом CRBN за умови його зв'язування з леналідомідом та іншими IMiDs [32].

Рентгеноструктурним методом також були отримані кристали 8OIZ та 8OJH. Їхня роздільна здатність становить 2.72 та 2.9 Å відповідно. Кристали включають в себе CRBN, DDB1 та ліганди – помалідомід та CC-885 [14].

Найменшу роздільну здатність серед обраних кристалічних структур мають 8U16 та 8U15 – 2.9 та 2.95 Å відповідно. В структуру кристалу входять

CRBN, DDB1, SALL4. 8U16 зв'язаний з помалідомідом, а 8U15, у свою чергу, з похідним помалідоміду – CC885 [33].

Слідуючи загальним рекомендаціям, було відібрано 9 кристалічних структур білка CRBN, як мають роздільну здатність від 1.8 до 2.95 Å. Два кристали, а саме 7BQV та 7BQU, включають в себе лише частину амінокислотної послідовності CRBN, а саме TBD. Окрім того в цих кристалах є заміна цистеїну на серин у 366 позиції, яка не має впливу на сайт та його здатність зв'язувати IMiDs. Також у вибірку було включено два кристали, отримані за допомогою електронної мікроскопії, хоча MolSoft зазначають, що варто надавати перевагу структурам отриманим за допомогою рентгеноструктурного аналізу [29].

Наступним етапом буде підготовка та проведення молекулярного докінгу заздалегідь визначених лігандів, для оцінки якості передбачення зв'язування з різними кристалами.

2.3 Підбір речовин для контрольної вибірки

При підготовці моделі, яка буде основою для молекулярного докінгу нових сполук, необхідно переконатися, що модель надає правильну оцінку для кожної речовини. Для цього необхідно створити сет, в який будуть включені відомі ліганди CRBN та декої – сполук, що подібні до природніх лігандів за фізичними властивостями, проте сильно відрізняються за хімічною структурою, що виключає їхнє зв'язування з мішенню [34].

Одразу до списку відомих лігандів можна включити речовини наявні у відібраних вище кристалах – 5-гідрокситалідомід, талідомід, PT-17, леналідомід, помалідомід, CC-220 та CC-885. Окрім використання їх для оцінки співвідношення істинних та хибних результатів, повторний докінг речовин у структуру дає можливість порівняти вихідне положення ліганду у кишені з результатами молекулярного докінгу.

Для пошуку інших лігандів CRBN було використано такі біоінформатичні бази даних як «PDBе-KB» та «ChEMBL». За допомогою інструменту пошуку за запитом «Q96SW2» знайдено людський CRBN та його ліганди. Попередньо ліганди були завантажені та відібрані за критерієм «рChEMBL Value», значення якого було встановлено як більше 6. Отримані речовини, а також структури з бази даних «PDBе-KB» було додано до загального сету тестових речовин у якості лігандів.

Додатково у сет було додано сполуки, які попередньо тестувалися у реакціях *in vitro*. За умов, які були використані в есеї, відібрані речовини пригнічували зв'язування CRBN з талідомідом на 60% та більше. Оскільки усі протестовані сполуки були похідними талідоміду ми можемо стверджувати, що вони витісняли талідомід із сайту зв'язування та займали гідрофобну кишеню сайту TBD. У результаті було сформовано пул зі 97 лігандів, які зв'язуються із цільовим сайтом.

Наступним кроком було сформовано пул декоїв, що були використані як «обманки» для перевірки точності передбачення моделі. Для генерування декоїв на основі вибірки відомих лігандів було використано сервіс з відкритим доступом «DUD.docking». Сервіс містить базу мануально створених декоїв під деякі пептиди, а також дає можливість згенерувати декої для власної мішені на основі лігандів [34]. За допомогою генерації декоїв було створено 227 декої, що мали схожі до лігандів фізико-хімічні властивості, проте відмінну хімічну структуру. Для тестування було обрано співвідношення ліганів до декоїв 30/70% – усереднене значенням співвідношення із літературних даних [35].

Таким чином було створено об'єднаний сет сполук, що включав в себе відомі ліганди білка CRBN та створені на їх основі декої. Загальне число лігандів у вибірці складало 97 речовин, декоїв – 227, а загальна кількість сполук у тестовому сеті складала 324 сполуки.

2.4 Молекулярний докінг контрольної вибірки сполук

Для перевірки специфічності та селективності моделі, а також точність прогнозування позиції ліганду у сайті зв'язування було проведено попередній докінг контрольної вибірки із 324 сполук.

Для докінгу було обрано програму ICM pro. ICM використовує метод Monte Carlo для генерування поз за допомогою обертання, трансляції твердого тіла та обертання зв'язків. Отримана конформація перевіряється на відповідність енергетичним критеріям відбору. Якщо поза відповідає критерію вона зберігається та використовується для створення наступної пози. Це повторюється до поки не набереться певна визначена кількість поз. Перевагою такого підходу є більша варіативність, оскільки ліганд може подолати енергетичний бар'єр на поверхні потенційної енергії молекули [36]. Після безпосереднього докінгу за формулою обчислюється оцінка для кожної стикованої сполуки. Чим вищою є оцінка тим гірше сполука стикується з мішенню. Для програми ICM pro рекомендується відкидати сполуки оцінка якої більша за -32 [29].

В цій частині роботі ми проводили аналіз кожної кристалічної структури окремо, у кінці порівнюючи результати. Перед початком роботи необхідно провести підготовку кожної структури до проведення докінгу. Перш за все необхідно завантажити кристалічну структуру, що нас цікавить. Для цього, використовуючи інструмент пошуку, за допомогою PDB ID знаходимо кристал. Програма автоматично завантажує дані про усю кристалічну структуру з бази «PDB». У структурі необхідно видалити усю зайву воду та елементи, що не приймають участь у докінгу, тобто усе окрім білка-мішені та ліганда. Окрім амінокислотної послідовності у структурі білка міститься іон цинку. Хоча він не знаходить в безпосередній близькості до сайту зв'язування, його не видаляли, щоб максимально зберегти структуру CRBN. Окрім того. У кристалах 7BQV та 7BQU наявна заміна 366 цистеїну на серин, проте ця заміна розміщена віддалено від сайту зв'язування ліганду і тому не була повернена до дикого типу.

Далі структуру перетворюємо з PDB файлу на стандартний для програми формат. Після цього відділяємо структуру ліганду від білка, щоб програма не вважала ліганд частиною поверхні зв'язування та використовувала для докінгу саме гідрофобну кишень. Наступним етапом створюється проєкт докінгу за допомогою модуля «Docking». Попередньо обравши в списку об'єктів ліганд, в налаштуваннях обираємо визначення сайту навколо обраної молекули. Ця функція автоматично обирає поверхню сайту зв'язування навколо вже наявного у структурі ліганда. Налаштування усіх позицій проби було обрано за замовчуванням. Налаштування розміру коробки, що оточує сайт могли змінюватися для різних кристалів за потреби.

Після підготовки проєкту було проведено докінг кристалів CRBN підготовленої вибірки з 324 сполук. Докінг проводився за допомогою інструменту «Batch docking», який підходить для докінгу невеликих вибірок. Кожна речовина стикувалася з білком 3 рази та отримувала окрему оцінку стикування. Процедуру було повторено для кожної кристалічної структури. В результаті для кожної структури було отримано 972 результати докінгу речовин з кристалом.

2.5 Побудова та аналіз ROC кривих

Оцінка специфічності та селективності обраного методу важлива для проведення будь-яких досліджень. Особливо це стосується біоінформатичні методи, які роблять передбачення про процеси, що відбуваються у складних живих системах. Одним з методів оцінки специфічності та селективності методу є побудова ROC кривих, або крива робочої характеристики приймача. ROC крива – це метод візуалізації роботи класифікатора. Графік представляє собою 2 осі. Вісь x вказує на рівень хибнопозитивних значень, або специфічність, тобто наскільки добре тест відсіює негативні результати із загальної вибірки. Вісь y, яка вказує на рівень істиннопозитивних значень, або

чутливість, тобто наскільки добре тест відбирає позитивні значення із загальної вибірки [37]. Баланс між цими двома показниками завжди є ключовим для отримання оптимальної кількості хітів, які в подальшому покажуть себе як робочі речовини.

Для побудови ROC кривих було використано програмне забезпечення GraphPad Prism. Оцінки для кожної сполуки за результатами докінгу до кристалу були розділені на оцінки лігандів та оцінки декоїв і внесені у таблицю програми в окремі колонки. Використовуючи функцію «Analyzed», а саме «ROC curve», дані було проаналізовано та отримано криву, а також значення площі під кривою. Процедура була повторена для кожної кристалічної структури.

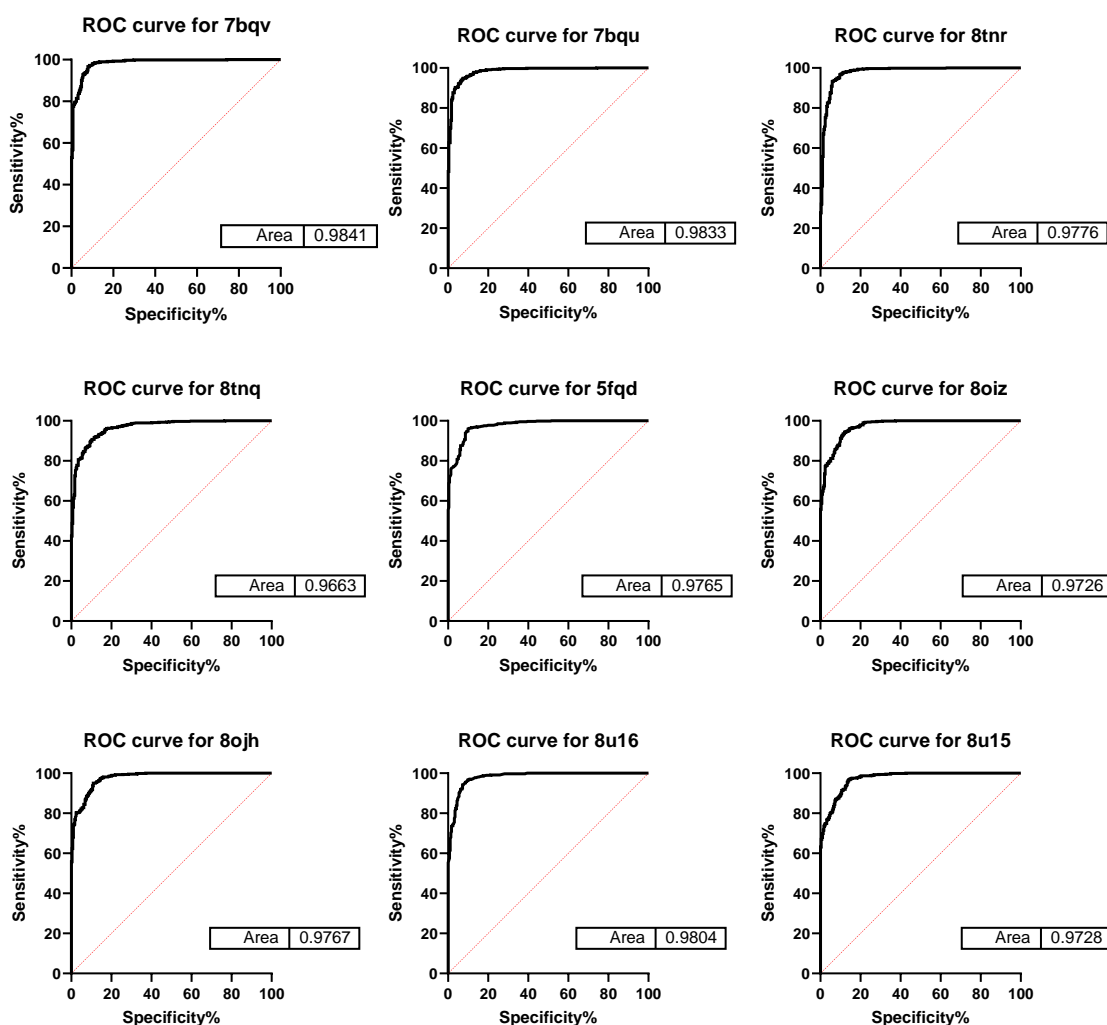


Рисунок 2.1 – ROC криві та AUC кожної кристалічної структури. Червона пунктирна лінія – крива випадкового присвоєння оцінки.

Побудувавши для кожної кристалічної структури ROC криві (рисунок 2.1), ми можемо зазначити що усі кристали мають площу під кривою (AUC) більше 0.9, що вже є високим показником. Кристал 7BQV показав найбільше значення AUC – 0.9844, що разом із високою роздільною здатністю кристала може свідчити про кращі можливості прогнозування стикування з лігандом, у порівнянні з іншими структурами.

2.6 Порівняння повторного стикування лігандів

Для молекулярного докінгу важливим є не лише загальна специфічність та чутливість, а і те наскільки точно визначається поза стикування ліганду з білком. Для порівняння розміщення 2 молекул використовують таку характеристику, як RMSD – середнє квадратичне відхилення позиції атомів. На рисунку 2.2 зображено приклад неправильного розміщення ліганда у кишені.

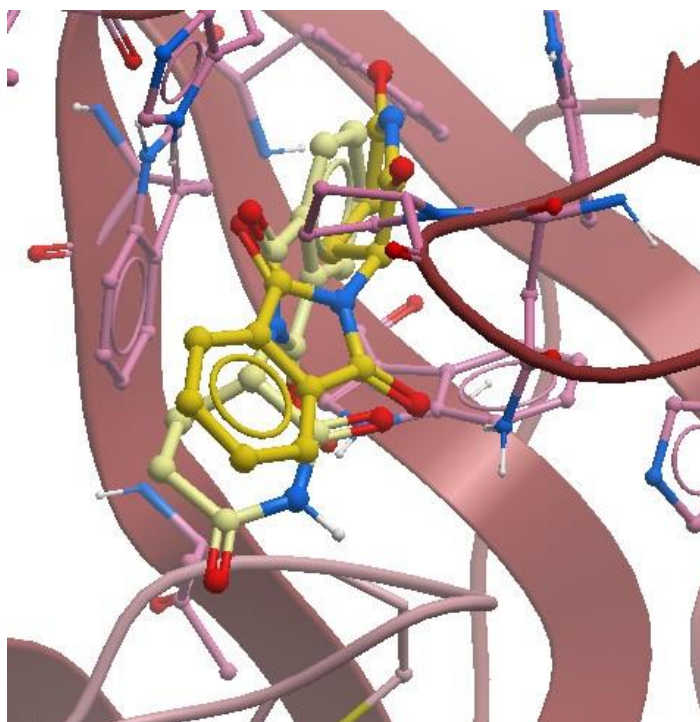


Рисунок 2.2 – приклад неправильного докінгу (S)-5-гідроксیتالідоміду з CRBN. Темно-жовтим позначено позу, що було отримано

рентгеноструктурним аналізом. Світло-жовтим – отриману за результатами докінгу.

Як зазначено вище, кожна речовина із вибірки стикувалася з білком тричі. Відповідно для кожного кристалу було визначено з позиції ліганду та вираховано 3 значення RMSD. Отримані результати стикування для кожного кристалу, а також значення AUC наведені у таблиці 2.2.

Таблиця 2.2 – результати проведеного докінгу контрольної вибірки для кожної кристалічної структури.

| PDB ID | Роздільна здатність | AUC | RMSD середнє | RMSD стандартне відхилення | RMSD 1 | RMSD 2 | RMSD 3 |
|--------|---------------------|--------|--------------|----------------------------|--------|--------|--------|
| 7bqv | 1.8 Å | 0.9841 | 1.6558 | 1.0333 | 2.0476 | 0.2408 | 2.6791 |
| 7bqu | 1.9 Å | 0.9833 | 2.6242 | 2.4536 | 0.1701 | 1.7268 | 5.9758 |
| 8tnr | 2.5 Å | 0.9776 | 2.1296 | 1.1872 | 3.6644 | 1.9518 | 0.7728 |
| 8tnq | 2.41 Å | 0.9663 | 2.4285 | 1.3222 | 0.7436 | 3.9733 | 2.5687 |
| 5fqd | 2.45 Å | 0.9765 | 1.4916 | 0.837 | 0.322 | 2.2341 | 1.9188 |
| 8oiz | 2.5 Å | 0.9726 | 1.5444 | 0.8247 | 0.3936 | 2.2839 | 1.9559 |
| 8ojh | 2.72 Å | 0.9767 | 2.1183 | 1.3545 | 0.2135 | 2.8954 | 3.246 |
| 8u16 | 2.9 Å | 0.9804 | 1.5033 | 0.8711 | 0.2987 | 2.329 | 1.8823 |
| 8u15 | 2.95 Å | 0.9728 | 5.9809 | 3.6716 | 7.0352 | 1.0506 | 9.8569 |

Слід зробити примітку стосовно результатів, отриманих для кристалу 8U15. CC-220 містить у своїй структурі «хвіст», що виступає за межі сайту зв'язування та не стикується з поверхнею білка. Ця частина молекули вільно розміщена за межами білка, тому не можна стверджувати, що вона має єдино правильне положення в просторі, яке зафіксоване у кристалічній структурі. Однак різне розташування цієї вільної частини молекули впливає на вимірювання RMSD, що проявляється у вигляді високих значень відхилення. Проте навіть якщо оцінювати інші показники, кристал погано підходить для проведення молекулярного докінгу, оскільки низька роздільна здатність та AUC вказують на нижчу здатність передбачати зв'язування ліганда з білком.

З отриманих даних ми бачимо, що кристал 7BQV показує найкращу здатність до передбачення позитивних та негативних результатів (AUC – 0.9841), а також має одне з найнижчих RMSD (середнє 1.6558 ± 1.0333).

2.7 Кореляційний аналіз оцінки стикування та інгібування сполук

Наявність контрольної вибірки зі 86 сполук, що була протестована попередньо у біохімічному тесті, дозволяє нам не лише порівняти здатність моделі до передбачення стикування та пози, а і оцінити кореляцію між отриманою оцінкою та відсотком інгібування зв'язування CRBN з міченим флуоресцентною міткою талідомідом. Ми можемо стверджувати, що чим вищим є відсоток інгібування реакції, тим краще речовина повинна зв'язуватися з білком. Оскільки усі сполуки, що тестувалися у біохімічному есеї, є похідними талідоміду, зв'язування з талідомідною пробою виключено.

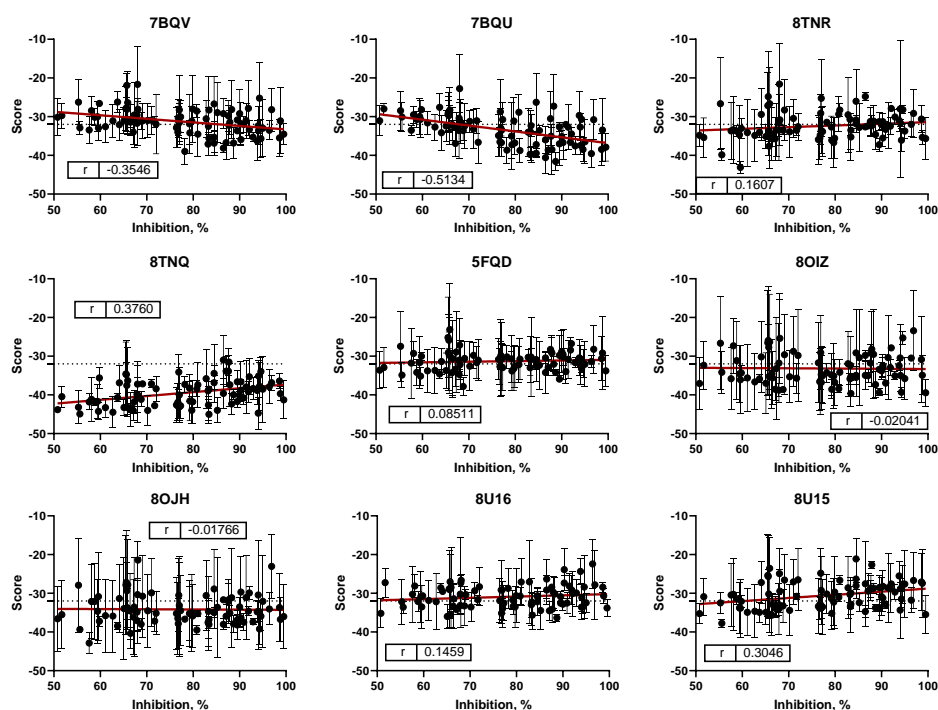


Рисунок 2.3 – Графік інгібування сполук, отриманого в біохімічному есеї проти оцінки докінгу. Аналіз за допомогою лінійної регресії (червона лінія) та кореляційного аналізу. Чорна пунктирна лінія – -32, оцінка, усі значення

нижче якої оцінюються як ліганди.

Кореляційний аналіз було виконано у «GraphPad Prism». Було побудовано графік, лінійну регресію та вираховано коефіцієнт кореляції для кожного протестованого кристалу.

За результатами, представленими на рисунку 2.3, ми можемо спостерігати слабку обернену залежність між оцінкою докінгу та відсотком інгібування реакції для двох кристалічних структур – 7BQV та 7BQU, коефіцієнт кореляції -0.355 та -0.513 відповідно. Саме такої кореляції ми очікували від порівняння оцінки стикування та інгібування ми очікували [29]. У той же час кристали 8TNR та 8U15 показали слабку пряму залежність між оцінкою та інгібуванням. Усі інші кристали не показали кореляції.

Спираючись на проаналізовані дані, отримані завдяки докінгу кристалічних структур та контрольної вибірки сполук, було прийнято рішення використовувати кристал 7BQV для подальшого докінгу сполук. Хоча структура 7BQU показала вищу кореляцію між оцінкою стикування та інгібуванням реакції в біохімічному есеї, вищий показник AUC та нижче RMSD, разом із вищою роздільною здатністю кристалічної структури схиляє нас до вибору саме 7BQV для проведення молекулярного докінгу вибірки сполук.

2.8 Віртуальний скринінг бібліотеки

Для докінгу було обрано дві бібліотеки компанії «Снамін» – «ННЛ-100» та «DDS-50». Їхній загальний розмір складає 150 400 різноманітних сполук, що мають молекулярну масу від 195 до 500 г/моль. Спільна бібліотека включає також сполуки, що містять у своїй структурі глютаримідне кільце. Їх було вирішено залишити, щоб перевірити та порівняти їхню оцінку з іншими сполуками, що не мають у своєму складі глютаримідного кільця.

Перед проведенням докінгу файл, що містить інформацію про

речовини бібліотеки формату .sdf, необхідно підготувати. Структури повинні не містити у собі солей та бути у нейтральному положенні. Для редагування структур було використано програмне забезпечення «DataWarrior».

Після видалення зі структур усіх лишніх компонентів структури у файлі було рандомізовано за допомогою скрипта, написаного мовою python. Наступним кроком було використання іншого скрипта, написаного також мовою python, який перетворював нейтральні групи у заряджені, для їхньої взаємодії із зарядженими групами білка під час докінгу. Далі файл з бібліотекою було розділено на окремі файли по 1000 сполук та проіндексовано для зручнішого проведення обчислень. Після цього бібліотека, що налічує загалом 150 400 сполук готова до проведення віртуального скринінгу.

Для проведення віртуального скринінгу було використано програмне забезпечення «ICM Pro», а саме програмний модуль «ICM-VSL», розроблене компанією «MolSoft». Докінг було проведено на стандартних налаштуваннях. За допомогою скрипта відбувався докінг кожного файлу, що містив у собі до 1000 сполук, після чого результати було об'єднано у спільний файл та відібрано 30 000 речовин, що показали найкращу оцінку докінгу.

Під час проведення експериментальної частини роботи було зроблено відбір кристалічної структури, підготовку бібліотеки зі 150 400 та проведено віртуальний скринінг. Для відбору кристалічної структури було знайдено в біоінформатичних базах даних 9 кристалів. Також було підготовлено бібліотеку контрольних сполук, яка включала в себе уже досліджені ліганди CRBN та декої, для перевірки правильності докінгу та оцінки кожної з кристалічних структур. Після докінгу контрольних сполук було побудовано ROC криві та обраховано AUC для оцінки селективності та специфічності. Окрім цього було проведено порівняння пози ліганда, що була зафіксовано під час створення кристалічної структури та пози, що були отримано за результатами докінгу. Також було проведено кореляційний аналіз оцінки докінгу та інгібування контрольної вибірки сполук у біохімічному тесті. За результатами проведеної роботи було обрано кристалічну структуру 7BQV.

Після цього було підготовлено скринінгову бібліотеку, що складалася сумарно зі 150 400 сполук. По завершенню роботи було проведено віртуальний скринінг підготовленої бібліотеки та отримано 30 000 сполук, що отримали найвищі оцінки за результатами докінгу.

РОЗДІЛ 3

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХНЄ ОБГОВОРЕННЯ

За результатами віртуального скринінгу бібліотеки було отримано 30 000 сполук, що мали найвищі оцінки. На рисунку 3.1 ми можемо бачити розподіл оцінки докінгу серед усіх 150 400 протестованих речовин. Оскільки бібліотека, що була обрана для докінгу не була таргетною, ми можемо спостерігати нормальний розподіл оцінок.

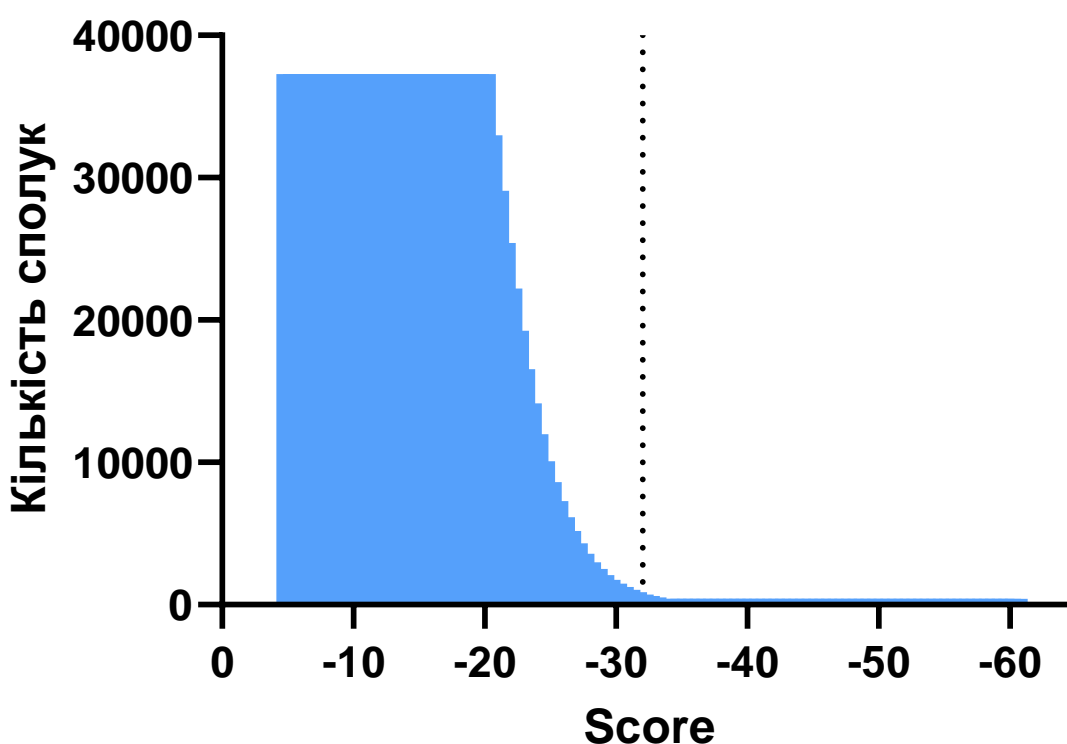


Рисунок 3.1 – розподіл оцінок докінгу серед усієї вибірки з 150 400 сполук. Чорна пунктирна лінія – -32, оцінка, усі значення нижче якої оцінюються як ліганди.

Для відбору хітів рекомендується відсіювати усі сполуки, що мають оцінку вищу ніж -32 [29]. Проте також зазначається, що для деяких білків, це значення може бути вищим. Тому було вираховано медіану оцінок, які були отримані в результаті докінгу контрольної вибірки сполук – -32.22 ± 5.21 . Це підтверджує, що в середньому модель оцінює сполуку як хіт – тобто речовину, що зв'язується з CRBN, на в діапазоні -27 - -37.4. Отже, оскільки дані

контрольної вибірки збігаються із загальними рекомендаціями, необхідно відкинути усі сполуки, що отримали оцінку меншу за -32. Кількість сполук, що отримали таку оцінку складає 732, що є 0.5 % від загальної кількості проаналізованих структур.

До вибірки сполук було включено у тому числі і речовини, що мають структурні елементи талідоміду, приклади яких наведено на рисунку 3.2.

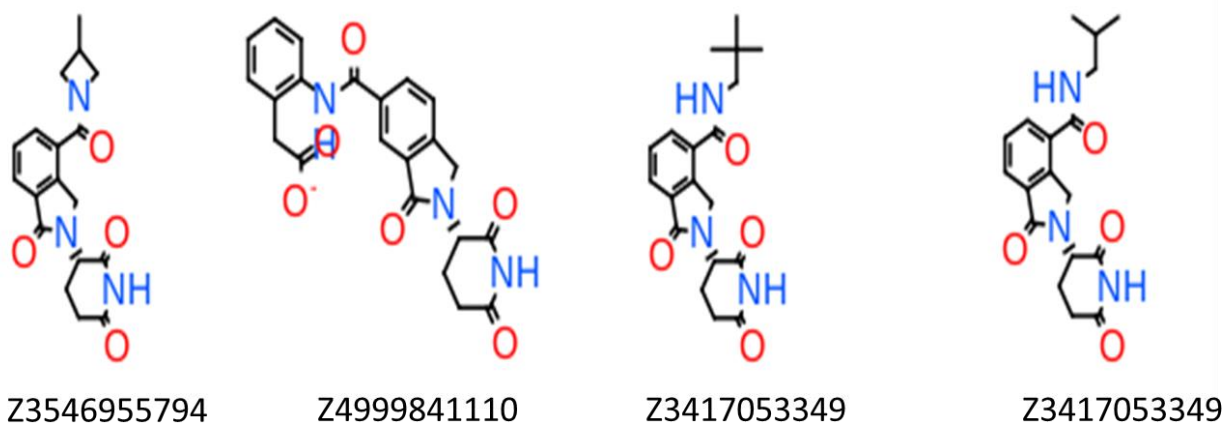


Рисунок 3.2 – приклади талідомідоподібних сполук у вибірці.

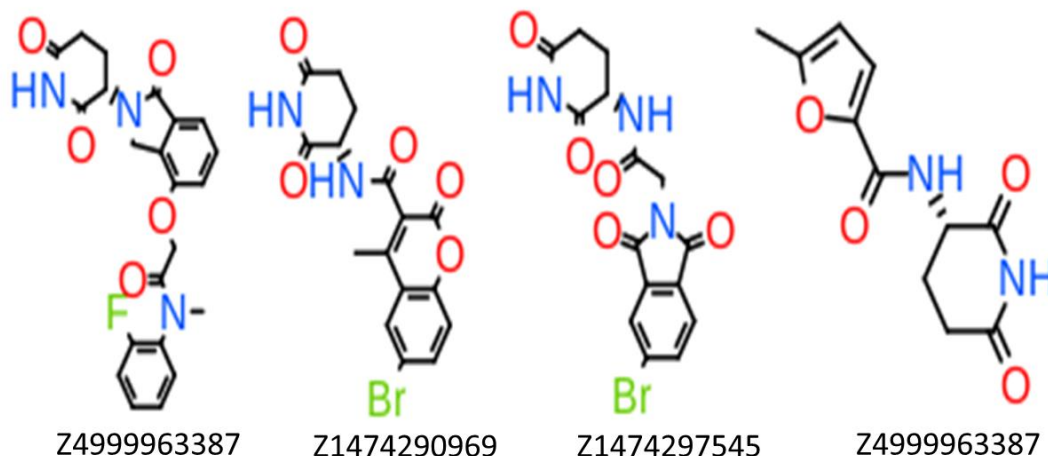


Рисунок 3.3 – приклади сполук, що містять глютаримідне кільце.

Серед усіх сполук, що мають оцінку докінгу менше -32 талідомідоподібних сполук налічується 40, тобто 4.9% від усіх сполук, що. Вони мають високі оцінки в діапазоні між -46.2 та -35.3. Ще 56 сполук (7.7%) містять у своїй структурі глютаримідне кільце, проте в них відсутні аналоги фталімідного кільця, як наведено на рисунку 3.3. Ці структури мають оцінки докінгу в діапазоні від -42 до -31.5, що може вказувати на певний вплив

фталімідного кільця в стикуванні сполук до сайту, хоч воно і не формує безпосередні водневі зв'язки з білком. Можна припустити, що елементи кільця утворюють водневі зв'язки з CRBN за посередництва води.

Оскільки наше дослідження має за мету пошук нових не-талідомідо-подібних сполук, ми вилучили з вибірки для аналізу усі сполуки, що містять глютарімідне кільце. В результаті в нас залишилося 640 сполук або 87.4% від вибірки хітів (0.43% від проаналізованої вибірки загалом).

Наступним кроком є аналіз отриманої вибірки, на наявність груп, для яких встановлені небажані побічні ефекти, таких як акриламіди, енаміни, гідразони, карбоніли та інші. Для цього було використано фільтри, що надаються у пакеті аналізу даних програми «ICM Pro», а саме фільтр хімічних сполук «Bad groups». У результаті відсіяно ще близько третини сполук і для подальшого аналізу відібрано 410 сполук, що не мають негативних побічних ефектів.

Оскільки ми шукаємо речовини, що в подальшому можуть бути додатково модифіковані для більшої афінності або створення молекулярного клею чи молекул PROTAC, одним з важливих фільтрів для відбору сполук це молекулярна маса сполук. Саме тому було прийнято рішення обмежити аналіз сполуками, що мають молекулярну масу не більше за 350 г/моль. Майже половина сполук мала більшу масу, тому їх було відкинуто. Кількість речовин з меншою масою складає 239 лігандів, що становить 0.16% від загальної вибірки протестованих речовин.

Після максимального обмеження вибірки хітів за допомогою фільтрів ми можемо безпосередньо розглянути отримані сполуки. Отримані речовини можна розділити за структурними особливості на певні кластери.

Найвищу оцінку докінгу отримала структура, що розміщується у 3-триптофановій кишені піролідонічним кільцем, структурою, схожою на глютарімідне кільце, рисунок 3.4. Структури з подібним кільцем, а також структури з сукцинімідним кільцем було об'єднано в одну групу для аналізу.

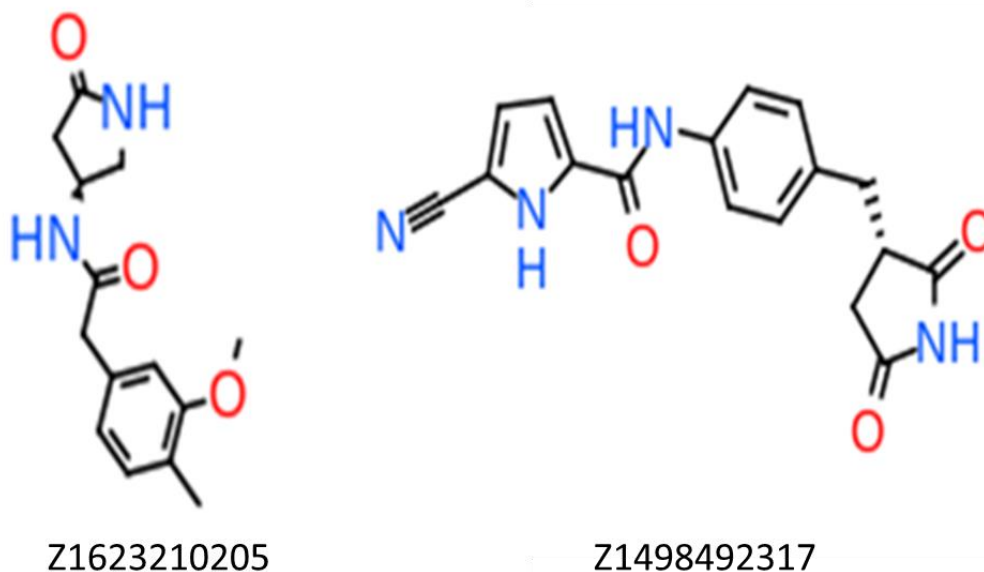


Рисунок 3.4 – приклади структур, що містять піролідоннове або сукцинімідне кільце.

За результатами докінгу було отримано 24 сполуки такого типу. Кільце, що розміщується у гідрофобній кишені утворює 2-3 водневі зв'язки, подібні до тих, що утворює глутаримідне кільце. Інші частини ліганді також можуть утворювати водневі зв'язки з білком, що може позитивно впливати на зв'язування. Серед цієї групи сполук ми можемо виділити сполуки Z1623210205 та Z2090768097, зображених на рисунку 3.5.

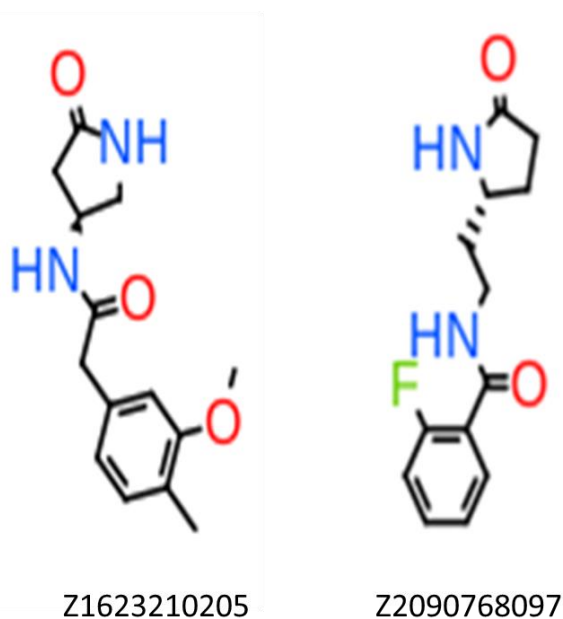


Рисунок 3.5 – схематичне зображення обраних сполук.

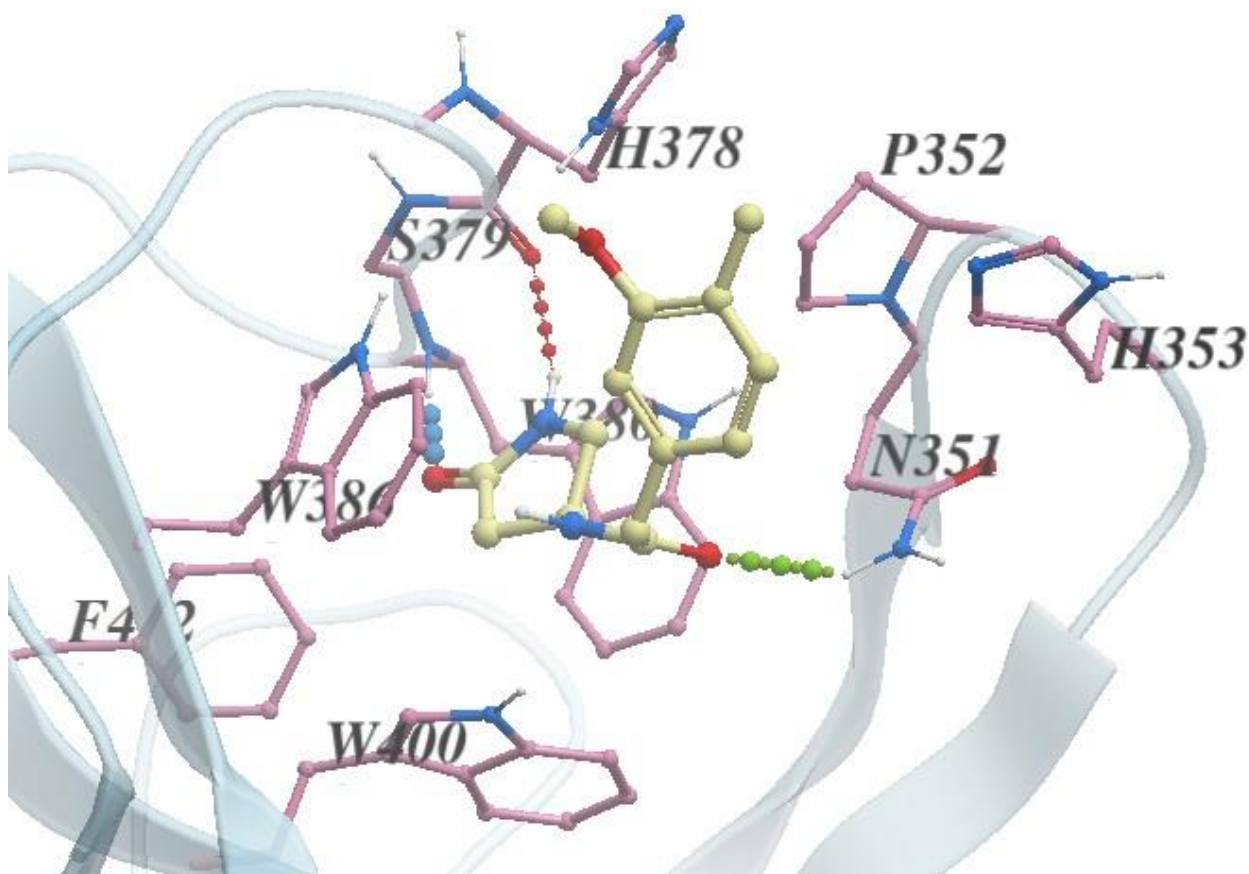


Рисунок 3.6 – стикування сполуки Z1623210205 з сайтом зв'язування.

Обидві сполуки мають доволі низьку молекулярну масу – 262.1 (Z1623210205) і 250.1 (Z2090768097) г/моль, що робить їх потенційно привабливими кандидатами для подальших модифікацій для створення лігандів. Структури утворюють 2 водневі зв'язки з гідрофобною кишенею за рахунок піролідонного кільця. Хоча на рисунку 3.7 видно, що відстань між атомом водню кільця та атомом кисню протеїну знаходяться на відстані 2.42 Å і не утворюють водневого зв'язку на рисунку 3.6 ми бачимо утворення зв'язку. Порівняння стикування кільця з талідомідом показало, що структура Z1623210205 стикувалася подібно до талідоміду, з чого можна зробити висновок, про наявність такого самого зв'язку із кільцем у структурі Z2090768097. Окрім того, обидві молекули містять у своїй структурі атом кисню, який утворює додатковий зв'язок, чим стабілізує структуру. Структури також містять частини, що виходять за межі сайту. Ми вважаємо, що ці частини не мають впливу на стикування молекул з білком, а тому за потреби можуть бути видалені чи модифіковані.

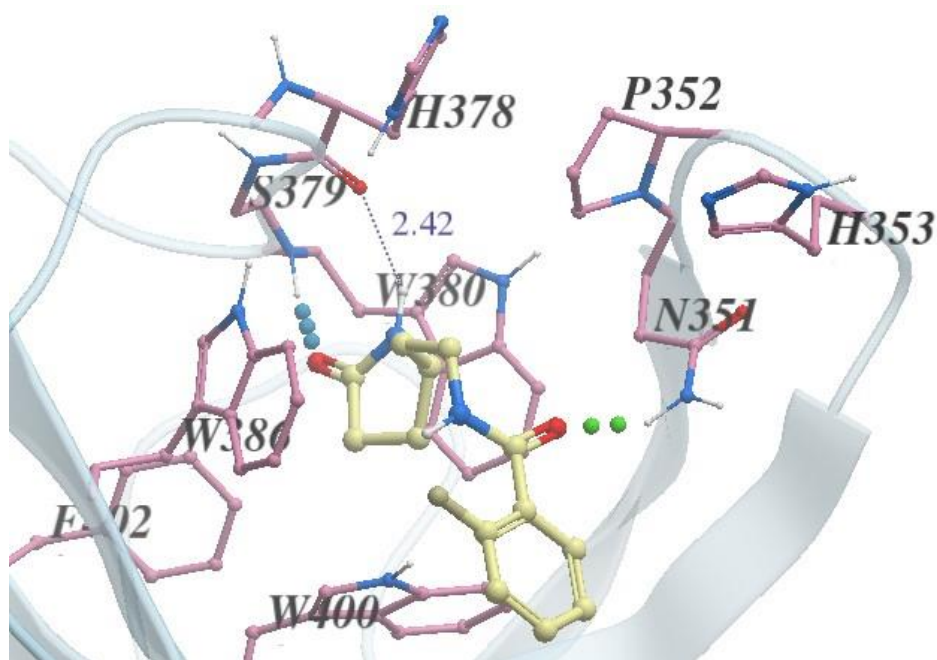


Рисунок 3.7 – стикування сполуки Z2090768097 з сайтом зв'язування.

Додатково потрібно відмітити групу сполук, що містять у своєму складі гексагідроіндолінонове кільце, як на рисунку 3.8. Такі сполуки мають більшу масу, у порівнянні зі структурами з піролідононовим. В той же час атоми, що містяться у цьому кільці не утворюють ніяких додаткових зв'язків, хоча така структура може створювати додаткову поверхню для вандервальсових взаємодій.

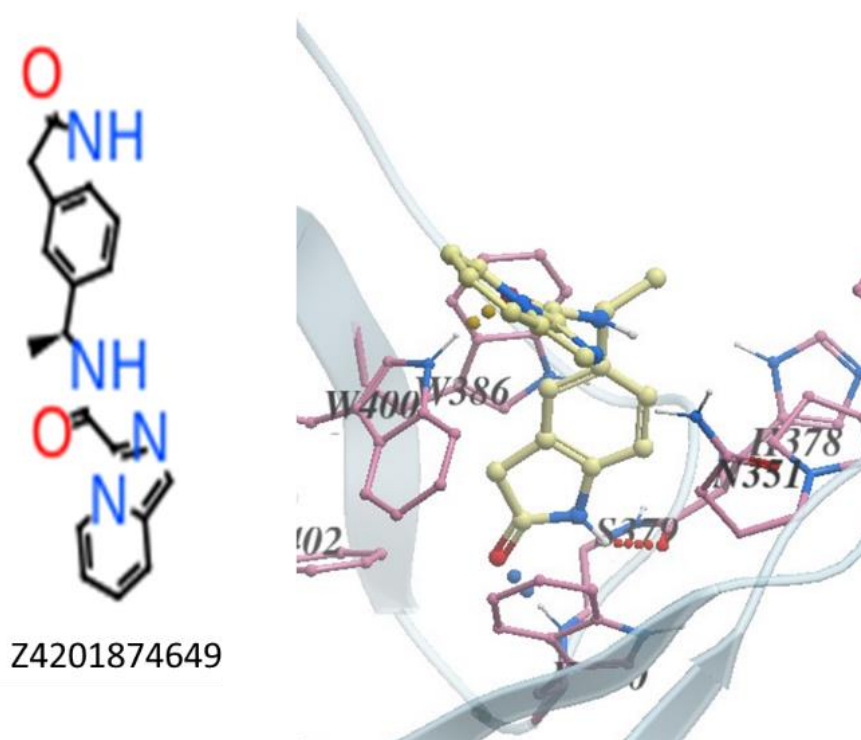


Рисунок 3.8 – структура та стикування структури з піролідононовим кільцем.

Серед протестованих сполук цікавим кандидатом на роль ліганда є сполука Z1343761507, представлена на рисунку 3.9. Ця речовина має невеликі розміри та молекулярну масу 222.1 г/моль, компактно розміщуючись у гідрофобній кишені. Дигідроурацил утворює 3 водневі зв'язки з амінокислотами, а додатковий зв'язок між N та NH групами посилює стабільність комплексу.

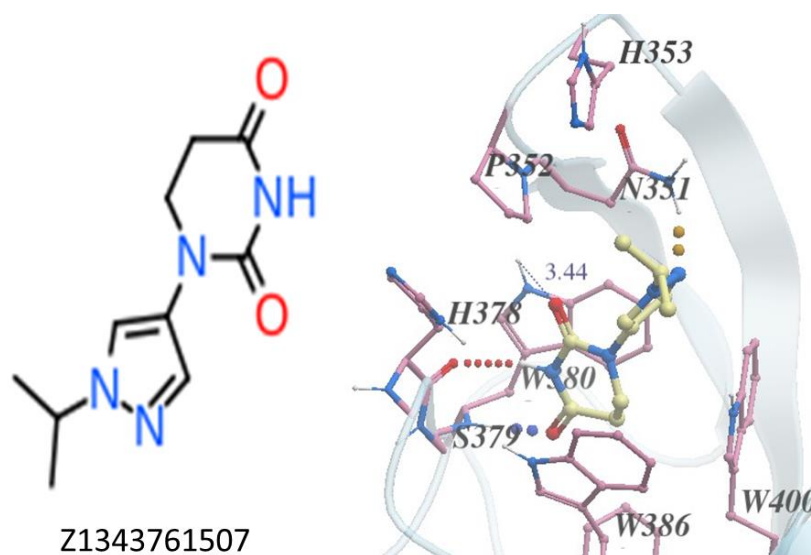


Рисунок 3.9 – структура та стикування речовини Z1343761507.

Також серед протестованих сполук можна відмітити сполуку Z906123502. Як видно на рисунку 3.10, вона в своєму складі містить більшу кількість оксигенів, що забезпечує утворення додаткових зв'язків. Проте сама структура кільця не дуже відрізняється від загальних мотивів.

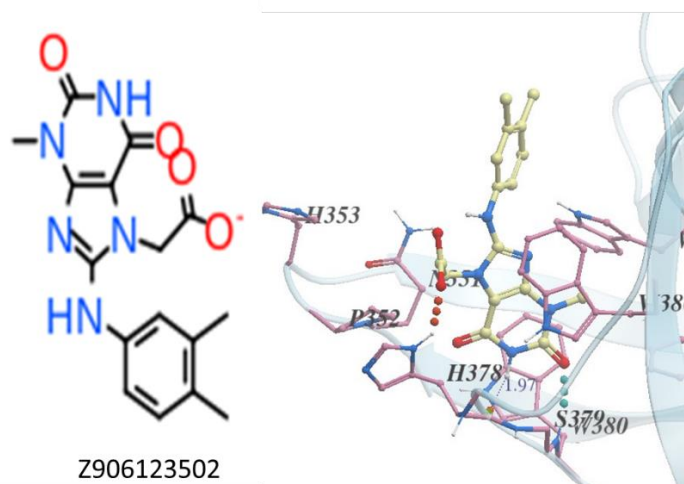


Рисунок 3.10 - структура та стикування речовини Z906123502.

Після аналізу усієї вибірки можна зробити висновок, що сполуки, які містять у своєму складі більше акцепторних груп краще стикуються із сайтом утвореним амінокислотами, що містять донорні групи. Серед хіті переважають структури, що розміщуються у 3-триптофановому сайті кільцем, що містить розміщений поруч кисень та нітроген, які виступають відповідно акцептором та донором водневого зв'язку. У свою чергу додаткові зв'язки за межами кільця забезпечують стабільність комплексу білок-ліганд.

За результатами докінгу загалом було відібрано 5 сполук – Z1623210205, Z2090768097, Z4201874649, Z1343761507 Z906123502, що потенційно можуть проявити хорошу активність у наступних стадіях дослідження.

ВИСНОВКИ

Розробка нових лікарських засобів вимагає від нас постійного пошуку нових методів та підходів. Одним із таких справді нових підходів є використання CRBN у таргетному протеасомному деградуванні білків. Моделювання роботи E3-лігазного комплексу, завдяки підбору підходящих лігандів, дозволяє не лише впливати на перебіг фізіологічних процесів, в яких задіяний сам CRBN, а і спрямовувати механізм протеасомної деградації на мішені, що є причинами різноманітних генетичних захворювань.

Проте перед пошуком власне самих лігандів важливим етапом є підготовка інструменту відбору цих лігандів. Саме тому в нашій роботі було протестовано 9 кристалічних структур білка CRBN та обрано 7BQV, яка показала найкращі результати у здатності передбачати стикування ліганді з білком.

Після підбору кристалічної структури було проведено сам віртуальний скринінг за результатами якого було отримано 30 000 сполук з найвищою оцінкою стикування. Після використання фільтрів, було отримано вибірку 239 сполук, стикування яких було проаналізовано.

Визначено, що структури, які містять у своєму складі кільце, що має хоча б одну донорну та одну акцепторну групу отримують високі оцінки стикування. Загалом спостерігається тенденція до патерну донор-акцептор-донор при утворенні водневих зв'язків із білком. Також було виявлено, що більша кількість акцепторних груп у структурі речовини сприяє більшій стабільності комплексу за рахунок утворення додаткових водневих зв'язків. Після аналізу було отримано 5 сполук – Z1623210205, Z2090768097, Z4201874649, Z1343761507 Z906123502, що потенційно можуть бути взяті у подальші етапи досліджень лігандів CRBN.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] H. K. Kim *et al.*, “Cereblon in health and disease,” *Pflugers Arch*, vol. 468, no. 8, pp. 1299–1309, Aug. 2016, doi: 10.1007/S00424-016-1854-1/METRICS.
- [2] S. Toma-Fukai and T. Shimizu, “Structural Diversity of Ubiquitin E3 Ligase,” *Molecules*, vol. 26, no. 21, Nov. 2021, doi: 10.3390/MOLECULES26216682.
- [3] K. M. Kortüm, Y. X. Zhu, C. X. Shi, P. Jedlowski, and A. K. Stewart, “Cereblon binding molecules in multiple myeloma,” *Blood Rev*, vol. 29, no. 5, pp. 329–334, Sep. 2015, doi: 10.1016/J.BLRE.2015.03.003.
- [4] P. Moreau *et al.*, “Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up,” *Annals of Oncology*, vol. 28, pp. iv52–iv61, Jul. 2017, doi: 10.1093/annonc/mdx096.
- [5] J. Lee, Y. Lee, Y. M. Jung, J. H. Park, H. S. Yoo, and J. Park, “Discovery of E3 Ligase Ligands for Target Protein Degradation,” *Molecules*, vol. 27, no. 19, Oct. 2022, doi: 10.3390/MOLECULES27196515.
- [6] E. R. Watson *et al.*, “Molecular glue CELMoD compounds are regulators of cereblon conformation,” *Science (1979)*, vol. 378, no. 6619, pp. 549–553, Nov. 2022, doi: 10.1126/SCIENCE.ADD7574/SUPPL_FILE/SCIENCE.ADD7574_MДАР_REPRODUCIBILITY_CHECKLIST.PDF.
- [7] Y. Zou, D. Ma, and Y. Wang, “The PROTAC technology in drug development,” *Cell Biochem Funct*, vol. 37, no. 1, p. 21, Jan. 2019, doi: 10.1002/CBF.3369.
- [8] T. Ishida and A. Ciulli, “E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones,” *Slas Discovery*, vol. 26, no. 4, p. 484, Apr. 2021, doi: 10.1177/2472555220965528.
- [9] R. J. Deshaies and C. A. P. Joazeiro, “RING domain E3 ubiquitin ligases,” *Annu Rev Biochem*, vol. 78, pp. 399–434, Jun. 2009, doi: 10.1146/annurev.biochem.78.101807.093809.
- [10] J. Barankiewicz, A. Salomon-Perzyński, I. Misiewicz-Krzemińska, and E. Lech-Marańda, “CRL4CRBN E3 Ligase Complex as a Therapeutic Target in Multiple Myeloma,” *Cancers (Basel)*, vol. 14, no. 18, Sep. 2022, doi: 10.3390/CANCERS14184492.
- [11] T. Ito *et al.*, “Identification of a primary target of thalidomide teratogenicity,” *Science (1979)*, vol. 327, no. 5971, pp. 1345–1350, Mar. 2010, doi: 10.1126/SCIENCE.1177319/SUPPL_FILE/ITO.SOM.REVISION1.PDF.
- [12] P. P. Chamberlain *et al.*, “Structure of the human Cereblon–DDB1–lenalidomide complex reveals basis for responsiveness to thalidomide analogs,” *Nature Structural & Molecular Biology* 2014 21:9, vol. 21, no. 9, pp. 803–809, Aug. 2014, doi: 10.1038/nsmb.2874.
- [13] B. Iovine, M. L. Iannella, and M. A. Bevilacqua, “Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions,” *Int J*

- Biochem Cell Biol*, vol. 43, no. 12, pp. 1664–1667, 2011, doi: 10.1016/J.BIOCEL.2011.09.001.
- [14] H. Bouguenina *et al.*, “A Degron Blocking Strategy Towards Improved CRL4CRBN Recruiting PROTAC Selectivity**,” *ChemBioChem*, vol. 24, no. 23, Dec. 2023, doi: 10.1002/CBIC.202300351.
- [15] S. Angers, T. Li, X. Yi, M. J. MacCoss, R. T. Moon, and N. Zheng, “Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery,” *Nature* 2006 443:7111, vol. 443, no. 7111, pp. 590–593, Oct. 2006, doi: 10.1038/nature05175.
- [16] M. E. Matyskiela *et al.*, “Crystal structure of the SALL4–pomalidomide–cereblon–DDB1 complex,” *Nature Structural & Molecular Biology* 2020 27:4, vol. 27, no. 4, pp. 319–322, Apr. 2020, doi: 10.1038/s41594-020-0405-9.
- [17] H. Furihata *et al.*, “Structural bases of IMiD selectivity that emerges by 5-hydroxythalidomide,” *Nature Communications* 2020 11:1, vol. 11, no. 1, pp. 1–11, Sep. 2020, doi: 10.1038/s41467-020-18488-4.
- [18] J. Yamamoto, T. Ito, Y. Yamaguchi, and H. Handa, “Discovery of CRBN as a target of thalidomide: a breakthrough for progress in the development of protein degraders,” *Chem. Soc. Rev.*, vol. 51, p. 6234, 2022, doi: 10.1039/d2cs00116k.
- [19] X. Nie *et al.*, “Development of Phenyl-substituted Isoindolinone- and Benzimidazole-type Cereblon Ligands for Targeted Protein Degradation,” *ChemBioChem*, vol. 25, no. 4, p. e202300685, Feb. 2024, doi: 10.1002/CBIC.202300685.
- [20] T. Asatsuma-Okumura, T. Ito, and H. Handa, “Molecular mechanisms of cereblon-based drugs,” *Pharmacol Ther.*, vol. 202, pp. 132–139, Oct. 2019, doi: 10.1016/J.PHARMTHERA.2019.06.004.
- [21] M. Krasavin *et al.*, “Synthesis of novel glutarimide ligands for the E3 ligase substrate receptor Cereblon (CRBN): Investigation of their binding mode and antiproliferative effects against myeloma cell lines,” *Eur J Med Chem.*, vol. 246, p. 114990, Jan. 2023, doi: 10.1016/J.EJMECH.2022.114990.
- [22] A. Kazantsev and M. Krasavin, “Ligands for cereblon: 2017–2021 patent overview,” *Expert Opin Ther Pat.*, vol. 32, no. 2, pp. 171–190, Feb. 2022, doi: 10.1080/13543776.2022.1999415.
- [23] F. Xiong, L. Yun Zhou, L. Chen, F. Cao, S. Zhang, and Z. Zuo, “Discovery of novel potential CRBN modulators through structure-based virtual screening and bioassay,” *J Mol Graph Model.*, vol. 117, p. 108325, Dec. 2022, doi: 10.1016/J.JMGM.2022.108325.
- [24] F. Xiong *et al.*, “Discovery of potential novel CRBN modulators by virtual screening and bioassay,” *Eur J Med Chem.*, vol. 236, Jun. 2022, doi: 10.1016/J.EJMECH.2022.114355.
- [25] J. Yan and Z. Zheng, “Discovery of Highly Potent CRBN Ligands and Insight into Their Binding Mode through Molecular Docking and Molecular Dynamics Simulations,” *ChemMedChem.*, vol. 18, no. 5, p. e202200573, Mar. 2023, doi: 10.1002/CMDC.202200573.

- [26] J. Yan, S. Sun, W. Zhang, P. Li, and Z. Zheng, “Combined scaffold hopping, molecular screening with dynamic simulation to screen potent CRBN ligands,” *J Cell Biochem*, vol. 122, no. 9, pp. 1207–1215, Sep. 2021, doi: 10.1002/JCB.29941.
- [27] M. H. Bohari and G. N. Sastry, “FDA approved drugs complexed to their targets: Evaluating pose prediction accuracy of docking protocols,” *J Mol Model*, vol. 18, no. 9, pp. 4263–4274, Sep. 2012, doi: 10.1007/S00894-012-1416-1.
- [28] “UniProtKB | Statistics | UniProt.” Accessed: Apr. 25, 2024. [Online]. Available: <https://www.uniprot.org/uniprotkb/statistics>
- [29] “ICM User’s Guide: Small Molecule Docking.” Accessed: Apr. 29, 2024. [Online]. Available: <https://www.molsoft.com/gui/start-dock.html>
- [30] H. Furihata *et al.*, “Structural bases of IMiD selectivity that emerges by 5-hydroxythalidomide”, doi: 10.1038/s41467-020-18488-4.
- [31] J. A. M. Mercer *et al.*, “Continuous evolution of compact protein degradation tags regulated by selective molecular glues,” *Science (1979)*, vol. 383, no. 6688, Mar. 2024, doi: 10.1126/SCIENCE.ADK4422/SUPPL_FILE/SCIENCE.ADK4422_DATA_S1.ZIP.
- [32] G. Petzold, E. S. Fischer, and N. H. Thomä, “Structural basis of lenalidomide-induced CK1 α degradation by the CRL4CRBN ubiquitin ligase,” *Nature 2016 532:7597*, vol. 532, no. 7597, pp. 127–130, Feb. 2016, doi: 10.1038/nature16979.
- [33] X. Ma *et al.*, “Structural and biophysical comparisons of the pomalidomide- and CC-220-induced interactions of SALL4 with cereblon,” *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 1–8, Dec. 2023, doi: 10.1038/s41598-023-48606-3.
- [34] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, “Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking,” *J. Med. Chem*, vol. 55, p. 32, 2012, doi: 10.1021/jm300687e.
- [35] M. Lapillo, T. Tuccinardi, A. Martinelli, M. Macchia, A. Giordano, and G. Poli, “Extensive reliability evaluation of docking-based target-fishing strategies,” *Int J Mol Sci*, vol. 20, no. 5, Mar. 2019, doi: 10.3390/IJMS20051023.
- [36] K. M. Sakamoto, K. B. Kim, A. Kumagai, F. Mercurio, C. M. Crews, and R. J. Deshaies, “Protacs: Chimeric molecules that target proteins to the Skp1–Cullin–F box complex for ubiquitination and degradation,” *Proc Natl Acad Sci U S A*, vol. 98, no. 15, p. 8554, Jul. 2001, doi: 10.1073/PNAS.141230798.
- [37] T. Fawcett, “ROC Graphs: Notes and Practical Considerations for Researchers,” 2004, Accessed: May 13, 2024. [Online]. Available: <http://www.purl>.