

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА  
Інститут високих технологій

Завідувач кафедри молекулярної біотехнології та біоінформатики

доц. Олексій Юрійович Нипорко

Протокол № \_\_\_\_\_ засідання кафедри

від “ \_\_\_\_\_ ” \_\_\_\_\_ 2021 р.

**Дослідження димеризаційного сайту SARS-CoV-2 специфічної M<sup>pro</sup>  
протеази та його взаємодії з потенційними інгібіторами**

Випускна кваліфікаційна робота бакалавра  
студента напрямку підготовки 091 «Біологія»

**Пономарьової Дарії Сергіївни**

Науковий керівник від кафедри:

к.б.н., доцент **Драган Анатолій Іванович**

Оцінка захисту роботи

---

**Київ – 2021 р.**

## РЕФЕРАТ

Дипломна робота бакалавра: 93 сторінок, 52 рисунки, 6 таблиць, 4 додатки, 78 джерел.

У даному дослідженні був проведений аналіз консервативності сполуки Mpro SARS CoV 2, виявлено потенційні інгібітори димеризації даного білка із подальшим визначенням їх потенційної токсичності та біодоступності.

Для підтвердження гомології структури SARS CoV Mpro 2 із структурами протеаз споріднених вірусів було проведено вирівнювання послідовностей 4 білків. Згідно з показниками MatchAlign та RMSD досліджувані білки мають дуже високу схожість, отже є консервативними. Висока консервативність білка SARS CoV 2 Mpro робить його одною з найперспективніших цілей при пошуку антикоронавірусних лікарських препаратів.

У ході дослідження, для виявлення потенційних інгібіторів димеризації була проаналізована бібліотека лігандів що містить більш ніж 8 мільйонів сполук, отримана із бази даних комерційно доступних сполук для віртуального скринінгу ZINC15. Було ідентифіковано 8229 сполук із показниками XP G Score менше за -5.8365 ккал / моль. На основі значень вільної енергії взаємодії було відібрано 6 речовин-кандидатів: кверцетин 3- (3-р-кумароїлглюкозид), кроцин, фагопіритол A2, пармоміцину сульфат, 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразин та ZINC5273784 .

Було створено 10 нс. симуляції молекулярної динаміки що описують комплекси 6 обраних сполук із Mpro. За допомогою аналізу графіків РМСД та значень вільної енергії взаємодії, отриманих з симуляцій молекулярної динаміки було ідентифіковано 4 сполуки здатні формувати найбільш міцні та стабільні комплекси із димеризаційним сайтом Mpro: кверцетин 3- (3-р-кумароїлглюкозид), кроцин, фагопіритол A2 та ZINC5273784.

Було створено профілі ADME речовин з найкращими показниками XP G Score та вільної енергії взаємодії із визначенням їх LD50, Caco2 проникності, ліпофільності, кількості порушень правила Ліпінського.

Ключові слова: Mpro SARS CoV 2; МОЛЕКУЛЯРНИЙ ДОКІНГ;  
СИМУЛЯЦІЯ МОЛЕКУЛЯРНОЇ ДИНАМІКИ; ZINC15; ДИМЕРИЗАЦІЯ;  
ADME

## ЗМІСТ

ВСТУП .....	6
СПИСОК СКОРОЧЕНЬ.....	7
<b>РОЗДІЛ 1</b>	
Огляд літератури .....	8
1.1 Структура віріона SARS CoV 2 .....	8
1.2 Життєвий цикл SARS CoV 2 .....	10
1.2.1 Проникнення вірусу до клітини.....	10
1.2.2 Трансляція вірусного геному .....	10
1.2.3 Синтез вірусної РНК.....	11
1.2.4 Експресія структурних та допоміжних білків.....	12
1.2.5 Утворення реплікаційних компартментів.....	12
1.3 Мпро SARS CoV 2.....	13
1.3.1 Каталітична активність Мпро SARS CoV 2.....	15
1.3.2 Архітектура активного центру Мпро SARS CoV 2 .....	16
1.3.2 (а) Субстрат-зв'язуючий підсайт S1 .....	18
1.3.2 (б) Субстрат-зв'язуючий підсайт S2.....	19
1.3.2 (в) Субстрат-зв'язуючі підсайти S3–S5 .....	19
1.3.3 Четвертинна структура Мпро SARS CoV 2.....	20
1.3.3 (а) Архітектура димерного інтерфейсу Мпро SARS CoV 2.....	22
1.3.4 Вплив димеризації на конформацію активного центру Мпро.....	23
1.4. Біоінформатичні методи дослідження у пошуці потенційних ліків. Методи молекулярного докінгу.....	25
<b>РОЗДІЛ 2</b>	
Об'єкти, матеріали та методи дослідження.....	29
2.1 Об'єкт дослідження.....	29
2.2 Система молекулярної візуалізації PyMol.....	29
2.1.1 Структурне вирівнювання у PyMol з використанням функції Align.....	29
2.3 Бази даних хімічних речовин вільного доступу .....	30
2.3.1 Архів PDB (Protein Data Bank).....	30
2.3.2 База даних вільного доступу ZINC15.....	30
2.4 Розрахункова платформа віртуального скринінгу <i>Schrödinger Suites 2018-1</i> .....	31

2.3.1 Утворення сітки рецептора (Receptor grid generation).....	31
2.3.2 Підготовка лігандів до докіну за допомогою функції LigPrep .....	31
2.3.3 Віртуальний скринінг бібліотеки лігандів (Virtual screening workflow) .....	33
2.5 Пакет програм для моделювання фізико-хімічних процесів GROMACS .....	34
2.5.1 Моделювання молекулярної динаміки комплексу білок-ліганд.....	35
2.5.1 (а) Отримання структури білка Mpro із PDB .....	35
2.5.1 (а) Створення топології з файлу PDB.....	35
2.5.1 (б) Підготовка структури ліганду .....	36
2.5.1 (в) Побудова комплексу ліганд-Mpro.....	37
2.5.1 (г) Створення коробки моделювання .....	38
2.5.1 (д) Внесення води та розчинника у систему.....	38
2.5.1 (е) Нейтралізація системи.....	38
2.5.1 (ж) Мінімізація енергії системи.....	39
2.5.1 (з) Отримання MD симуляції комплексу Mpro-ліганд .....	42
2.5.1 (с) Подальша обробка системи .....	42
2.5.2 Отримання значень RMSD .....	42
2.5.3 Отримання значень RMSF.....	43
2.5.4 Визначення енергії взаємодії Mpro-ліганди .....	44
2.6 Програмні забезпечення для визначення біодоступності та потенційної токсичності досліджуваних речовин. FAF-Drugs4. pkCSM.....	44
<b>РОЗДІЛ 3</b>	
Результати та їх обговорення .....	46
3.1 Вирівнювання послідовностей Mpro SARS CoV 2, Mpro SARS CoV, Bat CoV 3CLpro та MERS CoV nsp5 протеази у системі PyMol для визначення структурних подібностей.....	46
3.2 Ідентифікація потенційних інгібіторів димеризації SARS-CoV-2 Mpro інгібіторів шляхом молекулярного докіну у програмному пакеті Schrodinger .....	48
3.2.1 Дослідження енергій зв'язування речовин, схвалених FDA та іншими великими юрисдикціями із димеризаційним сайтом Mpro.....	48
3.2.2 Дослідження енергій зв'язування речовин що не містять інтерференційних структур (PAINS) із димеризаційним сайтом та активним центром Mpro.....	57
3.2.3 Дослідження енергій зв'язування речовин що споживаються людиною/містяться у харчових продуктах із димеризаційним сайтом Mpro	65

3.3 Дослідження молекулярної динаміки молекул-кандидатів та їх комплексів із димеризаційним сайтом Mpro .....	66
3.3.1 Дослідження середньквдратичних відхилень (RMSD) комплексів Mpro-ліганд .....	66
3.3.2 Дослідження середньквдратичних флуктуацій (RMSF) комплексів Mpro-ліганд .....	71
3.3.3 Дослідження вільної енергії взаємодії ліганд-білок .....	73
3.4 Визначення лікоподібності сполук із найкращими показниками енергії зв'язування з димеризаційним сайтом Mpro .....	79
3.4.1 Дослідження потенційної токсичності речовин-кандидатів.....	80
3.4.2 Дослідження біодоступності речовин-кандидатів.....	81
3.4.3 Поглиблений аналіз ADMET характеристик обраних лігандів.....	81
ВИСНОВКИ.....	83
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	84
ДОДАТКИ.....	93

## ВСТУП

У грудні 2019 року новий штам коронавірусу спричинив спалах респіраторного захворювання в місті Вухань, столиці провінції Хубей у Китаї, і з тих пір поширився у всьому світі [1]. Етіологічним агентом COVID-19 є вірус SARS-CoV-2, що має геном, подібний (82% схожості) до генома раніше відомого SARS-CoV (коронавірус SARS) [2]. На даний час не існує високо специфічних терапевтичних засобів націлених на SARS-CoV-2, і ефективні варіанти лікування залишаються дуже обмеженими.

Привабливою терапевтичною мішенню CoV є основна протеаза (Mpro) також відома як 3-хімотрипсин-подібна цистеїнова протеаза (3CLpro), оскільки цей фермент відіграє ключову роль у процесингу вірусних поліпротеїнів.

Інгібування Mpro перешкоджає реплікації вірусу, отже є великонадійною антикоронавірусною стратегією. Оскільки Mpro SARS-CoV-2 активна лише у димерній формі, потенційно ефективною терапевтичною стратегією є пошук інгібіторів, спрямованих на димеризацію ферменту [1-3].

Mpro є високо консервативним білком, оскільки більшість мутацій цього білка є летальними для вірусу. Таким чином, препарати, націлені на фермент Mpro, суттєво знижують ризик мутаційно-опосередкованої резистентності до лікарських препаратів та виявляють протівірусну активність широкого спектру дії [3].

## СПИСОК СКОРОЧЕНЬ

SARS-CoV-2 – важкий гострий респіраторний синдром коронавірусу 2

M<sub>pro</sub> SARS CoV 2 – SARS CoV 2 специфічна M<sub>pro</sub> протеаза

N<sub>sp</sub> – неструктурний білок

3CL<sub>pro</sub> – 3-хімотрипсин-подібна протеаза

S – білок шипа

E – білок вірусної оболонки

M – білок вірусної мембрани

N – білок нуклеокапсиду віруса

ORF – відкрита рамка зчитування

RTC – реплікаційно-трансляційний комплекс

RMSD – середньоквадратичне відхилення

RMSF – середньоквадратична флуктуація

MD – молекулярна динаміка

ADME (Absorption, Distribution, Metabolism, and Excretion) – показники поглинання, розподілу, метаболізму та екскреції речовини у організмі

Bat CoV 3CL<sub>pro</sub> – Bat CoV специфічна 3CL-подібна протеаза

MERS CoV nsp5 – MERS CoV специфічна протеаза

LD50 – середня доза речовини, що викликає загибель половини членів піддослідної групи

## РОЗДІЛ 1

### ОГЛЯД ЛІТЕРАТУРИ

#### 1.1 Структура віріона SARS CoV 2

Коронавіруси належать до сімейства Coronaviridae у порядку Nidovirales та містять одноланцюгову молекулу РНК в якості носія генетичного матеріалу. Розмір молекули РНК становить від 26 до 32 кб в довжину. Підгрупами сімейства коронавірусів є альфа ( $\alpha$ ), бета ( $\beta$ ), гамма ( $\gamma$ ) та дельта ( $\delta$ ) коронавірус. SARS CoV 2 належить до сімейства бетакоронавірусів[3].

Віріон коронавірусу складається із структурних білків, а саме шипа (S), оболонки (E), мембрани (M), нуклеокапсиду (N)[3,4]. Одноланцюговий генوم РНК (+ssRNA) інкапсидований N, M і E забезпечують його інкорпорацію у вірусну частинку під час процесу збирання. S-тримери виступають із вірусної оболонки, отриманої від клітини- господаря, і забезпечують специфічність до клітинних рецепторів при проникненні у клітину[5].

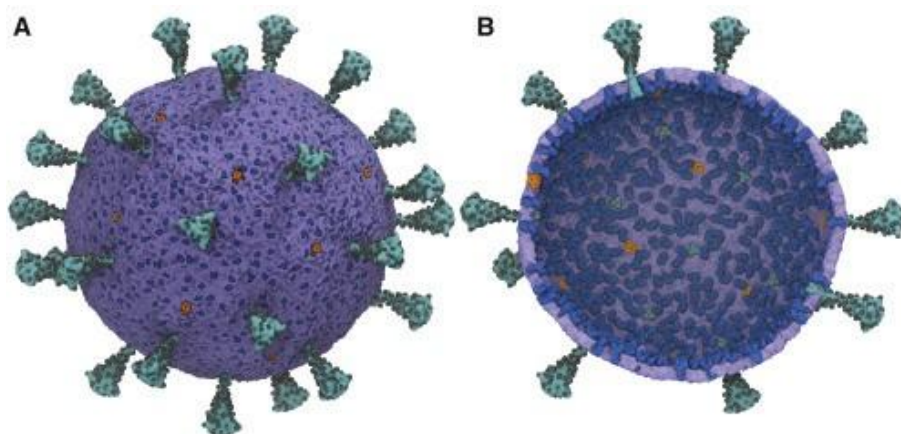


Рис 1.1. Модель віріона SARS-CoV-2. (A) Зовнішній вигляд віріона SARS-CoV-2. (B) Вид зсередини віріона SARS-CoV-2. Тримери S-білка відмічені блакитним кольором, а місця глікозилювання представлені як чорні сфери. Димери M-білка мають синій колір, а пентамерні іонні канали E - помаранчевий. Діаметр оболонки мембрани складає ~ 100 нм, 140 нм, враховуючи білки S на поверхні віріона.[5]

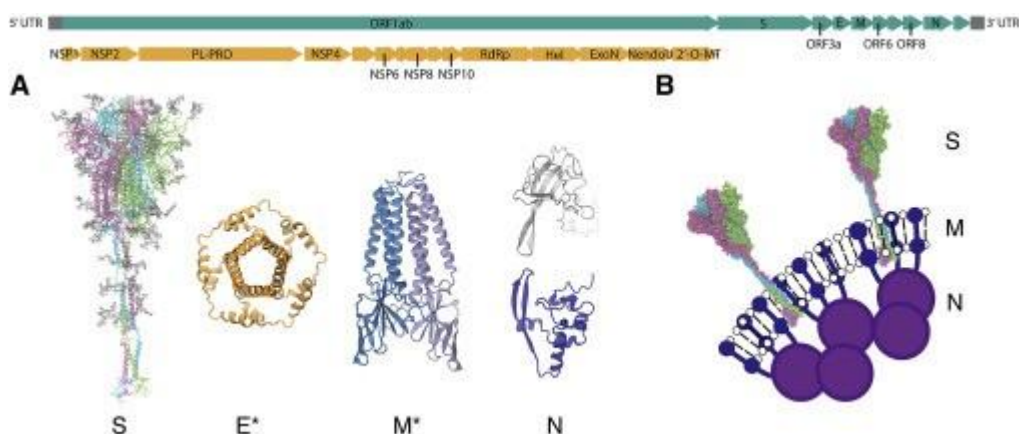


Рис 1.2 Вірусні білки SARS-CoV-2 . Геном SARS-CoV-2 показаний на верхній панелі. Неструктурні білки (NSP), кодовані у відкритій рамці зчитування (ORF) 1ab, пофарбовані в оранжевий колір, а весь геном - у блакитний. (A) Моделі структурних білків SARS-CoV-2 S, E, M та N. (B) Схема поверхні віріона отримана з крио-ЕМ зображень віріона. [5]

## 1.2 Життєвий цикл SARS CoV 2

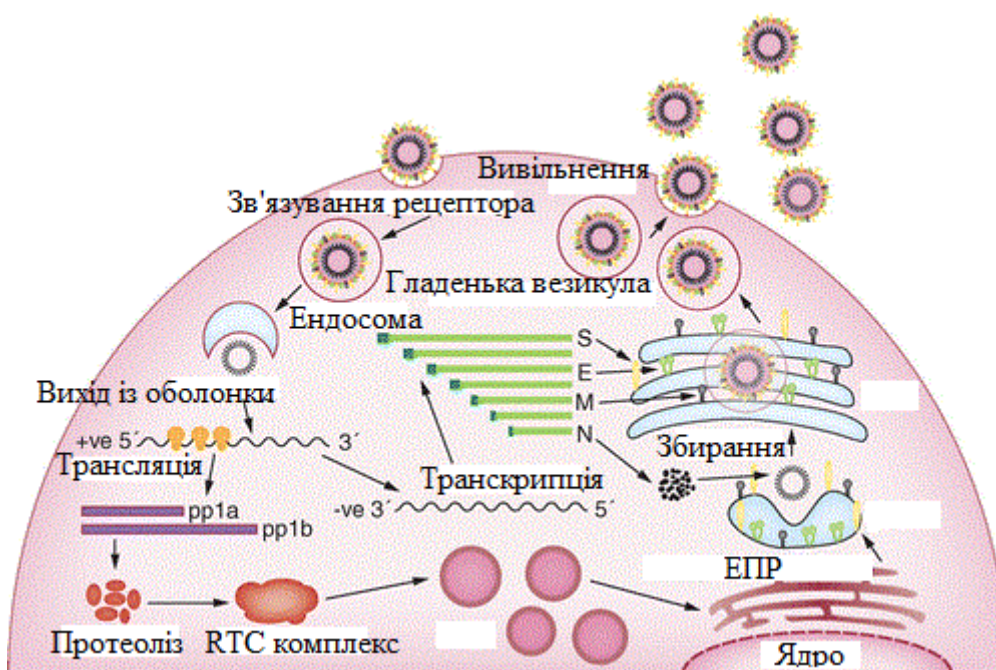


Рис 1.4. Життєвий цикл SARS CoV 2 [78]

### 1.2.1 Проникнення вірусу до клітини

S білки коронавірусу є гомотримерними глікопротеїнами злиття класу I, розділеними на дві функціонально різні частини (S1 та S2). Поверхневий S1 містить рецептор-зв'язуючий домен (RBD), який специфічно взаємодіє з рецептором клітини-хазяїна, тим самим визначаючи тропізм і патогенність вірусу. Трансмембранний домен S2 містить ділянки гептадного повторення

та пептид злиття, який, при великих конформаційних перебудовах, опосередковує злиття вірусних та клітинних мембран[6].

Окрім зв'язування з рецепторами, важливим етапом злиття є протеолітичне розщеплення S білків коронавірусу протеазами клітини господаря [6,7]. SARS-CoV використовує серинову протеазу TMPRSS2 на клітинній поверхні для праймування та проникнення в клітину, хоча ендосомні протеїнази цистеїну катепсин B (CatB) і CatL також можуть приймати участь в цьому процесі[8].

### 1.2.2 Трансляція вірусного геному

Вивільнення РНК коронавірусу в цитоплазму клітини хазяїна при проникненні є початком складної програми експресії вірусних генів. Трансляція ORF1a та ORF1b з геномної РНК призводить до утворення двох поліпротеїнів, pp1a та pp1ab, відповідно. Це є результатом запрограмованого -1 рибосомного зсуву рамки зчитування на короткому перекритті ORF1a та ORF1b 5. Профайлінг рибосом показав, що ефективність зсуву рамки зчитування між ORF1a та ORF1b становить від 45% до 70% у випадку SARS-CoV-2. Рр1a є приблизно в 1,4-2,2 рази більш вираженим, ніж рр1ab[9].

Шістнадцять неструктурних білків ко-трансляційно та посттрансляційно вивільнюються з pp1a (nsp1-11) та pp1ab (nsp1-10, nsp12-16) при протеолітичному розщепленні двома цистеїновими протеазами, що знаходяться в nsp3 (папаїноподібна протеаза; PLpro) та nsp5 (хімотрипсин-подібна протеаза)[10]. Протеазу, що знаходиться в nsp5 також називають 3С-подібною протеазою (3CLpro) через її схожість з пікорнавірусною 3С-протеазою або основною протеазою (Mpro), оскільки вона відповідає за протеолітичну обробку більшості місць розщеплення поліпротеїнів. Nsp2-16 є елементами вірусного RTC і націлені на визначені субклітинні зони, в яких взаємодія з факторами клітини-господаря визначає хід реплікаційного циклу. Вірогідно, Nsp2-11 забезпечують необхідні допоміжні функції для розміщення вірусного RTC, такі як модуляція внутрішньоклітинних мембран, ухилення від факторів імунітету господаря та забезпечення кофакторів для реплікації, тоді як nsp12-16 виконують основні ферментативні функції, беруть участь у синтезі РНК, корекції РНК та РНК модифікації [4,7].

Синтез РНК здійснюється за допомогою nsp12 РНК-залежної РНК-полімерази (RdRP) та двох її кофакторів nsp7 та nsp8, останній із запропонованою активністю праймази або 3'-кінцевої аденилілтрансферази [4,6,9,7]. Nsp14 забезпечує 3'-5' екзонуклеазну активність, яка сприяє синтезу РНК з унікальною функцією корекції РНК [1]. Коронавірусна система кепування складається з nsp10, який функціонує як кофактор, nsp13, який забезпечує активність РНК 5'-трифосфатази, та nsp14 та nsp16, які виконують

функції N7-метилтрансферази та 2'-O-метилтрансферази, відповідно [6,2,7,4].

### 1.2.3 Синтез вірусної РНК

Реплікація вірусного геному ініціюється шляхом синтезу негативно-направлених геномних копій у повному розмірі, які функціонують як шаблони для генерації нової позитивно-направленої геномної РНК. Ці новосинтезовані молекули РНК використовуються для трансляції, утворення більшої кількості нових nsp та RTC, або упаковуються у нові віріони. Відмітною ознакою коронавірусів та більшості представників порядку Nidovirales є переривчастий процес транскрипції вірусного геному, що супроводжується утворенням вкладених одна в одну 3' та 5' ко-термінальних субгеномних РНК (sgRNAs) [8,9].

Під час синтезу РНК з негативно направленим ланцюгом RTC перериває транскрипцію після зустрічі послідовностей регуляції транскрипції (TRS), розташованих вище за течією до більшості ORF в 3' третині вірусного геному. Біля цих TRS елементів, які також називають TRS-тілом, синтез негативно ланцюгової РНК зупиняється і відновлюється у TRS, сусідньому з лідуючою послідовністю (TRS-L), розташованою приблизно в 70 нуклеотидах від 5' кінця молекули [8,9,2].

Цей неперервний етап синтезу РНК коронавірусу включає взаємодію між комплементарними TRS утвореного негативно ланцюга РНК, і позитивної геномної РНК (TRS-L з + напрямком). Після повторного ініціювання синтезу РНК в області TRS-L, негативноланцюгова копія лідуючої послідовності додається до РНК що зароджується для завершення синтезу негативних ланцюгів sgРНК[10]. Переривчастий етап синтезу негативно-ланцюгової РНК призводить до утворення набору негативно-ланцюгових sgРНК, які потім використовуються як шаблони для синтезу характерного набору вкладених одна в одну позитивних sg-мРНК, які транслюються в подальшому у структурні та допоміжні білки. Незважаючи на те, що мРНКsg коронавірусу є структурно поліцистронними, передбачається, що вони функціонально моноцистронні, і що лише перша ORF на 5' кінці, яка відсутня в наступній меншій sgРНК, транслюється з кожної sgRNA [8,10]

### 1.2.4 Експресія структурних та допоміжних білків

Після реплікації та синтезу РНК білки S, E та M транслюються та проникають в ендоплазматичний ретикулум (ER). Ці білки рухаються в проміжний відділ ендоплазматичного ретикулума-Гольджі (ERGIC). У цьому

компартменті вірусні геноми, інкапсидовані білком N, починають проникати в мембрану, що призводить до утворення зрілого вірусу. Білок М спрямовує більшість білково-білкових взаємодій, необхідних для збирання віріонів. Після складання віріони транспортуються на поверхню клітини у везикулах і вивільнюються шляхом екзоцитозу.[12]

### 1.2.5 Реплікаційні компартменти

Первинні взаємодії між nsps та факторами клітини господаря під час раннього циклу реплікації коронавірусу ініціюють біогенез реплікаційних органел [6,10,12]. Для SARS CoV 2, як і для багатьох інших позитивно-ланцюгових РНК-вірусів, є характерною генерація ER-похідних та взаємопов'язаних перинуклеарних подвійних мембранних структур, таких як двомембранні везикули (ДМВ), звивисті мембрани та двомембранні сфери.

## 1.3 Мпро SARS CoV 2

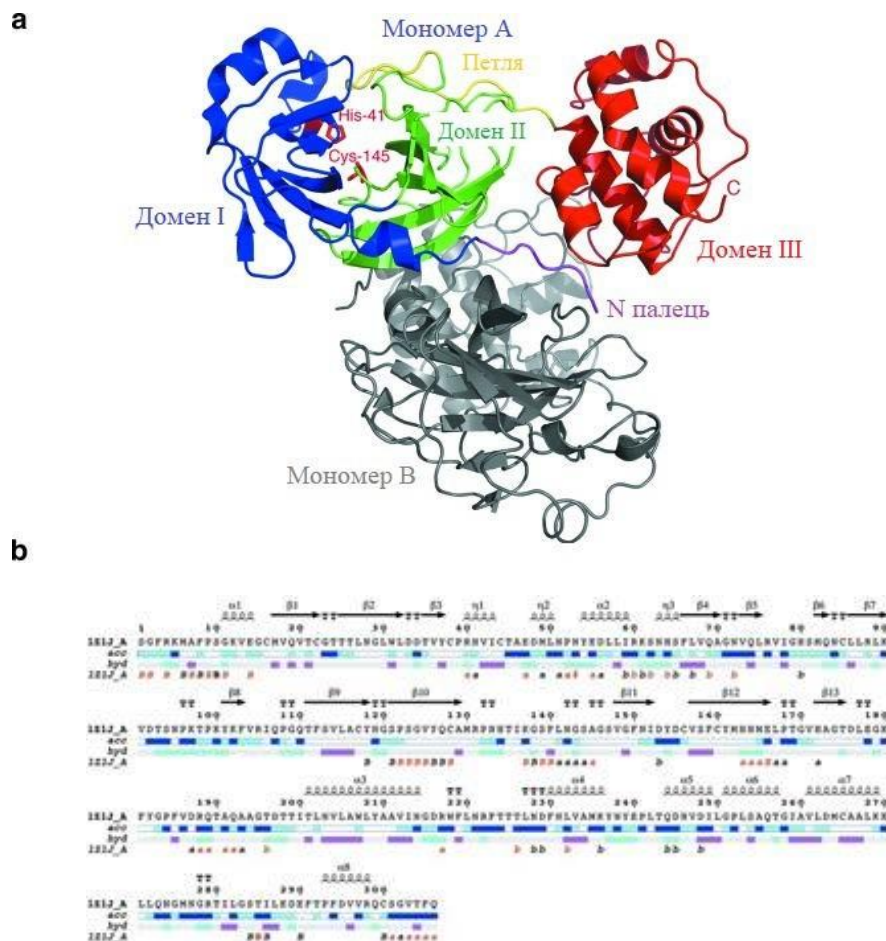


Рис. 1.5. Структура SARS-CoV Mpro. (a) Візуалізація димеру SARS-CoV Mpro (код pdb: 1Z1J). Домен I (синій колір) та II (зелений колір) з каталітичною діадою His41 та Cys145 (червоний колір). Каталітичні домени пов'язані з додатковим доменом III (червоний) довгою петлею (жовтий). N-

палець позначений пурпуровим кольором. Субодинаця В позначена сірим кольором. (b) Структурний аналіз M<sub>pro</sub>. Первинна амінокислотна послідовність відображається разом із вторинними структурними елементами. [13]

Білок SARS-CoV-2 M<sub>pro</sub> опосередковує дозрівання вірусних поліпротеїнів і є активним у димерній формі. [14]

SARS-CoV-2 M<sub>pro</sub> - це димер, що складається з двох мономерів, розташованих майже перпендикулярно один до одного. [16] Кожен мономер складається з трьох доменів і має каталітичну діаду (His41 та Cys145), розташовану в щілині між доменами I та II (залишки 10–99 та 100–182 відповідно). Каталітичні домени з'єднані довгою петлею із С-кінцевим доменом III (залишки 198–303), що складається з п'яти антипаралельних  $\alpha$ -спіралей. Між доменом II мономеру А та NH<sub>2</sub>-кінцевими залишками (N-пальцем) мономеру В в димерній структурі SARS-CoV-2 M<sub>pro</sub> існує димерний інтерфейс ( $\sim 1394 \text{ \AA}^2$ ).

Димеризація опосередковує ферментативну активність, оскільки N-палець кожного з двох мономерів взаємодіє з амінокислотним залишком Glu166 іншого мономеру, що сприяє правильній орієнтації S1 кишені місця зв'язування субстрату. С- і N-кінці мономерів утворюють димерний інтерфейс і тісно утримуються в димері, на відміну від мономерного стану, де рухливість цих кінцевих елементів вище.

Каталітичний N-кінцевий домен (I + II) та С-кінцевий домен III SARS-CoV M<sub>pro</sub> можуть піддаватися фолдингу незалежно один від одного. N-кінцевий домен (I + II) без домену III згортається в структуру, яка не відрізняється від інтактною хімотрипсиноподібної складки, але є ферментативно неактивною [17]. Додатковий домен III SARS-CoV M<sub>pro</sub> підвищує структурну стабільність каталітичного домену. Це, можливо, досягається за рахунок збільшення швидкості фолдингу доменів I і II, що, таким чином, збільшує загальну стабільність білка. До того ж, домен III приймає участь в утворенні четвертинної структури M<sub>pro</sub>.

M<sub>pro</sub> є перспективною ціллю для розробки анти-коронавірусних терапевтичних засобів широкого спектру дії завдяки своїй висококонсервативній тривимірній структурі у різних CoV. [13] CoVi постійно піддаються активному мутагенезу; однак ключові білки є дуже консервативними, оскільки мутації ключових білків часто є летальними для вірусу. [15] Було виявлено, що структура SARS-CoV-2 M<sub>pro</sub> схожа на кристалічну структуру SARS-CoV M<sub>pro</sub>. Лише 12 із 306 залишків відрізняються у SARS-CoV-2 M<sub>pro</sub> порівняно з SARS-CoV M<sub>pro</sub> (96%

ідентичності послідовності). Крім того, жоден з 12 варіантних залишків (T35V, A46S, S65N, L86V, R88K, S94A, H134F, K180N, L202V, A267S, T285A, I286L) не бере участі у будь-якій ферментативній активності SARS-CoV-2 Mpro. На загальну структуру SARS-CoV-2 Mpro мутації не впливали, і її структура може бути повністю накладена на структуру MAR SARS-CoV

Таким чином, препарати, націлені на консервативну Mpro, зазвичай здатні запобігати реплікації та розповсюдженню вірусу та виявляти противірусну активність широкого спектру. Крім того, ліки, орієнтовані на Mpro, можуть зменшити ризик опосередкованої мутаціями резистентності у майбутніх патогенних вірусних штаммах.

### 1.3.1 Каталітична активність Mpro SARS CoV 2

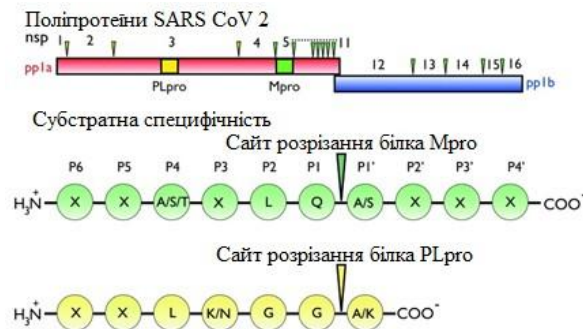


Рис 1.6 Поліпротеїни SARS-CoV. Автопроцесинг поліпротеїнів 1a та 1ab (1a + 1b) з допомогою Mpro відбувається на 11 ділянках (зелені трикутники) та з допомогою PLpro на трьох ділянках (жовті трикутники), що призводить до дозрівання 16 неструктурних білків (nsP). Mpro та PLpro знаходяться у nsP 5 та 3 відповідно. [16]

SARS-CoV Mpro розщеплює pp1a на 11 ділянках, що містять канонічну послідовність Leu – Gln– ↓ - (Ala / Ser). [18 ]

Зв'язування пептидного субстрату з активним центром утворює комплекс Міхаеліса. Електростатичний тригер, опосередкований Cys145 на чутливому пептидному зв'язку, ініціює хімічну реакцію. Ацилювання сульфгідрильної групи цього цистеїну призводить до утворення ковалентного зв'язку між С-кінцевим фрагментом субстрату та SH-групою та вивільненням N-кінцевого фрагменту. Субстрат розщеплюється при пептидному зв'язку Gln-Ala (або Gln-Ser). N-кінцевий напівпептид вивільняється як перший продукт, тоді як С-кінцева половина ацилює залишок Cys145 у активному центрі. Потім ацильований проміжний продукт дезацилюється, вивільняючи С-кінцевий пептид, і це завершує каталітичний цикл. His41 діє як загальна каталітична основа на етапі деацилювання. Запропонований фліп-флоп механізм пояснює роль четвертинної структури Mpro у регулюванні її діяльності. Лише одна з

двох субодиниць є каталітично активною в будь-який момент часу, і, отже, дві субодиниці існують в неперервному циклі активність-неактивність. Зв'язування субстрату на одній субодиниці індукує деациляцію іншої [19]

Група –SH Cys145 є іонно-спареною із сусіднім залишком гістидину (His41) з утворенням каталітичної діади (Cys145 – His41). Це відрізняє Mpro від більшості серинових протеаз, які мають каталітичну тріаду Ser – His – Asp у своїх активних центрах. У Mpro стабільна молекула води займає положення Asp типової тріади серинової протеази, і, можливо, відіграє роль у стабілізації імідазолієвого кільця під час каталізу. Мутації залишків каталітичної діади (H41A та C145A) майже повністю скасовують ферментативну активність, і ці мутантні ферменти існують виключно як димери [19,20]. Однак мутація Cys145 до Ser призводить до утворення частково активного ферменту. [20,21,22]. Обмежуючим швидкість гідролізу етапом у Mpro є етап ковалентного деацилювання. [28].

Для пояснення контролю активності ферменту Mpro було запропоновано механізм асоціації – активації – каталізу – дисоціації. Каталітично компетентна конформація в одному протомері індукується лише при утворенні димеру. У фізіологічних умовах Mpro існує у вигляді асиметричного димеру, який може мати напівсайтовий каталітичний цикл ациляції-деациляції; тобто коли одна субодиниця знаходиться в активній ацильованій формі, інша - у деацильованій. Димер є важливою функціональною одиницею Mpro, і регулює каталітичний оборот. [27,28,29].

Фліп-флоп механізм функціонування може бути описаний таким чином: дві субодиниці використовуються по черзі в реакціях ацилювання та деацилювання, завдяки чому зв'язування на субодиниці А індукує деациляцію на субодиниці В і навпаки. Таким чином, активність Mpro може регулюватися негативною кооперативністю.

### 1.3.2 Архітектура активного центру Mpro SARS CoV 2

Каталітичний сайт Mpro SARS CoV 2 самостійно набуває реактивної цвіттеріонної форми, в якій Cys145 знаходиться в негативно зарядженому тіолатному стані, а His41 подвійно протонований і позитивно заряджений.

В каталітичній ділянці 3CL Mpro використовується неканонічна діада Cys145 – His41, та молекула води, зв'язана водневим зв'язком з каталітичним гістидином [30,31]. Тіолова група цистеїну функціонує як нуклеофіл на першому етапі реакції гідролізу, атакуючи атом вуглецю розщеплюваного пептидного зв'язку. Для гідролізу субстрату необхідно, щоб каталітична діада перебувала у цвіттеріонному стані з депротонованим Cys145 та протонованим His41, який може бути утворений шляхом переносу протону від тіолу Cys145 до імідазолу His41 за загальним кислотно-основним механізмом [32,33]

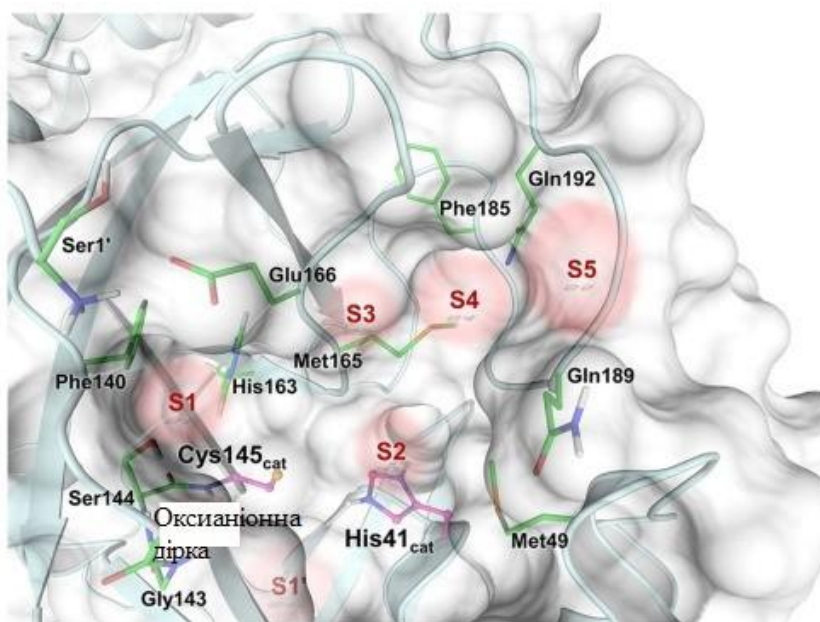


Рис 1.7 Архітектура активного центру SARS – CoV-2 3CL Mpro із зазначенням розташування субстрат-зв'язуючих субсайтів S1' – S5 та оксианіонного отвору.[30]

Розташований за 3,9 Å від Cys145, каталітичний залишок His41 протонований як на Nδ1 атомі азоту, так і на Nε2 бічного ланцюга імідазолу, і тому є позитивно зарядженим. His41 міцно зв'язаний водневими зв'язками з молекулою води (H<sub>2</sub>O<sub>cat</sub>), яка, вірогідно, відіграє роль третього каталітичного залишку [18, 25] із канонічної каталітичної тріади, стабілізуючи заряд і положення імідазолового кільця His41. Відстань Nδ1-H ... OH<sub>2</sub>O становить 1,7 Å. Позиція молекули H<sub>2</sub>O<sub>cat</sub> стабілізується ще кількома, можливо, більш слабкими водневими зв'язками з основним ланцюгом His41, His164 та Asp187. His164 є подвійно протонуваним і позитивно зарядженим. Asp187 не протонований, заряджений негативно і бере участь в утворенні міцного сольового містку з Arg40. His163, розташований поблизу каталітичного Cys145, є однократно протонуваним і незарядженим, утворюючи водневий зв'язок з протонуваним фенольним бічним ланцюгом Tyr161, група OH якого розвернена від імідазолу His163.

### 1.3.2 (а) Субстрат-зв'язуючий підсайт S1

Група P1 субстрату, зазвичай Gln, зв'язується в субсайті S1. З одного боку він оточений залишками 140–144, що утворюють поворот, формуючий оксианіонну дірку, а з протилежного боку - Met165, Glu166 та His172. Задня стінка підсайту S1 створюється бічними ланцюгами Phe140 та His163. N-кінець Ser1' другого протомера в межах активного гомодимера 3CL Mpro закриває підсайт S1 зверху. N-кінцевий амін є протонуваним позитивно зарядженим катіоном амонію -NH<sub>3</sub><sup>+</sup>. Він утворює три водневі зв'язки, по одному з карбонілом основного ланцюга Phe140, карбоксилатом бічного

ланцюга Glu166 та молекулою H<sub>2</sub>O. Обидва залишки гістидину, His163 та His172, у цьому підсайті є однопротонованими та нейтральними. Депротонований N $\delta$ 1 His172 пов'язаний через водень з амідним азотом основного ланцюга Gly138 з відстанню азот-гідроген 2,2 Å, тоді як протонований N $\epsilon$ 2 має довгий, можливо слабкий, водневий зв'язок з

карбоксилатом Glu166, гідроген-киснева відстань = 2,5 Å. [21]

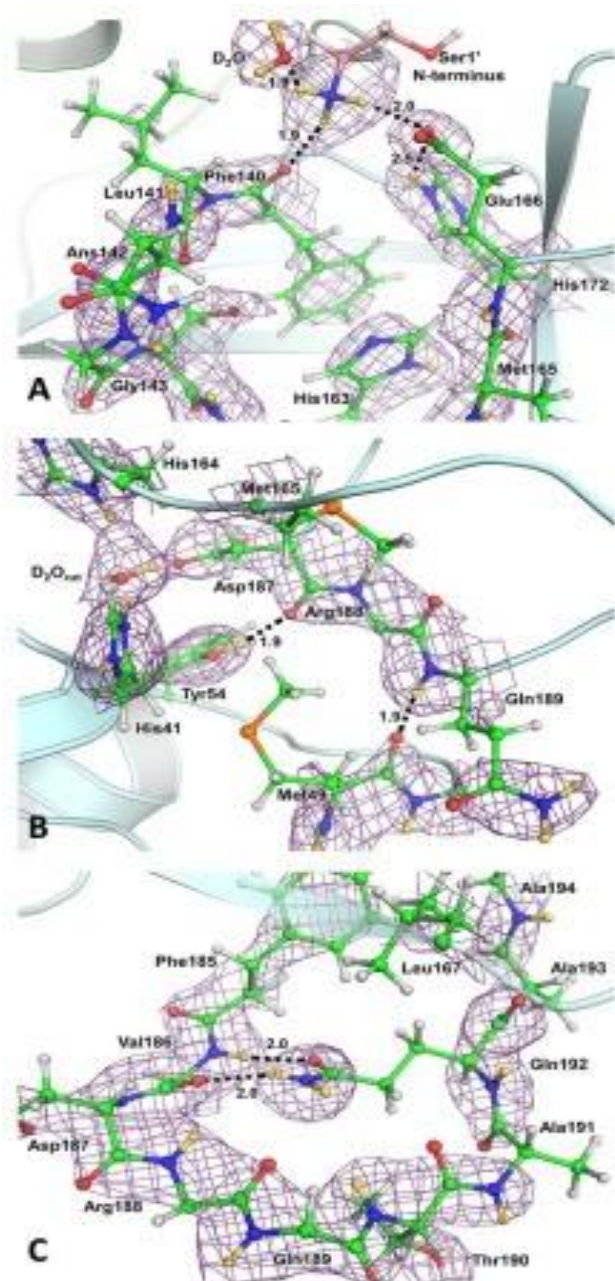


Рис 1.8 Архітектура та водневі зв'язки у субстрат-зв'язуючих підсайтах S1 (A), S2 (B) та S4 (C). Карта ядерної щільності 2Fo - Fc. Водневі зв'язки показані у вигляді чорних пунктирних ліній. Відстань гідроген-кисень вказана в Ангстремах. Бічні ланцюги Val186, Arg188 та Gln189 не візуалізовані. Атоми дейтерію пофарбовані в жовтий колір, тоді як атоми водню мають світло-сірий колір. [20]

### 1.3.2 (б) Субстрат-зв'язуючий підсайт S2

Підсайт S2 є більш гідрофобним, ніж підсайт S1, оскільки він зв'язує гідрофобні залишки Leu або Phe у положенні субстрату P2. Підсайт S2 оточений  $\pi$ -системами His41 та основними ланцюгами, що з'єднують Asp187, Arg188 та Gln189. Він обмежений Met165, тоді як Met49, розташований на короткій спіралі P2, що охоплює залишки Ser46 – Leu50. Met49 є

конформаційно гнучким, звільняючи своє місце в ферменті, що не містить ліганду, щоб дозволити різним групам P2 зайняти цей субсайт при зв'язуванні інгібіторів [35,34,36]. Tyr54 служить основою підсайту S2, і його фенольна гідроксильна група є донором гідрогену у водневому зв'язку з карбонілом Asp187 основного ланцюга.

### 1.3.2 (в) Субстрат-зв'язуючі підсайти S3–S5

Серед субсайтів S3, S4 і S5 лише субсайт S4 має чітко визначену архітектуру. Субсайти S3 і S5 знаходяться на поверхні білка, повністю піддаються впливу основного розчинника і мають чітко визначені межі. Підсайт S3 розташований між залишками Glu166 та Gln189, які знаходяться на відстані > 9 Å. Субсайт S5 знаходиться між Pro168 β-шпильки, що охоплює залишки Met165 – His172, і петлею P5, що складається із залишків Thr190 – Ala194. Субсайт S4 утворений між довгою петлею, що охоплює залишки Phe185 – Ala194, яка виконує роль його основи, і β-шпилькою зверху. [33] Петля повертається на 180 ° при Gln189, при цьому її вторинна структура стабілізується водневими зв'язками між амідом бічного ланцюга Gln192 та атомами головного ланцюга Val186. Бічні ланцюги Leu167 та Phe185 утворюють задню стінку цієї ділянки у внутрішній частині білка, створюючи глибоку, переважно гідрофобну кишеню.

### 1.3.3 Четвертинна структура Мрго SARS CoV 2

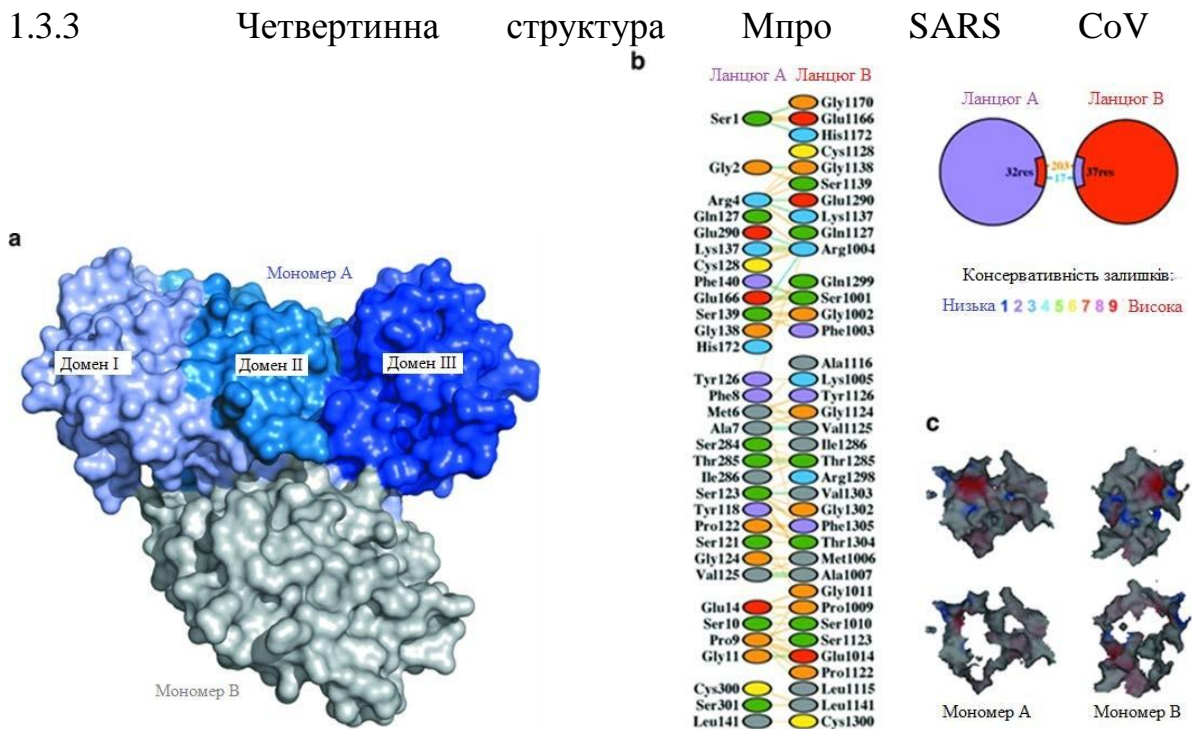


Рис 1.9 Четвертинна структура SARS-CoV Мрго. (А) Поверхнева модель Мрго. (В) Контакти субодиниць на інтерфейсі димеру Мрго. Водневі зв'язки (блакитний), не валентні взаємодії (помаранчевий). (С) Поверхневий потенціал контактуючих регіонів. На нижніх панелях порожнисті області представляють області безпосереднього контакту в субодиницях А і В. Позитивно заряджені області показані синім кольором, а негативно заряджені - червоним.

При утворенні димера N-палець кожного протомера стискається між двома протомерами і створює міжмолекулярні контакти, подібні до тих, що наявні у

TGEV Mpro [35]. Група NH<sub>3</sub> Ser1 одного протомера утворює сольовий місток з Glu166 та водневий зв'язок з карбонільною групою Phe140 іншого протомера. Arg4 утворює міжмолекулярний сольовий місток з Glu290. Ala7 і Val125 утворюють два міжмолекулярні водневі зв'язки. Крім того, NH остову Ser10 може утворювати водневий зв'язок з гідроксильною групою Ser10 на іншому протомері, а NH Gly11 може утворювати міжмолекулярний водневий зв'язок з карбоксильною групою Glu14. Тим часом, бічний ланцюг Met6 здійснює гідروفобну взаємодію з Tyr126 та Phe140 іншого протомера.

N- і C-кінці протеази важливі для підтримки олігомерного стану 3CL<sub>pro</sub>, при цьому більшість взаємодій є або іонними, або гідروفобними. Сольовий місток між Arg4 та Glu290 є важливим елементом міжмономерних взаємодій. Мутація Glu290 до Ala призводить до повної втрати димеризації та каталітичної активності. Видалення перших чотирьох залишків N-кінця призводить до різких втрат димеризації та каталітичної активності, підтверджуючи важливість Arg4 [28]. Мутація Met6 до Ala повністю перешкоджає димеризації та ферментативній активності [16]. Met6 вірогідно утворює гідروفобні взаємодії з Tyr126 та Phe140 протилежного ланцюга. Видалення третього альфа-спірального домену також заважає димеризації та зменшує активність [27-29]. Помічено, що лише третій домен димеризується сам по собі, що призвело до утворення гіпотези, що роллю цього домену може бути регуляція ферментативної активності через димеризацію. [29]

### 1.3.3 (a) Архітектура димерного сайту Mpro SARS CoV 2

Два протомери в гомодимері SARS – CoV-2 3CL Mpro взаємодіють через великий димерний сайт. Протомер 1 утворює складні мережі водневих взаємодій з N-кінцевими залишками 1' – 16', із залишками 118' – 125' β-ланцюга та залишками 137' – 142' петлі протомеру 2; Також існує багато гідروفобних взаємодій в межах димерного сайту. N-кінці двох протомерів зустрічаються на початку короткої α-спіралі, що охоплює залишки Gly11 – Cys16 та Gly11' – Cys16', утворюючи кілька водневих зв'язків із залученням атомів основного та бічного ланцюгів Ser10, Gly11, Ser10', Gly11', і Glu14'. Потім N-кінцева петля простягається через площу протилежного протомера до підсайту S1. Тут Ala7, Phe8, Arg4 і Ser1 створюють водневі зв'язки з Val125, Glu290, Lys137, Phe140 та Glu166. Ser123 та Asn119' β-ланцюга протомеру 2, що містять залишки 118 – 125, мають прямі водневі зв'язки та опосередковану водою взаємодію із C-кінцевими залишками протомеру 1, на додаток до зворотних водневих зв'язків між Val125 та Ala7, утвореними завдяки симетрії 2 порядку димера ензиму. Подібна взаємність водневого зв'язку виявлена також у петлі, що складається із залишків 137' – 144' у протомері 2, що взаємодіє з N-кінцевими залишками Ser1 та Arg4 протомеру

1. Також гідроксил бічного ланцюга Ser139' утворює водневий зв'язок із амідом бічного ланцюга Gln299. [30]

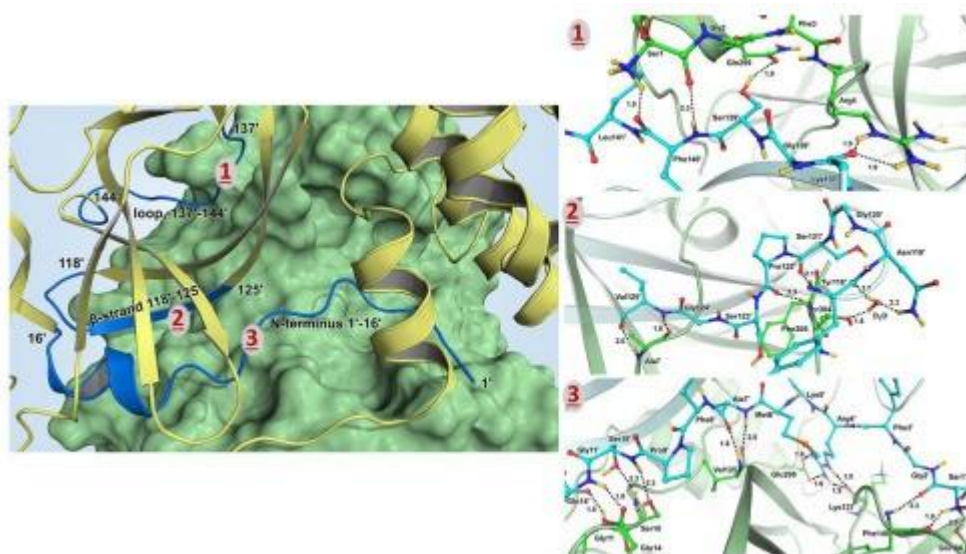


Рисунок 1.10 Водневі зв'язки та взаємодії, опосередковані водою, в димері SARS – CoV-2 3CL Mpro. Ліва панель, Візуалізація, протомер 1 (залишки 1–306) зображений у вигляді зеленої поверхні, протомер 2 (залишки 1'–306 ') забарвлений у жовтий колір. Частина протомера 2, що взаємодіють з протомером 1, пофарбовані в синій колір. Права панель, водневі зв'язки димерного сайту, утворені N-кінцевими залишками 1' – 16 ', зображені на рис. 1; взаємодії β-ланцюга 118' – 125 ', зображені на рис. 2; взаємодії, петлі 137' – 141–, зображені на рис. 3. Водневі зв'язки показані у вигляді чорних пунктирних ліній, відстань - у Ангстремах [20]

#### 1.3.4 Вплив димеризації на конформацію активного центру Mpro

N кінець білка (N палець) не лише виходить за межі каталітичної складки з утворенням зв'язків на додатковому домені того самого протомеру, але також формує взаємодії з залишками протилежного протомеру. З цих міжпротомерних взаємодій дві значно стабілізують димерну структуру, а саме сольовий місток між бічними ланцюгами Arg4 та Glu290 протилежного протомера та гідрофобно-ароматична взаємодія між бічними ланцюгами Met6 та Tyr126 протилежного протомера. [29]

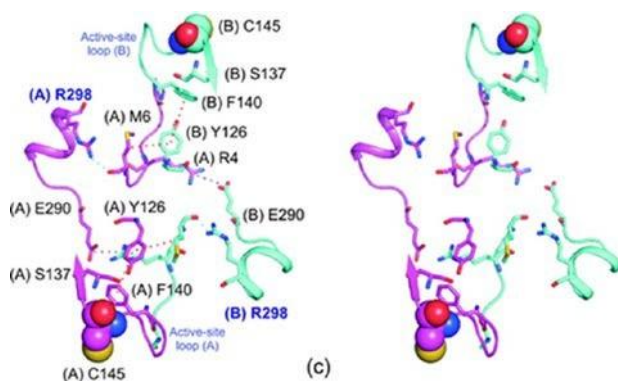


Рис 1.11 Стереозображення мережі взаємодій, відповідальних за підтримку димерної структури, один протомер позначений фіолетовим кольором, а інший блакитним. Водневі зв'язки позначені блакитними пунктирними лініями, гідрофобні взаємодії - червоними пунктирними лініями, а сольові містки - синіми пунктирними лініями. Залишки Cys145 в активних сайтах обох протомерів відображаються у вигляді сфер.

Tyr126 підтримує димеризацію за рахунок ароматично-гідрофобної взаємодії з Met6 протилежного протомеру, а також стабілізує каталітичний механізм за допомогою ароматичної стекінгової взаємодії з Phe140 того самого протомеру. [28]

Залишок Arg298 служить ключовим компонентом для підтримки димеризації шляхом інтеграції N пальця і додаткового домену одного протомеру, що стабілізує точне позиціонування та орієнтацію N пальця та додаткового домену, і є абсолютно необхідним для утворення димеру. [30] Фенільне кільце Phe140 взаємодіє з імідазольним кільцем His163 в нижній частині кишені S1, гарантуючи, що цей гістидиновий залишок залишається незарядженим у широкому діапазоні значень рН.

Багато залишків, утворюючих каталітичний механізм, зазнають значних змін в мутантних структурах протеази, що не містять залишку Arg298. Залишки петель над положеннями 138-144 мають набагато глибші зміни, ніж інші області, оскільки залишки петлі, Gly138-Ser139-Phe140-Leu141, самі безпосередньо контактують з поверхнею протилежного протомера. Подібним чином, антипаралельний  $\beta$ -лист над залишками від 111 до 129, що оточують каталітичні ділянки, також має великі перегрупування, оскільки багато залишків, таких як Tyr126 на цьому  $\beta$ -аркуші, мають прямі контакти з іншим протомером. Проте найбільш виразною характеристикою є утворення короткої 310-спіралі із залишків Ser139-Phe140-Leu141, які в активній структурі ферменту приймають конформацію петлі. Найголовніше, що це перетворення скасовує важливу стекінгову взаємодію між кільцями Phe140

та His163, а також суттєво перекичує конформацію залишків Gly143-Ser144-Cys145, приводячи таким чином до згортання субстрат-зв'язуючої кишенки та оксианіонного отвору.

Мономерні структури що не містять Arg298 також зазнають глибоких структурних змін в інших регіонах, критичних для каталізу. Наприклад, зміни виявляються також щодо залишків 180 - 200, що утворюють підсайт S2. У димерній структурі, що складається з двох асиметричних протомерів, активний протомер без характерної 310-спіралі має велику та глибоку порожнину над активним центром, подібну до повністю активного ферменту. На відміну від цього, згорнутий протомер з характерною 310-спіраллю має порожнину активної зони зі значно зменшеним об'ємом. Більш того, порожнина активної зони структури, не здатної до димеризації повністю фрагментована. Як наслідок, каталітичний механізм мономера не є придатним для зв'язування субстратів і не має місця для розміщення глутаміну на сайті аналога субстрату P1.

#### **1.4. Біоінформатичні методи дослідження у пошуці потенційних ліків. Методи молекулярного докінгу.**

Молекулярний докінг є широко використовуваним, відносно швидким та економічним обчислювальним інструментом для прогнозування *in silico* зв'язування та афінності сполук [37]. Важливість та актуальність методів молекулярного докінгу опосередкована тим, що на ранніх стадіях пошуку нових лікарських препаратів є сенс аналізувати великі бібліотеки сполук. Молекулярний докінг досліджує поведінку малих молекул при зв'язуванні з цільовим білком. Оскільки більша кількість білкових структур визначається експериментально за допомогою рентгенівської кристалографії або ЯМР-спектроскопії, молекулярне стикування все частіше використовується як інструмент пошуку ліків. [40,37,39]

У докінгових дослідженнях сайт зв'язування визначається, як правило, у якості жорсткого білку. Ліганд вкладається в цей сайт зв'язування шляхом гнучкого обертання функціональних груп всередині ліганду та обертання самого ліганду. Електростатична та ван-дер-ваальсова взаємодії розраховуються для різних конформацій, а оціночна функція визначає конформації ліганду, які найкраще підходять за енергією.

У Glide, одній з найбільш розповсюджених програм молекулярного докінгу, задіяній у даному дослідженні, використовується ієрархічна серія фільтрів

для пошуку можливих місць розташування ліганду в області досліджуваного сайту рецептора. Форма та властивості рецептора представлені на сітці кількома різними наборами полів, які забезпечують поступово більш точне оцінювання пози ліганду. Конформаційна гнучкість обробляється в Glide за допомогою розширеного конформаційного пошуку, доповненого евристичним скринінгом, який швидко усуває непридатні конформації, наприклад конформації, що мають внутрішні водневі зв'язки на великих діапазонах. [40]

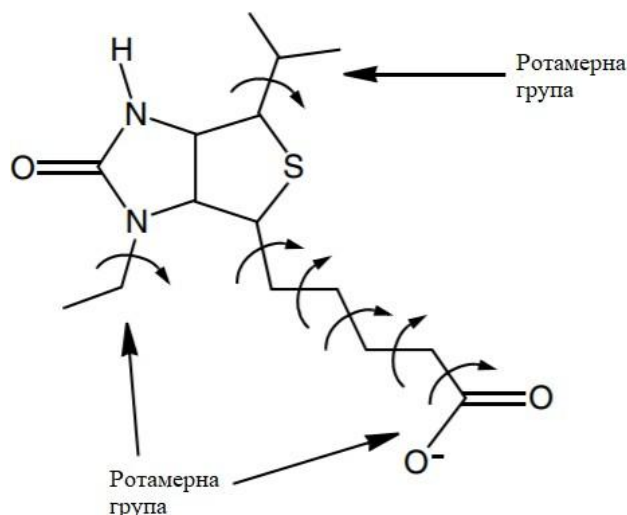


Рис. 1.12 Ротамерні групи ліганда

Кожен ліганд розділяється на область ядра і деяку кількість ротамерних груп. Кожна ротамерна група прикріплена до ядра зв'язком здатним до обертання, але не містить додаткових обертаючих зв'язків. Ядро - це та частина молекули, що залишається, коли кожен кінець ліганду розривається на «останньому» обертаючому зв'язку. Вуглецеві та азотні кінцеві групи, що закінчуються воднем ( $\text{—CH}_3$ ,  $\text{—NH}_2$ ,  $\text{—NH}_3^+$ ), не вважаються обертаючимися, оскільки їх конформаційні зміни мають невелике значення. Під час генерації конформації кожна область ядра представлена набором його конформацій, кількість яких залежить від кількості обертових зв'язків, конформаційно лабільних 5- і 6-членних кілець. Цей набір зазвичай містить менше 500 ядерних конформацій, навіть для досить великих і гнучких лігандів, і набагато менше для більш жорстких лігандів. Перераховується кожен стан ротамера для кожної групи ротамерів, прикріплених до ядра. Ядро та усі можливі конформації ротамерних груп піддаються докінгу як єдиний об'єкт. [42]

Для кожної конформації ядра (або, для жорсткого стикування, кожного ліганду) виконується пошук можливих місць розташування та орієнтації у досліджуваному сайті білка. Пошук починається з вибору «точок» на рівномірно розташованій сітці  $2 \text{ \AA}$ , що охоплює досліджувану область (етап

1). Цей вибір здійснюється наступним чином. Відстані від «точок» до поверхні білка обчислюють у ряді заздалегідь визначених напрямків та сортують у діапазони відстаней ("контейнери") шириною 1 Å. [41] Подібно до цього відстані від центру ліганду (середньої точки двох найбільш широко відокремлених атомів) до поверхні ліганду сортуються в "контейнери" шириною 1 Å. Для даної точки, відстань від точки до рецептора порівнюється з відстанню від центру ліганда до поверхні ліганду. Glide розміщує центр ліганду в певній точці, якщо є достатньо хороший збіг, але пропускає точку, якщо збігу нема. Другий етап ієрархії починається з вивчення розміщення атомів, які лежать на певній відстані від лінії, проведеної між найбільш широко розділеними атомами (діаметр ліганда). Це робиться для попередньо визначеного вибору можливих орієнтацій діаметра ліганду (етап 2). Якщо занадто багато стеричних зіткнень з рецептором, орієнтація пропускається. Далі (етап 2b) розглядається обертання навколо діаметра ліганда, підраховуються взаємодії підмножини, що складається з усіх атомів, здатних утворювати водневі зв'язки або взаємодії ліганд-метал з рецептором (тест підмножини). Якщо цей показник достатньо хороший, оцінюються всі взаємодії з рецептором (Крок 2c). Оцінка в цих трьох тестах проводиться за допомогою дискретизованої версії емпіричної функції оцінки ChemScore Шредінгера [41]. Як і ChemScore, цей алгоритм розпізнає сприятливі взаємодії гідрофобних, водневих зв'язків та зв'язування металів та карає стеричні зіткнення. Ця стадія називається "жадібною оцінкою", оскільки фактична оцінка кожного атома залежить не тільки від його положення щодо рецептора, але й від найкращого можливого балу, який він може отримати, перемістившись на  $\pm 1$  Å по x, y або z осі. Це робиться для того, щоб зменшити вплив на результат великих стрибків на 2 Å у положеннях точка / центр-ліганда. Завершальним етапом на стадії 2 є повторна оцінка найкращих поз "жадібною" оцінкою за допомогою процедури «уточнення» (етап 2г), при якій ліганд в цілому може жорстко рухатися на  $\pm 1$  Å в декартових напрямках. Лише невелика кількість найкращих поз (як правило, 100-400) передається на третій етап ієрархії - мінімізація енергії на попередньо розрахованих OPLS ван дер Ваальсових та електростатичних сітках для рецептора. [56-59]

Мінімізація енергії, як правило, починається з набору сіток ван-дер-Ваальса та електростатичних сіток, які були "згладжені", для зменшення великої енергетичні та градієнтні умови, що виникають в результаті занадто тісних міжатомних контактів. Вона закінчується на повномасштабній незв'язаній енергетичній поверхні OPLS ("відпал").

Невелика кількість поз лігандів з найвищими оцінками піддаються процедурі відбору проб, при якій досліджуються альтернативні кути обертання ядра локальних мінімумів та обертання груп ротамерів, щоб спробувати покращити показник енергії. Нарешті, мінімізовані пози перераховуються за допомогою власної функції підрахунку GlideScore програми Шредінгер.

Формула для обчислення GlideScore:

$$\text{GScore} = 0.05 * \text{vdW} + 0.15 * \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{Rewards} + \text{RotB} + \text{Site}$$

vdW позначає енергію ван дер Ваальса, Coul позначає енергію Кулона, Lipo позначає ліпофільний контакт, HBond позначає водневі зв'язки, Metal позначає зв'язування з металами, BuryP позначає штраф за поховані полярні групи, RotB позначає штраф за іммобілізовані зв'язки що обертаються, Site позначає полярні взаємодії в досліджуваному сайті,  $a = 0,065$  і  $b = 0,130$  - коефіцієнти vdW і Coul. [51]

Вибір найкраще стикованої структури для кожного ліганду здійснюється за допомогою модельної енергетичної оцінки (Emodel), яка поєднує оцінку енергії сітки, спорідненість зв'язування, передбачену GlideScore, та (для гнучкого стикування) енергію внутрішнього напруження для модельного потенціалу, що використовується для керувати алгоритмом конформаційного пошуку. Glide також обчислює спеціально створену оцінку взаємодії енергій Кулона-ван-дер-Ваальса (CvdW), що була сформульована таким чином, щоб уникнути надмірні оцінки взаємодій заряд-заряд за рахунок взаємодій заряд-диполь та диполь-диполь. Цей показник має бути більш придатним для порівняння афінності зв'язування різних лігандів, ніж звичайна енергія взаємодії Кулона-ван-дер-Ваальса. Разом із GlideScore, отримується також і «оцінка докінгу», що представляє собою оцінку GlideScore, доповнену штрафами E<sub>rik</sub>. [52]

Результати отримані за допомогою програми молекулярного докінгу зазвичай уточнюються шляхом симуляції молекулярної динаміки комплексу білок-ліганд.

Моделювання молекулярної динаміки (МД) описує залежні від часу зміни атомних координат білка та ліганду як у зв'язаній, так і в незв'язаній формі, що надзвичайно корисно для отримання інформації про зміну конформаційної ентропії після зв'язування. Моделювання МД дозволяє також досліджувати нерівноважні ефекти, що призводять до перехідних конформерів, які сприяють події зв'язування, але не можуть легко спостерігатися в експериментах. Методи докінга білка-ліганду дозволяють швидко передбачити найбільш сприятливу структуру комплексу та оцінити спорідненість до зв'язування. Більш точний прогноз спорідненості зв'язування можна отримати за допомогою розрахунків вільної енергії, які враховують усі термодинамічно відповідні явища, такі як динаміка / гнучкість білка, явне включення розчинника та різниця між взаємодією білка та ліганду в комплексі та їх взаємодією з водою та протиіони у їх незв'язаних формах. [51, 52, 55]

## РОЗДІЛ 2

### Об'єкти, матеріали та методи дослідження

#### 2.1 Об'єкт дослідження

Отримана з бази даних PDB модель Mpro SARS CoV 2 із кодом 6lu7 використовувалася як об'єкт дослідження.

Модель була описана у статті “Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors.” [53]

- Клас білка: вірусний протеїн
- Організм (и): Тяжкий гострий респіраторний синдром коронавірусу 2, синтетична конструкція
- Система експресії: кишкова паличка BL21 (DE3)
- Наявність мутацій: Ні
- Метод отримання: Рентгенівська дифракція
- Роздільна здатність: 2,16 Å
- Молекулярна вага: 34,51 кДа
- Кількість атомів: 2500

#### 2.2 Бази даних хімічних речовин вільного доступу

Для отримання 3D структур досліджуваних білків та лігандів були використані бази даних PDB та ZINC15

##### 2.2.1 Архів PDB (Protein Data Bank)

Білковий банк даних (PDB) - архів структурних даних біологічних макромолекул. Він включає дані, отримані за допомогою рентгенівської кристалографії та спектрометрії ядерно-магнітного резонансу (ЯМР), представлені біологами та біохіміками з усього світу. В даний час PDB перебуває у юрисдикції Всесвітнього банку білкових даних (wwPDB), мережі з чотирьох організацій - Дослідницька колабораторія з структурної біоінформатики (RCSB) PDB (США), PDB в Європі (PDBe) (Європа), PDB Японія (PDBj) (Японія) та Банк даних біологічного магнітно-резонансного банку (BMRB) (США) [54]

Із бази даних PDB було отримано структуру 6lu7 (The crystal structure of COVID-19 main protease in complex with an inhibitor N3 Method отримання: X-RAY DIFFRACTION Різкість: 2.16 Å) а також структури 4YLU (X-ray structure of MERS-CoV nsp5 protease bound with a non-covalent inhibitor Method: X-RAY DIFFRACTION Resolution: 2.10 Å) [74] 4YOG (HKU4-3CLpro bound to non-covalent inhibitor 3B Method: X-RAY DIFFRACTION Resolution: 2.00 Å) [75] що в подальшому використовувалися для вирівнювання послідовностей та дослідження гомології.

## 2.2.2 База даних вільного доступу ZINC15

Для пошуку лігандів та завантаження бібліотеки 3D структур для аналізу використовувалась база комерційних сполук ZINC15.

ZINC15 - це база даних загального доступу та набір інструментів, розроблений для забезпечення швидкого доступу до сполук готових для віртуального скринінгу, виявлення лігандів, скринінгу фармакофорів, бенчмаркінгу, та розробок полів. ZINC15 спирається на сторонні бази даних, такі як ChEMBL, HMDB, DrugBank та <https://ClinicalTrials.gov>, для анотації сполук. [48]

Властивості каталогів і, в свою чергу, молекул, які вони містять, класифікуються у трьох незалежних серіях: Біогенні: Ендогенні метаболіти людини> Нелюдські метаболіти> Природні продукти> Невідома біогенність. Тип ліків: затверджена FDA> Світові препарати> Дослідження> У людини> In vivo> У клітинах> In vitro> невідома біоактивність. У межах кожної серії кожна сполука класифікується за найвищим рівнем, який вона досягає за членством у каталозі.

Для виконання дослідження із бази даних були завантажені бібліотеки сполук із підкласів ZINC15 anodyne (не містить молекул зі структурами PAINS) (8 міл. Структур), world (Лікарські засоби, затверджені у великих юрисдикціях), fda (каталоги, що містять лише препарати, схвалені FDA) , biogenic (метаболіти та натуральні продукти) та in-man (речовини що споживаються людиною/мають відношення до фізіології людини, наприклад метаболіти, нутрицевтики або речовини що містяться у харчових продуктах).

## 2.3 Система молекулярної візуалізації PyMol

Для порівняння структури білка Mpro SARS CoV 2 із 3 аналогічними білками шляхом вирівнювання послідовностей була використана програма PyMol.

PyMOL, це крос-платформний інструмент для молекулярної графіки, що широко застосовується для тривимірної візуалізації білків, нуклеїнових кислот, малих молекул, електронної густини, поверхонь та траєкторій. [55]

### 2.3.1 Структурне вирівнювання у PyMol з використанням функції Align

Функція Align виконує вирівнювання білкових послідовностей з наступною структурною суперпозицією, а потім виконує нуль або більше циклів уточнення, щоб відкинути структурні неспівпадення, знайдені під час підгонки. align здатен якісно виконувати аналіз пар білків з високою схожістю послідовностей (ідентичність> 30%). [56] Для порівняння білків з нижчою ідентичністю послідовностей необхідно використовувати команди super чи sealign. Оскільки у попередніх дослідженнях гомології протеаз Mpro

SARS CoV 2, Mpro SARS CoV, Bat CoV 3CLpro та MERS CoV nsp5 була спрогнозована висока схожість послідовностей досліджуваних білків, для вирівнювання було обрано функцію align:

```
fetch blu7 4ylu, async=0
```

```
# 1) default with outlier rejection
align blu7, 4ylu
```

```
# 2) with alignment object, save to clustalw file
align blu7, 4ylu, object=alnobj
save alignment.aln, alnobj
```

```
# 3) all-atom RMSD (no outlier rejection) and without superposition
align blu7, 4ylu, cycles=0, transform=0
```

## 2.4 Розрахункова платформа віртуального скринінгу *Schrödinger Suites 2018-1*

Докінгові дослідження проводились із використанням програмного забезпечення Schrödinger на платформі молекулярного моделювання Maestro (Schrödinger Release 2019–4: Maestro, Schrödinger, LLC, New York, NY, 2019. версія 12.2.012). [60]

Програма віртуального скринінгу (Virtual Screening Workflow) у Шредингері запускає послідовність завдань для скринінгу великих бібліотек сполук на зв'язування з одною або кількома цілями. Робочий процес включає підготовку лігандів за допомогою LigPrep, фільтрування за допомогою ligfilter за властивостями QikProp або іншими структурними властивостями, а також стикування Glide на трьох рівнях точності, HTVS, SP та XP. [61]

Для виконання роботи використовувалася версія програми *Schrödinger Suites 2018-1* (<https://www.schrodinger.com/downloads/releases?release=276798>)

### 2.3.1 Утворення сітки рецептора (Receptor grid generation)

Для ефективного проведення розрахунків стикування, Glide працює не з самою структурою білка, а з сіткою, що представляє властивості структури (тобто електростатичний потенціал, що генерується в кожній точці сітки, взаємодії ван дер Ваальса). [57] Форма та властивості рецептора представлені на сітці кількома різними наборами полів, які забезпечують поступово більш точне оцінювання пози ліганду. Для дослідження було створено сітку рецептора на основі підготовленої структури blu7.

Перед використанням у якості рецептора для стикування, білкова структура була оброблена. Операції обробки включали (1) додавання атомів водню, (2) розподіл атомних зарядів, (3) елімінацію молекул води, які не беруть участі у зв'язуванні лігандів, (4) заміщення селеноцистеїнів цистеїнами. В ході даного дослідження необхідна підготовка була виконана через майстер підготовки білка з меню Workflows. [58]

Після усунення недоліків структури білка було здійснено протонування за допомогою модуля Ерік програмного пакету Schrödinger. Сітки для активного центру та димеризаційного сайту протеази були сформовані шляхом ручного відбору амінокислот що приймають участь у димеризації, формують активний центр, відповідно. Процес проходив при рН = 7 (+2)

Положення амінокислот, що приймають участь у димеризації Mpro:

1 **sgfrkmafps gkve**gcmvqv tegttnlgl wlddvycpr hvictsedml npnyedllir  
 61 ksnhnflvqa gnvqlrvigh smqncvklk vdtanpktpk ykfv**ri**qpgq **tfsvlacy**ng  
 121 spsgvyqcam rpnfti**kgsf lng**scgsvgf nidydcvsfc ymhmel**ptg v**hagtdlegn  
 181 fygpfvdrqt aqaagtdtti tvnvlawlya avingdrwfl nrftttlndf nlvamkynye  
 241 pltqdhvdil gplsaqtgia vldmcaslke llqngmng**rt ilgsallede** ftpfdv**rqc**  
 301 **sgvtfq**

Номери амінокислот, обраних для формування сітки димеризаційного сайту:

sgrkmapsgkerifsvlacykgsflnepghtgsalerqgvf  
 1,2,4,5,6,7,9,10,11,12,14  
 115,116, 122-128  
 137-142  
 166,168,170,172  
 280,283-286,290,298-299,302-303,305.

Номери амінокислот, обраних для формування сітки активного сайту:

20 – 28, 39 – 42 77 – 47 48 – 54 118 119 136 138 – 147 161 163 – 173 181 185 –  
 193

Для пом'якшення потенціалу неполярних частин рецептора, було виконано масштабування радіусів Ван дер Ваальса атомів білка з частковим атомним зарядом (абсолютним значенням), меншим за вказане відсічення. Усі інші атоми рецептора не були масштабовані. Коефіцієнт масштабування: 1.0  
 Часткове відсічення заряду: 0.25

### 2.3.2 Підготовка лігандів до докіну за допомогою функції LigPrep

Процес підготовки лігандів LigPrep складається з ряду етапів, які виконують перетворення, застосовують виправлення до структур, генерують зміни на структурах, усувають небажані структури та оптимізують їх.

#### 1. Перетворення формату структури.

файл вхідної структури перетворюється у формат Maestro за допомогою `sdconvert`. Відношення, зазначені у файлі `SD`, перетворюються на хіральності, які зберігаються як властивості у файлі Maestro.

#### 2. Додавання атомів водню.

Структурам, які мають невиражені атоми водню, необхідно додати атоми водню, перш ніж можна буде провести мінімізацію тривимірних структур. Атоми водню додаються таким чином, що відповідає певному силовому полю. Цей крок виконується програмою `applyhtreat`, що використовується панеллю Hydrogen Treatment в Maestro.

#### 3. Нейтралізація заряджених груп.

Заряджені групи необхідно нейтралізувати перед генеруванням станів іонізації. Нейтралізація здійснюється шляхом додавання або видалення йонів водню.

#### 4. Генерування станів іонізації.

Усі конформації, що існують у певному діапазоні рН мають бути доступні програмі. На даному етапі генеруються різні стани іонізації для кожної структури. Цьому кроку повинен передувати етап нейтралізації. Був обраний рН = 7 (+-2). Виконано з допомогою `Erik`

#### 5. Генерування таутомерів.

Як і з іонізацією, важлива наявність різноманітних тавтомерних станів. Програма-таутомеризатор генерує різні таутомери для кожної структури.

#### 6. Визначення хіральності.

На цьому етапі ідентифікуються всі хіральні атоми в структурах та надаються варіанти збереження хіральності або генерування додаткових стереохімічних ізомерів з тією ж молекулярною формулою. Крок виконується програмою-стереоізатором. Було згенеровано 4 або менше стереоізомерів на ліганд

#### 10. Утворення низькоенергетичної конформації кілець.

Коли інформація про конформацію кільця недоступна, необхідно спеціально генерувати конформації кілець з низькою енергією. Процес відбувається за допомогою `ring_conf`. Генерувалося по 1 низькоенергетичній конформації кілець на ліганд

11. Видалення проблематичних сполук.

Структури, які можуть спричинити подальші помилки в обробці або при мінімізації енергії структур видаляються за допомогою `premin`.

12. Оптимізація геометрії.

Геометрії сформованих структур оптимізуються з використанням обчислювальної програми MacroModel™, `bmin`.

### 2.3.3 Віртуальний скринінг бібліотеки лігандів (Virtual screening workflow)

Було проведено скринінг набору підготовлених лігандів, використовуючи сітку, яка була попередньо створена на основі структури PDB 6LU7.

Віртуальний скринінг попередньо підготовленої бібліотеки лігандів був розпочатий завантаженням структур лігандів у форматі `.sdf`. Файл у форматі `.zip` використовувався як вхідний файл сітки рецептора. Стикування рецептор-ліганд проводилося у три етапи:

HTVS: найшвидший метод аналізу молекул, був використаний через необхідність провести скринінг великої кількості малих з'єднань.

SP метод: стандартна форма стикування, більш жорстке оцінювання ніж на етапі HTVS.

XP метод: забезпечує найточніші прогнози молекулярних взаємодій.

Докінг відбувався з відбором 10 % сполук на етапах HTVS та SP і 40 % на XP етапі. Вихідні файли було збережено в форматі Maestro `.maegz`.

## 2.5 Пакет програм для моделювання фізико-хімічних процесів GROMACS.

Симуляції молекулярної динаміки проводилися для сполук з найкращими значеннями вільної енергії зв'язування отриманими після докінгу у Schrodinger. Молекулярна динаміка (МД) - це метод імітації молекулярного руху, який часто застосовують для дослідження великих біомолекул, таких як білки або нуклеїнові кислоти. Моделювання МД описує еволюцію в часі молекулярної системи, наприклад, білка, шляхом розв'язування рівнянь руху Ньютона для всіх атомів у системі. Подібне моделювання може точно описати динаміку біологічних систем, використовуючи три наближення; (а) наближення Борна-Оппенгеймера, (б) наближення, що полягає в розгляданні ядер як класичних частинок, і (в) використання емпіричного силового поля для опису взаємодії між частинками.

МД симуляції для цієї роботи проводились в програмному пакеті GROMACS

GROMACS – це програмний пакет молекулярної динаміки для моделювання білків, ліпідів та нуклеїнових кислот, розроблений на кафедрі біофізичної хімії Університету Гронінгена. GROMACS може працювати на центральних процесорах (ЦП) та графічних процесорах (ГП). [44]

### 2.6.1 Моделювання молекулярної динаміки комплексу білок-ліганд

Процес моделювання молекулярної динаміки поділяється на кілька етапів:

1. Налаштування (завантаження даних, сольватація, тобто додавання води та іонів)
2. Мінімізація енергії білка
3. Збалансування розчинника навколо білка (з допомогою NVT і NPT)
4. Створення симуляції, що виробляє траєкторію.

Траєкторія - це бінарний файл, у якому містяться атомні координати на декількох часових етапах і, отже, який відображає динамічний рух молекули.

#### 2.6.1 (a) Отримання структури білка Mpro із PDB

Для початку симуляції MD файл координат структури білка у форматі PDB було завантажено з веб-сайту RCSB. Після завантаження структури PDB її було перевірено та переформатовано за допомогою звичайного текстового редактора Notepad (Windows). Структура PDB була перевірена на відсутність певних атомів чи залишків та присутність лігандів або молекул води. Для моделювання MD була використана тривимірна структура з кодом PDB 6LU7. У структурі PDB (6LU7) наявні усі атоми / залишки. Ліганди та кристали води були видалені перед початком моделювання MD.

#### 2.6.1 (a) Створення топології з файлу PDB

Після підготовки вхідної структури PDB було використано перший модуль GROMACS, `pdb2gmx`, для перетворення файлу PDB у формат файлу молекулярної геометрії GROMACS (`.gro`), файл топології (`.top`) та файл обмеження положення (`.itp`). Для моделювання Mpro було обрано силове поле CHARMM36, модель води SPC / E, було обрано стандартні варіанти для всіх станів протонування залишків, дисульфідних містків тощо.

`Mpro.gro` - це структурований у форматі GROMACS файл, який містить координати всіх атомів, визначених у силовому полі. Відсутні атоми водню додаються `pdb2gmx` за замовчуванням. Файл `topol.top` - це системна топологія, що містить молекулярний опис, такий як молекулярні параметри, зв'язки, силове поле та заряди. Файл `posre.itp` містить інформацію, яка використовується для обмеження позицій важких атомів на етапі вирівнювання положення-обмеження.

### 2.6.1 (б) Підготовка структури ліганду

Топологія лігандів була створена на сервері CGenFF. CGenFF вимагає файл у форматі Sybyl .mol2 у якості вхідного файлу для збору елементарної інформації про тип атомів та зв'язків. Для створення файлу .mol2 була використана програма Avogadro.

Для сортування молекулярних зв'язків у файлі було використано скрипт `sort_mol2_bonds.pl`, оскільки можливе виникнення проблем при побудові правильної топології з відповідними координатами, якщо зв'язки не перераховані у порядку зростання.

```
perl sort_mol2_bonds.pl lig.mol2 lig_fix.mol2
```

Файл "lig\_fix.mol2" використовувався на наступному кроці.

Файл потоку CHARMM був сформований через сервер CGenFF. Файл потоку CHARMM містить всю інформацію про топологію - типи атомів, заряди та з'єднання. У ньому також є розділи для додаткових пов'язаних параметрів, які були створені за аналогією для будь-яких внутрішніх взаємодій, не охоплених силовим полем. CGenFF також надає оцінки показників для кожного параметра, тобто оцінку того, наскільки надійний призначений параметр. Всі оцінки менше 10 вважаються прийнятними для використання. Значення від 10 до 50 означають, що певна перевірка топології є обґрунтованою, і будь-які значення, більші за 50, як правило, вимагають ручної параметризації

Оцінки параметрів для всіх досліджуваних лігандів були нижчими за 10, що свідчить про те, що топологія має дуже хорошу якість і може бути використана безпосередньо для моделювання.

Скрипт `cgenff_charmm2gmx.py` був використаний для перетворення файлу у форматі, придатні для GROMACS

```
python cgenff_charmm2gmx.py lig lig_fix.mol2 lig.str charmm36-feb2021.ff
```

Цей скрипт вводить нові пов'язані параметри, які не є частиною існуючого силового поля, і ці параметри записуються у файл, який називається "lig.prm", який має формат .itp-файлу GROMACS. Топологія ліганду була записана у "lig.itp", що містить визначення ліганду [moleculetype].

### 2.6.1 (в) Побудова комплексу ліганд-Мpro

З pdb2gmx було отримано файл під назвою " Mpro.gro", який містить оброблену структуру білка, що відповідає силовим полям. Із cgenff\_charmm2gmx.py було отримано lig\_ini.pdb, який має всі необхідні атоми Н і відповідає іменам атомів у топології lig. Цей файл .pdb було перетворено у формат .gro за допомогою editconf:

```
gmx editconf -f lig_ini.pdb -o lig.gro
```

Mpro.gro було скопійовано до нового файлу complex.gro. Далі у файл complex.gro було скопійовано координатний розділ lig.gro, нижче останнього рядка атомів білка, і перед векторами коробки. Оскільки було додано певну кількість атомів у файл .gro, була збільшена цифра у другому рядку complex.gro, щоб відобразити цю зміну.

Для включення параметрів ліганду lig в топологію системи було вставлено рядок #include " lig.itp" у topol.top після включення файлу обмеження позиції. Включення обмежень положення означає кінець розділу молекулярного типу "Білок".

Ліганд вводить нові двогранні параметри, які були записані в "lig.prm" скриптом cgenff\_charmm2gmx.py. Вгорі topol.top було вставлено оператор #include, щоб додати ці параметри.

Розміщення виразу #include є важливим - воно повинно з'являтися перед будь-яким [ moleculetype], оскільки всі параметри повинні бути визначені перед тим, як молекули можуть бути побудовані. Він також повинен з'явитися після оператора #include для батьківського силового поля, оскільки всі типи атомів повинні бути відомі перед тим, як можуть бути введені зв'язані параметри, які їх використовують.

Останнє коригування було зроблено в директиві [ molecules ]. Щоб врахувати той факт, що в complex.gro є нова молекула, її було вказано у даній директиві.

### 2.6.1 (г) Створення коробки моделювання

Модулем для визначення розмірів коробки моделювання є editconf, а команда для створення коробки:

```
gmx editconf -f complex.gro -o newbox.gro -bt dodecahedron -d 1.0
```

Дана команда центрує білок у коробці (-c) та розміщує його принаймні на 1,0 нм від краю коробки (-d 1,0). Тип коробки визначається як дванадцятигранник. Нова конформація записується у файл newbox.gro

### 2.6.1 (д) Внесення води та розчинника у систему

Щоб імітувати фізіологічне середовище, білок повинен бути сольватованим. Це виконується за допомогою невеликої заздалегідь вирівняної системи координат води, яка повторюється над коробкою, з видаленням молекул води, що перекриваються. Для сольватації білка був використаний модуль `solvate`, який додає необхідну кількість молекул води навколо білка на основі типу коробки, вказаного на етапі `editconf`.

```
gmx solvate -cp newbox.gro -cs spc216.gro -p topol.top -o solv.gro
```

Сольватний модуль розчиняє білкову конфігурацію у ємності з молекулами розчинника. Використовується поле, вказане у файлі координат білка (`-cp`), що є результатом попереднього кроку `editconf`, а конфігурація розчинника (`-cs`) є частиною стандартної установки GROMACS. `Spc216.gro` - загальна врівноважена триточкова модель розчинника, яка може бути використана як конфігурація розчинника для моделей води SPC, SPC / E або TIP3P. Координати води беруться з водної системи SPC, а прапор `-p` додає нову топологію води до файлу топології (`.top`). Створена конфігурація буде записана на `solv.gro`, що містить координати білка та молекул води.

### 2.6.1 (є) Нейтралізація системи

На етапі `pdb2gmx` заряджений білок знаходиться в сольватованій системі. Перед моделюванням цю систему слід нейтралізувати. Для цього зрівноважуючі іони слід додавати відповідно до загального заряду системи за допомогою модуля `genion`. `genion` зчитує топології та замінює молекули води на іони.

Вхідний файл для модуля `genion` має розширення `.tpr`, яке також називається файлом попереднього процесора `gromacs`. Цей файл створений модулем `grompp` GROMACS, попереднім процесором GROMACS. `grompp` збирає параметри імітації (`.mdp`), топологію (`.top`) та координати (`.gro`) у єдиний вхідний файл (`.tpr`). Файл із розширенням `.mdp` - це файл параметрів молекулярної динаміки, що містить ряд налаштувань для керування ходом, мінімізації енергії, контролю вихідної потужності, пошуку сусідів, електростатики, VdW, температурного зв'язку, муфти тиску тощо.

```
gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr
```

Результатом є опис системи моделювання на атомному рівні у двійковому файлі `ions.tpr`, який містить усі параметри для всіх атомів у системі. Потім цей файл передається в модуль `genion` командою:

```
gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -pname NA -name CL -neutral
```

У команді `genion` файл `.tpr` надається як вхід (-s), файл `.gro` генерується як вихід (-o), а топологія (-p) оновлюється для відображення видалення молекул води та додавання іонів. Позитивні та негативні іонні імена визначаються прапорцями `-pname` та `-nname` відповідно.

### 2.6.1 (ж) Мінімізація енергії

Під час налаштування системи моделювання може виникнути неприродний стрес, наприклад, якщо два атоми випадково розташовуються занадто близько один до одного. Це може призвести до великих сил, що діють на частинки, які дестабілізують систему в процесі моделювання. Мінімізація енергії (ЕМ) зазвичай використовується для усунення цих стеричних зіткнень або невідповідної геометрії та вдосконалення експериментальних конструкцій з низькою роздільною здатністю. GROMACS підтримує різні алгоритми мінімізації, і найчастіше використовуються крутий спуск та спряжений градієнт. Алгоритм найкрутішого спуску є найшвидшим у видаленні найбільших деформацій у системі, але зближується повільно, коли наближається до мінімуму. Градієнт кон'югатів повільніший за найкрутіший спуск на ранніх стадіях мінімізації, але стає більш ефективним ближче до мінімуму енергії. Загальноприйнятим є мінімізація початкової енергії за допомогою ефективного методу найкрутішого спуску та подальша мінімізація за допомогою більш складного методу, такого як алгоритм спряженого градієнта. Процес для ЕМ нагадує додавання іонів. На цьому етапі знову було використано `grompp` для збірки структури, топології та параметрів моделювання у двійковий вхідний файл (`.tpr`). Замість того, щоб передавати файл `.tpr` до `genion`, було запущено мінімізацію енергії через GROMACS MD, `mdrun`. Команда `grompp` спочатку використовується для обробки та підготовки вхідного файлу, необхідного для найнижчого циклу мінімізації енергії спуску.

```
gmx grompp -f em1.mdp -c 1QGD_solv_ions.gro -p topol.top -o em1.tpr
```

Після того, як файл `.tpr` було підготовано, наступна команда:

```
gmx mdrun -v -deffnm em1
```

Прапор `-deffnm` визначає префікс імені файлу за замовчуванням, який використовується для всіх вхідних та вихідних файлів. Відповідні закінчення файлів будуть додані до різних вхідних та вихідних файлів, тому їх не потрібно вказувати окремо. На цьому кроці будуть виведені такі файли. Файл `em1.log` містить детальну інформацію про кілька енергетичних термінів протягом моделювання та дає аналіз продуктивності в кінці. Файл `em1.gro` містить координати остаточної конфігурації системи. Файл `em1.trr` містить координати та сили для кожного вихідного кадру. Файл `em1.edr` містить інформацію про різні енергетичні терміни та, якщо це можливо для моделювання, інформацію про температуру, тиск, об'єм та деякі інші величини під час моделювання. ЕМ із методом спряженого градієнта проводиться після методу найкрутішого спуску за допомогою наступних команд: `gmx grompp -f em2.mdp -c em1.gro -p topol.top -o em2.tpr gmx mdrun_d -v -deffnm em2` Для мінімізації з використанням метод спряженого градієнта, GROMACS, скомпільований з подвійною точністю, слід використовувати з модулем `mdrun_d`.

### Врівноваження з обмеженням положення

Після мінімізації була отримана стартова структура для моделювання. Для початку реального моделювання динаміки розчинник та іони навколо білка слід збалансувати та довести до температури, при якій можна проводити симуляцію. На цьому етапі положення атомів білка стримуються. Врівноваження проводиться у два етапи. Спочатку систему доводять до правильної температури на основі кінетичних енергій, а потім до системи застосовують правильний тиск, поки вона не досягне належної щільності. Перший етап проводиться в ансамблі NVT, де кількість частинок, об'єм і температура є постійними. У NVT температура системи повинна сягати плато при бажаному значенні. Другий етап проводиться в ансамблі NPT, де кількість частинок, тиск і температура постійні, що найбільше нагадує експериментальні умови. На цьому етапі тиск стабілізується, а отже, і щільність системи.

### Стримування ліганду

Щоб стримувати ліганд, необхідно створити для нього топологію обмеження положення. По-перше, було створено індексну групу для `lig`, яка містить лише його неводневі атоми:

```
gmx make_ndx -f lig.gro -o index_lig.ndx
```

...

> 0 &! a H \*

> q

Потім було запущено модуль `genrestr` і обрано нову індекс групу:

```
gmx genrestr -f jz4.gro -n index_jz4.ndx -o posre_jz4.itp -fc 1000 1000 1000
```

Дану інформацію необхідно включити до топології. Наступні рядки було додано до топології:

```
; Ligand position restraints
```

```
#ifdef POSRES
```

```
#include "posre_jz4.itp"
```

```
#endif
```

### Збалансування NVT

Файл `posre.itp`, згенерований `pdb2gmx`, використовується для вирівнювання обмеження положення, яке застосовує силу стримування положення до важких атомів білка. Це дозволяє збалансувати розчинник навколо білка, без додавання змінних структурних змін у білку. Зазвичай для вирівнювання обмеження положення достатньо 50–100 пс. Для ансамблю NVT температуру контролюють за допомогою V-масштабування, еталонна температура становить 300К, а муфта тиску вимкнена.

### Урівноваження NPT

Файл `.mdp` для вирівнювання NPT 50 пс суттєво не відрізняється від файлу параметрів, який використовується для вирівнювання NVT. Муфта тиску повинна додаватися і контролюватися за допомогою баростату Паррінелло-Рахмана. Швидкості зчитуються з попередньої траєкторії, тому генерація швидкості повинна бути вимкнена за допомогою `gen_vel=no`.

```
gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr gmx mdrun -deffnm npt
```

### 2.6.1 (з) Отримання MD симуляції комплексу Mpro-ліганд

Після завершення двох фаз врівноваження система збалансована при бажаній температурі та тиску і готова до запуску моделювання для збору даних. Файл

.mdp параметра потрібно відрегулювати відповідно. Під час циклу моделювання обмежувачі положення відключені. Моделювання також проводиться під ансамблем NPT, але час моделювання значно довший, ніж процес рівноваги NPT

### 2.6.1 (є) Подальша обробка системи

Для перекалібрації білка і зміни положення молекули в елементарній комірці, щоб відновити бажану ромбічну додекаедричну форму, було використано trjconv:

```
gmx trjconv -s md_0_10.tpr -f md_0_10.xtc -o md_0_10_center.xtc -center -pbc mol -ur compact
```

Для центрування було обрано опцію "Блок" , опція "Система" була обрана як бажаний результат. Для ще більш плавної візуалізації було виконано ще одну підгонку:

```
gmx trjconv -s md_0_10.tpr -f md_0_10_center.xtc -o md_0_10_fit.xtc -fit rot+trans
```

"Остов" було обрано для least-squares fitting до білкового остову, "Система" була обрана як бажаний результат.

### 2.6.2 Отримання значень RMSD

Середньоквадратичне відхилення (RMSD) певних атомів у молекулі щодо еталонної структури може бути розраховане за допомогою програми gmx rms шляхом least-square fitting досліджуваної структури до структури-референсу ( $t_2 = 0$ ) і подальшого обчислення середньоквадратичного значення.

$$RMSD(t_1, t_2) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \| \mathbf{r}_i(t_1) - \mathbf{r}_i(t_2) \|^2 \right]^{\frac{1}{2}}$$

де  $M = \sum N_i = 1 m_i$  і  $r_i(t)$  - положення атома  $i$  в момент часу  $t$ . Для припасування не потрібно використовувати ті самі атоми, що і для розрахунку RMSD; напр. білок, як правило, пасується на атомах хребта (N, C), але RMSD можна обчислити на хребті або на цілому білку.

Крім того, RMSD можна обчислити за допомогою методу, що не потребує припасування, за допомогою програми gmx rmsdist:

$$RMSD(t) = \left[ \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{r}_{ij}(t) - \mathbf{r}_{ij}(0)\|^2 \right]^{\frac{1}{2}}$$

де відстань  $r_{ij}$  між атомами в момент часу  $t$  порівнюється з відстанню між тими самими атомами в момент часу  $0$ .

У даній роботі використовувався метод `gmx rms` із припасуванням до комплексу білок-ліганд. Під час отримання RMSD за методом `gmx rms` кожна структура з траєкторії (`-f`) порівнюється з еталонною структурою. Еталонна структура взята з файлу структури (`-s`).

### 2.6.3 Отримання значень RMSF

RMSF означає середньоквадратичне коливання. Це чисельне вимірювання, подібне до RMSD, але замість того, щоб вказувати позиційні відмінності між цілими структурами в часі, RMSF - це розрахунок індивідуальної гнучкості залишків або того, наскільки конкретний залишок рухається (коливається) під час моделювання.

`gmx rmsf` обчислює середньоквадратичну флуктуацію (RMSF, тобто стандартне відхилення) атомних позицій на траєкторії (поставляється з `-f`) після (необов'язково) припасування до еталонної структури (поставляється з `-s`).

### 2.6.4 Визначення енергії взаємодії Mpro-ліганди

Для кількісної оцінки сили взаємодії між Mpro та лігандом було розраховано енергію незв'язаної взаємодії між цими двома речовинами. GROMACS має здатність розкладати незв'язані енергії короткого діапазону між будь-якою кількістю визначених груп.

Розрахунок енергії взаємодії здійснюється за допомогою `energygrps` у файлі `.mdp`.

```
gmx grompp -f ie.mdp -c npt.gro -t npt.cpt -p topol.top -n index.ndx -o ie.tpr
```

Потім було використано `mdrun` з опцією `-rerun` для перерахунку енергій з існуючої траєкторії моделювання:

```
gmx mdrun -deffnm ie -rerun md_0_10.xtc -nb cpu
```

Extract the energy terms of interest via the energy module. The terms we are interested in are Coul-SR:Protein-JZ4 and LJ-SR:Protein-JZ4.

Було визначено енергетичні умови Coul-SR: Protein-lig, LJ-SR: Protein-lig, Coul-14: Protein-lig та LJ-14: Protein-lig за допомогою енергетичного модуля.

```
gmx energy -f ie.edr -o interaction_energy.xvg
```

## **2.5 Програмні забезпечення для визначення біодоступності та потенційної токсичності досліджуваних речовин. FAF-Drugs4. pkCSM.**

### 2.5.1 FAF-Drugs4

Для отримання профілю ADME-Tox, досліджувані ліганди було завантажено у форматі структурних даних (SDF) у систему FAF-Drugs4.

FAF-Drugs4 (безкоштовний інструмент фільтрації ADME-Tox) - це програма для фільтрації великих складних бібліотек сполук. Інструмент може виконувати обчислювальне прогнозування деяких властивостей ADME-Tox (адсорбція, розподіл, метаболізм, екскреція та токсичність) з метою сприяння відбору речовин перед хімічним синтезом або впорядкуванням. FAF-Drugs4 використовує заздалегідь визначені фільтри. [63]

протокол включає три окремі етапи: (1) попередній етап курації даних перед процесом фільтрації, який необхідний для стандартизації молекулярних структур та видалення солей, протиіонів, неорганічних речовин, сумішей та повторюваних сполук, (2) основна процедура, яка обчислює різні параметри та (3) остання частину, яка збирає результати, генерує таблиці CSV (обчислювані дескриптори ПК, структурні оповіщення, PAINS) та пише вихідні файли, відфільтровані SDF. [64]

### 2.5.2 pkCSM.

Для подальшого, більш детального дослідження лікоподібності лігандів сполуки з найкращими показниками вільної енергії зв'язування були завантажені у програму pkCSM в форматі smiles.

pkCSM - це програма для прогнозування фармакокінетичних властивостей, яка спирається на визначення сигнатур за допомогою графіків. Вони кодують схеми відстані між атомами і використовуються для представлення малої

молекули та для підготовки прогнозувальних моделей. Сигнатури pkCSM успішно використовувались у п'яти основних різних класах фармакокінетичних властивостей для розробки моделей прогностичної регресії та класифікації.

## РОЗДІЛ 3

### Результати та їх обговорення

#### 3.1 Вирівнювання послідовностей Mpro SARS CoV 2, Mpro SARS CoV, Bat CoV 3CLpro та MERS CoV nsp5 протеази у системі PyMol для визначення структурних подібностей

Для підтвердження схожості структури SARS CoV Mpro 2 із структурами протеаз споріднених вірусів було проведено вирівнювання послідовностей у програмі PyMol за допомогою функції align із подальшим підрахунком RMSD та отриманням MatchAlign балів.

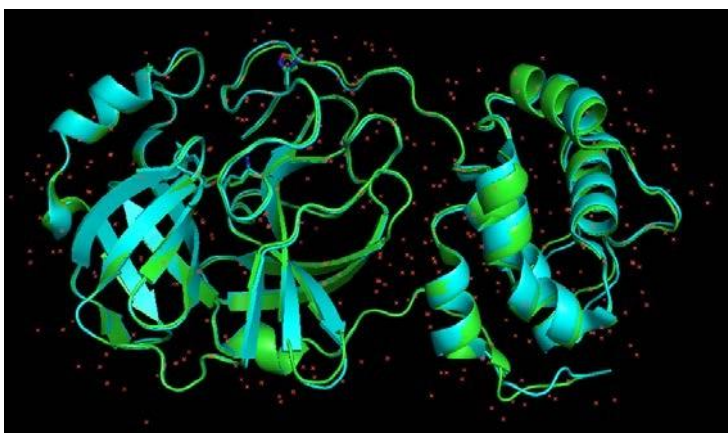


Рис. 3.1 Вирівнювання мономерів SARS CoV Mpro та SARS CoV 2 Mpro (виконано у PyMol). SARS CoV 2 Mpro (6lu7) відмічена зеленим кольором, SARS CoV Mpro () відмічена синім

MatchAlign бали: 1627.000

RMSD = 1.125 Å

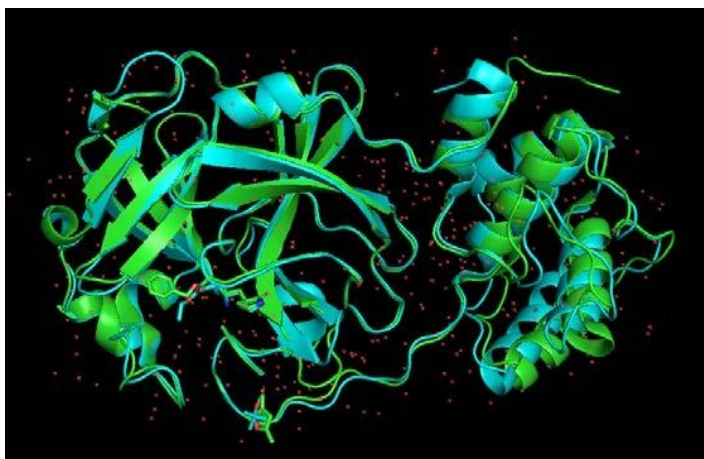


Рис. 3.2 Вирівнювання мономерів MERS CoV nsp5 протеази та SARS CoV 2 Mpro (виконано у PyMol). SARS CoV 2 Mpro (6LU7) відмічена зеленим кольором, MERS CoV nsp5 протеаза (4YLU) відмічена синім

MatchAlign бали: 829.000

RMSD = 2.132 Å

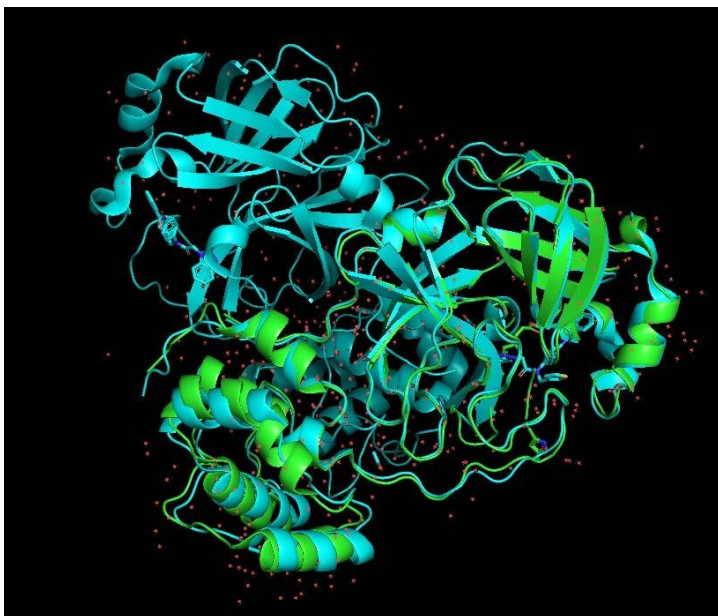


Рис. 3.3 Вирівнювання мономерів Bat CoV 3CLpro та SARS CoV 2 Mpro (виконано у PyMol). SARS CoV 2 Mpro (6LU7) відмічена зеленим кольором, Bat CoV 3CLpro (4YOG) відмічена синім

MatchAlign бали: 791.000

RMSD = 3.514 Å

Оптимальне вирівнювання - це вирівнювання, що дає найвищий бал. Це, оцінка вирівнювання X та Y = оцінка X та Y за оптимального вирівнювання.

RMSD менше приблизно 2 Å, як правило, вважається дуже близьким. Одинарна мутація може мати дуже глибокий вплив на структуру або, по суті, не матиме різниці залежно від того, де вона знаходиться в структурі. [56] Для білків роздільна здатність рентгенівських променів зазвичай знаходиться в діапазоні 2-3,5 Ангстрема, тому середньоквадратичне значення для шаблону в цьому діапазоні (навіть лише для основи) є великим. Згідно з показниками MatchAlign та RMSD досліджувані білки мають дуже високу схожість, отже є консервативними. Висока консервативність білка SARS CoV 2 Mpro робить його одною з найперспективніших цілей при пошуку антикоронавірусних ліків.

### 3.2 Ідентифікація потенційних інгібіторів димеризації SARS-CoV-2 Mpro шляхом молекулярного докінгу у програмному пакеті Schrodinger

У ході дослідження була проаналізована бібліотека лігандів що містить більш ніж 8 мільйонів сполук, отримана із бази даних комерційно доступних сполук для віртуального скринінгу ZINC15, що містить сполуки у готовому для докінгу 3D форматі.

Результати докінгу оцінювалися за функціями підрахунку. Функції підрахунку - це швидкі, приблизні математичні методи, що використовуються для оцінки сили зв'язування (як правило, шляхом вимірювання сили нековалентних взаємодій) між білком і лігандом після стикування. Особливий параметр що брався до уваги при аналізі результатів – XP G Score. [59]

Сполуки були зістиковані із димеризаційним сайтом Mpro протеази, що призвело до ідентифікації 8242 сполук із показниками XP G Score нижче за -5,0 ккал / моль.

Були проаналізовані такі підгрупи бази даних ZINC15 як anodyne (підгрупа не містить молекул із інтерференційними структурами PAINS) in-man (підгрупа містить речовини що споживаються людиною/мають відношення до фізіології людини, наприклад метаболіти, нутрицевтики або речовини що містяться у харчових продуктах), world (підгрупа містить лікарські засоби, затверджені у великих юрисдикціях), fda (підгрупа містить лише препарати, схвалені FDA), biogenic (підгрупа містить метаболіти та натуральні продукти).

#### 3.2.1 Дослідження енергій зв'язування речовин, схвалених FDA та іншими великими юрисдикціями із димеризаційним сайтом Mpro.

У першу чергу був проведений скринінг речовин що були схвалені для використання у якості лікарських препаратів на зв'язування із димеризаційним сайтом Mpro.

Із груп world, biogenic та fda було отримано відносно невелику кількість молекул-кандидатів із показниками XP G Score у діапазоні від -5 до -9.143. Речовина із найкращим показником вільної енергії взаємодії була обрана для більш детального аналізу.

## 1. Аналіз зв'язування паромоміцину сульфату з димеризаційним сайтом Mpro протеази.

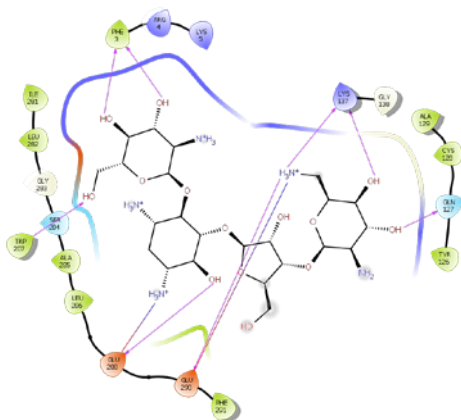


Рис. 3.4 Схематична репрезентація комплексу паромоміцину сульфату із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки позначені як червоно-сині лінії.

Паромоміцину сульфат є аміноглікозидним антибіотиком широкого спектру дії з амебіцидним та бактерицидним ефектом. Енергія докінгу паромоміцину сульфату була визначена як  $-9.143$  ккал / моль.

Амінокислоти GLU 290 та GLU 288 обидві утворюють один сольовий місток із молекулою ліганда. TRP207; GLN 127; GLU290; GLU288 формують по одному водневому зв'язку. PHE 3 та LYS 137 утворюють два водневі зв'язки.

### 3.2.2 Дослідження енергій зв'язування речовин що не містять інтерференційних структур (PAINS) із димеризаційним сайтом та активним центром Mpro

У підгрупі ZINC anodyne, з якої були отримані речовин що не містять інтерференційних структур (PAINS) було виявлено 8188 сполук із показниками XP G Score менше за  $-5.8365$  ккал / моль при зв'язуванні з димеризаційним сайтом SARS-CoV-2 Mpro. У подальшому був проведений аналіз зв'язування даних 8188 сполук із активним центром Mpro протеази. У 60 сполук із проаналізованих були виявлені прийнятні значення XP G Score при докінгу з активним центром Mpro протеази. 10 із даних сполук із показниками XP G Score для димеризаційного сайту не вище за  $-11.035$  ккал / моль було обрано для подальшого аналізу.

1. Аналіз зв'язування сполуки із ID ZINC000003922065 (C<sub>33</sub>H<sub>49</sub>N<sub>9</sub>O<sub>8</sub>) з активним сайтом та димеризаційним центром Mpro протеази.

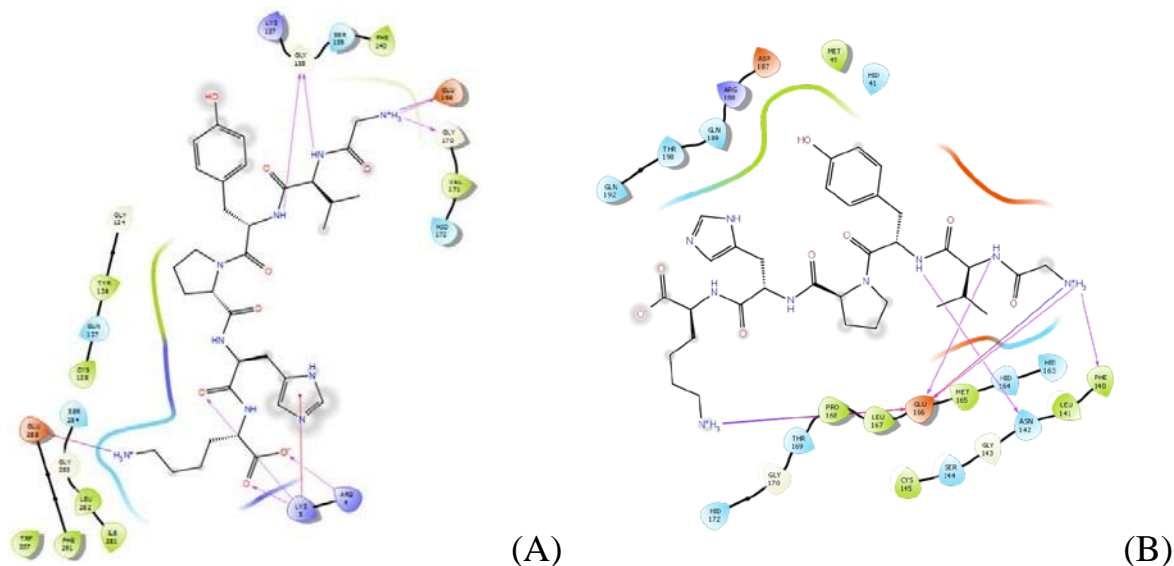


Рис. 3.5 Взаємодія комплексу ZINC000003922065-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Рі-рі катіонні взаємодії – червоними. Сольові містки вказані як червоно-сині стрілки (А) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (В) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Сполука ZINC000003922065 має показник XP G Score -11.891 ккал / моль при зв'язуванні з активним центром та -12.444 ккал / моль при зв'язуванні з димеризаційним сайтом. У формуванні водневих зв'язків з лігандом задіяні амінокислоти GLY138; GLY170; ARG4; LYS5; GLU166 при зв'язуванні з димеризаційним сайтом, та GLU 166; ASN 42; PHE 140 при зв'язуванні з активним центром. У комплексі з активним центром між лігандом та амінокислотою GLU 166 також формується сольовий місток. У комплексі із димеризаційним сайтом LYS 5 формує рі-рі катіонний зв'язок, GLU166; GLU 288 формують сольові містки

2. Аналіз зв'язування сполуки із ID ZINC000008551612 (C<sub>20</sub>H<sub>31</sub>N<sub>4</sub>O<sub>17</sub>P) з активним сайтом та димеризаційним центром Mpro протеази

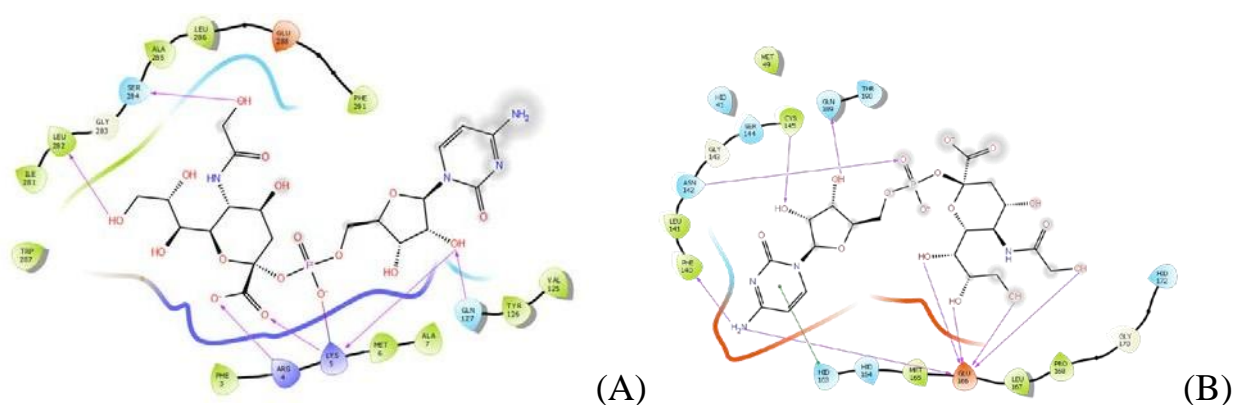


Рис. 3.6 Взаємодія комплексу ZINC000008551612-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Пі-пі стекінгові взаємодії – зеленими. Сольові містки вказані як червоно-сині стрілки. (А) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Мpro SARS-CoV-2 (В) Схематична репрезентація комплексу ліганда із активним центром Мpro SARS-CoV-2

Показник XP G Score ZINC000008551612 при зв'язуванні з активним центром був визначений як -8.260 ккал / моль, при зв'язуванні з димеризаційним сайтом – -11.759 ккал / моль. Амінокислоти SER284; LEU282; LYS5; ARG4; GLN127 димеризаційного сайту утворюють водневі зв'язки із лігандом. LYS 5 формує один сольовий місток.

У зв'язку ліганда із активним центром формується 9 водневих зв'язків із амінокислотами GLU 166; PHE 140; ASN 142; CYS 145; GLN 189 та одна Пі-пі стекінгова взаємодія із HIS 163

3. Аналіз зв'язування сполуки із ID ZINC000009574481 (C<sub>33</sub>H<sub>54</sub>N<sub>10</sub>O<sub>13</sub>) з активним сайтом та димеризаційним центром Мpro протеази

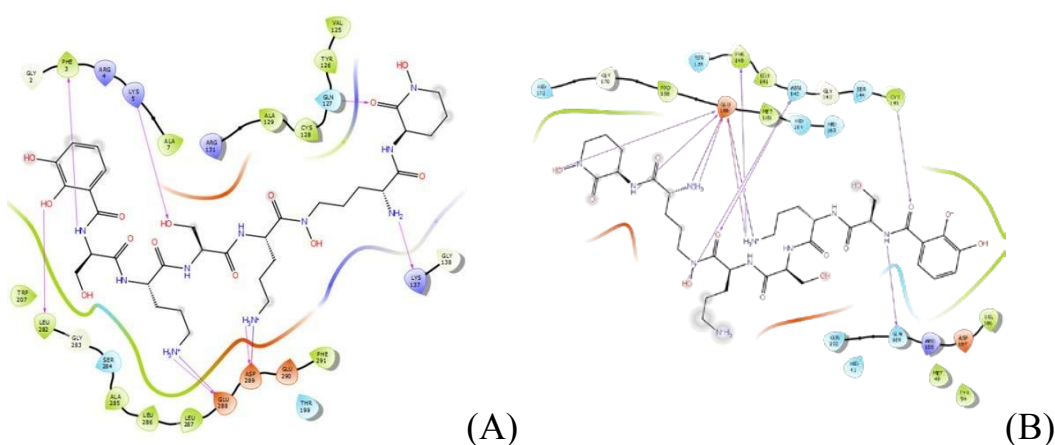


Рис. 3.7 Взаємодія комплексу ZINC000009574481-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки вказані

як червоно-сині стрілки. (A) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (B) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Сполука ZINC000009574481 має показник XP G Score при зв'язуванні з активним центром: -11.954 ккал / моль, при зв'язуванні з димеризаційним сайтом: -11.729 ккал / моль. Амінокислоти димеризаційного сайту PHE3; LEU282; LYS5; GLU288; ASP289; GLN127; LYS137 утворюють сім водневих зв'язків з досліджуваним лігандом. ASP289; GLN127 утворюють два сольових містки. У зв'язку ліганда із активним центром формується 9 водневих зв'язків із амінокислотами GLU 166; PHE 140; ASN 142; CYS 145; GLN 189 та 2 сольових містки із GLU 166.

4. Аналіз зв'язування сполуки із ID ZINC000004533632 (C<sub>33</sub>H<sub>47</sub>N<sub>9</sub>O<sub>7</sub>) з активним сайтом та димеризаційним центром Mpro протеази

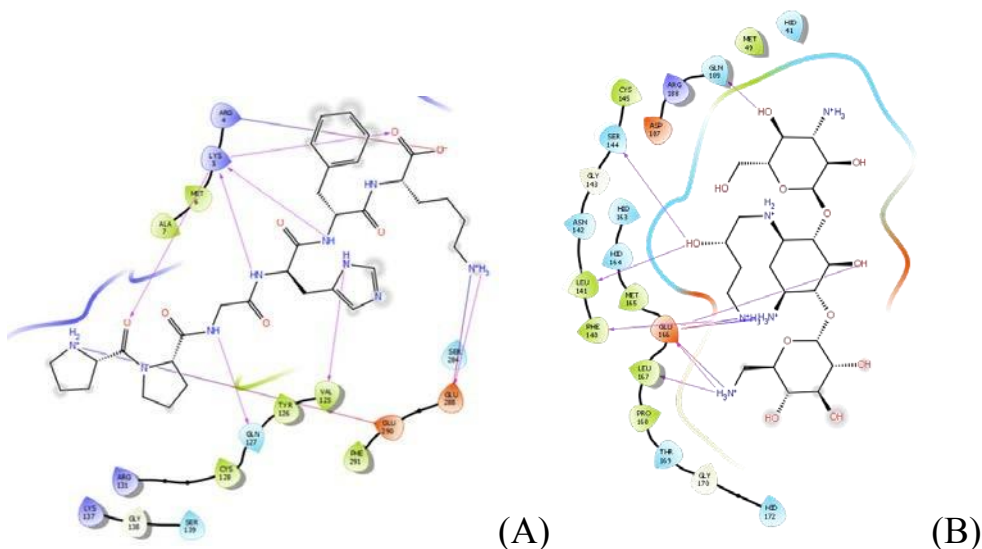


Рис. 3.8 Взаємодія комплексу ZINC000004533632-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки вказані як червоно-сині стрілки. (A) Схематична репрезентація комплексу ZINC000004533632 із димеризаційним сайтом Mpro SARS-CoV-2 (B) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Показник XP G Score ZINC000004533632 при зв'язуванні з активним центром дорівнює -15.141 ккал / моль, -11.653 ккал / моль при зв'язуванні з димеризаційним сайтом. Зв'язування з димеризаційним сайтом опосередковане амінокислотами LYS5; GLU288; GLN127; VAL125 (водневі зв'язки) та ARG4; GLU288; GLU290 (сольові містки). Зв'язок із активним центром також підтримується за рахунок 7 водневих зв'язків та 2 сольових

містків утворених із амінокислотами GLN 189; SER 144; LEU 141; PHE 140; GLU 166; LEU 167 та GLU 166 відповідно

5. Аналіз зв'язування сполуки із ID ZINC000003918259 (C<sub>27</sub>H<sub>44</sub>N<sub>6</sub>O<sub>7</sub>) з активним сайтом та димеризаційним центром Mpro протеази

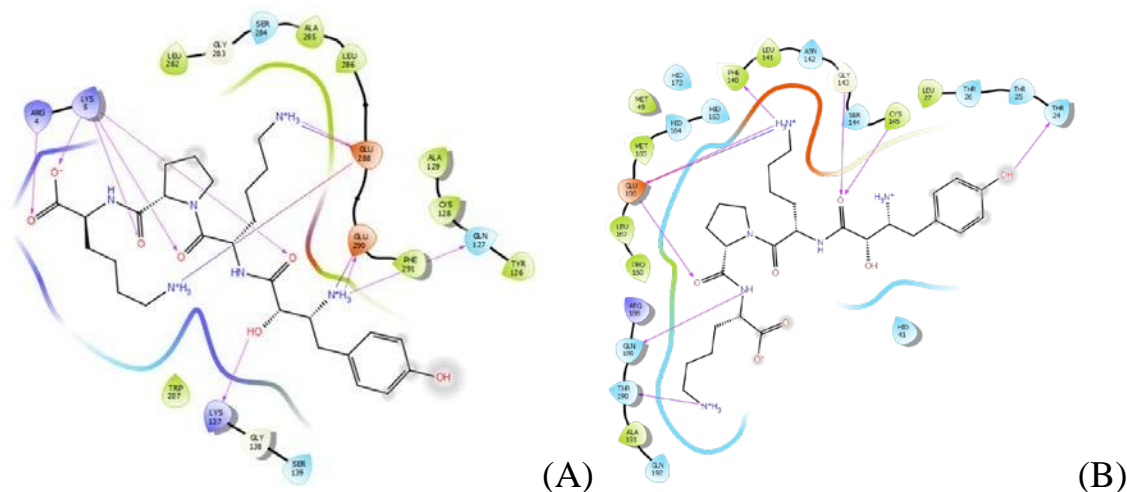


Рис. 3.9 Взаємодія комплексу ZINC000003918259-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки вказані як червоно-сині стрілки. (A) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (B) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Енергія зв'язування ZINC000003918259 була визначена як -11.482 ккал / моль при зв'язуванні з активним центром та -11.638 ккал / моль при зв'язуванні димеризаційним сайтом. Зв'язок з димеризаційним сайтом забезпечують 9 водневих зв'язків між лігандом та LYS5; ARG4; GLU288; GLU290; LYS137; GLN127 та 3 сольових містки між лігандом та GLU 290 GLU 288

8 водневих зв'язків CYS 145; GLY 143; PHE 140; GLU 166; GLN 189; THR 190; THR 24 та 1 сольовий місток із GLU 166 підтримують комплекс ліганд-активний центр

6. Аналіз зв'язування сполуки із ID ZINC000005412491 (C<sub>29</sub>H<sub>42</sub>N<sub>6</sub>O<sub>10</sub>) з активним сайтом та димеризаційним центром Mpro протеази

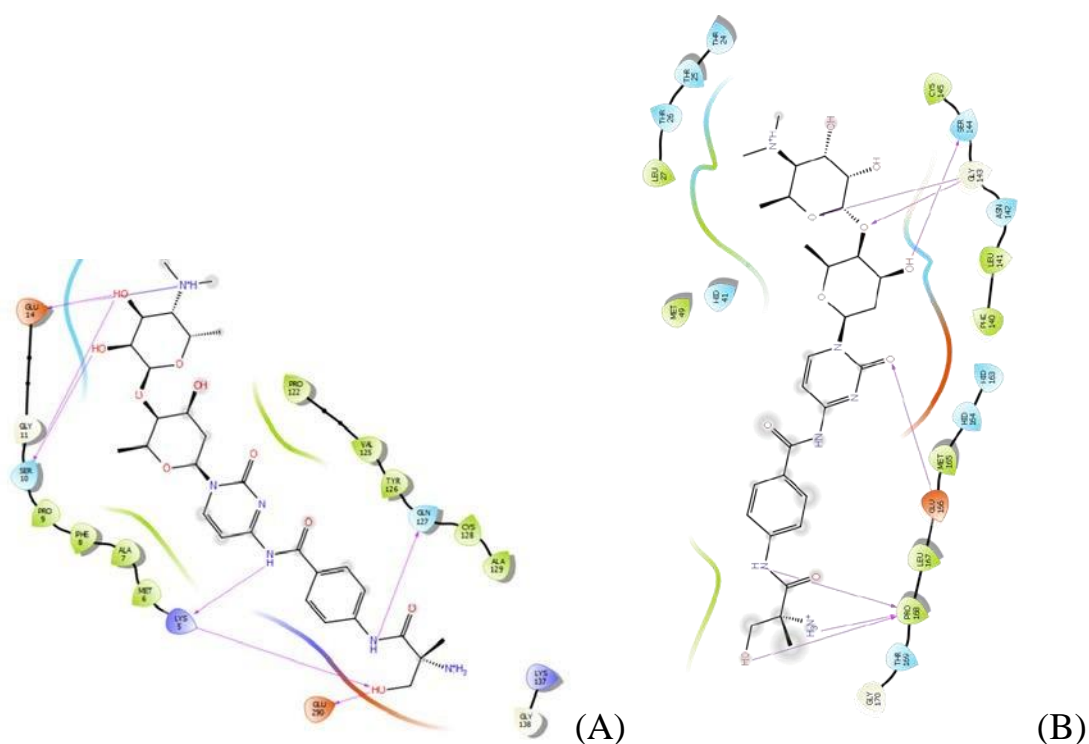


Рис. 3.10 Взаємодія комплексу ZINC000005412491-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки вказані як червоно-сині стрілки. (А) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (В) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Сполука ZINC000005412491 характеризується енергією зв'язування з активним центром: -7.196 ккал / моль за рахунок 7 водневих зв'язків із SER 144; GLY 143; GLU 166 та PRO 168. Та енергією зв'язування з димеризаційним сайтом: -11.561 ккал / моль. Як і попередні сполуки, C29H42N6O10 взаємодіє із залишками SER10; GLU14; LYS5; GLU290; GLN127 димеризаційного центру шляхом формування водневих зв'язків. C29H42N6O10 також утворює два сольових містки із SER10; GLU14.

7. Аналіз зв'язування сполуки із ID ZINC000003920109 (C25H42N8O7) з активним сайтом та димеризаційним центром Mpro протеази

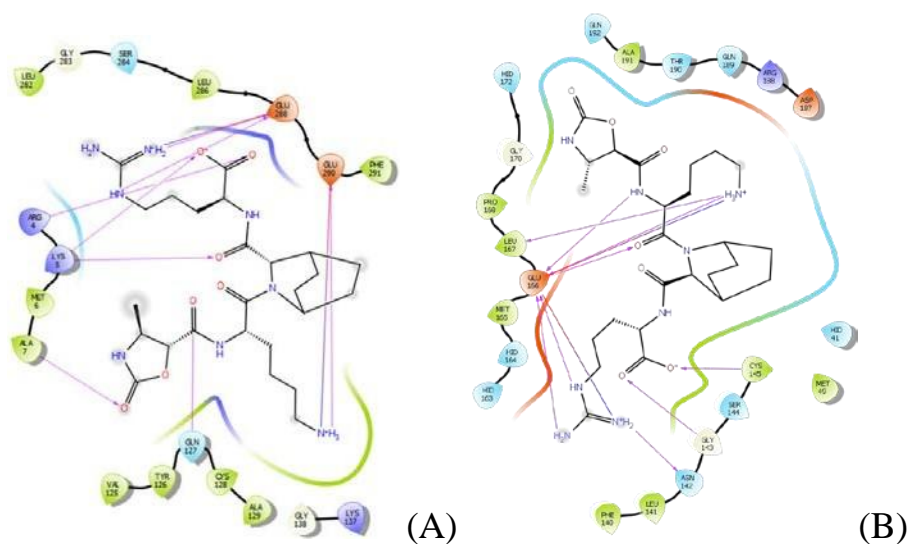


Рис. 3.11 Взаємодія комплексу ZINC000003920109-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки вказані як червоно-сині стрілки. (A) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (B) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Сполука ZINC000003920109 показала енергію зв'язування з активним центром: -11.574 ккал / моль та енергію зв'язування з димеризаційним сайтом: -11.283 ккал / моль. Комплекс протеїн-ліганд стабілізується 8 водневими зв'язками із залишками GLU288; GLU290; ALA7; GLN127; LYS5; ARG4 та 2 сольовими мітками із GLU288; GLU290 при зв'язуванні з димеризаційним сайтом, та 9 водневими зв'язками з LEU 167; GLU 166; ASN 142; GLY 143; CYS 145 і 2 сольовими мітками із GLU 166 при зв'язуванні з активним центром.

8. Аналіз зв'язування сполуки із ID ZINC000008551975 (C<sub>17</sub>H<sub>27</sub>N<sub>5</sub>O<sub>16</sub>P<sub>2</sub>) з активним сайтом та димеризаційним центром Mpro протеази

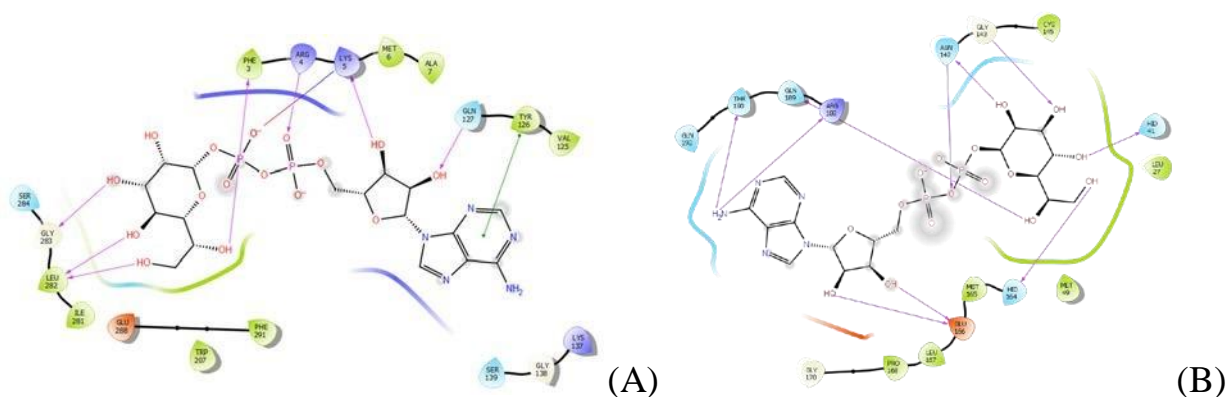


Рис. 3.12 Взаємодія комплексу ZINC000008551975-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Пі рі стекінгові

взаємодії – зеленими. Сольові містки вказані як червоно-сині стрілки. (A) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (B) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Сполука ZINC000008551975 показала енергію зв'язування з активним центром: -9.525 ккал / моль та енергію зв'язування з димеризаційним сайтом: -11.128 ккал / моль. Речовина вступає у водневі взаємодії з PHE3; ARG4; LYS5; GLN127; LEU282; GLY283 та утворює сольовий місток із LYS5 амінокислотою димеризаційного сайту, також утворюється одна  $\pi$   $\pi$  стекінгова взаємодія між амінокислотою TYR126 та лігандом.

У комплексі із активним центром Mpro формується 10 водневих зв'язків між амінокислотами GLU166; HIS164; HIS 41; GLY 143; ASN 142; ARG 188; GLN 189; THR 190 та лігандом

9. Аналіз зв'язування сполуки із ID ZINC000005273784 (C<sub>17</sub>H<sub>27</sub>N<sub>5</sub>O<sub>16</sub>P<sub>2</sub>) з активним сайтом та димеризаційним центром Mpro протеази

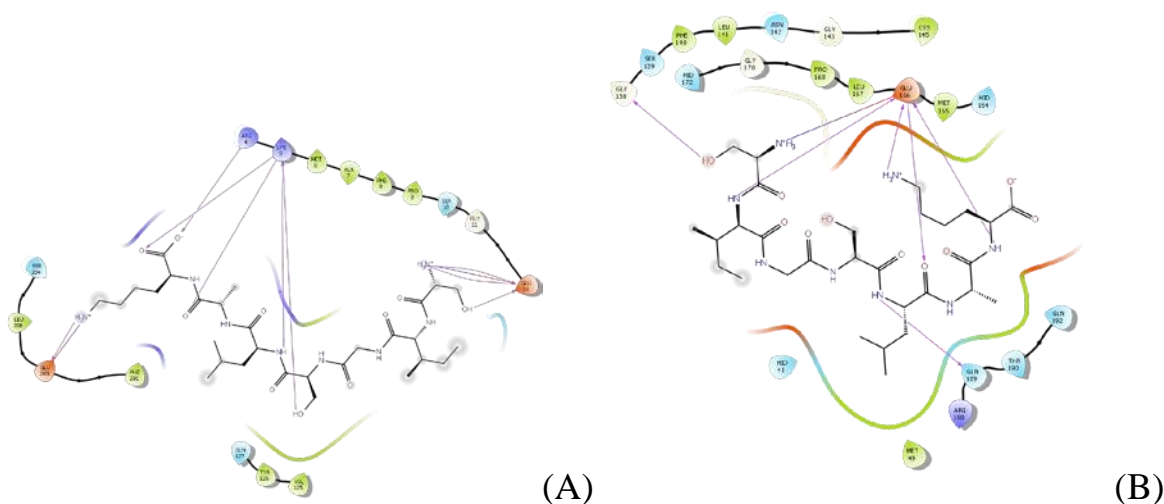


Рис. 3.13 Взаємодія комплексу ZINC000005273784-Mpro SARS-CoV-2. Водневі зв'язки представлені фіолетовими стрілками. Сольові містки вказані як червоно-сині стрілки. (A) Схематична репрезентація комплексу ліганда із димеризаційним сайтом Mpro SARS-CoV-2 (B) Схематична репрезентація комплексу ліганда із активним центром Mpro SARS-CoV-2

Сполука ZINC000005273784 показала енергію зв'язування з активним центром: -12.594 ккал / моль та енергію зв'язування з димеризаційним сайтом: -10.361 ккал / моль.

У активному центрі досліджені ліганди утворюють зв'язки з амінокислотами каталітичної діади Cys145, яка опосередковує електростатичний тригер,

ініціюючи хімічну реакцію, та His41. Комплекси також утворюються з His164 та Asp187, амінокислотами стабілізуючими позицію молекули H<sub>2</sub>Ocat, третього елемента каталітичної діади, через слабкі водневі зв'язки.

Зв'язки формуються з Met165, Glu166 та His172, амінокислотами що приймають участь в утворенні субсайта S1, зв'язуючого групу P1 субстрату, зазвичай Gln. Phe140 та His163, що створюють задню стінку підсайту S1 також приймають участь у зв'язуванні ліганда. Фенільне кільце Phe140 взаємодіє з імідазольним кільцем His163 в нижній частині кишені S1, забезпечуючи незаряджений стан гістидинового залишку у широкому діапазоні значень рН.

При зв'язуванні з димеризаційним сайтом утворюються комплекси між лігандом та амінокислотами Arg4 та Glu290, необхідними для забезпечення димеризації. Мутація Glu290 до Ala призводить до повної втрати димеризації та активності, тоді як мутація Arg4 призводить до приблизно п'ятикратної втрати димеризації.

В утворенні комплексів з лігандами також приймає участь Met6, мутація якого до Ala повністю перешкоджає димеризації та ферментативній активності. При утворенні димеру, Met6 вірогідно утворює гідрофобні взаємодії з Tyr126 та Phe140, амінокислотами які також приймають участь у зв'язуванні досліджуваних лігандів.

З лігандом утворюють зв'язки Ala7, Phe8, Arg4 що створюють водневі зв'язки з Val125, Glu290, Lys137, Phe140 та Glu166 при димеризації. Val125, Glu290, Lys137, Phe140 та Glu166 також активно формують зв'язки з досліджуваними лігандами.

Досліджені ліганди активно зв'язують амінокислоти Glu166 і Phe140, карбоксилатна та карбонільна групи яких утворюють міцні Н-зв'язки з NH-групою Ser1 протилежного протомеру. Ця взаємодія стабілізує фермент, сприяє правильній орієнтації петлі оксианіону та S1 кишені ділянки зв'язування субстрату і, отже, призводить до підвищеної каталітичної ефективності.

### 3.2.3 Дослідження вільних енергій зв'язування речовин що споживаються людиною/містяться у харчових продуктах із димеризаційним сайтом Mrgo

У групі in man бази даних ZINC15, з якої було отримано бібліотеку сполук що споживаються людиною/містяться у харчових продуктах було

ідентифіковано 41 структуру, із показником XP G Score не більше -8.063 ккал / моль. 9 із них було відібрано для поглибленого аналізу.

1. Аналіз зв'язування 1,4,6-тригалоїл-бета-D-глюкопіранози із димеризаційним сайтом SARS CoV 2 Mpro протеази.

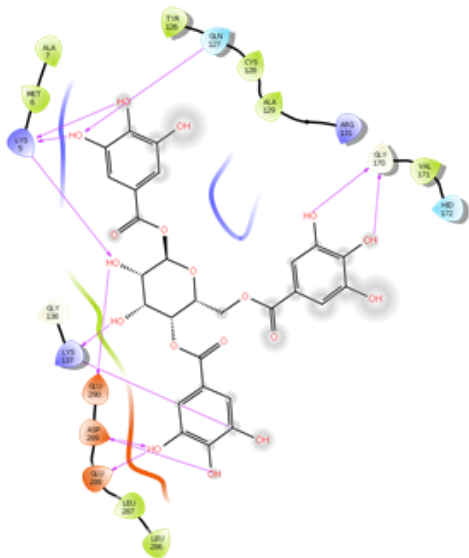


Рис. 3.14 Схематична репрезентація комплексу 1,4,6-тригалоїл-бета-D-глюкопіранози із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки представлені фіолетовими стрілками

Енергія докінгу 1,4,6-тригалоїл-бета-D-глюкопіранози була визначена як -10.998 ккал / моль. Речовина належить до танінів. У комплексі Mpro-ліганд амінокислоти GLN 127; LYS 137; GLU 288; GLU 290 утворюють по одному водневому зв'язку із молекулою досліджуваної речовини, LYS 5; GLY 170; ASP 289 формують по 2 водневі зв'язки. LYS 137 також вступає у один водневий зв'язок із молекулою 1,4,6-тригалоїл-бета-D-глюкопіранози

1,4,6-тригалоїл-бета-D-глюкопіраноза належить до класу органічних сполук, відомих як таніни. Це природні поліфеноли, які можна розподілити на чотири основних класи: таніни здатні до гідролізації (на основі елагової кислоти або галової кислоти), конденсовані таніни (складаються з олігомерних або полімерних проантоціанідинів), складні таніни (складаються з катехіну, зв'язаного з галлотаніном або елагітаніном) та флоротаніни (олігомери флороглюцинолу). 1,4,6-тригалоїл-бета-D-глюкопіраноза міститься в найвищих концентраціях в чаї, але також була виявлена в плодах гранатового дерева (*Punica granatum*).

2. Аналіз зв'язування кверцетину 3- (3-р-кумароїлглюкозиду) із димеризаційним сайтом SARS CoV 2 Mpro протеази.

Кверцетин 3- (3-р-кумароїлглюкозид) належить до класу органічних сполук, відомих як флавоноїд-3-о-глікозиди. Це фенольні сполуки, що містять флавоноїдну частину, О-глікозидно пов'язану з вуглеводною частиною в положенні С3.

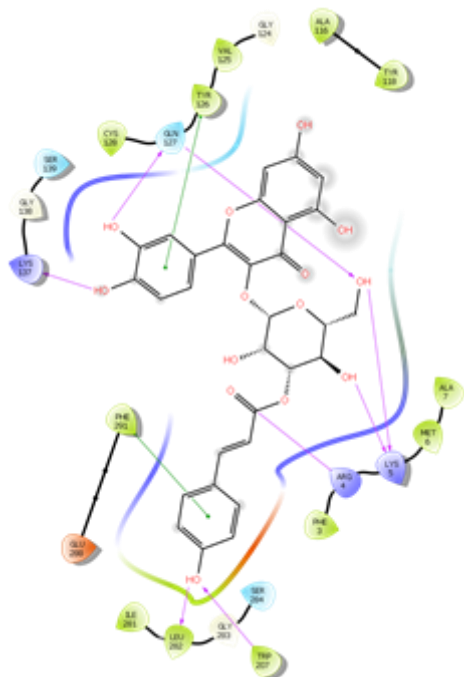


Рис. 3.15 Схематична репрезентація комплексу кверцетину 3- (3-р-кумароїлглюкозиду) із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки. Рі-рі стекінгові взаємодії — як зелені лінії.

Енергія докінгу кверцетину 3- (3-р-кумароїлглюкозиду) була визначена як -10.964 ккал / моль. Комплекс підтримується за рахунок водневих зв'язків утворених амінокислотами GLN 127; LYS 137; LYS 5; TRP 207; LEU 282; ARG 4 та рі-рі стекінгових взаємодій TYR 126; PHE 291.

3. Аналіз зв'язування 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину із димеризаційним сайтом SARS CoV 2 Mpro протеази.

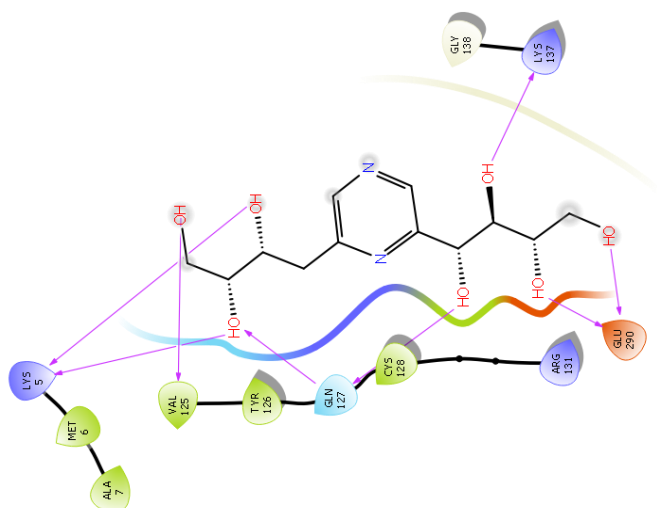


Рис. 3.16 Схематична репрезентація комплексу 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки.

Енергія докінгу речовини була визначена як -10.636 ккал / моль.

Ліганд формує водневі зв'язки із такими амінокислотами димеризаційного сайту: LYS 5 (2 зв'язки); VAL 125; GLN 127 (2 зв'язки); GLU 290 (2 зв'язки); LYS 137

2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразин належить до класу органічних сполук, відомих як піразини. Піразини - це сполуки, що містять піразинове кільце, яке є шестичленним ароматичним гетероциклом, що складається з двох атомів азоту (в положеннях 1 і 4) та чотирьох атомів вуглецю. Поза людським тілом було виявлено у горіхах.

4. Аналіз зв'язування фагопіритулу A2 із димеризаційним сайтом SARS CoV 2 Mpro протеази.

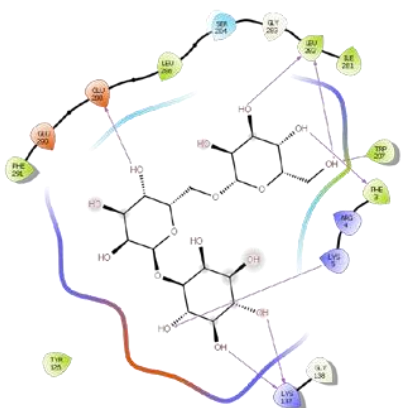


Рис. 3.17 Схематична репрезентація фагопіритулу A2 із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки.

Енергія докінгу фагопіритулу A2 була визначена як -10.198 ккал / моль.

Комплекс білок-ліганд підтримується завдяки одному водневому зв'язку що утворюється між лігандом та амінокислотами GLU 288; TRP 207; PHE 3, та двом водневим зв'язкам між лігандом та амінокислотами LEU 282 та LYS 137.

Фагопіритол A2 належить до класу органічних сполук, відомих як о-глікозилові сполуки. Поза людським тілом фагопіритол A2 був виявлений, але не кількісно визначений у зернових продуктах.

5. Аналіз зв'язування Кверцетин 3-О- (6''-малоніл-глюкозид) 7-О-глюкозиду із димеризаційним сайтом SARS CoV 2 Mpro протеази.

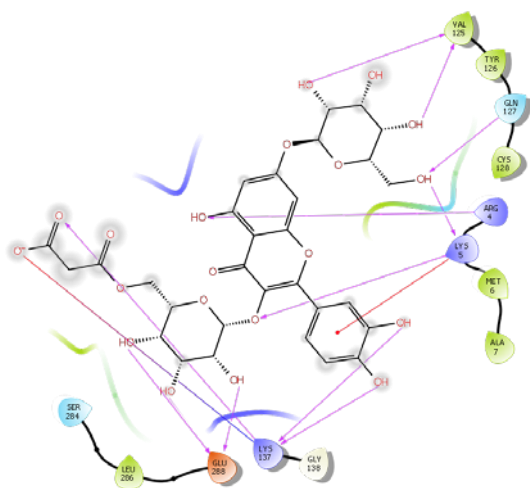


Рис 3.18 Схематична репрезентація комплексу кверцетин 3-О- (6''-малоніл-глюкозид) 7-О-глюкозиду із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки. Рі-рі катіонові взаємодії відображені як червоні лінії.

Енергія докінгу речовини була визначена як -10.101 ккал / моль.

Водневі зв'язки формуються амінокислотами VAL 125; GLN 127; ARG 4; LYS 5; LYS 137; GLU 288. LYS 5 утворює рі-рі катіонну взаємодію, LYS 137 утворює сольовий місток із COO групою ліганду.

Кверцетин 3-О- (6"-малоніл-глюкозид) 7-О-глюкозид належить до класу органічних сполук, відомих як флавоноїди-7-о-глікозиди. Це фенольні сполуки, що містять флавоноїдний фрагмент, О-глікозидно пов'язаний з вуглеводним фрагментом у положенні С7. Кверцетин 3-О- (6"-малоніл-глюкозид) 7-О-глюкозид виявляється в середньому в найбільшій концентрації в листі салату (*Lactuca sativa*). Кверцетин 3-О- (6"-малоніл-глюкозид) 7-О-глюкозид також був виявлений в ендівії (*Cichorium endivia*)

6. Аналіз зв'язування катехіну 3',5-диглюкозиду із димеризаційним сайтом SARS CoV 2 Mpro протеази.

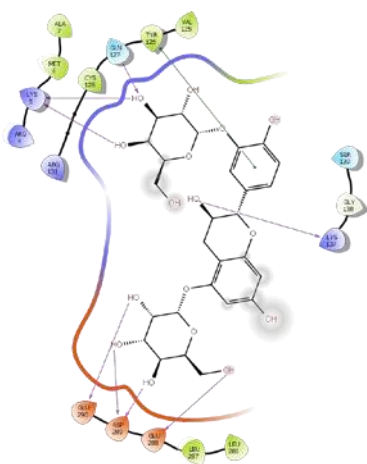


Рис 3.19 Схематична репрезентація комплексу катехіну 3',5-диглюкозиду із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки. Рі-рі стекінгові взаємодії — як зелені лінії.

Енергія докінгу речовини була визначена як -9.894 ккал / моль.

Було визначено 8 водневих зв'язків утворених із лігандом амінокислотами LYS 5; GLN 127; LYS 137; GLU 290; ASP 289; GLU 288, що підтримують комплекс. В підтримці комплексу також приймає участь амінокислота TYR 128, що утворює рі-рі стекінгову взаємодію із лігандом.

Катехін 3', 5-диглюкозид належить до класу органічних сполук, відомих як флавоноїдні о-глікозиди. Флавоноїдні О-глікозиди - це сполуки, що містять вуглеводну частину, яка О-глікозидно пов'язана з 2-фенілхромен-4-он-флавоноїдом. Катехін 3', 5-диглюкозид - надзвичайно слабка основна (по суті нейтральна) сполука (на основі його рKa). Речовину виявлено у зелених овочах.

## 9. Аналіз зв'язування актеозиду із димеризаційним сайтом SARS CoV 2 Mpro протеази.

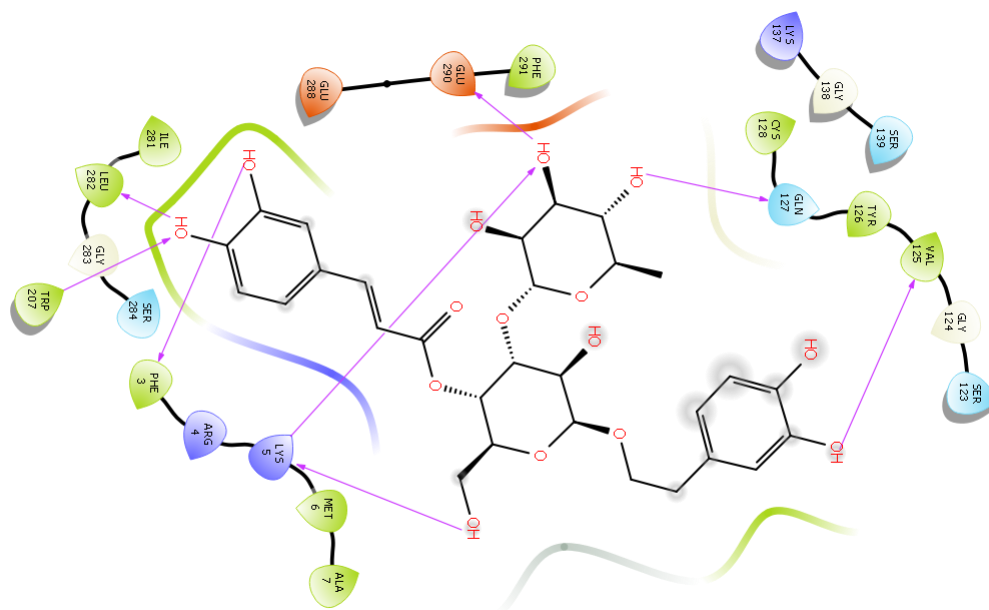


Рис 3.20 Схематична репрезентація комплексу актеозиду із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки. Пі-рі катіонові взаємодії відображені як червоні лінії.

Енергія докінгу актеозиду була визначена як  $-9.870$  ккал / моль.

Комплекс існує за рахунок водневих зв'язків, у яких більшість амінокислот (PHE 3; LEU 282; GLU 290; GLN 127; VAL 125) є акцепторами водневого зв'язку. TRP 207 виступає донором зв'язку. LYS 5 формує два водневі зв'язки один у якості донора, а інший – акцептора.

Актеозид, також відомий як кусагінін або вербаскозид, належить до класу органічних сполук, відомих як кумарові кислоти та їх похідні. Вербаскозид – це кофеїл-фенілетаноїдний глікозид, у якому фенілпропаноїдна кавова кислота та фенілетаноїд-гідрокситирозол утворюють складний ефір та ефірний зв'язок з рамнозною частиною дисахариду, а саме  $\beta$ - (3', 4'-дигідроксифеніл) етил-О- $\alpha$ -L-рамнопіранозил (1  $\rightarrow$  3) - $\beta$ -D- (4-О-кофеїл) -глюкопіранозид. У сімействі Lamiaceae речовину можна знайти в лікарських рослинах роду *Phlomis*, у Scrophulariaceae, у *Verbascum phlomoides*, *Verbascum mallophorum* або, родина Buddlejaceae, у *Buddleja globosa* або *Buddleja cordata*, у родини Bignoniaceae, у *Pithecoctenium* sp та *Tynanthus panurensis*, у родини Orobanchaceae, у *Cistanche* sp та *Orobanche rapumgenistae*, у Plantaginaceae, у *Plantago Vercea laceolatae* (звичайна вербена),

*Aloysia citrodora* (лимонна вербена) та *Lantana camara*, у *Oleaceae*, у *Olea europaea* (оливкова), у сочевиці, у хижої рослини *Pinguicula lusitanica* та у *Byblis liniflora*. В середньому актеозид виявляється у найвищій концентрації в кількох різних продуктах харчування, таких як вербена звичайна, вербена лимонна та оливки. Актеозид також був виявлений у гірчичках, фруктах та кунжуті. Незважаючи на те, що повідомлялося про деяку *in vitro* генотоксичність вербаскозиду на людських лімфоцитах із залученням білків PARP-1 та p53[45], подальші тести *in vivo* не повідомляли про генотоксичність для перорального прийому високих доз[46]. Речовина також виробляється в культурах волохатих коренів *Paulownia tomentosa* (імператричне дерево, *Paulowniaceae*) та в культурах рослинних клітин *Leucoscepterum* sp (*Lamiaceae*) та *Syringa* sp (*Oleaceae*). Похідні вербаскозиду можна знайти у *Verbascum undulatum* і особливо апіозиди у *Verbascum* sp. Вербаскозид можна зустріти у видів у всіх сімействах порядку *Lamiales* (син. *Scrophulariales*).

8. Аналіз зв'язування альфа-D-Ксилопіранозил- (1-> 6) -бета-D-глюкопіранозил- (1-> 4) -D-глюкози із димеризаційним сайтом SARS CoV 2 Mpro протеази.

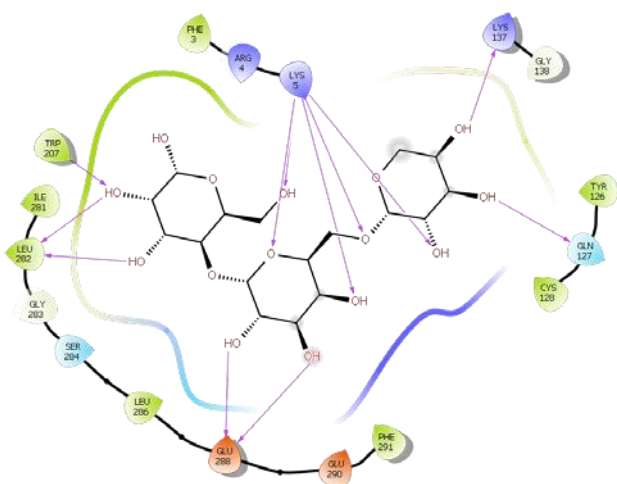


Рис 3.21 Схематична репрезентація комплексу альфа-D-Ксилопіранозил- (1-> 6) -бета-D-глюкопіранозил- (1-> 4) -D-глюкози із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки.

Комплекс утримується за рахунок водневих зв'язків амінокислот димеризаційного сайту із спиртовими та кето-групами досліджуваного ліганда.

В утворенні водневих зв'язків приймають участь амінокислоти LYS 137 (1 зв'язок); GLN 127 (1 зв'язок); LYS 5 (5 зв'язків); TRP 207 (1 зв'язок); LEU 282 (2 зв'язки); GLU 288 (2 зв'язки)

альфа-D-Ксилопіранозил- (1-> 6) -бета-D-глюкопіранозил- (1-> 4) -D-глюкоза належить до класу органічних сполук, відомих як олігосахариди. Це вуглеводи, що складаються з 3 - 10 моносахаридних одиниць, пов'язаних між собою за допомогою глікозидних зв'язків. альфа-D-ксилопіранозил- (1-> 6) - бета-D-глюкопіранозил- (1-> 4) -D-глюкоза була виявлена у зернових продуктах та фруктах.

### 9. Аналіз зв'язування кверцетин 3,4'-диглюкозиду із димеризаційним сайтом SARS CoV 2 Mpro протеази

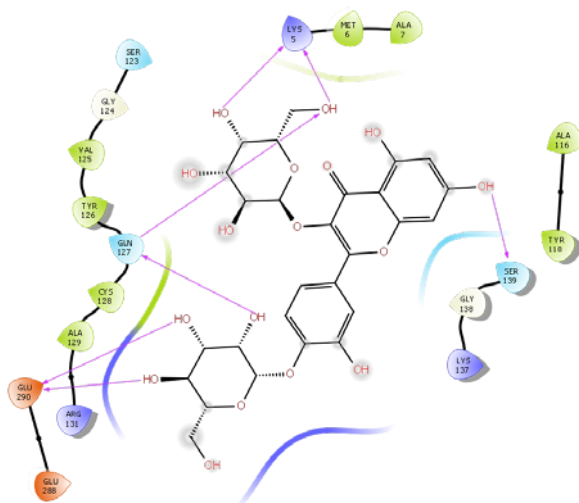


Рис 3.22 Схематична репрезентація комплексу кверцетин 3,4'-диглюкозиду із амінокислотними залишками димеризаційного сайту Mpro SARS CoV 2. Водневі зв'язки відображені як фіолетові стрілки.

У комплексі було визначено 7 водневих зв'язків, утворених амінокислотами LYS 5; GLN 127; GLU 290; SER 139 із спиртоими групами кверцетин 3,4'-диглюкозиду.

Кверцетин 3,4'-диглюкозид належить до класу органічних сполук, відомих як флавоноїд-3-о-глікозиди. Кверцетин 3,4'-диглюкозид міститься в середньому в найвищих концентраціях в цибулі-шалот (*Allium сера*) та садовій цибулі (*Allium schoenoprasum*)

Досліджені ліганди із групи ZINC15 in-man при зв'язуванні з димеризаційним сайтом SARS-CoV 2 Mpro зазвичай утворюють комплекси з амінокислотами ALA 7; LYS 5; Arg 4, які формують структури N-пальця

ферменту. N-кінцеві залишки (N-палець) протомера А при димеризації взаємодіють із залишками домену II протомеру В. N-кінці взаємодіють із амінокислотними залишками поблизу S1 субстрат-зв'язуючого підсайту. Ця взаємодія стабілізує фермент, сприяє правильній орієнтації оксианіонної петлі та S1 кишені, таким чином призводячи до підвищеної каталітичної ефективності ферменту. Залишки, прилеглі до N-кінця, також відіграють ключову роль у димеризації, зокрема Pro9 та Phe305 протомера А, які взаємодіють із залишками 122-127 протомера В.

Амінокислотні залишки Glu290; Tyr126; Met6; Arg4, зв'язувані дослідженими речовинами значно стабілізують димерну структуру шляхом міжпротомерних взаємодій, таких як сольовий місток між бічними ланцюгами Arg4 та Glu290 протилежного протомера та гідрофобно-ароматична взаємодія між бічними ланцюгами Met6 та Tyr126 протилежного протомера.

Під час докінгових досліджень було виявлено 17 сполук із високими показниками вільних енергій зв'язування, що здатні утворювати водневі, сольові, рі-рі катіонні та рі-рі стекінгові взаємодії із амінокислотними залишками, важливими для димеризації Mpro SARS CoV 2, такими як Arg4, LYS 5, GLN 127. Деякі з досліджених лігандів здатні також високо специфічно зв'язуватися з активним центром фермента. Дані речовини є перспективними потенційними лікарськими препаратами, націленими на інгібування висококонсервативного білка коронавірусу Mpro.

### **3.3 Дослідження молекулярної динаміки молекул-кандидатів та їх комплексів із димеризаційним сайтом Mpro.**

У програмному пакеті GROMACS було створено симуляції молекулярної динаміки комплексів білку Mpro із 6 лігандами із найкращими показниками XP G Score. Аналіз траєкторій, отриманих в результаті моделювання, проводили за допомогою інструментів програмного пакету Gromacs. Аналіз середньоквадратичного відхилення (RMSD), середньоквадратичної флуктуації (RMSF) атомів Ca білка щодо мінімізованої кристалічної структури, середньоквадратичної флуктуації важких атомів лігандів, вільної енергії взаємодії, енергії Леннарда-Джонса, Кулонівської енергії також проводили за допомогою Gromacs. Візуалізація результатів аналізу виконувалась за допомогою програмного забезпечення Origin.

#### **3.3.1 Дослідження середньоквадратичних відхилень (RMSD) комплексів Mpro-ліганд**

Визначення RMSD необхідно для аналізу стабільності комплексу. RMSD виконує порівняння положення досліджуваної структури протягом усього періоду симуляції із еталонною структурою.

Для вивчення динамічної стабільності систем аналізувалися RMSD ліганду, білку та комплексу ліганд-білок. Суперпозиція координат кожної складної структури в траєкторії на початкову структуру дозволяє проаналізувати прогресію середньоквадратичних відхилень.

### 1. Аналіз середньоквадратичних відхилень Mpro, ліганду пармоміцину сульфату та їх комплексу.

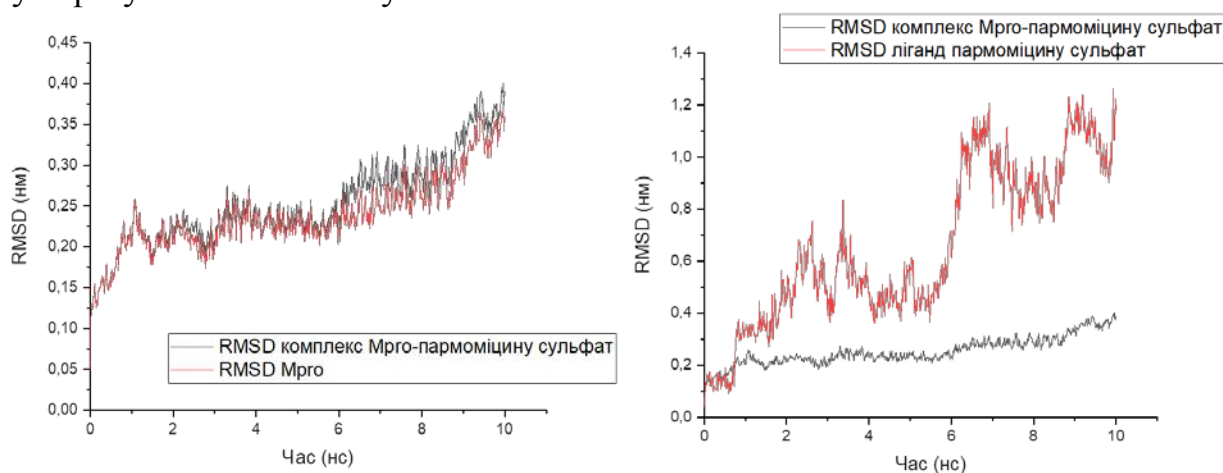


Рис 3.23 Молекулярна динаміка комплексу Mpro - пармоміцину сульфат (10 ns)

(а) RMSD протеази Mpro у комплексі з пармоміцину сульфатом, RMSD протеази Mpro без ліганду (б) RMSD ліганду пармоміцину сульфату і RMSD ліганду у комплексі з Mpro

XP GScore = -9.143

Значення середнього квадратного відхилення (RMSD) більше ніж 4 Å вказує на невисоку стабільність комплексу білок-ліганд.

### 2. Аналіз середньоквадратичних відхилень Mpro, ліганду кверцетин 3- (3-р-кумароїлглюкозиду) та їх комплексу.

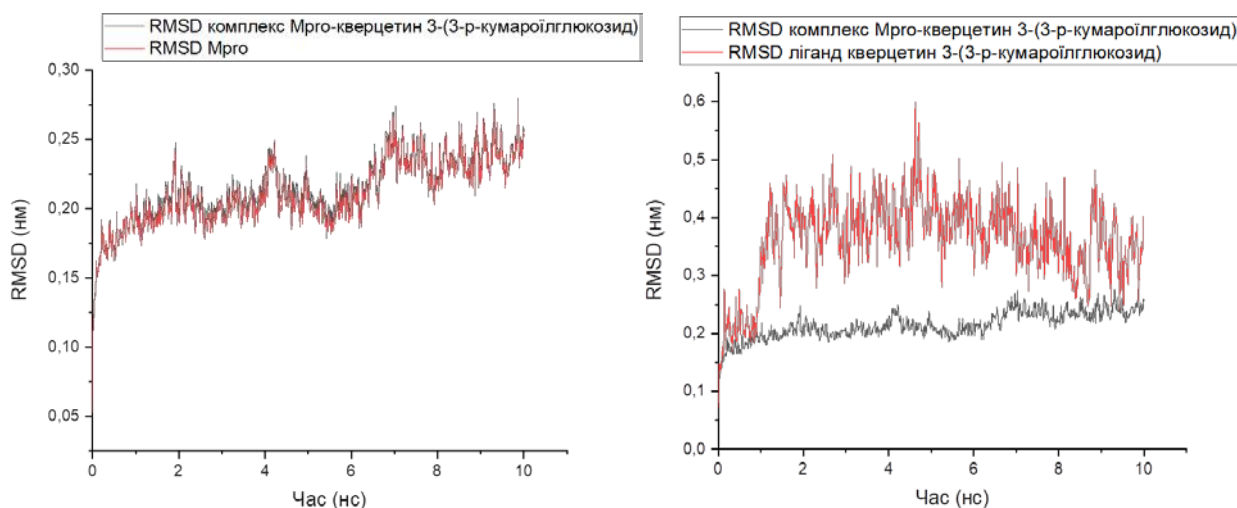


Рис 3.24 Молекулярна динаміка комплексу Mpro - кверцетин 3- (3-р-кумароїлглюкозид) (10 ns)

(а) RMSD протеази Mpro у комплексі з кверцетин 3- (3-р-кумароїлглюкозидом), RMSD протеази Mpro без ліганду (б) RMSD ліганду кверцетин 3- (3-р-кумароїлглюкозиду) і RMSD ліганду у комплексі з Mpro

Речовина належить до класу органічних сполук, відомих як флавоноїд-3-о-глікозиди. Це фенольні сполуки, що містять флавоноїдну частину, О-глікозидно пов'язану з вуглеводною частиною в положенні С3.

XP GScore = -10.964

Значення середнього квадратного відхилення (RMSD) менше, ніж 3 Å вказує на стабільність комплексу білок-ліганд.

3. Аналіз середньоквадратичних відхилень Mpro, ліганду 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину та їх комплексу.

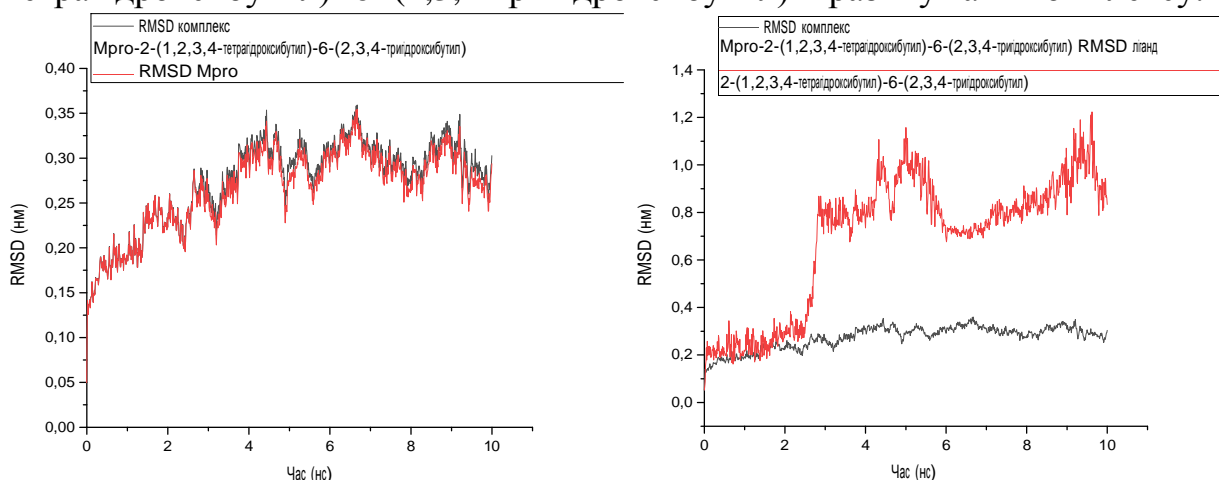


Рис 3.25 Молекулярна динаміка комплексу Mpro - 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразин (10 ns)

(а) RMSD протеази Mpro у комплексі з 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразином, RMSD протеази Mpro без ліганду (б) RMSD ліганду 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину і RMSD ліганду у комплексі з Mpro

Речовина належить до класу органічних сполук, відомих як піразини. Піразини - це сполуки, що містять піразинове кільце, яке є шестичленним ароматичним гетероциклом, що складається з двох атомів азоту (в положеннях 1 і 4) та чотирьох атомів вуглецю. Речовину було виявлено у горіхах.

XP GScore = -10.636

Значення середнього квадратного відхилення (RMSD) вище за 3.5 Å вказує на недостатню стабільність комплексу білок-ліганд.

4. Аналіз середньоквадратичних відхилень Mpro, ліганду кроцину та їх комплексу.

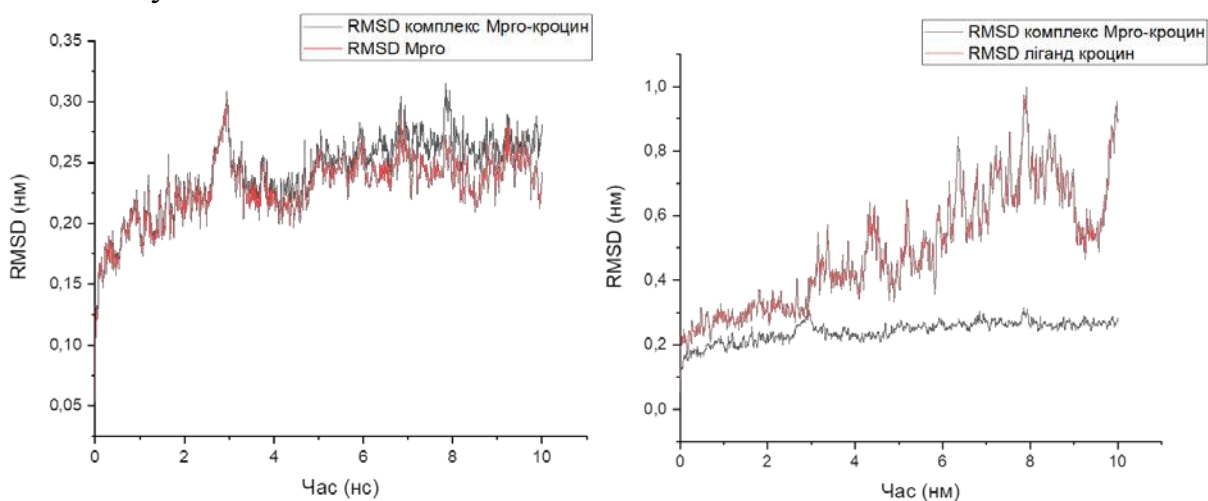


Рис 3.26 Молекулярна динаміка комплексу Mpro - Кроцин (10 ns)

(а) RMSD протеази Mpro у комплексі з кроцином, RMSD протеази Mpro без ліганду (б) RMSD ліганду кроцину і RMSD ліганду у комплексі з Mpro

Належить до класу органічних сполук, відомих як дитерпеноїди. Це терпенові сполуки, утворені чотирма одиницями ізопрену

XP GScore = -10.706

Значення середнього квадратного відхилення (RMSD) менше, ніж 3,5 Å вказує на достатню стабільність комплексу білок-ліганд.

5. Аналіз середньоквадратичних відхилень Mpro, ліганду фагопіритулу A2 та їх комплексу.

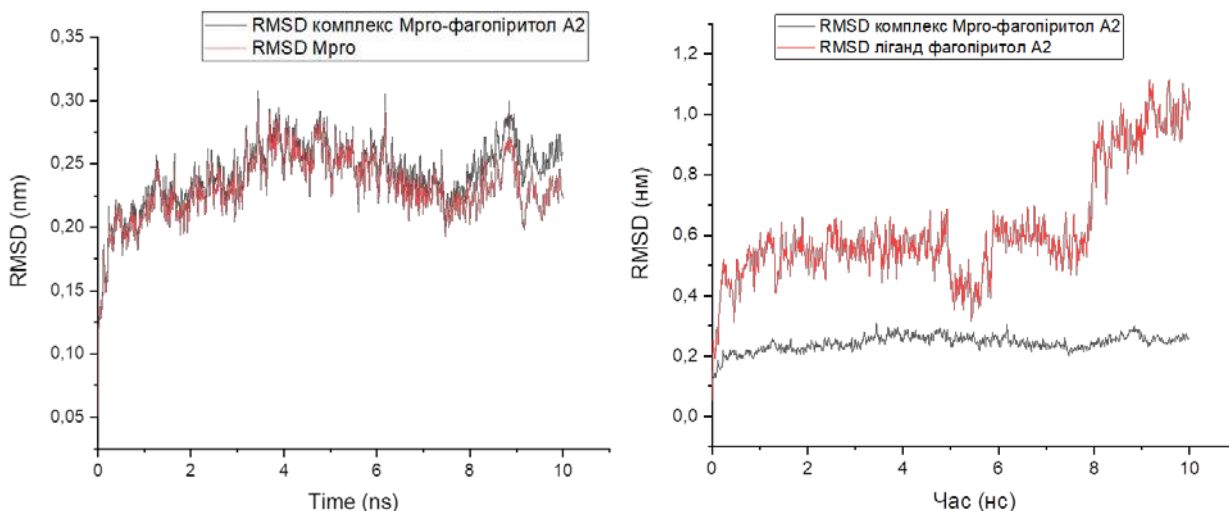


Рис 3.27 Молекулярна динаміка комплексу Mpro - Фагопіритол A2 (10 ns)

(а) RMSD протеази Mpro у комплексі з фагопіритолом A2, RMSD протеази Mpro без ліганду (б) RMSD ліганду фагопіритолу A2 і RMSD ліганду у комплексі з Mpro

Речовина належить до класу органічних сполук, відомих як о-глікозилові сполуки. Фагопіритол A2 був виявлений у зернових продуктах

XP GScore = -10.198

Значення середнього квадратного відхилення (RMSD) менше, ніж 3,5 Å вказує на достатню стабільність комплексу білок-ліганд.

б. Аналіз середньоквадратичних відхилень Mpro, ліганду ZINC5273784 та їх комплексу.

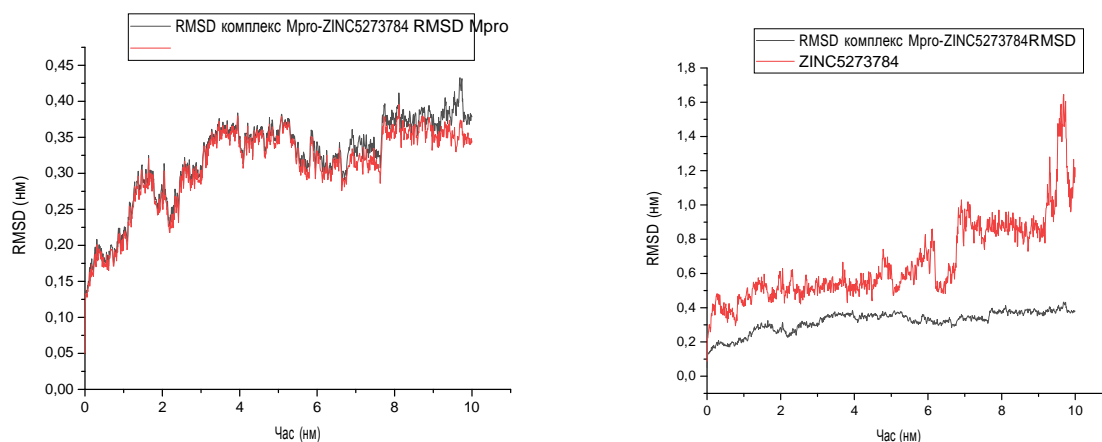


Рис 3.28 Молекулярна динаміка комплексу Mpro - ZINC5273784 (10 ns)

(а) RMSD протеази Mpro у комплексі з ZINC5273784, RMSD протеази Mpro без ліганду (б) RMSD ліганду ZINC5273784 і RMSD ліганду у комплексі з Mpro

Значення середнього квадратного відхилення (RMSD) що значно перевищує 3,5 Å вказує на недостатню стабільність комплексу білок-ліганд.

Графіки середньоквадратичних відхилень показали, що більшість комплексів є стабільними. Комплекси досягають рівноваги протягом початкової фази моделювання, а потім не демонструють значних змін положення протягом 10 нс моделювання.

### 3.3.2 Дослідження середньоквадратичних флуктуацій (RMSF) лігандів та комплексів Mpro-ліганд.

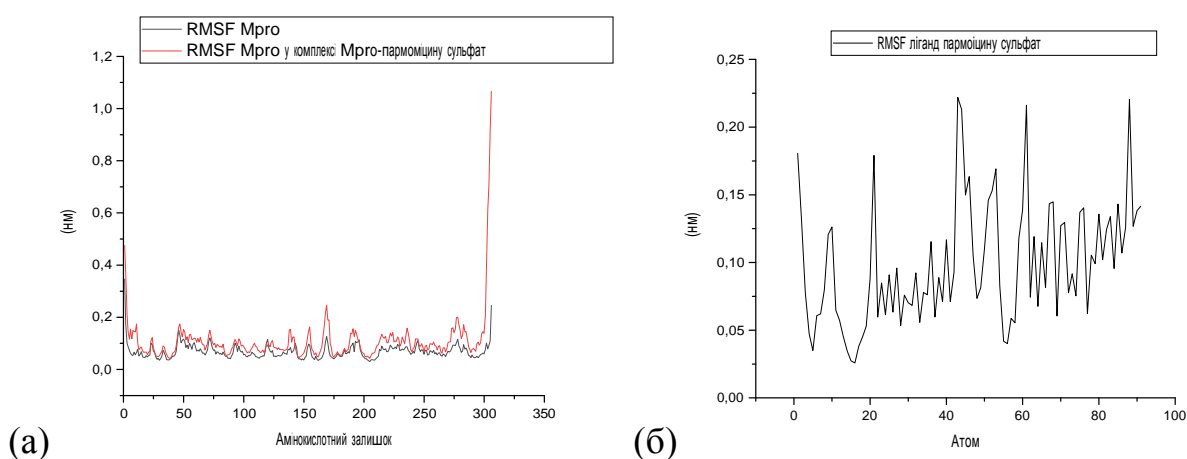


Рис. 3.29 Дослідження RMSF системи. (а) С-альфа RMSF вільної Mpro та Mpro у комплексі із пармоміцину сульфатом. (б) RMSF важких атомів пармоміцину сульфату.

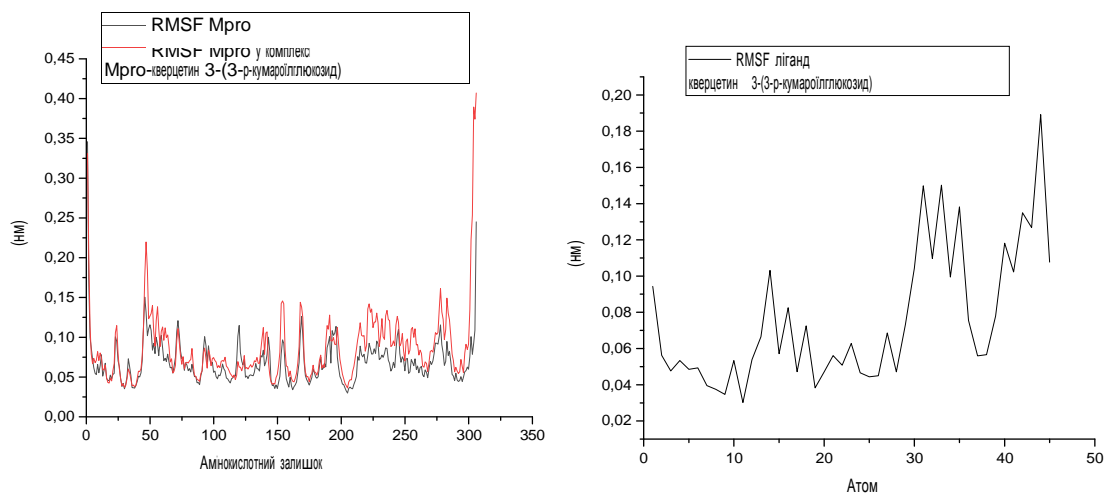


Рис. 3.30 Дослідження RMSF системи. (а) С-альфа RMSF вільної Мрго та Мрго у комплексі із кверцетин 3- (3-р-кумароїлглюкозидом). (б) RMSF важких атомів кверцетин 3- (3-р-кумароїлглюкозиду).

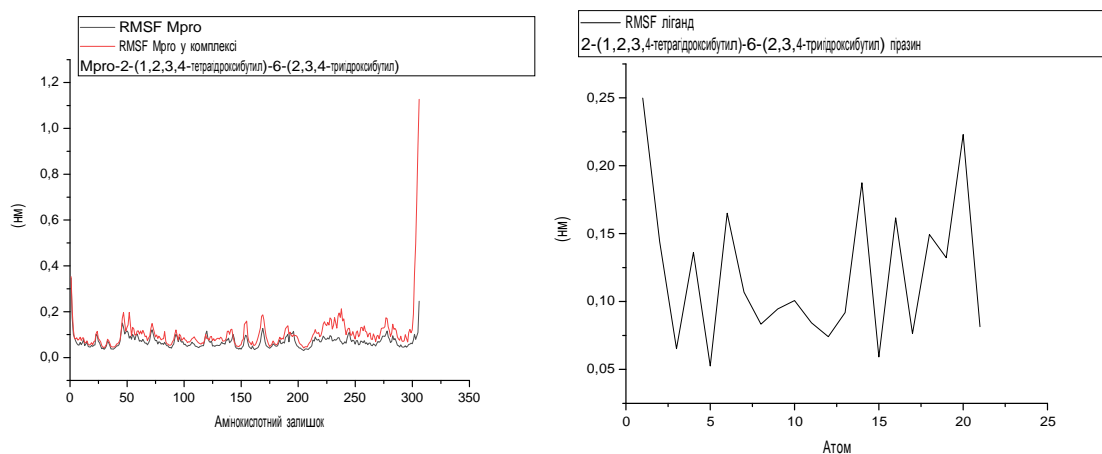


Рис. 3.31 Дослідження RMSF системи. (а) С-альфа RMSF вільної Мрго та Мрго у комплексі із 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразином. (б) RMSF важких атомів 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину.

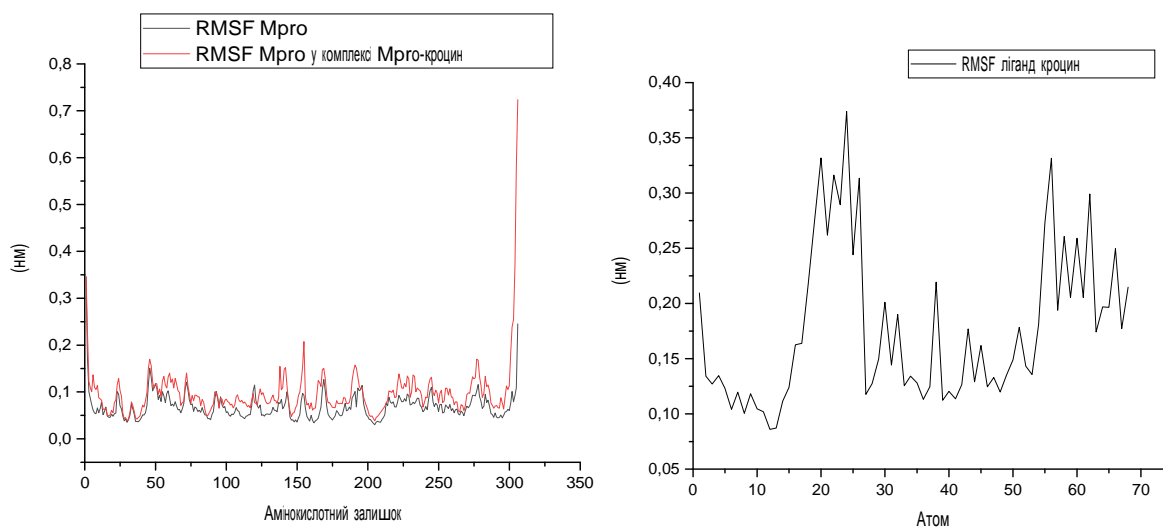


Рис. 3.32 Дослідження RMSF системи. (а) С-альфа RMSF вільної Мрго та Мрго у комплексі із кроцином. (б) RMSF важких атомів кроцину.

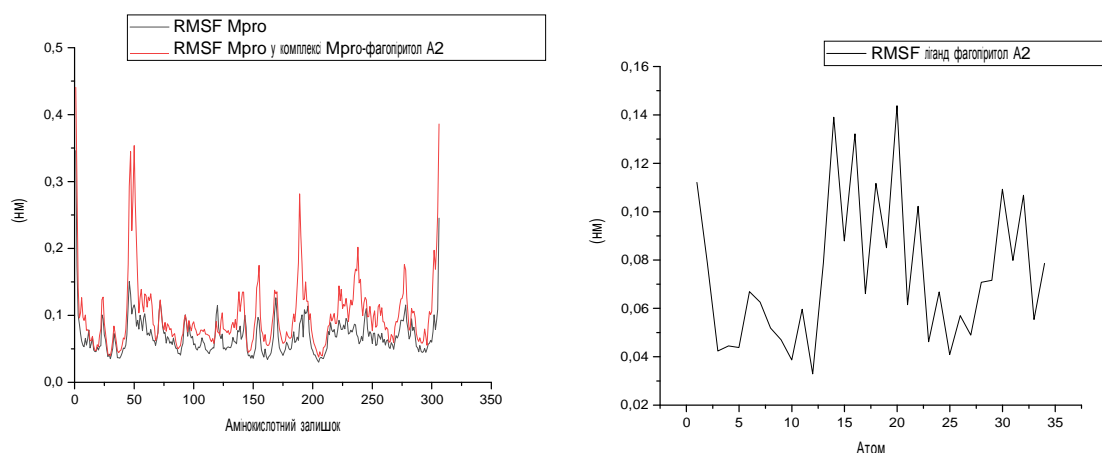


Рис. 3.33 Дослідження RMSF системи. (а) С-альфа RMSF вільної Mpro та Mpro у комплексі із фагопіритолом A2. (б) RMSF важких атомів фагопіритолу A2.

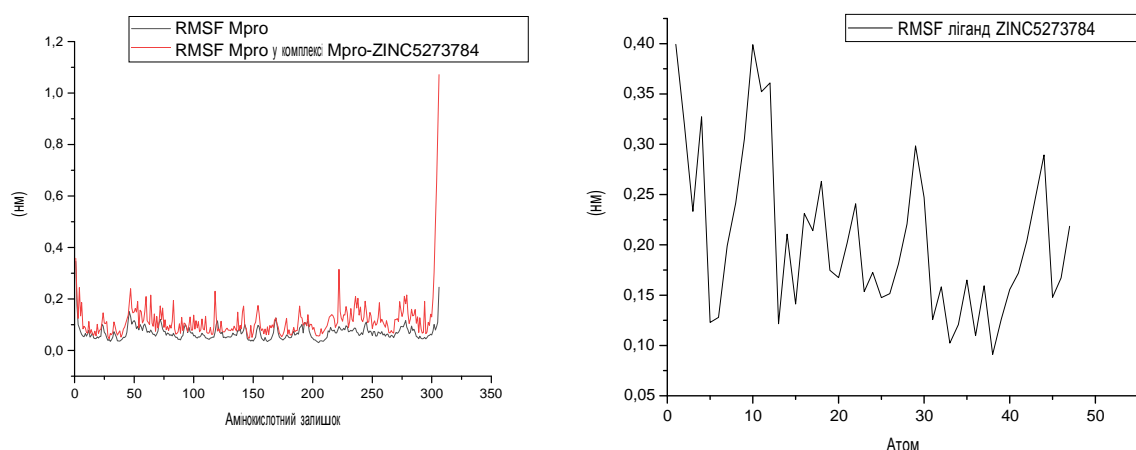


Рис. 3.34 Дослідження RMSF системи. (а) С-альфа RMSF вільної Mpro та Mpro у комплексі із ZINC5273784. (б) RMSF важких атомів ZINC5273784.

Для вивчення коливання залишків, середньоквадратичне коливання (RMSF) атомів С-альфа аналізували за допомогою інструменту `g_rmsf`. Порівняння між комплексом та незв'язаним білком показує, що утворення комплексів з досліджуваними лігандами призводить до посилення флуктуації амінокислотних залишків білка.

### 3.3.3 Дослідження вільної енергії взаємодії ліганд-білок.

Для аналізу комплексів було визначено енергію Леннарда-Джонса, Кулонівську енергію та вільну енергію взаємодії.

Потенціал Леннарда-Джонса описує потенційну енергію взаємодії двох незв'язаних атомів або молекул на основі їх відстані що їх розділяє. Рівняння потенціалу враховує різницю між силами притягання (диполь-диполь,



Coul-SR:Protein-fda (kJ/mol)	-166.766	38	100.557	81.8524	-14.4181
LJ-SR:Protein-fda (kJ/mol)	-44.9562	11	25.7365	70.5493	8.74457
Coul-14:Protein-fda	0	0	0	0	0 (kJ/mol)
LJ-14:Protein-fda	0	0	0	0	0 (kJ/mol)
Total	-211.722	43	--	152.402 (kJ/mol)	-59.3743
	152.347				

Табл 3.1 Визначення усередненого значення вільної енергії взаємодії комплексу пармоміцин сульфат-Mpro

Відсутність значень Кулонівської енергії та енергії Леннарда-Джонса що дорівнювали б або були вище 0 свідчить про стабільність комплексу.

Вільна енергія взаємодії комплексу ліганд-білок дорівнює -59.3743, що свідчить про зв'язування середньої сили.

## 2. Аналіз вільної енергії взаємодії ліганду кверцетин 3- (3-р-кумароїлглюкозиду) із Mpro.

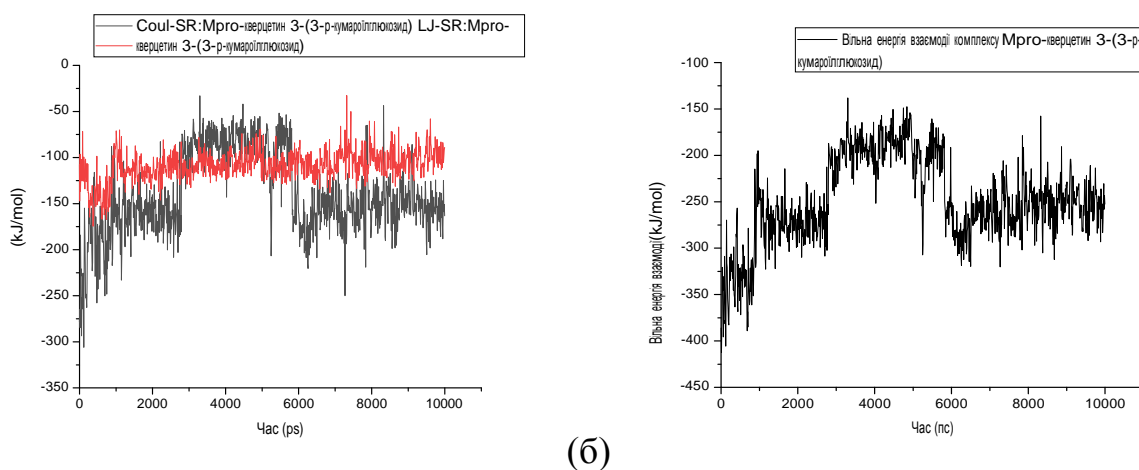


Рис. 3.36 Дослідження симуляції молекулярної динаміки комплексу кверцетин 3- (3-р-кумароїлглюкозид)-Mpro (а) Кулонівська енергія та енергія Ленарда-Джонса комплексу. (б) Вільна енергія взаємодії комплексу.

Енергія                      Середня    Похибка                      RMSD    Tot-Drift     $-kT \ln\langle e^{-(E/kT)} \rangle$



Coul-SR:Protein-zo (kJ/mol)	-96.5349	33	79.1182	196.684	10.5644
LJ-SR:Protein-zo (kJ/mol)	-49.3094	6.7	18.5273	29.5094	-21.2681
Coul-14:Protein-zo	0	0	0	0	0 (kJ/mol)
LJ-14:Protein-zo	0	0	0	0	0 (kJ/mol)
Total	-145.844	39	--	226.193 (kJ/mol)	-38.7449
	107.099				

Табл 3.3 Визначення усередненого значення вільної енергії взаємодії комплексу 2- (1,2,3,4-тетрагідроксибутил) піразин -Mpro

Відсутність значень Кулонівської енергії та енергії Леннарда-Джонса що дорівнювали б або були вище 0 вказує на стабільність досліджуваного комплексу. Вільна енергія взаємодії комплексу ліганд-білок дорівнює -38.7449, що свідчить про зв'язування середньої сили

#### 4. Аналіз вільної енергії взаємодії ліганду кроцину із Mpro

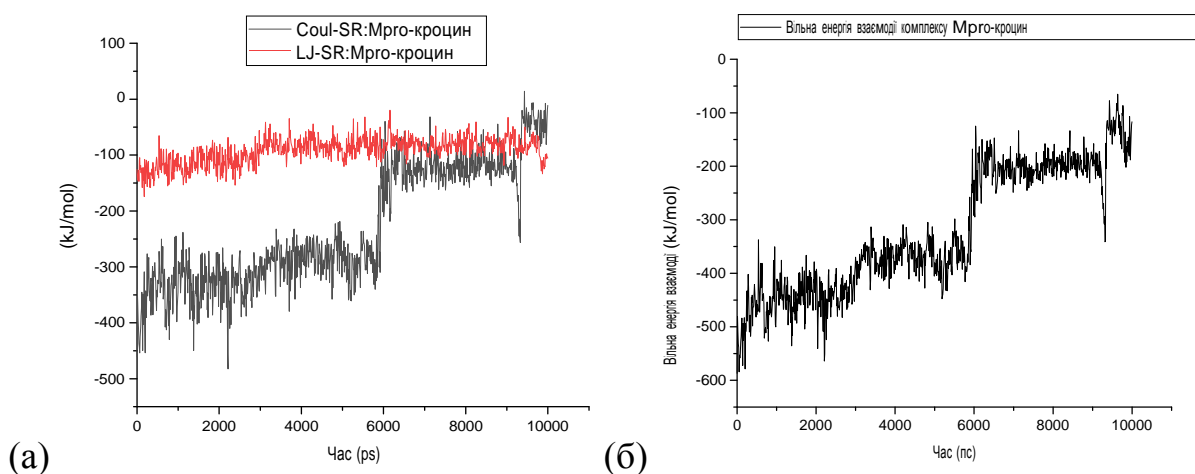


Рис. 3.38 Дослідження симуляції молекулярної динаміки комплексу кроцин-Mpro (а) Кулонівська енергія та енергія Ленарда-Джонса комплексу. (б) Вільна енергія взаємодії комплексу.

Енергія                      Середня    Похибка    RMSD   Tot-Drift    $-kT \ln \langle e^{-(E/kT)} \rangle$

Coul-SR:Protein-do (kJ/mol)	-229.52	50	107.798	331.848	-3.26285
LJ-SR:Protein-do (kJ/mol)	-92.5249	7.2	22.9565	43.7916	-36.8906
Coul-14:Protein-do	0	0	0	0	0 (kJ/mol)
LJ-14:Protein-do	0	0	0	0	0 (kJ/mol)
Total	-322.045	55	--	375.639	(kJ/mol) -95.7878
226.257					

Табл 3.4 Визначення усередненого значення вільної енергії взаємодії комплексу кроцин-Mpro

Відсутність значень Кулонівської енергії та енергії Леннарда-Джонса що дорівнювали б або були вище 0 свідчить про стабільність комплексу. Вільна енергія взаємодії комплексу ліганд-білок дорівнює -95.7878, що свідчить про задовільну силу зв'язування

### 5. Аналіз вільної енергії взаємодії ліганду фагопіритолу A2 із Mpro

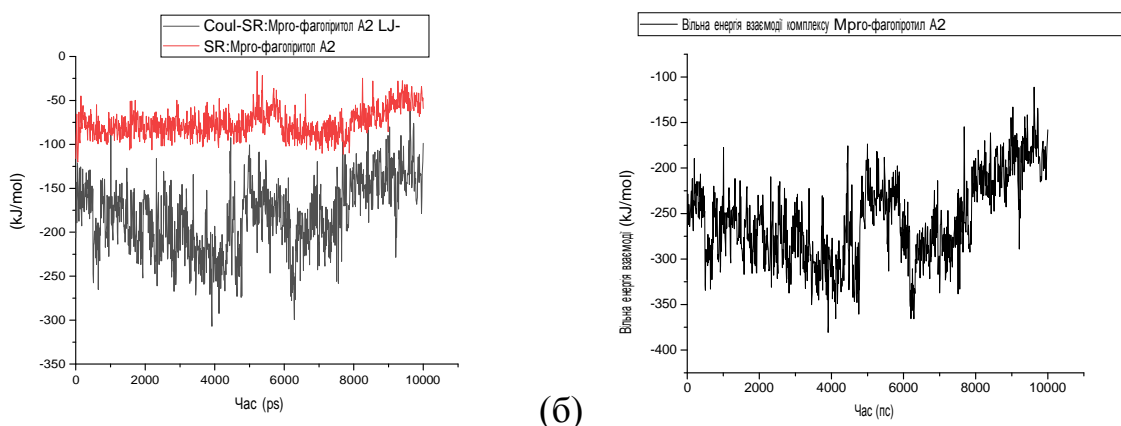


Рис. 3.39 Дослідження симуляції молекулярної динаміки комплексу фагопіритол A2-Mpro (а) Кулонівська енергія та енергія Леннарда-Джонса комплексу. (б) Вільна енергія взаємодії комплексу.

Енергія	Середня	Похибка	RMSD	Tot-Drift	$-kT \ln \langle e^{-(E/kT)} \rangle$
Coul-SR:Protein-po (kJ/mol)	-181.105	12	41.0248	52.7979	-70.6002
LJ-SR:Protein-po (kJ/mol)	-75.9519	4.9	16.1051	22.2658	-33.6372

Coul-14:Protein-po	0	0	0	0	0 (kJ/mol)
LJ-14:Protein-po	0	0	0	0	0 (kJ/mol)
Total	-257.057	16	--	75.0637	(kJ/mol) -146.552
	110.505				

Табл 3.5 Визначення усередненого значення вільної енергії взаємодії комплексу фагопіритол А2-Мрго

Відсутність значень Кулонівської енергії чи енергії Леннарда-Джонса що дорівнювали б, або перевищували 0 свідчить про стабільність комплексу фагопіритол А2-Мрго. Вільна енергія взаємодії комплексу ліганд-білок дорівнює -146.552, що свідчить про міцне зв'язування у комплексі.

#### 6. Аналіз вільної енергії взаємодії ліганду ZINC5273784 із Мрго

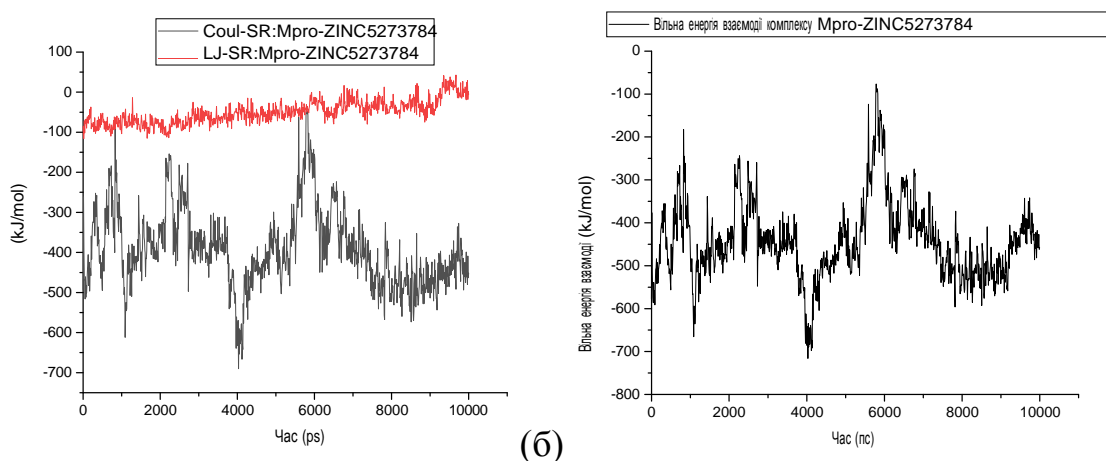


Рис. 3.40 Дослідження симуляції молекулярної динаміки комплексу ZINC5273784-Мрго (а) Кулонівська енергія та енергія Леннарда-Джонса комплексу. (б) Вільна енергія взаємодії комплексу.

Енергія	Середня	Похибка	RMSD	Tot-Drift	$-kT \ln \langle e^{-(E/kT)} \rangle$
Coul-SR:Protein-zoro (kJ/mol)	-393.362		18	97.7417	-91.2792
LJ-SR:Protein-zoro (kJ/mol)	-50.1723		11	29.3227	79.5013
Coul-14:Protein-zoro	0	0	0	0	0 (kJ/mol)
LJ-14:Protein-zoro	0	0	0	0	0 (kJ/mol)

Total	-443.534	11	--	-11.7779 (kJ/mol)	-110.103
333.431					

Табл 3.6 Визначення усередненого значення вільної енергії взаємодії комплексу ZINC5273784-Mpro

Відсутність значень Кулонівської енергії та енергії Леннарда-Джонса що дорівнювали б, або були вище 0 свідчить про стабільність комплексу ZINC5273784-Mpro. Вільна енергія взаємодії комплексу ліганд-білок дорівнює -110.103, що є показником сильного зв'язування у комплексі.

Найгірші показники енергії комплексів Mpro зі сполуками пармоміцину сульфатом та 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразином відповідають даним значення середнього квадратного відхилення, отриманим із графіків RMSD.

За допомогою аналізу молекулярних симуляцій було ідентифіковано 4 сполуки із найкращими показниками вільної енергії взаємодії: кверцетин 3- (3-р-кумаройлглюкозид), кроцин, фагопіритол A2 та ZINC5273784.

### 3.4 Визначення лікоподібності сполук із найкращими показниками енергії зв'язування з димеризаційним сайтом Mpro

Біодоступність є важливою властивістю потенційних лікарських препаратів, оскільки вона визначає частку введеної дози хімічної сполуки, яка поглинається. На цей параметр сильно впливають фізико-хімічні властивості сполук, особливо їх гідрофільність та розчинність. У даній роботі речовини-кандидати були відфільтровані за наступними параметрами: порушення правил-5 (LipinskiR05), правило Вебера, правило Егана та пероральна фізико-хімічна оцінка Bayer, за допомогою програми FAFDrugs. Правило Ліпінського враховує значення чотирьох параметрів, що вказують на хорошу біодоступність: молекулярна маса  $MW \leq 500$ ; Донори Н-зв'язків  $\leq 5$ ; Акцептори Н-зв'язків  $\leq 10$ ;  $\log P \leq 5$ . Один із цих параметрів за межами діапазону є прийнятним результатом; якщо два параметри виходять за межі діапазонів, можливе погане поглинання або проникність. Правило Вебера розглядає біодоступність для сполук з кількістю зв'язків здатних до обертання  $\leq 10$  і  $tPSA \leq 140 \text{ \AA}^2$ . Правило Егана розглядає біодоступність для сполук з  $0 \geq tPSA \leq 132 \text{ \AA}^2$  та  $-1 \geq \log P \leq 6$ . [74]

Деякі молекулярні властивості хімічної сполуки сильно впливають на її біодоступність. На приклад:

- (i) Кількість Н-донорів, вимірюється як кількістю атомів, що мають принаймні один пов'язаний водень.
- (ii) Кількість Н-акцепторів, виражається як кількість атомів, які можуть приймати водень: будь-який азот, кисень і фтор, за винятком N з позитивним формальним зарядом, вищими ступенями окиснення. Кількість Н-донорів та Н-акцепторів впливає на площу полярної поверхні молекули і, отже, на швидкість її проникнення через мембрани.
- (iii) Кількість зв'язків здатних до обертання, що визначається як будь-який одинарний зв'язок (не в кільці), зв'язаний з нетермінальним важким атомом. З цього числа амідні зв'язки C – N виключаються, оскільки вони мають високий енергетичний бар'єр обертання. Кількість зв'язків здатних до обертання має прямі наслідки для гнучкості речовини. Висока гнучкість не відповідає обмеженням, сумісним із цільовим зв'язуванням, зменшуючи біодоступність.
- (iv) Молекулярна маса використовується як параметр біодоступності завдяки її кореляції з кількістю обертаючихся та водневих зв'язків, молекули з більшою кількістю обертаючихся зв'язків та водневих зв'язків зазвичай мають більшу молекулярну масу. Абсорбція перорально введеного препарату залежить від його розчинності. З цієї причини властива розчинність є ще одним важливим параметром, який враховується при описі біодоступності. [75]

Окрім біодоступності, важливим є також профіль безпеки, який ілюструє можливі несприятливі ефекти хімічних сполук. Для оцінки профілю безпеки досліджуваних сполук була визначена схильність викликати фосфоліпідоз (надмірне накопичення фосфоліпідів у лізосомах).

Сполуки були також відфільтровані за правилом GlaxoSmithKline 4/400 (GSK 4/400), яке вважає, що сполуки з  $\log P > 4$  і  $M > 400$  Da мають менш сприятливий профіль безпеки, та правилом Pfizer 3/75, що пов'язує  $\log P > 3$  і  $tPSA < 75 \text{ \AA}^2$  з несприятливим впливом хімічних сполук.

### 3.4.1 Дослідження біодоступності речовин-кандидатів

Була проведена оцінка токсичності та біодоступності 46 лігандів з найвищими показниками вільної енергії зв'язування .

Більшість досліджуваних речовин із підгрупи сполук що не містять структур PAINS проявляють низьку пероральну біодоступність EGAN та 3 порушення правила ліпінського. Виключеннями є речовини ZINC000009574481 та ZINC000008220119 із 2 порушеннями правила Ліпінського. Ліганди ZINC000003914746; ZINC000008218858; ZINC000004748476; ZINC000003916648; ZINC000008830558 мають високу біодоступність EGAN та 3 порушення правила ліпінського. Речовина ZINC000005273784 проявляє

0 порушень правила, що вказує на її перспективність як потенційний лікарський препарат. (табл. 5)

Із 16 лігандів із групи сполук що споживаються людиною, обраних для аналізу потенційної токсичності та біодоступності 13 проявили високу пероральну біодоступність EGAN. Речовинами із низькою біодоступністю були ZINC936070151; ZINC000085815081 та ZINC100823327. Менш ніж 3 порушення правила ліпінського мали речовини ZINC100823369 (1) та ZINC000100828363 (2).

### 3.4.2 Дослідження потенційної токсичності речовин-кандидатів

Із 47 досліджених речовин 6 проявили здатність викликати фосфоліпідоз. Фосфоліпідоз, це порушення зберігання фосфоліпідів, що призводить до надмірного накопичення фосфоліпідів у лізосомах різних типів тканин. Інгібування лізосомальної фосфоліпази A1, A2 та / або C сприяє накопиченню у клітині комплексів катіонний амфифільний лікарський засіб (CAD) - фосфоліпід. Потенційними індукторами фосфоліпідозу є сполуки ZINC3914746; ZINC8218858 ; ZINC000004748476; ZINC8220483; ZINC830558; ZINC000004726430 . Лікарські засоби, які індукують фосфоліпідоз, часто ретельніше перевіряються контролюючими організаціями.

Усі 47 досліджених речовин мають хороший показник GSK правила 4/400 та оцінки 3\_75, що свідчить про їх низьку прогнозовану токсичність та перспективність у якості потенційних кандидатів для розробки лікарських препаратів.

### 3.4.3 Поглиблений аналіз ADMET характеристик обраних лігандів

Такі показники як Caco2 проникність, кишкова проникність, інгібування Р-глікопротеїнів I та II, VDss, повне виведення, субстрат ниркового OCT2, токсичність за Еймсом, максимальна нелетальна доза (для людини), гепатотоксичність були визначені для речовин з найкращими показниками вільної енергії зв'язування, та вільної енергії взаємодії.

Високу кишкову проникність було виявлено у сполук - (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину (20%) та кверцетину 3- (3-р-кумароїлглюкозиду) (50%).

## Інгібування hERG I

Функція потенціалзалежних калієвих каналів (кодований hERG), полягає у забезпеченні затриманого випрямного калієвого струму (IKr), що лежить в основі реполяризації потенціалу серцевої дії. Жодна з досліджених речовин не є інгібітором hERG I, що вказує на низьку токсичність речовин, оскільки фармакологічна блокада каналу hERG призводить до уповільнення реполяризації, а отже, подовження тривалості потенціалу дії та збільшення інтервалу QT, виміряного на електрокардіограмі. Це може стати причиною розвитку побічних ефектів, що призводить до відмови від багатьох потенційних лікарських препаратів на ранніх стадіях клінічного тестування. Тому скринінг на інгібування hERG є важливим елементом доклінічних досліджень.

## Зв'язування білками плазми

Більшість препаратів у плазмі існують у рівновазі між незв'язаним станом та зв'язаними з білками сироватки. На ефективність даного лікарського засобу може впливати ступінь зв'язування білків в крові, оскільки чим більше пов'язаного, тим менш ефективно він може проходити через клітинні мембрани або дифундувати. Прогнозовані вільні частки досліджуваних сполук у крові людини (Fu) після поглинання становлять: 0.585 для 2-(1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину, 0.144 для кверцетину 3- (3-р-кумароїлглюкозиду), 0.381 для кроцину, 0.408 для фагопіритулу A2.

## Субстрат OAT2

Органічний катіонний транспортер 2 - це транспортер ниркового захоплення, який відіграє важливу роль у розподілі та нирковому кліренсі ліків та ендогенних сполук. Оцінка потенціалу транспортування за допомогою OAT2 надає корисну інформацію стосовно потенційних протипоказань лікарського засобу. Із досліджуваних речовин субстратом OAT2 є кроцин. Ліганди 2-(1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразин; кверцетин 3- (3-р-кумароїлглюкозиду); фагопіритул A2 не є субстратами для OAT2, отже можливі протипоказання до використання даних сполук.

## Інгібування цитохрому P450

Цитохром P450 є важливим ферментом детоксикації організму і, головним чином, функціонує у печінці. Він окислює ксенобіотики, полегшуючи їх виведення. Багато лікарських речовин дезактивуються цитохромом P450.

Інгібітори цього ферменту можуть впливати на метаболізм ліків і мати побічні ефекти. Тому важливо оцінювати здатність сполук інгібувати цитохром P450. Сполука вважається інгібітором цитохрому P450, якщо концентрація, необхідна для 50-процентного інгібування, є меншою за 10 мкМ. Досліджувані речовини не є інгібіторами CYP2D6; CYP3A4 ; CYP1A2; CYP2C19; CYP2C9 ; CYP2D6; CYP3A4 на основі чого можна прогнозувати здатність речовин добре метаболізуватися в організмі. Препарати, що пригнічують активність CYP2D6 здатні призводити до несприятливих побічних ефектів.

#### Напівлетальна доза речовин (LD50)

Важливо враховувати можливу токсичну потенційної сполуки. LD50 - це стандартне вимірювання гострої токсичності, що використовується для оцінки відносної токсичності різних молекул. LD50 - це кількість сполуки у mol/kg що спричиняє загибель 50% групи досліджуваних тварин. Для речовин були визначені такі показники LD50: LD50 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразин = 2.816 mol/kg; LD50 кверцетин 3- (3-р-кумароїлглюкозид) = 2.685 mol/kg; LD50 кроцин = 2.477 mol/kg; LD50 фагопіритол A2 = 2.542 mol/kg; LD50 ZINC3922065 = 2.482 mol/kg

#### Потенційна мутагенність речовин

З допомогою теста Еймса було визначено потенційну мутагенність речовин. Тест Еймса - широко застосовуваний метод оцінки мутагенного потенціалу сполук. Позитивний тест вказує на те, що сполука є мутагенною і, отже, може діяти як канцероген. Жодна з досліджуваних речовин не має мутагенних властивостей згідно з результатами теста Еймса

#### Потенційна гепатотоксичність речовин

Важливим елементом дослідження ADMET характеристик є визначення потенційної гепатотоксичності сполук. Речовина може бути класифікована як гепатотоксична, якщо вона здатна спричинити принаймні одну патологічну або фізіологічну реакцію у печінці, що пов'язана з порушенням нормального функціонування органа. Із досліджених речовин жодна не проявляє гепатотоксичності окрім 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину та ZINC5273784.

Більшість досліджених речовин проявляють невисокі значення біодоступності, що може перешкоджати пероральному введенню препаратів.

Переважає частина досліджених лігандів не є токсичними, мутагенними, не викликають фосфоліпозу.

## ВИСНОВКИ

1. Показано, що білок Mpro SARS CoV 2 характеризується високою консервативністю послідовності, із RMSD = 1.125 Å при вирівнюванні із великою протеазою SARS CoV та RMSD = 2.132 Å при вирівнюванні з великою протеазою MERS CoV.
2. Виявлено ряд лігандів, що утворюють водневі зв'язки, сольові містки, пі-пі катіонні взаємодії з такими важливими для димеризації амінокислотами, як LYS4; ARG 5; Tyr126; Val125, Glu290, Lys137, Phe140; Glu166 та ін. потенційно інгібуючи димеризацію білка, необхідну для каталітичної активності.
3. Виявлено ряд лігандів, здатних формувати комплекси як з димеризаційним сайтом Mpro SARS CoV 2, так і з активним центром білка, зв'язуючи амінокислотні залишки каталітичної діади, а також Met165, Glu166 та His172, що приймають участь в утворенні субсайта S1, Phe140 та His163, що утворюють задню стінку підсайту S1.
4. Досліджено RMSD та значення вільної енергії взаємодії симуляцій молекулярної динаміки комплексів обраних лігандів із Mpro SARS CoV 2 із ідентифікацією чотирьох сполук із найкращими показниками: кверцетин 3-(3-р-кумароїлглюкозид), кроцин, фагопіритол A2 та речовина ZINC5273784
5. Проаналізовано ADME профілі обраних сполук із визначенням таких параметрів як LD50, Caco2 проникність, потенційна гепатотоксичність, мутагенність, інгібування цитохрому P450, інгібування hERG I. Більшість проаналізованих сполук мали сприятливі профілі токсичності, але достатньо низькі показники біодоступності.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Pal M, Berhanu G, Desalegn C, Kandi V. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. *Cureus*. 2020 Mar 26;12(3):e7423. doi: 10.7759/cureus.7423. PMID: 32337143; PMCID: PMC7182166.
2. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. 2014 Aug 7;6(8):2991-3018. doi: 10.3390/v6082991. PMID: 25105276; PMCID: PMC4147684.
3. Wang L, Xiang Y. Spike Glycoprotein-Mediated Entry of SARS Coronaviruses. *Viruses*. 2020 Nov 11;12(11):1289. doi: 10.3390/v12111289. PMID: 33187074; PMCID: PMC7696831.
4. Chambers JP, Yu J, Valdes JJ, Arulanandam BP. SARS-CoV-2, Early Entry Events. *J Pathog*. 2020 Nov 24;2020:9238696. doi: 10.1155/2020/9238696. PMID: 33299610; PMCID: PMC7707962.5. Philip V'kovski, Annika Kratzel, Silvio Steiner, Hanspeter Stalder & Volker Thiel Coronavirus biology and replication: implications for SARS-CoV-2
6. Astuti I, Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes Metab Syndr*. 2020 Jul-Aug;14(4):407-412. doi: 10.1016/j.dsx.2020.04.020. Epub 2020 Apr 18. PMID: 32335367; PMCID: PMC7165108.
7. Báez-Santos YM, St John SE, Mesecar AD. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res*. 2015 Mar;115:21-38. doi: 10.1016/j.antiviral.2014.12.015. Epub 2014 Dec 29. PMID: 25554382; PMCID: PMC5896749.8. Yunjeong Kim, Scott Lovell, [...], and Kyeong-Ok Chang Broad-Spectrum Antivirals against 3C or 3C-Like Proteases of Picornaviruses, Noroviruses, and Coronaviruses
9. Jin Z, Wang H, Duan Y, Yang H. The main protease and RNA-dependent RNA polymerase are two prime targets for SARS-CoV-2. *Biochem Biophys Res Commun*. 2021 Jan 29;538:63-71. doi: 10.1016/j.bbrc.2020.10.091. Epub 2020 Nov 21. PMID: 33288200; PMCID: PMC7680044.
10. Arya R, Kumari S, Pandey B, Mistry H, Bihani SC, Das A, Prashar V, Gupta GD, Panicker L, Kumar M. Structural insights into SARS-CoV-2 proteins. *J Mol Biol*. 2021 Jan 22;433(2):166725. doi: 10.1016/j.jmb.2020.11.024. Epub 2020 Nov 24. PMID: 33245961; PMCID: PMC7685130.

11. Hillen HS. Structure and function of SARS-CoV-2 polymerase. *Curr Opin Virol.* 2021 Jun;48:82-90. doi: 10.1016/j.coviro.2021.03.010. Epub 2021 Apr 6. PMID: 33945951; PMCID: PMC8023233.
12. Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells.* 2020 May 20;9(5):1267. doi: 10.3390/cells9051267. PMID: 32443810; PMCID: PMC7291026.
13. Snijder EJ, Decroly E, Ziebuhr J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res.* 2016;96:59-126. doi: 10.1016/bs.aivir.2016.08.008. Epub 2016 Sep 14. PMID: 27712628; PMCID: PMC7112286.
14. Sicari D, Chatziioannou A, Koutsandreas T, Sitia R, Chevet E. Role of the early secretory pathway in SARS-CoV-2 infection. *J Cell Biol.* 2020 Sep 7;219(9):e202006005. doi: 10.1083/jcb.202006005. Erratum in: *J Cell Biol.* 2020 Sep 7;219(9): PMID: 32725137; PMCID: PMC7480111.
15. Hidalgo P, Valdés M, González RA. Molecular biology of coronaviruses: an overview of virus-host interactions and pathogenesis. *Bol Med Hosp Infant Mex.* 2021;78(1):41-58. English. doi: 10.24875/BMHIM.20000249. PMID: 33661875.
16. Goyal B, Goyal D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb Sci.* 2020 Jun 8;22(6):297-305. doi: 10.1021/acscombsci.0c00058. Epub 2020 May 27. PMID: 32402186; PMCID: PMC7252589.
17. Suárez D, Díaz N. SARS-CoV-2 Main Protease: A Molecular Dynamics Study. *J Chem Inf Model.* 2020 Dec 28;60(12):5815-5831. doi: 10.1021/acs.jcim.0c00575. Epub 2020 Aug 3. PMID: 32678588; PMCID: PMC7409944.
18. Silvestrini, L., Belhaj, N., Comez, L. et al. The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci Rep* 11, 9283 (2021). <https://doi.org/10.1038/s41598-021-88630-9>
19. Shi J, Han N, Lim L, Lua S, Sivaraman J, Wang L, Mu Y, Song J. Dynamically-driven inactivation of the catalytic machinery of the SARS 3C-like protease by the N214A mutation on the extra domain. *PLoS Comput Biol.* 2011 Feb;7(2):e1001084. doi: 10.1371/journal.pcbi.1001084. Epub 2011 Feb 24. PMID: 21390281; PMCID: PMC3044768.

20. Cheng SC, Chang GG, Chou CY. Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. *Biophys J.* 2010 Apr 7;98(7):1327-36. doi: 10.1016/j.bpj.2009.12.4272. PMID: 20371333; PMCID: PMC2849084.
21. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* 2002 Jul 1;21(13):3213-24. doi: 10.1093/emboj/cdf327. PMID: 12093723; PMCID: PMC126080.
22. Du QS, Wang SQ, Zhu Y, Wei DQ, Guo H, Sirois S, Chou KC. Polyprotein cleavage mechanism of SARS CoV Mpro and chemical modification of the octapeptide. *Peptides.* 2004 Nov;25(11):1857-64. doi: 10.1016/j.peptides.2004.06.018. PMID: 15501516; PMCID: PMC7115412.
23. Carlos A. Ramos-Guzmán, J. Javier Ruiz-Pernía, and Iñaki Tuñón Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods // *ACS Catalysis* 2020 10 (21), 12544-12554 DOI: 10.1021/acscatal.0c03420
24. Świderek K, Moliner V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem Sci.* 2020 Jun 25;11(39):10626-10630. doi: 10.1039/d0sc02823a. PMID: 34094317; PMCID: PMC8162313.
25. Bzówka M, Mitusińska K, Raczyńska A, Samol A, Tuszyński JA, Góra A. Structural and Evolutionary Analysis Indicate That the SARS-CoV-2 Mpro Is a Challenging Target for Small-Molecule Inhibitor Design. *Int J Mol Sci.* 2020 Apr 28;21(9):3099. doi: 10.3390/ijms21093099. PMID: 32353978; PMCID: PMC7247150.
26. Ye Q, West AMV, Silletti S, Corbett KD. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* 2020 Sep;29(9):1890-1901. doi: 10.1002/pro.3909. Epub 2020 Aug 6. PMID: 32654247; PMCID: PMC7405475.
27. Kneller DW, Phillips G, Weiss KL, Pant S, Zhang Q, O'Neill HM, Coates L, Kovalevsky A. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J Biol Chem.* 2020 Dec 11;295(50):17365-17373. doi: 10.1074/jbc.AC120.016154. Epub 2020 Oct 15. PMID: 33060199; PMCID: PMC7832724.

28. Wioletta Rut, Mikolaj Zmudzinski , Zongyang Lyu, Digant Nayak Scott J. Snipas , Miklos Bekes, Tony T. Huang , Shaun K. Olsen , Marcin Drag Activity profiling of SARS-CoV-2-PLpro protease provides structural framework for anti-COVID-19 drug design // doi: <https://doi.org/10.1101/2020.04.29.068890>
29. Paasche A, Zipper A, Schäfer S, Ziebuhr J, Schirmeister T, Engels B. Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry*. 2014 Sep 23;53(37):5930-46. doi: 10.1021/bi400604t. Epub 2014 Sep 7. PMID: 25196915.
30. Daniel W. Kneller, Gwyndalyn Phillips, Kevin L. Weiss, Swati Pant, Qiu Zhang, Hugh M. O'Neill, View ORCID ProfileLeighton Coates, View ORCID ProfileAndrey Kovalevsky Protonation states in SARS-CoV-2 main protease mapped by neutron crystallography // doi: <https://doi.org/10.1101/2020.09.22.308668>
31. Daniel W. Kneller<sup>1,2</sup> , Gwyndalyn Phillips<sup>1,2</sup> , Kevin L. Weiss<sup>1,2</sup> , Swati Pant<sup>1,2</sup> , Qiu Zhang<sup>1,2</sup>, Hugh M. O'Neill<sup>1</sup>, Leighton Coates<sup>1,2,3</sup> , and Andrey Kovalevsky. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. // Received for publication, September 22, 2020, and in revised form, October 12, 2020 Published, Papers in Press, October 15, 2020, DOI 10.1074/jbc.AC120.016154
32. Robert MCnard, Euridice Carmona , Citline Plouffe, Dieter Briimme, Yasuo Konishi, Jean Lefebvre and Andrew C. Storer. The specificity of the S,' subsite of cysteine proteases. // Biotechnology Research Institute, National Research Council of Canada, 6100 Avenue Royalmount, Montrtal, Quebec, H4P 2R2 Canada Received 8 June 1993
33. Jin Z., Du X., Xu Y., Deng Y., Liu M., Zhao Y., Zhang B., Li X., Zhang L., Peng C., Duan Y., Yu J., Wang L., Yang K., Liu F., Jiang R., Yang X., You T., Liu X., Yang X., Bai F., Liu H., Liu X., Guddat L.W., Xu W., Xiao G., Qin C., Shi Z., Jiang H., Rao Z., Yang H. Structure of Mpro from COVID-19 virus and discovery of its inhibitors.
34. Anacleto Silva de Souza, Jose David Rivera, [...], and Cristiane Rodrigues Guzzo. Molecular Dynamics Reveals Complex Compensatory Effects of Ionic Strength on the Severe Acute Respiratory Syndrome Coronavirus 2 Spike/Human Angiotensin-Converting Enzyme 2 Interaction// *J Phys Chem Lett*. 2020 Dec 17; 11(24): 10446–10453.

35. Middleton CT, Woys AM, Mukherjee SS, Zanni MT. Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy. *Methods*. // 2010 Sep;52(1):12-22. doi: 10.1016/j.ymeth.2010.05.002. Epub 2010 May 22. PMID: 20472067; PMCID: PMC2933966.
36. Maik Gotte. Free energy calculations of protein-ligand complexes with computational molecular dynamics. // Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultäten der Georg-August-Universität zu Göttingen
37. Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, Ji XL, Liu SQ. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci*. 2016 Jan 26;17(2):144. doi: 10.3390/ijms17020144. PMID: 26821017; PMCID: PMC4783878.
38. Steinbrecher T, Labahn A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr Med Chem*. 2010;17(8):767-85. doi: 10.2174/092986710790514453. PMID: 20088755.
39. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*. 2007;36:21-42. doi: 10.1146/annurev.biophys.36.040306.132550. PMID: 17201676.
40. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004 Mar 25;47(7):1739-49. doi: 10.1021/jm0306430. PMID: 15027865.
41. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev*. 2017 Apr;9(2):91-102. doi: 10.1007/s12551-016-0247-1. Epub 2017 Jan 16. PMID: 28510083; PMCID: PMC5425816.
42. David P Elder. Solubility in Pharmaceutical R&D: Predictions and Reality // January 2015 American Pharmaceutical Review 16(2)
43. Dana Craciun<sup>1, a</sup>), Dorina Modra<sup>2</sup> and Adriana Isvoran. ADME-Tox Profiles of Some Food Additives and Pesticides // AIP Conference Proceedings 1694, 040007 (2015); <https://doi.org/10.1063/1.4937259>

44. Tsaïoun K, Blaauboer BJ, Hartung T. Evidence-based absorption, distribution, metabolism, excretion (ADME) and its interplay with alternative toxicity methods. *ALTEX*. 2016;33(4):343-358. doi: 10.14573/altex.1610101. PMID: 27806179.
45. Santoro A, Bianco G, Picerno P, Aquino RP, Autore G, Marzocco S, Gazzero P, Lioi MB, Bifulco M. Verminoside- and verbascoside-induced genotoxicity on human lymphocytes: involvement of PARP-1 and p53 proteins. *Toxicol Lett*. 2008 May 5;178(2):71-6. doi: 10.1016/j.toxlet.2008.02.006. Epub 2008 Feb 23. PMID: 18395372.
46. Santos-Cruz LF, Ávila-Acevedo JG, Ortega-Capitaine D, Ojeda-Duplancher JC, Perdigón-Moya JL, Hernández-Portilla LB, López-Dionicio H, Durán-Díaz A, Dueñas-García IE, Castañeda-Partida L, García-Bores AM, Heres-Pulido ME. Verbascoside is not genotoxic in the ST and HB crosses of the *Drosophila* wing spot test, and its constituent, caffeic acid, decreases the spontaneous mutation rate in the ST cross. *Food Chem Toxicol*. 2012 Mar;50(3-4):1082-90. doi: 10.1016/j.fct.2011.12.006. Epub 2011 Dec 11. PMID: 22197714.
47. Irina Kufareva and Ruben Abagyan. Methods of protein structure comparison // *Methods Mol Biol*. Author manuscript; available in PMC 2015 Feb 9. Published in final edited form as: *Methods Mol Biol*. 2012; 857: 231–257.
48. Teague Sterling and John J. Irwin. ZINC 15 – Ligand Discovery for Everyone // *Journal of Chemical Information and Modeling* 2015 55 (11), 2324-2337 DOI: 10.1021/acs.jcim.5b00559
49. Kodakkat Parambil Safna Hussan, Mohamed Shahin Thayyil, Thaikadan Shameera Ahamed and Karuvanthodi Muraleedharan. Biological Evaluation and Molecular Docking Studies of Benzalkonium Ibuprofenate // Submitted: June 10th 2019 Reviewed: October 17th 2019 Published: February 24th 2020 DOI: 10.5772/intechopen.90191
50. Xuan-Yu Meng, Hong-Xing Zhang, [...], and Meng Cui. Molecular Docking: A powerful approach for structure-based drug discovery // *Curr Comput Aided Drug Des*. Author manuscript; available in PMC 2012 Jun 1. Published in final edited form as: *Curr Comput Aided Drug Des*. 2011 Jun 1; 7(2): 146–157.
51. Zhiyong Zhou, Anthony K. Felts, [...], and Ronald M. Levy. Comparative Performance of Several Flexible Docking Programs and Scoring Functions: Enrichment Studies for a Diverse Set of Pharmaceutically Relevant Targets // *J Chem Inf Model*. Author manuscript; available in PMC 2008 Sep 23. Published in

final edited form as: *J Chem Inf Model.* 2007; 47(4): 1599–1608. Published online 2007 Jun 23. doi: 10.1021/ci7000346

52. Feng Ding, Shuangye Yin, and Nikolay V. Dokholyan. Rapid flexible docking using a stochastic rotamer library of ligands // *J Chem Inf Model.* Author manuscript; available in PMC 2011 Sep 27. Published in final edited form as: *J Chem Inf Model.* 2010 Sep 27; 50(9): 1623–1632. doi: 10.1021/ci100218t

53. Liu, X., Zhang, B., Jin, Z., Yang, H., Rao, Z. The crystal structure of COVID-19 main protease in complex with an inhibitor N3 // DOI: 10.2210/pdb6lu7/pdb Primary publication DOI: 10.1038/s41586-020-2223-y Initial deposition on: 26 January 2020

54. Stephen K Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Li Chen, Gregg V Crichlow, Cole H Christie, Kenneth Dalenberg, Luigi Di Costanzo, Jose M Duarte. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences // *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D437–D451, <https://doi.org/10.1093/nar/gkaa1038> Published: 19 November 2020

55. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

56. Charles W Westin, Brett Hanson, Paul A Craig. Using PyMOL's Align Feature to Create a Database of Ligand Binding Site Files for the Structural Analysis of Proteins. // First published: 01 April 2007 <https://doi.org/10.1096/fasebj.21.5.A296-d>

56. Virtual Screening With GLIDE Copyright © 2015 Schrödinger, LLC.

57. Glide Quick Start Guide Copyright © 2015 Schrödinger, LLC.

58. Glide User Manual Copyright © 2015 Schrödinger, LLC.

59. Glide, version 6.7, Schrödinger, LLC, New York, NY, 2015.

60. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem.* 2004, 47, 1739-1749.

61. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* 2004, 47, 1750–1759.
62. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* 2006, 49, 6177–6196.
63. David Lagorce, Lina Bousslama, Jerome Becot, Maria A Miteva, Bruno O Villoutreix. FAF-Drugs4: free ADME-tox filtering computations for chemical biology and early stages drug discovery // *Bioinformatics*, Volume 33, Issue 22, 15 November 2017, Pages 3658–3660, <https://doi.org/10.1093/bioinformatics/btx491>
64. Douglas E. V. Pires, Tom L. Blundell, and David B. Ascher. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures // *J Med Chem.* 2015 May 14; 58(9): 4066–4072. Published online 2015 Apr 10. doi: 10.1021/acs.jmedchem.5b00104 PMID: 25860834
65. Least-Squares Superpositions and Root-Mean-Squares Deviation
66. Graham F. Smith. Designing Drugs to Avoid Toxicity // *Progress in Medicinal Chemistry*, Elsevier, Volume 50, 2011, Pages 1-47, ISSN 0079-6468, ISBN 9780123812902, <https://doi.org/10.1016/B978-0-12-381290-2.00001-X>.
67. William J. Egan, Kenneth Merz, John J. Baldwin. Prediction of Drug Absorption Using Multivariate Statistics // November 2000 *Journal of Medicinal Chemistry*
68. National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 441375, Paromomycin sulfate. Retrieved June 9, 2021 from <https://pubchem.ncbi.nlm.nih.gov/compound/Paromomycin-sulfate>.
69. National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 31703, Doxorubicin. Retrieved June 9, 2021 from <https://pubchem.ncbi.nlm.nih.gov/compound/Doxorubicin>.
70. National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 3736, Iopromide. Retrieved June 9, 2021 from <https://pubchem.ncbi.nlm.nih.gov/compound/Iopromide>.

71. National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 3730, Iohexol. Retrieved June 9, 2021 from <https://pubchem.ncbi.nlm.nih.gov/compound/Iohexol>.
72. National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 493570, Riboflavin. Retrieved June 9, 2021 from <https://pubchem.ncbi.nlm.nih.gov/compound/Riboflavin>.
73. National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 3467, Gentamicin. Retrieved June 9, 2021 from <https://pubchem.ncbi.nlm.nih.gov/compound/Gentamicin>.
74. Tomar, S., Johnston, M.L., St John, S.E., Osswald, H.L., Nyalapatla, P.R., Paul, L.N., Ghosh, A.K., Denison, M.R., Mesecar, A.D. Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro): IMPLICATIONS FOR nsp5 REGULATION AND THE DEVELOPMENT OF ANTIVIRALS. // (2015) J Biol Chem 290: 19403-19422
75. St John, S.E., Tomar, S., Stauffer, S.R., Mesecar, A.D. Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4-The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS). // (2015) Bioorg Med Chem 23: 6036-6048
76. Ketan T. Savjani, Anuradha K. Gajjar, and Jignasa K. Savjani. Drug Solubility: Importance and Enhancement Techniques // Institute of Pharmacy, Nirma University, Sarkhej Gandhinagar Highway, Gujarat, Ahmedabad 382481, India
77. Daniel F Veber Stephen R Johnson Bristol-Myers Squibb Hung-Yuan Cheng Brian R Smith. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates // July 2002 Journal of Medicinal Chemistry 45(12):2615-23 DOI:10.1021/jm020017n SourcePubMed
78. SK Manirul Haque , Omar Ashwaq , Abdulla Sarief & Abdul Kalam Azad John Mohamed. A comprehensive review about SARS-CoV-2 // Published Online:14 Oct 2020 <https://doi.org/10.2217/fvl-2020-0124>

## ДОДАТОК А

А) Результати докінгу сполук, отриманих із підгрупи ZINC15 fda. Об'єм підгрупи: 1657 сполук

№	ZINC ID	XP GScore	Характеристика речовини.
1	ZINC000060183170	-9.143	Паромоміцину сульфат Похідна речовина неоміцину (НУ-В0470), є аміноглікозидним антибіотиком широкого спектру дії з амебіцидним та бактерицидним ефектом
2	ZINC000003918087	-8.122	Доксорубіцин Протипухлинний антибіотик. Хіміотерапевтичний препарат, що застосовується для лікування раку. Доксорубіцин вводять шляхом ін'єкції у вену.
3	ZINC000003830960	-7.746	Йопромід Йодовмістне контрастне середовище для рентгенівських знімків. Продається під назвою Ultravist, виробляється Bayer Healthcare. Низькоосмолярна, неіонна контрастна речовина для внутрішньосудинного застосування.
4	ZINC000003830943	-7.562	Іогексол Продається під торговою назвою Omnipaque, є контрастною речовиною, що використовується у рентгенівських знімках для візуалізації артерій, вен, шлуночків мозку, сечовидільної системи та суглобів, а також під час комп'ютерної томографії (КТ).
5	ZINC000002036848	-6.982	Рибофлавін Рибофлавін, також відомий як вітамін В2, - це вітамін, який міститься в їжі і використовується як харчова добавка.
6	ZINC000242437511	-7.107	Гентаміцин Гентаміцин, також відомий як g-мітицин або гараміцин, належить до класу органічних сполук, відомих як аміноциклітолові глікозиди. Це органічні сполуки, що містять

			амікоциклітовий фрагмент, глікозидно пов'язаний з вуглеводним фрагментом. Гентаміцин - препарат, який використовується для лікування інфекцій, спричинених такими мікроорганізмами: <i>s. aeruginosa</i> , <i>e. coli</i> , види клебсієла-ентеробактер-серратія, деякі штами цитробактерів та стафілококів (коагулазопозитивні та коагулазонегативні). Гентаміцин є дуже сильною основою.
7	ZINC000242548690	-6.92	Дигоксин Глікозид карденоліду, є β-гідроксильованим дигітоксином при C-12. Серцевий глікозид, отриманий з рослини наперстянки, <i>Digitalis lanata</i> , використовується для контролю частоти скорочення шлуночків при фібриляції передсердь та для лікування застійної серцевої недостатності при фібриляції передсердь.
8	ZINC000013585233	-6.846	Сапроптерин Сапроптерин дигідрохлорид - це синтетична форма кофактора ВН4.

Б) Результати докінгу сполук отриманих із підгрупи ZINC15 world. Об'єм підгрупи: 7136 сполук

№	ZINC ID	XP GScore	Характеристика речовини.
1	ZINC000255965156	-7.459	Гінзенозид Rg1 Гінсенозиди - це клас стероїдних глікозидів та тритерпенових сапонінів, що містяться виключно у рослині роду <i>Panax</i> (женьшень). Rg1, як видається, найпоширеніший у женьшені <i>Panax</i> (китайський / корейський женьшень).
2	ZINC000003860156	-6.29	adenosine 5'-monophosphate також відомий як 5'-аденилова кислота, є нуклеотидом. АМФ складається з фосфатної групи, цукрової рибози та нуклеоснови аденина; це ефір фосфорної кислоти та нуклеозиду аденозину

3	ZINC000003872805	-6.054	Гексопrenalін Гексопrenalін є селективним агоністом $\beta$ 2-адренергічних рецепторів, що використовується при лікуванні астми.
4	ZINC000034108591	-6.046	Мікофенольна кислота О-ацил-глюкуронід Складний ефір карбонової кислоти, що є результатом формальної конденсації групи карбонової кислоти мікофенольної кислоти з аномерною гідроксигрупою $\beta$ -D-глюкуронової кислоти.
5	ZINC000095618592	-5.352	Глюкуронід мікофенольної кислоти Глюкуронід мікофенольної кислоти належить до класу органічних сполук, відомих як фенольні глікозиди. Це органічні сполуки, що містять фенольну структуру, прикріплену до глікозильної частини. Деякі приклади фенольних структур включають лігнани та флавоноїди. Серед одиниць цукру, що містяться в природних глікозидах, є D-глюкоза, L-фруктоза та L рамноза.

## ДОДАТОК Б

Речовини із підгрупи ZINC15 anodyne, що проявили найкращі показники зв'язування з димеризаційним сайтом та з активним центром Mpro протеази

№	ZINC ID речовини	Хімічна формула речовини	XP GScore акт. центр (Kcal/mol)	XP GScore дим. сайт (Kcal/mol)
1	ZINC000003 922065	C33H49N9O8	-11.891	-12.444
2	ZINC000008 551612	C20H31N4O17P	-8.260	-11.759

3	ZINC000009 574481	C33H54N10O13	-11.954	-11.729
4	ZINC000004 533632	C33H47N9O7	-15.141	-11.653
5	ZINC000003 918259	C27H44N6O7	-11.482	-11.638
6	ZINC000005 412491	C29H42N6O10	-7.196	-11.561
7	ZINC000003 920109	C25H42N8O7	-11.574	-11.283
8	ZINC000008 551975	C17H27N5O16P2	-9.525	-11.128
9	ZINC000003 914747	C22H31N7O10	-10.564	-11.036
10	ZINC000005 706429	C31H37N3O11	-14.160	-11.035

### ДОДАТОК В

Результати докінгу сполук, отриманих із підгрупи ZINC15 in-man. Об'єм підгрупи: 1111777 сполук

№	XP GScore	ZINC ID	Характеристика речовини. Джерела у харчових продуктах
1	-10.998	ZINC0000	1,4,6-тригалоїл-бета-D-глюкопіраноза належить до

		26488096	класу органічних сполук, відомих як таніни. 1,4,6-тригалоїл-бета-D-глюкопіраноза міститься в найвищих концентраціях в чаї , але також була виявлена плодах гранатового дерева ( <i>Punica granatum</i> ).
2	-10.964	ZINC0001 00825227	Кверцетин 3- (3-р-кумароїлглюкозид) належить до класу органічних сполук, відомих як флавоноїд-3-о-глікозиди. Це фенольні сполуки, що містять флавоноїдну частину, О-глікозидно пов'язану з вуглеводною частиною в положенні С3.
3	-10.706	ZINC9360 70151	Кроцин належить до класу органічних сполук, відомих як дитерпеноїди. Це терпенові сполуки, утворені чотирма одиницями ізопрену
4	-10.636	ZINC1008 23369	2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразин належить до класу органічних сполук, відомих як піразини. Речовину було виявлено у горіхах.
5	-10.198	ZINC1008 28926	Фагопіритол А2 належить до класу органічних сполук, відомих як о-глікозилові сполуки. Поза людським тілом фагопіритол А2 був виявлений, але не кількісно визначений у зернових продуктах.
6	-10.101	ZINC0000 85815081	Кверцетин 3-О- (6 "-малоніл-глюкозид) 7-О-глюкозид належить до класу органічних сполук, відомих як флавоноїди-7-о-глікозиди. Кверцетин 3-О- (6 "-малоніл-глюкозид) 7-О-глюкозид виявляється в середньому в найбільшій концентрації в листі салату ( <i>Lactuca sativa</i> ). Кверцетин 3-О- (6" -малоніл-глюкозид) 7-О-глюкозид також був виявлений в ендівії ( <i>Cichorium endivia</i> )
7	-9.894	ZINC1008 26065	Катехін 3 ', 5-диглюкозид належить до класу органічних сполук, відомих як флавоноїдні о-глікозиди. Катехін 3 ', 5-диглюкозид - надзвичайно слабка основна (по суті нейтральна) сполука (на основі його рК <sub>a</sub> ). Речовину виявлено у зелених

			овочах.
8	-9.870	ZINC1008 23327	Актеозид, також відомий як кусагінін або вербаскозид, належить до класу органічних сполук, відомих як кумарові кислоти та їх похідні. Це ароматичні сполуки, що містять частину коричної кислоти (cinnamic acid) (або її похідне), гідроксильовану в атомі вуглецю С2 (орто-), С3 (мета-) або С4 (пара-) бензольного кільця. В середньому актеозид виявляється у найвищій концентрації в кількох різних продуктах харчування, таких як вербена звичайна, вербена лимонна та оливки. Актеозид також був виявлений у гірчичках, фруктах та кунжуті. Похідні вербаскозиду можна знайти у <i>Verbascum undulatum</i> і особливо апіозиди у <i>Verbascum sp.</i> Вербаскозид можна зустріти у видів у всіх сімействах порядку Lamiales (син. Scrophulariales).
9	-9.668	ZINC0001 00828363	альфа-D-Ксилопіранозил- (1-> 6) -бета-D-глюкопіранозил- (1-> 4) -D-глюкоза належить до класу органічних сполук, відомих як олігосахариди. альфа-D-ксилопіранозил- (1-> 6) -бета-D-глюкопіранозил- (1-> 4) -D-глюкоза була виявлена у зернових продуктах та фруктах.
10	-9.662	ZINC0001 00825216	Кверцетин 3,4'-диглюкозид належить до класу органічних сполук, відомих як флавоноїд-3-о-глікозиди. Кверцетин 3,4'-диглюкозид міститься в середньому в найвищих концентраціях в цибулі-шалот ( <i>Allium sera</i> ) та садовій цибулі ( <i>Allium schoenoprasum</i> )
11	-9.638	ZINC0001 00825670	Кверцетин 3-галактозид 7-рамнозид належить до класу органічних сполук, відомих як флавоноїди-7-о-глікозиди. Кверцетин 3-галактозид 7-рамнозид був виявлений у кількох різних продуктах харчування, таких як каперси, листя рослин роду <i>Tiliaceae</i> , особливо срібної липи ( <i>Tilia tomentosa</i> ).

12	-9.607	ZINC0001 00825341	3,6-диглюкопіранозил-5,7-дигідрокси-4'-метоксифлавоон, також відомий як 3,6-ді-С-глюкопіранозилацетин, належить до класу органічних сполук, відомих як флавоноїдні С-глікозиди. Флавоноїдні С-глікозиди - це сполуки, що містять вуглеводну частину, яка 3-глікозидно пов'язана з 2-фенілхромен-4-он-флавоноїдом. 3,6-диглюкопіранозил-5,7-дигідрокси-4'-метоксифлавоон було виявлено у фруктах.
13	-9.606	ZINC0001 00829483	Піцеатанол 4'-галлоїлглюкозид належить до класу органічних сполук, відомих як стилбенові глікозиди. Стілбенові глікозиди - це сполуки, що структурно характеризуються наявністю вуглеводного фрагменту, глікозидно пов'язаного зі стильбеновою основою. 4'-галлоїлглюкозид піцеатанолу був виявлений у зелених овочах.
14	-9.555	ZINC0001 00820603	b-D-галактопіранозил- (1-> 6) -b-D-галактопіранозил- (1-> 6) -D-галактоза належить до класу органічних сполук, відомих як олігосахариди. Це вуглеводи, що складаються з 3 - 10 моносахаридних одиниць, пов'язаних між собою за допомогою глікозидних зв'язків. Сполука міститься у фруктах
15	-9.497	ZINC0000 85859005	Ізоларицирезинол 4'-О-бета-D-глюкозид належить до класу органічних сполук, відомих як глікозиди лігнану. Це ароматичні поліциклічні сполуки, що містять вуглеводний компонент, глікозидно пов'язаний з лігнановим фрагментом. Вони включають 1-арилтетралінові лактони. Ізоларицирезинол 4'-О-бета-D-глюкозид було виявлено в алкогольних напоях.

## ДОДАТОК Г

### Визначення ADMET властивостей лігандів

#### А) Оцінка біодоступності лігандів

№	ZINC ID речовини	Порушення Ліпінськи	правила	Пероральна бідоступність EGAN
1	ZINC000003922065	3		Низька
2	ZINC000008551612	3		Низька
3	ZINC000009574481	2		Низька
4	ZINC000004533632	2		Низька
5	ZINC000003918259	3		Низька
6	ZINC000005412491	3		Низька
7	ZINC000003920109	3		Низька
8	ZINC000008551975	3		Низька
9	ZINC000003914747	3		Низька
10	ZINC000005706429	2		Низька
11	ZINC000008220119	2		Низька
12	ZINC000004726429	3		Низька
13	ZINC000008437048	3		Низька
14	ZINC000005706432	3		Висока
15	ZINC000003916647	3		Низька
16	ZINC000003914746	3		Висока
17	ZINC000008218858	3		Висока
18	ZINC000004748476	3		Висока
19	ZINC000003918256	3		Низька
20	ZINC000008220483	3		Низька
21	ZINC000012503837	3		Низька
22	ZINC000004915903	3		Низька

23	ZINC000005181312	3	Низька
24	ZINC000003916648	3	Висока
25	ZINC000008216330	3	Низька
26	ZINC000009331428	3	Низька
27	ZINC000008830558	3	Висока
28	ZINC000005273784	0	Низька
29	ZINC000004726430	3	Низька
30	ZINC000008217893	3	Низька
31	ZINC000026488096	3	Висока
32	ZINC000100825227	3	Висока
33	ZINC936070151	3	Низька
34	ZINC100823369	1	Висока
35	ZINC100828926	3	Висока
36	ZINC000085815081	3	Низька
37	ZINC100826065	3	Висока
38	ZINC100823327	3	Низька
39	ZINC000100828363	2	Висока
40		3	Висока
41	ZINC000100825216		Висока
42	ZINC000100825670	3	Висока
43	ZINC000100825341	3	Висока
44	ZINC000100829483	3	Висока
45	ZINC000100820603	3	Висока
46	ZINC000085859005	3	Висока

## Б) Оцінка токсичності лігандів

№	ZINC ID речовини	Фосфоліпідоз	Правило GSK 4/400	Оцінка 3_75
1	ZINC000003922065	Не індукує	Хороший бал	Хороша
2	ZINC000008551612	Не індукує	Хороший бал	Хороша
3	ZINC000009574481	Не індукує	Хороший бал	Хороша
4	ZINC000004533632	Не індукує	Хороший бал	Хороша
5	ZINC000003918259	Не індукує	Хороший бал	Хороша
6	ZINC000005412491	Не індукує	Хороший бал	Хороша
7	ZINC000003920109	Не індукує	Хороший бал	Хороша
8	ZINC000008551975	Не індукує	Хороший бал	Хороша
9	ZINC000003914747	Не індукує	Хороший бал	Хороша
10	ZINC000005706429	Не індукує	Хороший бал	Хороша
11	ZINC000008220119	Не індукує	Хороший бал	Хороша
12	ZINC000004726429	Не індукує	Хороший бал	Хороша
13	ZINC000008437048	Не індукує	Хороший бал	Хороша
14	ZINC000005706432	Не індукує	Хороший бал	Хороша
15	ZINC000003916647	Не індукує	Хороший бал	Хороша
16	ZINC000003914746	Індукує	Хороший бал	Хороша
17	ZINC000008218858	Індукує	Хороший бал	Хороша
18	ZINC000004748476	Індукує	Хороший бал	Хороша
19	ZINC000003918256	Не індукує	Хороший бал	Хороша
20	ZINC000008220483	Індукує	Хороший бал	Хороша

21	ZINC000012503837	Не індукує	Хороший бал	Хороша
22	ZINC000004915903	Не індукує	Хороший бал	Хороша
23	ZINC000005181312	Не індукує	Хороший бал	Хороша
24	ZINC000003916648	Не індукує	Хороший бал	Хороша
25	ZINC000008216330	Не індукує	Хороший бал	Хороша
26	ZINC000009331428	Не індукує	Хороший бал	Хороша
27	ZINC000008830558	Індукує	Хороший бал	Хороша
28	ZINC000005273784	Не індукує	Хороший бал	Хороша
29	ZINC000004726430	Індукує	Хороший бал	Хороша
30	ZINC000008217893	Не індукує	Хороший бал	Хороша
31	ZINC000026488096	Не індукує	Хороший бал	Хороша
32	ZINC000100825227	Не індукує	Хороший бал	Хороша
33	ZINC936070151	Не індукує	Хороший бал	Хороша
34	ZINC100823369	Не індукує	Хороший бал	Хороша
35	ZINC100828926	Не індукує	Хороший бал	Хороша
36	ZINC000085815081	Не індукує	Хороший бал	Хороша
37	ZINC100826065	Не індукує	Хороший бал	Хороша
38	ZINC100823327	Не індукує	Хороший бал	Хороша
39	ZINC000100828363	Не індукує	Хороший бал	Хороша
40		Не індукує	Хороший бал	Хороша
41	ZINC000100825216	Не індукує	Хороший бал	Хороша
42	ZINC000100825670	Не індукує	Хороший бал	Хороша
43	ZINC000100825341	Не індукує	Хороший бал	Хороша
44	ZINC000100829483	Не індукує	Хороший бал	Хороша

45	ZINC000100820603	Не індукує	Хороший бал	Хороша
46	ZINC000085859005	Не індукує	Хороший бал	Хороша

В) Поглиблене вивчення токсичності та біодоступності ліганда 2- (1,2,3,4-тетрагідроксибутил) -6- (2,3,4-тригідроксибутил) піразину

Характеристика	Назва моделі	Передбачуване значення	Міра вимірювання
Поглинання	Розчинність у воді	-1.619	log mol/L
Поглинання	Caco2 проникність	-0.463	log Papp in 10 <sup>6</sup> cm/s
Поглинання	Поглинання у кишківнику (людина)	20.986	% поглинуто
Поглинання	P-глікопротеїн субстрат	Так	Так/Ні
Поглинання	P-глікопротеїн I інгібітор	Ні	Так/Ні
Поглинання	P-глікопротеїн II інгібітор	Ні	Так/Ні
Розповсюдження	VDss (Об'єм розповсюдження у сталому середовищі)	-0.841	log L/kg
Розповсюдження	Незв'язана фракція (людина)	0.585	Fu
Розповсюдження	BBB проникність	-1.787	log BB
Розповсюдження	CNS проникність	-5.837	log PS
Метаболізм	CYP2D6 субстрат	Ні	Так/Ні
Метаболізм	CYP3A4 субстрат	Ні	Так/Ні
Метаболізм	CYP1A2 інгібітор	Ні	Так/Ні
Метаболізм	CYP2C19 інгібітор	Ні	Так/Ні
Метаболізм	CYP2C9 інгібітор	Ні	Так/Ні
Метаболізм	CYP2D6 інгібітор	Ні	Так/Ні
Метаболізм	CYP3A4 інгібітор	Ні	Так/Ні
Екскреція	Повне виведення	0.981	log ml/хвил./kg
Екскреція	Renal OCT2	Ні	Так/Ні

	субстрат		
<b>Токсичність</b>	AMES токсичність	Ні	Так/Ні
<b>Токсичність</b>	Максимальна доза (людина)	0.855	log mg/kg/день
	hERG I інгібітор	Ні	Так/Ні
<b>Токсичність</b>	hERG II інгібітор	Ні	Так/Ні
<b>Токсичність</b>	Гостра токсичність у пацієнтів (LD50)	2.816	mol/kg
<b>Токсичність</b>	Хронічна токсичність у пацієнтів (LOAEL)	4.006	log mg/kg_bw/день
<b>Токсичність</b>	Гепатотоксичність	Так	Так/Ні

Г) Поглиблене вивчення токсичності та біодоступності ліганда кверцетину 3- (3-р-кумароїлглюкозиду)

<b>Характеристика</b>	<b>Назва моделі</b>	<b>Передбачуване значення</b>	<b>Міра вимірювання</b>
<b>Поглинання</b>	Розчинність у воді	-2.988	log mol/L
<b>Поглинання</b>	Caco2 проникність	0.085	log Papp in 10 <sup>6</sup> cm/s
<b>Поглинання</b>	Поглинання у кишківнику (людина)	50.497	% поглинуто
<b>Поглинання</b>	P-глікопротеїн субстрат	Так	Так/Ні
<b>Поглинання</b>	P-глікопротеїн I інгібітор	Так	Так/Ні
<b>Поглинання</b>	P-глікопротеїн II інгібітор	Так	Так/Ні
<b>Розповсюдження</b>	VD <sub>ss</sub> (Об'єм розповсюдження у сталому середовищі)	-0.384	log L/kg
<b>Розповсюдження</b>	Незв'язана фракція (людина)	0.144	Fu
<b>Розповсюдження</b>	BBB проникність	-2.368	log BB
<b>Розповсюдження</b>	CNS проникність	-4.733	log PS
<b>Метаболізм</b>	CYP2D6 субстрат	Ні	Так/Ні
<b>Метаболізм</b>	CYP3A4 субстрат	Ні	Так/Ні

<b>Метаболізм</b>	CYP1A2 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2C19 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2C9 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2D6 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP3A4 інгібітор	Ні	Так/Ні
<b>Екскреція</b>	Повне виведення	-0.19	log ml/хвил./kg
<b>Екскреція</b>	Renal OСТ2 субстрат	Ні	Так/Ні
<b>Токсичність</b>	AMES токсичність	Ні	Так/Ні
<b>Токсичність</b>	Максимальна доза (людина)	0.612	log mg/kg/день
	hERG I інгібітор	Ні	Так/Ні
<b>Токсичність</b>	hERG II інгібітор	Так	Так/Ні
<b>Токсичність</b>	Гостра токсичність у пацюків (LD50)	2.685	mol/kg
<b>Токсичність</b>	Хронічна токсичність у пацюків (LOAEL)	3.899	log mg/kg_bw/день
<b>Токсичність</b>	Гепатотоксичність	Ні	Так/Ні

Д) Поглиблене вивчення токсичності та біодоступності ліганда кроцину

<b>Характеристика</b>	<b>Назва моделі</b>	<b>Передбачуване значення</b>	<b>Міра вимірювання</b>
<b>Поглинання</b>	Розчинність у воді	-2.862	log mol/L
<b>Поглинання</b>	Са <sub>сo</sub> 2 проникність	-1.345	log P <sub>app</sub> in 10 <sup>6</sup> cm/s
<b>Поглинання</b>	Поглинання у кишково- кишковнику (людина)	0	% поглинуто
<b>Поглинання</b>	P-глікопротеїн субстрат	Так	Так/Ні
<b>Поглинання</b>	P-глікопротеїн I інгібітор	Ні	Так/Ні
<b>Поглинання</b>	P-глікопротеїн II інгібітор	Ні	Так/Ні
<b>Розповсюдження</b>	VD <sub>ss</sub> (Об'єм розповсюдження у сталому середовищі)	-0.149	log L/kg

<b>Розповсюдження</b>	Незв'язана фракція (людина)	0.381	Fu
<b>Розповсюдження</b>	BBB проникність	-2.673	log BB
<b>Розповсюдження</b>	CNS проникність	-6.992	log PS
<b>Метаболізм</b>	CYP2D6 субстрат	Ні	Так/Ні
<b>Метаболізм</b>	CYP3A4 субстрат	Ні	Так/Ні
<b>Метаболізм</b>	CYP1A2 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2C19 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2C9 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2D6 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP3A4 інгібітор	Ні	Так/Ні
<b>Екскреція</b>	Повне виведення	1.78	log ml/хвил./kg
<b>Екскреція</b>	Renal OCT2 субстрат	Так	Так/Ні
<b>Токсичність</b>	AMES токсичність	Ні	Так/Ні
<b>Токсичність</b>	Максимальна доза (людина)	0.412	log mg/kg/день
	hERG I інгібітор	Ні	Так/Ні
<b>Токсичність</b>	hERG II інгібітор	Так	Так/Ні
<b>Токсичність</b>	Гостра токсичність у пацюків (LD50)	2.477	mol/kg
<b>Токсичність</b>	Хронічна токсичність у пацюків (LOAEL)	9.013	log mg/kg_bw/день
<b>Токсичність</b>	Гепатотоксичність	Ні	Так/Ні

Є) Поглиблене вивчення токсичності та біодоступності ліганда фагопіритулу А2

<b>Характеристика</b>	<b>Назва моделі</b>	<b>Передбачуване значення</b>	<b>Міра вимірювання</b>
<b>Поглинання</b>	Розчинність у воді	-1.803	log mol/L
<b>Поглинання</b>	Сасо2 проникність	-0.804	log Papp in 10 <sup>6</sup> cm/s
<b>Поглинання</b>	Поглинання у кишківнику (людина)	0	% поглинуто
<b>Поглинання</b>	P-глікопротеїн субстрат	Так	Так/Ні

<b>Поглинання</b>	Р-глікопротеїн I інгібітор	Ні	Так/Ні
<b>Поглинання</b>	Р-глікопротеїн II інгібітор	Ні	Так/Ні
<b>Розповсюдження</b>	VD <sub>ss</sub> (Об'єм розповсюдження у сталому середовищі)	-0.382	log L/kg
<b>Розповсюдження</b>	Незв'язана фракція (людина)	0.408	F <sub>u</sub>
<b>Розповсюдження</b>	BBB проникність	-1.937	log BB
<b>Розповсюдження</b>	CNS проникність	-9.055	log PS
<b>Метаболізм</b>	CYP2D6 субстрат	Ні	Так/Ні
<b>Метаболізм</b>	CYP3A4 субстрат	Ні	Так/Ні
<b>Метаболізм</b>	CYP1A2 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2C19 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2C9 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP2D6 інгібітор	Ні	Так/Ні
<b>Метаболізм</b>	CYP3A4 інгібітор	Ні	Так/Ні
<b>Екскреція</b>	Повне виведення	1.67	log ml/хвил./kg
<b>Екскреція</b>	Renal OAT2 субстрат	Ні	Так/Ні
<b>Токсичність</b>	AMES токсичність	Ні	Так/Ні
<b>Токсичність</b>	Максимальна доза (людина)	0.47	log mg/kg/день
<b>Токсичність</b>	hERG I інгібітор	Ні	Так/Ні
<b>Токсичність</b>	hERG II інгібітор	Так	Так/Ні
<b>Токсичність</b>	Гостра токсичність у пацюків (LD <sub>50</sub> )	2.542	mol/kg
<b>Токсичність</b>	Хронічна токсичність у пацюків (LOAEL)	7.275	log mg/kg_bw/день
<b>Токсичність</b>	Гепатотоксичність	Ні	Так/Ні

## Ж) Поглиблене вивчення токсичності та біодоступності ліганда ZINC5273784

Характеристика	Назва моделі	Передбачуване значення	Міра вимірювання
Поглинання	Розчинність у воді	-2.883	log mol/L
Поглинання	CaCo2 проникність	-0.536	log Papp in 10 <sup>6</sup> cm/s
Поглинання	Поглинання у кишківнику (людина)	0	% поглинуто
Поглинання	P-глікопротеїн субстрат	Так	Так/Ні
Поглинання	P-глікопротеїн I інгібітор	Ні	Так/Ні
Поглинання	P-глікопротеїн II інгібітор	Ні	Так/Ні
Розповсюдження	VD <sub>ss</sub> (Об'єм розповсюдження у сталому середовищі)	-1.735	log L/kg
Розповсюдження	Незв'язана фракція (людина)	0.71	Fu
Розповсюдження	BBB проникність	-1.727	log BB
Розповсюдження	CNS проникність	-4.705	log PS
Метаболізм	CYP2D6 субстрат	Ні	Так/Ні
Метаболізм	CYP3A4 субстрат	Ні	Так/Ні
Метаболізм	CYP1A2 інгібітор	Ні	Так/Ні
Метаболізм	CYP2C19 інгібітор	Ні	Так/Ні
Метаболізм	CYP2C9 інгібітор	Ні	Так/Ні
Метаболізм	CYP2D6 інгібітор	Ні	Так/Ні
Метаболізм	CYP3A4 інгібітор	Ні	Так/Ні
Екскреція	Повне виведення	-0.097	log ml/хвил./kg
Екскреція	Renal OСТ2 субстрат	Ні	Так/Ні
Токсичність	AMES токсичність	Ні	Так/Ні
Токсичність	Максимальна доза (людина)	1.329	log mg/kg/день
	hERG I інгібітор	Ні	Так/Ні
Токсичність	hERG II інгібітор	Ні	Так/Ні
Токсичність	Гостра	2.532	mol/kg

	токсичність у пацюків (LD50)		
<b>Токсичність</b>	Хронічна токсичність у пацюків (LOAEL)	4.218	log mg/kg_bw/день
<b>Токсичність</b>	Гепатотоксиність	Так	Так/Ні