

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Інститут високих технологій

Завідувач кафедри супрамолекулярної хімії  
д.х.н., професор Ігор Володимирович Комаров  
Протокол №\_\_\_\_засідання кафедри  
Від «\_\_»\_\_\_\_\_2021\_року

**ВІРТУАЛЬНИЙ СКРИНІНГ СПОЛУК ДЛЯ ПОШУКУ СЕЛЕКТИВНИХ  
ІНГІБІТОРІВ АНГІОТЕНЗИН-ПЕРЕТВОРЮВАЛЬНОГО ФЕРМЕНТУ II**

Випускна кваліфікаційна робота магістра  
денної форми навчання за спеціальністю  
102 «Хімія (Хемоінформатика)»  
**Кравець Ірини Олександрівни**

Науковий керівник від кафедри  
д. х. н. **Волочнюк Д.М.**

Робота виконана у ТОВ «Кемспейс»

Оцінка захисту роботи  
\_\_\_\_\_

**Київ – 2021 р.**

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ACE2 / АПФ2 – ангіотензин перетворювальний фермент 2

ACE / АПФ – ангіотензин перетворювальний фермент

SARS-CoV-2 - гострий респіраторний синдромом коронавірусу-2 (COVID-19)

ГРВІ – гостра респіраторна вірусна інфекція

TMPRSS2 – Transmembrane protease serine 2

RAS – ренін-аніотензинова система

Ang1 (2) – ангіотензин 1 (2)

AT1R – рецептор ангіотензину 1

ADME - Absorption, Distribution, Metabolism, and Elimination

DMPK - Drug Metabolism and Pharmacokinetics

HTVS – High-Throughput Virtual Screening

SP - Standard Precision (стандартна точність)

XP – Extra Precision (висока точність)

RDB - REAL DATA BASE

ВООЗ - Всесвітня організація охорони здоров'я

PDB – Protein Data Base

IC50 - inhibitory concentration with 50%

pKa - the negative log of the acid dissociation constant or Ka value

Ki – константа інгібування

Asp415 – залишок аспарагіну порядковим номером 415 у протеїні

Glu – глютамінова кислота

Lys – лізин

His – гістидин

Arg – аргінін

Gln – глютамін

Tyr – тирозин

Val – валін

Asn – аспарагін

Ser – серин

Thr – триптофан

Cys – цистеїн

Pro – пролін

(R273Q) - заміна аргініну залишком глутаміну

(R273K) - заміна аргініну залишком лізину

PAINS - Pan-Assay Interference Compounds

RMSD - середньо-квадратичне відхилення

## **ЗМІСТ**

|  |           |
|--|-----------|
| <b>АНОТАЦІЯ.....</b>   | <b>5</b>  |
| <b>ВСТУП.....</b>  | <b>7</b>  |
| <b>РОЗДІЛ I. Огляд літератури</b>                                    |           |
| 1.1. Епідеміологічні дані ВООЗ про COVID-19.....                     | 10        |
| 1.2. Роль ACE2 в інфекції ГРВІ-CoV-2.....                            | 10        |
| 1.3. Літературні данні щодо сайту зв'язування ACE2.....              | 13        |
| <b>РОЗДІЛ II. Розробка докінгової моделі</b>                         |           |
| 2.1. Огляд сайту зв'язування мішені та відмінностей ACE2 та ACE..... | 16        |
| 2.2. Формування моделей для докінгу та їх верифікація.....           | 19        |
| <b>РОЗДІЛ III. Віртуальний скринінг</b>                              |           |
| 3.1. Віртуальний скринінг стокової бази Enamine.....                 | 30        |
| 3.2. Віртуальний скринінг REAL DATA BASE Enamine.....                | 39        |
| 3.3. Молекулярна динаміка.....                                       | 46        |
| <b>ВИСНОВКИ.....</b>   | <b>56</b> |
| <b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....</b>                               | <b>57</b> |

## АНОТАЦІЯ

Методи молекулярного стикування (докінгу) та молекулярної динаміки є одними з основних підходів молекулярного моделювання, їх широко застосовують в drug-дизайні та медичній хімії для пошуку сполук, що будуть взаємодіяти з молекулярною мішенню. Ангіотензин-перетворювальний фермент 2 (далі ACE2), є рецептором, необхідним для проникнення COVID-19 в клітину. Роль пригнічення каталітичної активності ACE2 малодосліджена з причини відсутності низькомолекулярних хімічних зондів для ACE2. Задача роботи - відібрати приблизно 20 нових потенційних зондів на основі віртуального скринінгу *in silico* комерційно доступних колекцій хімічних сполук та 20 потенційних зондів з бази на замовлення за допомогою методів молекулярного докінгу та динаміки для подальшого дослідження ролі ACE2 за допомогою хіміко-генетичного підходу шляхом хімічно індукованого пригнічення каталітичної активності, але не в рамках цієї роботи.

Методы молекулярной стыковки (докинга) и молекулярной динамики являются одними из основных подходов молекулярного моделирования, их широко применяют в драг-дизайне и медицинской химии для поиска соединений, которые будут взаимодействовать с молекулярной мишенью. Ангиотензин-превращающий фермент 2 (далее ACE2) – рецептор, необходимый для проникновения COVID-19 в клетку. Роль ингибирования каталитической активности ACE2 малоисследована по причинам отсутствия низкомолекулярных химических зондов для ACE2. Задача работы – отобрать приблизительно 20 новых потенциальных зондов на основании виртуального скрининга *in silico* коммерчески доступных коллекций химических соединений и дополнительно 20 потенциальных зондов из базы «под заказ» с помощью методов молекулярного докинга и динамики для дальнейшего исследования роли ACE2 с помощью

химико-генетического подхода путем химически индуцированного ингибирования каталитической активности, но не в рамках этой работы.

Molecular docking and molecular dynamics are some of the main approaches of molecular modeling; they are widely used in drug-design and medicinal chemistry in researching compounds that will interact with a molecular target. Angiotensin-converting enzyme 2 (hereinafter ACE2) is a receptor required for COVID-19 to enter the cell. The role of inhibition of the catalytic activity of ACE2 has been pore studied due to the lack of small molecule chemical probes for ACE2. The task of the work is to select approximately 20 new potential probes based on in silico virtual screening of commercially available collections of chemical compounds and an additional 20 potential probes from the database "on demand" using molecular docking and dynamics methods for further investigating the role of ACE2 using a chemical-genetic approach by chemically induced inhibition of catalytic activity, but not within the scope of this work.

## ВСТУП

### Актуальність роботи

Дані щодо використання хімічно-генетичних підходів у дослідженні ролі ACE2, шляхом прямого пригнічення каталітичної активності ACE2 є обмеженими. Це можна пояснити відсутністю низькомолекулярних хімічних зондів для ACE2. Єдиним комерційно доступним селективним інгібітором ACE2 є поліпептид DX600 (Ac-GDYSHCSPLRYPWKCTYPDPEGGGNH<sub>2</sub>), який має серйозні обмеження для використання *in vivo* через невідповідний профіль ADME/DMPK параметрів [8]

Отже, залишаються відкритими питання «Наскільки серйозної шкоди тканинам та органам завдає тимчасове пряме пригнічення активності ACE2, яке спостерігається під час COVID-19? Чи може воно призвести до летальних наслідків?»

Був проведений всебічний аналіз бази даних речовин з потенційною лікарською активністю за допомогою платформи Reaxys і виявили дев'ять потенційних хімічних зондів з високою селективністю до ACE2 vs ACE1 [4,5,6]. Але всі ці сполуки виявилися комерційно недоступними, що відомі лише в аналітичних (кілька мг) кількостях, а їх використання як хімічних зондів не підтверджене.

Ключові слова: ACE2, COVID-19, каталітична активність, хімічні зонди.

### Мета роботи:

Планується відібрати приблизно 20 нових потенційних зондів на основі віртуального скринінгу *in silico* комерційно доступних колекцій хімічних сполук та 20 потенційних зондів з бази на замовлення. Для подальшої цілі дослідження ролі ACE2 за допомогою хіміко-генетичного підходу шляхом хімічно індукованого пригнічення каталітичної активності ACE2 як фармакологічної

моделі побічних ефектів COVID-19 в нервовій, серцево-судинній системах, кишечнику та системі гемостазу, але не в рамках цієї роботи.

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

1. Зіставлення послідовностей та просторових структур активних центрів ферментів ACE та ACE2, виявлення ключових відмінностей для будування моделі докінгу, що дозволить відібрати сполуки, що мають селективність ACE2 проти ACE;
2. Побудова моделей «сітки» та характеристик для докінгу ACE2 та ACE;
3. Перевірка моделей п.2. за допомогою докінгу референтних селективних інгібіторів ACE2 та так званих decoys, визначення enrichment;
4. Докінг стокової бібліотеки лігандів Enamine Ltd. на валідованих моделях;
5. Докінг Real Data Base бібліотеки лігандів Enamine Ltd. на валідованих моделях;
6. Молекулярна динаміка структур з найвищими оцінками молекулярного стикування.

**Методи досліджень**, що будуть застосовані в роботі:

1. Робота з бібліотеками (фільтрування, розрахунок дескрипторів і т.ін.)
2. Генерування decoys
3. Молекулярний докінг
4. Молекулярна динаміка

**Застосоване програмне забезпечення:**

1. Зіставлення просторових структур, 3d-візуалізація – Chimera

2. Хемоінформатична частина - Робота з бібліотеками, фільтрування – Python in Jupiter, пакет RDKit
3. Генерування decoys - <http://dude.docking.org/generate>
4. In silico віртуальний скринінг та молекулярна динаміка – Maestro 2020 від Schrodinger

## **РОЗДІЛ I. Огляд літератури**

### **1.1. Епідеміологічні дані ВООЗ про COVID-19. Роль ACE2 в інфекції SARS-CoV-2**

11 березня 2020 року ВООЗ оголосив хворобу COVID-19, що викликана новим тяжким гострим респіраторним синдромом коронавірус-2 (SARS-CoV-2), глобальною пандемією [1]. З моменту виявлення у грудні 2019 року Global Cases - 141 459 551, Global Deaths - 3 020 990, захворювань Україні - 2 006 187, смертей в Україні - 41755 (на 19.04.2021) [2].

Пандемія COVID-19 спричинена важким гострим респіраторним синдромом коронавірусу 2 (SARS-CoV-2), який поділяє високу гомологію послідовності амінокислот з коронавірусом ГРВІ, що виник у 2002 році. [3]

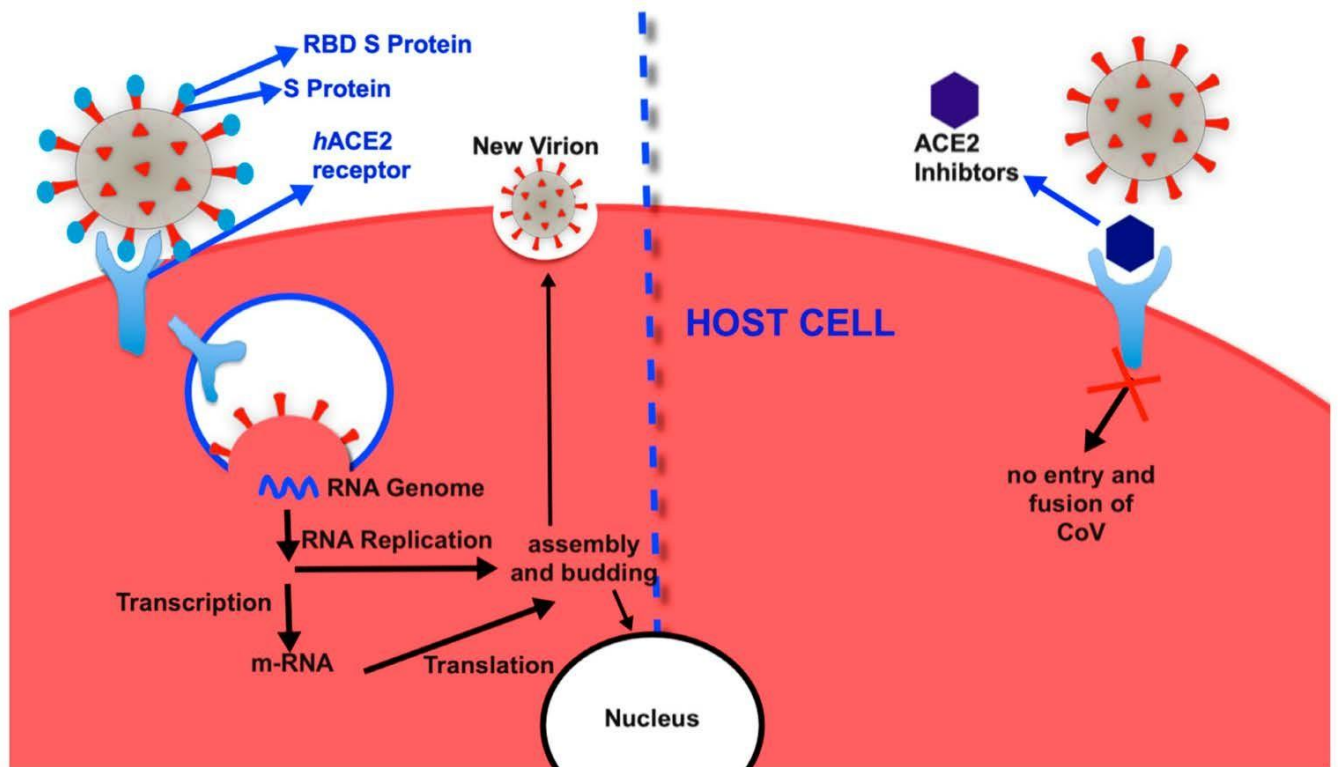
Поверхневий блок (S1) білку-шипа (S) SARS-CoV-2 використовує ACE2 як рецептор входу, а потім задіює серинову протеазу хазяїна TMPRSS2 для праймування S, що дозволяє злиття вірусних і клітинних мембран і потрапляння вірусу в клітину. [3]

### **1.2. Роль ACE2 в інфекції ГРВІ-CoV-2**

Ангіотензин-перетворювальний фермент 2 (далі ACE2), є рецептором, необхідним для проникнення COVID-19 в клітину. А з іншого боку, ангіотензин-перетворюючий фермент 2 (АПФ2), гомолог АПФ, являє собою нову і потенційно важливу мішень при серцево-ниркових захворюваннях[12].

Саме йому віддають переважне епідеміологічне значення щодо ризику важких форм перебігу захворювання на COVID-19 у пацієнтів із серцево-судинними захворюваннями та гіпертонічною хворобою. [3] На малюнку 3 показано життєвий цикл SARS-CoV. Білок (S) SARS-CoV зв'язується з рецептором ангіотензин-перетворюючого ферменту 2 (ACE2), щоб потрапити в клітини господаря і вивільнити геном РНК у клітини-мішені. Структурні та

неструктурні білки CoV та геном РНК збираються у віріони, які виділяються з клітин-мішеней.



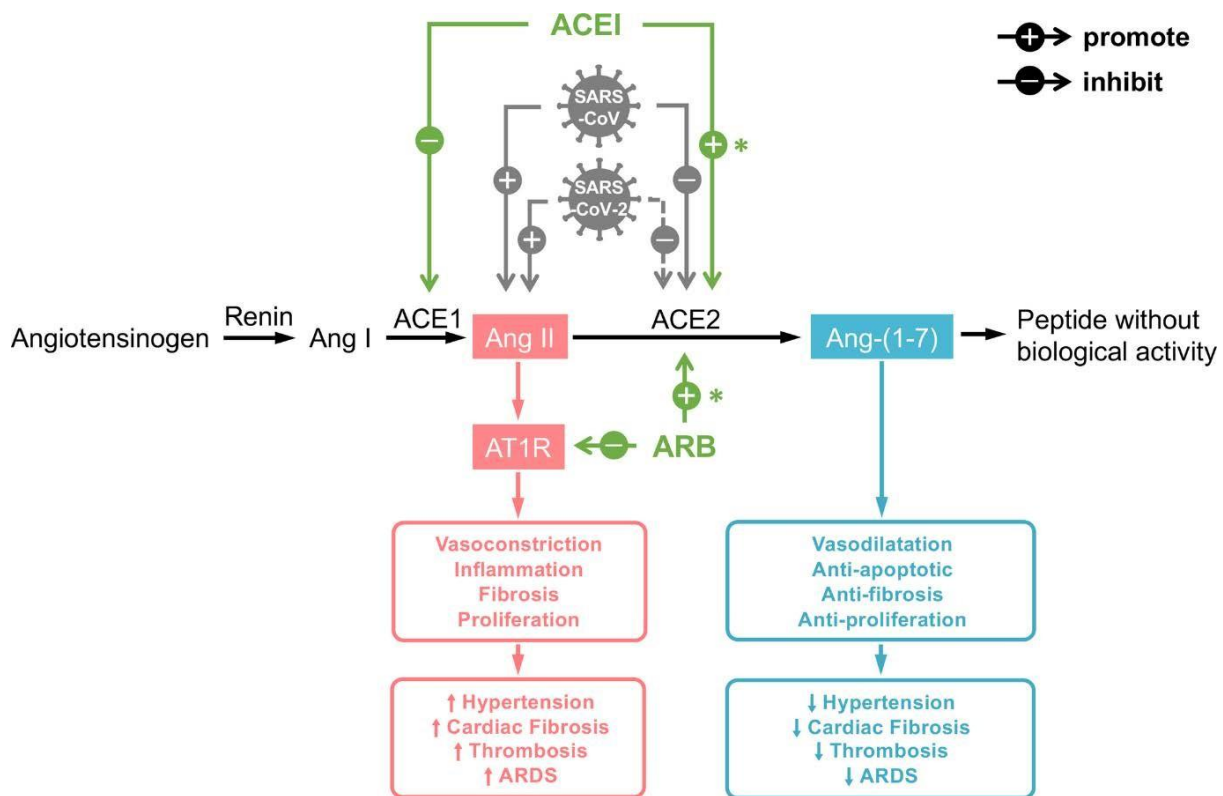
Малюнок 1. Життєвий цикл ГРВІ-CoV. Білок (S) SARS-CoV зв'язується з рецептором ангіотензинперетворюючого ферменту 2 (ACE2), щоб потрапити в клітини господаря і вивільнити геном РНК у клітини-мішені. Структурні та неструктурні білки CoV та геном РНК збираються у віріони, які виділяються з клітин-мішеней. Адаптовано з [25]

ACE2 - це металопротеаза, що функціонує як карбоксипептидаза. ACE2 складається з 805 амінокислотних залишків та має трансмембранний домен та одиночну HEXXH послідовність, що зв'язує цинк (положення амінокислотних залишків 374-378). Каталітичний домен ACE2 на 42% ідентичний домену ACE. [5]

ACE2 - це мембранний білок з ферментативним доменом, розташований на зовнішній поверхні клітин людини. Він був названий так, оскільки спочатку цей білок був ідентифікований як гомолог (або його варіант) ангіотензинперетворюючого ферменту (ACE), ферменту, який опосередковує утворення пептидного гормону, ангіотензину II з ангіотензину I. ACE був широко вивчений

і є добре відомим та проявляє судинозвужувальний ефект (тобто спричиняє скорочення м'язів у стінці судини та звуження просвіту судини). [7]

ACE2, який зараз відомий як вірусний рецептор, також діє як судинорозширювальний засіб, який врівноважує ACE і викликає розслаблення стінок судин. І ACE, і ACE2 є важливими гравцями в системі ренінгіотензинової системи (RAS), яка регулює кров'яний тиск і приплив крові до багатьох органів, включаючи легені, серце та нирки. RAS охоплює складну мережу ферментів, пептидних гормонів та рецепторів, як показано на малюнку 2.



Малюнок 2. Схематичне відображення регуляції ренінгіотензинової системи (RAS). [13]

Ангіотензиноген, попередник ангіотензину (Ang), який секретується печінкою, розщеплюється нирковим ферментом реніном для отримання ангіотензину I (Ang I). Потім Ang I перетворюється на Ang II за допомогою ACE. Ang II, гормональний пептид з восьми амінокислотами, зв'язується з рецепторами ангіотензину 1 типу (AT1R) на поверхні м'язових клітин у дрібних судинах, викликаючи звуження судин. Він також сприяє реабсорбції натрію

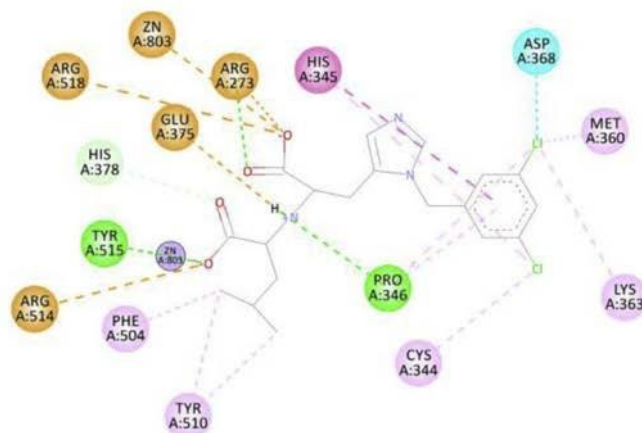
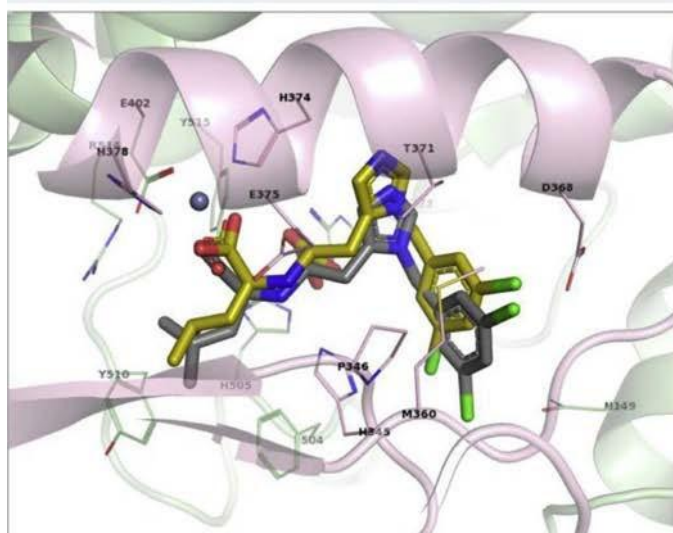
нирками. Як звуження судин, так і реабсорбція натрію призводять до підвищення артеріального тиску. Таким чином, аномально висока активність ACE призводить до підвищення рівня Ang II, що викликає гіпертонію. І навпаки, ACE2 каталізує перетворення пептиду восьми амінокислот Ang 2 в пептид семи амінокислот (Ang 1-7), який, здається, має протилежний ефект Ang II завдяки своїй дії на інший рецептор, званий Mas рецептор (MasR). Хоча точна роль Ang 1-7 у регуляції артеріального тиску не з'ясована до кінця, існують докази того, що він знижує артеріальний тиск та індукує розширення судин. Крім того, ACE2 розщеплює Ang I до Ang 1-9, і, отже, може додатково врівноважити ефект ACE, видаляючи його субстрат. Викликаючи перетворення Ang II в Ang (1-7) та Ang I в Ang 1-9, ACE2 може зіграти певну роль у підтримці балансу між звуженням судин та розширенням судин, щоб тримати артеріальний тиск у контролі. [9]

Узагальнені важливі дані щодо проникнення SARS-CoV-2 у клітину:

- S-білки SARS та SARS-2 опосередковують потрапляння вірусів у подібний спектр клітинних ліній;
- Як і SARS, SARS-CoV-2 використовує ACE2 у якості рецептора для проникнення в клітину;
- Серинові протеази TMPRSS2 праймують S білок SARS-CoV-2 для входу у клітину [3]

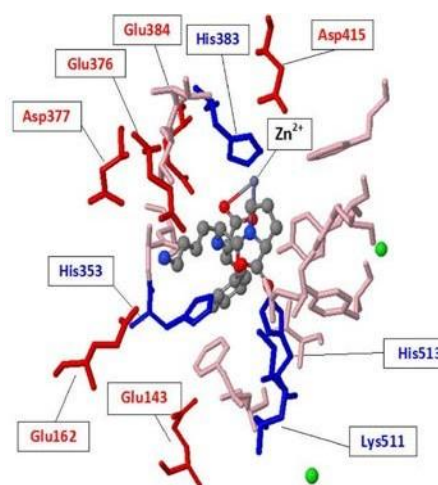
### **1.3. Літературні данні щодо сайту зв'язування ACE2**

Є дані щодо положення сайту зв'язування та хімічних зав'язків між амінокислотними залишками протеїну АЕС2 на прикладі комплексу ACE2 – MLN-4760 (PDB ID: 1R4L). Будучи потужним та селективним інгібітором ACE2 ( $IC_{50} < 1$  нМ), MLN-4760 утворює досить сильні спрямовані взаємодії



Малюнок 3. Ліворуч: Крупним планом мінімізований MLN-4760 (золотим), накладений на кристалографічно визначений комплекс ACE2 – MLN-4760 (PDB ID: 1R4L) (срібним). Праворуч: Двовимірний вигляд сприятливих взаємодій, що закріплюють кокрystalізований інгібітор в ділянці активного сайту ACE2 (PDB ID: 1R4L). Іонні взаємодії (сольові містки) та взаємодії заряду показані у вигляді помаранчевих пунктирних ліній. Водневі зв'язки показані у вигляді вапняно-зелених пунктирних ліній та у вигляді пунктирних ліній світло-бірюзового кольору (слабкі). Галогенні зв'язки показані у вигляді блакитних пунктирних ліній.  $\pi$  –  $\pi$ -взаємодії відображаються у вигляді пурпурових пунктирних ліній. [23,24]

Дані про активний сайт ACE наведені у літературі на прикладі взаємодії з широко використовуваним лікарським засобом – інгібітором ACE – лізиноприлом. На малюнках 3 та 4 зображено - лізиноприл у взаємозв'язку с ACE, пунктиром зазначені найважливіші зв'язки [26]



Малюнок 4. Активний сайт ACE у взаємозв'язку з лізиноприлом.

Карбоксильна група в'яжеться до активного центру молекули цинку. (Малюнок 4 ліворуч). Фенільне кільце лізиноприлу взаємодіє в підсайті S1. Лізин взаємодіє з S1'-підсайтом. Пролін взаємодіє з S2'-підсайтом. Лізиноприл утворює водневі та іонні зв'язки взаємодії з негативно зарядженими залишками амінокислот (Asp415, Glu162, Asp377, Glu384, Glu143, Glu376). Лізиноприл також взаємодіє з позитивно зарядженими залишками амінокислот через водневі зв'язки, іонні взаємодії або взаємодії з ентропічним сприянням (Lys511, His513, His353, His383). (Малюнок 4 праворуч) [26]

## РОЗДІЛ II. Розробка докінгової моделі

### 2.1. Огляд сайту зв'язування мішені та відмінностей ACE2 та ACE

Існують структурні відмінності між активним центром ACE (дипептидилкарбоксіпептидази) та ACE2 (карбоксіпептидази), які відповідають за відмінності в специфічності[10].

ACE2 має ряд характеристик з ACE, обидва є цинквмісними ферментами, які є чутливими до активації аніонів [4,17,18]. Однак, на відміну від ACE, ACE2 функціонує як карбоксіпептидаза і не сприйнятливий до інгібування класичними інгібіторами ACE [10].

Було проведено огляд активного сайту таргету та накладання (з вирівнюванням) молекулярних структур ACE2 (PDB код: 1R4L) та ACE (PDB код: 6F9T)

Ключові відмінності, що враховані для селективності це:

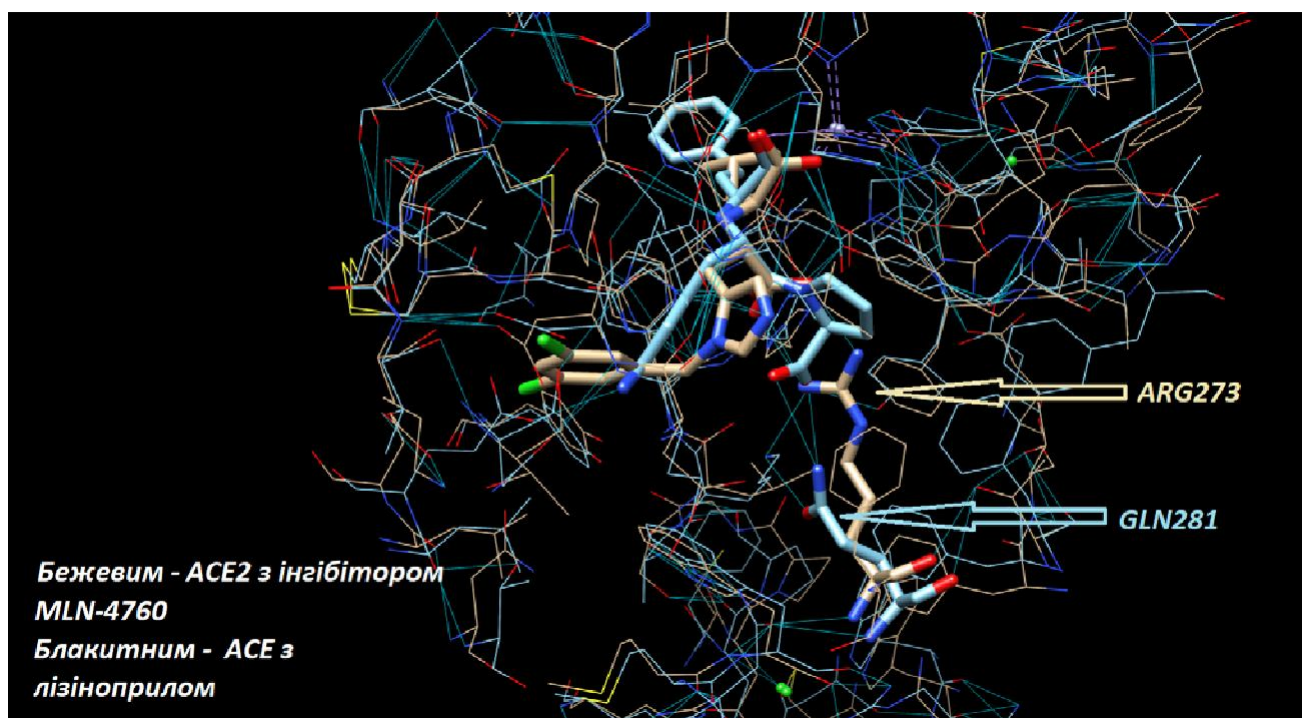
- Arg273 vs Gln281 (Малюнок 5)
- Tyr510 vs Val518 (Малюнок 6)

Візуалізацію створено з використанням програмного забезпечення UCSF Chimera - це програма для інтерактивної візуалізації та аналізу молекулярних структур та відповідних даних, включаючи карти щільності, траєкторії та вирівнювання послідовностей. Вона доступна безкоштовно для некомерційного використання.

Вирівнювання зроблене за допомогою функції MatchMaker. Візуальні обробки (виділення кольором певних залишків, зміна форми тощо) зроблено за допомогою швидких команд у Command Line [31,32].

Є дані щодо перевіреної гіпотези про те, що позитивний бічний ланцюг Arg273 є критичним для зв'язування субстрату. Гіпотеза була перевірена заміною аргініну залишком глутаміну (R273Q), за допомогою сайт-спрямованого мутагенезу. Також для порівняльних цілей залишок аргініну замінили лізином для підтримки заряду на бічному ланцюзі (R273K). За такими замінами не спостерігалось активності ферментів до відомих інгібіторів. Тому позитивний бічний ланцюг ARG273 є критичним для зв'язування субстрату. [10]

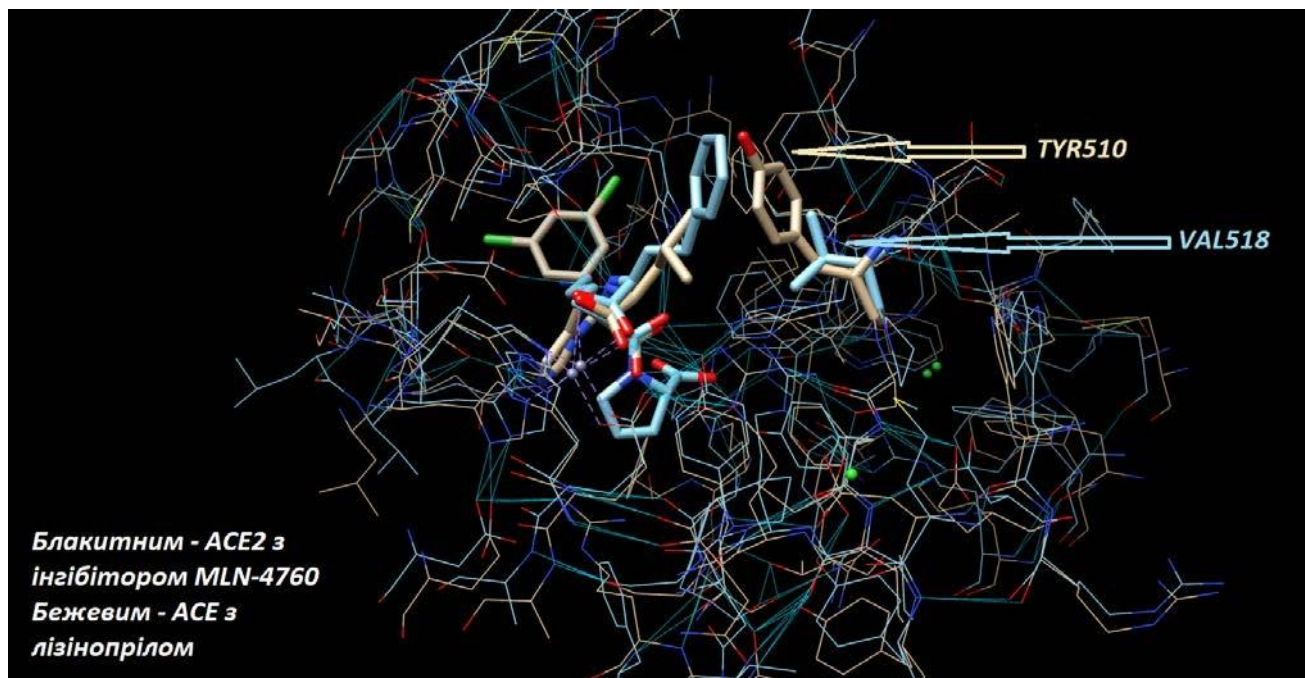
При вирівнюванні та зіставленні структур спостерігаємо, що на аналогічній позиції активного сайту ACE знаходиться Gln281, що має інші хімічні властивості. Тому саме цю відмінність візьмемо для моделювання «сітки» активного сайту для подальшого молекулярного докінгу структур.



Малюнок 5. Візуалізація відмінностей структур активних сайтів ACE та ACE2. ARG273 - світлокоричневим (ACE2) проти GLN281 – блакитним (ACE)

Також вважаю однією із головних відмінностей активних сайтів протеїнів ACE2 та ACE це Tyr510 (ACE2) та Val518 (ACE). Tyr510 обумовлює гідрофобну ділянку та може мати рі-зв'язок з лігандом. Val518 (ACE) також буде

взаємодіяти з гідрофобним замісником ліганду, а бічний ланцюг валіну набагато меншого розміру, тому замісник активного ліганду може бути меншим щоб зберігати зв'язок з тирозином. Ця відмінність теж буде використана під час моделювання «ґрід» ACE2 для молекулярного стикування.



Малюнок 6. Візуалізація відмінностей структур активних сайтів ACE та ACE2. TYR510 - світлокоричневим (ACE2) проти VAL518 – блакитним (ACE)

Є дані, які встановлюють важливу роль His505, і His345 для зв'язування активних лігандів (що відповідає His513 та His353 в tACE) [10] Тут немає відмінностей між структурами протеїнів, тому ці залишки не задіяні в якості обмежувачих критеріїв у моделі молекулярного стикування, але будуть використані для оцінки енергії зв'язування різних поз лігандів

## 2.2. Формування моделей для докінгу та їх верифікація

На схемі 1 показано хід роботи з формування моделі для докінгу.

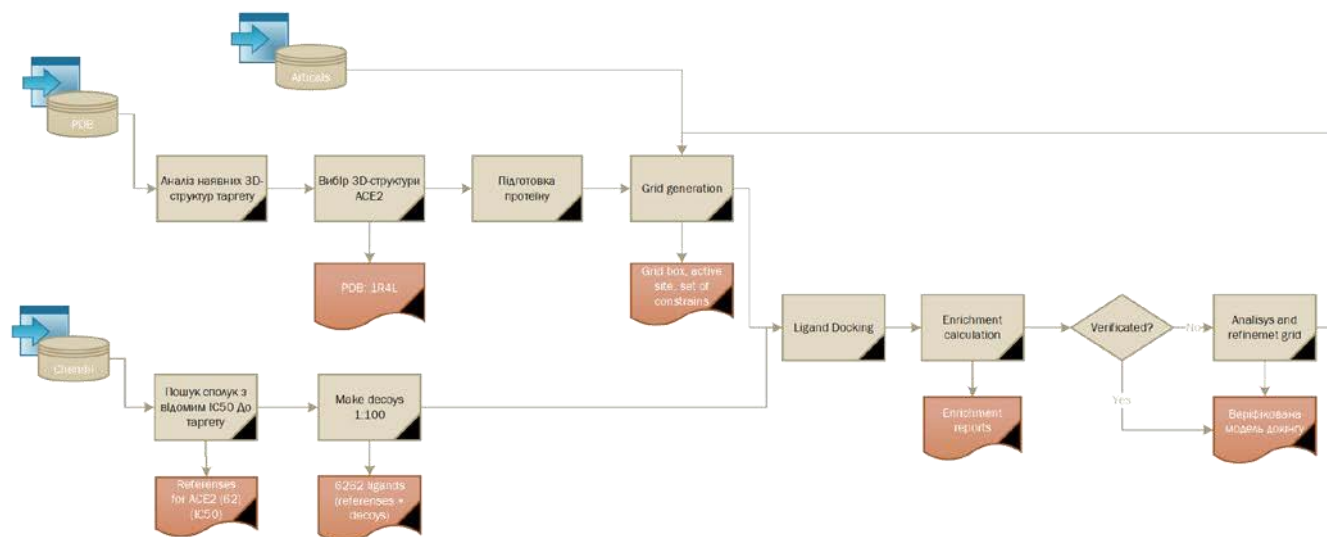


Схема 1. Хід роботи з формування моделі для докінгу

Розглянемо окремо кожний крок:

1) Вибір та підготовка протеїну проводилась за такими критеріями:

- висока якість (до 2~2.5 Å);
- наявність в кристалі активного ліганду (протеїн з лігандом краще, ніж “порожній”);
- організм: Homo sapiens;
- дата публікації та рейтинг журналу

2) Підготовка протеїну проводилась за допомогою The Protein Preparation Wizard.

Було видалені оригінальні водні та потім додані модулем The Protein Preparation Wizard, видалені всі молекули води, структура протеїну підготовлена (сгенеровані усі можливі стани при  $pH = 7.4 \pm 0.2$ ) та мінімізована в силовому полі OPLS3e (зсув важких атомів не більше ніж 0.3 ангстрема). [15]

Оптимізація Н-зв'язків (ProtAssign). Створюється етап оптимізації Н-зв'язку (званий ProtAssign) найкраща мережа водневих зв'язків шляхом розміщення / обертання воднів та функціональних груп, які не впливають на електронну густину (Asn, Gln та His) [14]. PROPKA [16] прогнозує значення рKa іонізуючих груп у білках та білково-лігандних комплексах на основі 3D-структур.

### 3) Підготовка докінгової моделі

Докінгова модель була створена за допомогою модулю Prime. Модуль створює файли «ґриду» (англ. Grid – сітка), які вміщують дані про активний сайт рецептора для подальшого стикування. Окрім місця стикування факмакофорна модель може містити «констрейнси» (англ. Constrains – обмеження), що можуть бути використані при подальшому докінгу для позначення певних атомів рецептора (позиційні / NOE, Н-зв'язок / металеві обмеження) для взаємодії ліганд-рецептор. Також у докінговій моделі задаються групи, що слід розглядати як поворотні пьд час генерації сітки. В даний час ці групи обмежені гідроксилами в Ser, Thr та Tyr та тиолами в Cys.

Підготовка протеїнів та докінгових моделей було проведено для двох протеїнів:

- для таргету (цільовому протеїну) – ACE2 – 1R4L
- для протеїна, до якого ліганді повинні бути неактивними – ACE - 6F9T

«ґриди» для ACE2 – 1R4L мають такі обмеження:

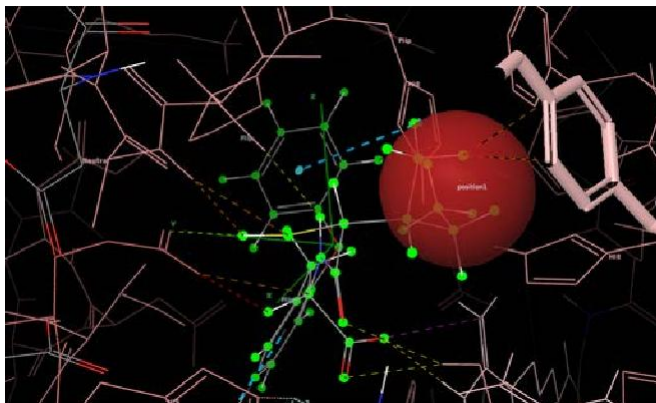
- Перша модель:
  - o ARG273 – водневий зв'язок

- Область біля Tyr510 – позиційне обмеження сферичної форми радіусом 2 Å
- Зв'язок з металом Zn<sup>2+</sup>
- Друга модель:
  - ARG273 – два позиційні обмеження – дві сфери радіусом по 1 Å навпроти атому азоту аргініну на відстані від кожного атому азоту 1,86 Å (місце атомів кисню в ліганді зі структури PDB)
  - Область біля Tyr510 – позиційне обмеження сферичної форми радіусом 2 Å
  - Зв'язок з металом Zn<sup>2+</sup>

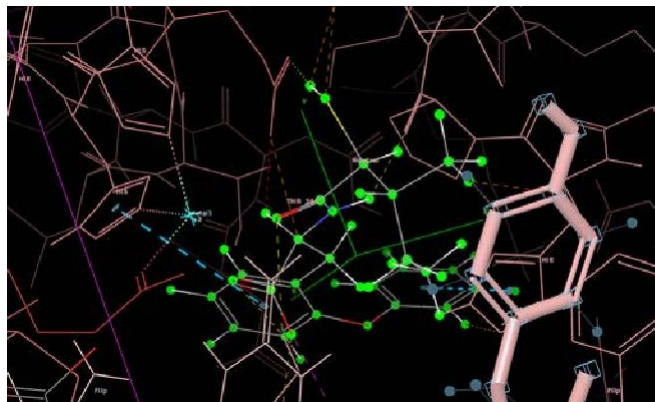
«Грід» для протеїну ACE (PDB код: 6F9T) має наступні констрейнси:

- Зв'язок з металом Zn<sup>2+</sup>
- LYS511 – водневий зв'язок
- TYR520 - водневий зв'язок

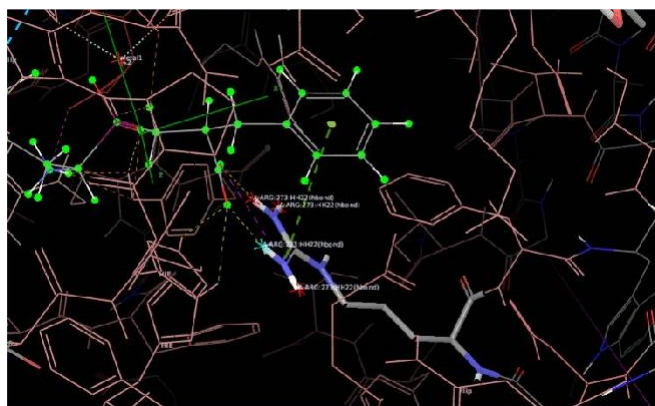
Малюнки ґридів



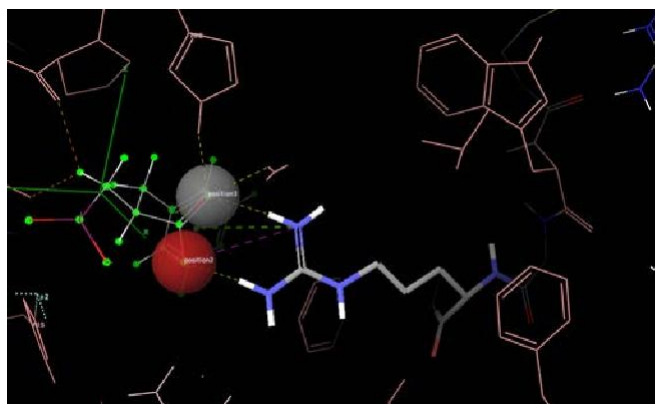
Малюнок 7. Область біля Tyr510 – позиційне обмеження сферичної форми радіусом 2 Å (для двох «ґридів» ACE2 – 1R4L)



Малюнок 8. Зв'язок з металом Zn<sup>2+</sup> (для двох «ґридів» ACE2 – 1R4L)



Малюнок 9. ARG273 – водневий зв'язок (для першої моделі «ґридів» ACE2 – 1R4L)



Малюнок 10. ARG273 – два позиційні обмеження – дві сфери радіусом по 1 Å навпроти атому азоту аргініну на відстані від кожного атому азоту (для другої моделі «ґридів» ACE2 – 1R4L)

Надалі дві моделі ґридів для ACE2 перевіримо за критерієм Enrichment factor, та виберемо більш точнішу (пункт 7 цього розділу)

#### 4) Підготовка референсних структур

Для перевірки докінгових моделей були відібрані всі структури з бази ChEMBL (<https://www.ebi.ac.uk/chembl/>), що мають фактично виміряні показники активності до ACE2 – константу  $K_i$  або IC50. Нарахувалося 62 ліганда (див. Додаток 1)

#### 5) Підготовка приманок (англ. decoys – приманка)

«Збагачення» вимірює, як відомі ліганди ранжуються на фоні молекул-приманок. Декої – набір молекул, які (ймовірно) не зв'язуються з заданою мішенню. Зазвичай це набір молекул, як правило, з ZINC, які схожі на відомі ліганди за хімічною та фізичною властивістю, але не схожі за індексом Танімото та 2d-фінгерпрінтами. Це інструмент, який використовують для визначення точності моделі ґриду таргету (мішені) – для його верифікації. Таким чином, для порівняння збагачення докінг-станцій необхідний набір порівняльних лігандів і приманок. [17]

В даній роботі використано безкоштовний он-лайн сервіс - каталог корисних манок (DUD) <http://dude.docking.org/>

На кожний відомий референтний ліганд було сформовано по 100 «декоїв» -всього 6200.

Суміш наборів лігандів та декоїв підготували за допомогою модулю Ligand Preparation (далі LigPrep) від Maestro. Ligand Preparation формує набір конформерів низькоенергетичної тривимірної структури з правильними хіральностями для кожної успішно обробленої вхідної структури. LigPrep також може створювати ряд структур з кожної вхідної структури з різними станами іонізації, таутомерами, стереохімією та кільцевими конформаціями, а також усунути молекули, використовуючи різні критерії, включаючи молекулярну масу або задану кількість та типи присутніх функціональних груп. [22]

б) Молекулярне стикування

Далі проводила молекулярні докінги HTVS та SP (Standard Precision) з вищезазначеними обмеженнями сітки рецептору та наступники обмеженнями щодо лігандів, а саме:

1. Сфера біля TYR510 – будь-яка гідрофобна ділянка (27 патернів)

PATTERN1 "[C]([CH3])([CH3])([CH3]) 1,2,3,4 include"

PATTERN2 "[CH]([CH3])([CH3]) 1,2,3 include"

PATTERN3 "[C,c]~[C,c]~[C,c]~[C,c]~[C,c]~[C,c]~[C,c]~[C,c] 1,2,3,4,5,6,7,8 include"

PATTERN4 "[C,c]~[C,c]~[C,c]~[C,c]~[C,c]~[C,c]~[C,c] 1,2,3,4,5,6,7 include"

PATTERN5 "[C,c]~[C,c]~[C,c]~[C,c]~[C,c]~[C,c] 1,2,3,4,5,6 include"

PATTERN6 "[C,c]~[C,c]~[C,c]~[C,c]~[C,c] 1,2,3,4,5 include"

PATTERN7 "[C,c]~[C,c]~[C,c]~[C,c] 1,2,3,4 include"

PATTERN8 "[C,c]~[C,c]~[C,c] 1,2,3 include"

PATTERN9 "[C,c]~[C,c] 1,2 include"

PATTERN10 "[CH3] 1 include"

PATTERN11 "C#C 1,2 include"

PATTERN12 "[CH2](C#[N-]) 1 include"

PATTERN13 "C=C 1,2 include"

PATTERN14 "[CX3](=[N])([NX3])[NX3] 1 include"

PATTERN15 "[C;X2] 1 include"

PATTERN16 "[#17] 1 include"

PATTERN17 "[#35] 1 include"

PATTERN18 "[#53] 1 include"

PATTERN19 "[Cl-] 1 exclude"

PATTERN20 "[Br-] 1 exclude"

PATTERN21 "[I-] 1 exclude"

PATTERN22 "C#N 1 exclude"  
 PATTERN23 "[c](~[n])~[n] 1 exclude"  
 PATTERN24 "[CX4](~[O,N,n])~[O,N,n] 1 exclude"  
 PATTERN25 "C(=O)[O] 1 exclude"  
 PATTERN26 "C(=[O,S])[N;X3] 1 exclude"  
 PATTERN27 "C=[O,N,S] 1 exclude")

2. H-bond Zn<sup>+</sup> - з будь-яким атомом-акцептором водневого зв'язку (34 патерни)

PATTERN1 "[N]#C 1 include"  
 PATTERN2 "[n] 1 include"  
 PATTERN3 "N(=N=N) 1 include"  
 PATTERN4 "N(=N)=N 1 include"  
 PATTERN5 "[N;X2]=C[N;X3] 1 include"  
 PATTERN6 "[N;X1] 1 include"  
 PATTERN7 "[N;X2] 1 include"  
 PATTERN8 "[N;X3] 1 include"  
 PATTERN9 "[#7] 1 include"  
 PATTERN10 "[O;X1]~[N;X3]~[O;X1] 1,3 include"  
 PATTERN11 "[O;X1;-]C=[O;X1] 1,3 include"  
 PATTERN12 "[O-][P,S] 1 include"  
 PATTERN13 "[O-] 1 include"  
 PATTERN14 "[o] 1 include"  
 PATTERN15 "[O;X1] 1 include"  
 PATTERN16 "[O;X2] 1 include"  
 PATTERN17 "[#8] 1 include"  
 PATTERN18 "[S-;X1] 1 include"  
 PATTERN19 "[F-] 1 include"

PATTERN20 "[Cl-] 1 include"  
 PATTERN21 "[n;X3] 1 exclude"  
 PATTERN22 "[N;X3][\*]=[\*] 1 exclude"  
 PATTERN23 "[N;X3][c,n]~[c,n,o] 1 exclude"  
 PATTERN24 "[NX3;+] 1 exclude"  
 PATTERN25 "[NH3] 1 exclude"  
 PATTERN26 "[NX4;+] 1 exclude"  
 PATTERN27 "[s] 1 exclude"  
 PATTERN28 "[SX1]=[\*] 1 exclude"  
 PATTERN29 "[S-]C 1 exclude"  
 PATTERN30 "[SX2;H1] 1 exclude"  
 PATTERN31 "[SX2] 1 exclude"  
 PATTERN32 "S(=O)(=O)([C,N])([C,N]) 1 exclude"  
 PATTERN33 "S(=O)([#6])([#6]) 1 exclude"  
 PATTERN34 "[SX4] 1 exclude

### 3. ARG273

- 3.1. Перша модель «ґрід» - з будь-яким атомом-акцептором водневого зв'язку
- 3.2. Друга модель «ґрід» - будь-який атом кисню в обох сферах ( PATTERN1 "[O;-]C=[O] 1,3 include")

Отримано чотири сети результатів:

Грід-1 – HTVS

Грід-2 – HTVS

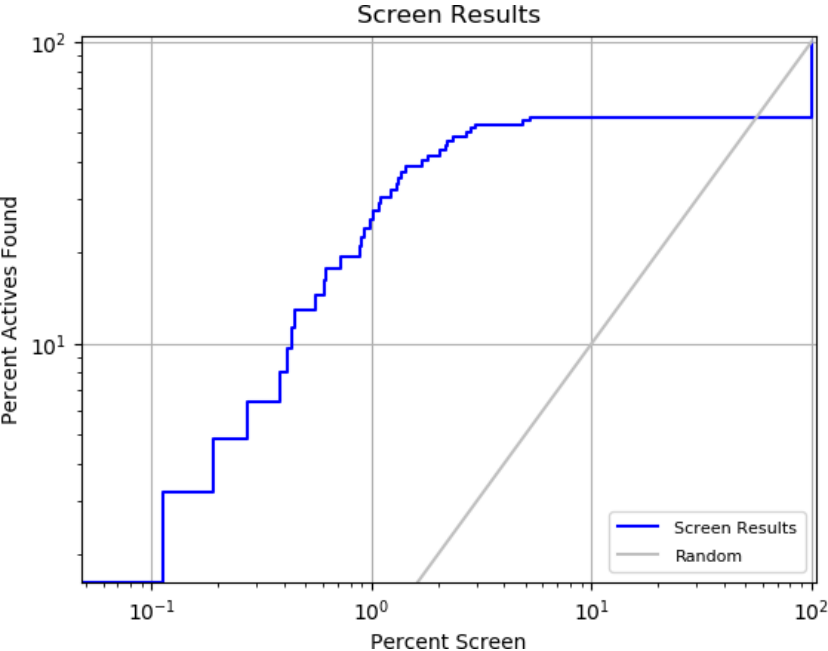
Грід-1 – SP

Грід-2 – SP,

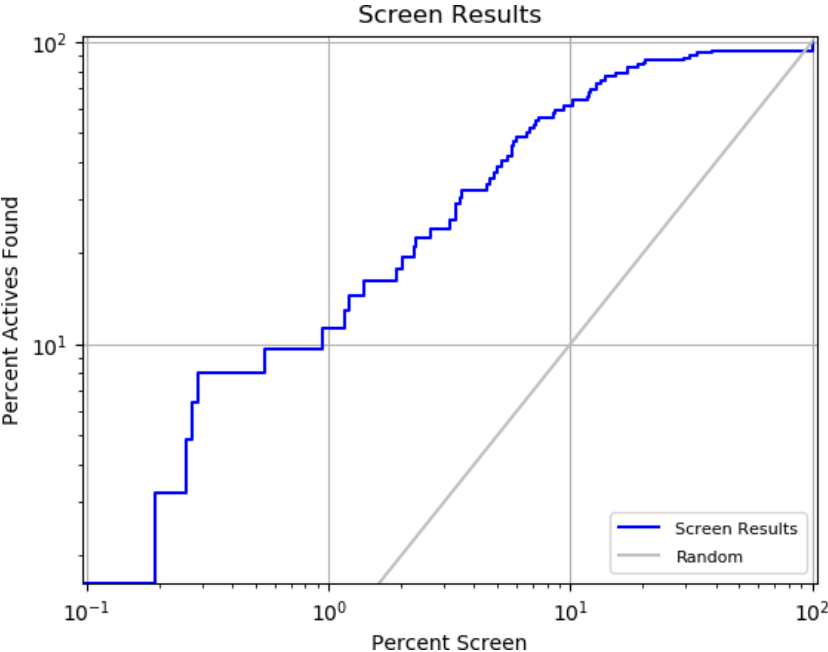
які надалі будемо оцінювати за допомогою Enrichment Calculator



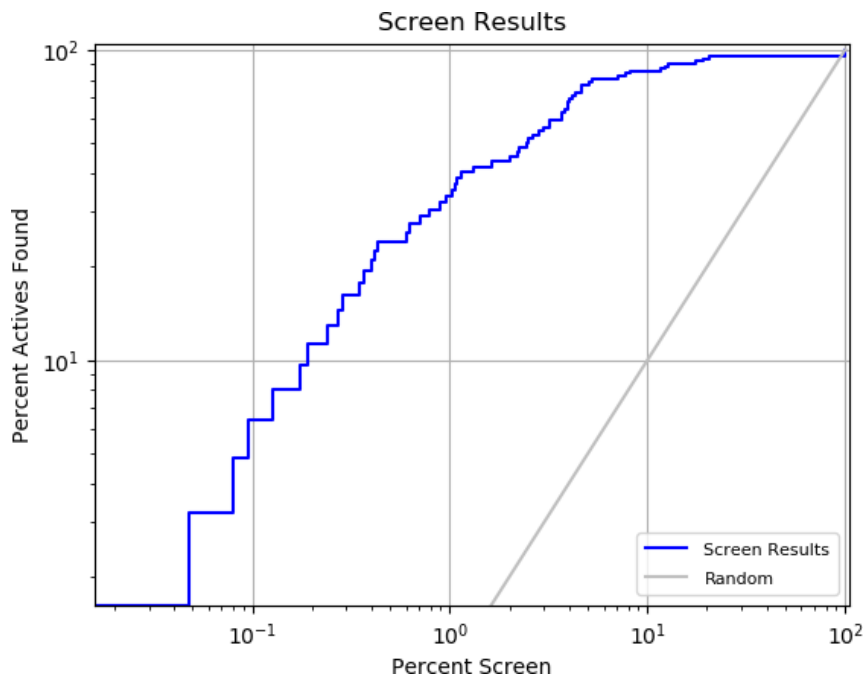
Другий сет: Грід-2 – HTVS - Area under accumulation curve: 0.76 -



Грід-1 – SP - Area under accumulation curve: 0.87 -



Грід-2 – SP - Area under accumulation curve: 0.94 -



Порівнявши результати бачимо, що позиційні обмеження (сферичні) біля ARG273 дають точнішу модель для докінгу в порівнянні з ARG273-H-bond як для HTVS так і для SP. Тому для подальшого скринінгу будемо використовувати Грід-2, будемо вважати цю модель верифікованою.

## РОЗДІЛ III. Віртуальний скринінг

### 3.1. Віртуальний скринінг стокової бази Enamine

В моїй роботі використовували скринінгові бібліотеки компанії Enamine. На момент роботи з бібліотекою для даного проекту стокова скринінгові колекція містила 3,2 мільйони низькомолекулярних органічних сполук.

На схемі 2 зображена схема роботи з цією бібліотекою

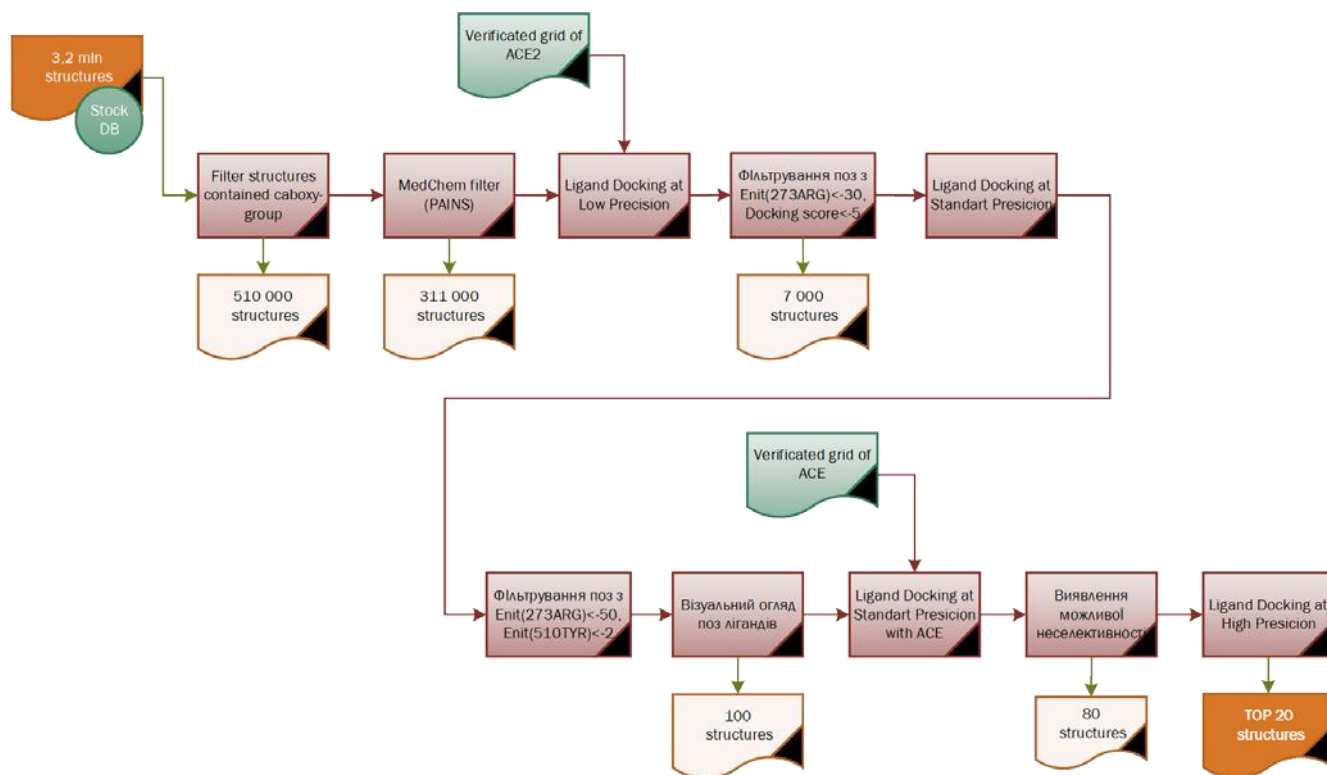


Схема 2. Схема віртуального скринінгу стокової бази Enamine

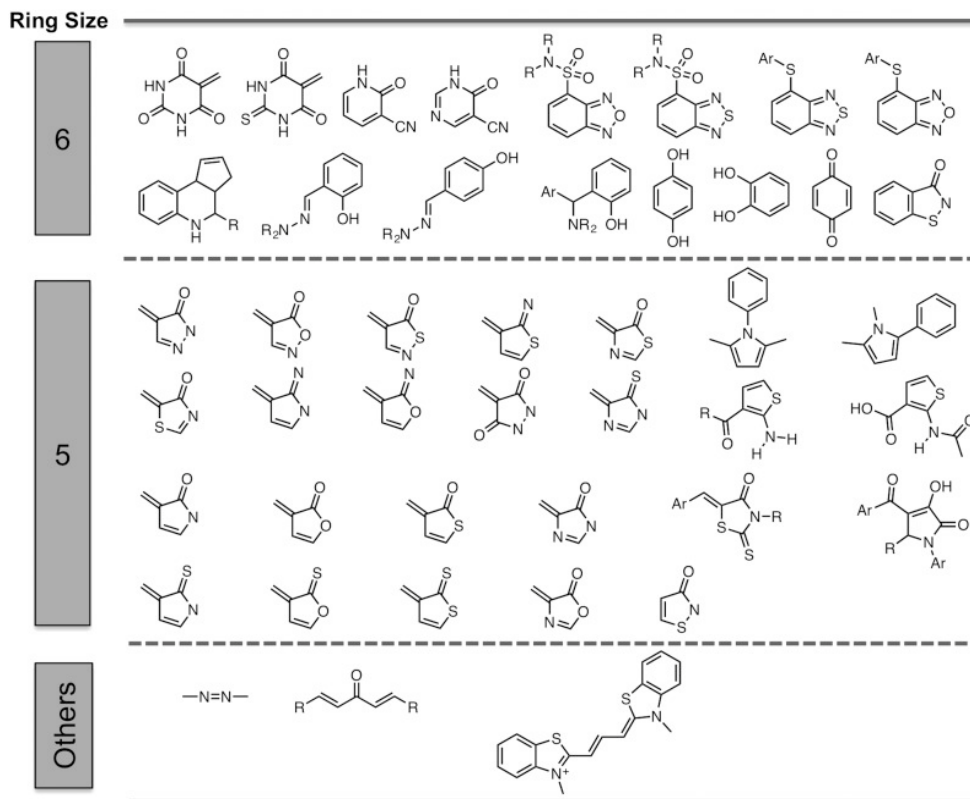
Бібліотека сполук розміром 3,2 мільйони досить велика для докінгу, тому вона повинна буди спочатку зменшена шляхом фільтрування (якщо є пріоритетність щодо субструктур лігандів що ймовірно будуть інгібувати нашу мішень). В нашому випадку більшість відомих лігандів мають по дві карбоксильні групи, або одну карбоксильну, а іншу – сульфат, або фосфатний залишок. Тому мною було вирішено спочатку відфільтрувати бібліотеку за

SMARTS субструктурою за допомогою скрипту на мові Python та модулю RDKit.

Використаний SMARTS - [O;H,-]C=O , що відфільтрує і нейтральні сполуки і заряджені (у вигляді солі). Після фільтрування залишилось – 510 000 сполук.

Наступним кроком мною було відфільтровано найпоширеніші фільтри, що використовуються в медичній хімії, а саме PAINS (Pan-Assay Interference Compounds) та LILLY.

Ці фільтри розроблені для вирішення проблем неселективності. Фільтри визначають чи є у сполуках підструктури, що з великою ймовірністю будуть інгібувати велику кількість протеїнів в організмі – як цільового так и багатьох інших. Це призведе до неможливості їх використання. PAINS – пов'язані з неконтрольованою біоактивністю та аналітичними інтерференціями в реальних та віртуальних кампаніях скринінгу з високою пропускнуою здатністю. На малюнку 11 наведені приклади субструктур PAINS.



Малюнок 11. Хімічні підструктури найгірших сполук, що впливають на аналіз (PAINS) та потенційних сполук.

Ці субструктури складають частину повної хімічної структури PAINS і припускають високий потенціал для цієї сполуки впливати на біологічні аналізи. Не кожна сполука, що містить одну з цих підструктур, буде інтерференційною сполукою, але діяти обережно, якщо активна сполука містить одну з цих підструктур. [27]

LILLY - набір з 275 правил, розроблених Lilly Research Laboratories, що використовуються для ідентифікації сполук, які можуть впливати безконтрольно на біологічні аналізи, що дозволяє видаляти їх із скринінгових наборів. Критерії цих правил включають реакційну здатність (наприклад, ацилгалогеніди), впливання на вимірювання аналізу (флуоресценція, поглинання, гасіння), діяльність, яка пошкоджує білки (окислювачі, миючі засоби), нестабільність (наприклад, приховані альдегіди) та відсутність властивості druglikeness (наприклад, сполуки бракує як кисню, так і азоту). [28] Після цього кроку кількість сполук у нашому наборі зменшилася до 311 000.

Наступний крок – це молекулярний докінг з низькою точністю (HTVS) використовуючи верифіковану модель (Розділ II). Запис результатів докінгу робила з розрахунком та фіксуванням енергій взаємодії лігандів з кожним амінокислотним залишком таргету (до 12 ангстрем).

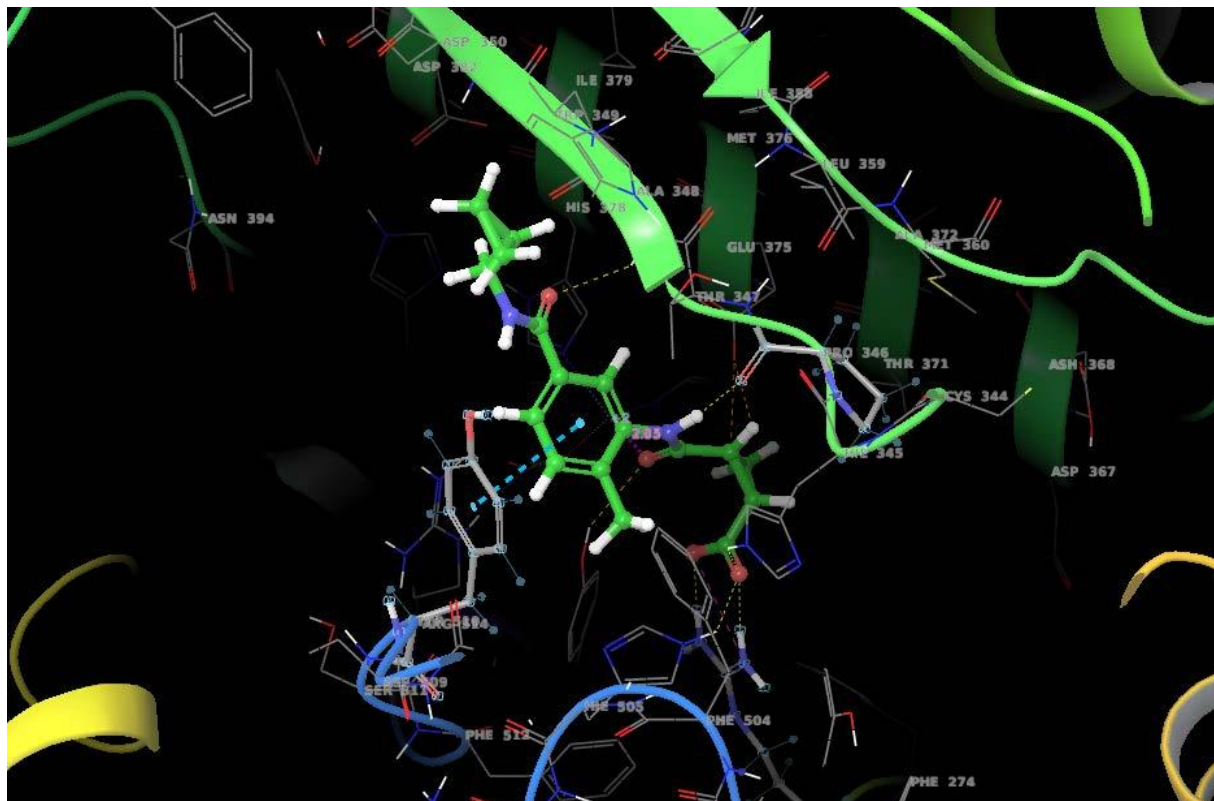
Візуальний огляд поз лігандів має велике значення у віртуальному скринінгу. [28] Запис енергій взаємодії лігандів з кожним амінокислотним залишком таргету дуже допомагає у візуалізації, як інструмент, що може відразу вивести ранжування лігандів не тільки за загальним скором, а й за енергією взаємодії у важливих субкишенях активного сайту мішені, що свідчить про правильну позу у цих ділянках.

Так відфільтрувала пози сполук за значеннями  $E_{\text{nit}}(273\text{ARG}) < -30$  та  $\text{Docking score} < -5$  - залишилось біля 7 000 сполук.

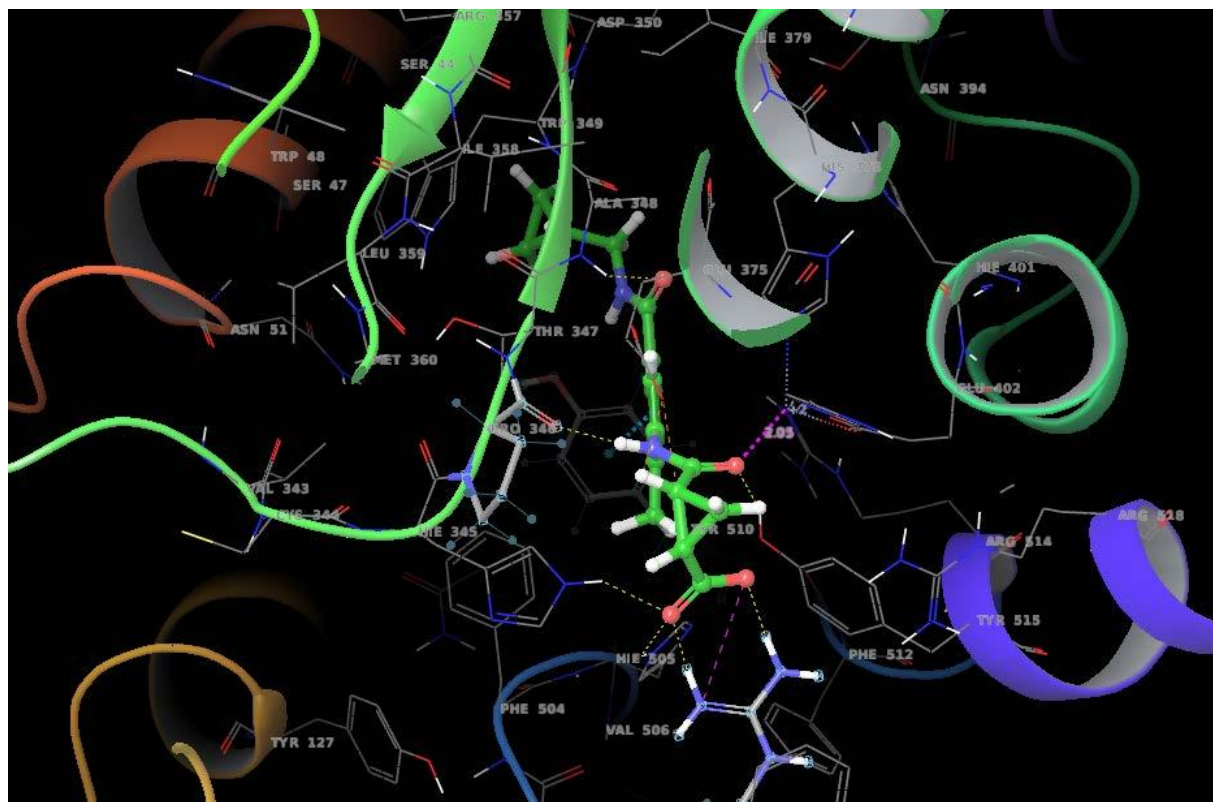
Молекулярний докінг зі стандартною точністю (SP) використовуючи верифіковану модель наступний крок нашого скринінгу. Для більш продуктивного віртуального огляду також спочатку відфільтровано за значеннями  $E_{\text{nit}}(273\text{ARG}) < -50$ ,  $E_{\text{nit}}(510\text{TYR}) < -2$ , а вже потім проводила «cherry picking».

Під час «cherry picking» найважливіше на що було звернуто увагу (Малюнок 12):

- Точність пози щодо Arg273
- Точність пози щодо Zn<sup>2+</sup>
- Позиція гідрофобного замісника ліганду біля Tyr510
- Наявність водневого зв'язку з проліном Pro346



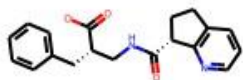
Малюнок 12а. Важливі критерії cherry picking - позиція гідрофобного замісника ліганду біля 510TYR, та наявність водневого зв'язку з проліном 346PRO



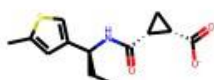
Малюнок 12б. Важливі критерії cherry picking - точність пози щодо 273ARG, точність пози відносно Zn<sup>2+</sup>

Для перевірки можливої неселективності лігандів по відношенню з ACE, далі провела молекулярне стикування здобутих ТОП100 сполук з мішенню ACE (PDB код: 6F9T). Були видалені сполуки, що «правильно» встали в сайт зв'язування ACE, тому що ймовірно, що ці сполуки не будуть мали селективності до ACE2.

Залишені 80 сполук досліджували молекулярним стикуванням на рівні XP (extra precision) – з високою точністю. Модель використана та ж, що і для SP. Експертною оцінкою та звертаючи увагу на Docking score, на специфічні взаємодії з важливими залишками було знайдено 20 TOP сполук. (Малюнок 12)



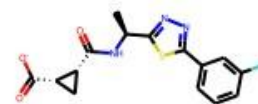
Z2831426920



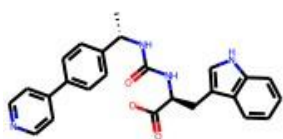
Z3681209542



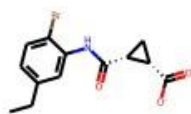
Z3681208436



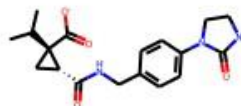
Z3681332013



Z3969355209



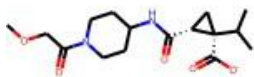
Z3681736600



Z3681207877



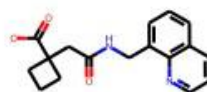
Z3681593370



Z3681369109



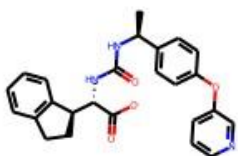
Z4188337437



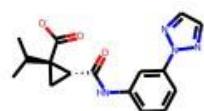
Z3037294184



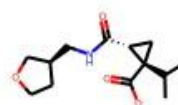
Z4549268235



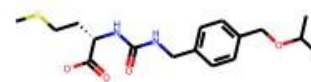
Z4361984504



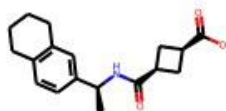
Z3681806363



Z3681799936



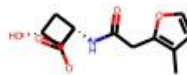
Z4602008618



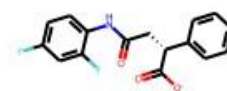
Z3466179106



Z4221819990



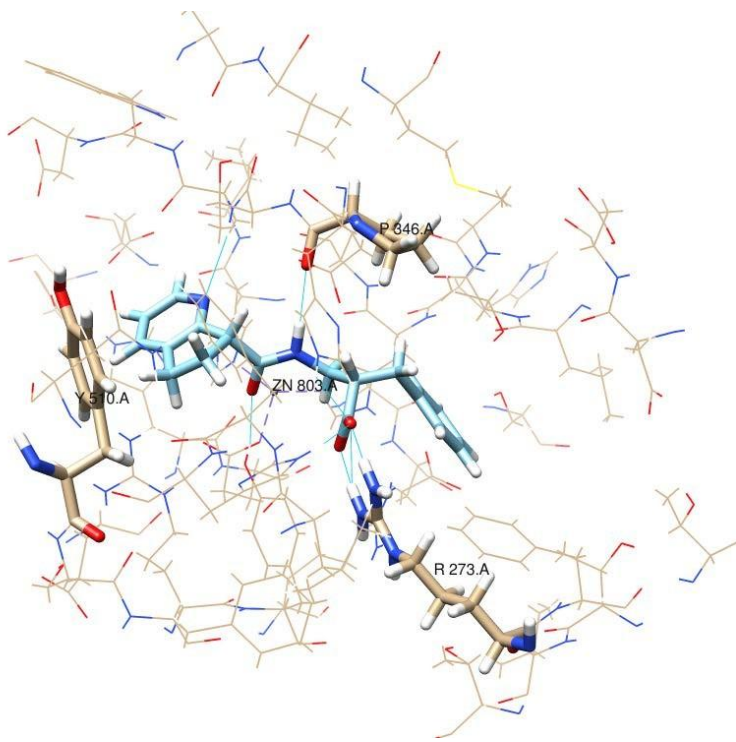
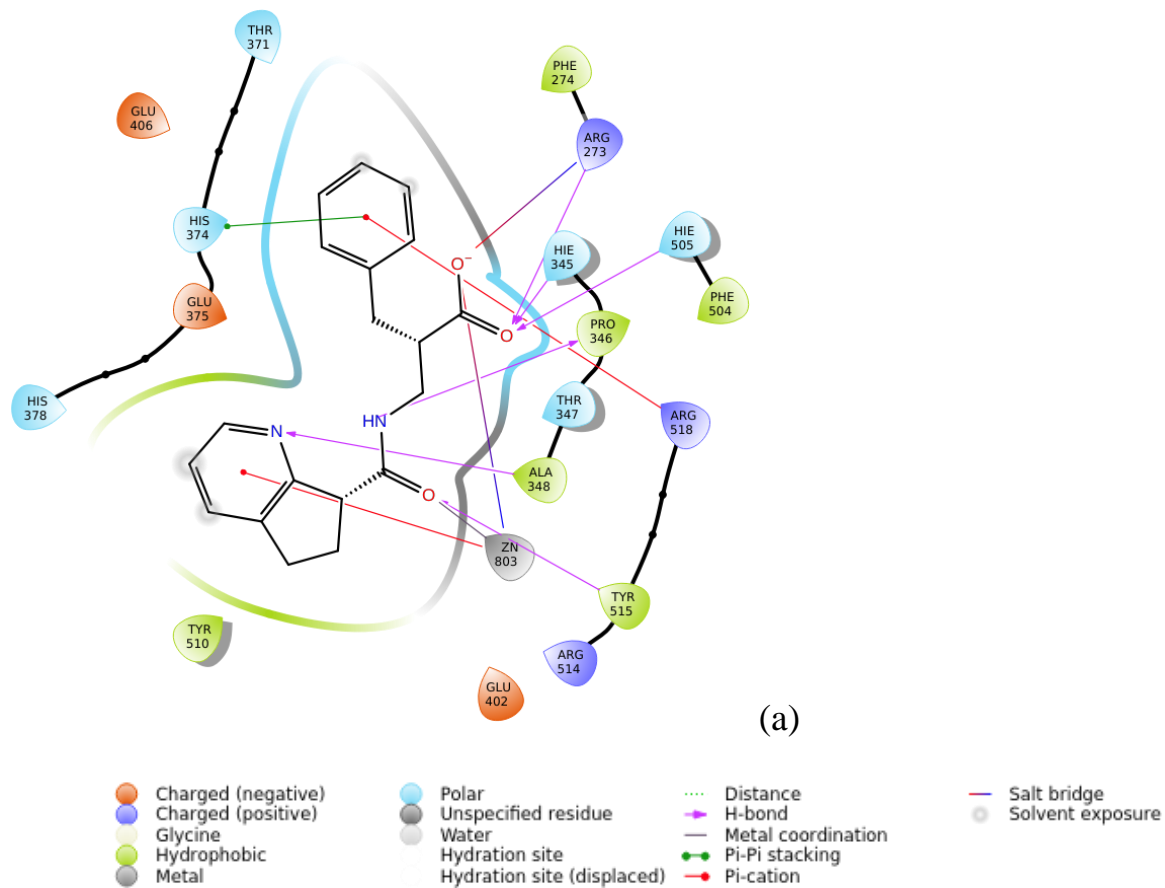
Z4561748167



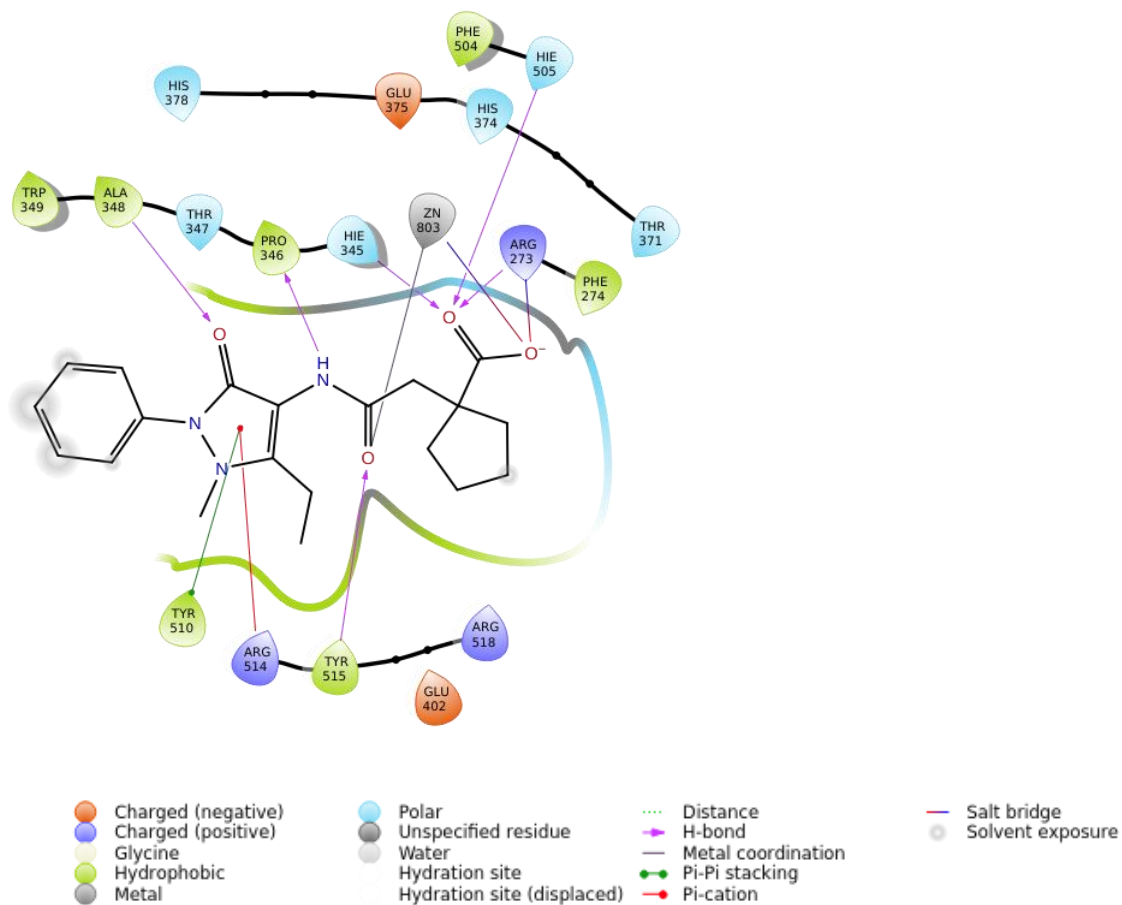
Z4119544425

Малюнок 12. Відібрані ТОП20 сполук з стокової бази

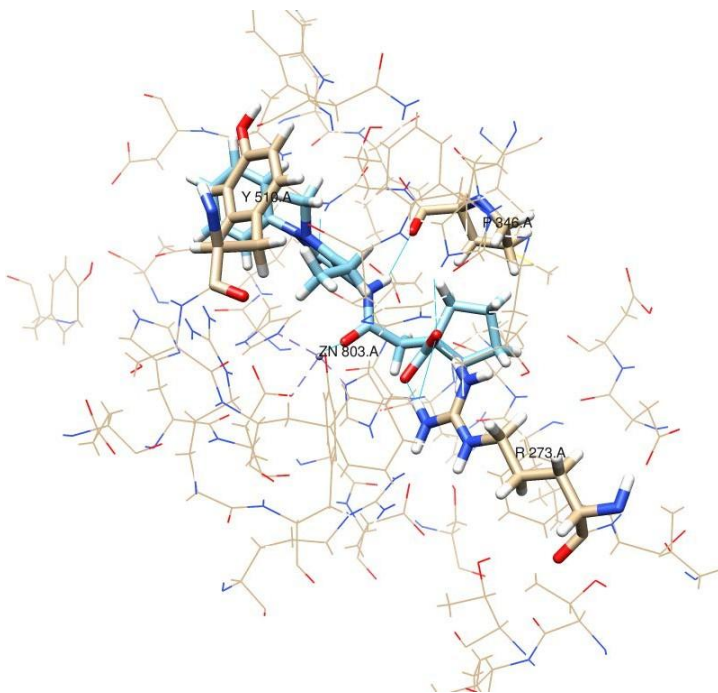
Деякі візуалізації наведені на Малюнках 13-14.



Малюнок 13. (а) 2d діаграма взаємодії ліганду Z2831426920 з активним сайтом мішені. (б) 3d візуалізація пози ліганду Z2831426920



(a)



(b)

Малюнок 14. (а) 2d діаграма взаємодії ліганду Z4188337437 з активним сайтом мішені. (б) 3d візуалізація пози ліганду Z4188337437

### 3.2. Віртуальний скринінг REAL DATA BASE Enamine

На вході - Enamine REAL DATA BASE в кількості 14,1 мільярдів сполук.

На схемі 3 зображена схема роботи з цією бібліотекою.

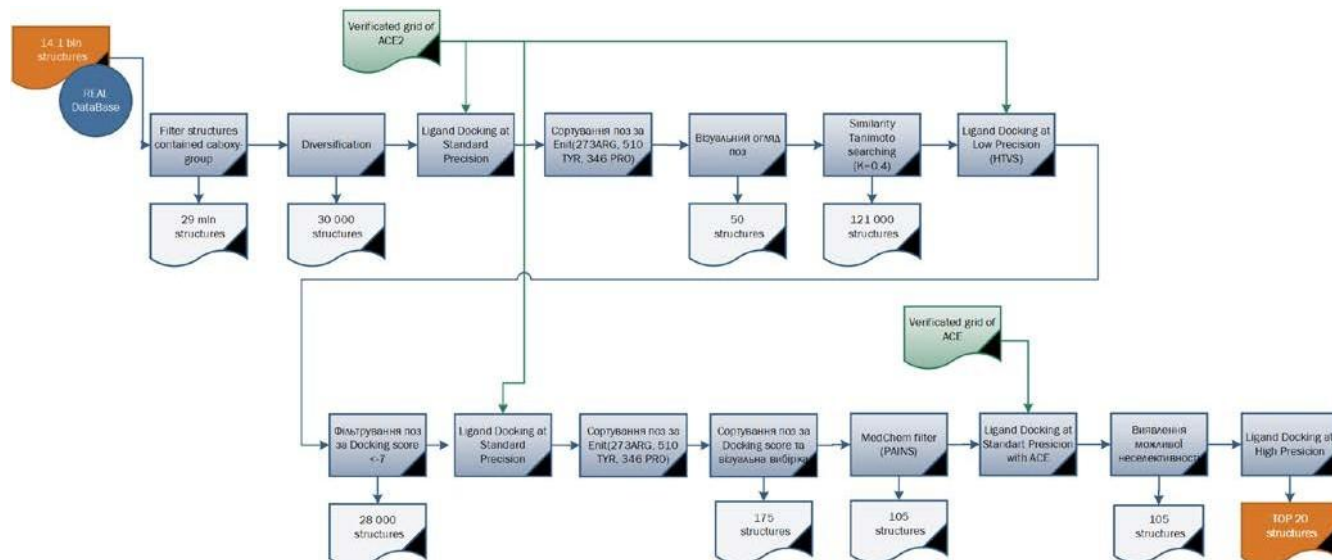


Схема 3. Хід роботи з REAL DATA BASE Enamine

Ще більша бібліотека сполук тим паче потребує попередньої фільтрації за певними визначеними ознаками. По-перше, було проведено фільтрацію за субструктурою - за допомогою скрипту на мові Python та модулю RDKit. Використаний SMARTS - [O;H,-]C=O, що відфільтрує і нейтральні сполуки і заряджені (у вигляді солей). Після фільтрування залишилось – 29 мільйонів сполук.

Наступним кроком була зроблена диверсифікація бібліотеки. Диверсифікація – вибір різноманітних молекул за допомогою «відбитків пальців» за Морганом (Morgan fingerprints). Загальним завданням є вибір невеликої підмножини різноманітних молекул із більшої сукупності. RDKit пропонує ряд підходів для цього, але найефективніший із них використовує алгоритм MaxMin. [29]

Спершу розбила на три сети по 9-10 мільйонів, кожену частину «зменшили» в тисячу разів зберігаючи хімічне різноманіття сету. Використано скрипт на мові Python та модулю RDKit (Додаток 3). Залишили сет з 30 000 структур.

Наступний крок – це молекулярний докінг зі стандартною точністю (SP) використовуючи верифіковану модель (Розділ II). Запис результатів докінгу робила з розрахунком та фіксуванням енергій взаємодії лігандів з кожним амінокислотним залишком таргету (до 12 ангстрем).

Сортування поз за енергіями Enit(273ARG, 510TYR, 346PRO) та візуальний огляд – вибрала 50 «топових» структур.

Ці 50 структур є, так би мовити, представниками кластерів. Тому подалі пропоную розширити набір структур, зробивши пошук структур з тієї ж бази Enamine REAL DATA BASE за подібністю (Shape Based Searching). Shape Based Searching робили за допомогою RDkit модулю oddt (Open Drug Discovery Toolkit) (Додаток 4), розраховували три показники:

- USR (Ultrafast Shape Recognition) - function usr(molecule)
- USRCAT (USR with Credo Atom Types) - function usr\_cat(molecule)
- Electroshape - function electroshape(molecule)

Скрипт вибрав молекули з бази, що мали електронну подібність до наших топових сполук на рівні 0,4 та вище.

Розширили сет до 121 000 структур.

Наступним кроком знову робимо молекулярний докінг з низькою точністю (HTVS) використовуючи верифіковану модель (Розділ II), результати якого фільтруємо за значенням docking score <-7 – залишили 28 000 лігандів.

Молекулярний докінг зі стандартною точністю (SP) використовуючи верифіковану модель наступний крок нашого скринінгу. Для більш продуктивного віртуального огляду також спочатку відфільтровано за

значеннями Enit(273ARG)<-50, Enit(510TYR)<-2, Enit(346PRO)<-1, відсортовано по docking score, а вже потім проводила «cherry picking» за тими ж критеріями що й під час роботи зі стоковою базою.

Розрахунок ChemFilters (PAINS, LILLY) та відсіювання сполук – наступний крок – залишили 105 лігандів

Для перевірки можливої неселективності лігандів по відношенню з ACE, далі провела молекулярне стикування здобутих ТОП105 сполук з мішенню ACE (PDB код: 6F9T). Лише одна сполука показала скор більше 7, але поза в неї неприродня, тому не вважаю що її треба відкинути.

105 сполук досліджували молекулярним стикуванням на рівні XP (extra precision) з високою точністю. Модель використана та ж, що і для SP. Експертною оцінкою звертаючи увагу на Docking score, на специфічні взаємодії з важливими залишками було знайдено 20 TOP сполук.

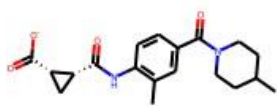
На малюнках 15-17 зображені ТОП20 сполук та їх взаємодії з активним сайтом мішені, на яких видно, що всі топові сполуки мають всі або майже всі зв'язки з наступних:

- ARG273 - H-bond, salt bridge
- TYR510 – Pi-pi stacking
- Zn – Metal coordination, salt bridge
- PRO346 - H-bond
- HIE505 - H-bond
- HIE345 - H-bond
- TYR515 - H-bond

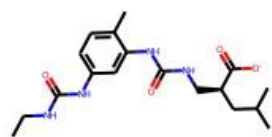
Та поодинокі випадки зв'язків:

- ALA348 - H-bond (рідко)
- TRP349 - Pi-pi stacking
- HIS378 - Pi-pi stacking

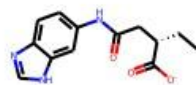
- ARG514 – Pi-cation
- HIE345 - Pi-pi stacking
- ARG273 - Pi-cation



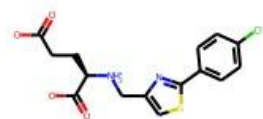
Z1991725061



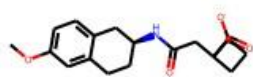
Z2396153450



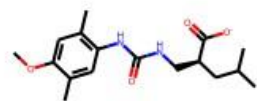
Z4422831009



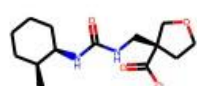
Z2899432355



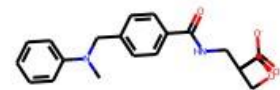
Z2793145343



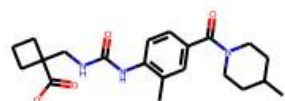
Z2396152156



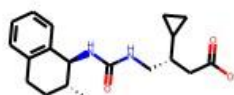
Z3510691032



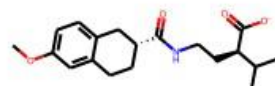
PV-002719568598



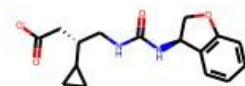
Z2610472638



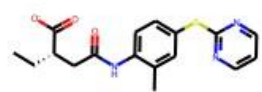
Z2596054371



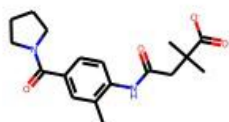
PV-002769423394



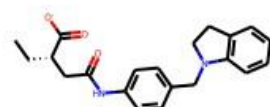
Z2596042648



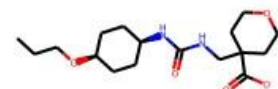
Z4422771675



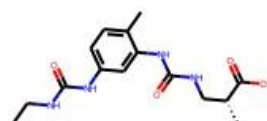
Z1213003608



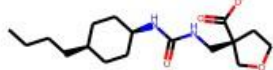
PV-002989187384



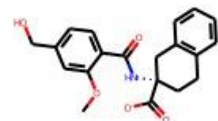
Z3654114554



Z2493059739



Z3510749222

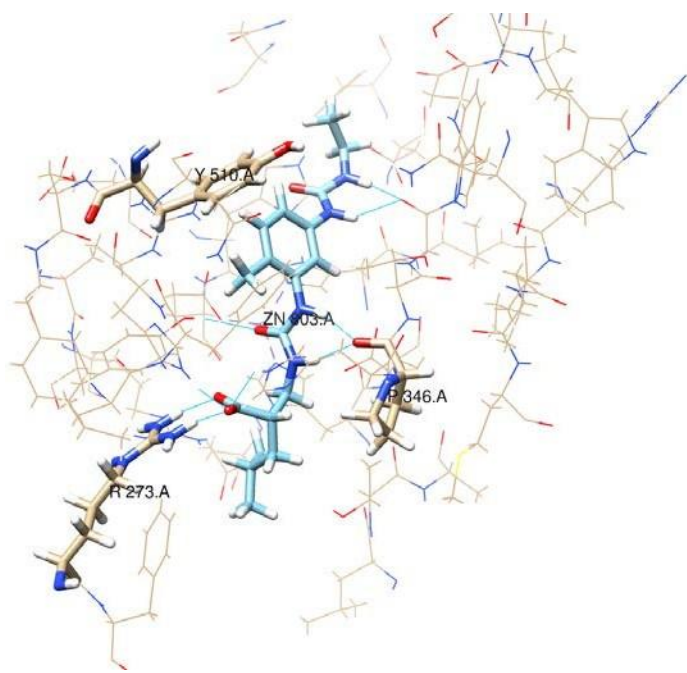
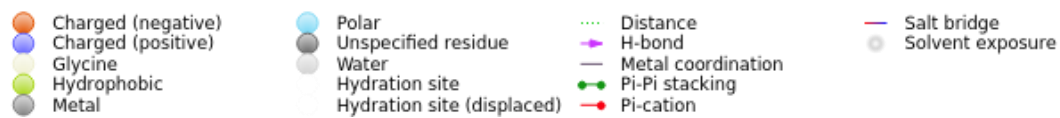
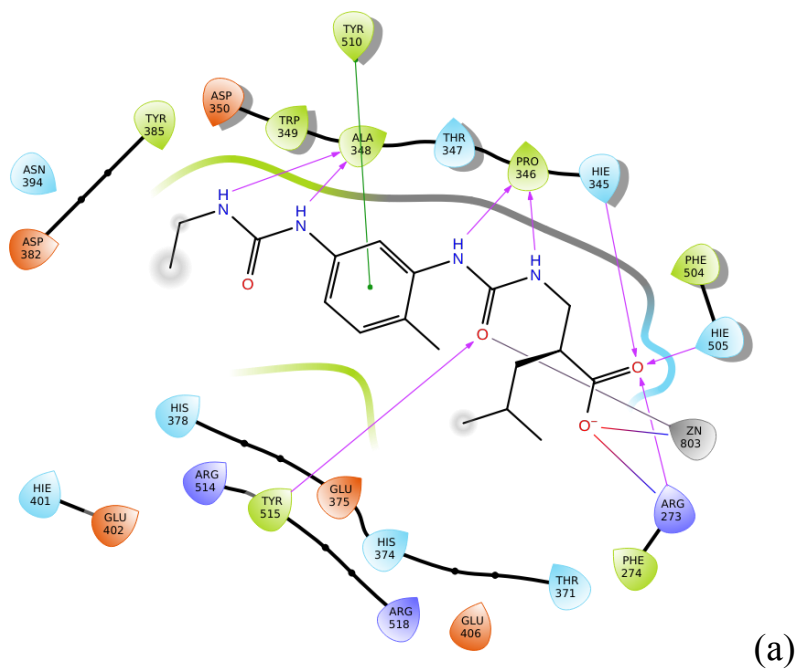


Z3390587460

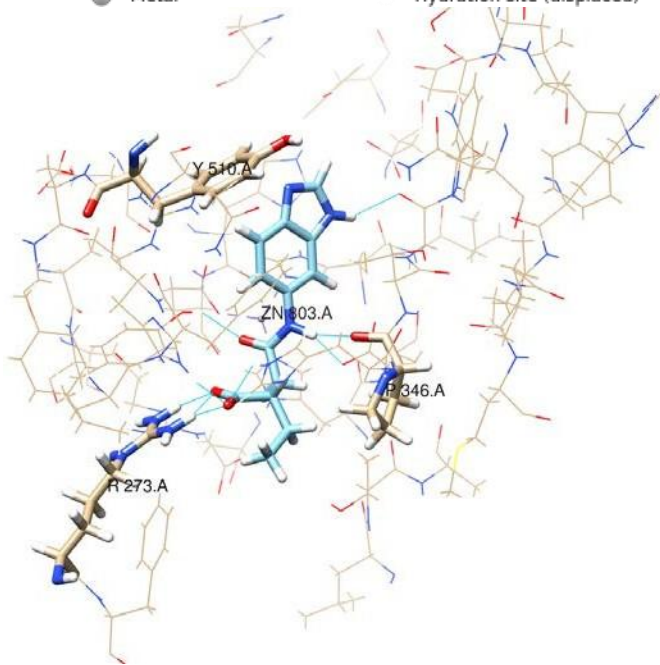
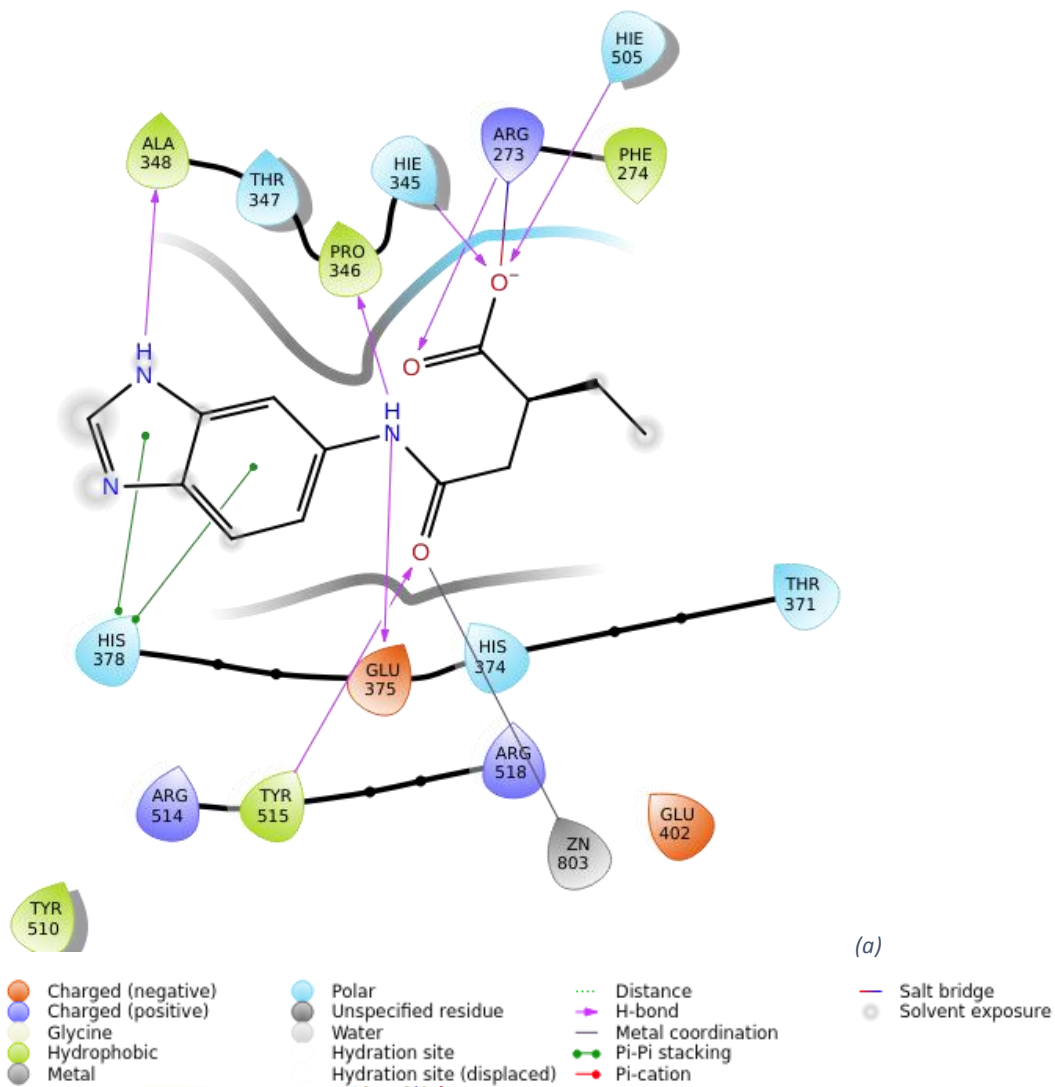


Z2362216591

Малюнок 15. Відібрані ТОП20 сполук з REAL Data Base

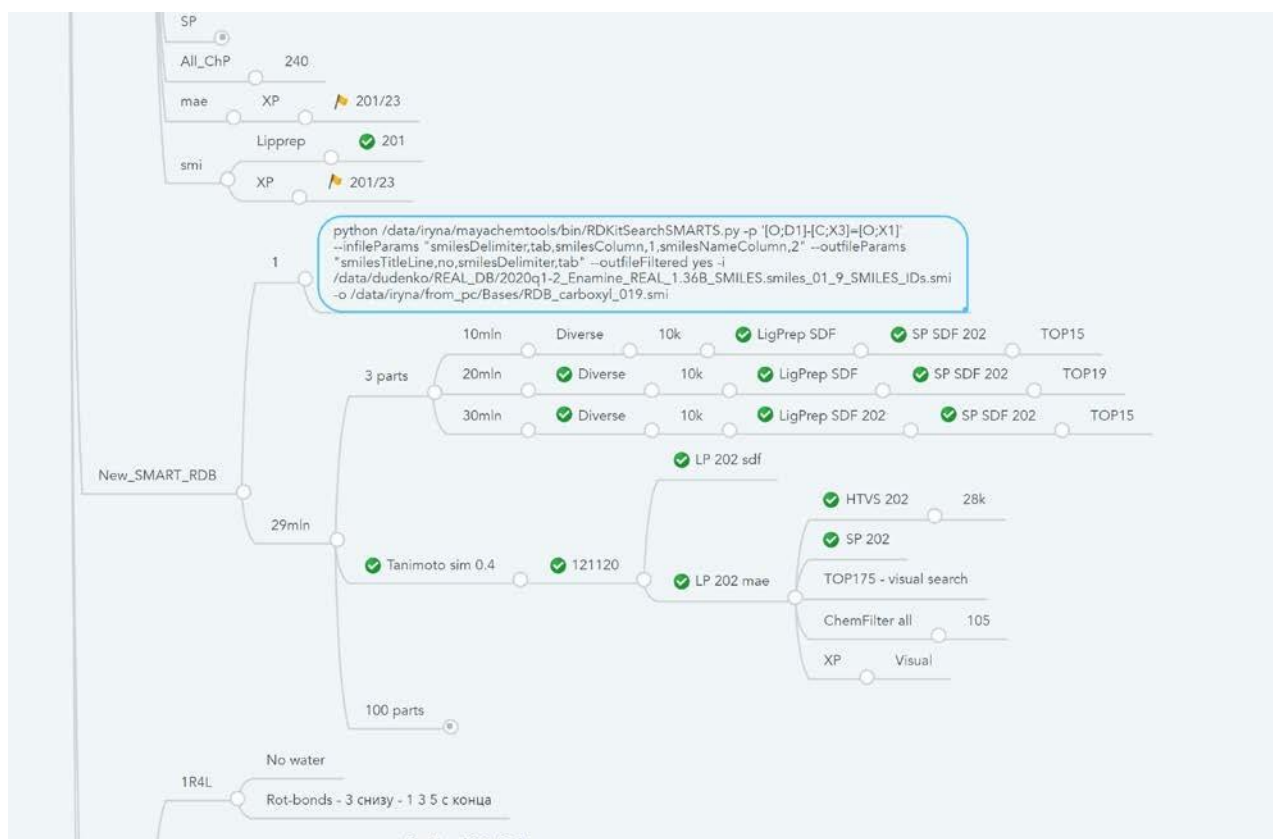


Малюнок 16. (а) 2d діаграма взаємодії ліганду Z2396153450 з активним сайтом мішені. (б) 3d візуалізація пози ліганду Z2396153450



Малюнок 17. (а) 2d діаграма взаємодії ліганду Z4422831009 з активним сайтом мішені. (б) 3d візуалізація пози ліганду Z4422831009.

Також хочу зауважити на тому питанні, що для роботи з кількістю сполук у десятки мільйонів треба розбивати на частини для більш раціонального використання процесорів. В той же час це потребує фіксації дій по роботі з частинами великих файлів. З попереднього досвіду роботи з великою кількістю складних багатокomпонентних задач зручно використовувати інтелектуальні карти. Під час роботи над цією задачею я використовувала систему інтелектуальних карт <https://www.mindmeister.com/> - безкоштовний його пакет.



Малюнок 18. Інтелектуальна карта для зручного фіксування ходу роботи з файлами, та їх частинами

### 3.3. Молекулярна динаміка

Для більш точної перевірки відібраних лігандів, було вибрано ТОП3 лігандів зі стоку та ТОП3 лігандів з RDB, та для них змодельована молекулярна динаміка.

Молекулярна динаміка ( МД ) є комп'ютерним моделюванням методу аналізу фізичних рухів з атомів і молекул . Атомам і молекулам дозволяється взаємодіяти протягом певного періоду часу, даючи уявлення про динамічну "еволюцію" системи. У найпоширенішій версії траєкторії руху атомів і молекул визначаються числовим розв'язком рівнянь руху Ньютона для системи взаємодіючих частинок, де сили між частинками та їх потенційні енергії часто обчислюються за допомогою міжатомних потенціалів або молекулярна механіка силові поля.

Схема моделювання молекулярної динаміки повинна враховувати наявну обчислювальну потужність. Розмір моделювання ( $n$  = кількість частинок), крок часу та загальна тривалість часу повинні бути вибрані так, щоб обчислення могло закінчитися протягом розумного періоду часу. Однак моделювання повинно бути достатньо тривалим, щоб відповідати часовим масштабам природних процесів, що вивчаються. Щоб зробити статистично обґрунтовані висновки з моделювання, проміжок часу, що моделюється, повинен відповідати кінетиці природного процесу.

В даній роботі молекулярна динаміка робилась за допомогою Desmond від Maestro (Schrodinger). Виконання дослідження, заснованого на моделюванні молекулярної динаміки Desmond включає кількість етапів, включаючи налаштування моделювання, розслаблення системи (це може бути просто мінімізацією), запуск моделювання, перегляд траєкторій та аналіз результатів.  
[30]

Виконання моделювання на водних біологічних системах вимагає підготовки біологічних такі молекули, як білки та ліганди, додавання зустрічних іонів для нейтралізації системи, вибір розміру вікна моделювання, сольватація розчинених речовин за допомогою явних молекул розчинника та вирівнювання білків до мембранного бішару (якщо він використовується).

Для оцінки різниці в зв'язуванні використовували параметри, отримані з аналізу молекулярної динаміки, такі як різниця RMSD (середньо-квадратичне відхилення) ліганду в комплексі протягом динаміки, оцінка кількості контактів між лігандом та протеїном в кожній точці протягом періоду динаміки, а також відсоток часу, в якому зв'язки були присутні. Для чотирьох комплексів зроблена динаміка на 10,2 нс.

Два ліганди зі стоку:

1) Z2831426920

2) Z3969355209

Два ліганди з бази RDB:

1) Z1991725061

2) Z2396153450

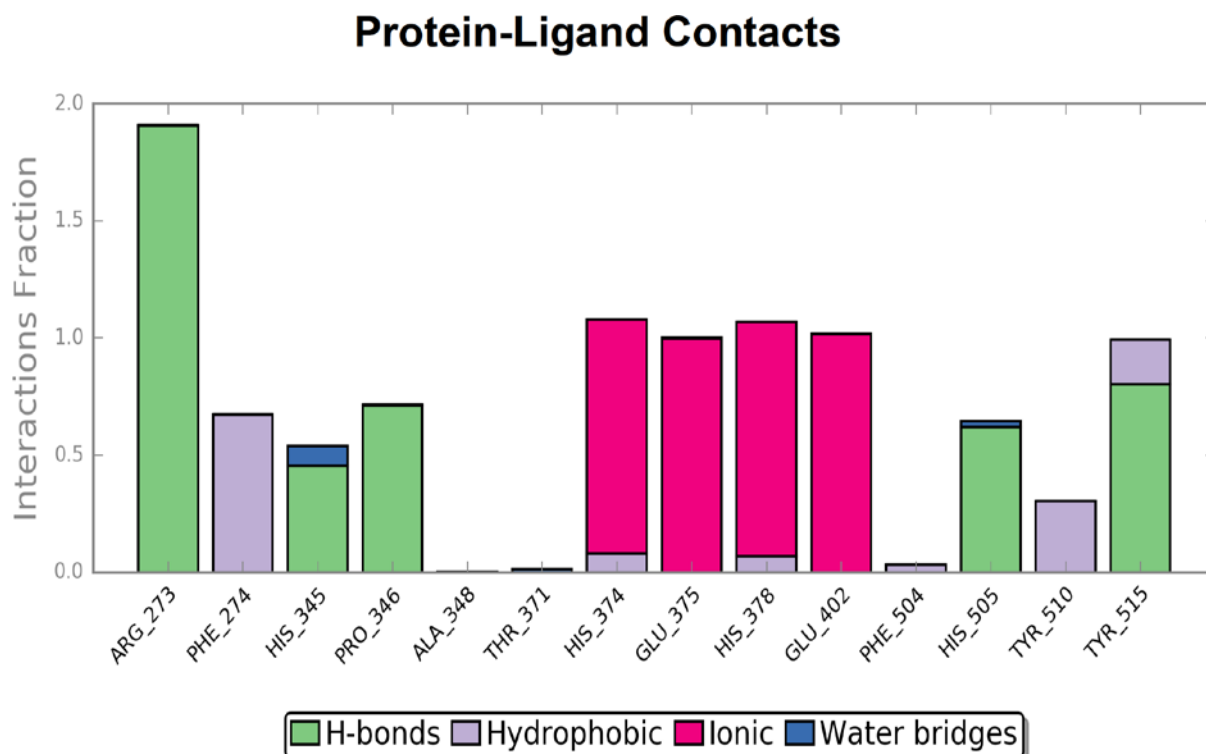
Аналізуватимемо дані за діаграмами, що зроблені за допомогою Simulation Interactions Diagram Report – по кожному моделюванню надаю по діаграмі взаємодії білку та ліганду, схемі контактів та графічному представленні сумарної кількості контактів у кожний момент симулювання.

(а) Взаємодія білка з лігандом може бути контрольована протягом усього моделювання. Білок-лігандні взаємодії (або "контакти") поділяються на чотири типи: водневі зв'язки, гідрофобні, іонні та водневі мости. Складені стовпчасті діаграми нормалізуються протягом траєкторії: наприклад, значення 0,7 означає, що 70% часу моделювання підтримується специфічна взаємодія. Можливі значення більше 1,0 оскільки деякі білкові залишки можуть утворювати кілька контактів одного і того ж підтипу з лігандом.

(б) Схема детальних взаємодій атомів лігандів із залишками білка. Взаємодії, які виникають більше ніж 30,0% часу моделювання у вибраній траєкторії (від 0,00 до 10,20 нс).

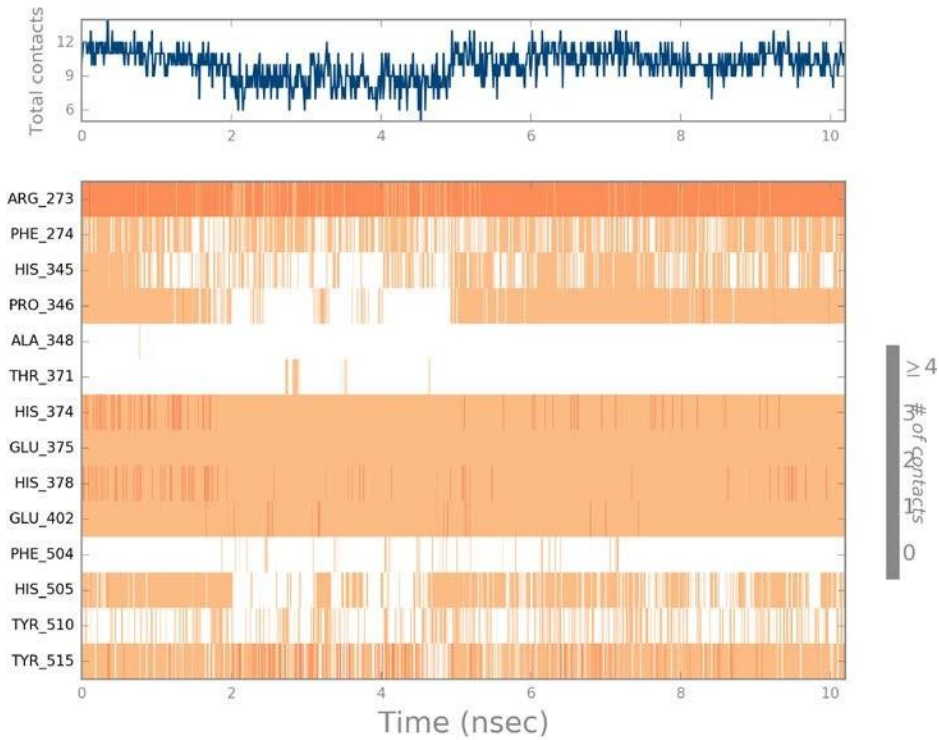
(в) Графічне представлення взаємодій та контактів (Н-зв'язки, гідрофобні, іонічні, водні мости) узагальнено на графіку. На одній панелі показано загальну кількість конкретних контактів з ліганд-білок протягом траєкторії. На нижній панелі показано, які залишки взаємодіють ліганд у кожному траєкторному кадрі. Деякі залишки здійснюють більше ніж один конкретний контакт з лігандом, який представлений темнішим відтінком оранжевого відповідно до шкали праворуч від ділянки..

На малюнках 19-24 зображені діаграми наявності контактів білок-ліганд в сайті зв'язуванні протягом періоду молекулярної динаміки



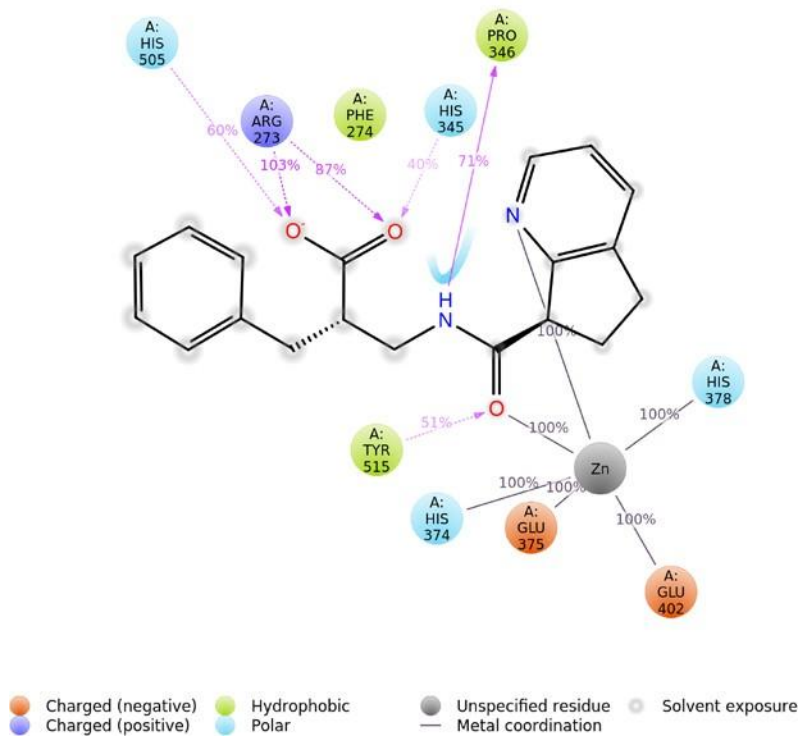
Малюнок 19. (а) Діаграма специфічних взаємодій ліганду Z2831426920 з амінокислотними залишками активного сайту протеїну ACE2.

## Protein-Ligand Contacts (cont.)



19(б)

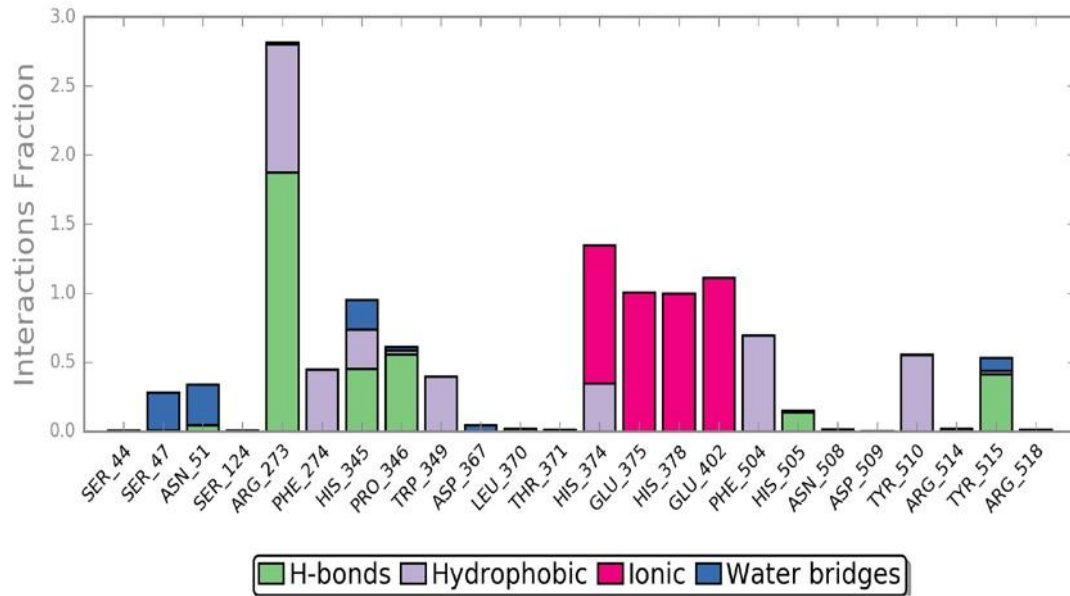
## Ligand-Protein Contacts



19(в)

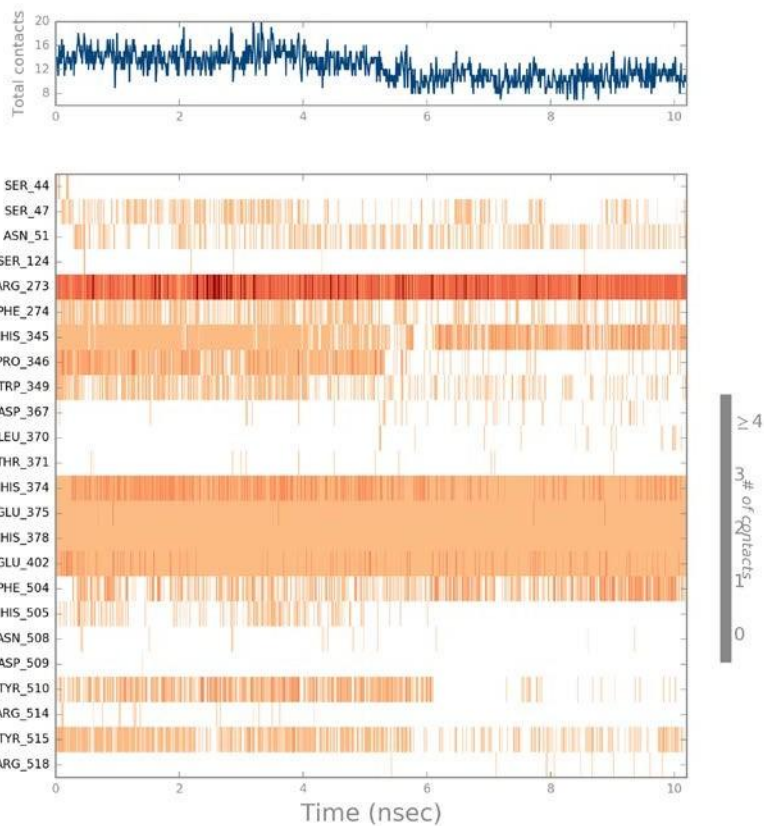
Малюнок 19. (а) Діаграма специфічних взаємодій ліганду Z283142692 з амінокислотними залишками активного сайту протеїну ACE2. (б) Схема детальних взаємодій атомів лігандів із залишками білка. (в) Узагальнене графічне представлення взаємодій та контактів (Н-зв'язки, гідрофобні, іонічні, водні мости)

## Protein-Ligand Contacts



20(a)

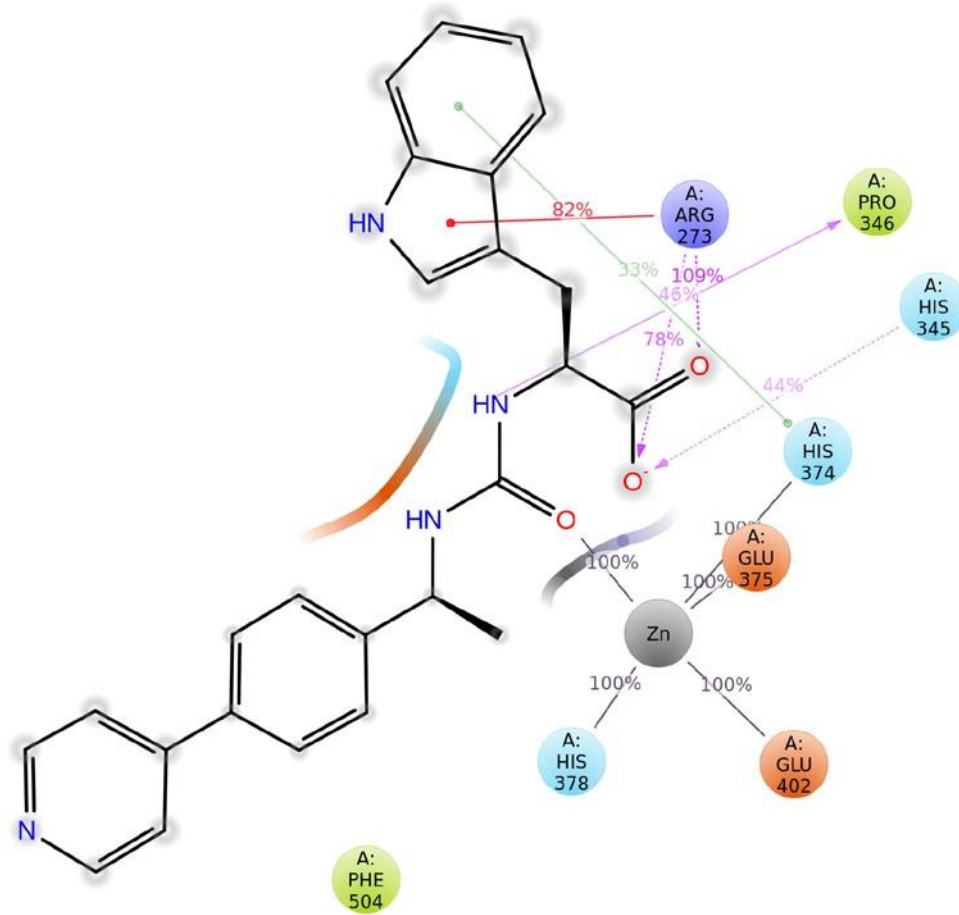
## Protein-Ligand Contacts (cont.)



20(б)

Малюнок 20. (а) Діаграма специфічних взаємодій ліганду Z3969355209 з амінокислотними залишками активного сайту протеїну ACE2. (б) Схема детальних взаємодій атомів лігандів із залишками білка.

# Ligand-Protein Contacts

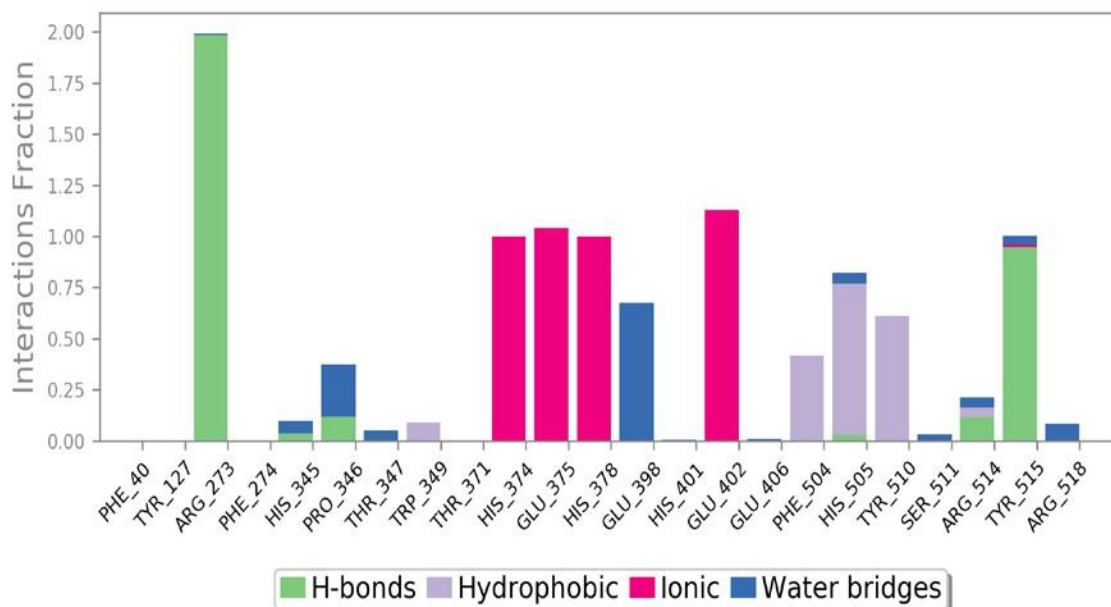


- Charged (negative)
- Polar
- Pi-Pi stacking
- Charged (positive)
- Unspecified residue
- Pi-cation
- Hydrophobic
- Metal coordination
- Solvent exposure

20(в)

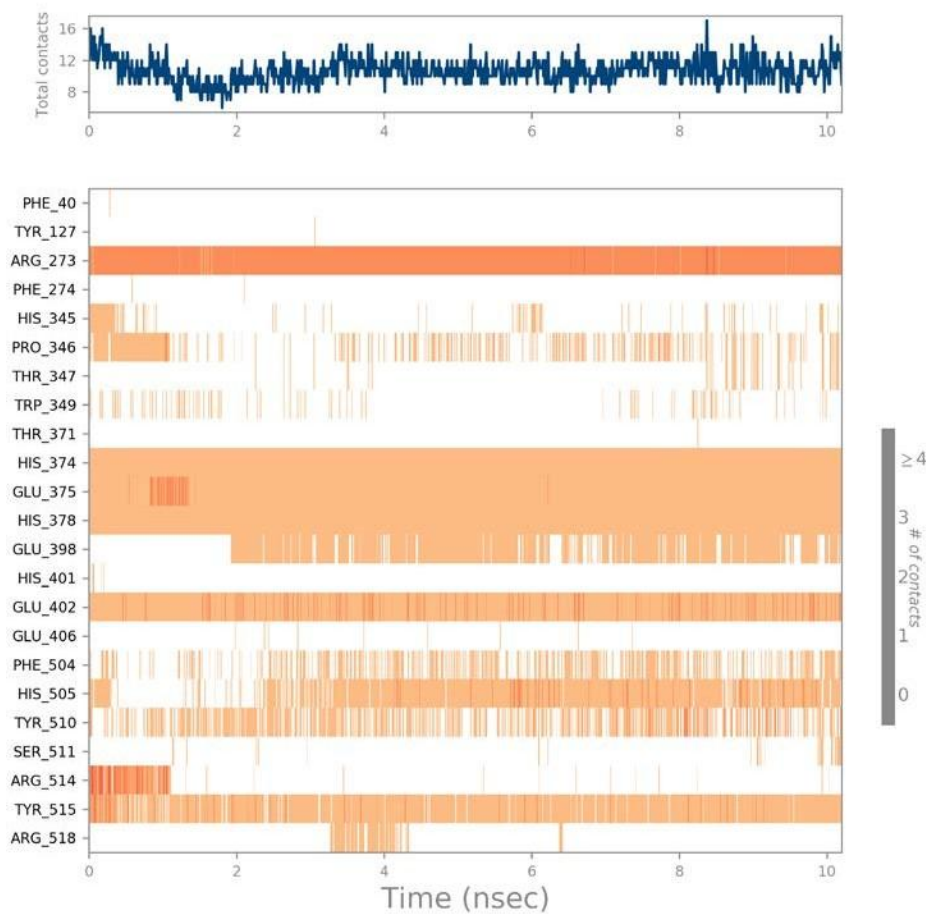
Малюнок 20. (а) Діаграма специфічних взаємодій ліганду Z3969355209 з амінокислотними залишками активного сайту протеїну ACE2. (б) Схема детальних взаємодій атомів лігандів із залишками білка. (в) Узагальнене графічне представлення взаємодій та контактів

## Protein-Ligand Contacts



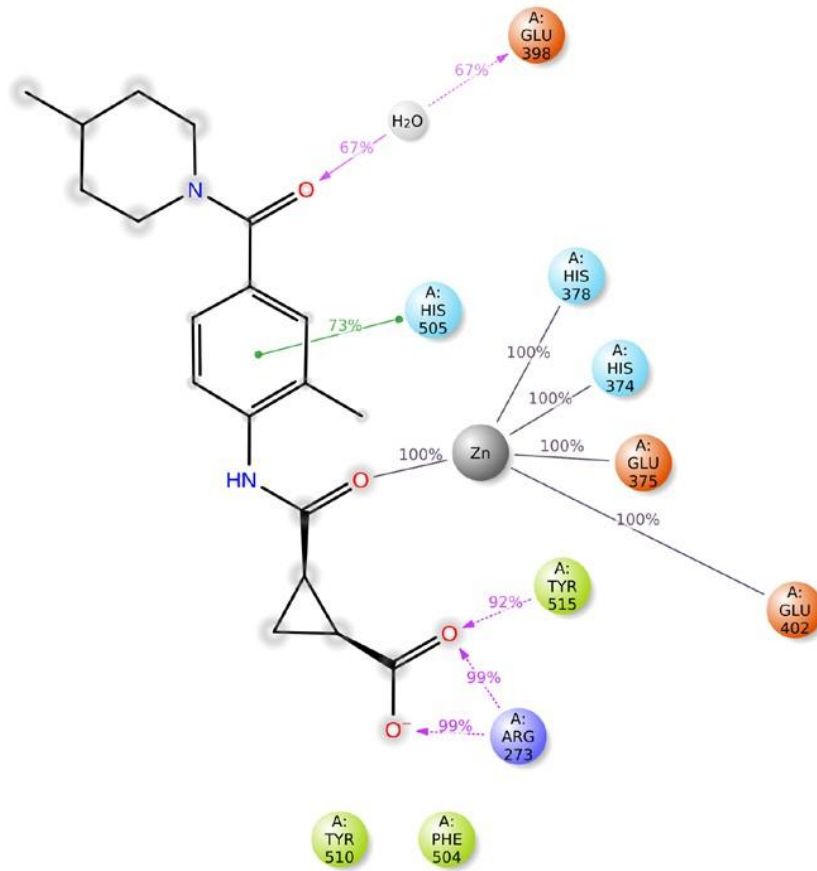
21(a)

## Protein-Ligand Contacts (cont.)



21(6)

## Ligand-Protein Contacts

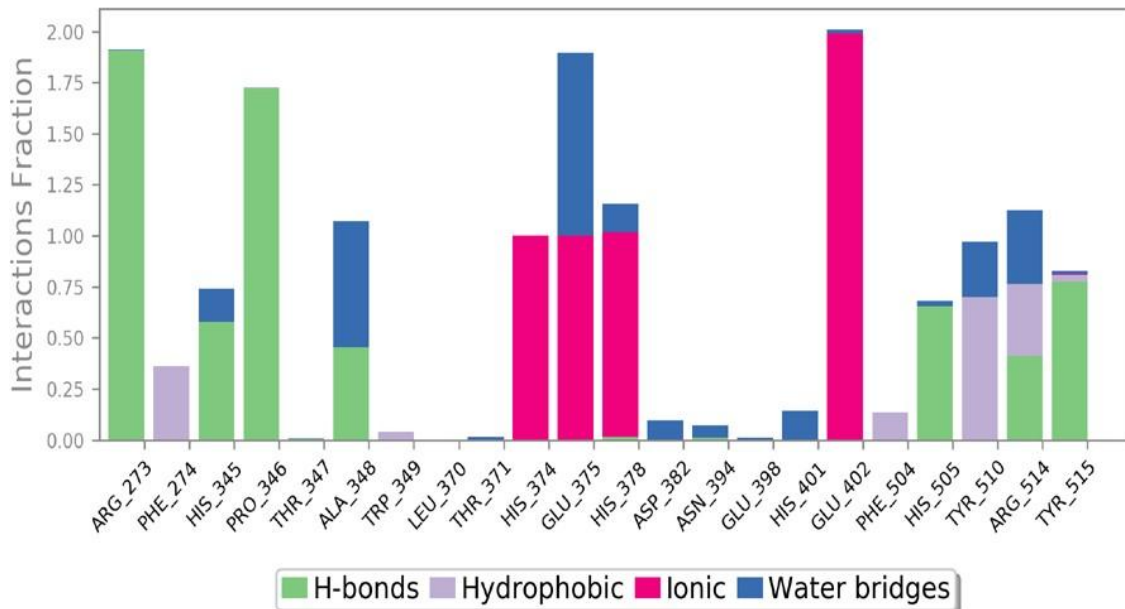


- |                    |       |                     |                    |
|--------------------|-------|---------------------|--------------------|
| Charged (negative) | Polar | Unspecified residue | Metal coordination |
| Charged (positive) | Water | Pi-Pi stacking      | Solvent exposure   |
| Hydrophobic        |       |                     |                    |

21(в)

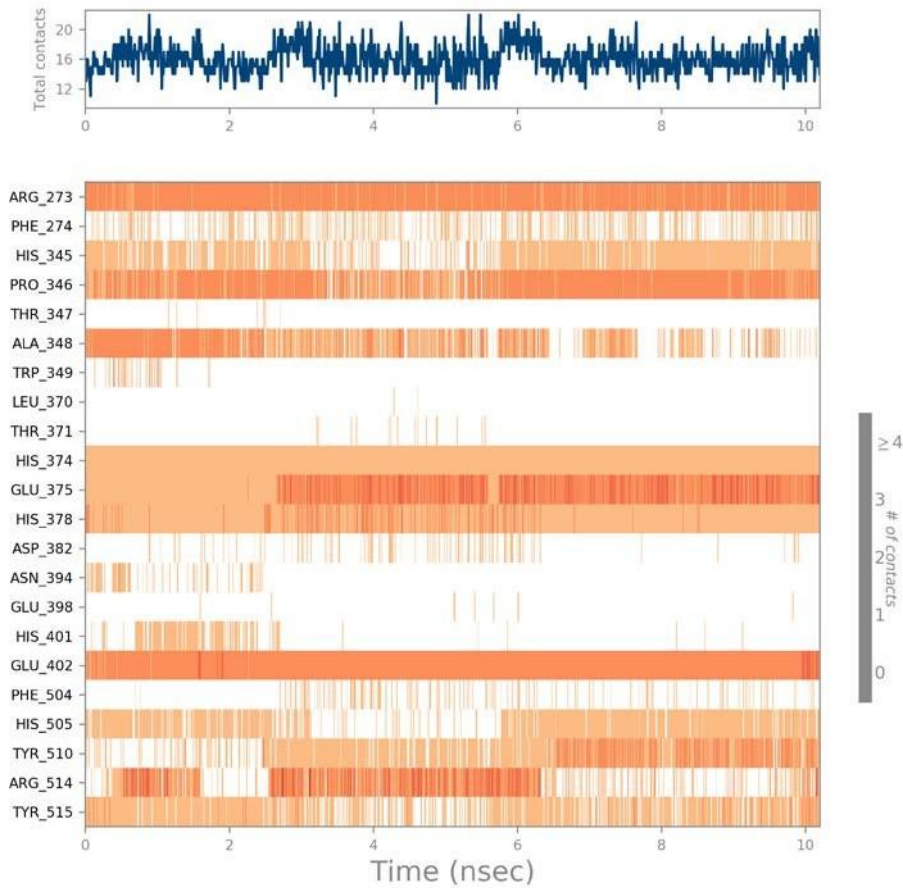
Малюнок 21. (а) Діаграма специфічних взаємодій ліганду Z1991725061 з амінокислотними залишками активного сайту протеїну ACE2. (б) Схема детальних взаємодій атомів лігандів із залишками білка. (в) Узагальнене графічне представлення взаємодій та контактів

## Protein-Ligand Contacts



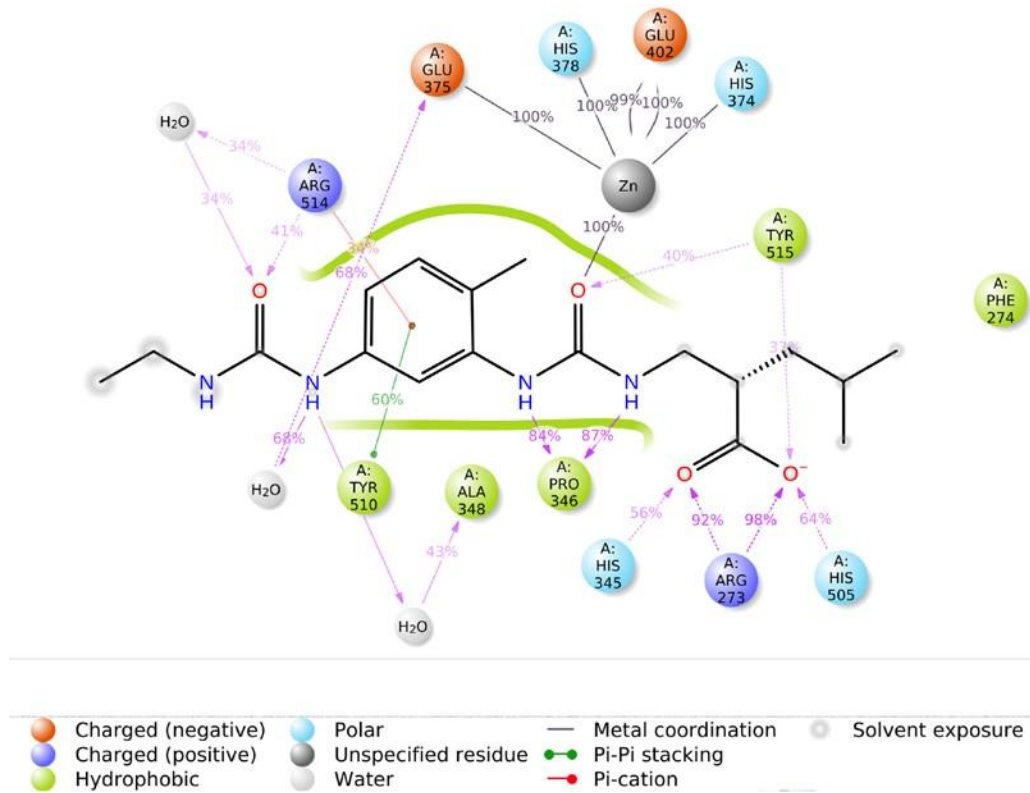
22(a)

## Protein-Ligand Contacts (cont.)



22(6)

## Ligand-Protein Contacts



22(в)

Малюнок 22. (а) Діаграма специфічних взаємодій ліганду Z2396153450 з амінокислотними залишками активного сайту протеїну ACE2. (б) Схема детальних взаємодій атомів лігандів із залишками білка. (в) Узагальнене графічне представлення взаємодій та контактів

Як бачимо всі сполуки мають в кожний момент траєкторії по 7-8 та більше контактів, тому вважаю сполуки мають великий потенціал бути активними під час перевірки *in vitro*.

Серед топ1000 сполук за завданням керувника проекту вибрали сет з 180 сполук з максимльним збереженням різноманітності. Всі переліки сполук передані для синтезування та перевірки у лабораторії їх фактичної активності *in vitro*.

## ВИСНОВКИ

В роботі досліджено відмінності активних сайтів цільового білку ACE2 та ACE як «антимішені». На цих відмінностях побудована докінгова модель, яка перевірена на відомих референтних сполуках з високим Enrichment Factor. Далі модель використовували для віртуального скринінгу баз сполук – стокова (3,2 мільйони сполук), та Real Data Base (14,1 мільярдів сполук).

За допомогою багатоступеневого скринінгу, який вміщує різні хемоінформатичні інструменти, такі як:

- пошук по підструктурі;
- докінг с трьома рівнями точності, а саме - HTVS, SP та XP;
- диверсифікацію;
- пошук за подібністю;
- медхімічні фільтри;
- молекулярна динаміка

В результаті відібрано по 20 сполук зі стокової бази та з Real Data Base.

Провели перевірку міцності зв'язку з таргетом трьох сполук з кожного сету за допомогою молекулярної динаміки періодом 10,2 нс. Всі 6 сполук мають в кожний момент траєкторії по 7-8 та більше контактів, тому вважаємо, що сполуки мають великий потенціал бути активними під час перевірки *in vitro*.

Всі переліки сполук передані для синтезування та перевірки у лабораторії їх фактичної активності.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). *J Gen Intern Med.* 2020 May;35(5):1545-1549. doi: 10.1007/s11606-020-05762-w. Epub 2020 Mar 4. PMID: 32133578; PMCID: PMC7088708.
2. COVID-19 Map - Johns Hopkins Coronavirus Resource Center (jhu.edu)  
<https://coronavirus.jhu.edu/map.html>
3. Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, Marcel A. Müller, Christian Drosten, Stefan Pöhlmann SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, *Cell*, Volume 181, Issue 2, 2020, Pages 271-280.e8, ISSN 0092-8674, <https://doi.org/10.1016/j.cell.2020.02.052>.
4. Substrate-Based Design of the First Class of Angiotensin-Converting Enzyme-Related Carboxypeptidase (ACE2) Inhibitors Natalie A. Dales,† Alexandra E. Gould,† James A. Brown,† Emily F. Calderwood,† Bing Guan,† Charles A. Minor,† James M. Gavin,‡ Paul Hales,‡ Virendar K. Kaushik,§ Michael Stewart,‡ Peter J. Tummino,§ Chad S. Vickers,§ Timothy D. Ocain,† and Michael A. Patane\*,† Millennium Pharmaceuticals, Inc., 75 Sidney Street, Cambridge, Massachusetts 02139 Received July 16, 2002
5. Vickers, C.; Hales, P.; Kaushik, V.; Dick, L.; Gavin, J.; Tang, J.; Godbout, K.; Parsons, T.; Baronas, E.; Hsieh, F.; Acton, S.; Patane, M.; Nichols, A.; Tummino, P. J. *Biol. Chem.* 2002, 277, 14838.
6. Sonja Groß, Christopher Jahn, Sarah Cushman, Christian Bär, Thomas Thum, SARS-CoV-2 receptor ACE2-dependent implications on the cardiovascular system: From basic science to clinical implications, *Journal of Molecular and Cellular Cardiology*, Volume 144, 2020, Pages 47-53, ISSN 0022-2828, <https://doi.org/10.1016/j.yjmcc.2020.04.031>.

7. Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System Mahmoud Gheblawi, Kaiming Wang, Anissa Viveiros, Quynh Nguyen, Jiu-Chang Zhong, Anthony J. Turner, Mohan K. Raizada, Maria B. Grant, Gavin Y. Oudit Originally published 8 Apr 2020  
<https://doi.org/10.1161/CIRCRESAHA.120.317015> *Circulation Research*. 2020;126:1456–1474
8. Joshi S, Balasubramanian N, Vasam G, Jarajapu YP. Angiotensin converting enzyme versus angiotensin converting enzyme-2 selectivity of MLN-4760 and DX600 in human and murine bone marrow-derived cells. *Eur J Pharmacol*. 2016 Mar 5;774:25-33. doi: 10.1016/j.ejphar.2016.01.007. Epub 2016 Feb 3. PMID: 26851370; PMCID: PMC4804635.
9. Raz A, Gamliel-Lazarovich A, Bogner I, Strigevsky A, Keidar S. [The importance of ACE2 in regulating the cardiovascular system]. *Harefuah*. 2007 Sep;146(9):703-6, 733. Hebrew. PMID: 17969309.
10. Guy JL, Jackson RM, Acharya KR, Sturrock ED, Hooper NM, Turner AJ. Angiotensin-converting enzyme-2 (ACE2): comparative modeling of the active site, specificity requirements, and chloride dependence. *Biochemistry*. 2003 Nov 18;42(45):13185-92. doi: 10.1021/bi035268s. PMID: 14609329.
11. *Biochemistry* 2003, 42, 45, 13185–13192 Publication Date: October 21, 2003  
<https://doi.org/10.1021/bi035268s>
12. Guy JL, Jackson RM, Acharya KR, Sturrock ED, Hooper NM, Turner AJ. Angiotensin-converting enzyme-2 (ACE2): comparative modeling of the active site, specificity requirements, and chloride dependence. *Biochemistry*. 2003 Nov 18;42(45):13185-92. doi: 10.1021/bi035268s. PMID: 14609329.
13. Guang C, Phillips RD, Jiang B, Milani F. Three key proteases--angiotensin-I-converting enzyme (ACE), ACE2 and renin--within and beyond the renin-angiotensin system. *Arch Cardiovasc Dis*. 2012 Jun-Jul;105(6-7):373-85. doi:

- 10.1016/j.acvd.2012.02.010. Epub 2012 Jun 15. PMID: 22800722; PMCID: PMC7102827.
14. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des.* 2013 Mar;27(3):221-34. doi: 10.1007/s10822-013-9644-8. Epub 2013 Apr 12. PMID: 23579614.
  15. Schrödinger Suite 2012 Protein Preparation Wizard; Epik version 2.3, Schrödinger, LLC, New York, NY, 2012; Impact version 5.8, Schrödinger, LLC, New York, NY, 2012; Prime version 3.1, Schrödinger, LLC, New York, NY, 2012.
  16. Protein pKa predictions were performed using PROPKA. Olsson, M. H. M.; Søndergard, C. R.; Rostkowski, M.; Jensen J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa predictions. *J. Chem. Theor. Comput.*, 2011, 7, 525–537.
  17. Shoichet, B., Mysinger, MM, Carchia, M., Irwin, JJ, & Shoichet, BK. (2012). Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *UCSF*. Retrieved from <https://escholarship.org/uc/item/3j66v5w2>
  18. Kirkpatrick P. Virtual screening: gliding to success. *Nat Rev Drug Disc* 2004;3:299.
  19. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). *Journal of Computer-Aided Molecular Design*, 11(5), 425–445. doi:10.1023/a:1007996124545 (<https://doi.org/10.1023/a:1007996124545>)
  20. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, 47(7), 1750–1759. doi:10.1021/jm030644s (<https://doi.org/10.1021/jm030644s>)

21. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739–1749. doi:10.1021/jm0306430 (<https://doi.org/10.1021/jm0306430>)
22. Maestro User Manual. Schrodinger 2019-4 Release
23. Terali, K., Baddal, B., & Gülcan, H. O. (2020). Prioritizing potential ACE2 inhibitors in the COVID-19 pandemic: insights from a molecular mechanics-assisted structure-based virtual screening experiment. *Journal of Molecular Graphics and Modelling*, 107697. doi:10.1016/j.jmgm.2020.107697 (<https://doi.org/10.1016/j.jmgm.2020.107697>)
24. Angiotensin-Converting Enzyme-2 (ACE2): Comparative Modeling of the Active Site, Specificity Requirements, and Chloride Dependence† Jodie L. Guy,‡ Richard M. Jackson,‡ K. Ravi Acharya,§ Edward D. Sturrock,| Nigel M. Hooper,‡ and Anthony J. Turner\*,‡ School of Biochemistry and Molecular Biology, UniVersity of Leeds, Leeds, LS2 9JT, United Kingdom, Department of Biology and Biochemistry, UniVersity of Bath, ClaVerton Down, Bath, BA2 7AY, United Kingdom, and DiVision of Medical Biochemistry and MRC/UCT LiVer Research Centre, UniVersity of Cape Town Medical School, ObserVatory 7925, South Africa ReceiVed July 18, 2003; ReVised Manuscript ReceiVed September 3, 2003
25. Muchtaridi, M., Fauzi, M., Khairul Ikram, N. K., Mohd Gazzali, A., & Wahab, H. A. (2020). Natural Flavonoids as Potential Angiotensin-Converting Enzyme 2 Inhibitors for Anti-SARS-CoV-2. *Molecules*, 25(17), 3980. doi:10.3390/molecules25173980 (<https://doi.org/10.3390/molecules25173980>)
26. Selective Binding of Lisinopril in Angiotensin-Converting Enzyme Domains  
Poster Team: Lexiee Carpenter, Gail Kingsbury, and Patrick Tierney Jmol Team: Jennifer Dippel, Whitney Hendrickson, and Sarah Hunt Faculty Advisor and

Mentor: Daniel S. Sem, Ph.D. Concordia University School of Pharmacy, 12800 North Lake Shore Drive, Mequon, WI 53097

27. Baell, J. B., & Holloway, G. A. (2010). New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry*, 53(7), 2719–2740. doi:10.1021/jm901137j (<https://doi.org/10.1021/jm901137j>)
28. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results André Fischer, Martin Smiesko, Manuel Sellner, and Markus A. Lill <https://doi.org/10.1021/acs.jmedchem.0c02227>
29. Ashton, M. et al. “Identification of Diverse Database Subsets using Property-Based and Fragment-Based Molecular Descriptions.” *Quantitative Structure-Activity Relationships* 21:598-604 (2002).
30. Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, “Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters,” Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, November 11-17, 2006.
31. <https://www.cgl.ucsf.edu/chimera/current/docs/UsersGuide/framecontrib.html>
32. <https://www.cgl.ucsf.edu/chimera/current/docs/UsersGuide/quickref.pdf>

## ДОДАТКИ

### Додаток 1. Перелік референтних структур – інгібіторів ACE2

Smiles Molecule ChEMBL ID

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)O CHEMBL253224

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(Oc2cccc2)c1)C(=O)O CHEMBL252003

CCCC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc2)cc1)C(=O)O CHEMBL257270

CC[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc2)cc1)C(=O)O CHEMBL257726

O=C(CS)N[C@@H](Cc1ccc(-c2cccc2)cc1)C(=O)O CHEMBL258333

O=C(O)[C@H](Cc1ccc(-c2cccc2)cc1)NC(=O)[C@H](S)Cc1cccc1 CHEMBL163454

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc1O)Cc1cccc1)C(=O)O CHEMBL254493

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc2cccc12)C(=O)O CHEMBL400527

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc(F)cc2F)cc1)C(=O)O CHEMBL398771

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc(F)cc2)cc1)C(=O)O CHEMBL254911

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](C)C(=O)O CHEMBL254900

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc2)cc1)C(=O)O CHEMBL252391

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2cc(C(F)(F)F)cc(C(F)(F)F)c2)cc1)C(=O)O CHEMBL252009

CC[C@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc2)cc1)C(=O)O CHEMBL269996

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(-c2cccc2)c1)C(=O)O CHEMBL401397

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc1)C(=O)O CHEMBL400526

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc(OCc2cccc2)c1)C(=O)O CHEMBL254495

CC(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2cccc2)cc1)C(=O)O CHEMBL271225

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1cccc1-c1cccc1)C(=O)O CHEMBL398545

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(Oc2cccc2)cc1)C(=O)O CHEMBL254282

O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](S)C1CCCC1 CHEMBL404117

O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H](S)Cc1ccccc1 CHEMBL350414

O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H](S)CC1CCCCC1 CHEMBL271224

CC(C)(C)C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O CHEMBL271223

O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](S)C1CCCCC1 CHEMBL257026

CC(=O)N[C@@H](Cc1c[nH]cn1)C(=O)N[C@H](CC(C)C)P(=O)(O)CC(Cc1cc(-c2ccccc2)no1)C(=O)O  
CHEMBL408182

CC(C)C[C@H](NC(=O)[C@@H]1CCCN1C(=O)[C@H](CO)NC(=O)[C@H](CS)NC(=O)[C@H](Cc1cnc[nH]1)NC(=O)[  
C@H](CO)NC(=O)[C@H](Cc1ccc(O)cc1)NC(=O)[C@@H](CC(=O)O)NC(=O)CN)C(=O)N[C@@H](CCNC(=N)N)C(=  
O)N[C@@H](Cc1ccc(O)cc1)C(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)N1CCC[C@H]1C(=O)N[C@@H](Cc1c[nH]c2ccc  
cc12)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CS)C(=O)N[C@@H](C(=O)N[  
C@@H](Cc1ccc(O)cc1)C(=O)N1CCC[C@H]1C(=O)N[C@@H](CC(=O)O)C(=O)N1CCC[C@H]1C(=O)O)[C@@H](C  
O CHEMBL1240682

CC(=O)N[C@@H](CC(C)C)C(=O)N1CCC[C@H]1P(=O)(O)CC(Cc1cc(-c2ccccc2)no1)C(=O)O  
CHEMBL261121

CC(C)C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)OCHEMBL405232

CC(C)(S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O CHEMBL404044

C[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)OCHEMBL257727

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc(C(F)(F)F)cc2)cc1)C(=O)O CHEMBL251809

CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccccc1Oc1ccccc1)C(=O)O CHEMBL251804

CC(C)C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL264665

O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H](S)Cc1ccc2ccccc2c1 CHEMBL272925

O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](S)c1ccccc1 CHEMBL437595

O=C(O)[C@H](Cc1ccccc1)CP(=O)(O)[C@H]1CCCN1 CHEMBL261033

CC(=O)N[C@@H](Cc1c[nH]cn1)C(=O)N1CCC[C@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O  
CHEMBL261423

CC(=O)N[C@@H](CCCCN)C(=O)N1CCC[C@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL411942

C[C@@H](N)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL405913

N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL412123

CC(=O)N[C@@H](CC(C)C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL260273  
C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O CHEMBL409713  
CC(=O)N[C@@H](CCC(=O)O)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O  
CHEMBL263663  
CC(=O)N[C@@H](Cc1c[nH]cn1)C(=O)N[C@H](CC(C)C)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O  
CHEMBL408448  
CC(=O)N[C@H](C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O)C(C)C CHEMBL411052  
CC(=O)N[C@@H](Cc1ccc(O)cc1)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O  
CHEMBL410509  
CC(C)C[C@H](CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1)C(=O)O CHEMBL411298  
CC(=O)N[C@@H](Cc1c[nH]cn1)C(=O)N1CCC[C@@H]1P(=O)(O)CC(Cc1cc(-c2ccccc2)no1)C(=O)O  
CHEMBL409105  
CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccccc2)cc1)C(=O)O CHEMBL254703  
CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc2ccccc2c1)C(=O)O CHEMBL253428  
CC[C@@H](C)[C@H](S)C(=O)N[C@@H](Cc1ccc(OCc2ccc(F)c(F)c2)cc1)C(=O)O CHEMBL251808  
CC[C@@H](C)[C@H](S)C(=O)NCC(=O)O CHEMBL401086  
CC(=O)N[C@@H](CC(C)C)C(=O)N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL408705  
CC(=O)N[C@@H](Cc1ccccc1)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O  
CHEMBL258698  
O=C(N[C@H](Cc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O)OCc1ccccc1 CHEMBL258464  
CC(=O)N[C@@H](C)C(=O)N1CCC[C@@H]1P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL409721  
O=C(O)[C@H](Cc1ccccc1)CP(=O)(O)[C@H]1CCCN1C(=O)OCc1ccccc1 CHEMBL260677  
C[C@@H](NC(=O)OCc1ccccc1)P(=O)(O)C[C@@H](Cc1ccccc1)C(=O)O CHEMBL258683  
O=C(N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O)[C@@H](S)CCc1ccccc1 CHEMBL402987  
O=C(O)[C@H](Cc1ccc(-c2ccccc2)cc1)NC(=O)[C@@H](S)C1CCC1 CHEMBL257229  
CC(C)[C@H](S)C(=O)N[C@@H](Cc1ccc(-c2ccccc2)cc1)C(=O)O CHEMBL269997  
CC(=O)CNC(=O)[C@H](CC(=O)O)NC(=O)[C@H](Cc1ccc(O)cc1)NC(=O)[C@H](CO)NC(=O)[C@H](Cc1c[nH]cn1)N  
C(=O)[C@@H]1CSSC[C@H](NC(=O)[C@@H](NC(=O)[C@H](Cc2ccc(O)cc2)NC(=O)[C@@H]2CCCN2C(=O)[C@H]  
(CC(=O)O)NC(=O)[C@@H]2CCCN2C(=O)[C@H](CCC(=O)O)NC(=O)CNC(=O)CNC(=O)CN[C@@H](C)O)C(=O)N[C

@@H](CCCCN)C(=O)N[C@@H](Cc2c[nH]c3cccc23)C(=O)N[C@@H](Cc2c[nH]c3cccc23)C(=O)N2CCC[C@H]2  
C(=O)N[C@@H](Cc2ccc(O)cc2)C(=O)N[C@@H](Cc2ccc(O)cc2)C(=O)N[C@@H](CCNC(=N)N)C(=O)N[C@@H](  
CC(C)C)C(=O)N2CCC[C@H]2C(=O)N[C@@H](CO)C(=O)N1 CHEMBL436639

Cc1nnc(S/C(=C/c2ccc(-c3cc(Cl)ccc3C)o2)C(=O)O)[nH]1 CHEMBL252417

CC(C)C[C@H](N[C@@H](Cc1cncn1Cc1cc(Cl)cc(Cl)c1)C(=O)O)C(=O)O CHEMBL429844

CNC(=O)[C@H](Cc1c[nH]cn1)NC(=O)CN(CCCc1cccc1)CC(=O)O CHEMBL3235416

NCCNCCN1CC1 CHEMBL398940

## Додаток 2. Enrichment Reports

### Додаток 2 (а) Грід-1 – HTVS

=====

Enrichment Report

-----

Actives file: 3.csv

Results: enrichment\_ACE2\_ARGHB\_HTVS\_enrichment\_from\_project\_selection.maegz

Total actives: 62

Total ligands(actives+decoys): 6262

Number of ranked actives: 35

BEDROC(alpha=160.9, alpha\*Ra=1.5931): 0.293

BEDROC(alpha=20.0, alpha\*Ra=0.1980): 0.482

BEDROC(alpha=8.0, alpha\*Ra=0.0792): 0.527

ROC: 0.56

RIE: 8.75

Area under accumulation curve: 0.76

Ave. Number of outranking decoys: 70

Minimum Tc over all active pairs: 0.033

Count and percentage of actives in top N% of decoy results.

% Decoys | 1%| 2%| 5%| 10%| 20%|

# Actives | 22| 30| 35| 35| 35|

% Actives | 35.5| 48.4| 56.5| 56.5| 56.5|

Count and percentage of actives in top N% of results.

% Results | 1% | 2% | 5% | 10% | 20% |

# Actives | 16 | 26 | 34 | 35 | 35 |

% Actives | 25.8 | 41.9 | 54.8 | 56.5 | 56.5 |

Enrichment Factors with respect to N% sample size.

% Sample | 1% | 2% | 5% | 10% | 20% |

EF | 26 | 21 | 11 | 5.6 | 2.8 |

EF\* | 35 | 24 | 11 | 5.6 | 2.8 |

EF' | 32 | 30 | 20 | 11 | 5.7 |

DEF | 25 | 20 | 11 | 5.6 | 2.8 |

DEF\* | 35 | 24 | 11 | 5.6 | 2.8 |

DEF' | 31 | 29 | 20 | 11 | 5.7 |

Eff | 0.945 | 0.921 | 0.837 | 0.699 | 0.477 |

Enrichment Factors with respect to N% actives recovered.

% Actives | 40% | 50% | 60% | 70% | 80% | 90% | 100% |

EF | 24 | 19 | n/a | n/a | n/a | n/a | n/a |

EF\* | 31 | 22 | n/a | n/a | n/a | n/a | n/a |

EF' | 32 | 29 | n/a | n/a | n/a | n/a | n/a |

FOD | 0.006 | 0.008 | n/a | n/a | n/a | n/a | n/a |

## Додаток 2 (б) Грід-1 – SP

=====

Enrichment Report

-----

Actives file: 3.csv

Results: enrichment\_ACE2\_ARGHB\_SP\_enrichment\_from\_project\_selection.maegz

Total actives: 62

Total ligands(actives+decoys): 6262

Number of ranked actives: 58

BEDROC(alpha=160.9, alpha\*Ra=1.5931): 0.155

BEDROC(alpha=20.0, alpha\*Ra=0.1980): 0.372

BEDROC(alpha=8.0, alpha\*Ra=0.0792): 0.568

ROC: 0.86

RIE: 6.74

Area under accumulation curve: 0.87

Ave. Number of outranking decoys: 525

Minimum Tc over all active pairs: 0.027

Count and percentage of actives in top N% of decoy results.

% Decoys | 1%| 2%| 5%| 10%| 20%|

# Actives | 7| 12| 25| 40| 54|

% Actives | 11.3| 19.4| 40.3| 64.5| 87.1|

Count and percentage of actives in top N% of results.

% Results | 1%| 2%| 5%| 10%| 20%|

# Actives | 7| 11| 24| 38| 52|

% Actives | 11.3| 17.7| 38.7| 61.3| 83.9|

Enrichment Factors with respect to N% sample size.

% Sample | 1%| 2%| 5%| 10%| 20%|

EF | 11| 8.9| 7.7| 6.1| 4.2|

EF\* | 11| 9.7| 8.1| 6.5| 4.4|

EF' | 16| 13| 10| 8.5| 6.5|

DEF | 11| 8.6| 7.5| 6| 4.2|

DEF\* | 11| 9.4| 7.8| 6.3| 4.4|

DEF' | 16| 12| 9.7| 8.3| 6.5|

Eff | 0.837| 0.813| 0.779| 0.732| 0.627|

Enrichment Factors with respect to N% actives recovered.

% Actives | 40%| 50%| 60%| 70%| 80%| 90%| 100%|

EF | 7.8| 7.5| 6.5| 5.6| 4.7| 2.9| n/a|

EF\* | 8.3| 8.1| 7.3| 6| 4.9| 3| n/a|

EF' | 10| 9.6| 9.1| 8.1| 7.1| 5| n/a|

FOD | 0.02| 0.03| 0.04| 0.04| 0.06| 0.08| n/a|

## Додаток 2 (в) Грід-2 – HTVS

=====

Enrichment Report

-----

Actives file: 3.csv

Results: enrichment\_ACE2\_ARGPos\_HTVS\_enrichment\_from\_project\_selection.maegz

Total actives: 62

Total ligands(actives+decoys): 6262

Number of ranked actives: 26

BEDROC(alpha=160.9, alpha\*Ra=1.5931): 0.382

BEDROC(alpha=20.0, alpha\*Ra=0.1980): 0.400

BEDROC(alpha=8.0, alpha\*Ra=0.0792): 0.410

ROC: 0.42

RIE: 7.26

Area under accumulation curve: 0.69

Ave. Number of outranking decoys: 37

Minimum Tc over all active pairs: 0.04

Count and percentage of actives in top N% of decoy results.

% Decoys | 1%| 2%| 5%| 10%| 20%|

# Actives | 21| 24| 26| 26| 26|

% Actives | 33.9| 38.7| 41.9| 41.9| 41.9|

Count and percentage of actives in top N% of results.

% Results | 1%| 2%| 5%| 10%| 20%|

# Actives | 21| 24| 26| 26| 26|

% Actives | 33.9| 38.7| 41.9| 41.9| 41.9|

Enrichment Factors with respect to N% sample size.

% Sample | 1%| 2%| 5%| 10%| 20%|

EF | 34| 19| 8.4| 4.2| 2.1|

EF\* | 34| 19| 8.4| 4.2| 2.1|

EF' | 55| 36| 16| 8.5| 4.2|

DEF | 34| 19| 8.4| 4.2| 2.1|

DEF\* | 34| 19| 8.4| 4.2| 2.1|

DEF' | 54| 36| 16| 8.5| 4.2|

Eff | 0.943| 0.902| 0.787| 0.615| 0.354|

Enrichment Factors with respect to N% actives recovered.

% Actives | 40%| 50%| 60%| 70%| 80%| 90%| 100%|

EF | 13| n/a| n/a| n/a| n/a| n/a| n/a|

EF\* | 15| n/a| n/a| n/a| n/a| n/a| n/a|

EF' | 28| n/a| n/a| n/a| n/a| n/a| n/a|

FOD | 0.005| n/a| n/a| n/a| n/a| n/a| n/a|

## Додаток 2 (г) Грід-2 – SP

=====  
Enrichment Report  
-----

Actives file: 3.csv

Results: enrichment\_ACE2\_ARGPos\_SP\_enrichment\_from\_project\_selection.maegz

Total actives: 62

Total ligands(actives+decoys): 6262

Number of ranked actives: 59

BEDROC(alpha=160.9, alpha\*Ra=1.5931): 0.439

BEDROC(alpha=20.0, alpha\*Ra=0.1980): 0.652

BEDROC(alpha=8.0, alpha\*Ra=0.0792): 0.781

ROC: 0.92

RIE: 11.83

Area under accumulation curve: 0.94

Ave. Number of outranking decoys: 197

Minimum Tc over all active pairs: 0.027

Count and percentage of actives in top N% of decoy results.

% Decoys | 1% | 2% | 5% | 10% | 20% |

# Actives | 26 | 31 | 50 | 53 | 59 |

% Actives | 41.9 | 50.0 | 80.6 | 85.5 | 95.2 |

Count and percentage of actives in top N% of results.

% Results | 1%| 2%| 5%| 10%| 20%|

# Actives | 21| 27| 48| 53| 58|

% Actives | 33.9| 43.5| 77.4| 85.5| 93.5|

Enrichment Factors with respect to N% sample size.

% Sample | 1%| 2%| 5%| 10%| 20%|

EF | 34| 22| 15| 8.6| 4.7|

EF\* | 42| 25| 16| 8.5| 4.8|

EF' | 67| 43| 26| 16| 8.9|

DEF | 32| 21| 15| 8.6| 4.7|

DEF\* | 40| 24| 16| 8.5| 4.8|

DEF' | 65| 42| 26| 16| 8.9|

Eff | 0.953| 0.923| 0.883| 0.791| 0.653|

Enrichment Factors with respect to N% actives recovered.

% Actives | 40%| 50%| 60%| 70%| 80%| 90%| 100%|

EF | 35| 20| 17| 17| 15| 7.1| n/a|

EF\* | 53| 25| 23| 21| 18| 7.6| n/a|

EF' | 75| 44| 36| 31| 27| 14| n/a|

FOD | 0.003| 0.005| 0.008| 0.01| 0.02| 0.02| n/a|

### Додаток 3. Скрипт для диверсифікації бібліотеки сполук

```
from rdkit.Chem import AllChem
from rdkit import Chem
from rdkit import SimDivFilters, DataStructs
import numpy as np
from rdkit.Chem.SaltRemover import SaltRemover
import sys
import os

in_file = os.path.basename(sys.argv[1])
basename = os.path.splitext(in_file)[0]
out_file = str(basename) + '.out'
print('IN: {} OUT: {} BASE: {}'.format(in_file, out_file, basename))

ids_label = 'ID'
smi_label = 'SMILES'
radius = 4
nBits = 1024
separator = '\t'
final_size = 30000

fps = []
ids = []
mmp = SimDivFilters.MaxMinPicker()
remover = SaltRemover()
```

```

print('open file {}'.format(in_file))
count = 0
suppl = Chem.SmilesMolSupplier(in_file,delimiter=separator,titleLine=True)
for mol in suppl:
    if mol:
        mol = remover.StripMol(mol)
        # fps.append(AllChem.GetMorganFingerprint(mol, radius))
        fps.append(AllChem.GetMorganFingerprintAsBitVect(mol, radius, nBits))
        ids.append(mol.GetProp('{}'.format("_Name")))
        count += 1
        if count % 10000 == 0:
            print('{} structures done'.format(count))
print('{} structures loaded '.format(len(ids)))
nfps = len(fps)

print('Start Diversity Picking')
bv_ids = mmp.LazyBitVectorPick(fps, nfps, final_size)
#bv_ids = mmp.LazyPick(fps, nfps, final_size)
print('Diversity Picking done')

print('Analyzing results')
fps_f = []
for i in range(len(bv_ids)):
    fps_f.append(fps[bv_ids[i]])
nns = []

```

```

fps_1 = []
for i in range(len(bv_ids)):
    fps_1 = fps_f[:i] + fps_f[i+1:]
    tans = DataStructs.BulkTanimotoSimilarity(fps[bv_ids[i]], fps_1)
    nns.append(max(tans))
mean = np.mean(nns)
dev = np.std(nns)
print("{} structure, {:.7f} max NNS, NNS Mean: {:.7f}, std dev: {:.7f}".
      format(len(bv_ids), max(nns), mean, dev))

print('Writing to file {}'.format(out_file))

with open(out_file, 'w', buffering=262144) as fout:
    fout.write('{}\n'.format(ids_label))
    for i in range(len(bv_ids)):
        j = bv_ids[i]
        line = ids[j] + '\n'
        fout.write(line)
print('Done')

```

## Додаток 4. Скрипт для Shape Based Searching

```
import oddt
from oddt import shape
from rdkit import Chem
from rdkit.Chem import AllChem
import pickle
import os
import shutil
import multiprocessing as mp
from tqdm import tqdm
import time
import sys
import gzip
from rdkit.Chem.SaltRemover import SaltRemover

base_folder = '/media/dudenko/TRANCHES/REAL_DB/Bases'
headers = "SMILES\tidnumber\tsimilarity_usr\tsimilarity_usr_cat\tsimilarity_electroshape\n"
min_sim = 0.4
result_file = f'name_min_sim_{min_sim}.smi'
#smiles = '&&&&'
SDFFile = 'filename.sdf'

remover = SaltRemover()

def initializerSD(SDFFile):
    global shape1
    global shape2
    global shape3
    mol = next(oddt.toolkit.readfile('sdf', SDFFile))
```

```
mol.addh()
shape1 = oddt.shape.usr(mol)
shape2 = oddt.shape.usr_cat(mol)
shape3 = oddt.shape.electroshape(mol)
```

```
def initializer(smiles):
    global shape1
    global shape2
    global shape3
    mol = oddt.rdk.readstring('smi', smiles)
    mol = remover.StripMol(mol) if '.' in smiles else mol
    mol.addh()
    mol.make3D(forcefield="mmff94", steps=8192)
    shape1 = oddt.shape.usr(mol)
    shape2 = oddt.shape.usr_cat(mol)
    shape3 = oddt.shape.electroshape(mol)
```

```
def list_to_line(_list, sep: str = '\t') -> str:
    return f'{sep.join(_list)}\n'
```

```
def file_list(folder: str):
    fl = list()
    for root, dirs, files in os.walk(folder):
        for file in files:
            fl.append(os.path.join(root, file))
    return fl
```

```
def human_format(num: int) -> str:
```

```
    num = float(f'{num:.3g}')
```

```
    magnitude = 0
```

```
    while abs(num) >= 1000:
```

```
        magnitude += 1
```

```
        num /= 1000.0
```

```
    num = f'{num:.1f}'.rstrip('0').rstrip('.')
```

```
    let = ['', 'K', 'M', 'B', 'T'][magnitude]
```

```
    return f'{num}{let}'
```

```
def add_count_to_file(file: str, count: int) -> str:
```

```
    name, ext = os.path.splitext(os.path.basename(file))
```

```
    if ext.lower() in ['.gz', '.bz2']:
```

```
        ext = '.' + name.split('.')[-1] + ext
```

```
        name = '.'.join(name.split('.')[:-1])
```

```
    folder = os.path.dirname(file)
```

```
    new_file = os.path.join(folder, f'{name}_{human_format(count)}{ext}')
```

```
    os.replace(file, new_file)
```

```
    return new_file
```

```
def worker(file):
```

```
    results = list()
```

```
    with gzip.open(file, 'rb') as f:
```

```
        while True:
```

```
            try:
```

```
                line, usr_shape, usr_cat_shape, electroshape_shape = pickle.load(f)
```

```
                similarity_usr = oddt.shape_usr_similarity(shape1, usr_shape)
```

```
                similarity_usr_cat = oddt.shape_usr_similarity(shape2, usr_cat_shape)
```

```

similarity_electroshape = oddt.shape.usr_similarity(shape3, electroshape_shape)
# if any(x >= min_sim for x in (similarity_usr, similarity_usr_cat, similarity_electroshape)):
if similarity_usr_cat >= min_sim:
    r = line.rstrip('\n').split('\t')
    results.append(list_to_line([r[0], r[1],
                                f'{similarity_usr:.4f}',
                                f'{similarity_usr_cat:.4f}',
                                f'{similarity_electroshape:.4f}']))
except EOFError:
    break
return results

if __name__ == '__main__':
    c = 0
    start = time.time()
# with open(result_file, 'wt') as f, mp.Pool(processes=24, initializer=initializer, initargs=(smiles,)) as pool:
    with open(result_file, 'wt') as f, mp.Pool(processes=4, initializer=initializerSD, initargs=(SDFile,)) as pool:
        f.write(headers)
        for res in tqdm(pool.imap(worker, sorted(file_list(base_folder))), total=len(file_list(base_folder))):
            if res:
                c += len(res)
                f.writelines(res)
    add_count_to_file(result_file, c)
    end = time.time()
    print(f'Done, {c:_} results, {end - start:.1f} seconds')

```