

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

Факультет комп'ютерних наук,
управління та адміністрування
Кафедра інформаційних технологій

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему: АНАЛІЗ ГЕНОМІВ ПРЕДСТАВНИКІВ РОДУ АРГЕМОНА
(*ARGEMONE L.*) БІОІНФОРМАТИЧНИМИ МЕТОДАМИ

Виконав студент 2 курсу групи МІС-22
спеціальності 122 Комп'ютерні науки
Сліщук Георгій Іванович

Керівник Доктор технічних наук,
професор
Мещеряков Володимир Іванович

Рецензент Доктор технічних наук,
професор
Казакова Надія Феліксівна

АНОТАЦІЯ

Робота присвячена біоінформатичному аналізу геномів представників роду Аргемона, відомого своєю значущістю у фармакології. Рід Аргемона містить види, які виробляють унікальні алкалоїди, важливі для розробки лікарських препаратів. Ці алкалоїди мають потенціал у лікуванні ряду захворювань, що робить генетичне дослідження цього роду особливо актуальним для відкриття нових терапевтичних можливостей.

В ході дослідження проведено філогенетичний аналіз представників роду Аргемона. Для цього використовувались нуклеотидні послідовності ITS1, що дозволило встановити еволюційні зв'язки між різними видами. Розроблений та валідований дизайн праймерів та зонду, що дозволяє ідентифікувати ДНК представників роду Аргемона в продуктах харчування.

Також проведені гомологічне моделювання та молекулярний докінг ключових ензимів (CYP719A14, CYP719A13, TyPPPO, TyDC, NCS1, NCS2, VBE1 та VBE2), відповідальних за біосинтез алкалоїдів в роду Аргемона. Це дозволило виявити потенційні мішені для фармакологічного втручання та зрозуміти механізми виробництва цих біоактивних речовин на молекулярному рівні.

В рамках дослідження було розроблено спеціалізоване програмне забезпечення для симуляції молекулярної динаміки. Завдяки цьому програмному забезпеченню вдалося провести валідацію раніше отриманих моделей ензимів. Використання таких симуляцій дозволяє не лише перевірити правильність структурних моделей, але й отримати додаткове розуміння про динаміку та функціональні аспекти взаємодії ензимів з молекулами алкалоїдів.

Ключові слова: БІОІНФОРМАТИКА, ГОМОЛОГІЧНЕ МОДЕЛЮВАННЯ, ФІЛОГЕНЕТИЧНИЙ АНАЛІЗ, СИМУЛЯЦІЇ МОЛЕКУЛЯРНОЇ ДИНАМІКИ.

Магістерська робота містить 101 сторінок, 4 таблиць, 33 рисунків, 62 посилань.

ABSTRACT

This study is dedicated to the bioinformatics analysis of genomes of the Argemone genus, known for its significance in pharmacology. The genus Argemone contains species that produce unique alkaloids, important for the development of medicinal drugs. These alkaloids have potential in treating a variety of diseases, making genetic research of this genus particularly relevant for discovering new therapeutic opportunities.

During the research, a phylogenetic analysis of the Argemone genus was conducted. For this, ITS1 nucleotide sequences were used, allowing the establishment of evolutionary links between different species. A design for primers and a probe was developed and validated, enabling the identification of Argemone DNA in food products.

Additionally, homology modeling and molecular docking of key enzymes (CYP719A14, CYP719A13, TyPPO, TyDC, NCS1, NCS2, BBE1, and BBE2), responsible for alkaloid biosynthesis in the Argemone genus, were conducted. This allowed for the identification of potential targets for pharmacological intervention and understanding of the mechanisms of production of these bioactive substances at the molecular level.

As part of the study, specialized software for simulating molecular dynamics was developed. Thanks to this software, it was possible to validate previously obtained enzyme models. The use of such simulations not only verifies the correctness of structural models but also provides additional understanding of the dynamics and functional aspects of enzyme interactions with alkaloid molecules.

Keywords: BIOINFORMATICS, HOMOLOGY MODELING, PHYLOGENETIC ANALYSIS, MOLECULAR DYNAMICS SIMULATIONS.

The master's thesis consists of 101 pages, 4 tables, 33 figures, 62 references.

ЗМІСТ

	Стор.
СПИСОК УМОВНИХ СКОРОЧЕНЬ.....	5
ВСТУП.....	6
РОЗДІЛ 1. Аналіз геномів представників роду Аргемона (<i>Argemone</i> L.) біоінформатичними методами.....	10
1.1 Біоінформатичний аналіз у сучасній біологічній науці.....	10
1.2 Біоінформатичне дослідження геномів рослин.....	15
1.3 Біоінформатичне дослідження геномів роду Аргемона (<i>Argemone</i> L.).....	18
1.3.1 Таксономія роду Аргемона.....	20
1.3.2 Вторинні метаболіти представників роду Аргемона.....	21
1.3.3 Практичне значення представників роду Аргемона.....	23
1.4 Інженерінг алколоїдів.....	24
РОЗДІЛ 2. БІОІНФОРМАТИЧНІ МЕТОДИ АНАЛІЗУ.....	27
2.1 Матеріали.....	27
2.1.1 Нуклеотидні послідовності аргемони.....	27
2.1.2 Види аргемони.....	30
2.2 Біоінформатичні методи.....	32
2.2.1 Вирівнювання нуклеотидних та амінокислотних послідовностей.....	32
2.2.2 Філогенетична реконструкція.....	32
2.2.3 Гомологічне моделювання структури протеїнів.....	33
2.2.4 Дизайн праймерів і зондів, <i>in silico</i> ПЛР.....	34
2.3 Екстрагування та очистка ДНК аргемони.....	34
2.4. Полімеразна ланцюгова реакція.....	35
2.5 Загальна схема роботи.....	36
РОЗДІЛ 3. ФІЛОГЕНЕТИЧНИЙ АНАЛІЗ ПРЕДСТАВНИКІВ РОДУ.....	37

	АРГЕМОНА.....	
3.1	Розробка праймерів і зондів та <i>in silico</i> ПЛР.....	37
3.2	<i>In vitro</i> ПЛР-аналіз	39
РОЗДІЛ 4.	БІОІНФОРМАТИЧНИЙ АНАЛІЗ ГЕНІВ, АСОЦІЙОВАНИХ З БІОСИНТЕЗОМ ВТОРИННИХ МЕТАБОЛІТІВ.....	42
4.1	Біоінформатичний аналіз гена <i>CYP719A14</i>	42
4.2	Біоінформатичний аналіз гена <i>CYP719A13</i>	47
4.3	Біоінформатичний аналіз гена <i>TyPPO</i>	50
4.4	Біоінформатичний аналіз гена <i>TyDC</i>	53
4.5	Біоінформатичний аналіз гена <i>NCS1</i>	57
4.6	Біоінформатичний аналіз гена <i>NCS2</i>	59
4.7	Біоінформатичний аналіз гена <i>BBE1</i>	62
4.8	Біоінформатичний аналіз гена <i>BBE2</i>	65
4.9	Інженірінг алкалоїдів представників роду Аргемона.....	68
4.10	Аналіз та узагальнення результатів.....	70
РОЗДІЛ 5.	РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ МОДЕЛЮВАННЯ МОЛЕКУЛЯРНОЇ ДИНАМІКИ.....	73
	ВИСНОВКИ ТА ПРАКТИЧНІ РЕКОМЕНДАЦІЇ.....	82
	СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	83
	ДОДАТОК А.....	91
	ДОДАТОК Б.....	98

СПИСОК УМОВНИХ СКОРОЧЕНЬ

ГМО	– генетично модифікований організм
ДНК	– дезоксирибонуклеїнова кислота
МД	– молекулярна динаміка
ПЛР	– полімеразна ланцюгова реакція
BLAST	– Basic Local Alignment Search Tool, програмне забезпечення для локального вирівнювання
ITS спейсер	– internal transcribed spacer, внутрішній транскрибований спейсер
NCBI	– National Center for Biotechnology Information, національний центр біотехнологічної інформації, генбанк
TLS	– Targeted Locus Study, цільове дослідження локусів
TSA транскриптому	– Transcriptome Shotgun Assembly, «шотган» складання транскриптому
WGS	– Whole Genome Shotgun, «шотган» цілого геному

ВСТУП

Актуальність теми. Аналіз геномів представників роду Аргемона (*Argemone* L.) за допомогою біоінформатичних методів набуває великої актуальності у сучасній ботаніці та біотехнології, враховуючи різноманіття медичного та аграрного застосування цих рослин. Глибинний геномний аналіз може розкрити ключові гени, відповідальні за синтез алкалоїдів, які мають анальгетичні властивості, та інших біоактивних сполук. Це відкриває нові можливості для розробки лікарських препаратів та стійких сільськогосподарських культур. Крім того, використання біоінформатики дозволяє прискорити процес ідентифікації та клонування цільових генів, що сприяє розширенню нашого розуміння молекулярних механізмів рослинної адаптації та еволюції.

Вивчення ключових ферментів, які каталізують біосинтез алкалоїдів у роді Аргемона, відкриває нові перспективи для медицини та сільського господарства. Знання конкретних шляхів та ферментативних механізмів дозволяє науковцям не тільки покращити біосинтетичні процеси у рослинах для отримання більш ефективних лікарських засобів, але й створювати ГМО-культури з покращеними властивостями. Зокрема, модифікація шляхів біосинтезу алкалоїдів може призвести до створення рослин з вищим вмістом цих сполук, що є важливим для фармацевтичної індустрії. Крім того, здатність маніпулювати метаболічними шляхами відкриває шлях для розробки рослин-біофабрик, що виробляють рідкісні або важко синтезуємі сполуки, що знижує залежність від складних хімічних синтезів. Таким чином, біоінформатичний аналіз сприяє створенню нових підходів у селекції та фармацевтиці, що базуються на глибокому розумінні генетичних та біохімічних основ живих організмів.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконана у відділі молекулярної генетики і фітосанітарної експертизи компанії

КОТЕКНА УКРАЇНА ЛІМІТЕД в рамках проекту дослідження геномів представників роду Аргемона, протягом 2022-2023 років.

Мета і задачі дослідження. Мета полягала у розробці та застосуванні комп'ютерних алгоритмів та біоінформатичних методів для аналізу молекулярно-генетичного поліморфізму геномів представників роду Аргемона та дослідженні молекулярно-генетичного поліморфізму геномів представників роду Аргемона біоінформатичними методами. Для дослідження вищенаведеної мети були поставлені наступні задачі:

1. Дослідження молекулярно-генетичного поліморфізму нуклеотидних послідовностей *internal transcribed spacer* (ITS) представників роду Аргемона та дизайн послідовностей праймерів та зонду.

2. Моделювання тривимірної структури основних ензимів, асоційованих із біосинтезом алкалоїдів у представників роду Аргемона.

3. Валідація побудованих моделей за допомогою методів симуляції молекулярної динаміки.

4. Дослідження аффінітету побудованих моделей до субстратів та продуктів ензиматичної реакції за допомогою молекулярного докінгу.

Об'єкт дослідження: аналіз геномів представників роду Аргемона.

Предмет дослідження: гени, що асоційовані із синтезом алкалоїдів.

Методи дослідження: біоінформатичні методи (вирівнювання нуклеотидних послідовностей, реконструкція філодендрограм, кластерний аналіз, дизайн праймерів, *in silico* ПЛР) використовували для вивчення нуклеотидних та амінокислотних послідовностей, отриманих з міжнародних баз даних, а також нуклеотидних послідовностей, отриманих в результаті секвенування, для вивчення поліморфізму нуклеотидних та амінокислотних послідовностей, гомологічне моделювання, симуляції МД та молекулярний докінг використовували для дослідження шляхів біосинтезу алкалоїдів, ПЛР в

реальному часі використовували для оцінки розробленого дизайну праймерів та зонду у якості тест-системи.

Наукова новизна одержаних результатів полягає у постановці наукової задачі, яка полягає у біоінформатичному дослідженні геномів представників роду Аргемона. Вперше проведено моделювання тривимірної структури ключових ензимів біосинтезу алкалоїдів Аргемони. Проведена валідація побудованих моделей за допомогою методу симуляції молекулярної динаміки. Досліджена взаємодія побудованих моделей із субстратами та продуктами ензиматичних реакцій за допомогою молекулярного докінгу. Створений унікальний дизайн праймерів та зонду для детекції ДНК представників роду Аргемона.

Практичне значення одержаних результатів полягає в розробці системи молекулярних маркерів для детекції ДНК представників роду Аргемона. Ця маркерна система може бути використана для детекції представників роду Аргемона у зерні, продуктах її переробки та харчових продуктах, зокрема у олії.

Особистий внесок здобувача. Здобувачем освітнього рівня магістра особисто проаналізовано літературні джерела за темою роботи, виділено ДНК з 8 зразків представників роду Аргемона, отриманих із банку гермоплазми USDA-ARS Germplasm Resources Information Network (GRIN), а також із Національного Ботанічного Саду ім. Гришка. Біоінформатичними методами проаналізована філогенія більш 20 нуклеотидних послідовностей ITS. Розроблено унікальний дизайн праймерів та зонду та проведений ПЛР *in silico* та *in vitro*. За допомогою ПЛР з розробленими праймерами та зондом проведена валідація тест-системи детекції та ідентифікації представників роду Аргемона. Проведено моделювання ензимів, асоційованих із синтезом алкалоїдів, проведена їх валідація та взаємодія із субстратами та продуктами ензиматичних реакцій.

Обговорення, аналіз, інтерпретацію отриманих результатів, а також підготовку публікацій до друку здійснювали спільно з д.т.н., професором, професором кафедри автоматизованих систем моніторингу навколишнього

середовища та інформатики Одеського державного екологічного університету Мещеряковим В.І., якому автор висловлює щирю подяку.

Апробація результатів. Результати дослідження представлено на конференції молодих учених Одеського державного екологічного університету 2023 року.

Публікації. Оpubліковано наукову статтю у журналі Фактори експериментальної еволюції організмів.

Структура та об'єм роботи. Робота складається з вступу, огляду літератури, розділу матеріалів та методів, двох розділів власних досліджень, аналізу та узагальнення результатів, висновків, практичних рекомендацій, списку використаної літератури.

Робота складається з п'яти розділів, її повний обсяг становить 101 сторінка машинописного тексту, з них 88 складають обсяг основного тексту, включаючи 4 таблиць і 33 рисунка. Список використаної літератури включає 62 джерело.

1 АНАЛІЗ ГЕНОМІВ ПРЕДСТАВНИКІВ РОДУ АРГЕМОНА (*ARGEMONE L.*) БІОІНФОРМАТИЧНИМИ МЕТОДАМИ

1.1 Біоінформатичний аналіз у сучасній біологічній науці

Біоінформатика, перетинаючи межі біології та інформатики, змінила парадигму сучасних біологічних досліджень. Її можливості в аналізі великих обсягів даних відкрили нові горизонти для розуміння складних біологічних систем і процесів на молекулярному рівні. Біоінформатика є галуззю, де злиття різних наук відкриває шлях для розробки нових підходів та програмних рішень, спрямованих на розшифровку біологічної інформації, особливо коли мова йде про обсяжні та складні дані. Ця галузь використовує знання та методики з біології, хімії, фізики, комп'ютерних наук, програмування, інформаційної інженерії, математики та статистики для того, щоб аналізувати та трактувати біологічні дані. Процес обробки та вивчення цих даних відомий як обчислювальна біологія.

Термін "біоінформатика" з'явився у науковому дискурсі наприкінці 1970-х років, однак його корені сягають ще глибше у історію комп'ютерних наук та біології. Інтенсивний розвиток біоінформатики розпочався у 1980-х, коли стрімке зростання об'ємів біологічних даних, зокрема геномних послідовностей, вимагало нових методів аналізу та інтерпретації. Цей період характеризується злиттям традиційної молекулярної біології з методами обчислювальної математики, статистики та інформатики, що сприяло виникненню біоінформатики як самостійної галузі наукових досліджень. Початок 21-го століття ознаменувався бурхливим розвитком цієї дисципліни, коли біоінформатика стала невід'ємною частиною біомедичних досліджень та біотехнологічної промисловості, забезпечуючи критичне розуміння біологічних процесів на молекулярному рівні [1].

Одним із найяскравіших прикладів інтенсивного розвитку біоінформатики є її роль у Проекті генома людини (HGP), що стартував у 1990 році. Цей амбітний проект, метою якого було визначення послідовності всіх 3 мільярдів пар основ ДНК у людському геномі, став каталізатором для стрімкого росту біоінформатичних методів і технологій. Завдання картування та секвенування великих обсягів генетичних даних вимагало розробки нових комп'ютерних алгоритмів, баз даних та інструментальних ресурсів для зберігання, аналізу та інтерпретації інформації. В результаті, біоінформатика стала невід'ємною частиною геномних досліджень, забезпечуючи необхідні інструменти для розуміння структури та функцій генів, що сприяло подальшим проривам у генетиці, медицині та біотехнологіях [2].

Однією із найпотужніших біоінформатичних баз даних є генбанк. Стан генбанку GenBank випуску 257.0 на дату 15 серпня 2023 року включає в себе 25,10 трильйонів баз та 3,69 мільярда записів. Цей реліз містить:

- 246,119,175 традиційних записів з 2,112,058,517,945 базами даних послідовностей;
- 2,631,493,489 WGS записів з 22,294,446,104,543 базами даних послідовностей;
- 686,271,945 масово орієнтованих TSA записів з 646,176,166,908 базами даних послідовностей;
- 124,421,006 масово орієнтованих TLS записів з 48,289,699,026 базами даних послідовностей.

За 59 днів між закриттям релізів 256.0 та 257.0, традиційна частина GenBank зросла на 145,578,541,799 баз та на 2,558,312 записів послідовностей. Також було оновлено 34,840 записів.

Загальна кількість файлів даних послідовностей збільшилася на 441 у цьому випуску, при цьому зросла кількість файлів у різних розділах, таких як BCT, ENV, INV та інші.

Необхідно зазначити, що незапаковані плоскі файли даних послідовностей GenBank релізу 257.0 вимагають приблизно 3,575 ГБ місця для завантаження. Файли даних ASN.1 потребують приблизно 1,669 ГБ [3].

Історично першими у біоінформатиці з'явилися методи роботи із нуклеотидними та амінокислотними послідовностями, а саме обслуговування молекулярно-генетичних методів визначення послідовностей, або секвенування. Такими є методи складання геному. Виділяють *de novo* складання геномів, коли із коротких фрагментів, що генерують методи секвенування шляхом вирівнювання формується цілісний геном. Вони в свою чергу поділяються на “жадібні” алгоритми (greedy algorithm) та метод складання за допомогою графів. Жадібний алгоритм складання геному працює за принципом пошуку найбільшої спільної послідовності між фрагментами ДНК, які потрібно з'єднати. Він ітеративно вибирає фрагменти, що мають найбільшу взаємну збіжність кінців, та з'єднує їх в один довший фрагмент. Цей процес повторюється, поки всі частини не будуть з'єднані в єдину послідовність. Такий підхід ефективний для розробки швидких рішень, але може не завжди знаходити найкращий можливий результат, оскільки він не здатен враховувати більш складні перекриття чи повторення в послідовності ДНК [4]. Методи складання графів для збірки геному використовують графові структури для представлення перекриттів між фрагментами ДНК. Вони створюють структуру, звану графом перекриттів або де Брюйн графом, де вершини представляють всі можливі послідовності ДНК певної довжини (k-мери), а ребра показують, як ці k-мери перекриваються один з одним. Знайшовши шлях через цей граф, який проходить через кожне ребро рівно один раз (Єйлерів шлях) або проходить через кожну вершину рівно один раз (Гамільтонів шлях), можна реконструювати весь геном. Цей метод дозволяє ефективно обробляти великі масиви даних та знаходити оптимальні шляхи навіть при наявності помилок у послідовності або складних повторень [5]. Якщо є доступ до референтного геному, використовують мапінг. Процес відображення

(мапінгу) зчитувань на геномну послідовність полягає у визначенні місць розташування коротких фрагментів ДНК (зчитувань), отриманих під час секвенування, у рамках більшої референсної геномної послідовності. Це робиться шляхом порівняння кожного фрагмента з референсним геномом для виявлення найбільш відповідного місця у послідовності. Мапінг важливий для визначення структурних варіацій у геномі, ідентифікації мутацій, а також для розуміння генетичної варіабельності. Цей процес вимагає складних алгоритмів і обчислювальних ресурсів, особливо при роботі з великими геномами та великою кількістю зчитувань [6]. Комбінацією першого (*de novo*) та другого підходів є послідовність керованого *de novo* складання геному. Послідовність керованого *de novo* складання геному використовує поєднання інформації з референсного геному та нових даних секвенування для створення більш точної та повної геномної послідовності. У цьому процесі, зчитування ДНК спочатку мапуються на відомий референсний геном для ідентифікації відомих областей, після чого використовуються алгоритми *de novo* складання для заповнення прогалів і виправлення помилок, викликаних структурними варіаціями чи мутаціями. Такий підхід дозволяє ефективніше виявляти нові варіанти у геномі та забезпечує високу точність складання за рахунок використання референсної послідовності як шаблону [7].

Після того, як секвеновані послідовності складені, наступним етапом біоінформатичного аналізу є вирівнювання послідовностей з метою їх порівняння. Вирівнювання послідовностей є фундаментальною задачею в біоінформатиці, яка дозволяє визначити ступінь подібності між двома або більше біологічними послідовностями (нуклеотидами або амінокислотами). Цей процес є ключовим для ідентифікації консервативних областей, що можуть мати важливе біологічне значення, виявлення еволюційних відносин між організмами, розуміння функціональних змін, а також для виявлення і розуміння мутацій та їхнього впливу на фенотип. Вирівнювання може бути глобальним, коли

порівнюються послідовності в цілому, або локальним, коли зосередження відбувається на найбільш подібних фрагментах. Для автоматизації цього процесу розроблено безліч алгоритмів та програмних засобів, серед яких найбільш відомі — BLAST (Basic Local Alignment Search Tool) для швидкого локального вирівнювання та Clustal для розрахунку багатопослідовного вирівнювання [8]. Традиційно розрізняють локальне вирівнювання та глобальне вирівнювання. Локальне вирівнювання послідовностей — це процес порівняння та зіставлення двох біологічних послідовностей, наприклад, ДНК, РНК або білків, для ідентифікації подібних або ідентичних регіонів. Відмінно від глобального вирівнювання, яке намагається вирівняти дві послідовності по всій їх довжині, локальне вирівнювання фокусується на знаходженні найбільш схожої підпослідовності між ними. Це корисно, коли дві послідовності мають лише невелику схожу ділянку або коли важливо знайти домени або функціональні мотиви, що збігаються. Алгоритми, такі як BLAST або Smith-Waterman, широко використовуються для локального вирівнювання, дозволяючи дослідникам виявляти значущі біологічні зв'язки та структурні взаємодії між молекулами [9]. Глобальне вирівнювання послідовностей — це метод порівняння двох біологічних послідовностей, таких як ДНК, РНК або білкові послідовності, із метою виявлення максимальної подібності по всій їх довжині. Цей процес включає вирівнювання від початку до кінця обох послідовностей, щоб знайти найкраще можливе збігання, використовуючи штрафи за пропуски та неспівпадіння. Він особливо корисний для порівняння геномів близькоспоріднених видів або для визначення консервативних регіонів у білках. Алгоритми глобального вирівнювання, такі як Needleman-Wunsch, дозволяють отримати повне уявлення про ступінь гомології та еволюційні відносини між двома послідовностями [10].

Структурна біоінформатика — це галузь біоінформатики, яка зосереджується на аналізі, визначенні та моделюванні тривимірних структур

біомолекул, таких як білки та нуклеїнові кислоти. Використовуючи дані, отримані з таких методів як кристалографія рентгенівськими променями, ЯМР-спектроскопія та кріо-електронна мікроскопія, фахівці у цій області розробляють комп'ютерні алгоритми та програмне забезпечення для передбачення і візуалізації просторових структур молекул. Завдяки цьому можна зрозуміти фундаментальні біологічні процеси на молекулярному рівні, розробляти нові ліки, та глибше вивчати механізми хвороб. Структурна біоінформатика відіграє ключову роль у сучасних біомедичних дослідженнях та біотехнології [11].

Прогнозування структури білка — це процес визначення тривимірної форми білка на основі його амінокислотної послідовності. Ця задача є однією з найважливіших в біоінформатиці, оскільки просторова структура білка визначає його функції та взаємодії з іншими молекулами. Прогнозування використовує комп'ютерні алгоритми для моделювання складання білка, включаючи методи гомологічного моделювання, *ab initio*, та з використанням штучного інтелекту, як-от система AlphaFold. Ефективність цих методів дозволяє швидко отримувати дані про структуру білків, що має важливе значення для розуміння біологічних процесів та розробки нових лікарських препаратів [12].

Гомологічне моделювання за допомогою Swiss-model є широко використовуваною комп'ютерною технікою для прогнозування тривимірної структури білка на основі його амінокислотної послідовності. Цей метод використовує відомі структури білків, що є гомологічними до цільового білка, як шаблони для створення його структурної моделі. Swiss-model аналізує послідовність білка, шукає найкращі шаблони у базах даних структурних білків, таких як Protein Data Bank, і використовує їх для побудови просторової моделі цільового білка. Цей процес дозволяє вченим передбачити, як білок буде виглядати та як він може функціонувати, що є ключовим для багатьох досліджень у біомедицині та фармакології [13].

1.2 Біоінформатичне дослідження геномів рослин

Біоінформатичне дослідження геномів рослин використовується для комплексного аналізу генетичного матеріалу рослин за допомогою комп'ютерних технологій для вивчення їх спадковості, еволюції та функціонування. Цей процес включає секвенування ДНК, анотацію генів, порівняльний аналіз геномів та ідентифікацію генетичних варіантів, які можуть впливати на різноманітність рослин, їх захист від шкідників та адаптацію до змін умов довкілля. Такі дослідження допомагають покращити розуміння біологічних механізмів, що лежать в основі розвитку рослин, їх відповідей на стрес та можуть сприяти створенню нових сортів рослин з покращеними характеристиками для сільського господарства. Так, в статті [14] автори розглядають застосування біоінформатики для дослідження генетичних механізмів у кукурудзи. Дослідження фокусується на аналізі вторинної структури транскриптів інтрону 1 гену *whp1*, який може мати важливе значення у відновленні фертильності кукурудзи з S-типом ЦЧС (цитоплазматичної чоловічої стерильності). Використовуючи вирівнювання і *in silico* фолдінг 74 нуклеотидних послідовностей, автори встановили структуру гену і виявили складні SSR (прості послідовні повтори). Результати дослідження можуть пролити світло на молекулярні механізми, що впливають на репродуктивні здібності рослин і демонструють важливість біоінформатики у сучасній геноміці рослин.

Біоінформатика інтенсивно вивчає і органельні геноми рослин. Так, еукаріотичні органели, зокрема мітохондрії мають власні геноми, що функціонують відносно незалежно від ядерного геному. Зазначається, що мутації у мітохондріальному геномі можуть призвести до серйозних змін у фенотипі організмів, включаючи патології у людей та ЦЧС у рослин. Це підкреслює важливість біоінформатичного аналізу для розуміння мутацій, які впливають на розвиток рослинних тканин, особливо андроцею при нормальному розвитку

гінецею. Аналіз, проведений на основі послідовностей генів альдегіддегідрогеназ, важливих у метаболічних шляхах мітохондрій, підсилює розуміння взаємодії між мітохондріальними і ядерними геномами, що є ключовим для розвитку біотехнологій у рослинництві [15].

Зі зміною кліматичних умов, зокрема зростанням середньорічної температури та частішими посухами, важливість біоінформатики та молекулярної генетики у розробці стратегій адаптації та селекції високоврожайних та стресостійких сортів кукурудзи стає все більшою. Важливе значення має аналіз гена AASR1, який, ймовірно, відіграє ключову роль у адаптації кукурудзи до змінних умов навколишнього середовища. Біоінформатичний та молекулярний аналіз цього гена може допомогти в ідентифікації генетичних маркерів для селекції, сприяючи розвитку нових сортів, здатних витримувати екстремальні погодні умови [16].

Структурна біоінформатика теж відіграє ключову роль у дослідженні геномів рослин. Використовуючи методи, такі як вирівнювання нуклеотидних послідовностей, філогенетичний аналіз та гомологічне моделювання, дослідники змогли ідентифікувати гомологи цього гена у різних родинах рослин, виявити одонуклеотидний поліморфізм, асоційований зі стійкістю до гербіцидів, та визначити рівень консервативності послідовностей між видами. Це дозволило провести маркерний скринінг селекційного матеріалу для відбору генотипів з бажаними властивостями, підкреслюючи важливість біоінформатичних підходів в сучасній агробіотехнології та рослинництві [17].

Біоінформатичні дослідження мають вирішальне значення для сучасної молекулярної генетики та селекції рослин, оскільки вони забезпечують глибоке розуміння генетичних процесів та механізмів, що стоять за розвитком та адаптацією рослин. Аналізуючи великі набори геномних даних, біоінформатики можуть визначати ключові гени, асоційовані з важливими агрономічними ознаками, такими як стійкість до хвороб, посухи, високий урожай. Ці знання

прискорюють розробку нових сортів рослин з покращеними характеристиками, сприяючи сталому розвитку сільського господарства та забезпеченню продовольчої безпеки у відповідь на зростаючі виклики зміни клімату та зростання населення планети.

1.3 Біоінформатичне дослідження геномів роду Аргемона (*Argemone* L.)

Біоінформатичне дослідження геномів роду *Argemone* є значущим, адже цей рід містить види, які мають потенційну лікарську та економічну вартість. Використовуючи сучасні методи секвенування та аналітичні алгоритми, дослідники можуть розшифровувати геном цих рослин, ідентифікувати генетичні маркери, відповідальні за виробництво алкалоїдів, та інші цінні метаболіти. Крім того, аналіз геномної варіації в межах роду може сприяти виявленню генів, зв'язаних з адаптацією до різних екологічних умов, що відкриває шлях для розробки нових сортів з підвищеною стійкістю до стресів. Це може мати важливе значення для фармацевтичної індустрії та для розробки стратегій збереження видів.

Argemone mexicana (мексиканський колючий мак) є важливою модельною рослиною у вивченні біосинтезу алкалоїдів, зокрема антимікробного сангвінаріну, що робить її значущим об'єктом для біохімічних та молекулярно-генетичних досліджень. Протягом останніх десятиліть вчені інтенсивно досліджують ці види, що призвело до майже повного розуміння основних шляхів біосинтезу на біохімічному та молекулярно-генетичному рівнях. Було клоновано та охарактеризовано більшість або всі ферменти, відповідальні за біосинтез таких алкалоїдів, як папаверин, морфін, сангвінарін, берберин та носкапін, що відкриває нові можливості для розробки лікарських препаратів та розуміння метаболічних процесів у рослинах [18].

Біоінформатичні дослідження роду *Argemone* відіграють ключову роль у розумінні регуляції генетичної експресії та метаболізму в рослинному царстві. Зокрема, в рамках вивчення двох видів сімейства *Papaveraceae* — *P. somniferum* L. та *P. bracteatum* L., було кількісно визначено рівень транскрипції прекурсорів та зрілих форм кластеризованих мікроРНК методом зворотної транскрипції полімеразної ланцюгової реакції (RT-PCR). Результати показали диференційовану експресію цих кластеризованих мікроРНК у різних тканинах цих видів. Крім того, прогнозування мішеней та функціональна класифікація на основі Gene Ontology (GO) вказали на те, що ці мікроРНК можуть відігравати вирішальну роль у різноманітних біологічних процесах та метаболічних шляхах. Це дослідження також відкрило кластеризовані мікроРНК, похідні від пре-рРНК, що може висвітлити важливість мікроРНК у рослинному царстві [19].

Біоінформатичні дослідження роду *Argemone* значно сприяють пошуку та розумінню природних хімічних сполук, які можуть викликати програмовану некрозу клітин. Традиційно такі сполуки виявляються за допомогою скринінгу концентрованих рослинних екстрактів без виділення активних компонентів, але за останні два десятиліття сучасна медицина все більше спирається на ізоляцію та очищення одного або двох складних активних та ізомерних з'єднань. Концепція багатоцільових ліків стрімко та вражаюче розвинулася з інноваційної моделі на початку 2000-х до одного з популярних напрямків розробки лікарських препаратів у 2021 році. Як альтернатива, розробка ліків на основі фрагментів також досліджується для ідентифікації цілеспрямованих препаратів, особливо потужних природних протипухлинних агентів, на основі добре визначених фрагментів, у протипухлинному використанні природно зустрічаючихся сумішей. Дослідження підсумовує ключові досягнення у сфері природних протипухлинних сполук, комп'ютерно-асистоване та фрагмент-базоване структурне роз'яснення, а також багатоцільовий підхід до дослідження природних з'єднань [20].

1.3.1 Таксономія роду Аргемона

Argemone є родом квіткових рослин у сімействі Макові (Papaveraceae), які зазвичай відомі як колючі маки. Включає 34 види, що є корінними для Америки та Гаваїв. Назва роду походить від грецького слова ἀργεμόνη, яке Діоскорид використовував для позначення рослини, схожої на мак, яка використовувалася для лікування катаракти [21]. Род описаний в 1853 році шведським систематиком, батьком сучасної систематики Карлом Лінеєм в *Species Plantarum* [22]. Ревізія роду була зроблена ботаніком Оунбі в 1968 році [23].

До роду входять наступні види:

Argemone aenea G. B. Ownbey

Argemone albiflora Hornem.

Argemone arida Rose

Argemone arizonica G. B. Ownbey

Argemone brevicornuta G. B. Ownbey

Argemone chisosensis G. B. Ownbey

Argemone corymbosa Greene

Argemone fruticosa Thurber ex A. Gray

Argemone glauca L. ex Pope

Argemone gracilentata Greene

Argemone hispida A. Gray

Argemone hybrida R. Otto & Verloove

Argemone mexicana L.

Argemone munita Dur. & Hilg.

Argemone pinnatisecta (G. B. Ownbey) S. D. Cerv. & C. D. Bailey

Argemone pleiacantha Greene

Argemone sanguinea Greene

Argemone subfusiformis G. V. Ownbey

Argemone subintegrifolia G. V. Ownbey

Argemone turnerae A. M. Powell

Типовим видом роду є Аргемона мексиканська, що походить з тропічної Америки, але розповсюджена в тропічних та субтропічних регіонах світу. Ця однорічна, прямостояча трав'яниста рослина висотою до 100 см, має колюче покриття та блакитно-зелене листя з білими та зеленими ділянками. Жовті квіти мають діаметр 4-5 см, а плід - овальну колючу капсулу. Аргемона мексиканська використовується як лікарська рослина завдяки антибактеріальним, антидіабетичним, антиоксидантним та гепатопротекторним властивостям, а також є об'єктом дослідження вторинних метаболітів, зокрема бензилізохінолінових алкалоїдів [24].

1.3.2. Вторинні метаболіти представників роду Аргемона

Рід *Argemone* відомий своїми вторинними метаболітами, які включають різноманітні біоактивні сполуки з потенційними фармакологічними властивостями. Вторинні метаболіти, такі як алкалоїди, що виявляються у цих рослинах, зокрема сангвінарин, мають антимікробні та протипухлинні властивості. Сангвінарин особливо цікавий завдяки його здатності викликати програмовану клітинну смерть, що може бути використане у розробці протиракових препаратів. Також в рослинах цього роду знаходять берберин, морфін, та інші сполуки, які вивчаються за їх потенційні терапевтичні ефекти. Вторинні метаболіти *Argemone* також вивчаються за їх роль у захисті рослин від шкідників і хвороб, вказуючи на їх важливість як для рослин, так і для можливих додатків у медицині та сільському господарстві.

Ідентифіковано більше ніж 20 алкалоїдів, серед яких особливе значення мають похідні бензилізохіноліну, зокрема берберин, протоберберин, сангвинарін [25, 26].

Охарактеризовано численні ензими, що залучені у біосинтезі похідних бензилізохіноліну, зокрема CYP719A14, що проявляє хейліантіфолінсинтазну активність і формує метилендіоксовий міст [27].

Алкалоїди, які містяться у представниках роду *Argemone*, можуть відігравати значну роль у пошуку нових терапевтичних підходів до лікування хвороби Альцгеймера — вікового нейродегенеративного розладу, що характеризується когнітивними порушеннями. Ізолювання таких сполук, як (+)-лауданозин, протопін, (–)-аргемонін, аллокриптопін, (–)-платицерин, (–)-мунітагін та (–)-нораргемонін із коренів та надземних частин *Argemone platyceras*, проведене з використанням етанолу та фракціонування на алюмінієвій колонці, відкрило нові можливості для лікування. Ці сполуки були проаналізовані за допомогою оптичного вращення, спектроскопічних та спектрометричних аналізів (ЯМР, мас-спектрометрія) і порівняно з літературними даними, що вперше дозволило ізолювати (+)-лауданозин з *A. platyceras*. Виявлено, що алкалоїди інгібували ацетилхолінестеразу, бутирилхолінестеразу та рекомбінантну проліл олігопептидазу у дозозалежний спосіб, зокрема (–)-мунітагін показав найвищу активність, інгібуючи обидва ферменти з IC_{50} 62.3 ± 5.8 мкМ і 277.0 ± 31.3 мкМ відповідно [28].

Аргемонін — це природний алкалоїд групи ізохіноліну, який був ізольований з рослин та насіння кількох видів, включаючи деякі рослини роду *Argemone*, такі як *A. gracilentata*, *A. platyceras*, *A. sanguinea* та інші. Крім того, його також знайдено в видів рослин *Buxifolia berberis*, *Thalictrum revolutum*, і *T. strictum*. Найпоширеніше застосування аргемоніну — контроль шкідників у сільському господарстві у комбінації з берберином та рицином завдяки його антибактеріальним, фунгіцидним та інсектицидним властивостям. Проте

біологічні дослідження аргемоніну є обмеженими, і його біологічні активності досі не були повністю визначені [29].

1.3.3. Практичне значення представників роду Аргемона

Представники роду Аргемона мають значний практичний потенціал, що робить їх об'єктами важливих досліджень у фармакології та агрономії. Зокрема, завдяки вмісту біоактивних алкалоїдів, таких як сангвінарин, ці рослини вивчаються як потенційні джерела нових лікарських препаратів з антимікробними, протипухлинними та інсектицидними властивостями. У сільському господарстві вони можуть використовуватися для розробки натуральних пестицидів для боротьби зі шкідниками та хворобами рослин. Крім того, деякі види Аргемони використовуються в традиційній медицині для лікування ряду захворювань, в тому числі для полегшення болю та запалення. Таким чином, різноманітність та ефективність біологічно активних речовин в роду Аргемона ставить його в центр уваги як для наукових досліджень, так і для практичного застосування.

Аргемона мексиканська (*A. mexicana*), відома своїм жовтим молочком, традиційно використовується в народній медицині для лікування різноманітних захворювань. Цей жовтий ексудат застосовують як засіб проти гідропсії (набряків), жовтяниці, сверблячки та інших шкірних захворювань. Завдяки антисептичним та протизапальним властивостям, молочко цієї рослини служить ефективним народним лікарським засобом, який використовується в різних культурах для поліпшення стану шкіри та лікування шкірних інфекцій. Таке широке використання свідчить про потенційну цінність Аргемони у фітотерапії та можливості подальших досліджень для ідентифікації та використання її лікувальних компонентів у сучасній медицині [30]. Квіти, листя та насіння рослини *Argemone* вже давно використовуються в народній медицині для

лікування різноманітних захворювань. Ці частини рослини застосовують при дизентерії, виразках, кашлі та високому кров'яному тиску. Таке використання базується на багатому спектрі фітохімічних сполук, які містяться в *Argemone*, і може мати антимікробну, протизапальну, знеболювальну та гіпотензивну дію. Це робить *Argemone* цінним ресурсом для традиційної медицини та потенційним об'єктом для фармацевтичних досліджень з метою розробки нових лікарських засобів [31-35].

Листя та стебла *A. mexicana* використовуються в народній медицині для лікування малярії та набряків. Вони мають протибольові, антиспазматичні, протипаразитарні та наркотичні властивості, а також проявляють протигрибкову, гепатопротективну, личинкову та хемотрилізуючу активність. Ці характеристики роблять рослину потенційно цінним компонентом для розробки фітопрепаратів та інших лікарських засобів, здатних боротися з широким спектром захворювань та станів, що збільшує інтерес до подальших досліджень її медичного потенціалу [36,37].

1.4 Інженерінг алкалоїдів

Інженерія алкалоїдів — це передова галузь біотехнології, яка включає маніпулювання шляхами біосинтезу в рослинах та мікроорганізмах для збільшення виробництва або створення нових алкалоїдних сполук. Ці сполуки мають широке застосування в медицині як потужні лікарські препарати завдяки їх анальгетичним, антимікробним, протипухлинним та іншим терапевтичним властивостям. За допомогою метаболічної інженерії, генної інженерії та синтетичної біології вчені прагнуть збагатити шляхи біосинтезу або вставити нові генетичні шляхи в рослинні або мікробні системи для ефективного виробництва цінних алкалоїдів. Це дозволяє оптимізувати виробництво вже відомих алкалоїдів, а також розробляти нові молекули з потенційною

біологічною активністю, що може відкрити шлях до нових фармацевтичних інновацій.

Ще багато чого належить дізнатися про алкалоїди та їх корисну роль у системі охорони здоров'я. Морські та наземні біоми мають величезний, ще не вивчений потенціал для виявлення нових хімічних структур. Ера четвертої промислової революції (4IR) та модель П'ятикратної спіралі об'єднують можливості великих даних, штучного інтелекту, машинного навчання, лабораторних AI систем та обміну знаннями на основі спільних інтегрованих та екологічно чутливих програм. Це сприяє створенню нових, біологічно значущих молекул через сталу оцінку відомих алкалоїдів.

Необхідна нова міжнародна парадигма відкриття та розвитку природних продуктів, щоб задовольнити потреби у лікувальних агентах для боротьби з захворюваннями, стійкими до багатьох ліків, рідкісними та занедбаними тропічними хворобами, а також малярією та туберкульозом. Природні, фізичні та технологічні ресурси вже наявні або швидко розвиваються. Для трансформації існуючої парадигми та реалізації підходу, який оптимізує швидкі розробки 4IR з інноваціями, створеними в рамках П'ятикратної спіралі для цих захворювань, необхідні фінансові зобов'язання та міжнародні консорціуми. Третій клас ліків сприятиме цим розробкам і стане натхненням для інвестицій та інновацій. За допомогою застосування сучасних технологій через людську творчість, нові, переосмислені та перепрофільовані алкалоїди зроблять важливий внесок в ці та інші зусилля з відкриття ліків. Нова та захоплююча ера дослідження алкалоїдів вже розпочалася [38].

Алкалоїдні сполуки та їхні синтетичні аналоги проявляють міцну противірусну активність проти широкого спектру інфекційних та смертельних ДНК/РНК вірусів. Деякі з цих сполук мають концентрації IC50 менше ніж 5 мкМ і можуть запобігати вірусним інфекціям на понад 90 відсотків. Деякі алкалоїди

мають вищу противірусну активність, ніж деякі стандартні противірусні медикаменти, при нижчих концентраціях [39].

Гени унікальних ферментів, ізольовані з різноманітних організмів, є перспективним генетичним ресурсом для модифікації метаболічних шляхів. Величезна кількість реакцій, каталізованих ферментами, активними у вторинному метаболізмі, надає безпрецедентні можливості для вибору підходящих ферментів для використання в метаболічній інженерії. Інгібування метаболічного шляху може виявитися набагато більш прямолінійним. Існує кілька ефективних технік для придушення ендогенного генного вираження: косупресія, антисенс-придушення, інтерференція РНК та генне розривання з використанням оберненої генетики. Ці техніки повинні відкрити шлях для метаболічної інженерії лікарських рослин [40].

2. БІОІНФОРМАТИЧНІ МЕТОДИ АНАЛІЗУ

Науково-дослідну роботу проводили протягом 2022-2023 років у відділі молекулярної генетики та фітосанітарної експертизи ТОВ КОТЕКНА УКРАЇНА ЛІМІТЕД (м. Одеса).

2.1 Матеріали

2.1.1. Нуклеотидні послідовності аргемони

Нуклеотидні послідовності ITS1 (Internal Transcribed Spacer 1) представників роду Аргемона є значущим об'єктом для досліджень у сфері комп'ютерних наук, зокрема в біоінформатиці, з кількох причин:

1. Біорізноманіття і систематика: ITS1 є ділянкою в геномі, яка варіюється між видами, але є консервативною в межах виду. Аналіз цих послідовностей дозволяє точно ідентифікувати види і родини рослин, у тому числі різні види Аргемона. Це сприяє розумінню біорізноманіття і відносин між видами.

2. Філогенетичний аналіз: Комп'ютерні методи дозволяють проводити філогенетичний аналіз, який допомагає визначити еволюційні зв'язки між різними видами Аргемона. Це важливо для розуміння еволюційних шляхів та історії розвитку видів.

3. Молекулярна ідентифікація: У сучасній біотехнології та фармацевтиці точна ідентифікація видів є ключовою для забезпечення якості рослинних матеріалів. Комп'ютерний аналіз ITS1 послідовностей може допомогти у цьому, забезпечуючи точну і швидку ідентифікацію.

4. Розробка алгоритмів біоінформатики: Вивчення ITS1 послідовностей Аргемона може сприяти розвитку нових алгоритмів і методів у біоінформатиці, що будуть корисними для аналізу інших видів і геномних даних.

Комп'ютерні методи, такі як вирівнювання послідовностей, філогенетична реконструкція, і статистичний аналіз, дозволяють обробляти великі набори даних ITS1, отримуючи точні та значущі результати, які є неможливими для отримання вручну через об'єм даних та складність аналізу.

Використовували 40 нуклеотидних послідовностей внутрішнього транскрибованого спейсеру 1 (internal transcribed spacer 1, ITS1) представників роду Аргемона, отриманих із бази даних у базі даних Національного центру біотехнологічної інформації (National Center for Biotechnology Information, NCBI) [41] для дослідження філогенетичних взаємозв'язків та розробки дизайну праймерів для ПЛР у реальному часі, представлені у таблиці 2.1

Таблиця 2.1 - Нуклеотидні послідовності ITS1 представників роду Аргемона

Назва	Номер доступу
<i>Argemone mexicana</i>	MH734909.1
<i>Argemone mexicana</i>	MH768272.1
<i>Argemone munita</i>	MF963932.1
<i>Argemone albiflora</i>	JX078976.1
<i>Argemone mexicana</i>	AY328303.1
<i>Argemone mexicana subsp. mexicana</i>	AF057653.1
<i>Argemone ochroleuca</i>	AF057654.1
<i>Argemone mexicana subsp. mexicana</i>	AF057652.1
<i>Argemone superba</i>	AF057655.1
<i>Argemone hunnemannii</i>	AF057659.1
<i>Argemone corymbosa subsp. corymbosa</i>	AF057661.1
<i>Argemone corymbosa subsp. arenicola</i>	AF057660.1
<i>Argemone glauca</i>	AF057651.1
<i>Argemone rosea</i>	AF057658.1
<i>Argemone squarrosa</i>	AF057688.1
<i>Argemone subfusiformis</i>	AF057656.1
<i>Argemone echinata</i>	AF057689.1
<i>Argemone aurantiaca</i>	AF057667.1
<i>Argemone albiflora</i>	AF057686.1

Продовження таблиці 2.1

Назва	Номер доступу
<i>Argemone munita subsp. rotundata</i>	AF057679.1
<i>Argemone munita subsp. rotundata</i>	AF057678.1
<i>Argemone munita subsp. argentea</i>	AF057676.1
<i>Argemone grandiflora</i>	AF057674.1
<i>Argemone pleiacantha subsp. pinnatisecta</i>	AF057682.1
<i>Argemone platyceras</i>	AF057681.1
<i>Argemone munita subsp. robusta</i>	AF057677.1
<i>Argemone hispida</i>	AF057675.1
<i>Argemone gracilentata</i>	AF057672.1
<i>Argemone arida</i>	AF057664.1
<i>Argemone albiflora</i>	AF057663.1
<i>Argemone albiflora</i>	AF057662.1
<i>Argemone albiflora</i>	AF057670.1
<i>Argemone pleiacantha subsp. pleiacantha</i>	AF057684.1
<i>Argemone albiflora</i>	AF057685.1
<i>Argemone pleiacantha subsp. ambigua</i>	AF057683.1
<i>Argemone albiflora</i>	AF057673.1
<i>Argemone echinata</i>	AF057671.1
<i>Argemone albiflora</i>	AF057668.1
<i>Argemone arizonica</i>	AF057666.1
<i>Argemone arizonica</i>	AF057665.1

Для гомологічного моделювання використовували *in silico* трансльовані нуклеотидні послідовності генів *Argemone mexicana*, асоційовані із синтезом алкалоїдів, представлені у таблиці 2.2.

Таблиця 2.2 - Нуклеотидні послідовності генів, асоційованих із синтезом алкалоїдів *Argemone mexicana*

Назва гену	Номер доступу
<i>CYP719A14 cheilanthifoline synthase</i>	EF451152.1
<i>CYP719A13 stylophine synthase</i>	EF451151.1
<i>TyPPO tyrosinase polyphenol oxidase</i>	EU881892.1
<i>TyDC tyrosine/DOPA decarboxylase</i>	EU881888.1

Продовження таблиці 2.2

Назва гену	Номер доступу
<i>NCS1 S-norcochlorogenic acid synthase 1</i>	EU881891.1
<i>NCS2 S-norcochlorogenic acid synthase 2</i>	EU881893.1
<i>BBE1 berberine bridge enzyme 1</i>	EU881889.1
<i>BBE2 berberine bridge enzyme 2</i>	EU881890.1

Для пошуку тривимірних структур лігандів використовували базу даних PubChem [42-50].

2.1.2 Види аргемони

Для дослідження специфічності розробленої тести-системи використовували зразки цільового матеріалу (представників роду Аргемона) та нецільового матеріалу.

1. Цільовий референтний матеріал: референтний матеріал *Argemone albiflora*, *A. aurantiaca*, *A. hispida*, *A. munita*, *A. polyanthemos*, сертифікований референтний матеріал *A. mexicana*, *A. ochroleuca*, *A. squarrosa*, що супроводжується сертифікатом.

2. Нецільовий матеріал, що належить до цієї ж родини: контрольні зразки *Chelidonium majus*, *Eschscholzia californica* cv Crimson King, *Eschscholzia californica* cv Purple, *Papaver croceum*.

3. Нецільовий матеріал, що належить до інших родин: контрольні зразки *Arachis hypogaea*, *Helianthus annuus*, *Hordeum vulgare*, *Lupinus* spp., *Ricinus communis*, *Secale cereale*, *Triticum aestivum*, сертифікований референтний матеріал *Brassica napus*, *Glycine max*, *Oryza sativa*, *Zea mays*, що супроводжується сертифікатом і статистичними характеристиками.

Список зразків, що використано в роботі представлено у таблиці 2.3.

Таблиця 2.3 - Референтний матеріал та контрольні зразки для перевірки специфічності

Назва	Родина	Джерело
Цільовий матеріал		
<i>Argemone albiflora</i>	<i>Papaveraceae</i>	USDA*
<i>Argemone aurantiaca</i>	<i>Papaveraceae</i>	USDA
<i>Argemone hispida</i>	<i>Papaveraceae</i>	USDA
<i>Argemone mexicana</i>	<i>Papaveraceae</i>	Ботанічний сад ім. Гришка**
<i>Argemone munita</i>	<i>Papaveraceae</i>	USDA
<i>Argemone ochroleuca</i>	<i>Papaveraceae</i>	Ботанічний сад ім. Гришка
<i>Argemone polyanthemus</i>	<i>Papaveraceae</i>	USDA
<i>Argemone squarrosa</i>	<i>Papaveraceae</i>	Ботанічний сад ім. Гришка
Нецільовий матеріал, що належить до цієї ж родини		
<i>Chelidonium majus</i>	<i>Papaveraceae</i>	Контрольний зразок
<i>Eschscholzia californica</i> cv Crimson King	<i>Papaveraceae</i>	Контрольний зразок
<i>Eschscholzia californica</i> cv Purple	<i>Papaveraceae</i>	Контрольний зразок
<i>Papaver croceum</i>	<i>Papaveraceae</i>	Контрольний зразок
Нецільовий матеріал, що належить до інших родин		
<i>Arachis hypogaea</i>	<i>Fabaceae</i>	Контрольний зразок
<i>Brassica napus</i> 73496	<i>Brassicaceae</i>	ERM-BF434a***
<i>Glycine max</i> DAS-4406-6	<i>Fabaceae</i>	ERM-BF436a
<i>Helianthus annuus</i>	<i>Asteraceae</i>	Контрольний зразок
<i>Hordeum vulgare</i>	<i>Triticaceae</i>	Контрольний зразок
<i>Lupinus</i> spp.	<i>Fabaceae</i>	Контрольний зразок
<i>Oryza sativa</i>	<i>Triticaceae</i>	USDA
<i>Ricinus communis</i>	<i>Euphorbiaceae</i>	Контрольний зразок
<i>Secale cereale</i>	<i>Triticaceae</i>	Контрольний зразок
<i>Triticum aestivum</i>	<i>Triticaceae</i>	USDA
<i>Zea mays</i> 5307	<i>Triticaceae</i>	AOCS 0411-C2****

*- United States Department of Agriculture (USDA-ARS) Germplasm Resources Information Network (GRIN)

** - Національний ботанічний сад імені М.М. Гришка, Київ, Україна

***- European Reference Materials (ERM) consortium

**** - American Oil Chemists' Society (AOCS)

2.2 Біоінформатичні методи

2.2.1 Вирівнювання нуклеотидних та амінокислотних послідовностей

В процесі вирівнювання нуклеотидних послідовностей були використані два ключові алгоритми: локальне вирівнювання за алгоритмом Сміта-Уотермана та глобальне вирівнювання за алгоритмом Нідлмана-Вунша. Алгоритм Сміта-Уотермана [51], ефективно виявляє найбільш схожі ділянки між двома послідовностями, забезпечуючи локальне вирівнювання. Це досягається шляхом оцінки подібності сегментів послідовностей і вибору оптимального шляху через матрицю оцінок. Для цього була застосована онлайн-програма «blastn» від National Center for Biotechnology Information (NCBI).

Алгоритм Нідлмана-Вунша [52], використався для глобального вирівнювання, тобто визначення оптимального вирівнювання всієї довжини обох послідовностей. Цей алгоритм враховує всі можливі вирівнювання і вибирає те, яке максимізує загальний бал подібності. Глобальне вирівнювання було здійснено за допомогою підпрограми «clustalW» програми «Mega».

2.2.2 Філогенетична реконструкція

Реконструкцію філогенетичної дендрограми за результатами вирівнювання проводили за UPGMA (Unweighted Pair Group Method with Arithmetic mean), описаний у джерелі [53]. Метод UPGMA є одним з найпростіших методів для побудови філогенетичних дерев, він базується на концепції мінімізації середнього арифметичного відстані між всіма парами елементів у кожній групі.

UPGMA використовує матрицю відстаней, отриману з результатів вирівнювання нуклеотидних або амінокислотних послідовностей, для побудови дендрограми. У цій дендрограмі гілки відображають еволюційні відносини, а

довжина гілок корелює з еволюційною відстанню. Метод UPGMA припускає константну швидкість мутацій по всьому дереву, що може бути обмеженням для деяких наборів даних. [53]. Філогенетичну дендрограму будували за допомогою програми MEGA 5.2 [54].

2.2.3 Гомологічне моделювання структури протеїнів

Гомологічне моделювання є ключовим методом у структурній біоінформатиці, який дозволяє прогнозувати тривимірну структуру білків на основі їх амінокислотних послідовностей. У цьому дослідженні використовувався сервер SWISS-MODEL [55-59], для гомологічного моделювання структури ензимів, асоційованих з синтезом алкалоїдів. SWISS-MODEL — це автоматизований веб-сервіс, який дозволяє користувачам здійснювати моделювання білкових структур, виходячи з їхніх амінокислотних послідовностей, шляхом порівняння з відомими структурами гомологічних білків.

Для аналізу взаємодій між молекулами білків та іншими молекулами (наприклад, лігандами) використовувався молекулярний докінг за допомогою веб-сервісу SwissDock [60]. SwissDock дозволяє виконувати докінг молекул у віртуальному середовищі, що дає можливість оцінити потенційні молекулярні взаємодії та їхній вплив на функціональні властивості білків.

Для візуалізації отриманих результатів гомологічного моделювання та молекулярного докінгу було використано програму UCSF Chimera [61]. UCSF Chimera є потужним інструментом для візуалізації молекулярних структур, що дозволяє детально відображати тривимірні моделі білків, їхні взаємодії з іншими молекулами та аналізувати структурні особливості. Цей інструмент дозволяє не лише візуально оцінити моделі білків, але й проводити різноманітні аналітичні

процедури, такі як вимірювання відстаней між атомами, оцінка взаємодій між білками та лігандами, а також інші типи структурних аналізів.

2.2.4 Дизайн праймерів та *in silico* ПЛР

Дизайн праймерів проводили за допомогою програми Primer3 за результатами глобального вирівнювання нуклеотидних послідовностей ITS1. Праймери добирали до ділянок, що виявили найбільший рівень поліморфізму. Цей процес включає вибір оптимальних послідовностей праймерів, які забезпечують найкращу ефективність та специфічність для ПЛР (полімеразної ланцюгової реакції). Особливу увагу приділяли параметрам, таким як температура плавлення, довжина праймера, вміст GC та уникнення формування димерів. *In silico* ПЛР проводили також в програмі Primer3. *In silico* ПЛР дозволяє моделювати ПЛР-реакцію у віртуальному середовищі, що дає змогу оцінити потенційну ефективність та специфічність праймерів, перш ніж проводити реальні експерименти. Цей етап є критично важливим для перевірки та оптимізації умов ПЛР, забезпечуючи високу точність та надійність експериментальних даних.

2.3 Екстрагування та очистка ДНК аргемони

Екстрагування ДНК проводили за допомогою набору для екстракції ДНК ДНК «SureFood PREP Advanced» (CONGEN Biotechnologie GmbH, Німеччина) згідно керівництву користувача.

2.4 Полімеразна ланцюгова реакція

Полімеразну ланцюгову реакцію (ПЛР) проводили в 200 мкл мікропробірках в ампліфікаторі Термоциклер QuantStudio 5 Real-Time PCR System (Applied Biosystems by Thermo Fisher Scientific, США). Реакційна суміш містила iTaq Universal Probes Supermix (2x) (BioRad, США).

Температурний та часовий режим для проведення ПЛР такий: перша денатурація 95 °C 5 хв; 40 циклів: денатурація 95 °C 1 хв; відпалювання праймерів і елонгація 60 °C 1 хв; зчитування флуоресцентного сигналу.

Праймери 10 мкМ (прямий і зворотний) і зонд 10 мкМ, синтезовані компанією Metabion, Німеччина.

Результати ПЛР у реальному часі аналізували за допомогою програмного забезпечення QuantStudio TM Design and Analysis Software VERSION 1.4.1.

2.5 Загальна схема роботи

Загальна схема роботи представлена на рис. 2.1

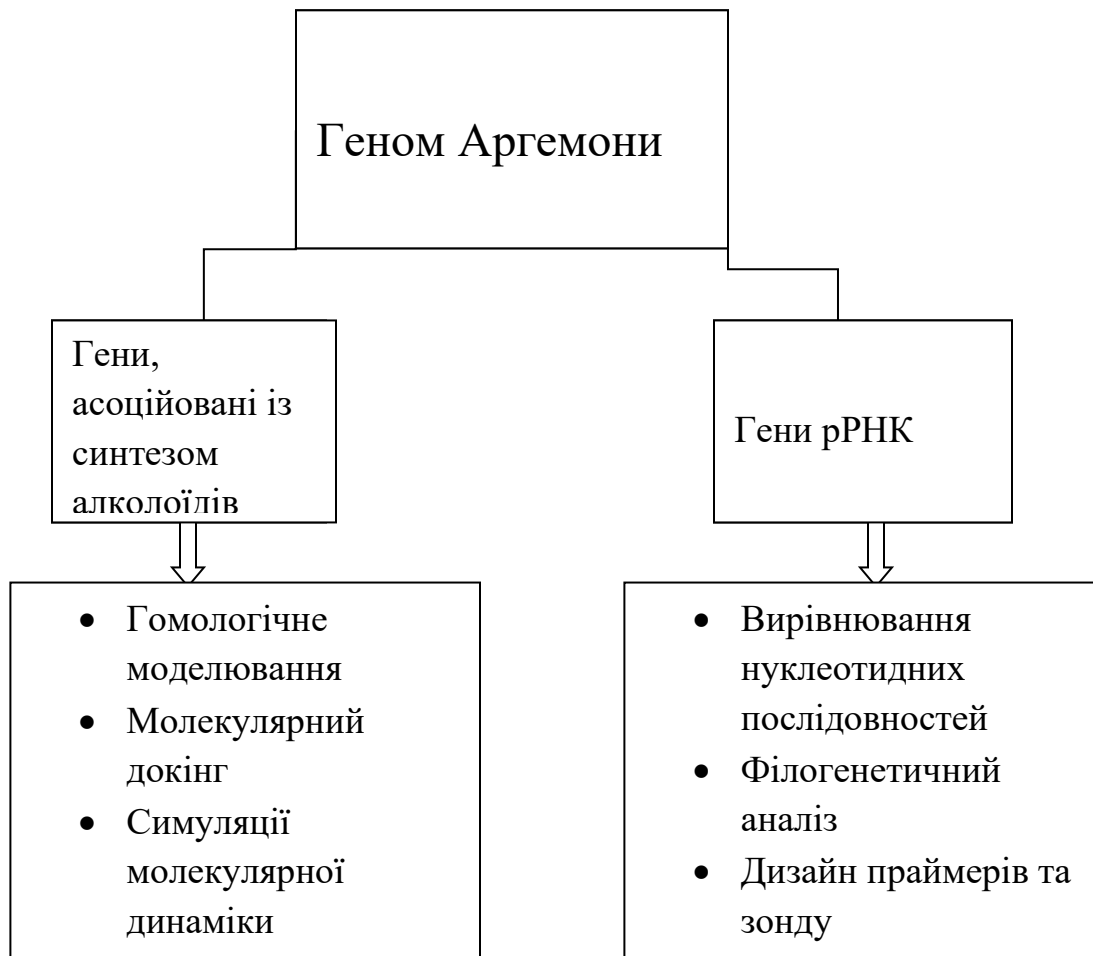


Рисунок 2.1 - Загальна схема молекулярно-генетичного та біоінформатичного аналізу геному Аргемони

3 ФІЛОГЕНЕТИЧНИЙ АНАЛІЗ ПРЕДСТАВНИКІВ РОДУ АРГЕМОНА

3.1 Розробка праймерів і зондів та *in silico* ПЛР

Проведено вирівнювання 40 нуклеотидних послідовностей ITS1 представників роду Аргемона. За результатами вирівнення проведена реконструкція філогенетичної дендрограми (рисунок 3.1). Еволюційна історія була вивчена за допомогою методу UPGMA. Найкраще дерево представлено. Дерево намальовано у масштабі, з довжинами гілок у тих же одиницях, що й еволюційні відстані, які були використані для виведення філогенетичного дерева. Еволюційні відстані були розраховані за допомогою методу Максимальної Складної Ймовірності та подані у одиницях числа заміни основ на сайт. Цей аналіз включав 39 нуклеотидних послідовностей. Позиції кодонів, що були включені, це 1-ша+2-га+3-тя+Некодуєчі. Всі неоднозначні позиції були вилучені для кожної пари послідовностей (опція парного видалення). У фінальному наборі даних було всього 727 позицій. Еволюційний аналіз проводився у MEGA11.

Вивчення нуклеотидних послідовностей ITS1 (Internal Transcribed Spacer 1) у представників роду *Argemone* є важливою частиною молекулярної таксономії. Ці послідовності відіграють ключову роль у визначенні відносин між видами, оскільки вони часто містять унікальні варіації, що дозволяють розрізняти навіть дуже близькі види. Результати вирівнювання послідовностей ITS1 роду *Argemone* показали високий рівень відповідності із традиційною таксономічною класифікацією, заснованою на морфологічних ознаках та інших біологічних даних. Це свідчить про те, що молекулярні методи можуть ефективно підтримувати і навіть уточнювати таксономічне розміщення видів в рамках роду, забезпечуючи більш глибоке розуміння їх еволюційних зв'язків.

В ході молекулярних досліджень було виявлено, що регіон ITS1, що є

частиною рибосомної ДНК, містить унікальні послідовності, специфічні для представників роду *Argemone*. Ця знахідка стала основою для розробки спеціалізованих праймерів та зондів, які можуть бути використані для швидкого та точного ідентифікування видів цього роду за допомогою полімеразної ланцюгової реакції (ПЛР) та інших молекулярно-діагностичних методів. Дизайн праймеру був оптимізований для ампліфікації саме цього унікального регіону, а TaqMan зонд дозволяє високоспецифічно гібридизуватися з ним, забезпечуючи точність визначення. Цей підхід значно підвищує ефективність таксономічної ідентифікації та вивчення філогенії в межах роду, а також може служити зразком для подібних досліджень в інших групах організмів.

Проведення *in silico* ПЛР з використанням спеціально розроблених праймерів і зондів для роду *Argemone* підтвердило їхню високу родо-специфічність. Цифрове моделювання реакцій ампліфікації виявило, що праймери та зонди реагують виключно на ДНК, що є характерною для представників роду *Argemone*, без перехресної реакції з ДНК інших видів. Такий результат свідчить про високу точність і працездатність розробленої тест-системи, яка може бути використана для надійної ідентифікації ДНК представників роду *Argemone*.

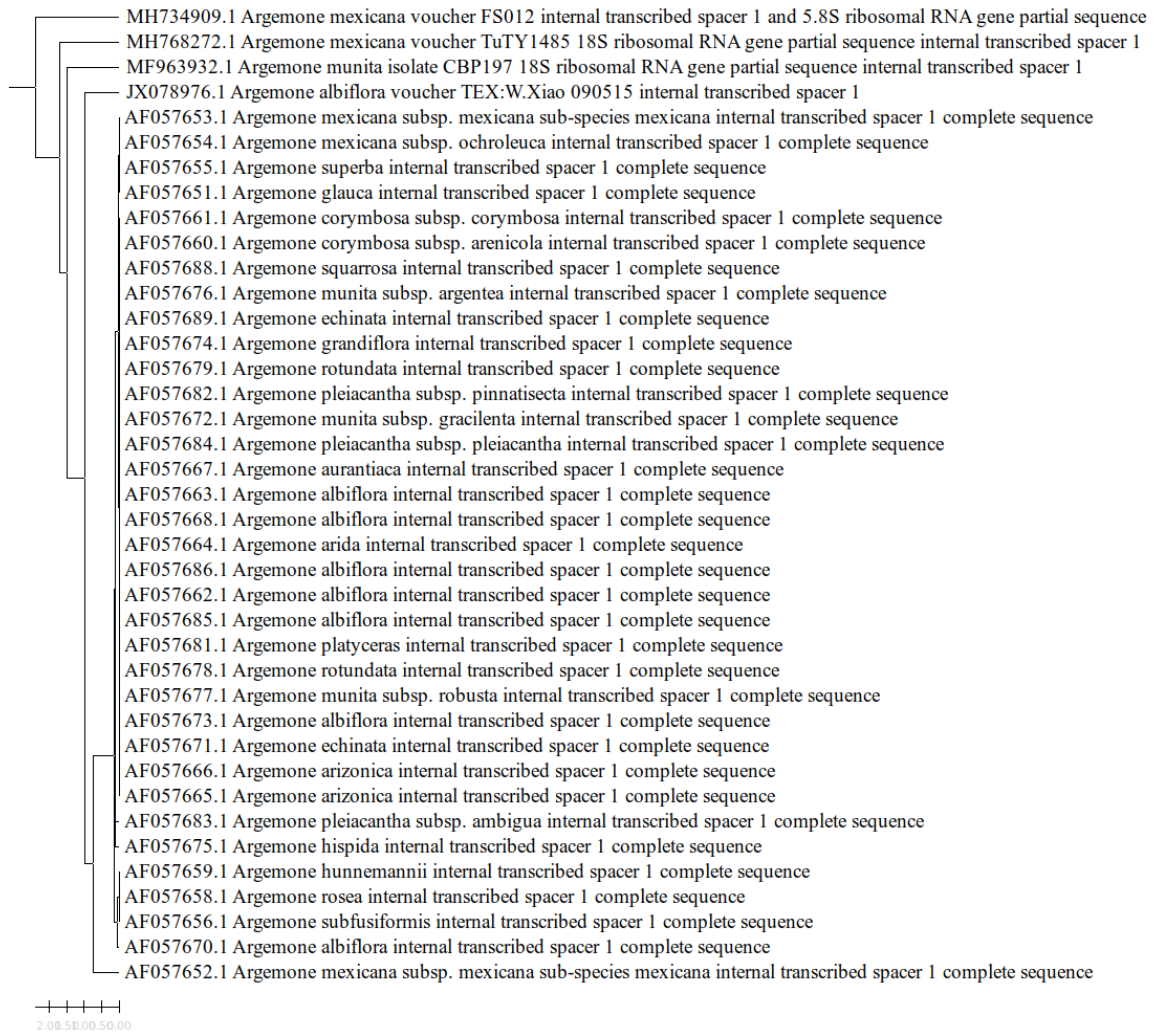


Рисунок 3.1 - Філогенетична дендрограма

3.2 *In vitro* ПЛР-аналіз

Показана працездатність розроблених праймерів та зонду у *in vitro* ПЛР аналізі. З метою оцінки якості розробленого дизайну праймерів та зонду проведена перевірка специфічності розроблених праймерів та зонду.

Перевірку специфічності (перехресної реакційної здатності системи виявлення) шляхом детекції наявності ампліфікації здійснено на цільовому та нецільовому матеріалі:

- цільовому матеріалі *Argemone albiflora*, *A. aurantiaca*, *A. hispida*, *A. mexicana*, *A. munita*, *A. ochroleuca*, *A. polyanthemos*, *A. squarrosa*;

- нецільовому матеріалі, що належить до цієї ж родини: *Chelidonium majus*, *Eschscholzia californica* cv *Crimson King*, *Eschscholzia californica* cv *Purple*, *Papaver croceum*;

- нецільовому матеріалі, що належить до інших родин: контрольні зразки *Arachis hypogaea*, *Brassica napus*, *Glycine max*, *Helianthus annuus*, *Hordeum vulgare*, *Lupinus* spp., *Oryza sativa*, *Ricinus communis*, *Secale cereale*, *Triticum aestivum*, *Zea mays*.

Результати представлено в таблиці 3.1.

Таблиця 3.1 - Результати перевірки специфічності системи *Argemone* spp.

Матеріал для тестування	Наявність (+) / відсутність (-) флуоресцентного сигналу при дослідженні послідовності ДНК, специфічної для	
	<i>Argemone</i> spp.	еукаріот
<i>Argemone albiflora</i>	+	+
<i>Argemone aurantiaca</i>	+	+
<i>Argemone hispida</i>	+	+
<i>Argemone mexicana</i>	+	+
<i>Argemone munita</i>	+	+
<i>Argemone ochroleuca</i>	+	+
<i>Argemone polyanthemos</i>	+	+
<i>Argemone squarrosa</i>	+	+
<i>Chelidonium majus</i>	-	+
<i>Papaver croceum</i>	-	+
<i>Eschscholzia californica</i> cv <i>Crimson King</i>	-	+
<i>Eschscholzia californica</i> cv <i>Purple</i>	-	+
<i>Arachis hypogaea</i>	-	+
<i>Brassica napus</i>	-	+
<i>Glycine max</i>	-	+
<i>Helianthus annuus</i>	-	+
<i>Hordeum vulgare</i>	-	+

Продовження таблиці 3

Матеріал для тестування	Наявність (+) / відсутність (-) флуоресцентного сигналу при дослідженні послідовності ДНК, специфічної для	
	<i>Argemone</i> spp.	еукаріот
<i>Lupinus</i> spp.	-	+
<i>Oryza sativa</i>	-	+
<i>Ricinus communis</i>	-	+
<i>Secale cereale</i>	-	+
<i>Triticum aestivum</i>	-	+
<i>Zea mays</i>	-	+

Здатність до ампліфікації методом ПЛР-аналізу ДНК, що екстрагована із зразків для перевірки відповідності параметрам валідації, підтверджено наявністю ампліфікації при використанні системи, специфічної для гену 18S рРНК еукаріот.

Наявність ампліфікації при використанні тест-системи *Argemone* spp. встановлено тільки для послідовності ДНК, специфічної для представників роду *Argemone*.

Висновок: наявність флуоресцентного сигналу встановлено при аналізі тільки цільового матеріалу – *Argemone albiflora*, *A. aurantiaca*, *A. hispida*, *A. mexicana*, *A. munita*, *A. ochroleuca*, *A. polyanthemos*, *A. squarrosa*. Система є родоспецифічною.

4 БІОІНФОРМАТИЧНИЙ АНАЛІЗ ГЕНІВ, АСОЦІЙОВАНИХ З БІОСИНТЕЗОМ ВТОРИННИХ МЕТАБОЛІТІВ

4.1 Біоінформатичний аналіз гена *CYP719A14*

За результатами гомологічного моделювання побудовано модель ензиму *CYP719A14*, що проявляє хейліантіфолінсинтазну активність (рис. 4.1). У якості моделі для гомологічного моделювання слугував протеїн 6b82.1.B Cytochrome P450, family 17, subfamily A, polypeptide 1 Zebra Fish CYP-450 17A1 Mutant Abiraterone Complex. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,60, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,63 \pm 0,05$, що відповідало критеріям прийнятності.

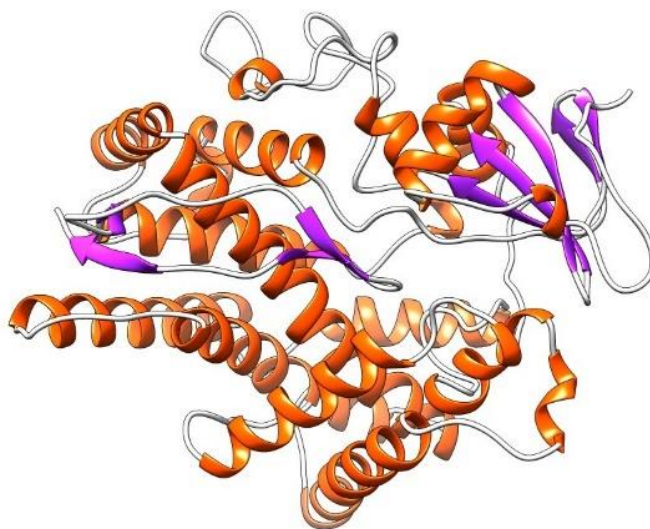


Рис. 4.1 - Модель ензиму *CYP719A14*, що проявляє хейліантіфолінсинтазну активність

Побудовану методом гомологічного моделювання структуру ензиму *CYP719A14* порівняли із структурою, реконструйованою за допомогою

штучного інтелекту AlphaFold B1NF20. Структури були майже тотожні, відрізнялася суттєво лише конформація ділянки протеїну MDETIWLIISTVHIVLGLIAKFLLGKSSSSSLSTMEWPVGPKK (рис. 4.2, 4.3)

Враховуючи той факт, що для цитохромів є характерним наявність порожнини, що є скритою від розчинника, а також те, що активний центр протеїну розташований глибоко в корі протеїну, проводили ідентифікацію порожнини, що проявляє афінитет до потенційного субстрату та продукту – каналу доступу ліганда. Для ідентифікації порожнини, що проявляє афінитет до потенційного субстрату та продукту – каналу доступу ліганда, проводили молекулярний докінг моделі ензиму із ймовірними субстратами-похідними бензилізохиноліну: (S)-Скоулеріном, (S)-Хейліантіфоліном, (S)-Тетрагідраколумбаміном, (S)-Ретикуліном. (S)-Скоулерін є субстратом для хейліантіфолінсинтазної активності, (S)-Хейліантіфолін є продуктом, а (S)-Тетрагідраколумбамін та (S)-Ретикулін є субстратами споріднених ензимів.

Для (S)-Скоулеріну ідентифіковано 30 кластерів докінгу. Показник FullFitness варіював від -2526,65 до -2550,23 ккал/моль, вільна енергія Гібса ΔG варіювала від -7,88 до -5,92 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками PHE57-GLY69, ALA285-ARG310, THR369-HSD390, що ймовірно й формують потенційний канал доступу ліганду (рис. 4.4, А). У потенційного каналу доступу лігандів показник FullFitness варіював від -2536,62 до -2550,23, вільна енергія Гібса ΔG варіювала від -7,88 до -6,95.

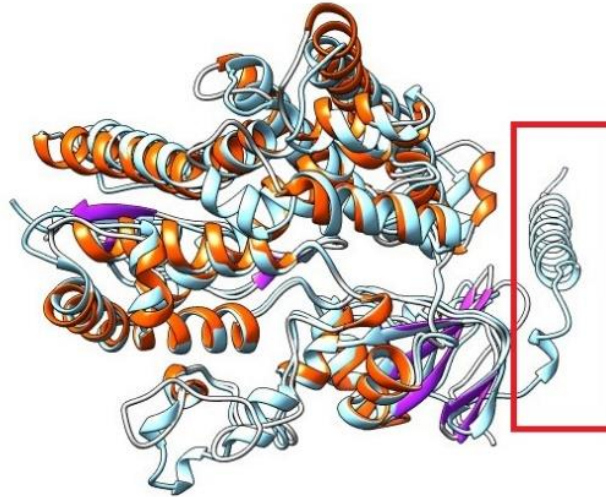


Рис. 4.2 - Порівняння структур, побудованих методом гомологічного моделювання і AlphaFold. Регіон, що відрізнявся, помічений прямокутником

	1	11	21	31	41
RMSD: ca					
model.pdb, chain B	35 E				W . P V G P K K . L P I I G N
AF-B1NF20-F1-m...v4.pdb, chain A	1 M D E T I W L I I S T V I I V L G I A K F L L G K S S S S L S T M E W P V G P K K L P I I				
RMSD: ca					
model.pdb, chain B	49 L H Q L G . . G D . . V F H V V L A N L A K V Y G S V F T I W V G S . W R P M I I V S D I D K A W E				
AF-B1NF20-F1-m...v4.pdb, chain A	46 I G . N L H Q L G G D V F H V V L A N L A K V Y G S V F T I W V G S W R P M I I V S D I D K A W E				
RMSD: ca					
model.pdb, chain B	94 V L V N K S S D Y S A R D M P D I T K I I S A N W K N I S C S D S G P F W H N L R K G L Q . G V A .				
AF-B1NF20-F1-m...v4.pdb, chain A	94 V L V N K S S D Y S A R D M P D I T K I I S A N W K N I S C S D S G P F W H N L R K G L Q G V A L T				
RMSD: ca					
model.pdb, chain B	142 . . L T P L N V . . A S Q Y H L Q E R D M K N L I N S M Y K D . . A S R K N G . . I L . K P L D				
AF-B1NF20-F1-m...v4.pdb, chain A	144 P L N V A S Q Y H L Q E R D M K N L I N S M Y K D A S R K N G . I L K P L D				
RMSD: ca					
model.pdb, chain B	181 Y L K E E T V R L L S R L I F . . G Q D F Q D E K L V V G . . M H H A L D D L V R I S G Y				
AF-B1NF20-F1-m...v4.pdb, chain A	181 Y L K E E T V R L L S R L I F G Q D F Q D E K L V V . G M H H I A L D D L V R I S G Y				
RMSD: ca					
model.pdb, chain B	222 A S L A D A . F I K F . . C E N L P S H . . K K S I R E V H E V K K R V E N . L I R P . H I V S . N				
AF-B1NF20-F1-m...v4.pdb, chain A	222 A S L A D A F I K . . F C E . . N L P S H K K S I R E V H E V K K R V E N L I R P H I V . S N				
RMSD: ca					
model.pdb, chain B	263 P P T N . . T Y L Y F L K T Q . D F N E D I I I S A I L E V Y D L G V D S T A S T T V W A L T F L V				
AF-B1NF20-F1-m...v4.pdb, chain A	263 P P T N T Y L Y F L K T Q D F N E D I I I S A I L E V Y D L G V D S T A S T T V W A L T F L V				

Рис. 3 - Порівняння структур, побудованих методом гомологічного моделювання і AlphaFold – амінокислотне вирівнювання

Для (S)-Хейліантіфоліну ідентифіковано 33 кластери докінгу. Показник FullFitness варіював від -2531,56 до -2553,06 ккал/моль, вільна енергія Гібса ΔG варіювала від -7,61 до -5,58 ккал/моль. Більшість кластерів докінгу групувалася

також у потенційному каналі доступу лігандів, як і у випадку із (S)-Скоулеріном (рис. 4.4, Б).

Для (S)-Тетрагідраколумбаміну ідентифіковано 30 кластерів докінгу. Показник FullFitness варіював від -2515,15 до -2537,23 ккал/моль, вільна енергія Гібса ΔG варіювала від -8,18 до -6,31 ккал/моль. Більшість кластерів докінгу групувалася також у потенційному каналі доступу лігандів, як і у випадку із (S)-Скоулеріном (рис. 4.4, В).

Для Ретикуліну ідентифіковано 26 кластерів докінгу. Показник FullFitness варіював від -2517,81 до -2542,82 ккал/моль, вільна енергія Гібса ΔG варіювала від -8,18 до -6,31 ккал/моль. Всі кластери докінгу групувалася лише у потенційному каналі доступу лігандів, як і у випадку із (S)-Скоулеріном (рис. 4.4, Г).

Зважаючи на те, що більшість кластерів для всіх лігандів групувалося саме у порожнині, що формувалася амінокислотними залишками PHE57-GLY69, ALA285-ARG310, THR369-HSD390, можна зробити припущення, що саме цю порожнину ідентифіковано як потенційний канал доступу ліганду і вказує на адекватність побудованої моделі ензиму.

З метою валідації побудованої моделі провели симуляцію молекулярної динаміки за допомогою розробленої нами програми для симуляції МД. Аналізували середньоквадратичне відхилення (Root-mean-square deviation, RMSD) та стабільність протеїну в ході візуалізації траєкторій атомів.

Проводили симуляцію в 10000 ітерацій, силове поле amber14-all, у якості температур використовували 300 К, 350 К, 400 К.

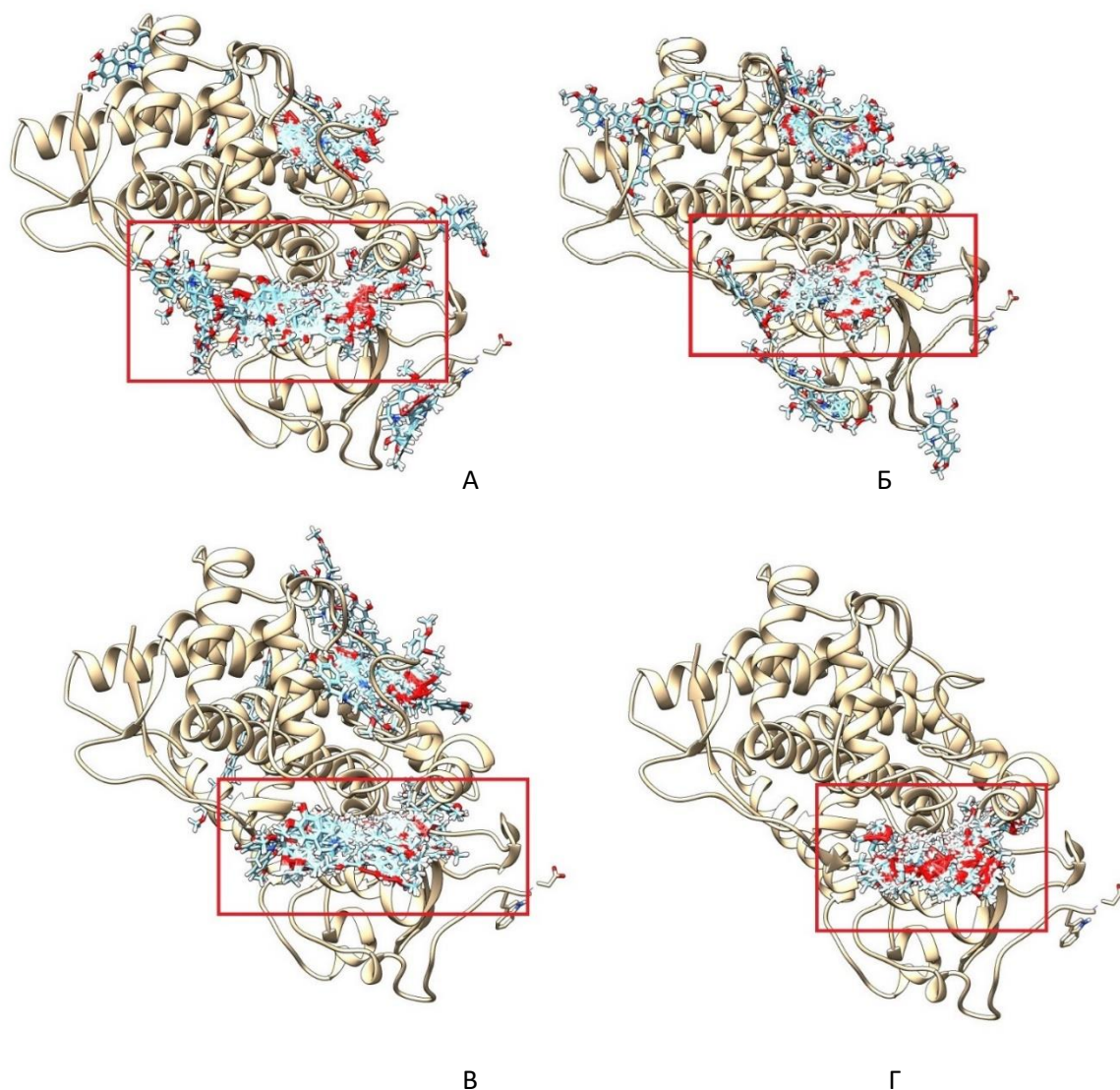


Рис. 4.4 - Кластери молекулярного докінгу А - (S)-Скоулеріну, Б - (S)-Хейліантіфоліну, В - (S)-Тетрагідраколумбаміну, Г - Ретикуліну. Потенційний канал доступу ліганду позначено прямокутником

За 300 К значення RMSD не перевищувало 2.26 А (рисунок 4.5, А), за 350 К - 2.59 А (рисунок 4.5, Б), за 400 К - 2.56 А (рисунок 4.5, В).

За всіх досліджених температурах спостерігалась цілісність протеїну, рух атомів не призводив до порушення структури протеїну.

Низькі значення RMSD, які не перевищували 4.0 та мало зростали із зростанням температури симуляції, а також той факт, що конформація протеїну зберігалася за підвищення температури симуляції, рух атомів не призводив до порушення цілісності протеїну говорить про адекватність побудованої моделі протеїну CYP719A14

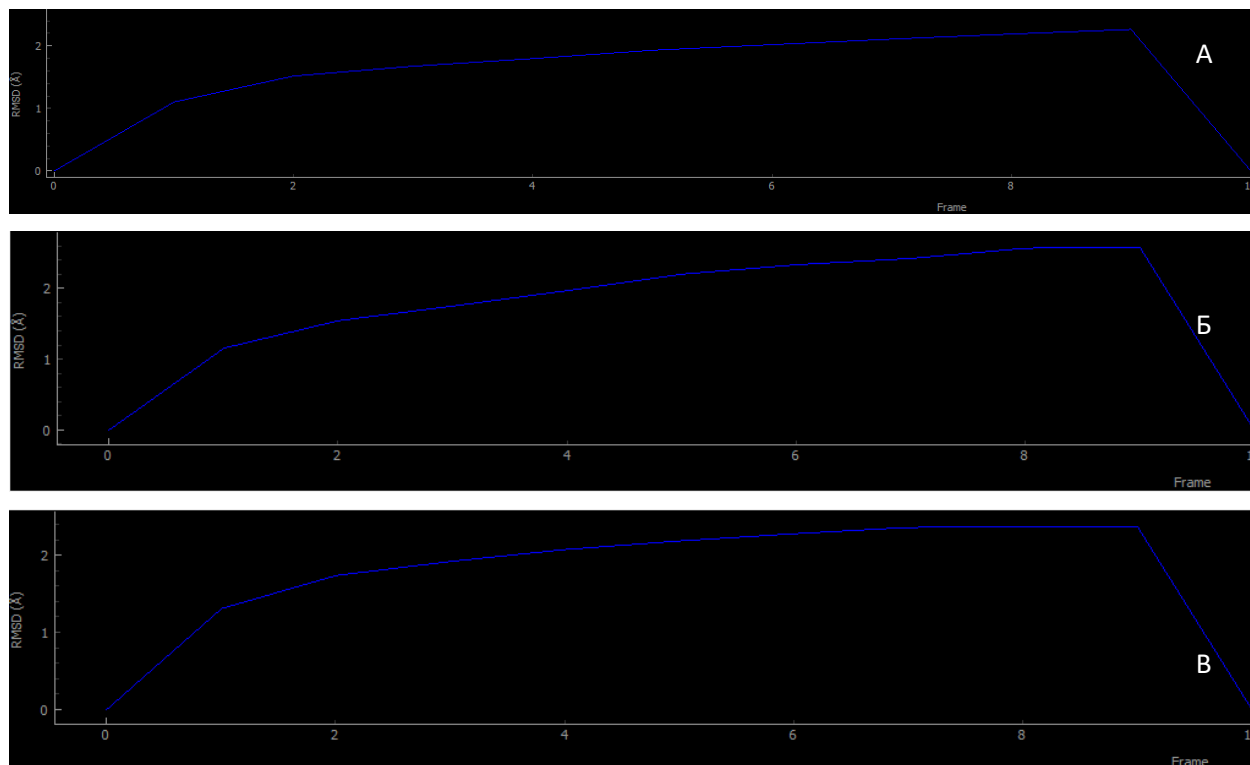


Рисунок 4.5 - Графік RMSD, температура А – 300 К, Б - 350 К, В – 400 К.

4.2 Біоінформатичний аналіз гена CYP719A13

За результатами гомологічного моделювання побудовано модель ензиму CYP719A13, що проявляє стилопінсинтазну активність (рис. 4.6). У якості моделі для гомологічного моделювання слугував протеїн B1NF19.1.A Trifunctional (S)-stylopine synthase/(S)-nandinine synthase/(S)-canadine synthase. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,90, аналіз якості енергії моделі (Qualitative Model

Energy Analysis, QMEANDisCo Global) дорівнював $0,65 \pm 0,05$, що відповідало критеріям прийнятності.

Як і для протеїну, що виявляє хейліантіфолінсинтазну активність, який теж є цитохромом, провели ідентифікацію порожнини. Що проявляє афінитет до субстрату і продукту реакції. Для ідентифікації порожнини, що проявляє афінитет до потенційного субстрату та продукту – каналу доступу ліганда, проводили молекулярний докінг моделі ензиму із ймовірними субстратами-похідними бензилізохиноліну: (S)-Хейліантіфоліном та (S)-Стилопіном. (S)-Хейліантіфолін є субстратом для хейліантіфолінсинтазної активності, а (S)-Стилопін є продуктом.

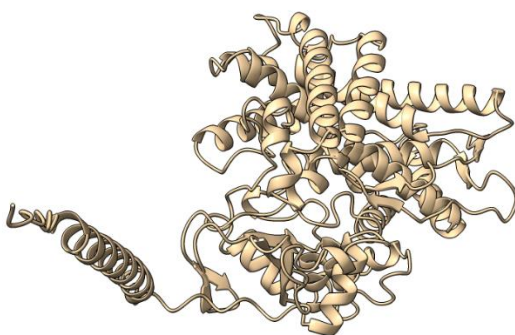


Рисунок 4.6 - Модель ензиму CYP719A13, що проявляє стилопінсинтазну активність

Для (S)-Хейліантіфоліну ідентифіковано 31 кластерів докінгу. Показник FullFitness варіював від -2106,73 до -2119,93 ккал/моль, вільна енергія Гібса ΔG варіювала від -6,2 до -7,81 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками GLN157-HIS221, TYR289-GLN327, що ймовірно й формують потенційний канал доступу ліганду

(рис. 4.7). У потенційного каналу доступу лігандів показник FullFitness варіював від -2112,88 до -2119,93, вільна енергія Гібса ΔG варіювала від -7,17 до -7,81.

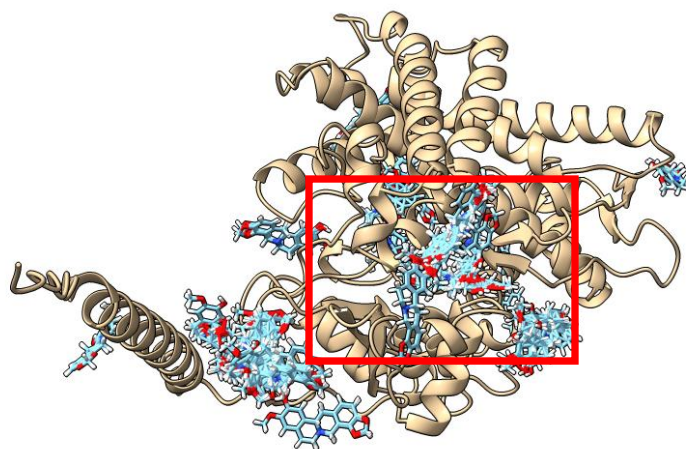


Рисунок 4.7 - Кластери молекулярного докінгу (S)-Хейліантіфоліну.
Потенційний канал доступу ліганду позначено прямокутником

Для (S)-Стилопіну ідентифіковано 33 кластери докінгу. Показник FullFitness варіював від -2109,05 до -2121,88 ккал/моль, вільна енергія Гібса ΔG варіювала від -7,73 до -5,72 ккал/моль. Більшість кластерів докінгу групувалася також у потенційному каналі доступу лігандів, як і у випадку із (S)-Хейліантіфоліном (рис. 4.8).

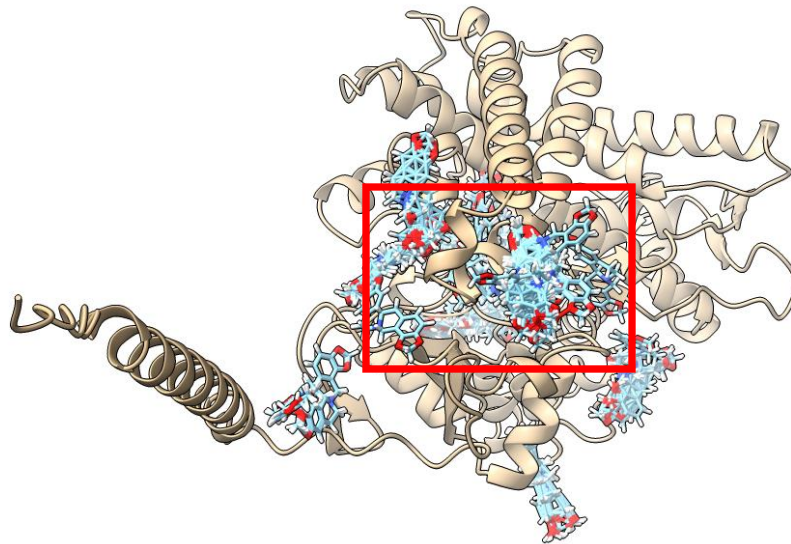


Рисунок 4.8 - Кластери молекулярного докінгу (S)-Стилопіну. Потенційний канал доступу ліганду позначено прямокутником

Зважаючи на те, що більшість кластерів для всіх лігандів групувалося саме у порожнині, що формувалася амінокислотними залишками GLN157-HIS221, TYR289-GLN327, можна зробити припущення, що саме цю порожнину ідентифіковано як потенційний канал доступу ліганду і вказує на адекватність побудованої моделі ензиму.

4.3 Біоінформатичний аналіз гена TyPPO

За результатами гомологічного моделювання побудовано модель ензиму TyPPO, що проявляє тирозинполіфенол-оксидазну активність (рис. 4.9). У якості моделі для гомологічного моделювання слугував протеїн A0A2G5EJL9.1.A Tyrosinase copper-binding domain-containing protein. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate,

GMQE) дорівнював 0,86, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,68 \pm 0,05$, що відповідало критеріям прийнятності.

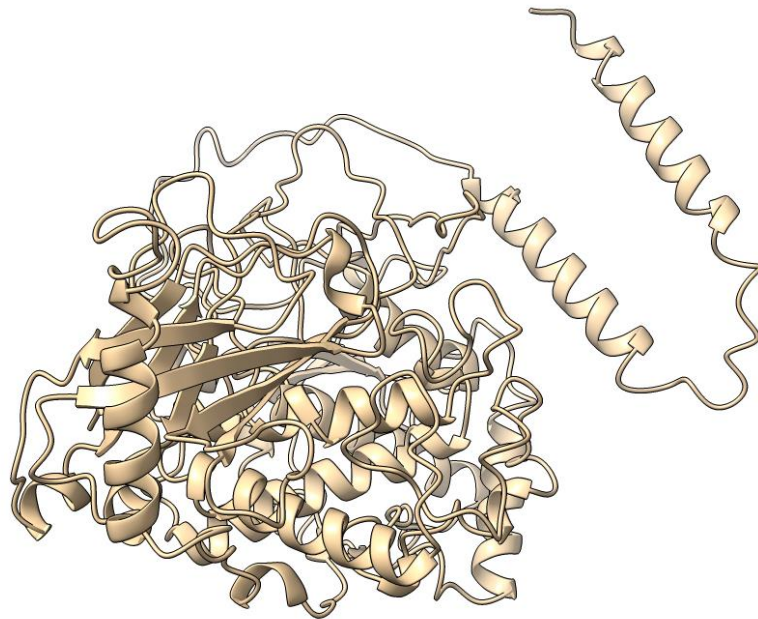


Рисунок 4.9 - Модель ензиму TyRPO, що проявляє тирозинполіфенол-оксидазну активність

Для ідентифікації потенційного активного центру побудованої моделі проведено молекулярний докінг із двома субстратами — тирозином та леводопа, а також вирівнювання із відомим активним центром тирозинази *Aspergillus spp.*

Для тирозину ідентифіковано 62 кластерів докінгу. Показник FullFitness варіював від -2583,81 до -2582,61 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,48 до -7,15 ккал/моль. Так як всі кластери докінгу були примірно рівні і не відрізнялися суттєво по термодинамічним характеристиками і кількістю позицій, ідентифікацію активного центру проведено за результатами вирівнювання. За результатами вирівнювання активний центр ензиму формувався

амінокислотними залишками HIS83-ARG100, GLN21-VAL54 та LEU313-TYR341 (рис. 4.10). У потенційного активного центру показник FullFitness варіював від -2589,59 до -2578,49, вільна енергія Гібса ΔG варіювала від -5,75 до -6,43.

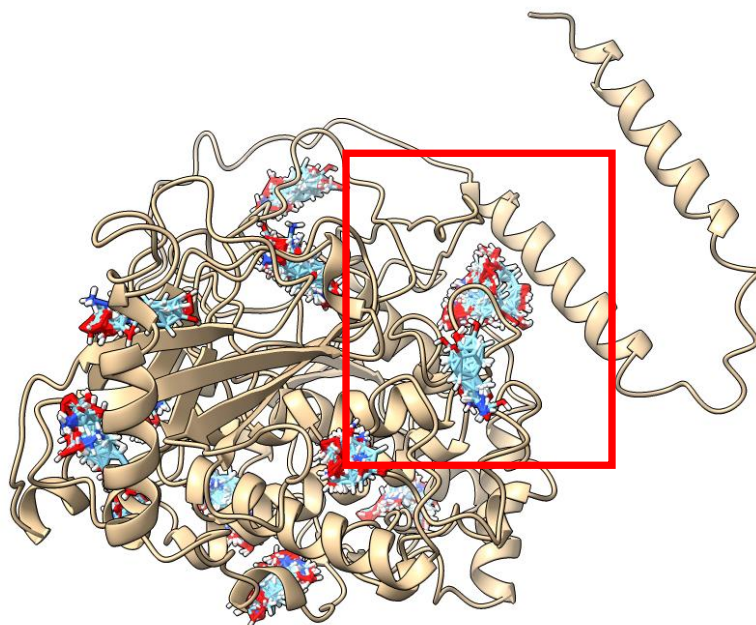


Рисунок 4.10 - Кластери молекулярного докінгу тирозину. Потенційний активний центр позначено прямокутником

Для левадопи ідентифіковано теж 62 кластерів докінгу. Показник FullFitness варіював від -2554,26 до -2578,54 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,39 до -7,28 ккал/моль. У потенційного активного центру, що формувався амінокислотними залишками HIS83-ARG100, GLN21-VAL54 та LEU313-TYR341 (рисунок 4.11) показник FullFitness варіював від -2572,70 до -2573,47, вільна енергія Гібса ΔG варіювала від -6,32 до -6,47.

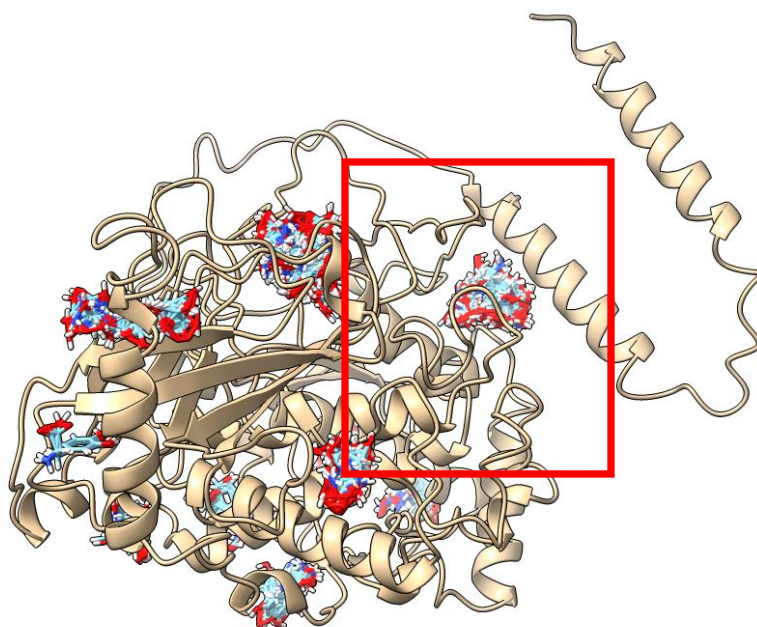


Рисунок 4.11 - Кластери молекулярного докінгу левадопи. Потенційний активний центр позначено прямокутником

4.4 Біоінформатичний аналіз гена TyDC

За результатами гомологічного моделювання побудовано модель ензиму TyDC, що проявляє тирозин декарбоксилазну активність (рис. 4.12). У якості моделі для гомологічного моделювання слугував протеїн D2SMM8.1.A tyrosine decarboxylase. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,92, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,83 \pm 0,05$, що відповідало критеріям прийнятності.

Також для TyDC побудовано гомодимерну модель ензиму, яка структурно є поєднанням двох субодиниць ензиму TyDC (рисунок 4.13). У якості моделі для гомологічного моделювання слугував протеїн bliu.1.A Tyrosine/DOPA decarboxylase 2. Показники якості моделі: стандартний показник точності

реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,87, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,83 \pm 0,05$, що також відповідало критеріям прийнятності.



Рисунок 4.12 - Модель ензиму TyDC, що проявляє тирозин декарбоксилазну активність

У зв'язку з тим, що з одного боку гомодимерна модель була побудована із двох субодиниць, що не мали суттєвих відмінностей від мономерної моделі та, відповідно мають бути ідентичними функціонально, а з другого боку у зв'язку із обмеженням сервером молекулярного докінгу на 1500 амінокислотних залишків, молекулярний докінг із метою ідентифікації активного центру проведено саме із мономерною моделлю. Також треба зазначити, що гомодимерна модель мала менші показники якості і можливо не є біологічно активною моделлю, але може мати значення у процесах регуляції синтезу вторинних метаболітів.



Рисунок 4.13 - Гомодимерна модель ензиму TyDC, що проявляє тирозин декарбоксилазну активність

Для ідентифікації потенційного активного центру побудованої моделі проведено молекулярний докінг із двома субстратами — тирозином та леводопа.

Для тирозину ідентифіковано 52 кластерів докінгу. Показник FullFitness варіював від -2532,49 до -2547,90 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,23 до -6,73 ккал/моль. Активний центр ензиму формувалася амінокислотними залишками GLY112-MET140, THR223-ARG383 (рис. 4.14). У потенційного активного центру показник FullFitness варіював від -2539,70 до -2547,90, вільна енергія Гібса ΔG варіювала від -5,62 до -6,73.

Для леводопи ідентифіковано 62 кластерів докінгу. Показник FullFitness варіював від -2518,52 до -2542,32 ккал/моль, вільна енергія Гібса ΔG варіювала від -4,70 до -6,82 ккал/моль. У потенційного активного центру, що формувалася амінокислотними залишками GLY112-MET140, THR223-ARG383 (рисунок 4.15)

показник FullFitness варіював від -2530,82 до -2539,43, вільна енергія Гібса ΔG варіювала від -5,64 до -6,26.

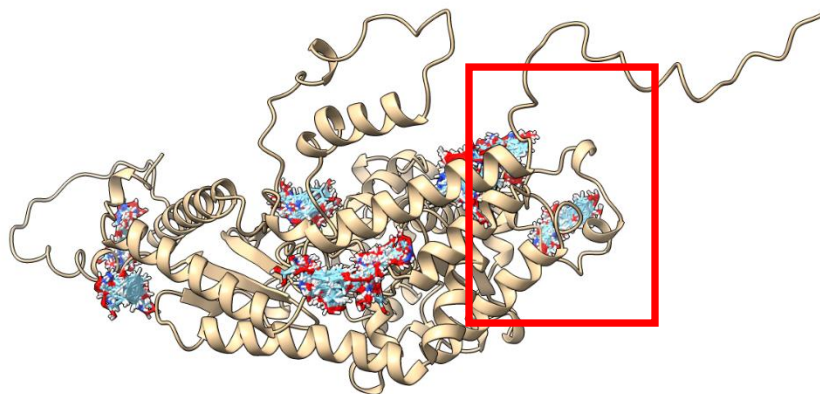


Рисунок 4.14 - Кластери молекулярного докінгу тирозину. Потенційний активний центр позначено прямокутником

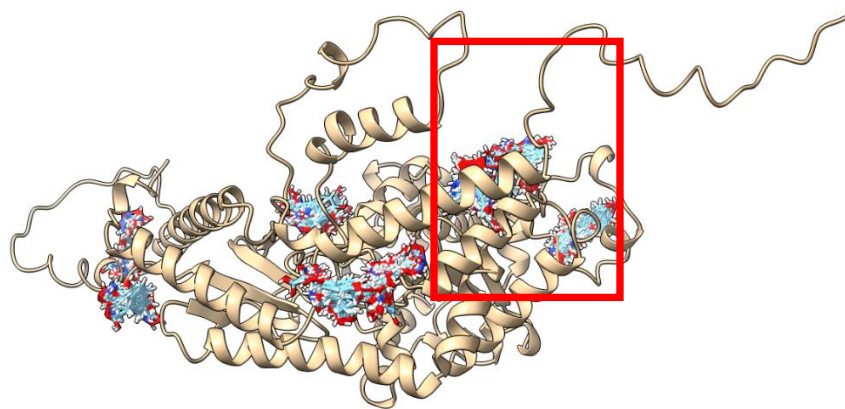


Рисунок 4.15 - Кластери молекулярного докінгу левадопи. Потенційний активний центр позначено прямокутником

4.5 Біоінформатичний аналіз гена NCS1

За результатами гомологічного моделювання побудовано модель ензиму NCS1, що проявляє S-норкоклаурин-синтазну активність (рис. 4.16). У якості моделі для гомологічного моделювання слугував протеїн D2SMN1.1.A S-norcoclaurine synthase 1. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,81, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,81 \pm 0,05$, що відповідало критеріям прийнятності.

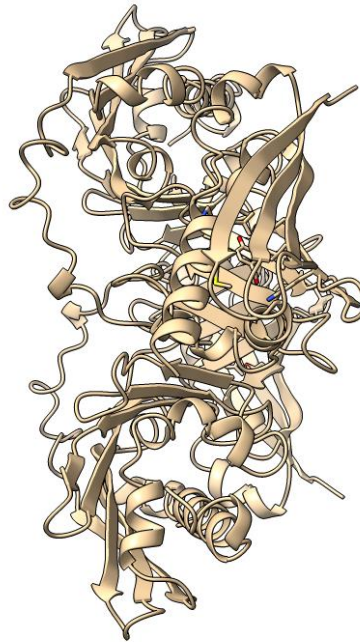


Рисунок 4.16 - Модель ензиму NCS1, що проявляє S-норкоклаурин-синтазну активність

Для ідентифікації активного центру ензиму, проводили молекулярний докінг моделі ензиму із ймовірними субстратами-похідними бензилізохіноліну:

(S)- Силопіном та (S)-Норкоклаурином. (S)-Силопін є субстратом для ензиму, а (S)-Норкоклаурин є продуктом.

Для (S)-Силопіну ідентифіковано 33 кластерів докінгу. Показник FullFitness варіював від -3639,58 до -3631,24 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,90 до -7,30 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками ALA42-VAL61, ILE166-SER184, що ймовірно й формують потенційний активний центр (рис. 4.17). У потенційного активного центру показник FullFitness варіював від -3638,56 до -3661,24, вільна енергія Гібса ΔG варіювала від -7,04 до -7,30.

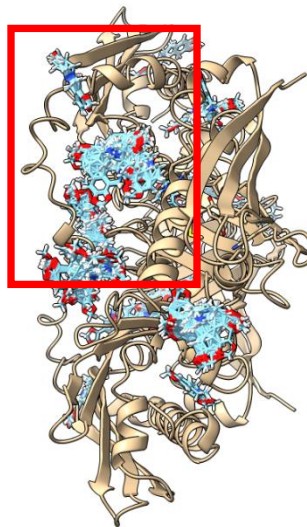


Рисунок 4.16 - Кластери молекулярного докінгу S-Силопіну. Потенційний активний центр позначено прямокутником

Для (S)-Норкоклаурину ідентифіковано 38 кластерів докінгу. Показник FullFitness варіював від -3660,81 до -3673,23 ккал/моль, вільна енергія Гібса ΔG варіювала від -6,01 до -6,70 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками ALA42-VAL61, ILE166-SER184 (рис. 4.18). У потенційного активного центру показник

FullFitness варіював від -3671,11 до -3673,23, вільна енергія Гібса ΔG варіювала від -6,75 до -6,70.

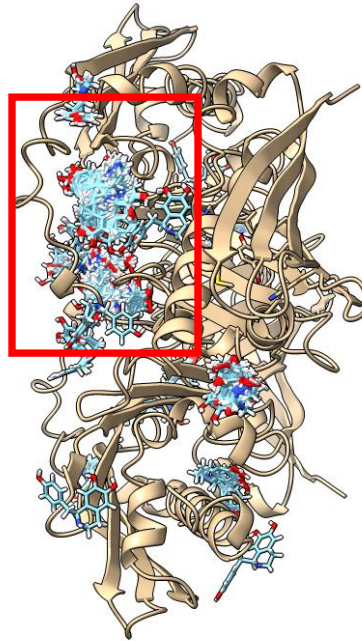


Рисунок 4.18 - Кластери молекулярного докінгу S-Норкоклаурину. Потенційний активний центр позначено прямокутником

4.6 Біоінформатичний аналіз гена NCS2

За результатами гомологічного моделювання побудовано модель ензиму NCS2, що також проявляє S-норкоклаурин-синтазну активність та є гомологом ензиму NCS1 і є подібною до нього на 78,44% (рис. 4.19). У якості моделі для гомологічного моделювання слугував також протеїн D2SMN1.1.A S-noroclaurine synthase 1. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,87, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,66 \pm 0,05$, що відповідало критеріям прийнятності.

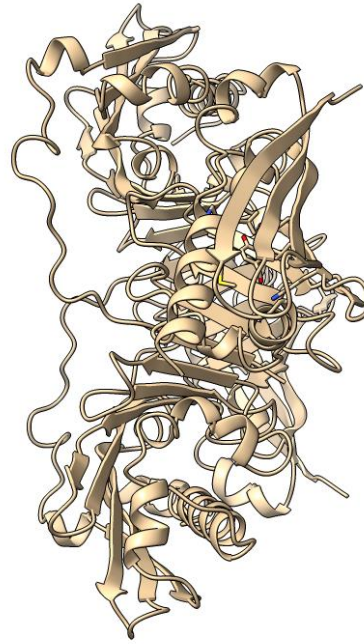


Рисунок 4.19 - Модель ензиму NCS2, що проявляє S-норкоклаурин-синтазну активність

Для ідентифікації активного центру ензиму, проводили молекулярний докінг моделі ензиму із ймовірними субстратами-похідними бензилізохиноліну: (S)- Стиллопіном та (S)-Норкоклаурином. (S)-Стиллопін є субстратом для ензиму, а (S)-Норкоклаурин є продуктом.

Для (S)-Стиллопіну ідентифіковано 34 кластерів докінгу. Показник FullFitness варіював від -3860,64 до -3876,93 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,89 до -7,24 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками ALA42-VAL61, VAL166-PHE184, що ймовірно й формують потенційний активний центр і майже ідентичний ензиму NSC1 (рис. 4.20). У потенційного активного центру показник FullFitness варіював від -3874,33 до -3876,93, вільна енергія Гібса ΔG варіювала від -7,32 до -7,24.

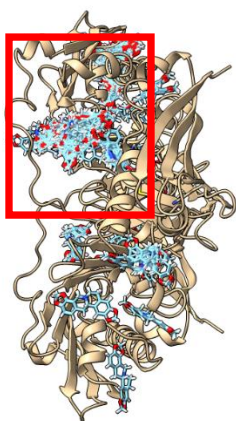


Рисунок 4.20 - Кластери молекулярного докінгу S-Стилопіну. Потенційний активний центр позначено прямокутником

Для (S)-Норкоклаурину ідентифіковано 35 кластерів докінгу. Показник FullFitness варіював від -3877,66 до -3891,31 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,87 до -7,14 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками ALA42-VAL61, VAL166-PHE184 (рис. 4.21). У потенційного активного центру показник FullFitness варіював від -3885,96 до -3891,31, вільна енергія Гібса ΔG варіювала від -6,19 до -7,14.

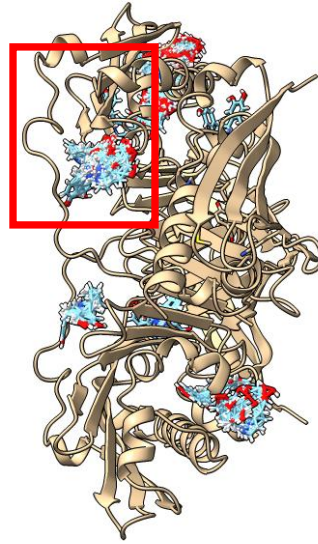


Рисунок 4.21 - Кластери молекулярного докінгу S-Норкоклаурину. Потенційний активний центр позначено прямокутником

NSC1 та NSC2 є паралогами і виникли в результаті дуплікації предкового гена, і зараз вони майже ідентичні і виконують ідентичну функцію. Найбільш ймовірним є те, що вони збільшують стабільність за рахунок того, що у них дещо різні термодинамічні характеристики (що показали результати гомологічного моделювання та докінгу) та оптимальними для них є дещо різні параметри — температура, концентрація субстрату та продукту, що збільшує спектр умов за яких рослини роду Аргемона можуть функціювати. Це яскравий приклад надлишковості (англ. Redundancy) системи.

4.7 Біоінформатичний аналіз гена VBE1

За результатами гомологічного моделювання побудовано модель ензиму VBE1, що формує бербериновий мостик (рис. 4.22). У якості моделі для гомологічного моделювання слугував протеїн D2SMM9.1.A Berberine bridge enzyme 1. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,91, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,88 \pm 0,05$, що відповідало критеріям прийнятності.

Для ідентифікації активного центру ензиму, проводили молекулярний докінг моделі ензиму із ймовірними субстратами-похідними бензилізохиноліну: берберином та бензофенантридином. Берберин субстратом для ензиму, а (S)-бензофенантридин є продуктом.

Для берберину ідентифіковано 34 кластерів докінгу. Показник FullFitness варіював від -2308,63 до -2325,65 ккал/моль, вільна енергія Гібса ΔG варіювала від -6,45 до -8,59 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками, ILE22-ILE132, що ймовірно й формують потенційний активний центр (рис. 4.23). У потенційного активного центру показник FullFitness варіював від -2312,33 до -2318,36 вільна енергія Гібса ΔG варіювала від -7,10 до -7,27.

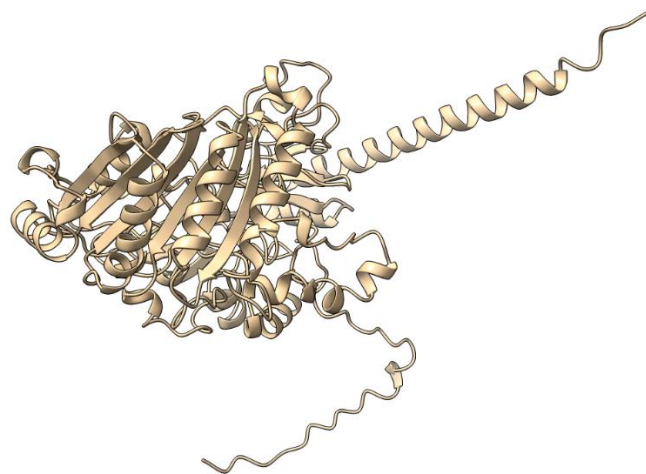


Рисунок 4.22 - Модель ензиму ВВЕ1, що формує бербериновий мостик

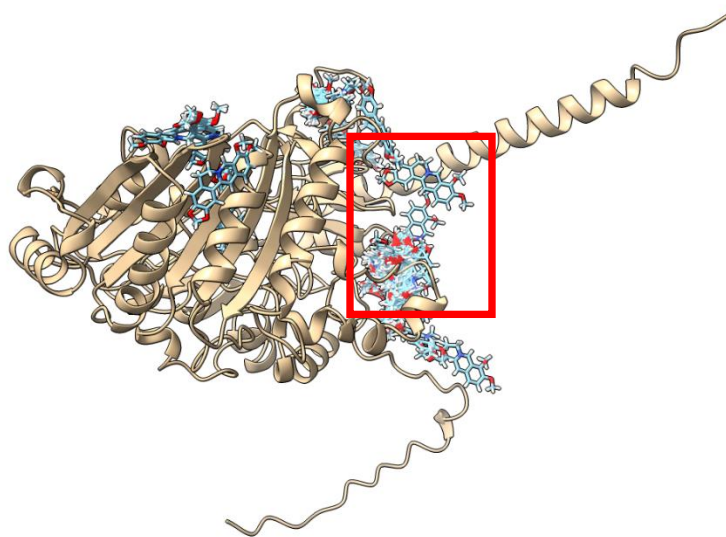


Рисунок 4.23 - Кластери молекулярного докінгу берберину. Потенційний активний центр позначено прямокутником

Для бензофенантридину ідентифіковано 38 кластерів докінгу. Показник FullFitness варіював від -2171,34 до -2297,59 ккал/моль, вільна енергія Гібса ΔG варіювала від -4,53 до -7,50 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками ILE22-ILE132 (рис.

4.24). У потенційного активного центру показник FullFitness варіював від -2286,74 до -2290,93, вільна енергія Гібса ΔG варіювала від -6,30 до -6,78.

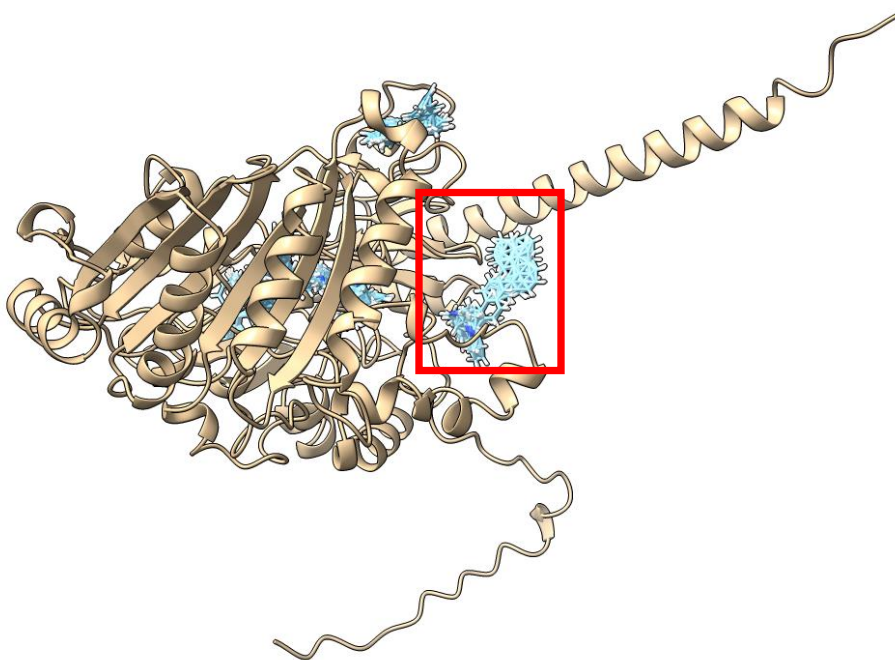


Рисунок 4.24 - Кластери молекулярного докінгу бензофенантридину. Потенційний активний центр позначено прямокутником

4.8 Біоінформатичний аналіз гена VBE2

За результатами гомологічного моделювання побудовано модель ензиму VBE2, що також формує бербериновий мостик (рис. 4.25). У якості моделі для гомологічного моделювання слугував протеїн P30986.1.A Reticuline oxidase. Показники якості моделі: стандартний показник точності реконструкції (Global Model Quality Estimate, GMQE) дорівнював 0,94, аналіз якості енергії моделі (Qualitative Model Energy Analysis, QMEANDisCo Global) дорівнював $0,88 \pm 0,05$, що відповідало критеріям прийнятності.

Для ідентифікації активного центру ензиму, проводили молекулярний докінг моделі ензиму із ймовірними субстратами-похідними бензилізохиноліну:

берберином та бензофенантридином. Берберин субстратом для ензиму, а (S)-бензофенантридин є продуктом.

Для берберину ідентифіковано 31 кластерів докінгу. Показник FullFitness варіював від -2486,61 до -2506,12 ккал/моль, вільна енергія Гібса ΔG варіювала від -6,72 до -8,16 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками SER17-LEU129, що ймовірно й формують потенційний активний центр (рис. 4.26). У потенційного активного центру показник FullFitness варіював від -2501,26 до -2506,12 вільна енергія Гібса ΔG варіювала від -6,76 до -8,16.

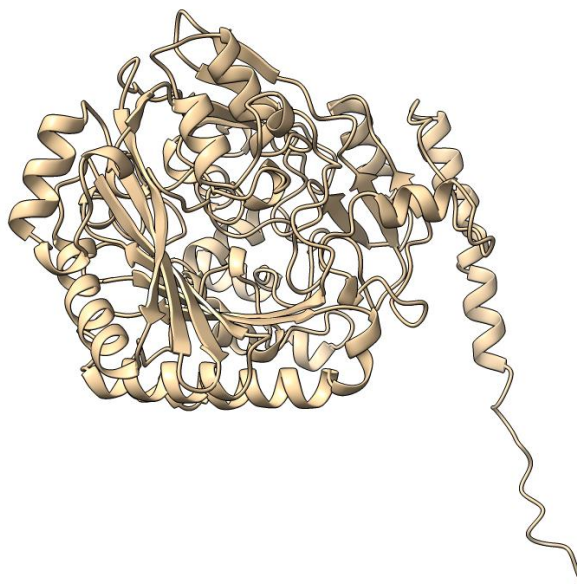


Рисунок 4.25 - Модель ензиму VBE1, що формує бербериновий мостик

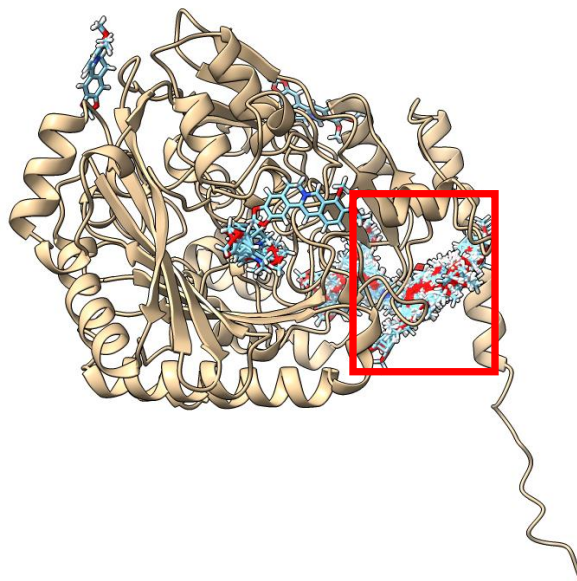


Рисунок 4.26 - Кластери молекулярного докінгу берберину. Потенційний активний центр позначено прямокутником

Для бензофенантридину ідентифіковано 31 кластерів докінгу. Показник FullFitness варіював від -2471,30 до -2482,11 ккал/моль, вільна енергія Гібса ΔG варіювала від -5,89 до -7,46 ккал/моль. Більшість кластерів докінгу групувалася у порожнині, яка формувалася амінокислотними залишками SER17-LEU129 (рис. 4.27). У потенційного активного центру показник FullFitness варіював від -2474,41 до -2477,55, вільна енергія Гібса ΔG варіювала від -6,17 до -6,55.

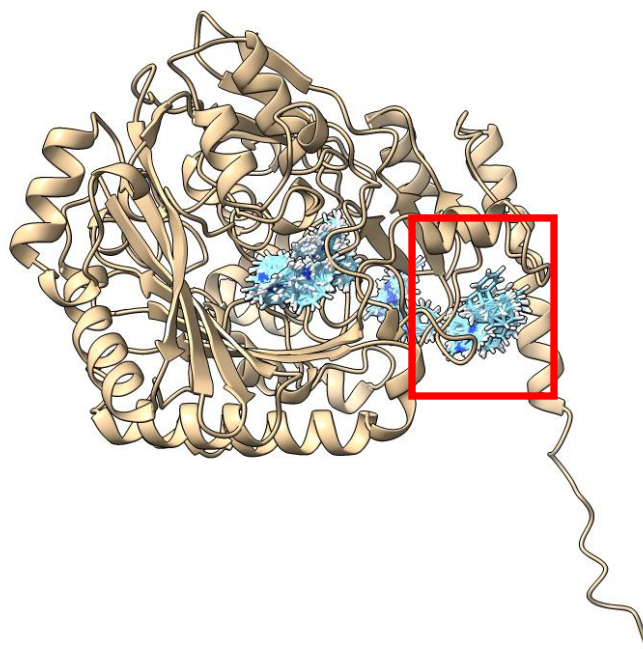


Рисунок 4.27 - Кластери молекулярного докінгу бензофенантрину. Потенційний активний центр позначено прямокутником

Незважаючи на те, що як і у випадку із білками NSC1 та NSC2 білки VBE1 та VBE2 є гомологами, конформація протеїну суттєво відрізнялася у VBE2 у порівнянні із VBE1, що проявлялося і у термодинамічних характеристиках моделей, так і у базовій моделі, на основі якої відбувалося гомологічне моделювання.

4.9 Інженірінг алкалоїдів представників роду Аргемона

Враховуючи велике значення представників роду Аргемона як джерела алкалоїдів для потреб фармакології, модифікація шляхів синтезу вторинних метаболітів є перспективною та важливою задачею. В даному дослідженні в результаті проведеного біоінформатичного аналізу було побудовано тривимірні моделі ензимів CYP719A14, CYP719A13, TyPP0, TyDC, NCS1, NCS2, VBE1 та

ВВЕ2, проведений молекулярний докінг та ідентифіковано потенційні активні центри цих ензимів. Подальша робота по удосконаленню алкалоїдного профілю може бути спрямована у таких напрямках. Перше і найпростіше — шляхом вимикання певних ензимів через сайленсінг відповідних генів можна змінити шляхи біосинтезу молекул, і таким чином привести до накопичення у тканинах проміжних продуктів біосинтезу, які далі можуть бути екстраговані із тканин й або безпосередньо використовуватися у фармакологічних цілях, або після подальшої, вже хімічної модифікації.

Другий шлях — це трансгенез генів, що асоційовані із синтезом алкалоїдів із інших рослин для того, щоб корінним чином змінити алкалоїдний профіль.

Третій — це модифікація існуючих нуклеотидних послідовностей генів, що відповідальні за вторинні метаболіти Аргемони, наприклад за допомогою системи CRISPR/Cas і для цього безцінними є побудовані нами моделі і ідентифіковані активні центри. Модифікація активних центрів кардинальним шляхом може змінити афінитет до субстрату, можливо навіть із зміною продуктів реакції на новітні, і це також можна спочатку дослідити *in silico* за допомогою гомологічного моделювання та молекулярного докінгу. Альтернативні кластери молекулярного докінгу можуть вказувати на сайти протеїну, що відповідають за регуляцію ензиматичної активності протеїну, наприклад при накопиченні продуктів реакції викликаючи пригнічення активності через механізм зворотнього зв'язку. Модифікація цих ділянок може підвищити активність ензимів і привести до накопичення більшої кількості бажаного продукту.

4.10 Аналіз та узагальнення результатів

Біоінформатичні дослідження є критично важливими в сучасному науковому просторі, особливо в контексті вивчення біологічних і хімічних процесів у рослин роду Аргемона. Ці дослідження включають в себе аналіз та інтерпретацію біологічних даних, які можуть включати генетичні послідовності,

структури білків, метаболічні шляхи тощо. Вони дозволяють визначити, яким чином різні біомолекули взаємодіють між собою, а також розкрити механізми утворення та функціонування алкалоїдів, які виробляються цими рослинами.

Важливість біоінформатичних досліджень полягає в можливості створення тривимірних моделей ензимів, що беруть участь у синтезі алкалоїдів. Наприклад, ензими CYP719A14, CYP719A13, TyPPPO, TyDC, NCS1, NCS2, VBE1, та VBE2 є ключовими у біосинтезі цих біологічно активних сполук. Тривимірне моделювання допомагає розуміти структуру та динаміку ензимів, їх активні центри та потенційні місця для зв'язування з субстратами або інгібіторами, що відкриває шлях для розробки нових лікарських препаратів.

Застосування молекулярного докінгу в рамках біоінформатичних досліджень дозволяє не тільки підтвердити структури, побудовані на підставі тривимірного моделювання, але й ідентифікувати потенційні активні центри. Це сприяє розумінню того, як можливе впливання на ці ензими з боку різних молекул може змінити їх активність. Цей аналіз дає важливу інформацію для розробки інгібіторів або активаторів ензимів, що може бути використано для керування метаболічними шляхами в рослинах роду *Argemone* і, як результат, для покращення виробництва корисних алкалоїдів.

Побудова тривимірних моделей ензимів, таких як CYP719A14, CYP719A13, TyPPPO, TyDC, NCS1, NCS2, VBE1 та VBE2, які беруть участь у біосинтезі алкалоїдів рослин роду *Argemone*, відіграє ключову роль в алкалоїдному інженірінгу. Ці моделі дають детальне розуміння структурної конформації та активних центрів ензимів, що є важливим для планування біотехнологічних експериментів. Використання цих моделей дозволяє дослідникам визначати, як модифікації у структурі ензимів можуть вплинути на їхню активність і специфічність у відношенні до субстратів і продуктів реакцій.

Молекулярний докінг ензимів з їх субстратами і продуктами реакцій є основою для розуміння механізмів каталізу та для виявлення молекулярних

взаємодій, що відбуваються під час біохімічних реакцій. Ця інформація є вкрай важливою для інженерії алкалоїдів, оскільки вона дозволяє оптимізувати існуючі шляхи біосинтезу та створювати нові, що можуть призвести до утворення алкалоїдів з новими або покращеними властивостями. Дослідження цих взаємодій на молекулярному рівні може допомогти в ідентифікації потенційних нових лікарських мішеней.

Використання побудованих моделей та даних молекулярного докінгу в біоінформатиці створює можливості для розробки комп'ютерно-орієнтованих методів відбору та оптимізації молекул, які можуть діяти як ефективні модулятори активності цих ензимів. Завдяки таким методам можна прогнозувати вплив потенційних інгібіторів або активаторів на метаболізм алкалоїдів без необхідності виробництва та тестування великої кількості експериментальних зразків, що значно прискорює процес розробки нових фармацевтичних препаратів.

5 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ МОДЕЛЮВАННЯ МОЛЕКУЛЯРНОЇ ДИНАМІКИ

Симуляції молекулярної динаміки є надзвичайно важливими в сучасній науці та інженерії, зокрема у таких галузях як матеріалознавство, біохімія, фармацевтика та нанотехнології. Ці симуляції дозволяють дослідникам віртуально моделювати та вивчати поведінку атомів та молекул у реальному часі. Це важливо для розуміння складних біологічних процесів, таких як складання білків, а також для розробки нових матеріалів з унікальними властивостями. Використання симуляцій молекулярної динаміки може значно прискорити процес розробки нових ліків, дозволяючи точно передбачати взаємодію між лікарськими речовинами та біологічними молекулами.

Незважаючи на величезний потенціал, симуляції молекулярної динаміки часто зустрічаються з викликами у вигляді високих вимог до обчислювальних ресурсів та складності програмного забезпечення. Багато існуючих інструментів вимагають глибоких знань у галузі комп'ютерного моделювання та теоретичної хімії, що обмежує їх доступність для широкого кола дослідників. Розвиток програмного забезпечення з графічним інтерфейсом для симуляцій молекулярної динаміки має потенціал значно спростити цей процес, роблячи ці цінні інструменти доступними для більш широкої аудиторії дослідників, які можуть не мати спеціалізованих знань у цій області.

Створення інтуїтивно зрозумілого графічного інтерфейсу для програмного забезпечення симуляцій молекулярної динаміки відкриває двері для інновацій та колаборації між різними науковими дисциплінами. Це не тільки полегшить роботу дослідників, які вже використовують ці технології, але й дозволить новим користувачам легше інтегруватися в цю область. Програмне забезпечення з графічним інтерфейсом може включати такі функції, як візуалізація траєкторій молекул, інтерактивне редагування структур та налаштування параметрів.

На сучасному ринку існує ряд продуктів для симуляцій молекулярної динаміки, проте вони часто демонструють певну полярність у доступності та зручності використання. З одного боку, існують високоякісні комерційні програми, які широко використовуються у фармацевтичній промисловості та біохімічних дослідженнях. Ці інструменти, хоча і надзвичайно потужні, часто виявляються досить дорогими через складність розробки та високу цінність, яку вони пропонують у комерційних дослідженнях. Така вартість може бути непосильною для деяких академічних лабораторій або індивідуальних дослідників, що обмежує їхню можливість використання цих передових інструментів.

З іншого боку, існують безкоштовні альтернативи, які, хоча і забезпечують базові можливості для симуляцій молекулярної динаміки, часто поступаються в зручності використання та інтуїтивності інтерфейсу. Багато з цих безкоштовних інструментів вимагають від користувачів глибоких знань у програмуванні та теоретичної хімії, а також володіння спеціалізованими командами та скриптами. Відсутність дружнього графічного інтерфейсу може ускладнити процес навчання та адаптації для новачків, які не мають технічного фону, знижуючи загальну доступність цих важливих наукових інструментів.

Тому нами був розроблений програмний продукт, якій з одного боку поєднує точність моделювання, а з іншого – зручність використання (див. додаток А). Для розробки програмного забезпечення симуляцій молекулярної динаміки було обрано мову програмування Python, і це рішення має вагомі переваги. Python відомий своєю високою читабельністю та простотою синтаксису, що робить його ідеальним вибором для наукових досліджень, де кларовність та легкість зрозумілості коду є критично важливими. Ця мова програмування також підтримує широкий спектр наукових бібліотек та інструментів, таких як NumPy, SciPy та Matplotlib, які вже широко використовуються у спільноті наукових досліджень. Ці бібліотеки дозволяють

легко виконувати складні математичні обчислення та візуалізації, що є невід'ємною частиною симуляцій молекулярної динаміки.

Ще однією перевагою Python є його висока масштабованість та інтеграція з іншими мовами програмування та технологіями. Python дозволяє легко інтегруватися з більш швидкими мовами, такими як C та C++, що може бути корисним для підвищення продуктивності в ресурсоємних розрахунках. Крім того, Python є кросплатформним, що означає, що розроблене на його основі програмне забезпечення може працювати на різних операційних системах, таких як Windows, macOS та Linux. Це робить Python чудовим вибором для створення доступних та універсальних наукових інструментів, які можуть використовуватися широким колом дослідників з різними технічними потребами та умовами.

MD Simulation програма є комплексним інструментом для молекулярно-динамічної симуляції та аналізу білків. Вона використовує MDAnalysis - потужну бібліотеку для аналізу траєкторій молекулярної динаміки, а також OpenMM для налаштування та виконання симуляцій. Серед основних особливостей програми - можливість обчислення Root-Mean-Square Deviation (RMSD), що дозволяє оцінити структурні зміни в молекулах під час симуляцій. Важливим аспектом є також використання потоків (через QThread) для ефективного управління довготривалими обчислювальними процесами.

Програма інтегрує функціональність PyQt5, що дозволяє створити інтуїтивно зрозумілий та зручний графічний інтерфейс користувача. Це включає створення вікон, кнопок, вибірних списків та текстових полів для введення параметрів симуляції, таких як температура та крок часу. Інтерфейс також надає можливість вибору певних атомів або груп атомів для детального аналізу (рисунк 5.1). Важливим елементом є інтеграція ruqtgraph та OpenGL для візуалізації тривимірних структур та траєкторій атомів, що дає користувачам можливість безпосередньо спостерігати за результатами своїх симуляцій.

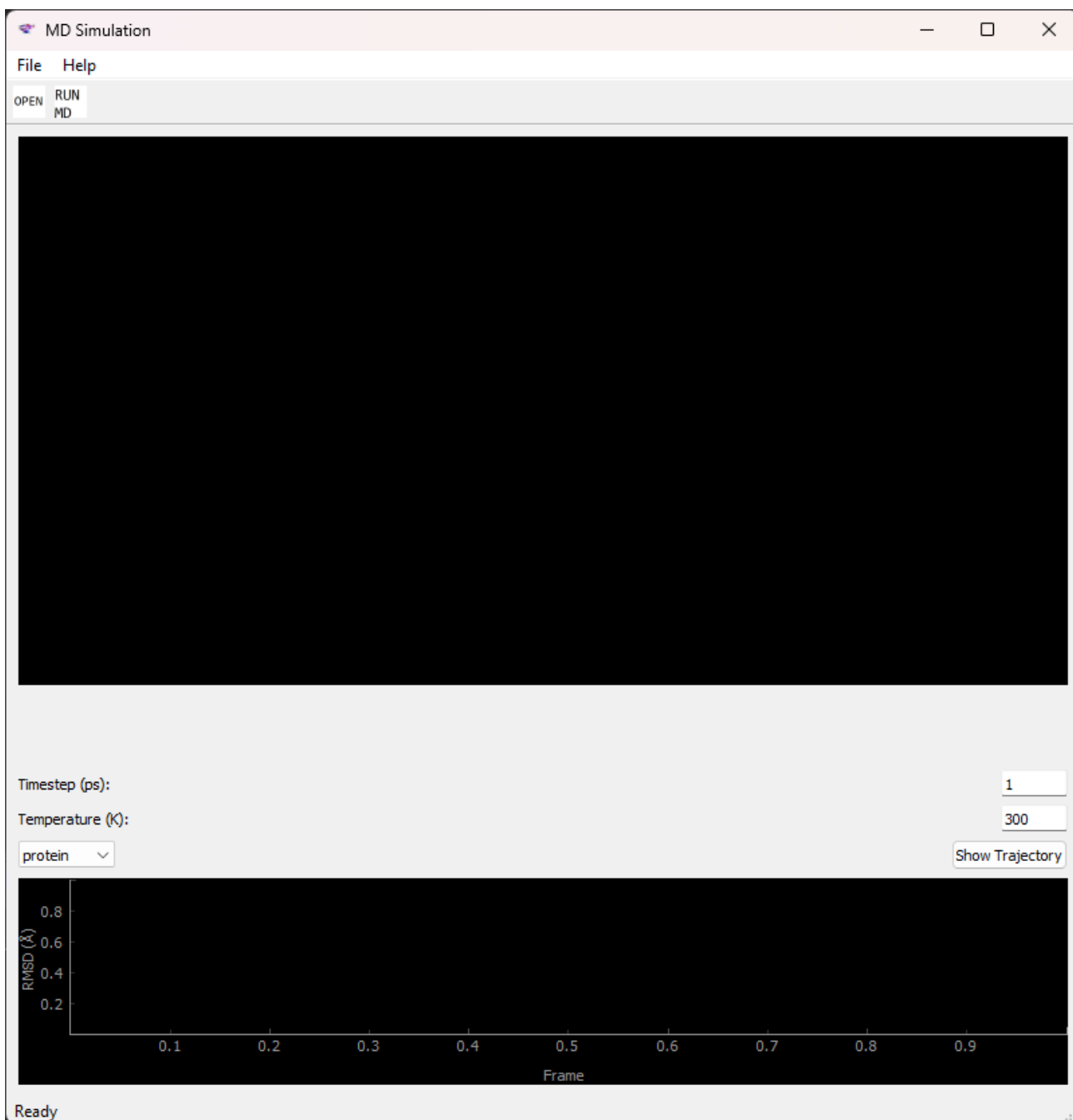


Рисунок 5.1 – інтерфейс програми MD simulation

MD Simulation має наступні блоки:

1. Блок Молекулярно-Динамічної Симуляції:
 - Використання OpenMM для налаштування та виконання симуляцій.
 - Обробка даних симуляції, таких як траєкторії руху атомів та молекул.
2. Аналітичний Блок:

- Використання MDAnalysis для аналізу траєкторій молекулярної динаміки.
- Функції для обрахунку RMSD та інших статистичних параметрів.
- 3. Потокова Обробка Даних (Клас MDRun):
- Використання потоків (QThread) для ефективної обробки даних симуляцій.
- Сигнали PyQt для асинхронного оновлення даних та візуалізації.
- 4. Графічний Інтерфейс Користувача (GUI):
- Створення вікон, кнопок, вибірних списків за допомогою PyQt5.
- Інтерактивність та зручність управління параметрами симуляції.
- 5. Візуалізація Даних:
- Інтеграція pyqtgraph та OpenGL для 3D-візуалізації.
- Відображення графіків RMSD та інших аналітичних показників.
- 6. Конфігурація та Налаштування:
- Введення та зміна параметрів симуляції через графічний інтерфейс.
- Можливість адаптації симуляцій під конкретні наукові потреби.

Програма здатна відкривати та візуалізувати файли моделей протеїнів у .pdb форматі (рисунок 5.2). З метою збереження обчислювальних ресурсів візуалізація спрощена, кожний атом є напівпрозорою сферою із наступною кольоровою кодировкою:

Атоми Вуглецю (C): Колір зелений, представлений як (0, 1, 0, 1) у форматі RGBA.

Атоми Азоту (N): Колір синій, представлений як (0, 0, 1, 1) у форматі RGBA.

Атоми Кисню (O): Колір червоний, представлений як (1, 0, 0, 1) у форматі RGBA.

Атоми Сірки (S): Колір жовтий, представлений як (1, 1, 0, 1) у форматі RGBA.

Атоми Водню (H): Колір світло-сірий, представлений як (0.8, 0.8, 0.8, 0.5) у форматі RGBA.

Інші Атоми: Для інших атомів використовується стандартний сірий колір, представлений як (0.5, 0.5, 0.5, 0.5) у форматі RGBA.

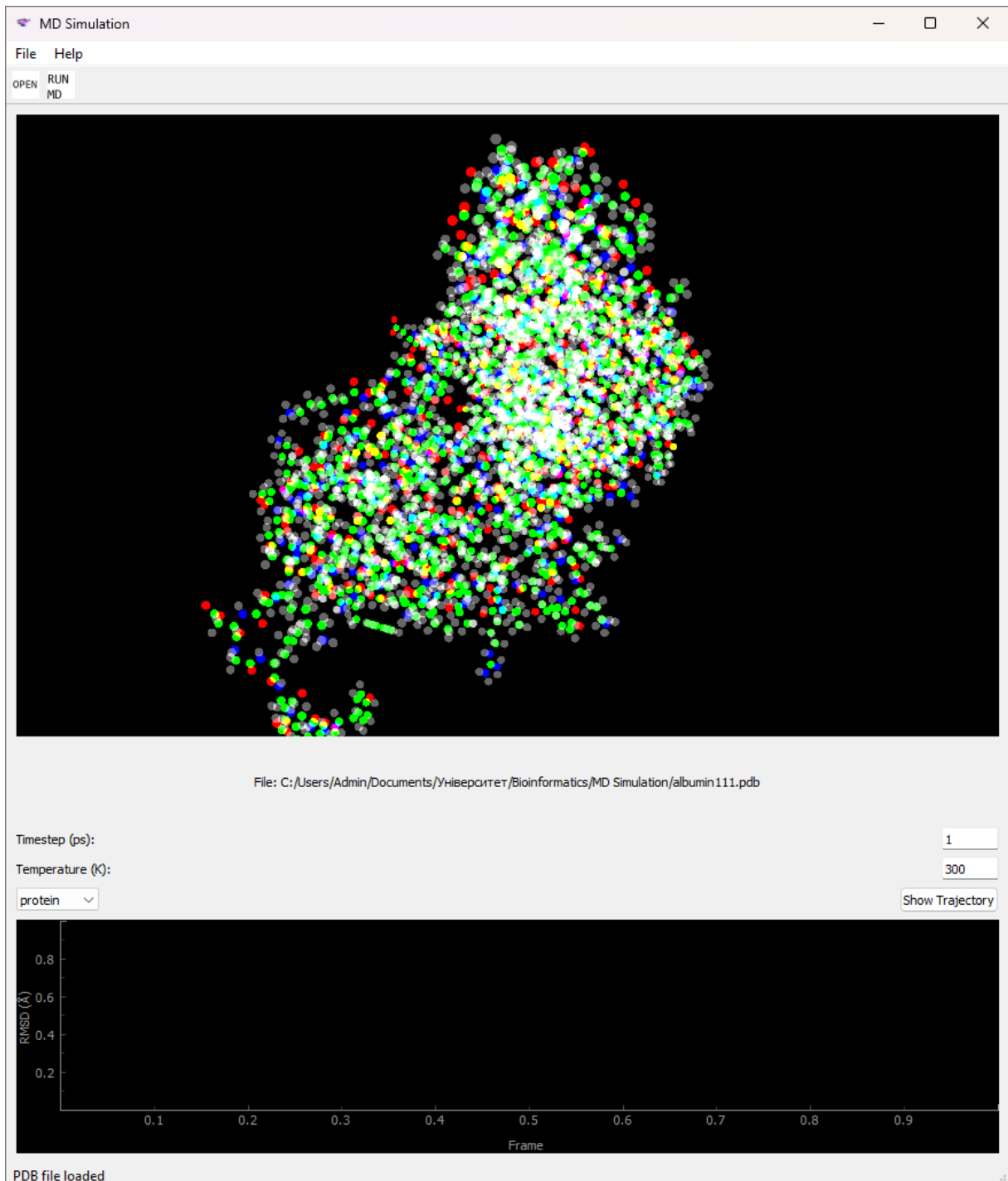


Рисунок 5.2 – Візуалізація протеїну (людський альбумін)

Далі можливо провести симуляцію молекулярної динаміки. Симуляція відбувається за рахунок використання наступних бібліотек:

MDAnalysis: Це потужна бібліотека для аналізу даних молекулярно-динамічних симуляцій. MDAnalysis дозволяє зчитувати, писати та обробляти траєкторії та структури, отримані з широкого спектру популярних форматів файлів молекулярної динаміки. Бібліотека надає зручні інструменти для детального аналізу молекулярних структур, включаючи обчислення таких параметрів, як RMSD, радіуси гірації, вторинні структури білків тощо.

OpenMM: Ця бібліотека використовується для виконання симуляцій молекулярної динаміки з високою продуктивністю. OpenMM забезпечує підтримку для швидких симуляцій на GPU, що дозволяє ефективно обробляти складні молекулярні системи. Вона включає інструменти для створення та оптимізації систем білків, нуклеїнових кислот та інших біомолекул, а також для управління симуляціями з різними умовами, такими як температура, тиск, сили міжмолекулярної взаємодії.

Для проведення симуляції використовуються наступні силові поля: 'amber14-all.xml' та 'amber14/tip3pfb.xml'. Важливим етапом симуляції є створення «коробки», у яку потрапляє модель протеїну, додається достатня кількість молекул води і проводиться симуляція. Код містить рядок `box_vectors = [Vec3(3.0, 0, 0), Vec3(0, 3.0, 0), Vec3(0, 0, 3.0)] * unit.nanometer`, який вказує на створення векторів для визначення геометрії коробки. Ці вектори задають розміри коробки у трьох вимірах (x, y, z), кожен з яких має довжину 3 нанометри.

Далі, в коді використовується метод `modeller.topology.setPeriodicBoxVectors(box_vectors)`, що встановлює періодичні граничні умови для коробки з визначеними векторами. Це означає, що коробка матиме форму паралелепіпеда з розмірами 3x3x3 нанометри і буде використовувати періодичність для моделювання нескінченного простору. Такий розмір коробки достатній для коректного моделювання.

Після проведення симуляції ми отримуємо графік RMSD, який вказує на відхилення атомів під час симуляції, і візуалізувати траєкторії руху атомів під час симуляції для того щоб оцінити, чи зберігається цілісність протеїну під час симуляції (рисунок 5.3). Також формується файл траєкторії trajectory.dcd, який може бути проаналізований.

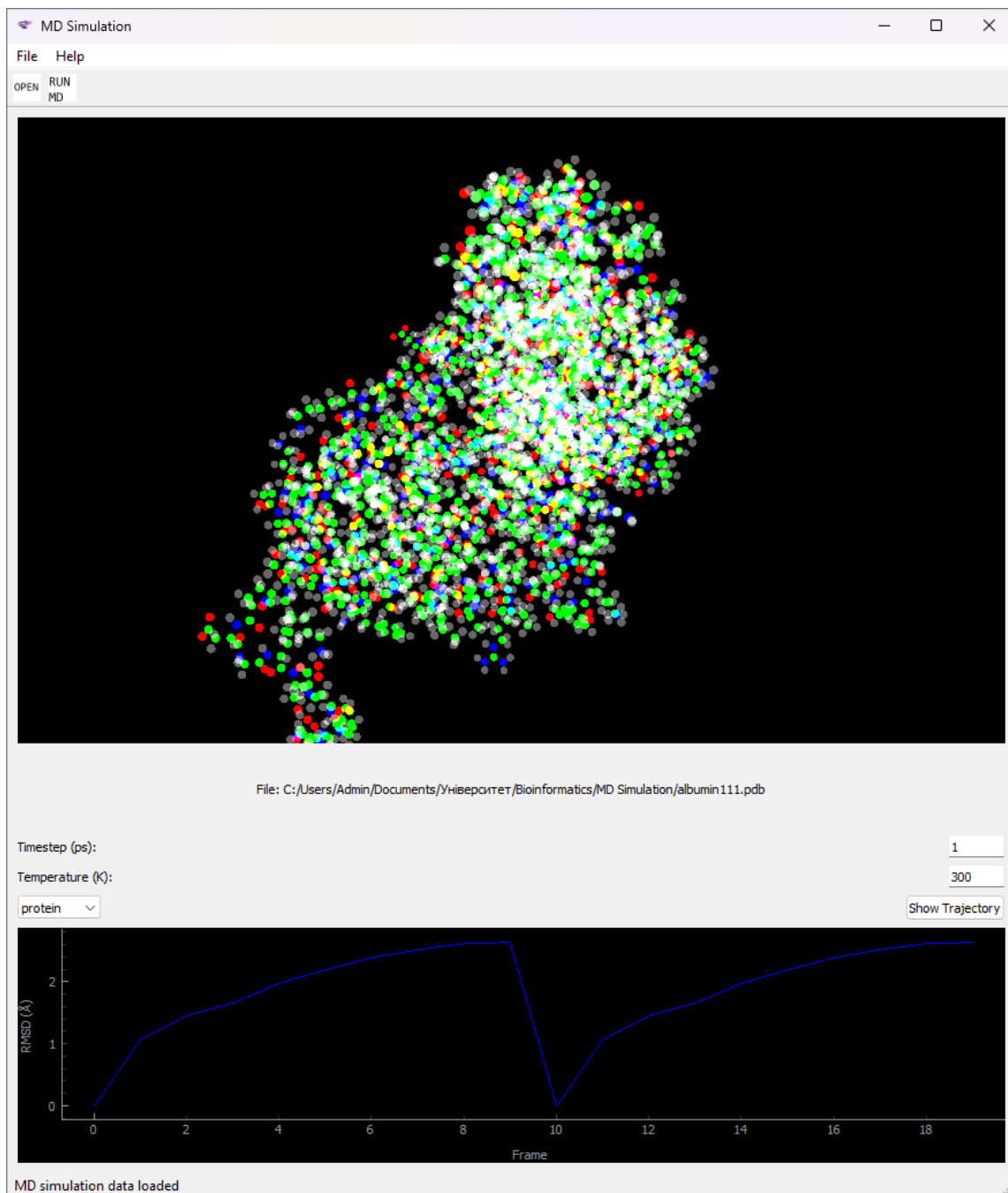


Рисунок 5.3 – Результат проведення симуляції молекулярної динаміки

Одною із особливостей програми є те, що вона потребує того, щоб модель протеїну у форматі .pdb містила атоми водню, у той час як для того, щоб зменшити розмір часто моделі не містять атомів водню. Моделі, що генеруються у результаті гомологічного моделювання також не містять атоми водню. Тому нами також розроблений простий скрипт для додання атомів водню (див. додаток Б). Цей скрипт містить наступні елементи:

Іморти: Сценарій імпортує модулі sys та tkinter (зокрема filedialog з tkinter). Також імпортується app з simtk.openmm, що є бібліотекою для молекулярної динаміки.

Функція для Роботи з Білками, що проводить обробку білкових структур, додавання водневих атомів до структур білків у форматі PDB (Protein Data Bank).

Графічний Інтерфейс: використан tkinter (рисунок 5.4).

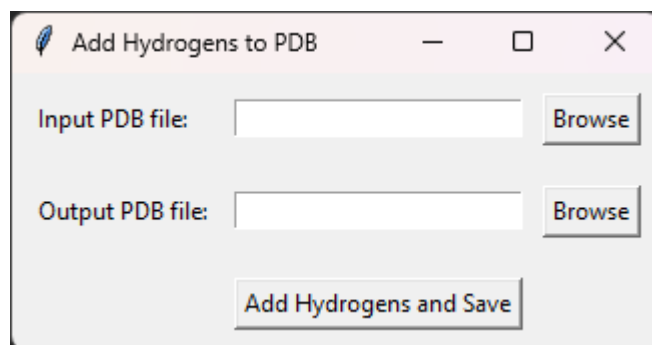


Рисунок 5.4 – інтерфейс скрипту для додавання атомів водню

Програма симуляцій спочатку валідована на наборі простих білків (кристалोगрами та ЯМР моделі із бази даних PDB), після чого використовувалася для проведення біоінформатичних досліджень ензимів Аргемони, що асоційовані із синтезом вторинних метаболітів. На адекватність моделей вказували занчення RMSD, які не перевищували 4 Å, а також той факт, що моделі зберігали цілісність під час симуляцій навіть за високих температур

ВИСНОВКИ ТА ПРАКТИЧНІ РЕКОМЕНДАЦІЇ

1. Розроблене програмне забезпечення для аналізу біоінформатичної інформації, який включав симуляції молекулярної дитнаміки, який дозволив отримати нову інформацію щодо біосинтезу вторинних метаболітів у представників роду Аргемона.

2. За результатами біоінформатичного дослідження нуклеотидних послідовностей ITS1 представників роду Аргемона реконструйовано філогенетичну дендрограму, яка співвідповідала таксономічним взаємовідносинам у роді Аргемона.

3. Ідентифіковано ділянку ITS1, яка була унікальна саме для представників роду Аргемона, що дозволило розробити дизайн праймерів для зонду, які можна використовувати як маркерну систему для ідентифікації ДНК представників роду Аргемона у харчових продуктах.

4. Проведено гомологічне моделювання ензимів CYP719A14, CYP719A13, TuPPO, TuDC, NCS1, NCS2, VBE1 та VBE2, що відповідальні за біосинтез алкалоїдів, проведено валідацію моделей.

5. За допомогою молекулярного докінгу ідентифіковано потенційні активні центри цих ензимів.

6. Розроблені та реалізовані моделі та методи аналізу біологічного об'єкта, засновані на комплексному використанні інформаційних та біологічних технологій, що сприяє поглибленню розуміння складних біоінформатичних процесів.

7. Розроблену систему молекулярних маркерів можна використовувати для ідентифікації контамінації харчових продуктів представниками роду Аргемона.

3. Дані біоінформатичних досліджень ензимів, що асоційовано із біосинтезом алкалоїдів можна використовувати для подальших досліджень інжинірінгу алкалоїдів із метою модифікації алкалоїдного профілю.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ouzounis C. A., Valencia A., Early bioinformatics: the birth of a discipline—a personal view. *Bioinformatics*. 2003. Vol. 19, № 17. P. 2176–2190. <https://doi.org/10.1093/bioinformatics/btg309>.
2. "History of the Human Genome Project". https://web.ornl.gov/sci/techresources/Human_Genome/project/hgp.shtml.
3. Release Notes For GenBank Release 257 <https://www.ncbi.nlm.nih.gov/genbank/release/257/>.
4. Bang-Jensen J., Gutin G., Yeo A. When the greedy algorithm fails, *Discrete Optimization*. 2004. Vol. 1, № 2. P. 121-127. DOI: 10.1016/j.disopt.2004.03.007.
5. Phillip C.E.C, Pevzner P.A., Tesler G. How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*. 2011. Vol. 29 № 11. P. 987–991. doi:10.1038/nbt.2023.
6. Schbath S., Martin V., Zytynicki M., Fayolle J., Loux V., Gibrat J.F.. Mapping reads on a genomic sequence: an algorithmic overview and a practical comparative analysis. *J. Comput. Biol.* 2012. Vol. 19. № 6. P. 796-813. doi: 10.1089/cmb.2012.0022.
7. Lischer, H.E.L., Shimizu, K.K. Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*. 2017. Vol. 18, P. 474 <https://doi.org/10.1186/s12859-017-1911-6>.
8. Mount D.M. *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 2004. NY. ISBN 978-0-87969-608-5.
9. Eddy, S. R. A Probabilistic Model of Local Sequence Alignment That Simplifies Statistical Significance Estimation. *PLOS Computational Biology*. 2008, Vol. 4. №. 5. P. e1000069. <https://doi.org/10.1371/journal.pcbi.1000069>.

10. Huang, M., Shah, N.D. Yao, L. Evaluating global and local sequence alignment methods for comparing patient medical records. BMC Med. Inform. Decis. Mak. 2019. Vol. 19. № 6. P. 263 <https://doi.org/10.1186/s12911-019-0965-y>.

11. Xu Q, Dai H, Zhao T, Wei D. Introduction to structural bioinformatics. Adv Exp Med Biol. 2015, Vol. 827. P. 1-7. DOI:10.1007/978-94-017-9245-5_1.

12. Carlos Outeiral, CASP14: what Google DeepMind's AlphaFold 2 really achieved, and what it means for protein folding, biology and bioinformatics, Oxford Protein Informatics Group. <https://www.blopig.com/blog/2020/12/casp14-what-google-deepminds-alfafold-2-really-achieved-and-what-it-means-for-protein-folding-biology-and-bioinformatics/>.

13. Schwede T., Kopp J., Guex N., Peitsch M.C. "SWISS-MODEL: an automated protein homology-modeling server". Nucleic Acids Res. 2003, Vol. 31 №13. P. 3381–3385. DOI:10.1093/nar/gkg520.

14. Сліщук Г.І., Волкова Н.Е., Сиволап Ю.М. Біоінформатичний аналіз вторинної структури тринскриптів інтрону 1 гену *whp1* кукурудзи/ Biopolymers and cell. 2012. Vol. 28, N 2. P. 156-160.

15. Сліщук Г.І., Кожухова Н.Е. Біоінформатичний аналіз генів мітохондріальних альдегіддегідрогеназ представників родини *Roaceae*. Збірник наукових праць «Фактори експериментальної еволюції». 2011. Т. 11. С. 175-180.

16. Сліщук Г.І., Куклін А.В., Волкова Н.Е. Біоінформатичний та молекулярно-генетичний аналіз гена *aasr1* кукурудзи. Всеукраїнська науково-практична Інтернет-конференція «Підвищення ефективності функціонування сільського господарства в умовах зміни клімату». 09.12.2016 р., Херсон, Україна. С. 134-136.

17. Сліщук Г.І., Волкова Н.Е., Захарова О.О., Корчмарьова А.В. Біоінформатичний аналіз гена, що кодує ацетогидроксиацидсинтазу, нуту звичайного. Збірник наукових праць «Фактори експериментальної еволюції організмів». 2019. Т. 24. С. 345-349.

18. Hagel J.M., Morris J.S., Lee E.J., Desgagné-Penix I., Bross C.D., Chang L., Chen X., Farrow S.C., Zhang Y., Soh J., Sensen C.W., Facchini P.J. Transcriptome analysis of 20 taxonomically related benzyloisoquinoline alkaloid-producing plants. *BMC Plant Biol.* 2015. Vol. 15. P. 227. DOI: 10.1186/s12870-015-0596-0.

19. Davoodi Mastakani F., Pagheh G., Rashidi Monfared S., Shams-Bakhsh M. Identification and expression analysis of a microRNA cluster derived from pre-ribosomal RNA in *Papaver somniferum* L. and *Papaver bracteatum* L. *PLoS One.* 2018. Vol. 13 № 8. P. e0199673. DOI: 10.1371/journal.pone.0199673.

20. Chavda V.P., Ertas Y.N., Walhekar V., Modh D., Doshi A., Shah N., Anand K., Chhabria M. Advanced Computational Methodologies Used in the Discovery of New Natural Anticancer Compounds. *Front Pharmacol.* 2021. Vol. 12. P. 702611. DOI: 10.3389/fphar.2021.702611.

21. Gledhill, D. (2008). *The Names of Plants* (4 ed.). Cambridge University Press. p. 55. ISBN 978-0-521-86645-3.

22. Linnaeus, C. 1753. *Species Plantarum*. Tomus I: 503.

23. Ownbey G.B. Monograph of the genus *Argemone* for North America and the West Indies. *Memoirs of the Torrey Botanical Club.* 1958, Vol. 21 № 1. P. 1-159.

24. Priya C.L., Rao K.V.B. Ethanobotanical and current ethanopharmacological aspects of *Argemone mexicana* Linn: an overview. *International Journal Of Pharmaceutical Sciences And Research.* 2012. Vol. 3 № 7. P. 2143-2148. DOI:10.13040/IJPSR.0975-8232.3.

25. Mirón-López G. Tissue distribution of transcripts involved in biosynthesis of benzyloisoquinoline alkaloid in mature plants of *Argemone mexicana* L. (*Papaveraceae*). *Revista Fitotecnia Mexicana.* 2018. Vol. 41(1), P. 13-21. DOI:10.35196/rfm.2018.1.13-21.

26. Laines-Hidalgo J.I., Muñoz-Sánchez J.A., Loza-Müller L., Vázquez-Flota F. An Update of the Sanguinarine and Benzophenanthridine Alkaloids' Biosynthesis and

Their Applications. *Molecules*. 2022. Vol. 27 № 4. P. 1378. DOI:10.3390/molecules27041378.

27. Díaz Chávez M.L., Rolf M., Gesell A., Kutchan T.M. Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy *Argemone mexicana*. *Arch. Biochem. Biophys.* 2011. Vol. 507 № 1. P. 186-93. DOI: 10.1016/j.abb.2010.11.016.

28. Siatka T., Adamcová M., Opletal L., Cahlíková L., Jun D., Hrabínová M., Kuneš J., Chlebek J. Cholinesterase and Prolyl Oligopeptidase Inhibitory Activities of Alkaloids from *Argemone platyceras* (*Papaveraceae*). *Molecules*. 2017. Vol. 22 № 7. P. 1181. doi: 10.3390/molecules22071181.

29. Leyva-Peralta M.A., Robles-Zepeda R.E., Garibay-Escobar A., Ruiz-Bustos E., Alvarez-Berber L.P., Gálvez-Ruiz J.C. *In vitro* anti-proliferative activity of *Argemone gracilentia* and identification of some active components. *BMC Complement Altern. Med.* 2015. Vol. 5 № 15 P. 13. doi: 10.1186/s12906-015-0532-8.

30. Chopra R.N., Nayar S.L., Chopra I.C. Glossary of Indian Medicinal Plants. NISCOM CSIR; New Delhi, India: 1956. p. 23.

31. Sharma J., Gairola S., Gaur R.D., Painuli R.M. The treatment of jaundice with medicinal plants in indigenous communities of the Sub-Himalayan region of Uttarakhand, India. *J. Ethnopharmacol.* 2012 Vol. 143 P. 262–291. doi: 10.1016/j.jep.2012.06.034.

32. Prajapati N.D., Purohit S.S., Sharma A.K., Kumar T. A Handbook of Medicinal Plants. Agrobios; Jodhpur, India: 2003. p. 59–60.

33. Savithramma N., Sulochana C., Rao K.N. Ethnobotanical survey of plants used to treat asthma in Andhra Pradesh. *India J. Ethnopharmacol.* 2007 Vol. 113 P. 54–61. doi: 10.1016/j.jep.2007.04.004.

34 De Albuquerque U.P., Monteiro J.M., Ramos M.A., de Amorim E.L. Medicinal and magic plants from a public market in northeastern Brazil. *J. Ethnopharmacol.* 2007. Vol. 110 P. 76–91. doi: 10.1016/j.jep.2006.09.010.

35. Bieski I.G., Rios Santos F., de Oliveira R.M., Espinosa M.M., Macedo M., Albuquerque U.P., de Oliveira Martins D.T. Ethnopharmacology of medicinal plants of the pantanal region (Mato Grosso, Brazil). *Evid. Based Complement. Altern. Med.* 2012 Vol. 2012 P. 272749. doi: 10.1155/2012/272749.

36. Singh S.K., Pandey V.D., Singh A., Singh C. Antibacterial activity of seed extracts of *Argemone mexicana* L. on some pathogenic bacterial strain. *Afr. J. Biotechnol.* 2009 Vol. 8 P. 7077–7081.

37. Malik C., Mohanty J.P., Pradhan S., Sharma C. Phytochemistry and pharmacology of *Argemone mexicana* Linn-An Indian medicinal plant. *Res. J. Pharm. Phytochem.* 2023 Vol. 15 P. 27–32.

38. Daley S.K., Cordell G.A. Alkaloids in Contemporary Drug Discovery to Meet Global Disease Needs. *Molecules.* 2021 Vol. 26 № 13 P. 3800. doi: 10.3390/molecules26133800.

39. Faisal S., Badshah S.L., Kubra B., Emwas A.H., Jaremko M. Alkaloids as potential antivirals. A comprehensive review. *Nat. Prod. Bioprospect.* 2023. Vol. 13 № 1 P. 4. doi: 10.1007/s13659-022-00366-9.

40. Sato, F., Hashimoto, T., Hachiya, A., Yamada, Y. Metabolic engineering of plant alkaloid biosynthesis. *Proceedings of the National Academy of Sciences.* 2000. Vol. 98 № 1. P. 367-372. <https://doi.org/10.1073/pnas.98.1.367>.

41. National Center for Biotechnology Information Database. URL: <http://www.ncbi.nlm.nih.gov/> Retrieved April 3, 2023.

42. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 439654, (S)-Scoulerine. Retrieved April 3, 2023 from https://pubchem.ncbi.nlm.nih.gov/compound/S_-Scoulerine.

43. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 440229, (S)-Tetrahydrocolumbamine. Retrieved April 3, 2023 from https://pubchem.ncbi.nlm.nih.gov/compound/S_-Tetrahydrocolumbamine.

44. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 439653, Reticuline. Retrieved April 3, 2023 from <https://pubchem.ncbi.nlm.nih.gov/compound/Reticuline>.

45. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 440582, (S)-Cheilanthifoline. Retrieved April 3, 2023 from https://pubchem.ncbi.nlm.nih.gov/compound/S_-Cheilanthifoline.

46. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 2353, Berberine. Retrieved November 25, 2023 from <https://pubchem.ncbi.nlm.nih.gov/compound/Berberine>.

47. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 440927, (S)-Norcoclaurine. Retrieved November 25, 2023 from https://pubchem.ncbi.nlm.nih.gov/compound/S_-Norcoclaurine.

48. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 6057, Tyrosine. Retrieved November 25, 2023 from <https://pubchem.ncbi.nlm.nih.gov/compound/Tyrosine>.

49. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 6047, Levodopa. Retrieved November 25, 2023 from <https://pubchem.ncbi.nlm.nih.gov/compound/Levodopa>.

50. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 12407248, Benzophenanthridine. Retrieved November 27, 2023 from <https://pubchem.ncbi.nlm.nih.gov/compound/Benzophenanthridine>.

51. Smith T. F. Identification of Common Molecular Subsequences [Текст] / T. F. Smith, M. S. Waterman // Journal of Molecular Biology. - 1981. - Vol. 147. - P. 195–197.

52. Needleman S. B. A general method applicable to the search for similarities in the amino acid sequence of two proteins [Текст] / S. B. Needleman, C. D. Wunsch // Journal of Molecular Biology . - 1970. - Vol. 3: № 48. - P. 443–53.

53. Sneath, P.H.A. and Sokal, R.R. (1973) Numerical Taxonomy: The Principles and Practice of Numerical Classification. WF Freeman & Co., San Francisco, 573 p.

54. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004 Jul 27;101(30):11030-5. doi: 10.1073/pnas.0404206101.

55. Waterhouse A., Bertoni M., Bienert S. et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018. Vol. 46 (W1). P. W296-W303. DOI: 10.1093/nar/gky427.

56. Bienert S., Waterhouse A., de Beer T. et al. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res. 2017. Vol. 45. P. D313-D319. DOI:10.1093/nar/gkw1132.

57. Guex N., Peitsch M., Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis. 2009. Vol. 30. P. S162-S173. DOI:10.1002/elps.200900140.

58. Benkert P., Biasini M., Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011. Vol. 27. P. 343-350. DOI:10.1093/bioinformatics/btq662.

59. Bertoni M., Kiefer F., Biasini M., Bordoli L., Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci. Rep. 2017. Vol. 7 (1). P. 10480. DOI:10.1038/s41598-017-09654-8.

60. Grosdidier A., Zoete V., Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res. 2011, Vol. 39. P. W270-W277. DOI:10.1093/nar/gkr366.

61. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 2004. Vol. 25(13). P. 1605-1612. DOI:10.1002/jcc.20084.

62. Сліщук Г.І. Дослідження за допомогою методу симуляції молекулярної динаміки протеїну СҮР719А14 Аргемони мексиканської (*Argemone mexicana* L.). XXII Наукова конференція молодих учених ОДЕКУ. 23-31 травня 2023 р. / Одеський державний екологічний університет. Одеса. 2023. С. 307-309.

ДОДАТОК А

Код скрипту для симуляції молекулярної динаміки

```
#-----  
# Name:          MD Simulation  
# Purpose:  
#  
# Author:        SLISHCHUK  
#  
# Created:       14.04.2023  
# Copyright:     (c) SLISHCHUK 2023  
# Licence:       GNU FDL  
#-----  
def main():  
    pass  
if __name__ == '__main__':  
    main()  
import os  
import sys  
import time  
import numpy as np  
import MDAnalysis as mda  
from MDAnalysis import Universe  
from MDAnalysis.analysis.rms import rmsd, RMSD  
from MDAnalysis.analysis import align  
from MDAnalysis.core.groups import AtomGroup  
from MDAnalysis.topology.guessers import guess_atom_mass  
from PyQt5.QtCore import Qt, QThread, pyqtSignal  
from PyQt5.QtGui import QIcon, QPixmap  
from PyQt5.QtWidgets import (QApplication, QMainWindow, QFileDialog,  
                             QMessageBox, QLabel, QVBoxLayout, QHBoxLayout,  
                             QWidget, QAction, QPushButton, QLineEdit,  
                             QComboBox)  
from pyqtgraph.opengl import GLViewWidget, GLScatterPlotItem  
import pyqtgraph as pg  
import openmm as mm  
from openmm import app, Vec3  
from openmm.unit import nanometers, unit  
class MDRun(QThread):  
    rmsd_calculated = pyqtSignal(list)  
    positions_updated = pyqtSignal(np.ndarray)  
    def __init__(self, filename, selections, timestep, temperature):  
        super().__init__()  
        self.filename = filename  
        self.universe = None  
        self.selections = selections  
        self.timestep = timestep  
        self.temperature = temperature  
        self.rmsd_values = []  
        self.rmsd_calc = None  
        self.trajectory = None  
        self.positions_data = []  
    def run(self):
```

```

self.run_openmm_simulation(self.temperature)
try:
    self.universe = Universe(self.filename, 'trajectory.dcd', dt=self.timestep
)
    alignment = align.AlignTraj(self.universe, self.universe, select=self.sele
ctions, in_memory=True)
    alignment.run()
    protein = self.universe.select_atoms(self.selections)
    self.trajectory = protein.universe.trajectory
    self.rmsd_calc = RMSD(protein, protein)
    self.rmsd_calc.run()
    for i in range(len(self.trajectory)):
        self.trajectory[i]
        rmsd_value = self.rmsd_calc.rmsd[i, 2]
        self.rmsd_values.append(rmsd_value)
        self.plot_rmsd(self.rmsd_values)
except Exception as e:
    print(f"Exception in MDRun.run(): {e}")
    self.rmsd_values = None
topology = self.filename
self.universe = Universe(topology, 'trajectory.dcd')
for positions in self.positions_data:
    self.universe.atoms.positions = positions.value_in_unit(app.angstrom)
    self.universe.trajectory.write(self.universe.atoms)
self.universe.trajectory.rewind()
for i in range(len(self.trajectory)):
    self.trajectory[i]
    rmsd_value = self.rmsd_calc.rmsd[i, 2]
    self.rmsd_values.append(rmsd_value)
self.rmsd_calculated.emit(self.rmsd_values)
def run_openmm_simulation(self, temperature):
    from openmm import app
    import openmm as mm
    from openmm import unit
    forcefield = app.ForceField('amber14-all.xml', 'amber14/tip3pfb.xml')
    pdb = app.PDBFile(self.filename)
    modeller = app.Modeller(pdb.topology, pdb.positions)
    modeller.addHydrogens(forcefield)
    box_vectors = [Vec3(3.0, 0, 0), Vec3(0, 3.0, 0), Vec3(0, 0, 3.0)] * unit.nano
meter
    modeller.topology.setPeriodicBoxVectors(box_vectors)
    system = forcefield.createSystem(modeller.topology, nonbondedMethod=app.Cutoff
fNonPeriodic, nonbondedCutoff=1.0 * unit.nanometer, constraints=app.HBonds)
    integrator = mm.LangevinIntegrator(temperature * unit.kelvin, 1.0 / unit.pico
second, 1.0 * unit.femtoseconds)
    platform = mm.Platform.getPlatformByName('CPU')
    simulation = app.Simulation(modeller.topology, system, integrator, platform)
    simulation.context.setPositions(modeller.positions)
    simulation.minimizeEnergy()
    class CustomReporter(app.StateDataReporter):
        def __init__(self, parent, file, report_interval):
            super().__init__(file, report_interval)
            self.parent = parent
        def report(self, simulation, state):
            super().report(simulation, state)
            positions = state.getPositions(asNumpy=True)._value
            self.parent.positions_updated.emit(positions)
    simulation.reporters.append(app.DCDReporter('trajectory.dcd', 1000))

```

```

        simulation.reporters.append(app.StateDataReporter('data.csv', 1000, step=True
, potentialEnergy=True, temperature=True))
        simulation.step(10000)
    def plot_rmsd(self, rmsd_values):
        pass
class ProteinViewer(QMainWindow):
    def __init__(self):
        super().__init__()
        self.setWindowTitle('MD Simulation')
        self.setGeometry(100, 100, 800, 800)
        self.setWindowIcon(QIcon('icon.png'))
        self.create_menu()
        self.create_toolbar()
        self.create_statusbar()
        self.filename = None
        self.universe = None
        self.filename2 = None
        self.universe2 = None
        self.timestep = 1
        self.gl_widget1 = None
        self.create_layout()
        self.md_thread = MDRun('', '', 1, 300)
        self.md_thread.rmsd_calculated.connect(self.update_rmsd_plot)
        self.md_thread_rmsd_values = []
    def update_rmsd_plot(self, rmsd_values):
        self.md_thread_rmsd_values = rmsd_values
        self.plot_rmsd(self.md_thread_rmsd_values)
    def show_about_dialog(self):
        QMessageBox.about(self, 'About', 'This is a simple MDAnalysis-
based tool for calculating RMSD.')
    def create_menu(self):
        menu_bar = self.menuBar()
        file_menu = menu_bar.addMenu('File')
        open_action = QAction('Open', self)
        open_action.setShortcut('Ctrl+O')
        open_action.setStatusTip('Open PDB file')
        open_action.triggered.connect(self.open_file_dialog)
        file_menu.addAction(open_action)
        file_menu.addSeparator()
        exit_action = QAction('Exit', self)
        exit_action.setShortcut('Ctrl+Q')
        exit_action.setStatusTip('Exit application')
        exit_action.triggered.connect(self.close)
        file_menu.addAction(exit_action)
        help_menu = menu_bar.addMenu('Help')
        about_action = QAction('About', self)
        about_action.setStatusTip('About this application')
        about_action.triggered.connect(self.show_about_dialog)
        help_menu.addAction(about_action)
    def create_toolbar(self):
        self.toolbar = self.addToolBar('Toolbar')
        self.toolbar.setMovable(False)
        load_pdb_action = QAction(QIcon('open.png'), 'Load PDB', self)
        load_pdb_action.triggered.connect(self.open_file_dialog)
        self.toolbar.addAction(load_pdb_action)
        run_md_action = QAction(QIcon('play.png'), 'Run MD', self)
        run_md_action.triggered.connect(self.run_md)
        self.toolbar.addAction(run_md_action)

```

```

def create_statusbar(self):
    self.statusbar = self.statusBar()
    self.statusbar.showMessage("Ready")
def visualize_protein(self, universe, gl_widget):
    print("Trying to visualize protein...")
    if not universe:
        return
    if not gl_widget:
        gl_widget = GLViewWidget()
    self.gl_widget1 = gl_widget
    gl_widget.clear()
    gl_widget.setCameraPosition(distance=100)
    gl_widget.setBackgroundColor('k')
    selection = self.selection_combo.currentText()
    if selection == 'all':
        atoms = universe.atoms
    else:
        atoms = universe.select_atoms(selection)
    coords = atoms.positions
    coords = coords.astype(np.float32)
    scatter = GLScatterPlotItem(pos=coords, size=1, pxMode=False)
    gl_widget.addItem(scatter)
    colors = []
    for atom in atoms:
        if atom.element == 'C':
            colors.append((0, 1, 0, 1))
        elif atom.element == 'N':
            colors.append((0, 0, 1, 1))
        elif atom.element == 'O':
            colors.append((1, 0, 0, 1))
        elif atom.element == 'S':
            colors.append((1, 1, 0, 1))
        elif atom.element == 'H':
            colors.append((0.8, 0.8, 0.8, 0.5))
        else:
            colors.append((0.5, 0.5, 0.5, 0.5))
    colors = np.array(colors, dtype=np.float32)
    scatter.setData(color=colors)
    return gl_widget
def update_visualization(self, positions):
    if not self.universe or not self.gl_widget1:
        return
    protein_atoms = self.universe.select_atoms("protein")
    protein_atoms.positions = positions
    self.visualize_protein(self.universe, self.gl_widget1)
def load_file(self):
    filename = self.filename
    if filename is None:
        QMessageBox.warning(self, 'Error', 'Please load a PDB file first')
        return
    try:
        self.universe = Universe(filename)
        self.statusBar().showMessage('PDB file loaded')
        self.visualize_protein(self.universe, self.gl_widget1)
    except Exception as e:
        self.statusBar().showMessage('Failed to load PDB file')
        QMessageBox.warning(self, 'Error', 'Failed to load PDB file')
        self.visualize_protein(None, self.gl_widget1)

```

```

def load_pdb_finished(self, load_thread):
    if load_thread.universe is not None:
        self.universe = load_thread.universe
        self.statusBar().showMessage('PDB file loaded')
        self.visualize_protein(self.universe, self.gl_widget1)
    else:
        self.statusBar().showMessage('Failed to load PDB file')
        QMessageBox.warning(self, 'Error', 'Failed to load PDB file')
        self.visualize_protein(None, self.gl_widget1)
def open_file_dialog(self):
    options = QFileDialog.Options()
    options |= QFileDialog.ReadOnly
    filename, _ = QFileDialog.getOpenFileName(self, 'Open PDB File', '', 'PDB
Files (*.pdb);;All Files (*)', options=options)
    if filename:
        self.filename = filename
        self.statusBar().showMessage(f'Loaded {filename}')
        self.label1.setText(f'File: {filename}')
        self.load_file()
def run_md(self):
    if self.universe is None:
        QMessageBox.warning(self, 'Error', 'Please load a PDB file first')
        return
    try:
        self.timestep = float(self.timestep_edit.text())
    except ValueError:
        QMessageBox.warning(self, 'Error', 'Invalid timestep value. Please ent
er a valid number.')
        return
    try:
        self.md_thread.positions_updated.disconnect()
    except TypeError:
        pass
    self.statusBar().showMessage('Running molecular dynamics simulation...')
    temperature = float(self.temperature_edit.text())
    self.md_thread = MDRun(self.filename, self.selection_combo.currentText(),
self.timestep, temperature)
    self.md_thread.rmsd_calculated.connect(self.update_rmsd_plot)
    self.md_thread.positions_updated.connect(self.update_visualization)
    self.md_thread.finished.connect(lambda: self.md_calculation_finished(self.
md_thread))
    self.md_thread.start()
def md_calculation_finished(self, md_thread):
    rmsd_values = md_thread.rmsd_values
    if rmsd_values is not None:
        self.statusBar().showMessage(f'RMSD: {rmsd_values[-1]:.3f}')
        self.plot_rmsd(rmsd_values)
    else:
        self.statusBar().showMessage('Failed to run MD simulation')
        QMessageBox.warning(self, 'Error', 'Failed to run MD simulation')
    if md_thread.universe is not None:
        self.universe = md_thread.universe
        self.statusBar().showMessage('MD simulation data loaded')
        self.visualize_protein(self.universe, self.gl_widget1)
    else:
        self.statusBar().showMessage('Failed to load MD simulation data')
        QMessageBox.warning(self, 'Error', 'Failed to load MD simulation data')
        self.visualize_protein(None, self.gl_widget1)

```

```

def show_trajectory(self):
    if self.universe is None:
        QMessageBox.warning(self, 'Error', 'Please load a PDB file first')
        return
    self.universe.load_new('trajectory.dcd')
    self.visualize_protein(self.universe, self.gl_widget1)
    protein = self.universe.select_atoms(self.selection_combo.currentText())
    u = self.universe
    protein = u.select_atoms('protein')
    masses = [1.008 for _ in protein]
    for atom, mass in zip(protein, masses):
        atom.mass = mass
    protein.masses = masses
    for _ in self.universe.trajectory:
        protein_centered = protein.positions - protein.center_of_mass()
        self.update_visualization(protein_centered)
        QApplication.processEvents()
        time.sleep(0.1)
        print(f"Trajectory visualisation")
def plot_rmsd(self, rmsd_values):
    self.graph_widget.clear()
    self.graph_widget.plot(y=rmsd_values, pen='b')
def create_layout(self):
    vbox = QVBoxLayout()
    gl_layout = QHBoxLayout()
    self.gl_widget1 = GLViewWidget()
    gl_layout.addWidget(self.gl_widget1, stretch=2)
    vbox.addLayout(gl_layout, stretch=8)
    hbox1 = QHBoxLayout()
    self.label1 = QLabel(self)
    self.label1.setAlignment(Qt.AlignCenter)
    hbox1.addWidget(self.label1)
    hbox2 = QHBoxLayout()
    self.timestep_label = QLabel('Timestep (ps):')
    self.timestep_edit = QLineEdit('1')
    self.timestep_edit.setFixedWidth(50)
    hbox2.addWidget(self.timestep_label)
    hbox2.addWidget(self.timestep_edit)
    hbox4 = QHBoxLayout()
    self.temperature_label = QLabel('Temperature (K):')
    self.temperature_edit = QLineEdit('300')
    self.temperature_edit.setFixedWidth(50)
    hbox4.addWidget(self.temperature_label)
    hbox4.addWidget(self.temperature_edit)
    vbox.addLayout(hbox1, stretch=1)
    vbox.addLayout(hbox2, stretch=1)
    vbox.addLayout(hbox4, stretch=1)
    hbox3 = QHBoxLayout()
    self.selection_combo = QComboBox(self)
    self.selection_combo.addItem('protein')
    self.selection_combo.addItem('backbone')
    self.selection_combo.addItem('name CA')
    self.selection_combo.addItem('all')
    hbox3.addWidget(self.selection_combo)
    hbox3.addStretch()
    show_trajectory_button = QPushButton('Show Trajectory', self)
    show_trajectory_button.clicked.connect(self.show_trajectory)
    hbox3.addWidget(show_trajectory_button)

```



```
vbox.addLayout(hbox3, stretch=1)
graph_widget = pg.PlotWidget()
graph_widget.setLabel('left', 'RMSD (Å)')
graph_widget.setLabel('bottom', 'Frame')
self.graph_widget = graph_widget
vbox.addWidget(self.graph_widget, stretch=3)
central_widget = QWidget(self)
central_widget.setLayout(vbox)
self.setCentralWidget(central_widget)

qt_app = QApplication(sys.argv)
window = ProteinViewer()
window.show()
sys.exit(qt_app.exec_())
```

ДОДАТОК Б

Код скрипту для додавання атомів водню

```
#-----  
# Name:      ProteinH  
# Purpose:  
#  
# Author:    SLISHCHUK  
#  
# Created:   16.04.2023  
# Copyright: (c) SLISHCHUK 2023  
# Licence:   <GNU FDL>  
#-----  
def main():  
    pass  
  
if __name__ == '__main__':  
    main()  
import sys  
import tkinter as tk  
from tkinter import filedialog  
from simtk.openmm import app  
def add_hydrogens_to_pdb(input_pdb_path, output_pdb_path):  
    pdb = app.PDBFile(input_pdb_path)  
    forcefield = app.ForceField('amber14-all.xml', 'amber14/tip3pfb.xml')  
    modeller = app.Modeller(pdb.topology, pdb.positions)  
    modeller.addHydrogens(forcefield)  
    app.PDBFile.writeFile(modeller.topology, modeller.positions, open(output_pdb_p  
ath, 'w'))  
def browse_input_file():  
    input_file_path.set(filedialog.askopenfilename(title="Select a PDB file"))  
def browse_output_file():  
    output_file_path.set(filedialog.asksaveasfilename(title="Save updated PDB file  
"))  
def process_pdb():  
    add_hydrogens_to_pdb(input_file_path.get(), output_file_path.get())  
  
root = tk.Tk()  
root.title("Add Hydrogens to PDB")  
  
input_file_path = tk.StringVar()  
output_file_path = tk.StringVar()  
  
input_label = tk.Label(root, text="Input PDB file:")  
input_label.grid(row=0, column=0, padx=(10, 0), pady=10, sticky='w')  
  
input_entry = tk.Entry(root, textvariable=input_file_path)  
input_entry.grid(row=0, column=1, padx=(10, 0), pady=10, sticky='we')  
  
input_button = tk.Button(root, text="Browse", command=browse_input_file)  
input_button.grid(row=0, column=2, padx=(10, 10), pady=10)  
  
output_label = tk.Label(root, text="Output PDB file:")  
output_label.grid(row=1, column=0, padx=(10, 0), pady=10, sticky='w')
```

```
output_entry = tk.Entry(root, textvariable=output_file_path)
output_entry.grid(row=1, column=1, padx=(10, 0), pady=10, sticky='we')

output_button = tk.Button(root, text="Browse", command=browse_output_file)
output_button.grid(row=1, column=2, padx=(10, 10), pady=10)

process_button = tk.Button(root, text="Add Hydrogens and Save", command=process_pd
b)
process_button.grid(row=2, column=1, padx=(10, 0), pady=10)

root.columnconfigure(1, weight=1)
root.mainloop()
```