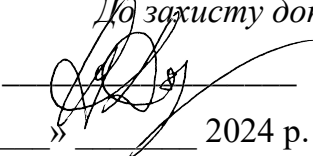


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
імені В. Н. Каразіна

Кафедра органічної хімії

До захисту допускаю

Завідувач кафедри
«__» ____ 2024 р. д.х.н., проф. А. О. Дорошенко

**ПРОГНОЗУВАННЯ ТРИВИМІРНУ БУДОВИ БІЛКІВ-МІШЕНЕЙ З
ВИКОРИСТАННЯМ МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ DEERMIND
ALPHAFOLD 3**

Кваліфікаційна робота
студентки 2 курсу магістратури
хімічного факультету



Думи Єлизавети Володимирівни

Науковий керівник:

д.х.н., с.н.с.



О.В. Кириченко

ХАРКІВ 2024

РЕФЕРАТ

Кваліфікаційна робота викладена на 44 сторінках, містить 22 рисунків та 63 літературних посилань.

Нейронна мережа AlphaFold 3 стала проривом у прогнозування третинної будови білка на молекулярному рівні. Точність таких прогнозів відіграє ключову роль у розробці нових ліків, зокрема в контексті інгібіторів, що зв'язуються з активним центром протеази. Окрім того, отримані структури можуть значно покращити розуміння механізмів антибіотикорезистентності, що відкриває нові горизонти в боротьбі з інфекціями. В роботі показано, що AlphaFold 3 продемонстрував свій потенціал як потужний інструмент для розвитку терапевтичних стратегій, підвищуючи ефективність пошуку ліків та покращення результатів досліджень у біології та фармацевтиці. Однак є потенціал для подальшого вдосконалення методу, що дозволить ще точніше прогнозувати структури для більш складних молекул.

Об'єкти дослідження – білкові структури, зокрема β -глюкозидаза, сіртруїн, білки вірусів SARS-CoV-2 та протитуберкульозні білки

Мета роботи – дослідження теоретичних основ роботи AlphaFold 3 і визначення його значення для фармацевтичної хімії, зокрема в контексті розробки нових лікарських засобів, прогнозування білкових структур та аналізу їх взаємодії з молекулами.

Методи дослідження – теоретичний аналіз алгоритмів прогнозування білкових структур (AlphaFold 3), порівняння результатів моделювання з експериментальними даними.

Ключові слова: AlphaFold 3, прогнозування білкових структур, фармацевтична хімія, SARS-CoV-2, антибіотикорезистентність, структурне моделювання, методи докінгу, білкові взаємодії, розробка ліків.

ABSTRACT

The MSc diploma thesis consists of 44 pages, 22 figures, and 63 references.

The AlphaFold 3 neural network was a breakthrough in predicting the tertiary structure of a protein at the molecular level. The accuracy of such predictions plays a key role in the development of new drugs, particularly in the context of inhibitors that bind to the active site of the protease. In addition, the obtained structures can significantly improve the understanding of the mechanisms of antibiotic resistance, which opens new horizons in the fight against infections. The thesis shows that AlphaFold 3 has demonstrated its potential as a powerful tool for the development of therapeutic strategies, increasing the efficiency of drug discovery and improving the results of research in biology and pharmaceuticals. However, there is potential for further improvement of the method, which will allow even more accurate prediction of structures for more complex molecules.

Objects of a study - protein structures, in particular β -glucosidase, sirtuin, proteins of SARS-CoV-2 viruses and anti-tuberculosis proteins

The aim of the investigation was to investigate the theoretical basis of AlphaFold 3 and determine its importance for pharmaceutical chemistry, in particular in the context of drug discovery, prediction of protein structures and analysis of their interaction with molecules.

Research methods - Research methods - theoretical analysis of protein structure prediction algorithms (AlphaFold 3), comparison of modeling results with experimental data.

Key words: AlphaFold 3, protein structure prediction, pharmaceutical chemistry, SARS-CoV-2, antibiotic resistance, structural modeling, docking methods, protein interactions, drug development

ЗМІСТ

	Стор.
ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	5
ВСТУП	6
1. ОГЛЯД НАУКОВОЇ ЛІТЕРАТУРИ	7
1.1 Білки та їх роль у біологічних процесах	7
1.1.1 Основні рівні структурної організації білків: первинна, вторинна, третинна та четвертинна структури	7
1.1.2 Роль тривимірної структури в функціонуванні білків	9
1.1.3 Визначення структури за допомогою експериментальних методів (рентгеноструктурний аналіз, ЯМР, криоелектронна мікроскопія) та їх обмеження	11
1.2 Комп'ютерне моделювання білків	15
1.2.1 Методи передбачення структури: гомологічне моделювання, аб ініцію, де ново)	15
1.2.2 Роль штучного інтелекту в сучасній хімії та біології	16
1.3 Алгоритми глибокого навчання та AlphaFold	18
1.3.1 DeepMind та розробка AlphaFold: основи роботи нейронних мереж	18
1.3.2 Технологічні принципи AlphaFold 3: покращення точності моделювання в порівнянні з попередніми версіями	21
1.3.3 Роль AlphaFold 3 у вирішенні проблеми передбачення структури білків	24
2. РОЗРАХУНКОВА ЧАСТИНА	26
2.2 Приклад моделювання певних білків	27
2.3 Оцінка результатів: порівняння з експериментальними даними та іншими методами	28
ВИСНОВКИ	38
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	39

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

AlphaFold 3 (AF3)	остання версія програми DeepMind AlphaFold для прогнозування структури білків
pLDDT	predicted Local Distance Difference Test, оцінка точності передбаченої структури білка
RMSD	Root Mean Square Deviation, міра відхилення між передбаченою та реальною структурами білка
Homology Modeling (HM)	гомологічне моделювання для прогнозування структури білка
Molecular Docking (MD)	докінг молекул для прогнозування взаємодій білків з ліками
Template Modeling score (TM-score)	оцінка схожості між прогнозованою та встановленою структурами білка.
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2 коронавірус, що викликає COVID-19
PDB	Protein Data Bank, база даних структур білків

ВСТУП

Прогнозування тривимірної структури білків є однією з найбільш важливих задач у сучасній біоінформатиці та фармацевтичній хімії. Розуміння просторової організації молекул білків є ключовим для дослідження їх функцій та механізмів взаємодії з ліками, що відкриває нові можливості для розробки ефективних лікарських засобів. Традиційні методи моделювання, такі як гомологічне моделювання (НМ) та молекулярний докінг (MD), використовуються для передбачення структур білків, однак вони часто мають обмеження, зокрема через необхідність наявності близьких гомологів для моделювання або обмежену точність прогнозів.

З приходом новітніх технологій, таких як AlphaFold 3 (AF3), процес прогнозування структур білків значно покращився. AlphaFold, завдяки використанню глибоких нейронних мереж, може передбачити третинну структуру білків з високою точністю, навіть у випадках, коли традиційні методи не дають задовільних результатів. Ця технологія вже показала себе як потужний інструмент у біоінформатиці, що є важливим для розробки нових противірусних препаратів. У 2024 році за досягнення у галузі прогнозування структури білків команда, що стоїть за розвитком AlphaFold, отримала Нобелівську премію з хімії.

У цій роботі ми розглянемо використання AlphaFold 3 для прогнозування структури білків тих, що мають значення для фармацевтичних досліджень, таких як білки вірусів (зокрема SARS-CoV-2) та бактеріальні білки, які можуть бути причиною антибіотикорезистентності; порівняємо точність отриманих структур з існуючими експериментальними даними та оцінимо потенціал цієї технології в контексті розробки нових лікарських засобів та покращення розуміння механізмів вірусної інфекції.

1. ОГЛЯД НАУКОВОЇ ЛІТЕРАТУРИ

1.1. Білки та їх у біологічних процесах

1.1.1. Основні рівні структурної організації білків: первинна, вторинна, третинна та четвертинна структури

Білки відіграють центральну роль у забезпеченні молекулярних функцій, будучи кінцевими продуктами численних інформаційних шляхів — складних біохімічних каскадів, за допомогою яких клітини отримують, обробляють і передають інформацію для підтримки життєдіяльності. Ці шляхи включають передачу сигналів, регуляцію генетичної активності, відповіді на зовнішні подразники та підтримання гомеостазу.

Білки виконують більшість життєво важливих клітинних процесів. Вони каталізують майже всі молекулярні реакції. Також білки виступають як важливі структурні елементи клітини, надаючи їй форми та міцності. Визначення функцій білка тісно пов'язане з розумінням його структури: особливості складу й форми дозволяють передбачити, які саме функції виконує кожен білок і як він вписується в загальний контекст клітинної діяльності.

Білкова молекула складається з ланцюга амінокислот, кожна з яких пов'язана з сусідньою за допомогою ковалентного пептидного зв'язку. Такий зв'язок утворюється внаслідок дегідратації карбоксильної групи однієї амінокислоти та аміногрупи іншої. Переважна більшість поліпептидів містить від 100 до 1000 залишків амінокислот [1].

Такий діапазон розмірів зумовлено тим, що ланцюгу з менш ніж 40 амінокислот зазвичай недостатньо для формування стабільної структури, яка б могла виконувати специфічні функції, для ланцюгів з понад 1000 амінокислот так само знижується стабільність та ефективність.

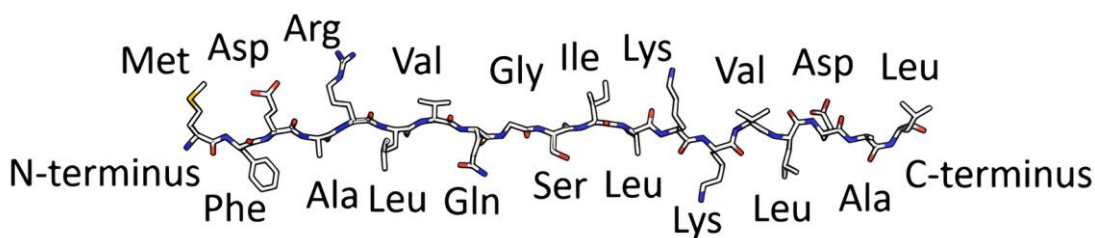


Рис. 1.1 - Поліпептидний ланцюг залишків амінокислот

Багато білків складаються не лише з амінокислотних залишків. Вони можуть утворювати комплекси з іонами металів, можуть ковалентно або нековалентно зв'язувати певні невеликі органічні молекули, а також можуть бути ковалентно модифіковані шляхом посттрансляційного приєднання груп, таких як фосфати і вуглеводи [1]. Неамінокислотну частину кон'югованого білка називають простетичною групою. Зазвичай простетична група відіграє важливу роль у біологічній функції білка.

Властивості конкретного білка визначаються головним чином порядком амінокислот у його ланцюгу, а не просто їхньою наявністю чи складом.

Послідовність амінокислот у пептидному ланцюзі білка називається його первинною структурою (Рис. 2). Різні ділянки послідовності утворюють вторинні структури, відомі як альфа (α) спіралі або бета (β) листки (Рис. 3). Такі патерни згортання характеризують локальну геометрію і це відіграє певну роль у поведінці різних сегментів білка.

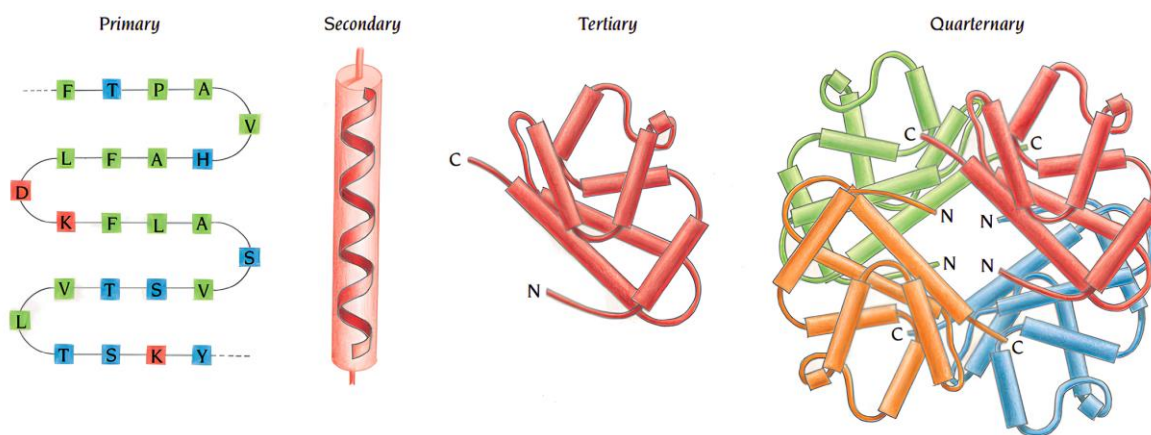


Рис. 1.2 - Первинна, вторинна, третинна та четвертинна структури білка [2].



Рис. 1.3 - Сегменти вторинної структури білка: α -спіралі та β -листки

Третинна структура білка утворюється завдяки щільному укладанню альфа-спіралей та бета-ниток, формуючи глобулярні одиниці, відомих як домени.

Четвертинна структура формується, коли кілька поліпептидних ланцюгів (субодиниць) об'єднуються в одну функціональну молекулу. У таких білках функціональні ділянки, зокрема активний центр, де відбуваються основні хімічні реакції, часто розташовані на межі між субодиницями або в межах одного з доменів, забезпечуючи ключову роль у біологічній активності білка.

Кожен рівень структури є критично важливим для стабільності, функції та специфічності білків у біологічних процесах.

1.1.2. Роль тривимірної структури в функціонуванні білків

Згортання білка — це процес, за допомогою якого поліпептидний ланцюг набуває правильної тривимірної структури (конформації), щоб досягти біологічно активного нативного стану. Поліпептидні ланцюги згортаються в нативний стан як спонтанно, так і за допомогою ферментів для каталізації утворення та обміну дисульфідних зв'язків. Велика кількість білків також потребує допомоги класу білків – шаперонів.

Білок не є статичним. Зокрема, його сегменти у багатьох випадках є внутрішньо невпорядкованими, хоча і повністю функціональними. Багато білків зберігають природну структуру в нативному стані, але загалом можуть приймати незліченну кількість конформацій за рахунок коливань атомів або ж груп атомів. До таких конформацій входять ті, які можуть бути досягнені без

розриву ковалентних зв'язків. Білки можуть мати схожі функції, проте унікальність тривимірної структури кожного білка доводиться наявністю специфічної структурної або хімічної функції.

Класифікація білків здійснюється за співвідношенням α -спіралей і β -листіків.

Як правило, конформація поліпептидного ланцюга мінімізує його вільну енергію. При обробці певними розчинниками руйнуються нековалентні взаємодії, що призводить до денатурації білка. Проте після видалення денатуруючого розчинника, білок зазвичай ренатурується у свою початкову тривимірну форму, що свідчить про те, що уся необхідна інформація для визначення конформації білка міститься в амінокислотній послідовності. [3]

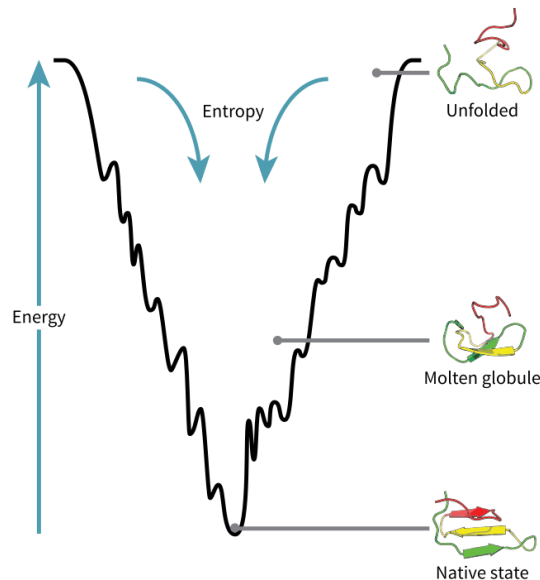


Рис. 1.4 - Енергетичний ландшафт згортання білка. Білок в стані з високою ентропією (невпорядкованістю), має найбільш нестабільний стан. Потенційна енергія білка залежить від його конформацій [4].

Зв'язування білка з лігандом у місці сайту на поверхні білка відбувається завдяки великій кількості водневих і слабких нековалентних зв'язків, електростатичних і ван-дер-ваальсових притягань. Такі взаємодії провокують білок підвищувати хімічну активність шляхом утворення міцніших зв'язків завдяки взаємодії сусідніх ділянок поліпептидного ланцюга. Так відбувається,

оскільки молекули води можуть конкурувати з лігандами. А якщо при згортанні білка відбувається зближення негативно заряджених бічних ланцюгів всупереч їх взаємному відштовхуванню, спорідненість сайту до позитивно зарядженого іона значно підвищується. Структурна адаптація, що відбувається між білком і лігандом називається індукованою відповідністю. Тобто, розташування бічних ланцюгів відносно один одного впливають на активність молекули, що показує деяку відмінність властивостей дещо різних конформацій одного і того ж білка.

1.1.3. Визначення структури за допомогою експериментальних методів (рентгеноструктурний аналіз, ЯМР, криоелектронна мікроскопія) та їх обмеження

Для роботи з білками та для їх дослідження використовують різні лабораторні методи. Концентрацію білка можна визначити за допомогою абсорбційної спектроскопії, а масу завдяки мас-спектрометрії. Для розділення білків на фракції використовують, наприклад, колонкову хроматографію. Електрофорез розділяє молекули відповідно до заряду та розміру.

Ще у другій половині 20-го століття вченими застосовувався метод Едмана (*Edman degradation*) [5] задля визначення первинної послідовності (секвенування) пептидів. Він дозволяв встановлювати послідовність амінокислот у поліпептидному ланцюзі білків. Основне обмеження методу — він підходить лише для відносно коротких пептидів, до 50 амінокислот, тому на зміну йому прийшли альтернативні дослідження іншими методами.

Один із них – мас-спектрометрія. Білки досліджують іонізаційною мас-спектрометрією з лазерною десорбцією/іонізацією (MALDI MS) [6][7] (Рис. 1.5а) або ж електророзпилювальною іонізаційною мас-спектрометрією (ESI MS) [8] (Рис. 1.5в) для запобігання розпаду таких масивних молекул, як білки. Цей метод обмежує дослідження, оскільки не дає уявлення про

третинну структуру білка, а дає змогу визначити амінокислотну послідовність первинної структури.

Про тривимірну структуру білка високоякісні дані дає рентгенівська кристалографія (Рис. 1.6). Проте вона має суттєві обмеження — білок повинен утворювати кристали, що не завжди можливо, оскільки деякі білки важко кристалізувати. Крім того, кристалографія не надає інформації про білки в природному, розчиненому стані. Також в багатьох випадках метод РСА не є ефективним при вивченні білок-білкових та білок-нуклеїнових комплексів.

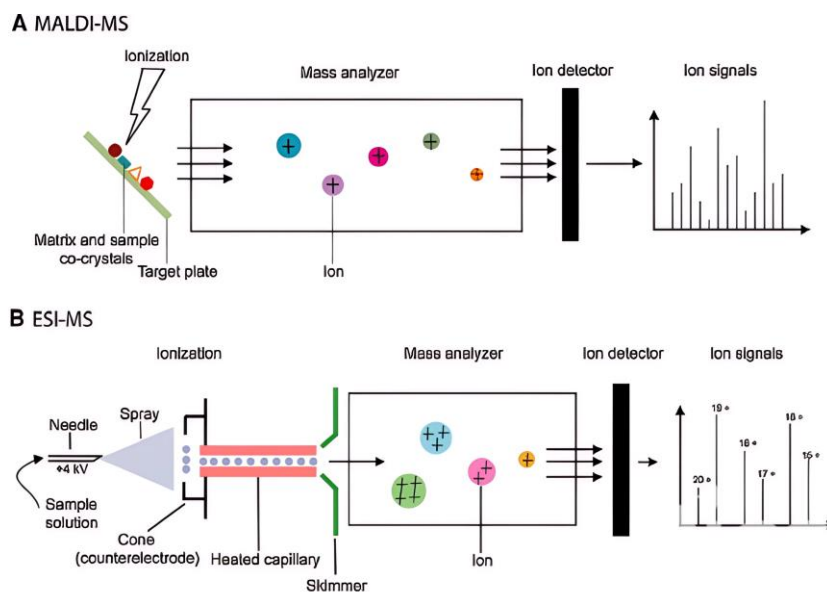


Рис. 1.5 - а) У MALDI-MS зразок кристалізують з органічною матрицею на металевій пластині. Лазерний імпульс нагріває кристали, що спричиняє десорбцію йонів у газову фазу.

У ESI-MS молекули зразка іонізуються прямо в розчині, проходячи через нагріту капілярну трубку, що утворює аерозоль крапель у вакуумі з високим електричним полем [9].

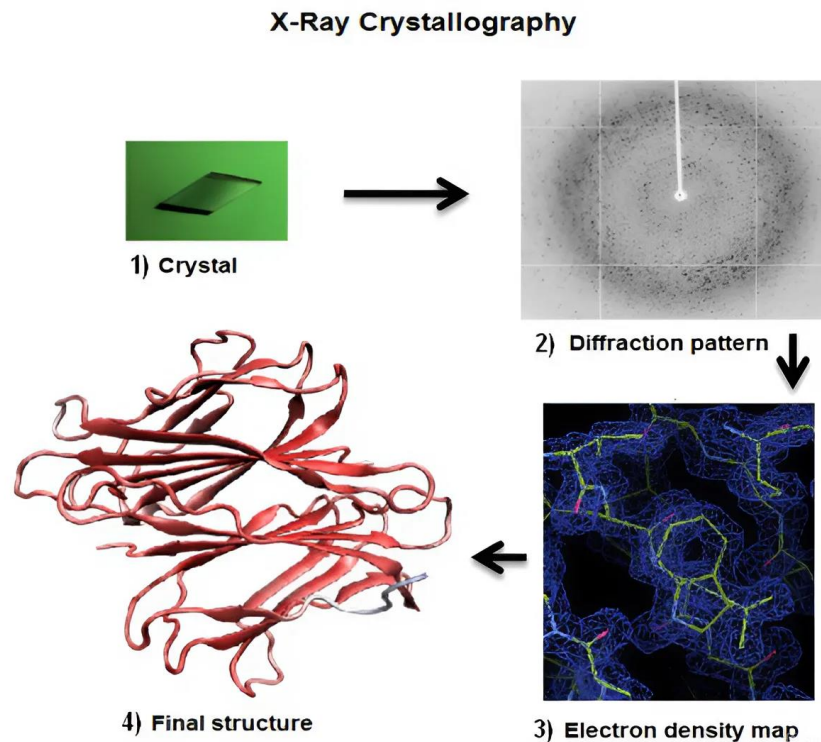


Рис. 1.6 - Визначення структури білка за допомогою кристалографічного дослідження [10].

Вирішенням проблеми необхідності кристалізації стає малокутове рентгенівське та нейтронне розсіювання (SAXS та SANS) – методи, що дозволяють аналізувати великі комплекси безпосередньо в розчині, що дозволяє краще контролювати умови експерименту. Проте, обидва методи обмежені роздільною здатністю та чутливістю до домішок і гетерогенності.

Спектроскопія ядерного магнітного резонансу (ЯМР) є альтернативним методом визначення просторової структури білка. Цей метод (рис. 1.7) добре підходить для дослідження білків у розчиненому стані, дозволяючи аналізувати їхню динаміку та конформаційні зміни, хоча вимагає відносно великої кількості матеріалу, а також білок має бути стабільним при кімнатній температурі під час збору даних.

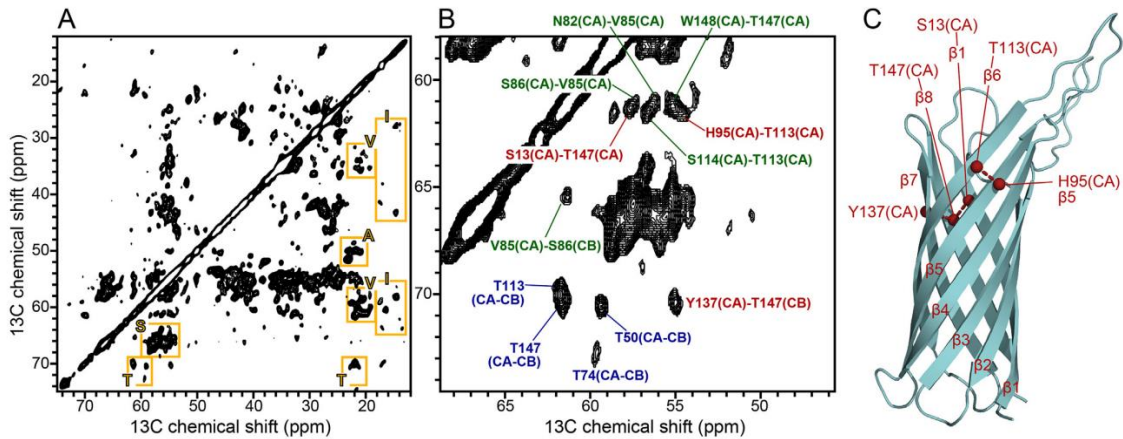


Рис. 1.7 - 2D ^{13}C - ^{13}C спектр (^{15}N , ^{13}C , ^2H) Ail в ліпосомах [11].

Кріоелектронна мікроскопія (кріо-ЕМ) (Рис. 1.8) дозволяє отримати зображення білків у близькому до природного стану без потреби кристалізації. Кріо-ЕМ особливо корисна для великих білкових комплексів. У 2017 році [12] навіть було присуджено Нобелівську премію за розробку кріоелектронної мікроскопії для візуалізації біомолекул. Хоча метод і доволі популярний, він все ж таки передбачає деякі складності підготовки проби, а сама процедура вимагає спеціальних знань та програмного забезпечення.

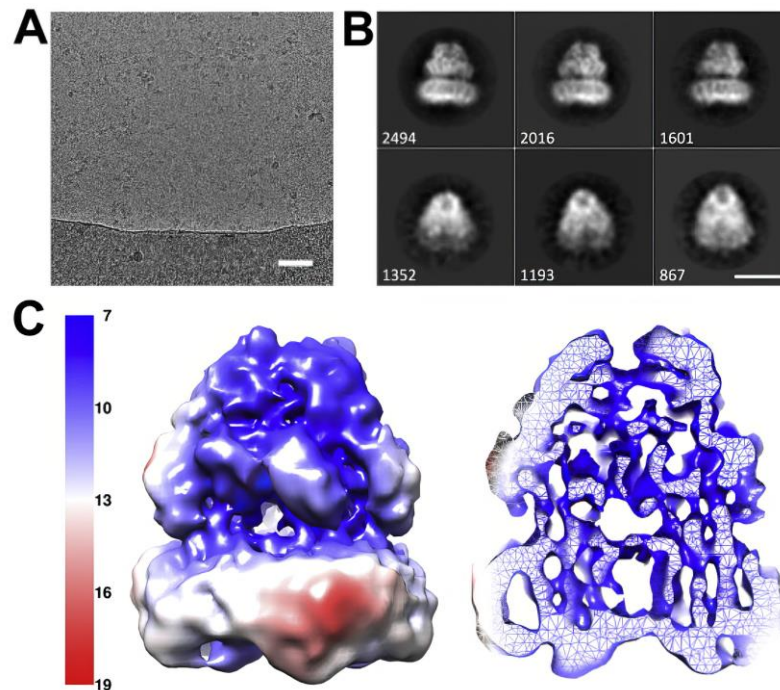


Рис. 1.8 - Набір даних сітки кріо-ЕМ для білку AcrV та репрезентативна реконструкція частинки [13].

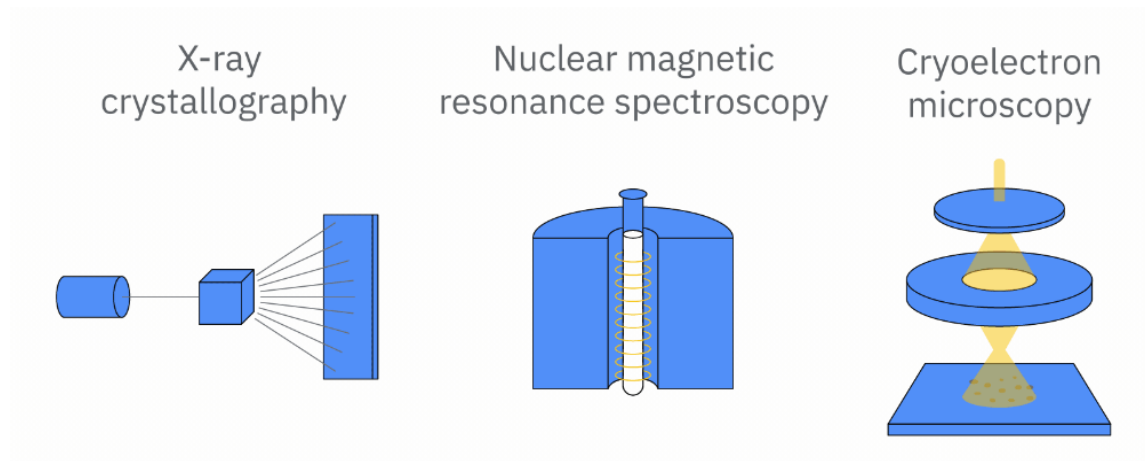


Рис. 1.9 - Традиційні методи дослідження структури білка.

1.2. Комп'ютерне моделювання білків

1.2.1. Методи передбачення структури: гомологічне моделювання, *ab initio*, *de novo*

Гомологічне моделювання [14-15] базується на припущенні, що білки зі схожими амінокислотними послідовностями матимуть подібні тривимірні структури. Цей підхід є особливо ефективним, якщо існують експериментально визначені структури (шаблони) білків. За наявності близьких гомологів у базах даних, метод є швидким і точним. Відбувається пошук шаблону, вирівнювання амінокислотної послідовності, побудова моделі, оптимізація і валідація.

Методи *ab initio* [16] прагнуть передбачити структуру білка, виходячи виключно з його амінокислотної послідовності, без використання шаблонів. Цей підхід базується на фізико-хімічних принципах взаємодії між атомами та молекулами та на припущенні того, що найнижче значення вільної енергії відповідає нативній структурі. Основні кроки включають моделювання енергетичного ландшафту (визначення конформації з найнижчою вільною енергією) і пошук конфігурації. Такі методи потребують значних

обчислювальних ресурсів і часто застосовуються для невеликих білків або фрагментів.

Методи *de novo* [17] є подібними до *ab initio*, але можуть використовувати інформацію з баз даних структур білків. Це дозволяє створювати моделі, використовуючи елементи вже відомих структур (наприклад, вторинні структурні мотиви) для зменшення обчислювальної складності.

Всі три методи традиційно відносять до структурного моделювання[18], що дозволяє моделювати специфічні взаємодії молекул у сайті зв'язування білків, що є важливим для розробки ліків, здатних ефективно взаємодіяти з біологічними мішенями.

Методи структурного моделювання також забезпечують основу для лігандного моделювання, що використовується для передбачення афінності ліганду до білка (наскільки добре молекула зв'язується), що має значення для розробки ліків. Такі методи, зокрема кількісний зв'язок структура-активність (QSAR) та аналіз хімічної схожості, базуються на інформації про вже відомі молекули, що проявляють терапевтичну активність.

1.2.2. Роль штучного інтелекту в сучасній хімії та біології

На основі експериментально визначених структур білків були розроблені певні емпіричні правила та алгоритми передбачення їх 3D-структури, які згодом були покладені в основу молекулярного моделювання просторової організації білків. Ці методи розвивалися спочатку як допоміжні при уточненні RSA- та ЯМР-структур, що дало змогу встановити загальні правила упаковки поліпептидних ланцюгів, а згодом розвинути новий підхід на основі конформаційного аналізу та мінімізації вільної енергії молекул білків [19].

Проблема виникає на етапі визначення дійсної конформації досліджуваного білка. Тут на допомогу приходить штучний інтелект. Сучасні методи машинного навчання можуть допомогти визначити складні

взаємозв'язки у великих наборах даних, дозволяючи передбачати структури білків. Хіміки та біологи все частіше звертаються до методів ШІ як альтернативи довгому та трудомісткому процесу вивчення складних білків класичними методами.

Методи передбачення структури білків зазвичай поділяються на два основні типи: моделювання на основі шаблонів (ТВМ) та вільне моделювання (FM), залежно від того, чи використовуються шаблони для побудови моделі [20]. Хоча ці методи можуть бути різними, останнім часом деякі методи ТВМ включають енергетично орієнтоване вдосконалення моделей, а частина FM використовує підходи на основі фрагментів, залучаючи дані з бази даних Protein Data Bank (PDB) через машинне навчання.

Останнім часом з'являються нові гібридні методи, що комбінують різні підходи або вдосконалюють відомі методи за допомогою штучного інтелекту, як-от AlphaFold 3.

Важливим аспектом є використання штучного інтелекту (ШІ) та машинного навчання, які значно покращили точність прогнозування. ШІ, зокрема глибоке навчання, активно використовується для моделювання складних білкових структур, дозволяючи значно скоротити час на їх обчислення і покращити передбачення для білків з новими або складними структурами, що не мають шаблонів у базах даних.

Розвиток нейронних мереж та алгоритмів,[21] таких як AlphaFold, дозволяє вирішувати задачі, що були важкодоступними для традиційних фізичних методів моделювання. Вони здатні аналізувати великі обсяги даних та отримувати точні прогнози про білкові структури на основі амінокислотних послідовностей. Це має величезне значення для сучасної хімії та біології, оскільки допомагає в розробці нових лікарських засобів, вакцин і вивченні біологічних процесів.

Використання ШІ також допомагає в оптимізації структури білків, що відкриває нові можливості для біомедичних досліджень і терапевтичних застосувань. Важливо, що ці технології використовуються не лише для

розпізнавання відомих структур, але й для створення нових функціональних білків, що може сприяти розвитку синтетичної біології та біотехнології

1.3. Алгоритми глибокого навчання та AlphaFold

1.3.1. DeepMind та розробка AlphaFold: основи роботи нейронних мереж

Нейронні мережі — це моделі обчислень, що використовують ідеї, натхненні функціонуванням людського мозку. Вони складаються з штучних нейронів, які організовані у шари: вхідний шар, приховані шари та вихідний шар. Основна мета нейронної мережі — навчитися знаходити залежності та закономірності у вхідних даних і використовувати їх для прогнозування чи класифікації.

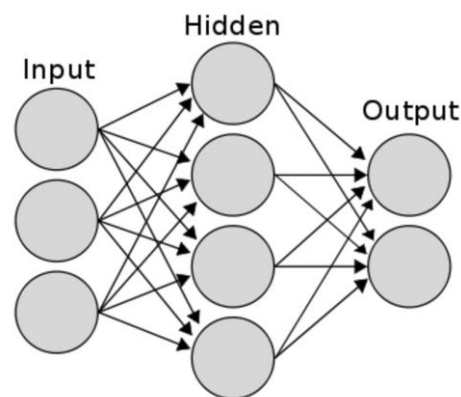


Рис. 1.10 - Базовий принцип роботи нейронних мереж [22]

Кожен нейрон є моделлю лінійної регресії – математичною моделлю, що використовується для прогнозування майбутніх подій. Вхідні сигнали помножуються на вагові коефіцієнти, які в процесі навчання мережі коригуються для досягнення оптимальних результатів. Активаційні функції нейронної мережі визначають, чи буде нейрон активований, і впливають на вихід. Нейронна мережа також навчається, щоб із часом також покращувати точність даних [23].

Згорткові нейронні мережі (Convolutional Neural Networks, CNN) мають більш складну архітектуру і класично використовувалися для різних моделей для прогнозування структури білка як особливо ефективні для аналізу даних із просторовою структурою, таких як зображення та 3D-структури білків.

У контексті передбачення структури білків, CNN можуть аналізувати як просторову організацію амінокислот, так і послідовність білка. Згорткові шари виділяють шаблони взаємодій між амінокислотами, допомагаючи нейронній мережі зрозуміти, як ці шаблони сприяють тривимірній структурі білка. У більш складних моделях CNN можуть бути поєднані з іншими архітектурами, включаючи трансформери, для моделювання взаємодій між частинами білкової послідовності, що дозволяє враховувати як локальні, так і глобальні зв'язки в структурі.

В основі розробленої у 2018 році [24] нейронної мережі AlphaFold лежить новий підхід до машинного навчання, який включає фізичні та біологічні знання про структуру білка. AlphaFold взяла участь у міжнародному конкурсі CASP (Critical Assessment of Structure Prediction) у 2018 році та продемонструвала значні покращення точності у порівнянні з іншими методами на той час, що стало великим проривом у цій сфері.

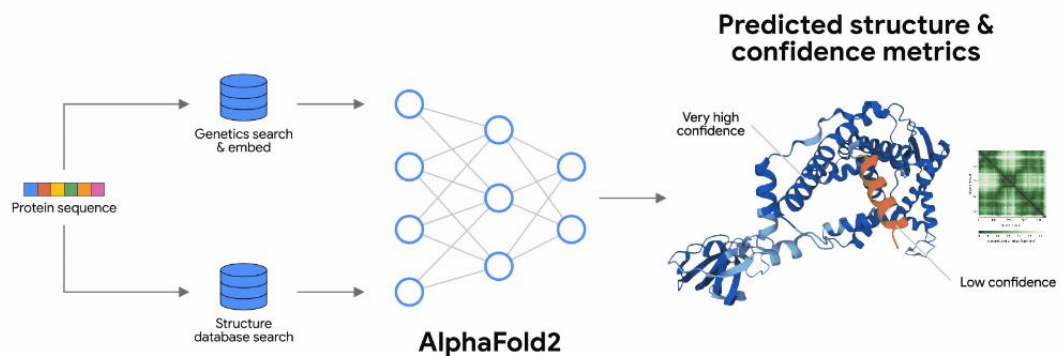


Рис. 1.11 - AlphaFold2, потужний метод, керований штучним інтелектом, зробив революцію в цій галузі, передбачаючи структури білка з надзвичайною точністю [25].

Замість класичних згорткових нейронних мереж (CNN), як це було в багатьох попередніх моделях, AlphaFold застосовує інші, більш спеціалізовані підходи, такі як MSA (multiple sequence alignment) та ключовий принцип будівельного блоку мережі під назвою Evoformer, щоб забезпечити точне передбачення структури білка. Оскільки мережа AlphaFold містить кілька рівнів вузлів штучних нейронів, її називають алгоритмом «глибокого навчання».

MSA — це вирівнювання кількох послідовностей амінокислот, що дає можливість зрозуміти еволюційні зв'язки між ними. Ідея полягає в тому, що білки, які мають схожі послідовності, часто мають схожі тривимірні структури. MSA служить важливим джерелом інформації про взаємозв'язки між амінокислотними залишками і допомагає прогнозувати, які частини білка стабільні або функціонально важливі. Evoformer — це компонент, який поєднує різні механізми, щоб враховувати зв'язки всередині послідовності та між послідовностями, і блоки зворотного зв'язку, щоб підсилити навчання моделі. Замість аналізу простих просторових патернів, як це було у CNN, MSA та Evoformer враховують специфічні еволюційні та просторові особливості амінокислотних послідовностей. Використання таких технологій дозволяє AlphaFold створювати точніші моделі тривимірних структур білків завдяки інтеграції еволюційних даних і глибоких взаємозв'язків між амінокислотами.

Архітектура AlphaFold здатна навчатися на даних з бази PDB та підвищувати точність даних за допомогою методу самодистиляції з шумом, який включає створення нових наборів даних із передбачуваних структур і повторне навчання з використанням як початкових, так і цих додаткових даних.

AlphaFold2 бере амінокислотну послідовність нового білка та вирівнює її з послідовністю інших подібних білків. Це ідентифікує ділянки послідовності, які мають тенденцію змінюватися разом у еволюційному часі, і тому, ймовірно, взаємодіють і, таким чином, фізично близькі в 3D-структурі білка.

Спочатку цільова форма передається в генеративну систему штучного інтелекту, так звану RFdiffusion, яка створює основи білкової структури, що приймають форму цільової структури. Як тільки алгоритм дифузії створює трьохвимірну модель структури, інша частина програмного забезпечення виявляє потенціальні амінокислотні послідовності. Ці послідовності передаються до ШІ для прогнозування, щоб виявити, які кандидати найімовірніше звернуться так, як задумано.

Протягом кількох хвилин (або десятків хвилин для більших білків або комплексів) AlphaFold2 представляє прогноз 3D-структури послідовності [26].

1.3.2. Технологічні принципи AlphaFold 3: покращення точності моделювання в порівнянні з попередніми версіями

AlphaFold Server — це веб-сервіс,[27] який може створювати високоточні прогнози біомолекулярної структури, що містять білки, ДНК, РНК, ліганди, іони, а також моделювати хімічні модифікації білків і нуклеїнових кислот на одній платформі. Він оснащений новітньою моделлю AlphaFold 3. Модель AlphaFold 3 є результатом співпраці Google DeepMind і Isomorphic Labs.

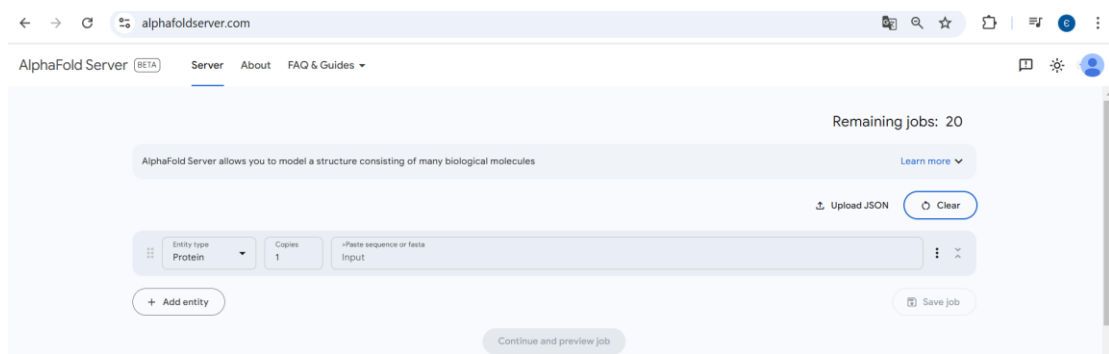


Рис. 1.12 - Сервер AlphaFold 3

На відміну від AlphaFold2, що передбачає лише структуру білків і білково-білкових комплексів, AlphaFold3 може генерувати прогнози, що

містять білки, ДНК, РНК, іони, ліганди та хімічні модифікації. Нова модель також покращує точність моделювання білкового комплексу.

Загальна структура AlphaFold 3 [28] повторює структуру AlphaFold 2, але є великі відмінності в кожному основному компоненті.

Обробка MSA спрощена, а Evoformer замінено на «парний формувач» для обробки парного представлення. Крім того, тепер використовується «дифузійна мережа», що дозволяє більш ефективно обробляти координати атомів без складних систем обертання. Цей модуль навчається за допомогою підходу шумозаглушення на основі шуму, зменшуючи залежність від стереохімічних затруднень і представлень на основі крутильних кутових коливань. Авторами його описано, як подібний до інструментів, які використовуються для створення зображень ШІ [29]. Процес починається із хмари атомів, і через певні кроки призводить до кінцевої структури з високою точністю.

Для взаємодії білків з іншими типами молекул AlphaFold 3 показує на 50% покращення, порівняно з існуючими методами прогнозування, а для деяких важливих категорій взаємодії точність прогнозування навіть подвоєно. [30]

Окрім методу навчання нейронної мережі [31] знаходити нові фрагменти білка, використовувався також математичний метод градієнтного спуску [32] для створення високоточних структур.

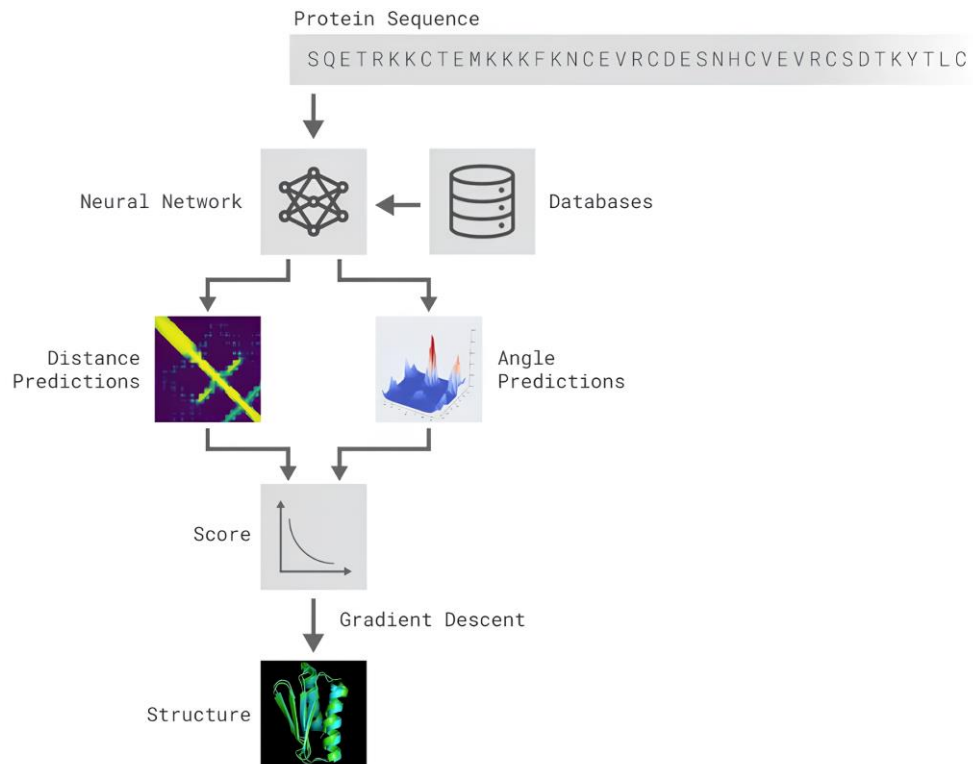


Рис. 1.13 - Градієнтний спуск — математична техніка, яка зазвичай використовується в машинному навчанні для внесення невеликих поступових покращень [33].

AlphaFold 3 встановлює нову планку,[34] перевершуючи найкращі відомі методи докінгу, не вимагаючи референтної структури білка або розташування ліганду, і це дозволяє прогнозувати нові, не охарактеризовані раніше структурно білки.

Нова версія також надає можливість аналізувати гнучкість, притаманну білкам і нуклеїновим кислотам при взаємодії з іншими молекулами - те, що неможливо при використанні методів докінгу.

Серед недоліків AlphaFold 3 можна виділити декілька обмежень. Вихідні дані моделі не завжди враховують хіральність, а також наявна схильність моделі час від часу створювати в прогнозах атоми, що перекриваються (зіштовхуються). У деяких випадках змодельований конформаційний стан може бути неправильним або неповним.

Незважаючи на значний прогрес у точності моделювання в AlphaFold 3, все ще залишається багато викликів та необхідність покращення моделі.

1.3.3. Роль AlphaFold 3 у вирішенні проблеми передбачення структури білків

В кожній клітині кожної живої істоти є велика кількість біоактивних структур, які поєднуються або взаємодіють між собою. Ці взаємодії мають вирішальне значення для розуміння біологічних процесів і просування зусиль для розробки нових ліків.[35] В цьому на допомогу приходить AlphaFold 3, виходячи навіть за рамки дослідження білків, охоплюючи велику кількість біомолекул. В області пошуку нових лікарських засобів це відкриває нові можливості, точно передбачаючи взаємодію молекул, подібних до ліків, таких як ліганди та антитіла, які зв'язуються з білками, впливаючи на їхню роль у здоров'ї та хворобах. Завдяки підвищеній точності прогнозування навіть складних молекулярних збірок, AlphaFold 3 полегшує застосування нейронних мереж в більш широкому спектрі біомедичних досліджень, включаючи лікування раку і розробку вакцин [36].

Передбачення зв'язування білок-антитіло з такою високою точністю, як пропонує AlphaFold 3, має ключове значення для розуміння імунної відповіді та дизайну нових ліків. вчені шукають молекули, які можуть приєднуватися до хвороботворних білків і змінювати їхню функціональність або сприятимуть їх здоровому функціонуванню [37].

Опубліковані дослідження з використанням AlphaFold 3 (Рис.1.14),[38] доводять високий рівень точності нейронної мережі.

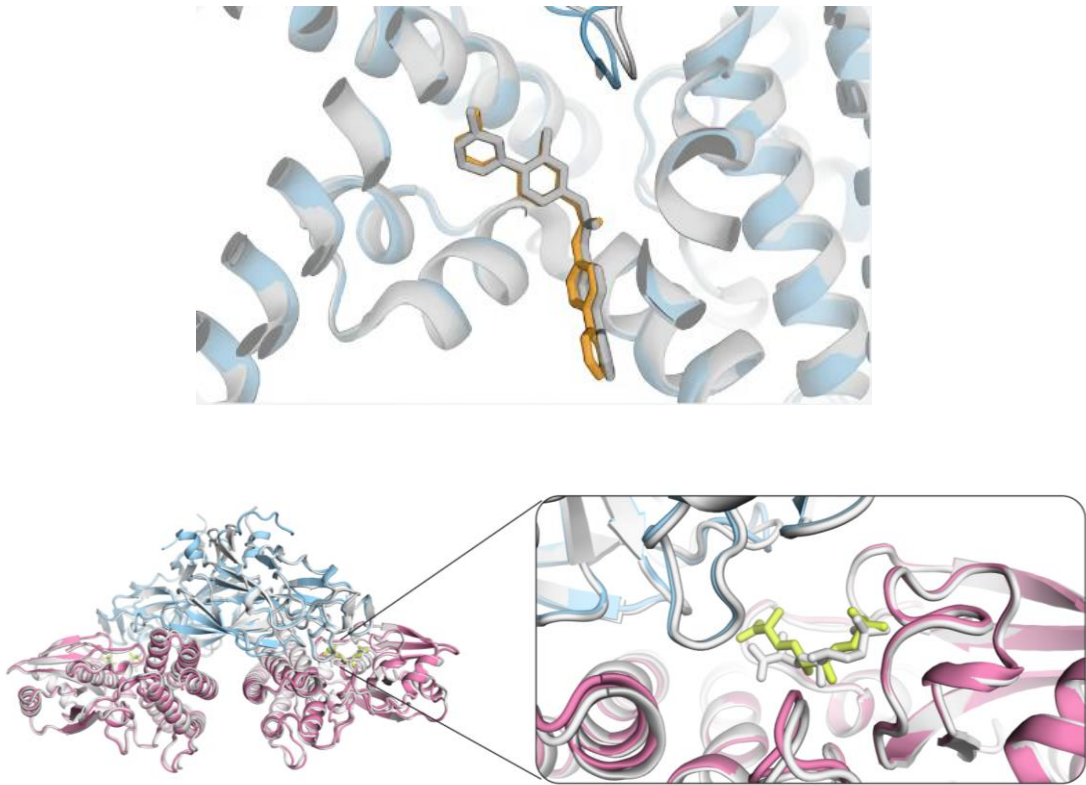


Рис.1.14 - а) PORCN : протиракова молекула клінічної стадії, пов'язана зі своєю мішенню разом з іншим білком [39]. б) Білки з унікальними складками: AziU3/U2 білковий комплекс каталізує утворення кілець азиридину на новому каталітичному місці, утвореному комплексом з новою складкою [40].

2. Результати та обговорення

2.1 Об'єкти дослідження

Для дослідження було обрано наступні білки різних класів.

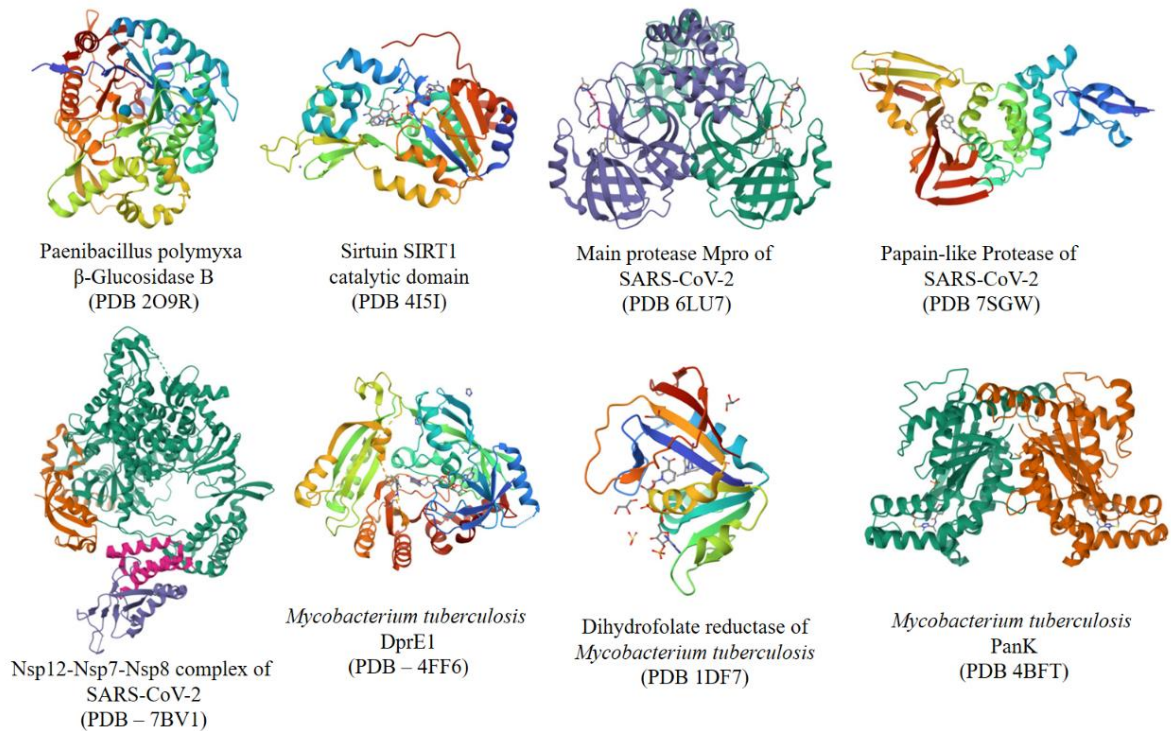


Рис. 2.1. - Рентгенівська структура досліджених білків. У дужках наведено відповідні PDB-коди.

Для прогнозування структур білків було використано метод AlphaFold 3, який продемонстрував високу точність у передбаченні третинної структури, з оцінкою pLDDT у переважній більшості ділянок понад 90, що підтверджує надійність отриманої моделі.

2.1.1 β-Глюкозидаза (*Paenibacillus polymyxa* β-Glucosidase B, PDB 2O9R)

Білок є структурою BglB, ферменту, що виробляється бактерією *Paenibacillus polymyxa*. Цей фермент є еко-бета-глюкозидазою, що розщеплює целобіозу та целодекстрини в бактеріях, що беруть участь у деградації

целюлозних субстратів. Структура VglV демонструє, як фермент розпізнає й гідролізує молекули, що складаються з глюкози, завдяки специфічним взаємодіям в агліконовому сайті [41-43].

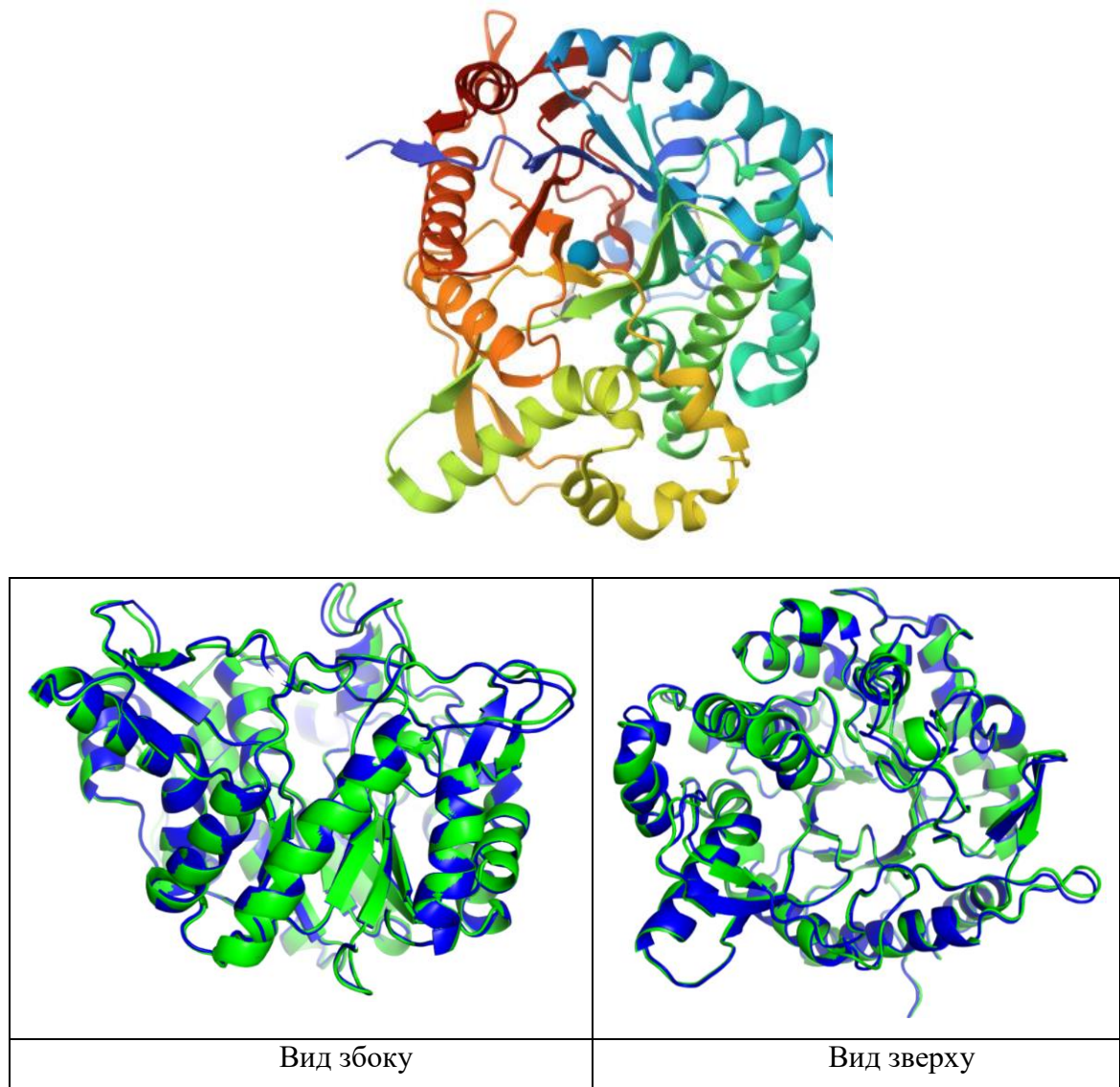


Рисунок 2.2. - Накладання нативної рентгеноструктури (PDB 2O9R, синій) β -глюкозидази *B. Paenibacillus polymyxa* та передбаченої AF3-структури (зелений), вид збоку (а) та зверху (б), відповідно.

Основні елементи вторинної структури, такі як альфа-спіралі та бета-листи, відповідають у більшості областей. Можна помітити невеликі розходження в окремих петлях чи кінцевих ділянках, що є типовим для порівняння експериментальних і теоретичних структур (рис. 2.2). Рівень

збіжності – високий. (PyMOL RMSD=0.371 nm; Discover Studio RSMD(main chain)=0.736)

2.1.2 Sirtuin SIRT1 catalytic domain, PDB – 4I5I

NAD(+)-залежна деацетилаза, що бере участь у регуляції різних клітинних процесів, включаючи старіння, реакцію на стрес та метаболізм.

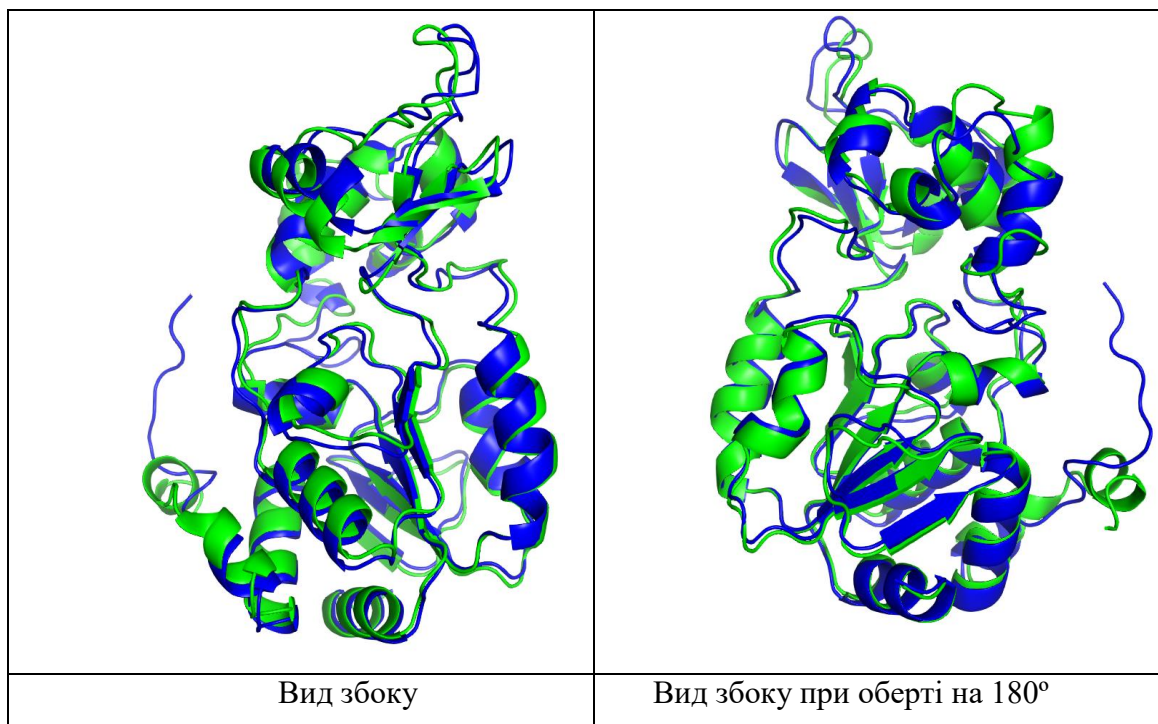
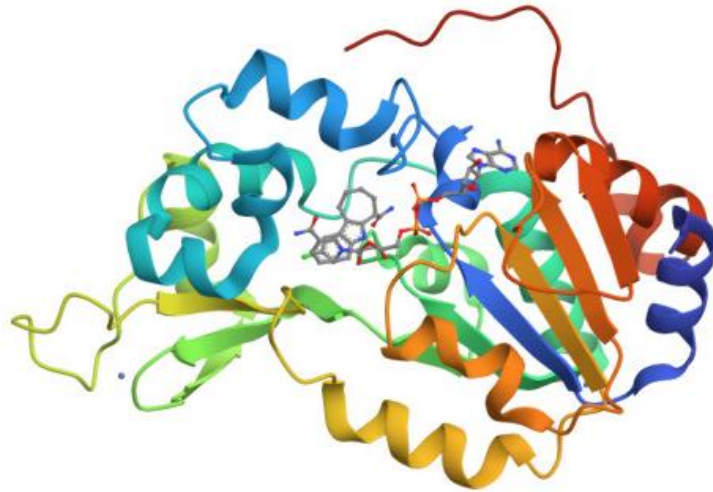


Рисунок 2.3. - Накладання нативної рентгенівської структури (PDB 4I5I, синій) каталітичного домену сиртуїну SIRT1 та AF3-передбаченої структури (зелений) у боковому (а) та поверненому на 180° (б) ракурсах, відповідно.

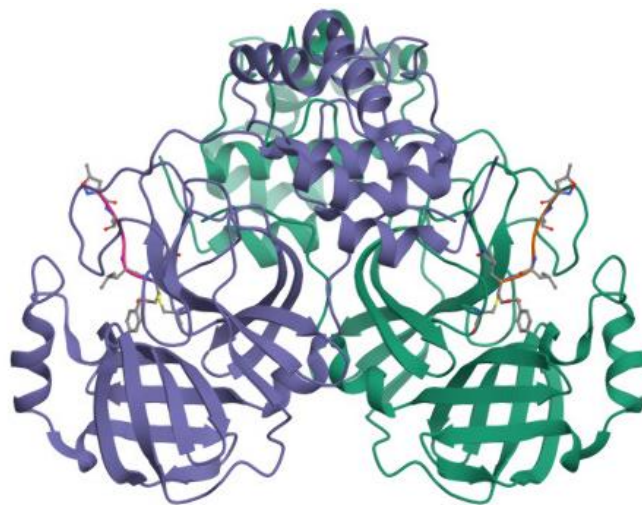
Структура розкриває деталі того, як SIRT1 взаємодіє зі своїми субстратами та інгібіторами, забезпечуючи розуміння, корисне для розробки ліків, спрямованих на метаболічні та вікові захворювання [44-46].

Результати розрахунку показали непоганий збіг для основних елементів (рис. 2.3). Локальні відмінності є переважно в петлях і гнучких сегментах. Це очікувано, оскільки ці регіони часто важче прогнозувати. Рівень точності – низький. (PyMOL RMSD=1.279 nm; Discover Studio RSMD(main chain)=3.301).

2.1.3 Main protease Mpro of SARS-CoV-2, PDB – 6LU7

Цей білок є ключовим ферментом вірусу SARS-CoV-2 (збудника COVID-19). Це головна протеаза розщеплює поліпротеїновий ланцюг, утворений вірусом, на функціональні білки, необхідні для його життєвого циклу. Структура 6LU7 була використана для розробки інгібіторів протеази, що можуть стати основою для ліків проти COVID-19 [47-49].

Структура змодельованого білка добре узгоджується з експериментальною. Бета-листи й альфа-спіралі вирівняні (рис. 2.4). Невеликі відмінності спостерігаються в периферійних сегментах білка, особливо в петлях. Середній рівень точності. (PyMOL RMSD=0.411 nm; Discover Studio RSMD(main chain)=2.796).



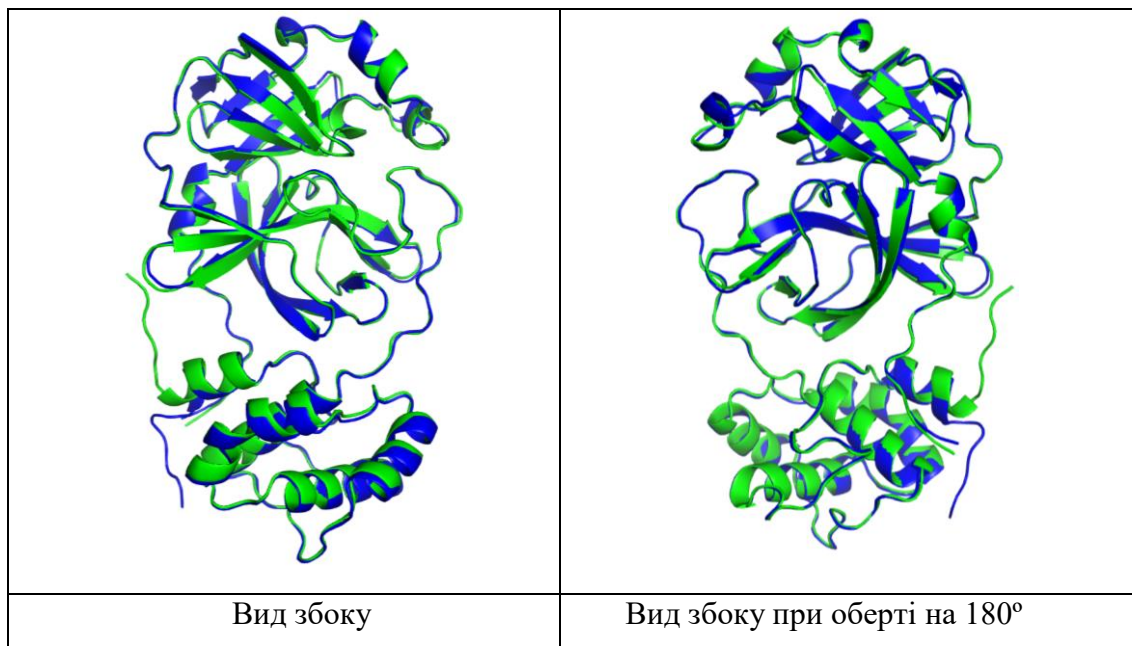


Рис 2.4 - Накладання нативної рентгенівської структури (PDB 6LU7, синій) протеази Mpro вірусу SARS-CoV-2 та передбаченої AF3-структури (зелений) у боковому (а) та поверненому на 180° (б) ракурсах, відповідно.

2.1.4 Папаїно-подібна протеаза вірусу SARS CoV-2, PDB – 7SGW

Субодиниця білка Spike SARS-CoV-2, варіант Omicron. Належить до поверхневого спайкового (S) глікопротеїну вірусу, який відповідає за прикріплення до рецепторів клітин людини (ACE2) і сприяє проникненню вірусу в клітину. Структура належить до варіанта Omicron, який має багато мутацій у спайковому білку. Ці мутації впливають на здатність вірусу уникати імунної відповіді та на підвищену інфекційність [50].

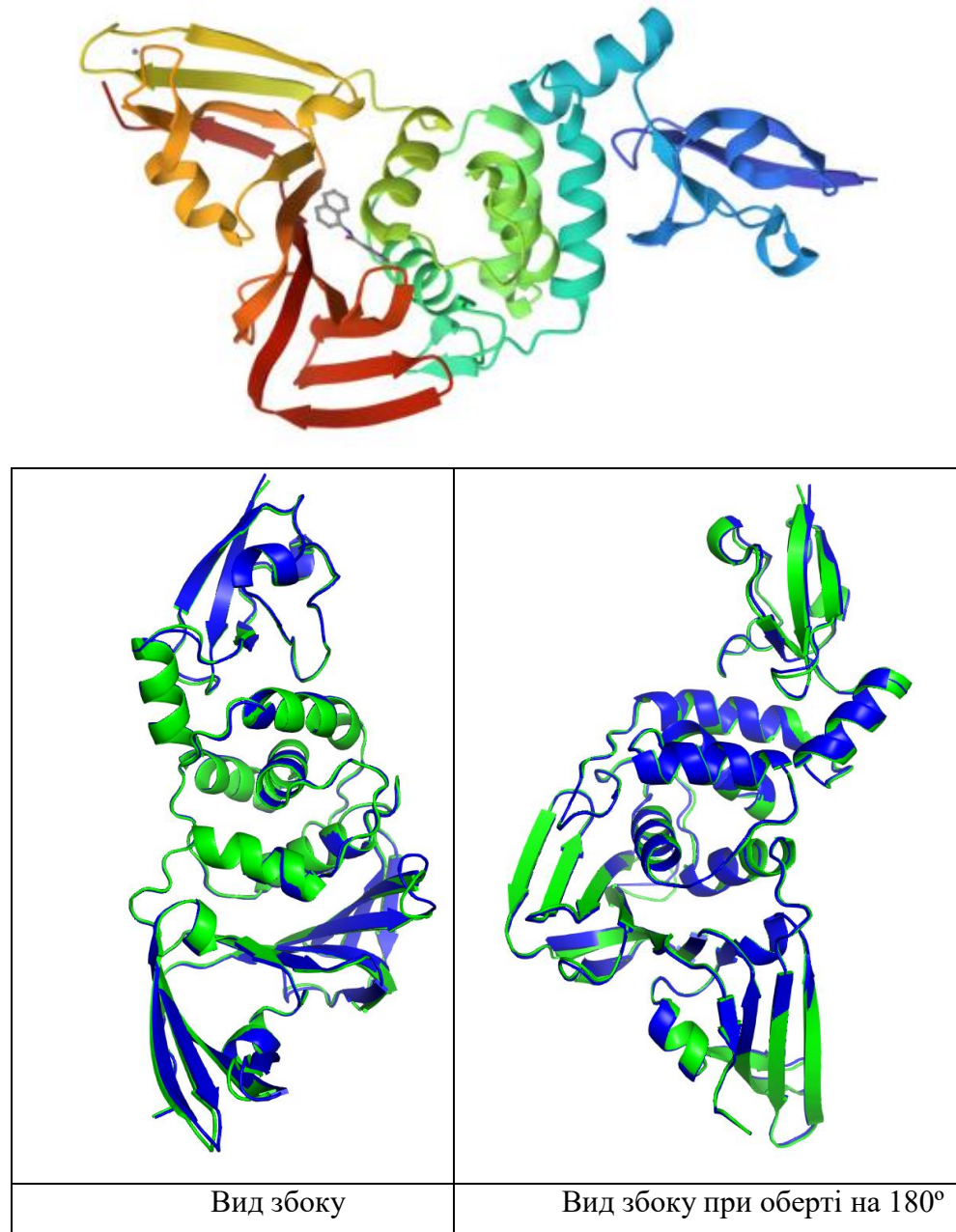


Рис 2.5 - Папаїно-подібна протеаза вірусу SARS CoV-2

Альфа-спіралі та бета-листки майже чітко збігаються з базою даних (рис. 2.5). Локальні зміщення в області бічних ланцюгів або невеликі деформації у петлях. Загалом збіжність сильна, рівень збіжності – високий. (PyMOL RMSD=0.456 nm; Discover Studio RSMD(main chain)=0.522)

2.1.5 Nsp12-Nsp7-Nsp8 complex of SARS-CoV-2, PDB – 7BV1

Моноклональне антитіло до SARS-CoV-2 у комплексі зі спайковим білком (рис. 2.6). Це структура комплексу антитіла (CB6) з рецептор-зв'язувальним доменом (RBD) спайкового білка SARS-CoV-2. Антитіло нейтралізує вірус, блокуючи його зв'язування з рецептором ACE2 [51-52].

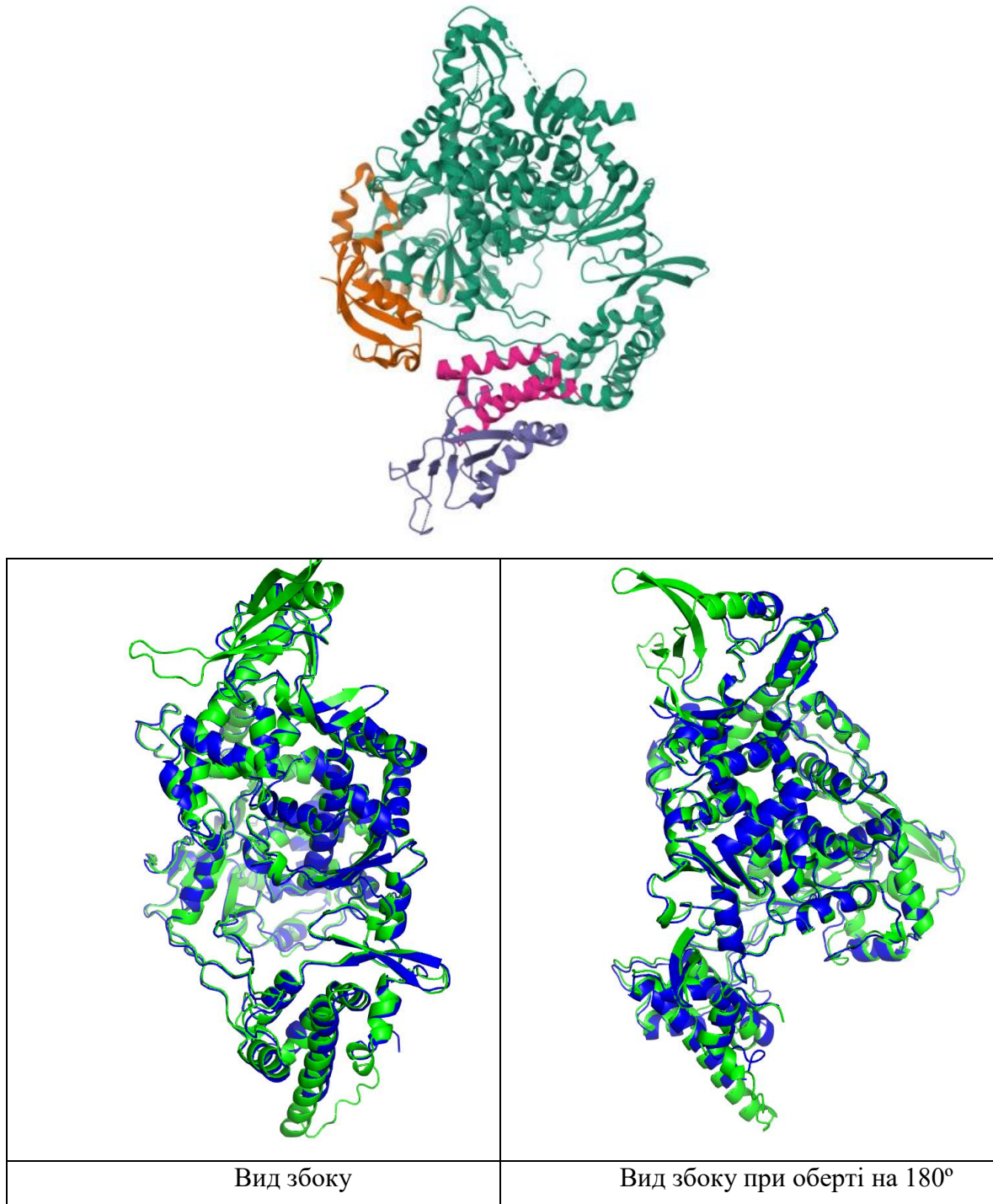


Рис 2.6 - Nsp12-Nsp7-Nsp8 complex of SARS-CoV-2

Просторове розташування основних доменів збігається, хоча можливі розбіжності в кінцевих ділянках або петлях, які є менш структурованими. Середній рівень точності. (PyMOL RMSD=0.614 nm; Discover Studio RSMD(main chain)=1.303)

2.1.6 *Mycobacterium tuberculosis* DprE1, PDB – 4FF6

Цей фермент (InhA) є критично важливою мішенню для розробки ліків проти туберкульозу, оскільки він відіграє ключову роль у синтезі клітинної стінки мікобактерій, каталізуючи редукцію подвійного зв'язку в ненасичених жирних кислотах в процесі їх синтезу [53-54].

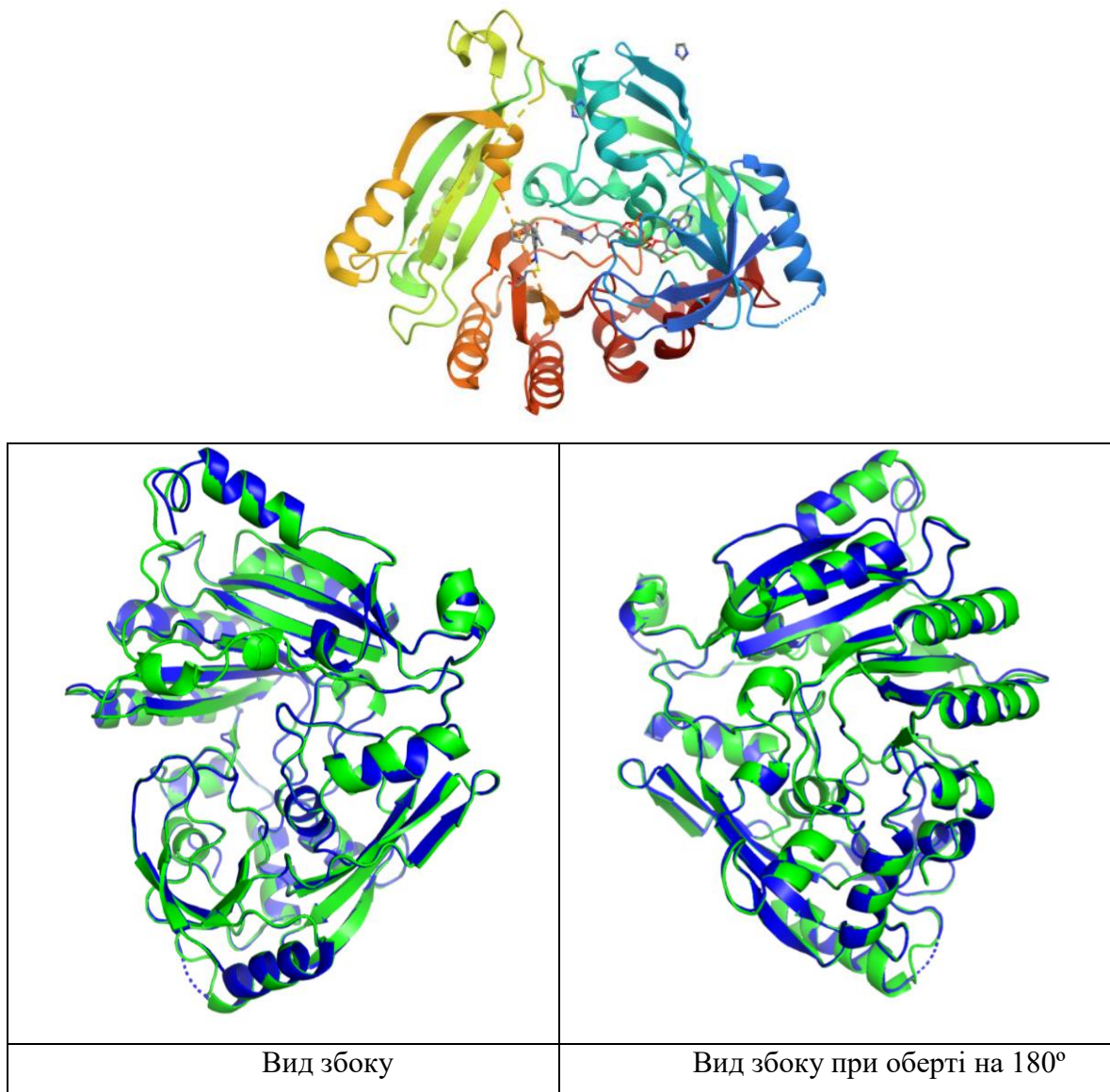


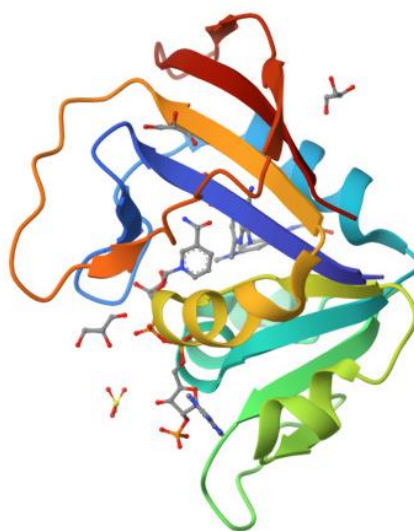
Рис 2.7. - *Mycobacterium tuberculosis* DprE1

Високий рівень точності для глобальної структури, однак наявні розбіжності в основному ланцюгу, що свідчить про неточність у динамічних областях (рис. 2.7). Рівень точності – середній. (PyMOL RMSD=0.236 nm; Discover Studio RSMD(main chain)=1.355)

2.1.7 Dihydrofolate reductase of *Mycobacterium tuberculosis*, PDB – 1DF7

Бета-лактамази - це ферменти, які гідролізують бета-лактамі антибіотики, такі як пеніциліни та цефалоспорини, роблячи їх неактивними. Мікобактерії туберкульозу можуть виробляти ці ферменти як частину механізмів резистентності до антибіотиків, які зазвичай застосовуються. Структура бета-лактамази (1DF7) важлива для розробки інгібіторів, які могли б протидіяти цій резистентності і зробити бета-лактамі антибіотики знову ефективними проти туберкульозу [55-57].

Надзвичайно низькі значення RMSD вказують на високу точність передбачення AF3, передбачення майже співпадає з експериментальними даними (рис. 2.8). Один з найкращих результатів. (PyMOL RMSD=0.408 nm; Discover Studio RSMD(main chain)=0.486)



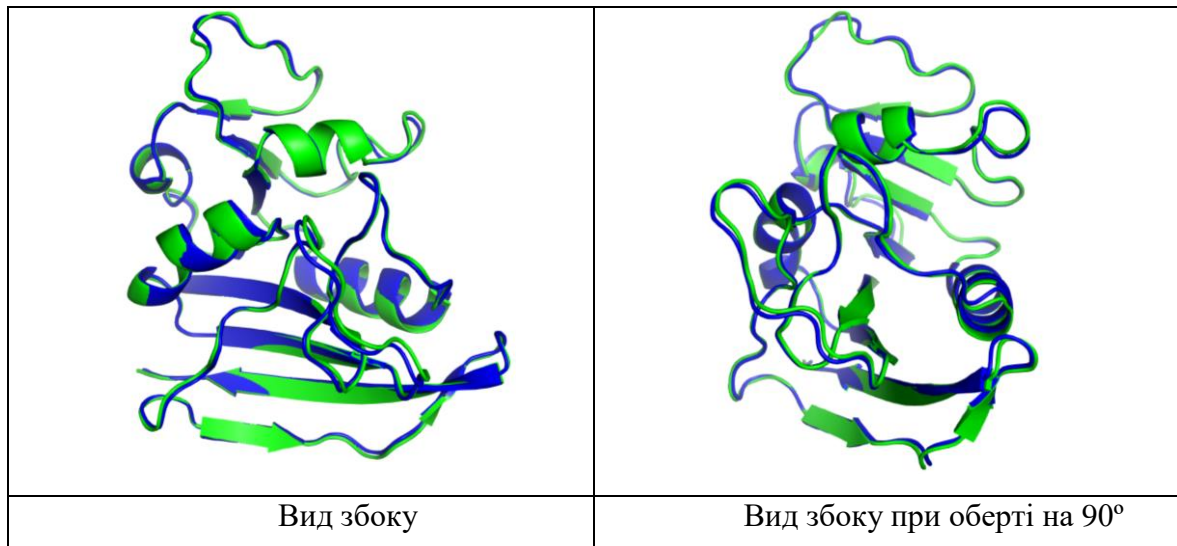
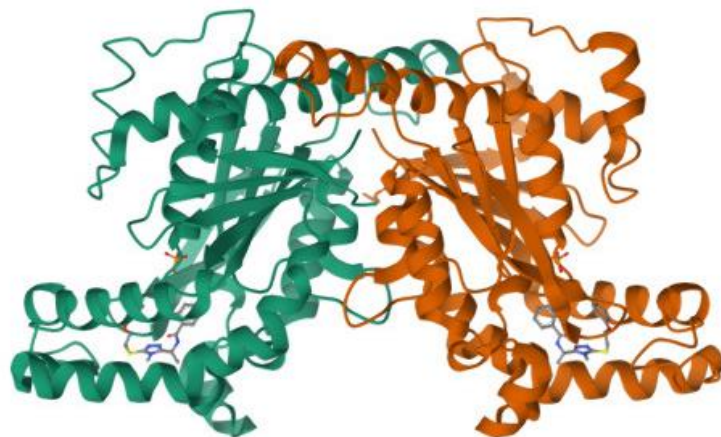


Рис 2.8 - Накладання нативної рентгенівської структури (PDB 1DF7, синій) дигідрофолатредуктази *Mycobacterium tuberculosis* та передбаченої AF3-структури (зелений) у боковому (а) та поверненому на 180° (б) зображенні, відповідно.

2.1.8 *Mycobacterium tuberculosis* PanK, PDB – 4BFT

Пантотенаткіназа (PanK) є важливим ферментом у біосинтезі коензиму А в мікобактерії туберкульозу. Він каталізує фосфорилування пантотенату (вітаміну В5), що є першим етапом біосинтезу КоА. Кристалічна структура демонструє її комплекс з інгібіторами, які націлені на її активну ділянку, щоб порушити біосинтетичний шлях коензиму А, пропонуючи перспективний шлях для розробки протитуберкульозних препаратів [58-60].



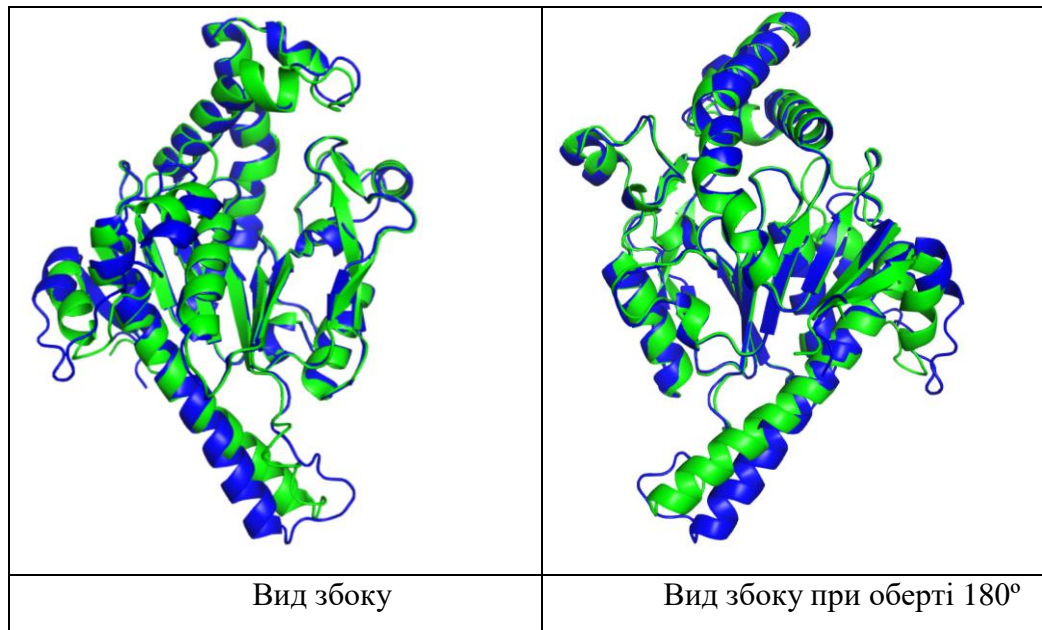


Рис 2.9 - Накладання нативної рентгенівської структури *Mycobacterium tuberculosis* PanK (PDB – 4BFT) та передбаченої AF3-структури (зелений) у боковому (а) та поверненому на 180° (б) зображенні, відповідно

Невелике значення відхилення вказує на гарний збіг, хоча в нестабільних ділянках деякі відмінності (рис. 2.9). Середній рівень точності. (PyMOL RMSD=0.629 nm; Discover Studio RSMD(main chain)=2.560).

2.1.9 SARS-CoV-2 Omicron Spike-protein-open-2, PDB – 7WVO

Описана структура білка-шипа SARS-CoV-2 є ключовою для розуміння механізму взаємодії вірусу з клітинними рецепторами людини, зокрема з рецептором ACE2, а також для розробки вакцин і терапевтичних засобів. Omicron-варіант має велику кількість мутацій у цьому білку (понад 30 змін), що впливають на його поведінку, включно зі зв'язуванням рецептора та імунною відповіддю [61-63].

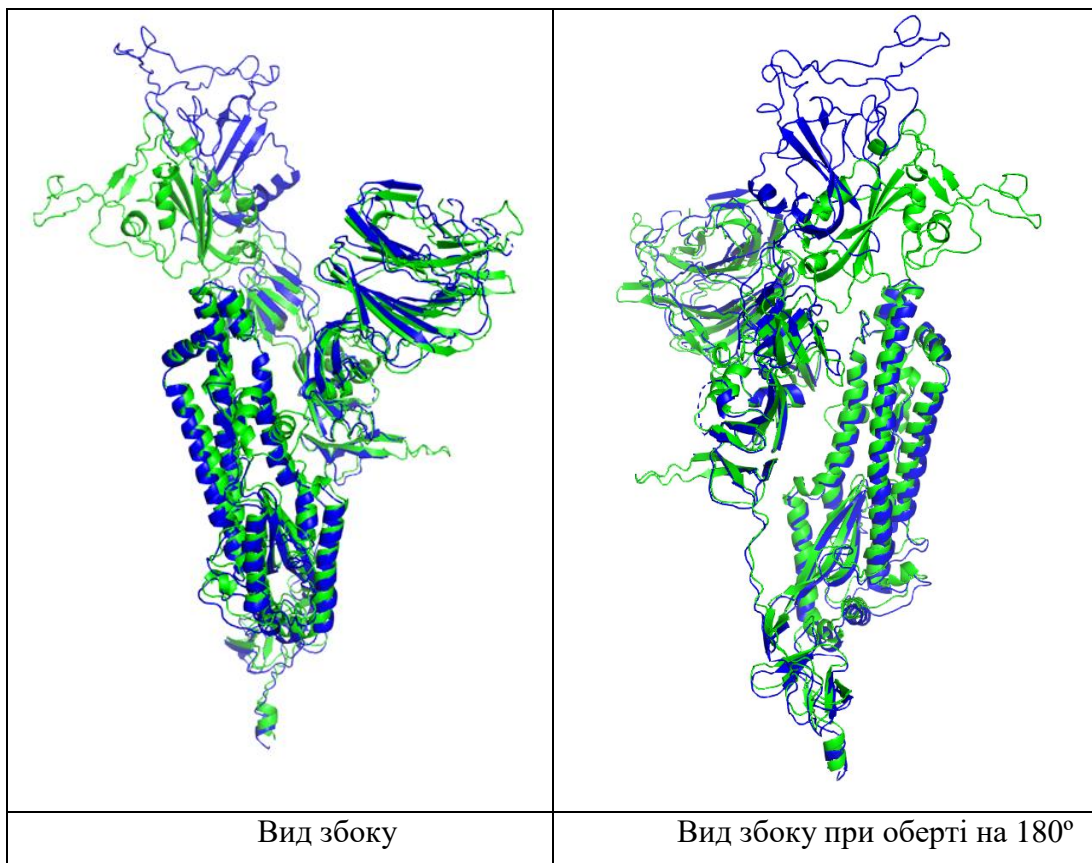


Рис 2.10 Накладання нативної рентгенівської структури Спайк-протеїн коронавірусу SARS-CoV-2 (PDB – 7WVO) та передбаченої AF3-структури (зелений) у боковому (а) та поверненому на 180° (б) зображенні, відповідно.

Доволі високі значення відхилень свідчать про труднощі моделювання складних структур з великою кількістю динамічних регіонів (рис. 2.10). Один із найгірших результатів. (PyMOL RMSD=2.046 nm; Discover Studio RSMD(main chain)=9.488).

3. Висновки

Нейронна мережа AlphaFold 3 стала проривом у прогнозування третинної будови білка на молекулярному рівні. Точність таких прогнозів відіграє ключову роль у розробці нових ліків, зокрема в контексті інгібіторів, що зв'язуються з активним центром протеази. Окрім того, отримані структури можуть значно покращити розуміння механізмів антибіотикорезистентності, що відкриває нові горизонти в боротьбі з інфекціями.

З використанням нейронної мережи AlphaFold 3 було досліджено тривимірну будову білків різних класів та встановлено наступне:

- використання AlphaFold 3 для прогнозування структури білка протеази SARS-CoV-2 дозволило отримати модель з високою точністю (pLDDT > 90), що підтверджує надійність цього методу для дослідження нових вірусів та мутацій. Прогнозовані структури були наближені до оригінальних з невеликими відхиленнями, що підтверджується результатами порівняння з раніше відомими структурами для малих молекул.
- у випадку великих та більш складних молекул, таких як Спайк-протеїн коронавірусу SARS-CoV-2, похибка прогнозу зростала, що вказує на потребу в подальших вдосконаленнях моделі, зокрема для складних багатоконпонентних систем.
- в цілому AlphaFold продемонстрував свій потенціал як потужний інструмент для розвитку терапевтичних стратегій, підвищуючи ефективність пошуку ліків та покращення результатів досліджень у біології та фармацевтиці. Однак є потенціал для подальшого вдосконалення методу, що дозволить ще точніше прогнозувати структури для більш складних молекул.



СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Voet, Donald, and Judith G. Voet. *Biochemistry*. 4th ed., Wiley, 2011.
2. Brändén, Carl-Ivar, and John Tooze. *Introduction to Protein Structure*. 1999.
3. Anfinsen, C.B. "Principles That Govern the Folding of Protein Chains." *Science*, vol. 181, no. 4096, 20 July 1973, pp. 223-30. [10.1126/science.181.4096.223](https://doi.org/10.1126/science.181.4096.223)
4. Bryngelson, Joseph D., José Nelson Onuchic, Nicholas D. Socci, and Peter G. Wolynes. "Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis." *Science*, vol. 21, no. 3, 1995, pp. 167-195.
5. Edman, P. "Method for Sequencing the Amino Acid Sequence of Peptides." *Biochemical Journal*, vol. 66, no. 3, 1957, pp. 1-10.
6. Fenselau, C. MALDI MS and Strategies for Protein Analysis. *Analytical Chemistry*, vol. 69, no. 21, 2011, pp. 661A–665A. DOI: [10.1021/ac971831z](https://doi.org/10.1021/ac971831z)
7. Hjernø, Karin, and Ole N. Jensen. *MALDI MS: A Practical Guide to Instrumentation, Methods and Applications*. 2007. DOI:10.1002/9783527610464
8. Siqueira, Walter L., et al. "Identification of Protein Components in in vivo Human Acquired Enamel Pellicle Using LC–ESI–MS/MS." *Journal of Proteome Research*, vol. 6, no. 7, 2007, pp. 2769-2776.
9. Lin, Ying, et al. "The Current State of Proteomics in GI Oncology." *Digestive Diseases and Sciences*, vol. 54, 2009, pp. 431-457. <https://doi.org/10.1007/s10620-008-0656-5>.
10. Mishra, Sushil, et al. "In Silico Engineering of Proteins That Recognize Small Molecules." *IntechOpen*, 2012, <https://doi.org/10.5772/28001>.
11. Opella, Stanley J., and Francesca M. Marassi. "Applications of NMR to Membrane Proteins." *Archives of Biochemistry and Biophysics*, vol. 628, 2017, pp. 1-10.
12. <https://www.nobelprize.org/prizes/chemistry/2017/press-release/>
13. Parmar, Mayuriben, et al. "Using a SMALP Platform to Determine a Sub-nm Single Particle Cryo-EM Membrane Protein Structure." *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1860, no. 2, 2018, pp. 378-383.
14. Biasini, Marco. *Accurate Modeling of Protein Structures by Homology*. Inauguraldissertation, 2013.

15. Rangwala, Huzefa, and George Karypis, editors. *Introduction to Protein Structure Prediction: Methods and Algorithms*. Wiley, 2010. DOI: 10.1002/9780470882207.
16. Lee, Jooyoung. "Ab Initio Protein Structure Prediction." *From Protein Structure to Function with Bioinformatics*, Springer, 2009, pp. 3-25. DOI: 10.1007/978-1-4020-9058-5_1.
17. Floudas, C.A., et al. "Advances in Protein Structure Prediction and De Novo Protein Design: A Review." *Chemical Engineering Science*, vol. 61, no. 3, 2006, pp. 966-988. DOI: 10.1016/j.ces.2005.04.009.
18. Valavanidis, Athanasios. "AlphaFold Protein Structure Database Predicted Millions of 3D Structures. Can AlphaFold Revolutionize the Discovery of New Drugs?" *Nature*, Oct. 2023, DOI: 10.1038/d41586-023-02982-w.
19. Одинець, К. О., & Корнелюк, О. І. "Методи аналізу та моделювання просторової структури білків." 2001.
20. Bertoline, Leticia M. F., Angélica N. Lima, Jose E. Krieger, and Samantha K. Teixeira. "Protein Bioinformatics Before and After AlphaFold2: An Overview of Protein Structure Prediction." *Frontiers in Bioinformatics*, vol. 3, 28 Feb. 2023, <https://doi.org/10.3389/fbinf.2023.1120370>.
21. Kuhlman, Brian, and Philip Bradley. "Advances in Protein Structure Prediction and Design." *Nature Reviews Molecular Cell Biology*, vol. 20, 2019, pp. 681–697. <https://doi.org/10.1038/s41580-019-0152-5>.
22. Graff, Philip, Farhan Feroz, Michael Hobson, and Anthony Lasenby. "SKYNET: An Efficient and Robust Neural Network Training Tool for Machine Learning in Astronomy." *Monthly Notices of the Royal Astronomical Society*, vol. 441, 2013, doi: 10.1093/mnras/stu642.
23. Jordan, M. I., and T. M. Mitchell. "Machine Learning: Trends, Perspectives, and Prospects." *Science*, vol. 349, no. 6245, 17 July 2015, pp. 255–260. <https://doi.org/10.1126/science.aaa8415>.
24. Jumper, John, Richard Evans, Alexander Pritzel, et al. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature*, vol. 596, 2021, pp. 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
25. <https://www.youtube.com/watch?v=7q8Uw3rmXyE>

26. <https://www.ebi.ac.uk/training/online/courses/alphafold/an-introductory-guide-to-its-strengths-and-limitations/strengths-and-limitations-of-alphafold/>
27. <https://alphafoldserver.com/about>
28. Abramson, J., J. Adler, J. Dunger, et al. "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3." *Nature*, vol. 630, 2024, pp. 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
29. <https://frontlinegenomics.com/alphafold-3-stepping-into-the-future-of-structure-prediction/>
30. <https://blog.google/technology/ai/google-deepmind-isomorphic-alphafold-3-ai-model>
31. <https://deepmind.google/discover/blog/alphafold-using-ai-for-scientific-discovery/>
32. Nielsen, Michael. *Neural Networks and Deep Learning*. Dec. 2019, <http://neuralnetworksanddeeplearning.com/>.
33. <https://deepmind.google/discover/blog/alphafold-using-ai-for-scientific-discovery/>
34. <https://deepmind.google/discover/blog/a-glimpse-of-the-next-generation-of-alphafold>
35. <https://frontlinegenomics.com/alphafold-3-stepping-into-the-future-of-structure-prediction/>
36. <https://blog.google/technology/ai/how-we-built-alphafold-3/>
37. Valavanidis, Athanasios. "AlphaFold Protein Structure Database Predicted Millions of 3D Structures. Can AlphaFold Revolutionize the Discovery of New Drugs?" *Nature*, Oct. 2023. <https://doi.org/10.1038/d41586-023-02982-w>.
38. Google DeepMind AlphaFold Team and Isomorphic Labs Team. "Performance and Structural Coverage of the Latest, In-Development AlphaFold Model." 2023.
39. Liu, Y., X. Qi, L. Donnelly, et al. "Mechanisms and Inhibition of Porcupine-Mediated Wnt Acylation." *Nature*, vol. 607, 2022, pp. 816–822. <https://doi.org/10.1038/s41586-022-04952-2>.
40. <https://www.isomorphiclabs.com/articles/rational-drug-design-with-alphafold-3>
41. Isorna, P., J. Polaina, L. Latorre-García, F. J. Cañada, B. González, and J. Sanz-Aparicio. "Crystal Structures of *Paenibacillus polymyxa* β -Glucosidase B Complexes Reveal the Molecular Basis of Substrate Specificity and Give New Insights into the Catalytic Machinery of Family I Glycosidases." *Journal of Molecular Biology*, vol. 371, no. 5, 2007, pp. 1204–1218. <https://doi.org/10.1016/j.jmb.2007.05.082>.

42. Chepeleva, L. V., D. O. Tarasenko, A. Y. Chumak, O. O. Demidov, A. D. Snizhko, O. O. Kolomoitsev, V. M. Kotliar, E. S. Gladkov, A. L. Tatarets, A. V. Kyrychenko, and A. D. Roshal. "4'-Benzyloxyflavonol Glucoside as Fluorescent Indicator for β -Glucosidase Activity." *Functional Materials*, vol. 30, no. 4, 2023, pp. 494–505. doi: 10.15407/fm30.04.494.
43. Chepeleva, L. V., O. O. Demidov, A. D. Snizhko, D. O. Tarasenko, A. Y. Chumak, O. O. Kolomoitsev, V. M. Kotliar, E. S. Gladkov, A. Kyrychenko, and A. D. Roshal. "Binding Interactions of Hydrophobically-Modified Flavonols with β -Glucosidase: Fluorescence Spectroscopy and Molecular Modelling Study." *RSC Advances*, vol. 13, no. 48, 2023, pp. 34107–34121. doi: 10.1039/D3RA06276G.
44. Zhao, X., Allison, D., Condon, B., Zhang, F., Gheyi, T., Zhang, A., Ashok, S., Russell, M., MacEwan, I., Qian, Y., Jamison, J. A., Luz, J. G. *Crystal structure of the SIRT1 catalytic domain bound to NAD and an EX527 analog.* **Journal of Medicinal Chemistry**, 56(3), 963–969 (2013). doi: 10.2210/pdb4i5i/pdb
45. Scarano, N., Brullo, C., Musumeci, F., Millo, E., Bruzzone, S., Schenone, S., & Cichero, E. *Recent Advances in the Discovery of SIRT1/2 Inhibitors via Computational Methods: A Perspective.* *Pharmaceuticals*, 17(5), 601 (2024). doi: [10.3390/ph1705060](https://doi.org/10.3390/ph1705060)
46. Shen, P., Deng, X., Chen, Z., Ba, X., Qin, K., Huang, Y., Huang, Y., & Li, T. *SIRT1: A Potential Therapeutic Target in Autoimmune Diseases.* *Frontiers in Immunology*, 12, Article 779177 (2021). doi: [10.3389/fimmu.2021.779177](https://doi.org/10.3389/fimmu.2021.779177)
47. Yevsieieva, L. V., K. O. Lohachova, A. Kyrychenko, S. M. Kovalenko, V. V. Ivanov, and O. N. Kalugin. "Main and Papain-like Proteases as Prospective Targets for Pharmacological Treatment of Coronavirus SARS-CoV-2." *RSC Advances*, vol. 13, no. 50, 2023, pp. 35500–35524. doi: 10.1039/D3RA06479D.
48. Hu, Q., Y. Xiong, G.-H. Zhu, Y.-N. Zhang, Y.-W. Zhang, P. Huang, and G.-B. Ge. "The SARS-CoV-2 Main Protease (MPro): Structure, Function, and Emerging Therapies for COVID-19." *MedComm*, vol. 3, no. 3, 2022, e151. <https://doi.org/10.1002/mco2.151>.
49. Lohachova, K. O., A. S. Sviatenko, A. Kyrychenko, V. V. Ivanov, T. Langer, S. M. Kovalenko, and O. N. Kalugin. "Computer-Aided Drug Design of Novel Nirmatrelvir Analogs Inhibiting Main Protease of Coronavirus SARS-CoV-2."

- Journal of Applied Pharmaceutical Science*, vol. 14, no. 5, 2024, pp. 232–239. doi: 10.7324/JAPS.2024.158114.
50. Ferreira, Juliana C., Adrian J. Villanueva, Kenana Al Adem, et al. "Identification of Novel Allosteric Sites of SARS-CoV-2 Papain-like Protease (PLpro) for the Development of COVID-19 Antivirals." *Volume 300, Issue 111078*, Nov. 2024, doi:10.1016/j.jmb.2024.111078.
51. Yin, W., et al. "Structural Basis for Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir." *Science*, vol. 368, 2020, pp. 1499–1504, doi:10.1126/science.abc1560.
52. Treviño, Miguel Á., et al. "SARS-CoV-2 Nsp8 N-terminal Domain Folds Autonomously and Binds dsRNA." *Nucleic Acids Research*, vol. 51, no. 18, 13 Oct. 2023, pp. 10041–10048. Oxford University Press, <https://doi.org/10.1093/nar/gkad714>.
53. Batt, S. M., et al. "Structural Basis of Inhibition of *Mycobacterium tuberculosis* DprE1 by Benzothiazinone Inhibitors." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, 2012, pp. 11354–11359, <https://doi.org/10.1073/pnas.1205735109>.
54. Yadav, Saloni, et al. "DprE1 Inhibitors: Enduring Aspirations for Future Antituberculosis Drug Discovery." *ChemMedChem*, vol. 18, no. 16, June 2023, article e202300099, Wiley, <https://doi.org/10.1002/cmdc.202300099>.
55. Hong, W., et al. "The Identification of Novel *Mycobacterium tuberculosis* DHFR Inhibitors and the Investigation of Their Binding Preferences by Using Molecular Modelling." *Scientific Reports*, vol. 5, 2015, article 15328, <https://doi.org/10.1038/srep15328>.
56. Snizhko, A. D., et al. "Synthesis of Novel Derivatives of 5,6,7,8-Tetrahydro-Quinazolines Using α -Aminoamidines and In Silico Screening of Their Biological Activity." *International Journal of Molecular Sciences*, vol. 23, no. 7, 2022, article 3781, <https://doi.org/10.3390/ijms23073781>.
57. Dal Molin, M., et al. "Identification of Novel Scaffolds Targeting *Mycobacterium tuberculosis*." *Journal of Molecular Medicine*, vol. 97, no. 11, 2019, pp. 1601–1613, <https://doi.org/10.1007/s00109-019-01840-7>.

58. Wang, S., Eisenberg, D. "Crystal structure of pantothenate kinase from *Mycobacterium tuberculosis*." *Journal of Structural Biology*, vol. 188, no. 3, 2014, pp. 270-278. <https://doi.org/10.1016/j.jsb.2014.10.005>.
59. Yang, K., Zhou, F., Sun, H. "Advances in structural and functional studies of pantothenate kinase, a target for antituberculosis drug development." *Current Pharmaceutical Design*, vol. 23, no. 8, 2017, pp. 1271-1280. <https://doi.org/10.2174/1381612823666170303150723>.
60. Choudhry, A. E., Mandichak, T. L., Broskey, J. P., et al. "Inhibition of pantothenate kinase from *Mycobacterium tuberculosis* by the antibiotic pantothenamide." *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 10, 2003, pp. 3171-3176. <https://doi.org/10.1128/AAC.47.10.3171-3176.2003>.
61. Wang, Q., et al. "Structural and functional basis of SARS-CoV-2 Omicron Spike-protein-open-2 binding to human ACE2." *Nature Communications*, vol. 13, no. 1, 2022, p. 3183. DOI: 10.1038/s41467-022-30981-6.
62. Huang, Y., et al. "SARS-CoV-2 spike protein open states: Implications for virus neutralization and therapeutic targeting." *Nature Communications*, vol. 13, no. 1, 2022, p. 3182. DOI: 10.1038/s41467-022-30982-5.
63. Lee, K. H., et al. "Structural insight into the Omicron variant of SARS-CoV-2: Implications for therapeutic development." *Cell*, vol. 185, no. 11, 2022, pp. 1887-1898. DOI: 10.1016/j.cell.2022.04.029.