

Голові разової спеціалізованої вченої ради
Інституту фізики НАН України
доктору фіз.-мат. наук Габовичу О. М.

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

на дисертаційну роботу Войціцького Тараса Васильовича
“Визначення параметрів нековалентних взаємодій малих органічних молекул із
протеїнами на основі моделей машинного навчання”
представлену до захисту на здобуття ступеня доктора філософії
за спеціальністю 104-фізика та астрономія

Вивчення взаємодії малих біологічно активних молекул з білками є важливим для розв’язання фундаментальних задач біофізики, молекулярної біології, біохімії та інших галузей науки. Це створює основу для покращення та розробки нових лікарських препаратів. Наразі вже існують цілі бібліотеки хімічних сполук, які потенційно можуть стати ліками від тієї чи іншої хвороби, проте подолання шляху від виявлення такої сполуки до впровадження у медичну практику в якості лікарського препарату проходить щонайменше десять років. Скорочення тривалості підготовчого періоду є одним з першочергових практичних завдань, що стоїть перед вченими. У зв’язку з цим розробка та застосування високоефективних числових методів, які би дозволяли відібрати серед великої кількості молекул ті, що мають найбільші шанси на успіх у подальшій розробці лікарського засобу, вже стало пріоритетним прикладним міждисциплінарним напрямом розвитку науки.

Останніми роками відбувся стрибок у розвитку обчислювальних методів із застосуванням методів машинного навчання та штучного інтелекту. Застосування цих методів відкриває нові можливості для швидкого та ефективного передбачення характеристик нековалентних взаємодій малих органічних молекул із протеїнами. Навчання моделей на великих масивах даних може забезпечити відтворення складних закономірностей молекулярного розпізнавання при значно менших обчислювальних витратах у порівнянні з деталізованими розрахунками в рамках методів класичної молекулярної динаміки, чи квантової хімії. Таким чином, тема дисертаційної роботи “Визначення параметрів нековалентних взаємодій малих органічних молекул із протеїнами на основі моделей машинного навчання” Войціцького Тараса Васильовича є актуальною і представляє неабиякий практичний інтерес.

У роботі автора розроблено графову штучну нейронну мережу 3DProtDTA для передбачення афінності зв’язування біологічно активних малих молекул з білками. В цьому методі структура білків представляється на рівні амінокислотних залишків та здійснюється навчання з використанням структур протеїнів, передбачених методом AlphaFold. Важливим є те, що в роботі було запропоновано новий підхід для подолання проблеми обмеженої кількості експериментальних даних, який пов’язаний з розширенням навчальних вибірок моделей молекулярного докінгу

шляхом генерації штучних комплексів протеїн-ліганд, що відтворюють статистичні закономірності нековалентних взаємодій, характерні для реальних комплексів. Також створено новий метод молекулярного докінгу ArtiDock, який поєднує точність сучасних штучних нейронних мереж із високою обчислювальною ефективністю, необхідною для проведення віртуального скринінгу великих хімічних баз даних. Розвинуті в дисертації методи мають велике практичне значення.

Дослідження по темі дисертації Войціцького Т.В. проводилися в рамках проєктів «Фізичні ефекти та молекулярні механізми взаємодії біологічних молекул та біологічних систем різного рівня структурної організації з малими молекулами, нанорозмірними та супрамолекулярними комплексами» РК №0123U100990 (2023-2027 рр.); «Низькорозмірні графеноподібні дихальогеніди перехідних металів з контрольованими полярними та електронними властивостями для перспективних застосувань у наноелектроніці та біомедицині» НФДУ № 2020.02/0027.

У вступі обґрунтовано актуальність обраної тематики досліджень, сформульовано мету роботи, основні задачі, об'єкт та предмет дослідження, вказується новизна отриманих результатів та їхнє практичне значення, описано особистий внесок здобувача в дослідженні, наводиться інформація про апробацію та публікації за темою дисертації.

У першому розділі проведено огляд сучасного стану досліджень взаємодій між білками та малими органічними молекулами. Описано фізико-хімічні механізми взаємодій між білками та лігандами, зроблено огляд методів їх досліджень. Окреслено роль сучасних методів машинного навчання, зокрема графових нейронних мереж, трансформерів та дифузійних моделей.

У другому розділі описано розроблену графову нейронну мережу 3DProtDTA для передбачення афінності зв'язування лігандів з білками. В рамках запропонованого методу було впроваджено новий підхід до кодування структури як білків, так і лігандів у вигляді графів, що інтегрує просторову інформацію молекул. Характеристики вхідних даних білків були отримані за допомогою бази даних передбачених білкових структур AlphaFold. Аналіз на еталонних наборах даних Davis та KIBA показав, що ефективність моделі 3DProtDTA перевершує наявні аналоги на 1-20% за середньоквадратичним відхиленням, коефіцієнтом конкордації та коефіцієнтом кореляції.

У третьому розділі описано генеративну дифузійну модель глибокого навчання PocketCFDM для передбачення позиціонування малих біологічно активних молекул в сайтах зв'язування білка. Розроблено спеціальний алгоритм, який генерує штучні комплекси білка з лігандами з фізико-хімічними властивостями комплексів, що є подібними до експериментально визначених. Розроблений підхід суттєво розширює об'єм даних, що може використовуватися для тренування моделей машинного навчання, і таким чином, розв'язує проблему обмеженої кількості наявних експериментальних даних. Запропонований підхід підвищує частку

правильних передбачень положення лігандів на 2-10% за критерієм середньоквадратичного відхилення координат атомів та знижує частку передбачень зі стеричними зіткненнями на 5-6%.

У четвертому розділі представлено метод протеїн-лігандного докінгу ArtiDock на основі глибинного навчання. Запропонований метод оптимізований для сценаріїв високопродуктивного віртуального скринінгу. Використовуючи тестові набори даних PLINDER, показано, що ArtiDock передбачає положення лігандів значно точніше, порівняно з іншими методами докінгу.

Серед одержаних у дисертації результатів найбільшу наукову цінність, на мою думку, мають наступні:

- Розроблено графову штучну нейронну мережу 3DProtDTA для передбачення афінності зв'язування лігандів, яка кодує структури протеїнів на рівні амінокислотних залишків та навчається на структурах, передбачених методом AlphaFold.
- Запропоновано новий підхід до збільшення кількості тренувальних даних для моделей молекулярного докінгу шляхом генерації штучних комплексів протеїн-ліганд, які імітують статистичні закономірності нековалентних взаємодій реальних комплексів, що розв'язує проблему обмеженої кількості наявних експериментальних даних.
- Створено новий метод молекулярного докінгу ArtiDock, який поєднує точність сучасних штучних нейронних мереж зі швидкістю, необхідною для віртуального скринінгу великих хімічних баз даних.

Усі одержані результати детально проаналізовано. Їх достовірність забезпечується застосуванням новітніх методів молекулярного моделювання із застосуваннями методів машинного навчання. Водночас аналіз дисертаційної роботи викликає деякі зауваження:

1. Запропоновані підходи доцільно пов'язати з урахуванням молекул води, іонів та, загалом, ефектів гідратації, які можуть суттєво впливати на структуру, стабільність і характер взаємодій досліджуваних молекулярних комплексів.
2. Для додаткової верифікації отриманих результатів та окремих висновків, зроблених на основі запропонованих моделей, доцільно було б провести їх порівняння з результатами, отриманими незалежними методами, зокрема методом класичної молекулярної динаміки. Це дозволило б оцінити відтворюваність отриманих результатів, межі застосовності розробленого підходу та додаткові деталі на молекулярному рівні.
3. В роботі основна увага зосереджена на оцінюванні прогностичної здатності розроблених моделей. Водночас було б цікаво детальніше проаналізувати випадки найбільших похибок прогнозування та з'ясувати, з якими структурними особливостями комплексів білок-ліганд вони можуть бути пов'язані. Це дозволило б краще оцінити межі застосовності

запропонованих моделей та визначити можливі напрями їх подальшого вдосконалення.

4. Представляється перспективним розроблення та застосування подібних методів машинного навчання для дослідження структурних і фізико-хімічних властивостей ДНК, зокрема для передбачення взаємодії ДНК з лігандами, іонами та іншими молекулами. Звичайно, це є швидше побажанням на майбутнє, аніж зауваженням.
5. Серед технічних зауважень, слід зазначити, що перша частина другого розділу значною мірою має оглядовий характер і містить опис загальновідомих положень щодо міжмолекулярних взаємодій. Цей матеріал доцільно було б частково перенести до першого розділу.

Водночас хочу наголосити, що сформульовані зауваження не ставлять під сумнів основних висновків дисертації та не зменшують загалом позитивного враження від роботи. Сподіваюся, що вони будуть корисними автору в подальших дослідженнях і стимулюватимуть його до розвитку цього наукового напрямку.

Основні результати роботи повністю опубліковано в 2 статтях RSC Advances (Q1) і 1 статті Journal of Chemical Information and Modeling (Q1). Дисертація написана грамотною мовою. Зміст дисертації відповідає спеціальності 104 – фізика та астрономія.

Враховуючи актуальність теми, новизну, наукове та практичне значення одержаних результатів, вважаю, що дисертаційна робота «Визначення параметрів нековалентних взаємодій малих органічних молекул із протеїнами на основі моделей машинного навчання» відповідає вимогам “Порядку присудження ступеня доктора філософії”, затвердженого Постановою № 44 Кабінету Міністрів України від 12 січня 2022 року. Порушень академічної доброчесності в дисертації та в наукових працях, у яких було представлено результати дисертації, мною не виявлено. Надана робота заслуговує на позитивну оцінку, а її автор, Войціцький Тарас Васильович, заслуговує на присудження ступеня доктора філософії за спеціальністю 104 – фізика та астрономія.

Директор
Інституту теоретичної фізики
ім. М.М. Боголюбова
Національної академії наук України
д-р. фіз.-мат. наук
старший дослідник



Сергій ПЕРЕПЕЛИЦЯ