

Технічні науки

УДК 577.322

Докашенко Богдан Вікторович

студент

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Докашенко Богдан Викторович

студент

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Dokashenko Bohdan

Student of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

**МОДЕЛЮВАННЯ МОЛЕКУЛЯРНОЇ СТРУКТУРИ КОМПЛЕКСУ
БІЛКА З ЛІГАНДОМ ЗА ДОПОМОГОЮ НЕЙРОМЕРЕЖІ
МОДЕЛИРОВАНИЕ МОЛЕКУЛЯРНОЙ СТРУКТУРЫ КОМПЛЕКСА
БЕЛКА С ЛИГАНДОМ С ПОМОЩЬЮ НЕЙРОСЕТИ
MOLECULAR MODELLING OF THE STRUCTURE OF A PROTEIN
COMPLEX WITH A LIGAND VIA A NEURAL NETWORK**

***Анотація.** Стаття представляє новий підхід до молекулярного докингу через використання нейронних мереж. Нейронні мережі здатні узагальнювати величезну кількість даних і генерувати на їх основі і знайдених в них законах в реальному часі нові приклади. Описано значення та проблеми молекулярного докингу. Класичні біохімічні підходи, їхні плюси та мінуси описані та порівнюються з запропонованим методом. Описано нову структуру методу та результати. Також пропонуються шляхи подальшого дослідження.*

Ключові слова: нейронна мережа, молекулярний докинг, ліганд, білок.

Аннотація. В даній статті представлено новий підхід до молекулярного докингу з використанням нейронних мереж. Нейронні мережі здатні обобщати велику кількість даних і генерувати на їх основі і знайденних закономірностях в реальному часі нові. Описано значимість і проблеми молекулярного докингу. Класичическіе біохімічні підходи, їх плюси і мінуси описуються і порівнюються з пропонуємым методом. Новіа структура методу і результати описані. Пропонуються також шляхи далінейших досліджень.

Ключевые слова: нейронная сеть, молекулярный докинг, лиганд, белок.

Summary. This paper presents the new approach for molecular docking through using neural networks. Neural networks are able to summarize a huge amount of data and generate real-time new ones based on them and the laws found. The significance and problems of molecular docking are described. Classical biochemical approaches, their pros and cons are described and compared with the proposed method. New method structure and results described. The ways for further research are proposed as well.

Key words: neural network, molecular docking, ligand, protein.

Постановка проблеми. На сьогоднішній день, молекулярний докинг є важливою частиною досліджень фізики та хімії. Основною проблемою постає неможливість проведення дослідів і експериментів такої великої кількості, а програми для моделювання працюють занадто довго і не завжди дають коректні результати.

Аналіз останніх досліджень і публікацій. Існують два основні підходи до моделювання докингу: взаємозалежність форми та симуляція. При першому підході білок та ліганд описуються як додаткові поверхні та

увага приділяється саме геометричній відповідності. При другому підході ліганд знаходить потрібне положення ітераційним шляхом. Кожна ітерація включає в себе переміщення, обертання та конформацію ліганд (конформаційний пошук), після чого заново обчислюється енергійна оцінка системи (оціночна функція). Виділяють такі оціночні функції як молекулярно-механічне поле сил, потенціал середніх сил, емпірична оціночна функція тощо. Зазвичай докинг видає не один результат сполуки, а цілий набір можливих варіантів, серед якого обирають найбільш правдоподібний варіант. Також при виборі найбільш підходящих критеріїв оцінки для дослідження систем білок-ліганд, потрібно враховувати такі міжмолекулярні взаємодії як водородні зв'язки, гідрофобні зв'язки (ефект яких складається з множини різноманітних міжмолекулярних взаємодій, таких як електростатичного, ван-дер-ваальсового, ентропійного ефектів тощо), стекінг тощо. Окрім оціночних функцій на результати докингу може впливати конформаційна рухливість білка-рецептора, але при цьому виникає проблема «комбінаторного вибуху», що в рази збільшує час докингу.

Отже, зазвичай при молекулярному докингу використовують підход симуляції, адже процес його роботи фізично ближчий до того, що відбувається насправді і зазвичай більш точний. Для підвищення точності може використовуватися консенсусний підход, який дає можливість переранжирувати отриманий після докингу набір по більш ефективному критерію.

Формулювання цілей статті (постановка завдання). Задача полягає в застосуванні методів машинного навчання, нейронної мережі для процесу молекулярного докингу задля підвищення точності та швидкості даного процесу.

Виклад основного матеріалу. На сьогоднішній день задача молекулярного моделювання є невід'ємною частиною фундаментальних досліджень, спрямованих на вивчення механізмів функціонування білків, а також прикладних фармацевтичних задач, таких як створення нових лікарських сполук, отримання високопродуктивних ферментів тощо. Молекулярний докинг – це метод молекулярного моделювання, метою якого є пошук та передбачення найбільш вірогідного та достовірного розташування та конформації молекул ліганда по відношенню до молекул протеїну для утворення нових стійких сполук.

Існує досить багато методів та алгоритмів для молекулярного докингу, які реалізовані в таких програмах як AUTODOCK, FLEXX, OPENBABEL тощо, але й досі не визначено остаточного підходу або алгоритму, який вважався би найефективнішим та найточнішим. Крім того, результати теоретичних досліджень досить часто не співпадають з експериментальними результатами та майже всі програми для молекулярного докингу працюють досить довго. Тож, знаходження оптимального, ефективного, швидкого та точного методу докингу залишається відкритим питанням, а сам процес докингу є областю активних наукових розробок.

Таким чином, для даної задачі було вирішено відійти від стандартних біохімічних алгоритмів та застосувати методи машинного навчання, а саме нейронну мережу.

Отже, як бачимо, з біохімічної точки зору процес молекулярного докингу дуже складний, довгий, не завжди дає точний результат, не має єдиного підходу для різних типів ліганд та білків, що змушує кожен раз перебирати та порівнювати результати різних методів та алгоритмів. Таким чином, в даній статті пропонується замість стандартних біохімічних підходів перейти до підходів машинного навчання, що узагальнить алгоритм роботи з різними типами молекул.

Існує багато можливостей електронного зберігання молекул. В нашому випадку маємо тривимірні структури молекул, що записані в форматі .pdb. Для подальшої роботи з даними було вирішено представляти молекули у вигляді кубів. Для кожного типу атомів будується куб, який розбитий сіткою, розмір комірки якої дорівнює 1 Å. В кожній комірці сітки одиницею визначається присутність атому та робиться розмиття за допомоги функції Гауса.

Нейромережа навчалася на прикладах скомпонованих структур білка з лігандом з бази *PDB*. Після тривалого навчання модель почала видавати результати. Ці розмиті координати які видає нейромережа потім подаються у другий етап докину на якому молекула ліганда, враховуючи структуру білка та передбачувані координати нейромережею, має знайти коректне значення позиції ліганда роблячі рухи у просторі. Причому порівнюється не тільки оригінал молекули ліганда, а й її конформації. Далі за допомогою методу induced fit підбирається конформація протеїну.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі. Отже, при порівнянні результатів молекулярного докину за допомогою стандартних біохімічних алгоритмів із запропонованим алгоритмом використання нейронних мереж було виявлено, що при використанні запропонованого алгоритму процес докину проходить в декілька разів швидше. Таким чином, тривалість докину зменшується з кількох годин до 20 хв. Крім того, запропонована модель узагальнює різні підходи до моделювання, враховує різні взаємозв'язки між атомами для різних моделей, таким чином досліднику не потрібно це робити власноруч.

Але на сьогодні розроблена модель має меншу точність ніж існуючі. У той час як звичайні алгоритми досягають точності у 0.3-0.7 Å, запропонована модель дає точність 0.8-1.2 Å, що не є критичним та може

бути цілковито поліпшено за допомогою навчання моделі на більшій кількості даних та за допомогою подальшого підбору параметрів.

Література

1. Joseph Gomes, Bharath Ramsundar, Evan N. Feinberg, and Vijay S. Pande Atomic Convolutional Networks for predicting Protein-Ligand binding affinity [Електронний ресурс]. – Режим доступу: <https://arxiv.org/pdf/1703.10603.pdf>
2. Пырков Т.В., Озеров И.В., Балицкая Е.Д., Ефремов Р.Г. (2010), Молекулярный докинг: роль невалентных взаимодействий в образовании комплексов белков с нуклеотидами и пептидами [Електронний ресурс]. – Режим доступу: [http://www.rjbc.ru/2010/4/2010_36_4\(2\).pdf](http://www.rjbc.ru/2010/4/2010_36_4(2).pdf)
3. Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron & Bengio, Yoshua (2014), "Generative Adversarial Networks", arXiv:1406.2661
4. Salimans, Tim; Goodfellow, Ian; Zaremba, Wojciech; Cheung, Vicki; Radford, Alec & Chen, Xi (2016), "Improved Techniques for Training GANs", arXiv:1606.03498