



КОРНЕЛЮК
Олександр Іванович —
член-кореспондент НАН
України, завідувач відділу
білкової інженерії та
біоінформатики Інституту
молекулярної біології і генетики
НАН України

ШТУЧНИЙ ІНТЕЛЕКТ І МОДЕЛЮВАННЯ ПРОСТОРОВОЇ СТРУКТУРИ ПРОТЕЇНІВ

Шановний Анатолію Глібовичу!

Шановні присутні!

Визначення просторової структури протеїнів є необхідним етапом у процесі встановлення взаємозв'язку між структурою та функцією білків. Потік даних про первинні структури протеїнів постійно зростає, що пов'язано зі значними успіхами в секвенуванні геномів організмів, однак експериментально встановлені просторові структури доступні лише приблизно для 10 % відомих послідовностей. Отримані методами рентгеноструктурного аналізу та ЯМР-спектроскопії дані про 3D-структури протеїнів депонуються і зберігаються у банках даних просторових структур.

Тому для аналізу більшості протеїнів особливого значення набувають альтернативні підходи, зокрема комп'ютерні методи передбачення їхньої просторової структури, які належать до методів біології *in silico*. Методи теоретичного передбачення структури протеїнів за повнотою фізичного підходу до їх опису можна поділити на методи розрахунку *ab initio*, які ґрунтуються на вихідних фізичних принципах, та методи класичної молекулярної механіки, в яких застосовують статистично встановлені правила у вигляді силових полів.

Для побудови структурної моделі протеїну зазвичай використовують найбільш ефективний серед емпіричних підходів — метод моделювання за гомологією, або метод порівняльного моделювання, оснований на фундаментальному принципі залежності між рівнем гомології амінокислотних послідовностей та подібністю їхньої просторової структури. Модель структури протеїну, побудована на основі матриці з гомологією понад 90 %, має такі ж малі похибки, як і кристалографічно визначена структура. Моделювання за гомологією можна проводити безпосередньо в інтернет-середовищі за допомогою таких систем, як Swiss-Model, Modeller, I-TASSER, CPHmodels, SDSC1, FAMS, 3D-JIGSAW та ін.

При комп'ютерному дослідженні структури білків виявилось, що з точки зору формування просторової структури білки можна поділити на дві групи: структуровані протеїни (70 %) та внутрішньо неупорядковані протеїни (30 %), остання з яких викликає в дослідників дуже великий інтерес.

Неупорядковані протеїни було відкрито порівняно недавно. Активне розшифрування повних геномів організмів засвідчило, що велика кількість геномних послідовностей в еукаріотичних геномах (до 30—40 %) можуть кодувати цілі протеїни чи їх сегменти, в яких відсутня впорядкована тривимірна структура. Проте неупорядковані ділянки протеїнів часто є функціональними, що суперечить класичній точці зору про детермінацію функції протеїну однією стабільною конформацією. Наприклад, фосфорильований кіназоіндукований домен (pKID) транскрипційного фактора CREB є неструктурованим у розчині, але при зв'язуванні з KIX-доменом CREB-зв'язувального білка (CBP) формує альфа-спіралі, які відповідають за функцію цього фрагмента.

Тому було сформульовано концепцію про існування неупорядкованих протеїнів, які утворюють комплекси з різними білковими партнерами і мають різні мінімуми енергії, тобто кожний комплекс є специфічним. Такий підхід істотно ускладнює картину взаємодії білків усередині клітини.

Наведу ще один приклад. С-кінцевий домен протеїну p53, який формує чотири різних типи структури в комплексах з різними білками: альфа-спіраль, бета-спіраль і два неупорядковані фрагменти.

Відкриття неструктурованих білків значно розширило парадигму, яка раніше зводилася тільки до каталізу структурованими білками, а тепер приводить до ансамблю неупорядкованих структур та множинності функцій цих структур у різних комплексах.

У відділі білкової інженерії та біоінформатики Інституту молекулярної біології і генетики НАН України ми досліджуємо два протеїни апарату трансляції еукаріотів — тирозил-тРНК синтетазу і AIMP1/p43. Виявилось, що окремі

фрагменти першого білка проявляють цитокінову активність. Це не канонічна активність, а тому такі фрагменти потенційно можуть стати лікарськими препаратами, оскільки, як було показано, вони спричиняють процеси апоптозу та антиангіогенні ефекти, і після впровадження відповідної біотехнології їх можна застосовувати для інгібування росту ракових клітин. Отже, ми маємо справу з терапевтичними білками, а тому важливо дослідити їхню структуру. Однак з цим виникає проблема, оскільки ці білки неструктуровані і не кристалізуються. Ми вперше провели комп'ютерне моделювання повнорозмірної структури тирозил-тРНК синтетази.

Поліпептид AIMP1/p43 є мультифункціональним протеїном, який проявляє як власності зв'язування тРНК, так і цитокінову активність. Виділений білок AIMP1/p43 є перспективним терапевтичним протеїном з проапоптотичними та антиангіогенними властивостями. Ми провели комп'ютерне моделювання, визначивши неупорядковані ділянки AIMP1/p43, і виявилось, що цей білок є димером. Далі нам було цікаво порівняти отримані результати з даними, одержаними методами штучного інтелекту.

Хотів би нагадати про значний внесок, який останніми роками зробили методи штучного інтелекту в розвиток комп'ютерного моделювання структури білків. З 2018 р. ці методи успішно застосовують для моделювання просторової структури протеїнів. У 2020 р. в конкурсі CASP14 програма на основі штучного інтелекту AlphaFold2, розроблена компанією DeepMind, показала найкращий результат: для двох третин пропонованих білків її точність передбачення була більшою за 90 %. AlphaFold2 змогла спрогнозувати майбутню структуру білків з похибкою лише в 1,6 Å, що відповідає точності експериментальних методів — ЯМР або рентгенівської кристалографії. База даних AlphaFold Protein Structure Database, створена в партнерстві з Європейським інститутом біоінформатики (EMBL-EBI), містить понад 200 млн моделей структури білків, які є у вільному доступі для світової наукової спільноти. Минуло-

го року розробників AlphaFold було відзначено Нобелівською премією.

В нашому інституті ми займаємося дизайном інгібіторів як потенційних лікарських препаратів. Зокрема, у відділах, якими керують член-кореспондент НАН України С.М. Ярмолюк і академік НАН України М.А. Тукало, розробляють інгібітори двох ферментів — лейцил-тРНК синтетази та метіоніл-тРНК синтетази. З використанням машинного навчання і штучних нейронних мереж було знайдено інгібітори з подвійною інгібувальною активністю до цих двох синтетаз.

Однак це приклад дизайну ліків на структурному рівні. Коли ж ми почали за допомогою AlphaFold2 моделювати структуру невідомого протеїну AIMP1/p43, результат виявився некоректним, оскільки отримана просторова структура не може утворювати димер. Річ у тім, що AlphaFold2 навчали на кристалографічних даних про структури, а невідомі білки не кристалізуються, і тому програма не може правильно передбачити структуру AIMP1/p43.

Проте наука не стоїть на місці. Нова версія моделі AlphaFold3, яку компанія DeepMind представила минулого року, дозволила ближче до реальності промодельовати димерну структуру протеїну AIMP1/p43. Втім, комп'ютерне моделювання для невідомих білків дає кілька різних варіантів структури. зокрема в невідомій ділянці формуються одна чи дві альфа-спіралі, і обидві ці структури можуть бути функціональними. Для дослідження невідомих протеїнів зазвичай застосовують метод ЯМР-спектроскопії, з використанням якого їх, власне, і було відкрито.

На ЯМР-спектрометрі в Інституті біохімії і біофізики (Варшава, Польща) ми вперше одержали просторову структуру С-кінцевого домену AIMP1/p43. Отримані дані підтвердили наявність стабільного ЕМАР-домену в його структурі та альфа-спіралі у невідомій ділянці.

Ще один метод, який ми використовуємо, це комп'ютерне моделювання молекулярної динаміки. Нагадаю, що в Україні свого часу за ініціативою академіка НАН України А.Г. Загороднього було створено потужну грид-інфраструктуру для розрахунків, зокрема й для розрахунків молекулярної динаміки.

Які ж ми бачимо перспективи розвитку досліджень внутрішньо невідомих білків? Це насамперед створення експериментальної бази просторових структур протеїнів у розчині на основі даних ЯМР-спектроскопії, бази даних динамічних структур на основі моделювання молекулярної динаміки; кріоелектронна мікроскопія; пошук функціонально важливих структур комплексів протеїнів методами штучного інтелекту. Все це в комплексі — інтегративна структурна біологія, яка надає можливості для вивчення невідомих білків.

Ми також вважаємо за потрібне відновити цільову комплексну програму наукових досліджень НАН України «Математичне моделювання у міждисциплінарних дослідженнях процесів і систем на основі інтелектуальних суперкомп'ютерних, грид- і хмарних технологій». Крім того, доцільно було б започаткувати нову програму «Використання методів штучного інтелекту для розроблення нових лікарських препаратів».

Дякую за увагу!

Alexander I. Kornelyuk

Institute of Molecular Biology and Genetics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

ORCID: <https://orcid.org/0000-0003-0146-2832>

ARTIFICIAL INTELLIGENCE AND MODELING THE SPATIAL STRUCTURE OF PROTEINS

Speech at the session of the General Meeting of the National Academy of Sciences of Ukraine, April 30, 2025

Cite this article: Kornelyuk A.I. Artificial intelligence and modeling the spatial structure of proteins (speech at the session of the General Meeting of the National Academy of Sciences of Ukraine, April 30, 2025). *Visn. Nac. Akad. Nauk Ukr.* 2025. (6): 72—74. <https://doi.org/10.15407/visn2025.06.072>