



**РОМАНЮК**

**Світлана Іванівна** — кандидат біологічних наук, старший науковий співробітник Інституту біохімії ім. О.В. Палладіна НАН України



**КОМІСАРЕНКО**

**Сергій Васильович** — академік НАН України, директор Інституту біохімії ім. О.В. Палладіна НАН України

## ПРОРИВ У ПРОГНОЗУВАННІ ПРОСТОРОВОЇ СТРУКТУРИ ТА ОБЧИСЛЮВАЛЬНОМУ ДИЗАЙНІ ПРОТЕЇНІВ, АБО ЧИ МОЖНА ВІРИТИ ПЕРЕДБАЧЕННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

Автори статті аналізують Нобелівську премію з хімії 2024 р., яку було присуджено американському біохіміку та комп'ютерному біологу Девіду Бейкеру (David Baker) за «комп'ютерний дизайн білків», а також представникам компанії Google DeepMind: британському фахівцю з систем штучного інтелекту Демісу Гассабісу (Demis Hassabis) і американському хіміку та інформатику Джону Джамперу (John M. Jumper) за «прогнозування структури білка». Досягнення нобелівських лауреатів у галузі обчислювального проектування протеїнів та прогнозування їхньої структури відкрили нову еру біохімічних і біологічних досліджень, що в поєднанні із застосуванням інструментів штучного інтелекту матиме далекосяжні наслідки для людства.

**Ключові слова:** Нобелівська премія з хімії 2024 року, Девід Бейкер, Деміс Гассабіс, Джон Джампер, проектування білків, прогнозування структури протеїнів.

9 жовтня 2024 р. Нобелівський комітет при Каролінському медичному інституті у Стокгольмі оголосив імена лауреатів Нобелівської премії з хімії. Напередодні компанія Clarivate за аналізом кількості цитувань визначила імена найбільш імовірних претендентів на Нобелівську премію з хімії 2024 р.<sup>1</sup>.

По-перше, це американець Роберто Кар (Roberto Car) — професор хімії Принстонського інституту матеріалів, директор Обчислювального хімічного наукового центру з хімії розчину та інтерфейсів при Принстонському університеті, а також швейцарський вчений Мікеле Паррінелло (Michele Parrinello) — почесний професор обчислювальної техніки на факультеті інформатики Італійського університету Швейцарії в Лугано, почесний професор кафедри хімії та прикладних біологічних наук у Швейцарській вищій технічній школі Цюриха, які розробили

<sup>1</sup> Citation Laureates 2024. Breakthroughs in Chemistry. <http://surf.li/ibfmbmf>

метод Кара—Паррінелло для розрахунку молекулярної динаміки *ab initio*, що спричинив революцію в комп'ютерній хімії.

По-друге, ймовірним кандидатом на премію назвали японця Казунарі Домена (Kazunari Domen) — професора Інституту регенерації води при Університеті Шиншу в Нагано та професора Токійського університету — за фундаментальні дослідження фотокаталізаторів для розщеплення води й побудови сонячних систем виробництва водню.

По-третє, передбачалося, що Нобелівську премію можуть отримати двоє британських вчених: Джон М. Джемпер і Деміс Гассабіс — директор і відповідно генеральний директор Google DeepMind — компанії з розроблення штучного інтелекту (ШІ), а також американець Девід Бейкер — професор біохімії, дослідник Медичного інституту Говарда Г'юза та директор Інституту дизайну протеїнів Медичної школи Університету Вашингтона у Сієтлі, які зробили значний внесок у прогнозування та проектування тривимірних структур протеїнів та їхніх функцій.

Неймовірно, але цього разу співробітникам компанії Clarivate вдалося абсолютно точно передбачити імена нобелівських лауреатів у галузі хімії, що трапляється не часто. Отже, у 2024 р. лауреатами 116-ї Нобелівської премії з хімії стали троє вчених — Девід Бейкер, Деміс Гассабіс та Джон Джемпер. Генеральний секретар Шведської королівської академії наук Ханс Еллегрен (Hans Ellegren) оголосив мотивування рішення про нагородження: вчених було відзначено цією престижною нагородою «за обчислювальний дизайн протеїну та передбачення його структури». Слід зауважити, що одну половину премії присудили Девіду Бейкеру «за обчислювальний дизайн протеїну», а іншу — Демісу Гассабісу та Джону Джемперу «за передбачення структури протеїну». Згідно з офіційним пресрелізом, «Нобелівська премія з хімії 2024 року присвячена протеїнам — геніальним хімічним інструментам життя. Девіду Бейкеру вдалося створити абсолютно нові види протеїнів. Деміс Гассабіс і Джон Джемпер розробили модель ШІ для розв'язання 50-річ-

ної проблеми: прогнозування складних структур протеїнів. Ці відкриття мають величезний потенціал»<sup>2</sup>.

У 2024 р. одразу дві Нобелівські премії (з фізики та з хімії) було присуджено вченим, дослідження яких стосувалися використання ШІ. Це викликало хвилю звинувачень на адресу Нобелівського комітету в тому, що ці дослідження не є достатньою мірою ні фізикою, ні хімією, а більше стосуються програмування та біології. Нобелівський лауреат з фізики 2024 р. та піонер машинного навчання Джеффері Гінтон (Geoffrey Hinton) так пояснив це: «Намагаючись зрозуміти, як реальні нервові системи досягають своїх дивовижних обчислювальних здібностей, виявилось необхідним вивчити дуже ідеалізовані моделі, які настільки ж відрізняються від реальних біологічних нейронних мереж, як яблука від планет». Цікаво, що рік тому Гінтон звільнився з Google, щоб вільно говорити про небезпеку застосування ШІ, і заявив, що частина його тепер шкодує про роботу свого життя<sup>3</sup>.

Однак повернімося до нобелівських лауреатів з хімії 2024 р. і познайомимося з ними ближче.

62-річний професор біохімії **Девід Бейкер** є дослідником Медичного інституту Говарда Г'юза та ад'юнкт-професором наук про геном, біоінженерії, хімічної інженерії, комп'ютерних наук і фізики в Університеті Вашингтона у Сієтлі (штат Вашингтон, США). Девід Бейкер народився 6 жовтня 1962 р. у Сієтлі в єврейській родині. Його батьки були викладачами Університету Вашингтона: зараз 92-річний батько Девіда — Маршалл Бейкер є почесним професором відділу фізики, що працює в галузі квантової фізики, зокрема теорії елементарних частинок, а матір — Марсія Бурджін Бейкер є професором геофізики та атмосферних наук на пенсії, а раніше вивчала механізм утворення блискавок і фізику хмар для розроблення глобальних кліматичних моделей. Девід Бейкер виріс неподалік від кампусу Університету

<sup>2</sup> The Nobel Prize in Chemistry 2024. Press release. <http://surl.li/axxhcd>

<sup>3</sup> Nature Briefing, 11 October 2024. <http://surl.li/kwloyf>

Вашингтона, закінчив середню школу Гарфілда в Сієтлі, а у 1984 р. здобув ступінь бакалавра мистецтв з біології в Гарвардському університеті у Кембриджі (штат Массачусетс). У 1989 р. Бейкер здобув ступінь доктора філософії з біохімії в Каліфорнійському університеті в Берклі за роботу з дослідження транспорту протеїнів у дріжджових екстрактах, виконану під керівництвом Ренді Шекмана (Randy Schekman), який у 2013 р. отримав Нобелівську премію з фізіології або медицини за відкриття механізму регулювання везикулярного транспорту в клітинах. Потім у 1993 р. Девід Бейкер закінчив докторантуру з біофізики у Девіда Агарда (David Agard) в Каліфорнійському університеті в Сан-Франциско та приєднався як викладач до факультету біохімії Медичної школи Університету Вашингтона у Сієтлі. У 2000 р. він став дослідником Медичного інституту Говарда Г'юза і зараз очолює Інститут дизайну протеїнів Університету Вашингтона. Цікаво, що Девід Бейкер став восьмим викладачем цього університету, який отримав Нобелівську премію.

Девід Бейкер є членом Національної академії наук США (з 2006) і Американської академії мистецтв і наук (з 2009), а також у 2024 р. він увійшов до списку 100 найвпливовіших людей у галузі охорони здоров'я журналу Time. Він є автором більш як 640 наукових праць з біохімії та понад 100 патентів, 90 його докторантів і постдокторантів успішно продовжили наукову та викладацьку кар'єру. Крім того, Бейкер є співзасновником 21 біотехнологічної компанії, серед яких найбільш відомими є Prospect Genomics (у 2001 р. її придбала дочірня компанія Eli Lilly, Icosavax, яку у 2023 р. придбала AstraZeneca), а також Sana Biotechnology, Lyell Immunotherapeutics і Xaira Therapeutics.

За свою роботу Девід Бейкер здобув численні нагороди, такі як стипендія Паккарда в галузі науки та інженерії (1994); премія Бекмана для молодих дослідників від Університету Вашингтона (1995); премія Овертона (2002); премія Ньюкомба Клівленда, премія Фейнмана з нанотехнологій (2004); міжнародна премія Саклера з біофізики (2008); премія «Зухвалий проект» від фонду TED (2019); премія «За про-

рив» у науках про життя (2021); премія Вайлі та премія Фонду BBVA «Межі знань» у категорії «Біологія та біомедицина» (2022).

У вільний час Девід Бейкер завзято мандрує. Він одружений з Ханнелою Руохола-Бейкер, яка має фінське походження і є професором біохімії, заступником директора Інституту стовбурових клітин і регенеративної медицини Університету Вашингтона в Сієтлі. Цікаво, що Ханнела вивчає диференціювання стовбурових клітин і роль мікроРНК у цьому процесі, а у 2024 р. Нобелівську премію з фізіології або медицини було присуджено саме за відкриття ролі мікроРНК у посттрансляційному регулюванні генів.

48-річний сер **Деміс Гассабіс** є розробником комп'ютерних ігор, програмістом, нейробіологом, шахістом світового класу, засновником і генеральним директором компанії DeepMind у Лондоні, а також урядовим радником Великої Британії з питань ШІ. Він народився 27 липня 1976 р. на півночі Лондона в сім'ї вчителів-мігрантів: грека-кіпріота та китайки з Сінгапуру. Його мати працювала в елітному універмазі John Lewis, а батько був співаком і автором пісень. Деміс був старшим із трьох дітей у сім'ї, зараз його сестра — композитор і піаністка, а брат — дослідник творчого письма. Ніхто з членів родини не захоплювався математикою, комп'ютерами чи загалом наукою. У віці чотирьох років Деміс Гассабіс почав грати в шахи, через рік він вже змагався на національному рівні, а згодом — і на міжнародному. У віці 8 років за гроші, виграні на шаховому турнірі, Деміс купив свій перший комп'ютер ZX Spectrum і самостійно опанував літературу з програмування. У 1989 р. Гассабіс став володарем найвищого рейтингу Ело серед шахістів віком до 13 років (набравши 2300 очок), тоді як найвищий у світі рейтинг Ело — 2882 пункти здобув у 2014 р. норвежець Магнус Карлсен. Деміс Гассабіс був капітаном багатьох юніорських шахових команд Англії та єдиний у світі п'ять разів перемагав на чемпіонаті світу Pentamind з інтелектуальних видів спорту. У 14 років Гассабіса відзначили на конкурсі дизайнерів-аматорів, проведеному журналом Amiga

Power, надавши місце тестера у відомій британській ігровій студії Bullfrog Productions та можливість познайомитися з культовим розробником комп'ютерних ігор Пітером Молінє.

Шкільну освіту Деміс спочатку отримував вдома, де його навчали батьки, у 12—14 років навчався у гімназії для хлопчиків — школі королеви Єлизавети в районі Барнет, а потім відвідував державну загальноосвітню школу Коледж Христа в районі Фінчлі, яку закінчив на два роки раніше, успішно склавши випускні іспити у 16 років. Спроба вступити до Кембриджського університету виявилася невдалою: йому відмовили через занадто юний вік і попросили зробити перерву.

Протягом цієї перерви Гассабіс продовжив свою кар'єру розробника ігор у Bullfrog Productions. Спочатку він розробляв рівні у грі Syndicate, а в 17 років очолив проект зі створення надпопулярного економічного симулятора Theme Park, продажі якого становили кілька мільйонів копій. Ця гра стала поштовхом для розвитку жанру імітаційних ігор-пісочниць, у яких гравець має можливість створювати, змінювати або знищувати віртуальний навколишній світ. Пітер Молінє назвав Деміса «найкращим співробітником століття». Деміс Гассабіс заробив достатньо, щоб самостійно оплатити навчання в університеті, і покинув Bullfrog заради вступу до Квінс-коледжу Кембриджського університету, в якому він у 1997 р. закінчив курс бакалаврату з інформатики Computer Science Tripos з подвійною відзнакою. Після закінчення університету Гассабіс рік працював у щойно створеній Пітером Молінє компанії Lionhead Studios, де був провідним програмістом ІІІ для гри Black & White (2001). У 1998 р. Деміс Гассабіс покинув Lionhead, щоб у віці 22 років заснувати власну компанію з розроблення ігор Elixir Studios, яка створила ще дві відомі стратегії: Republic: The Revolution (2003) і Evil Genius (2005), які номінувалися на премію BAFTA за інтерактивну музику, створену британським композитором Джеймсом Ганніганом, відомим передусім своєю музикою до ігор про Гаррі Поттера. У 2000 р. компанію Elixir Studios було оцінено в £12 млн. Однак уже че-

рез п'ять років Гассабіс, продавши всю інтелектуальну власність іншим студіям, закритим Elixir Studios, пішов з ігрової індустрії та занурився у світ нейронауки, проголосивши своєю метою розроблення нових технологій ІІІ.

У 2009 р. Деміс Гассабіс здобув ступінь доктора філософії з когнітивної нейронауки в Університетському коледжі Лондона, виконавши роботу з дослідження нервових процесів, які лежать в основі епізодичної пам'яті, під керівництвом професора когнітивної нейронауки ірландського походження Елеонор Магауір (Eleanor Maguire). Надалі він продовжив свої дослідження в галузі нейронауки та ІІІ в Кембриджі (штат Массачусетс, США) як запрошений вчений у Гарвардському університеті, а також у лабораторії відомого нейробіолога Томазо Поджіо (Tomaso A. Poggio) в Массачусетському технологічному інституті. Отримавши докторську стипендію Генрі Велкома, Гассабіс повернувся до Великої Британії і почав працювати в лабораторії Пітера Даяна (Peter Dayan) — директора підрозділу обчислювальної нейронауки Гетсбі в Університетському коледжі Лондона.

У 2010 р. Деміс Гассабіс разом з двома нейробіологами — Шейном Леггом (Shane Legg) та Мустафою Сулейманом (Mustafa Suleyman) — заснував компанію DeepMind Technologies, яка спеціалізувалася на технологіях ІІІ. У 2014 р. Google придбав британський стартап DeepMind, який після цього отримав назву Google DeepMind. Деміс Гассабіс залишився генеральним директором компанії, і саме тут на нього чекали найбільші наукові здобутки, які й було відзначено Нобелівською премією з хімії 2024 р.

Деміс Гассабіс є членом Королівського товариства мистецтв (з 2009), Науково-консультативної групи з питань надзвичайних ситуацій (з 2012), Королівської інженерної академії наук Великої Британії (з 2017), Лондонського королівського товариства (з 2018), Американської академії мистецтв і наук (з 2021), Папської академії наук і Європейської академії (з 2023). Крім того, Деміса Гассабіса було включено до багатьох рейтингових списків найвпливові-



ших людей світу, Європи та Великої Британії за версією провідних журналів, зокрема Nature 10 (2016) і Time 100 (2017), а також посвячено в лицарі за заслуги в галузі ШІ (2024).

Деміс Гассабіс за останні 10 років здобув близько 50 наукових відзнак і нагород, головними з яких є премія Малларда від Королівського товариства (2014); Срібна медаль Королівської інженерної академії (2016); нагорода «Золота тарілка» від Американської академії досягнень; премія The Asian Awards за видатні досягнення в науці і техніці (2017); премія Дена Девіда в категорії «Майбутнє — Штучний інтелект», медаль Пія ХІ від Папської академії наук (2020); медаль IRI від Інституту промислових досліджень (2021); премія Вайлі в галузі біомедичних наук, премія принцеси Астурійської за технічні та наукові дослідження, премія Фонду BBVA «Межі знань» у категорії «Біологія та біомедицина», Глобальна швейцарська премія в галузі ШІ, премія VinFuture для інноваторів з видатними досягненнями у нових сферах (2022); медаль Лавлейс від Британського комп'ютерного товариства, премія «За прорив» у науках про життя, міжнародна премія канадського Фонду Гайрднера, премія Альберта Ласкера за фундаментальні медичні дослідження (2023) і премія Кейо з медичних наук (2024).

Деміс Гассабіс живе зі своєю сім'єю у Хайгейті — північному районі Лондона, неподалік від місця, де він виріс. Його дружина, італійський молекулярний біолог, досліджує хворобу Альцгеймера. Гассабіс захоплюється футболом і протягом усього життя є фанатом ФК «Ліверпуль». Ще одним його великим захопленням є кіно. Гассабіс вважає своїм другом британського режисера Алекса Гарленда, перший фільм якого *Ex Machina* (2014) про ШІ одержав премію Оскар за найкращі візуальні ефекти. За мотивами життєвого шляху і досягнень Гассабіса американський режисер Грег Кос, десятиразовий лауреат премії «Еммі», зняв два документальних фільми: «Гра в мислення» (2024) і відзначений багатьма нагородами фільм «АльфаГо» (2017)<sup>4</sup>.

<sup>4</sup> Lourie A. Exclusive interview: meet Demis Hassabis, London's megamind who just sold his company to Google for £400m. <https://surl.li/setryi>

39-річний **Джон Майкл Джампер** є хіміком і комп'ютерним науковцем, директором компанії Google DeepMind, яку заснував і досі очолює Деміс Гассабіс. Джон Джампер народився у 1985 р. у пригороді міста Літл-Рок — адміністративного центру штату Арканзас (США), і йому судилося стати першим нобелівським лауреатом з Арканзасу та наймолодшим нобелівським лауреатом з хімії за останні 70 років. Джон з раннього віку виявляв інтерес до математики та природничих наук. Дитинство й підліткові роки він провів у рідному пригороді в оточенні домашніх тварин і чудової природи, навчався у приватній середній школі «Академія Пуласкі» у Літл-Рок, де займався в науково-технічному гуртку, шаховому клубі, гуртку «Модель ООН», а також грав у футбол. У старших класах за дипломну роботу «Модель для країн-ізоїв» Джон отримав відзнаку вчителя з соціальних наук Білла Топіча, який значно вплинув на становлення майбутнього вченого. Після закінчення школи у 2003 р. Джампер переїхав до сусіднього штату Теннессі, щоб вивчати фізику і математику в Університеті Вандербільта в Нешвіллі. У 2007 р. він закінчив університет і здобув ступінь бакалавра з фізики та математики.

Отримавши стипендію Маршалла, Джон Джампер переїхав до Великої Британії для навчання у Коледжі Святого Едмунда при Кембриджському університеті. Джампер потрапив до знаменитої Кавендішської лабораторії, яка зараз є фізичним факультетом університету, а в 1874 р. була створена як перша у світі навчально-наукова лабораторія і названа на честь Генрі та Вільяма Кавендішів, герцогів Девонширу, один з яких був відомим фізиком і хіміком, а інший — канцлером Кембриджського університету, який пожертвував велику суму на відкриття лабораторії. Засновником і першим професором лабораторії був Джеймс Максвелл, розробник електромагнітної теорії, а серед відомих відкриттів, зроблених у Кавендішській лабораторії, — відкриття електрона, нейтрона та структури ДНК. У 2008 р. Джампер завершив навчання в Кембриджському університеті, здобувши ступінь магістра філо-

софії з теоретичної фізики конденсованого середовища, і повернувся до США.

Через відсутність великого інтересу до тієї галузі фізики, якою він займався, Джон Джампер закинув докторську дисертацію і наступні три роки провів в обчислювальній лабораторії D.E. Shaw Research у Нью-Йорку, де розробив модель для вивчення динаміки протеїнів та інших молекул. У 2011 р. Джампер перейшов з лабораторії до Чиказького університету, де застосував машинне навчання для вивчення фізики згортання протеїну. У 2012 р. він здобув ступінь магістра наук з теоретичної хімії, а у 2017 р. — ступінь доктора філософії з теоретичної хімії за роботу, присвячену новим методам моделювання грубозернистого згортання та динаміки протеїну з використанням машинного навчання, яку він виконував під керівництвом Тобіна Сосніка (Tobin Sosnick) і Карла Фріда (Karl Freed) у Чиказькому університеті. Невдовзі після цього Джон Джампер влаштувався на роботу в компанію Google DeepMind у Лондоні, а у 2023 р. став її директором.

Джон Джампер отримав багато нагород, серед яких премія Вайлі у галузі біомедичних наук, премія Фонду BBVA «Межі знань» у категорії «Біологія та біомедицина», премія VinFuture для інноваторів з видатними досягненнями у нових сферах (2022); премія «За прорив» у науках про життя, міжнародна премія канадського Фонду Гайрднера, премія Альберта Ласкера за фундаментальні медичні дослідження (2023); лекція Арех в Університеті Вандербільта (2024). У 2021 р. Джон Джампер увійшов до списку найбільш впливових людей у науці Nature 10.

Хоча особисті відомості про вченого залишаються відносно конфіденційними, відомо, що Джон Джампер живе в Лондоні з сім'єю, його дружина також є науковцем і має ступінь доктора філософії.

Що ж таке обчислювальний дизайн протеїну? Чому визначення просторової структури протеїнів є таким важливим? І до чого тут ШІ?

Протеїни не дарма вважають основою життя на нашій планеті, адже вони є не лише матері-

алом, з якого побудовано всі живі організми, а й багатофункціональними молекулами, необхідними для реалізації важливих для життя процесів: ензиматичного каталізу хімічних реакцій, передавання клітинного сигналу, транспортування інших молекул, реагування на подразники, здійснення імунного захисту, контролю за процесами зчитування генної інформації та синтезу нових протеїнів тощо. Виконання цих різноманітних функцій можливе завдяки тому, що протеїни є макромолекулами зі складною та унікальною тривимірною структурою. Як вона утворюється?

Протеїни — це природні полімери, що синтезуються в клітині з 20 типів L- $\alpha$ -амінокислот, які різняться будовою бічних ланцюгів. Гнучкий поліпептидний ланцюг фактично складається з амінокислотних залишків, з'єднаних пептидними зв'язками  $-\text{NH}-\text{CO}-$ , які утворюються внаслідок реагування карбоксильної групи однієї амінокислоти з аміногрупою іншої амінокислоти. Причому пептидний зв'язок є полярним, оскільки група NH може діяти як донор водневого зв'язку, а група CO — як акцептор. Завдяки водневим зв'язкам між наближеними в просторі групами NH і CO різних ділянок ланцюга утворюються такі регулярні елементи вторинної просторової структури, як  $\alpha$ -спіралі та  $\beta$ -листи, послідовність яких своєю чергою згортається під впливом різних факторів з утворенням складної третинної структури. Основними факторами, що спонукають протеїновий ланцюг згоратися, є численні нековалентні взаємодії між амінокислотними залишками, зокрема водневі зв'язки, гідрофобні ефекти, сили Ван-дер-Ваальса та іонні зв'язки [1].

Вторинні структури залежно від властивостей амінокислот, з яких вони складаються, можуть бути щільно упаковані всередині молекули протеїну в гідрофобному середовищі або розміщуватися на гідрофільній поверхні. Бічний ланцюг кожного типу амінокислоти може займати лише певний обмежений об'єм і має обмежену кількість можливих взаємодій з іншими сусідніми бічними ланцюгами. З огляду на це, стає зрозумілим, чому просторова структура протеїну є унікальною і чому вона визна-

чається його амінокислотою послідовністю. Оскільки функції протеїнів тісно пов'язані з їх тривимірними структурами, вивчення біологічних процесів є фактично неможливим без знання просторової структури протеїнів, які беруть у них участь. Проблема полягає в тому, що для кожної амінокислової послідовності існує величезна кількість можливих варіантів просторової структури.

Лауреати Нобелівської премії з хімії 2024 р. відкрили спосіб швидкого, легкого і дешевого визначення просторової структури протеїну за амінокислотою послідовністю (прогнозування структури протеїнів), а також спосіб визначення амінокислової послідовності протеїну з бажаною просторовою структурою (обчислювальний дизайн протеїнів). Їм вдалося розв'язати цю надзвичайно складну та актуальну проблему завдяки не лише власній геніальності, а й використанню технології ШІ, яка невпинно розвивається протягом останніх 25 років.

Якою ж була історія успіху та попереднього розвитку відповідної галузі науки?

Ця історія розпочалася (як і у випадку багатьох інших видатних відкриттів) у 1950-х роках, коли відбувався бурхливий розвиток молекулярної біології та з'явилися нові технічні можливості для вивчення протеїнів. Уперше просторову структуру протеїну (міоглобіну) «побачив» за допомогою методу рентгенівської кристалографії британський біохімік Джон Кендрю (John Kendrew) у 1958 р. Через два роки його колега Макс Перуц (Max Perutz) з'ясував просторову структуру гемоглобіну. Це видатне відкриття було відзначено Нобелівською премією з хімії 1962 р. та привернуло увагу інших вчених до вивчення принципів побудови просторової структури протеїнів. Так, у 1961 р. американський біохімік Крістіан Анфінсен (Christian Anfinsen) помітив, що після розгортання пептидного ланцюга під впливом різних чинників він завжди згортається в одну й ту саму просторову структуру, яка, очевидно, визначається його амінокислотою послідовністю. Важливість цього відкриття було підкреслено врученням Анфінсену Нобелівської

премії з хімії 1972 р., хоча механізм згортання протеїну не був ще до кінця зрозумілим. Цю невизначеність підкреслив у 1968 р. американський молекулярний біолог Сайрус Левінтал (Cyrus Levinthal), звернувши увагу на парадокс у висновках Анфінсена, який було названо парадоксом Левінтала: «Проміжок часу, за який поліпептид згортається, є на багато порядків меншим, ніж час, протягом якого він просто перебирає би усі можливі конфігурації». Інакше кажучи, якби поліпептидний ланцюг згортався випадковим чином, знадобилося б більше часу, ніж вік Всесвіту, щоб знайти правильну просторову структуру, а в клітині цей процес відбувається лише за кілька мілісекунд<sup>5</sup>.

Так вчені зрозуміли, що згортання є заздалегідь визначеним процесом, зумовленим лише амінокислотою послідовністю протеїну, а отже, має існувати теоретичний спосіб визначити структуру протеїну за його амінокислотою послідовністю. Це завдання стало одним з найважливіших у молекулярній біології, особливо якщо взяти до уваги, що рентгенівська кристалографія потребувала дуже багато часу, коштів і зусиль для одержання результату, який до того ж не для кожного протеїну вдавалося отримати. З початку 1990-х років з розвитком методів секвенування ДНК у базах даних накопичилося понад 2,6 млрд послідовностей ДНК, водночас кількість ідентифікованих протеїнових послідовностей у базі UniProt становила приблизно 200 млн, а просторових структур протеїнів — усього 180 тис. [2]. Незважаючи на поширеність таких точних експериментальних методів, як рентгенівська кристалографія, ЯМР-спектроскопія та криоелектронна мікроскопія, нові дані щодо визначення просторової структури протеїнів з'являлися дуже рідко. Тому з часом розрив між відомими послідовностями протеїнів та їхніми просторовими структурами неминуче зростає. Це спонукало вчених використати обчислювальні методи, щоб віднайти спосіб надійного передбачення просторової структури протеїну на основі амінокислової послідовності. Зворотним боком

<sup>5</sup> Aqvist J. Computational protein design and protein structure prediction. <http://surl.li/jxzyz>

цієї проблеми, ще більш цікавим, стало визначення амінокислотної послідовності, яка могла б згорнутися в протеїн з певною визначеною структурою та потрібними людям властивостями.

Перші спроби вирішити ці завдання було спрямовано на передбачення вторинної структури протеїнів. У 1974 р. Пітер Чоу (Peter Chou) та Джеральд Фасман (Gerald Fasman) створили один із перших алгоритмів для передбачення вторинної структури протеїну, проаналізувавши 15 амінокислотних послідовностей протеїнів з відомою структурою [3]. Проте цей метод виявився не надто точним, адже в алгоритмі не враховували третинні взаємодії, які також впливали на формування вторинної структури. У 1988 р. Лінн Ріган (Lynne Regan) і Вільям ДеГрадо (William DeGrado) сконструювали протеїн із чотирьох спіралей, розрахувавши його амінокислотну послідовність, виходячи з того, що зовнішня грань кожної  $\alpha$ -спіралі має бути гідрофільною, а внутрішня — гідрофобною [4]. Після цього протеїни з чотирма спіралями стали стандартними об'єктами для досліджень у галузі протеїнового дизайну.

Невдовзі стало зрозумілим, що для прогнозування структури більших за розміром і складніших протеїнів необхідно застосовувати повністю автоматизовані обчислення. Вперше це зробили у 1997 р. американські вчені Бассіл Дахіят (Bassil Dahiyat) і Стівен Мейо (Stephen Mayo) з Каліфорнійського інституту технологій у Пасадені, які розробили дизайн нового невеликого протеїну з просторовою структурою, подібною до так званого мотиву цинкового пальця (Znf), що входить до складу багатьох протеїнів, забезпечуючи зв'язування з певною цільовою молекулою (ДНК, РНК, протеїном або ліпідом). Цей новий протеїн мав 6 із 28 спільних з природним протеїном амінокислотних залишків і, на відміну від нього, не містив іонів  $Zn^{2+}$  [5]. Цей алгоритм обчислювального дизайну було засновано на функціях фізико-хімічного потенціалу та стереохімічних обмеженнях, і його використовували для скринінгу комбінаторної бібліотеки. Це дослідження стало важливою віхою в розвитку дизайну протеїнів, але розв'язання аналогічно-

го завдання для великого протеїну тоді було абсолютно нездійсненним.

У 1990-х роках з'явилося ще два цікавих підходи до вивчення та передбачення структури протеїнів. По перше, це було моделювання молекулярної динаміки (MD) згортання, в якому за допомогою методу грубої сили намагалися правильно згорнути розчинений протеїн, що перебував у розгорнутому стані. В 1998 р. Йонгу Дуану (Yong Duan) і Пітеру Колману (Peter Kollman) вдалося згорнути в такий спосіб невеликий протеїн усього за мікросекунду [6]. Другий підхід полягав у вирівнюванні множинних послідовностей (MSA), в якому порівнювали послідовності протеїнів певної родини та знаходили корельовані мутації, що виникли під час еволюції і стосувалися пар амінокислотних залишків, які взаємодіяли та розташовувалися близько один від одного в межах просторової структури [7]. Цей підхід вважали дуже перспективним, але ефективність передбачення просторової структури протеїну з використанням MSA залишалася невисокою.

Штучні нейронні мережі почали застосовувати для прогнозування вторинної структури ще наприкінці 1980-х років [8], але активне використання цих технологій почалося лише через 10 років, коли закінчилася так звана «зима ШІ» — період розчарувань і зниження інтересу вчених та інвесторів до цієї технології. Для прогнозування структури протеїнів розробили численні підходи моделювання на основі шаблонів (TBM), які ґрунтувалися на пошуку подібних структур у банку даних протеїнів, а також вільне моделювання (відоме як підхід *ab initio*), що використовувало наскрізні методи глибокого машинного навчання та обходилося без використання глобальних шаблонних структур. Тривалий час методи TBM вважали надійнішими, однак їхня точність значною мірою залежала від якості вирівнювання між цільовим протеїном та ідентифікованими шаблонами, яка своєю чергою часто визначалася еволюційними відмінностями між ними [1].

Для тестування різних методів прогнозування просторової структури протеїнів і стимулювання прогресу в цій галузі у 1994 р. було запо-



чатковано конкурс «Критична оцінка прогнозування структури протеїну» (CASP). Заснували його Джон Молт (John Moulton) з Університету Мериленду в Коледж-Парку та Кшиштоф Фіделіс (Krzysztof Fidelis) з Ліверморської національної лабораторії при Каліфорнійському університеті. З 2004 р. одним з організаторів конкурсу є українець Андрій Криштафович (Andriy Kryshchakovich) — вчений-біоінформатик родом із Коломиї, який закінчив Львівський національний університет, до 2000 р. працював у НАН України, а зараз є співробітником Каліфорнійського університету в Девісі<sup>6</sup>. Учасники конкурсу з усього світу кожні два роки намагалися за амінокислотною послідовністю передбачити структури протеїнів, які було визначено, але ще не було опубліковано. Протягом перших десяти років прогрес був незначним: найкращі результати демонстрували точність визначення до 40 %. Однак саме завдяки участі в цьому конкурсі нобелівські лауреати 2024 р. змогли продемонструвати свої визначні досягнення та здобути всесвітню славу.

Яким шляхом доля привела їх до перемоги і що відбувалося в їхньому житті на той момент?

У 2010 р. Деміс Гассабіс, здобувши науковий ступінь з нейронауки, заснував компанію DeepMind, щоб використати одержані знання для вдосконалення нейромереж. Співзасновниками компанії були його близькі друзі — нейробиологи Шейн Легг та Мустафа Сулейман. Невдовзі DeepMind стала відомою, досягши певних успіхів у розробленні моделей ШІ для популярних настільних ігор, і в 2014 р. її купив Google за \$400 млн за умови, що Гассабіс залишиться на посаді генерального директора [9]. Через два роки Google DeepMind потрапила до заголовків усіх газет світу після того, як її програма ШІ AlphaGo перемогла з рахунком 4:1 Лі Седола — південнокорейського 18-разового чемпіона світу зі старовинної настільної гри го, який три роки потому оголосив про завершення своєї кар'єри через бурхливий розвиток ШІ,

<sup>6</sup> Андрій Криштафович: «Успішне моделювання структури білків — найбільше досягнення науки в цьому столітті загалом». *Тиждень*. 27.01.2021. <http://surl.li/ilrcmo>

що, за його словами, є «сутністю, яку неможливо перемогти».

Об'єднання з Google пішло на користь компанії DeepMind і спрямувало її діяльність на розв'язання найважливіших наукових проблем. Цьому сприяли провідні вчені світу, що входили до складу корпоративної ради, зокрема фахівець у галузі інженерії протеїнів Френсіс Арнольд (Frances Arnold), яка отримала Нобелівську премію з хімії 2018 р. за проведену вперше спрямовану еволюцію ензимів. Невдовзі передбачення просторової структури протеїнів за допомогою обчислювальних методів із використанням ШІ стало одним із важливих напрямів діяльності компанії.

Перші успіхи було продемонстровано у 2018 р., коли на конкурсі CASP13 компанія Google DeepMind представила свою програму AlphaFold, яка значно випередила конкурентів, продемонструвавши точність близько 70 %. Однак надалі спроби вдосконалити програму зайшли в глухий кут. Саме в цей час у компанію прийшов Джон Джампер, який використав свій досвід у теоретичній фізиці та моделюванні протеїнів і запропонував повністю переробити алгоритм AlphaFold, створивши принципово нову програму — AlphaFold2. Ця програма для прогнозування ймовірної структури протеїну використовувала нещодавно розроблений тип нейронної мережі під назвою трансформатор (замість згортової нейронної мережі), а також глибоке машинне навчання на базі даних амінокислотних послідовностей і відомих протеїнових структур. Представлення AlphaFold2 на наступному конкурсі CASP14 у 2020 р. стало справжнім проривом, адже програма продемонструвала приголомшливий результат — точність передбачення структури протеїнів сягнула 90 %, що приблизно відповідало точності експериментального визначення за допомогою рентгенівської кристалографії або ЯМР-спектроскопії. Проблему визначення обчислювальними методами просторової структури протеїнів за амінокислотною послідовністю було розв'язано [10].

У 2021 р. компанія Google DeepMind на чолі з Гассабісом за допомогою AlphaFold2 роз-

рахувала просторові структури протеїнів 20 модельних організмів, таких як бактерії кишкової палички, дріжджі та плодові мушки, а також майже 44 % усіх протеїнів людини. Для решти протеїнів людини передбачувана структура часто визначалася як «невпорядкована», тобто ці протеїни мали кілька конформацій, які змінювалися при зв'язуванні з іншими молекулами<sup>7</sup>. Наступним кроком було обчислення структур «усього протеїнового всесвіту» — 200 млн відомих протеїнів, що походять приблизно з 1 млн біологічних видів. Гассабіс і Джампер розмістили цю інформацію у відкритій базі даних UniProt і оприлюднили вихідний код AlphaFold2. Зараз цією програмою користуються понад 2 млн людей зі 190 країн світу.

За допомогою AlphaFold2 визначено структуру багатьох корисних протеїнів, зокрема природних ензимів, які можуть розкладати пластик і зупинити екологічну катастрофу на Землі (2022); частини величезної молекулярної структури, що утворює пори в ядерній мембрані клітини та складається з понад тисячі протеїнів (2022); бактеріального ензиму, який спричиняє стійкість бактерій до антибіотиків, що є важливим для розроблення способів запобігання стійкості до антибіотиків (2023), тощо. У травні 2024 р. команда Джампера представила новий варіант програми — AlphaFold3, що може передбачати структуру не лише протеїнів, а й інших молекул, зокрема ДНК і РНК. Поки що код нової програми є закритим з комерційних міркувань, але програму планують зробити доступною для некомерційного використання.

Девіда Бейкера доля привела до галузі вивчення просторової структури протеїнів випадково. Він вступив до Гарвардського університету з бажанням вивчати філософію та соціальні науки, однак зацікавився клітинною біологією і структурою протеїнів, коли під час занять з еволюційної біології натрапив на підручник «Молекулярна біологія клітини». Наприкінці 1990-х років Бейкер почав розробляти комп'ютерне програмне забезпечення Rosetta для передбачення структури протеїнів,

яке досягло певного успіху на конкурсі CASP у 1998 р. [11], а через рік він очолив компанію Rosetta Commons. Створена програма давала можливість не лише прогнозувати просторову структуру протеїну за його амінокислотною послідовністю, а й ввести бажану структуру протеїну та отримати відповідну амінокислотну послідовність. Це було початком нової галузі науки — дизайну протеїнів. Спочатку Бейкер орієнтувався на просторові структури природних протеїнів, але згодом прийшло розуміння, що їхній репертуар дуже обмежений і теоретично можна створювати протеїни з просторовою структурою, яка ніколи не існувала, та функціями, яких ніколи не було в природі. Він досить оригінально проілюстрував цю думку: «Якщо ви хочете побудувати літак, ви не починаєте з модифікації птаха. Замість цього ви, розуміючи основні принципи аеродинаміки, будете на цих принципах машину, що літає»<sup>8</sup>.

В основу роботи алгоритму програми Rosetta було покладено ідею, яку висловлювали раніше Елвін Джонс (Thomas Alwyn Jones) і Сорен Тіруп (Søren Skou Thirup) [12], щодо побудови нового протеїну на основі коротких фрагментів, наявних у відомих структурах інших протеїнів. Rosetta використовувала оптимізацію Монте-Карло в розрахунках з енергетичною функцією, яка брала до уваги взаємодії Вандер-Ваальса, водневі зв'язки та ефекти сольватації, а можливі орієнтації бічних ланцюгів амінокислот програма відбирала з великої бібліотеки ротамерів. У результаті Rosetta генерувала багато можливих рішень, ранжувала їх за енергією та обирала найбільш енергетично вигідний варіант.

Цікаво, що Rosetta постійно розвивалася завдяки тому, що велика кількість користувачів і співрозробників брали участь у розподіленому обчислювальному проєкті Rosetta@home, а також в онлайн відеогрі-головоломці Foldit [13], яку Бейкер розробив разом з колегами із Центру ігрознавства Університету Вашингтона. Це дозволило людям без наукового досвіду

<sup>8</sup> Fernholm A. Popular science background: They have revealed proteins' secrets through computing and artificial intelligence. <https://surl.li/hyesju>

<sup>7</sup> Service R.F. New public database of AI-predicted protein structures could transform biology. <https://surl.li/uzztho>

допомагати розв'язувати проблеми, пов'язані зі структурою протеїнів. У такий спосіб до реальних досліджень широкого спектра медичних проблем залучили понад 400 тис. людей. Найуспішніших учасників включили як співавторів до опублікованих робіт Бейкера.

Девід Бейкер наполегливо проводив дослідження в галузі дизайну протеїнів, і результат не забарився. У 2003 р. його дослідницька група розробила абсолютно нову структуру протеїну Top7, якої в природі не існувало і яка складалася з двох  $\alpha$ -спіралей і  $\beta$ -листа з п'яти  $\beta$ -ланцюгів; програма Rosetta запропонувала його амінокислотну послідовність на основі пошуку коротких фрагментів у базі даних усіх відомих структур і оптимізації послідовності за законами енергетичного ландшафту протеїнів; далі в бактеріях було синтезовано відповідний рекомбінантний протеїн Top7 і досліджено його просторову структуру за допомогою рентгенівської кристалографії [14]. Результат перевершив усі сподівання: теоретична і реальна структури протеїну повністю збіглися. До того ж протеїн мав довжину 93 амінокислотні залишки, що на той час було рекордним значенням для протеїнів, одержаних за допомогою дизайну *de novo*.

У 2008 р. Бейкер здійснив першу спробу спроектувати *de novo* ензими, що каталізують реакції, для яких не існує ензимів у природі [15]. Одержані ензими були здатні підвищити швидкість каталізу хімічної реакції, але значно поступалися за швидкістю реакції природним ензимам. Зараз для вдосконалення цих ензимів застосовують технологію експериментальної спрямованої еволюції протеїнів, яку запропонувала Френсіс Арнольд. Значно кращих результатів вдалося досягти в дизайні протеїнів, що зв'язують ліганди. Зокрема, було отримано протеїни, які з високою спорідненістю і специфічністю зв'язують стероїди [16]. Застосування лабораторної еволюції зробило ці протеїни надзвичайно чутливими. Крім того, дизайн протеїнів було успішно застосовано в галузях розроблення протеїнових перемикачів і сенсорів для різних аналізаторів [17], а також створення наноматеріалів на основі протеїнів, які самостійно збираються в ікосае-

дричні вірусоподібні частинки [18]. Після дивовижного успіху на конкурсі CASP2020 програми AlphaFold2 Бейкер удосконалив свою програму Rosetta, запозичивши елементи нейромережі на основі трансформаторів, і назвав нову версію RoseTTAFold [19]. Девід Бейкер також відкрив код RoseTTAFold, щоб усі охочі могли його вдосконалювати та знаходити нові застосування для програми.

У 2017 р. Інститут дизайну протеїнів, який очолює Бейкер, отримав понад \$11 млн від фонду Open Philanthropy, а у 2021 р. — ще \$3 млн. За допомогою програми Rosetta та її оновленого варіанту отримано низку корисних протеїнів із запланованим дизайном, зокрема наноматеріал, який сам збирається зі 120 протеїнів (2016); протеїн, що зв'язується з опіюдом фентанілом і може використовуватися для його виявлення в навколишньому середовищі (2017); наночастинки з протеїнами на поверхні, що імітують вірус грипу і можуть бути вакциною проти грипу (2021); протеїни геометричної форми, які можуть змінювати свою форму внаслідок зовнішніх впливів і використовуватися для виготовлення крихітних датчиків (2024). Крім того, команда Бейкера, використовуючи RoseTTAFold, створила структурну базу даних із сотень G-протеїнових рецепторів, що є класичними мішенями для багатьох ліків. Серед потенційних продуктів Бейкер назвав універсальну вакцину проти грипу, знеболювальні засоби, які не викликають звикання, розумні протеїни, здатні ідентифікувати та лікувати ракові клітини або клітини, що вийшли з-під контролю та спричиняють аутоімунні розлади, потенційні засоби лікування нейродегенеративних розладів та протеїни, що самі збираються, для сонячних елементів або виробництва наноструктур.

Успіхи програм з використанням ШІ у передбаченні просторової структури протеїнів дали привід сподіватися, що вони здатні спростити процес розроблення ліків і заощадити 25—50 % часу та коштів на розроблення. Це неабияка економія, адже вартість виведення лікарського препарату на ринок оцінюють приблизно у \$2,5 млрд, а створення препарату може тривати 12—15 років. Водночас 9 з 10 препаратів, які

пройшли клінічні випробування, так і не отримують схвалення [20]. Але чи дійсно програми на кшталт AlphaFold або RoseTTAFold можуть стати революційним інструментом для відкриття ліків? Наскільки можна вірити передбаченням ШІ?

Як і в усіх методах машинного навчання, диявол криється в базі даних. Розробники AlphaFold визнали, що точність передбачення погіршується для протеїнів з неадекватним вирівнюванням кількох послідовностей через малу кількість гомологів. Наприклад, це характерно для трансмембранних протеїнів, яких у базі даних значно менше, ніж водорозчинних. Крім того, точність AlphaFold стає низькою для доменів протеїну, структура яких визначається не внутрішньодоменими взаємодіями, а взаємодією з іншими доменами. Слід пам'ятати, що протеїни є динамічними молекулами, і під час функціонування вони можуть мати кілька стабільних конформацій та переходити від однієї до іншої при зв'язуванні навіть невеликого ліганду, наприклад при зв'язуванні гемоглобіном кисню. Зрештою, програмне прогнозування не враховує наявність органічних кофакторів, які є важливими для функціонування багатьох ензимів. Програма RoseTTAFold досягла успіху в передбаченні структури подвійних або навіть потрійних комплексів, але олігомери вищого порядку становлять проблему, адже кількість можливих способів розташування субодиниць швидко зростає разом із кількістю субодиниць. До того ж вчені визнали той факт, що не всі протеїни згортаються в стабільні структури, і до 50 % протеїнів мають внутрішні порушення структури, які стабілізуються після зв'язування з цільовою поверхнею або утворення комплексу з іншими протеїнами. Ба більше, деякі протеїни ніколи не згортаються в жодну структуру, навіть після зв'язування з іншим протеїном. Усі ці складнощі потребують подальшого вдосконалення програм [21].

Біотехнологічна компанія Recursion у Солт-Лейк-Сіті (штат Юта) використала власну програму MatchMaker на основі ШІ, щоб спрогнозувати, як 36 млрд потенційних лікарських сполук або лігандів з бази даних Enamine Real Space

можуть зв'язуватися з понад 15 тис. протеїнів, просторову структуру яких було передбачено AlphaFold. У результаті було отримано неймовірну кількість даних, однак дослідники мають певні сумніви щодо їхньої якості. При моделюванні процесу «стикування» маленької молекули з передбаченою структурою протеїну виявилось, що приєднання до моделей AlphaFold є набагато менш точним, ніж до протеїнових структур, визначених експериментально. Можливо, причиною цього є невеликі варіації в орієнтації бічних ланцюгів амінокислот. Крім того, при взаємодії з молекулою структура протеїну трохи змінюється, що не враховує AlphaFold. Останній недолік було усунуто в програмі DragonFold, створеній компанією Charm Therapeutics на основі програми Бейкера RoseTTAFold. До того ж DragonFold підказує, як модифікувати потенційний препарат для створення більш міцного та селективного зв'язування з протеїном. Незважаючи на ці успіхи та величезні зусилля, яких докладають, щоб використовувати програми на основі ШІ для відкриття ліків, поки що можна лише зазначити, що ми перебуваємо на початку довгого шляху [22].

Інша проблема полягає в тому, що для навчання та вдосконалення здатності програм на основі ШІ розробляти ліки необхідно, щоб хтось синтезував і перевіряв молекули, які вони пропонують (може виявитися, що запропоновану сполуку неможливо синтезувати). Однак фармацевтичні компанії, які мають засоби для виготовлення та тестування сполук, тримають свої результати в таємниці для захисту від конкурентів [20].

Прикладом ліків, створених «з нуля» за допомогою ШІ, є препарат проти ідіопатичного легеневого фіброзу, розроблений компанією Insilico Medicine за 18 місяців, який зараз проходить II фазу клінічних випробувань у Китаї та Новій Зеландії. Співробітники компанії визначили цільовий ензим TNIK за допомогою системи ШІ PandaOmics Insilico, навченої на біомолекулярних даних пацієнтів і науковій літературі. Потім вони використали платформу молекулярного дизайну Chemistry42, щоб знайти молекулу для блокування цього ензиму.



Після тестування кількох кандидатів за допомогою радіометричного ензиматичного аналізу, а також внесення змін щодо поліпшення всмоктування, розподілу, метаболізму та виведення було отримано кінцевий варіант препарату. Видається захопливою ідея, що така система ШІ могла б вчитися на своїх попередніх циклах розроблення ліків і розв'язувати подібні завдання щоразу все швидше, перетворившись на конвеєр з розроблення ліків<sup>9</sup> [23].

З метою розвитку галузі використання ШІ для розроблення ліків було започатковано конкурс CASPE, аналогічний змаганню з передбачення просторової структури протеїнів. У цьому конкурсі взяли участь 23 команди, які передбачили понад 2 тис. сполук, що теоретично могли б зв'язуватися з протеїном WDR (доменом повтору WD40), який входить до складу багатофункціонального ензиму LRRK2, пов'язаного з хворобою Паркінсона. Для тестування в лабораторії відібрані на роль ліків кандидати на замовлення синтезував український постачальник малих молекул — компанія «Енамін» (м. Київ). Виявилось, що лише близько дюжини сполук були здатні зв'язуватися з ензимом (менш як 1 % результатів), однак ці малі молекули мали достатньо хороші профілі зв'язування і за своїми властивостями могли б бути ліками. Серед семи найкращих команд шість використовували ШІ і лише одна — набір класичних обчислювальних інструментів VirtualFlow для пошуку ліків. Наступні два конкурси планується присвятити пошуку інгібіторів гелікази NSP13 і макродомену NSP3 коронавірусу SARS-CoV-2 [24].

Попри наявні проблеми, вчені не полишають спроб застосовувати програми на основі ШІ для розроблення нових фармацевтичних продуктів, зокрема терапевтичних антитіл, вакцин і систем доставлення ліків у клітини. Так, нова програма RFDiffusion з використанням ШІ, що нагадує інструменти для створення зображень Midjourney і DALL-E, допомогла створити тисячі абсолютно нових антитіл, які розпізнають

<sup>9</sup> Walrath R. Insilico reveals a 'soup to nuts' process for AI-generated lung fibrosis drug. *Chemical & Engineering News*. 08.03.2024. <https://surl.li/oxksdf>

певні бактеріальні, вірусні або ракові мішені. Ці антитіла являли собою мініпротеїни, подібні до однодоменних антитіл верблюдів і акул. Хоча під час лабораторних випробувань лише одне зі 100 антитіл спрацювало належним чином, вчені вважають це знаковим моментом і вірять, що ШІ в недалекому майбутньому дозволить значно скоротити витрати на розроблення терапевтичних антитіл [25].

У лабораторії Бейкера для боротьби з COVID-19 за допомогою програми RoseTTAFold розробили назальний спрей, що містив синтетичні мініпротеїни (розміром 60 амінокислотних залишків), які взаємодіяли з рецептор-зв'язувальним доменом (RBD) S протеїну коронавірусу у 6 разів міцніше, ніж найкращі відомі антитіла. Крім того, дослідження Бейкера сприяли створенню його колегами з Медичної школи Університету Вашингтона першої у світі розробленої комп'ютером протеїнової вакцини проти COVID-19. Також було створено вакцини на основі наночастинок розміром з вірус, що самі збираються з протеїнів, створених за допомогою ШІ, та несуть на своїй поверхні прив'язані через короткі пептидні ланцюжки рецептор-зв'язувальні домени, за допомогою яких коронавірус чи респіраторно-синцитіальний вірус проникають у клітину. Вдосконалений варіант вакцини проти COVID-19 містив наночастинки з чотирма варіантами домену RBD від різних коронавірусів<sup>10</sup>.

Програму EVEscape на основі ШІ намагаються застосувати для прогнозування еволюції вірусів, зокрема коронавірусу SARS-CoV-2, з метою виробництва ефективних вакцин, які б діяли на випередження проти нових мутантних штамів. Прогнозування потенційних мутацій SARS-CoV-2 та інших небезпечних вірусів може допомогти передбачити їх майбутнє. У 2023 р. Національний інститут алергії та інфекційних захворювань США оголосив про виділення \$100 млн на створення подібних бібліотек вакцин, з яких можна за потреби швид-

<sup>10</sup> Jacobsen R. Artificial proteins never seen in the natural world are becoming new COVID vaccines and medicines. *Scientific American*. 01.07.2021. <https://surl.li/gvbhfw>

ко взяти і використати найбільш перспективний дизайн вакцини [26].

Фармацевтичні компанії намагаються використовувати ШІ для розроблення ліків, націлених не лише на протеїни, а й на молекули нуклеїнових кислот. Так, стартап Стенфордського університету Atomic AI нещодавно зібрав \$35 млн, щоб передбачити та ідентифікувати усі молекули інформаційної РНК організму людини, які можуть бути мішенню для таких терапевтичних препаратів [27]. Водночас китайська компанія Baidu Research запропонувала програмне забезпечення, що використовує методи комп'ютерної лінгвістики для розроблення мРНК-вакцин нового покоління зі складнішою структурою. Цей інструмент було застосовано для оптимізації принаймні однієї дозволеної РНК-вакцини — SW-BIC-213 проти COVID-19 [23].

Дослідники використовують програми на основі ШІ для запобігання пандеміям, створюючи вакцини проти нових збудників або проти нових штамів, які виникають у результаті еволюції відомих збудників. Так, програму AlphaFold було використано для прогнозування структури вірусного протеїну G, завдяки якому проникає в клітину геніпавірус Langya (LayV), виявлений у 2018 р. у Східному Китаї, де він передався людям від землерийок. Цей вірус викликає проблеми з диханням і лихоманку, а також є родичем надзвичайно смертоносних вірусів Nipah та Hendra, які спричиняють спалахи захворювань у домашніх тварин і можуть бути летальними для людей [26].

Програма AlphaFold також допомогла Метту Хіггінсу з Оксфорду з'ясувати структуру поверхневого протеїну Pfs48/45 гамет малярійного плазмодія — компонента нової вакцини проти малярії [28], яка у 2024 р. пройшла першу фазу клінічних випробувань. Очікується, що ця вакцина блокуватиме кілька стадій розвитку малярійного плазмодія і буде ефективнішою за єдину дозволу для використання у 2021 р. вакцину проти малярії — Mosquirix, розроблену бельгійським підрозділом британської фармацевтичної компанії GlaxoSmithKline.

Одним з улюблених застосувань програми AlphaFold Деміс Гассабіс назвав молекуляр-

ний шприц, розроблений Феном Чжаном з Інституту Бродів у Кембриджі (Массачусетс, США), для доставлення протеїнів у клітини людини [9]. Для цього Чжан запропонував використовувати системи позаклітинної скоротливої ін'єкції (eCIS) симбіотичних бактерій, які доставляють корисні протеїни в клітини своїх господарів — комах. За допомогою AlphaFold було ідентифіковано ділянку хвостового волокна, яка опосередковує взаємодію eCIS з клітинами хазяїна, що допомогло перенацілити його на рецептор епідермального фактора росту людини (EGFR) і доставити в EGFR-позитивні клітини людини. Таку систему можна націлити на будь-що: наприклад, на епітопи вірусів людини або маркерів раку, та доставляти у відповідні клітини людини протеїнові токсини для їх знешкодження. Нещодавно Чжан відкрив компанію Aera Therapeutics, яка розробляє системи доставки для генної РНК-терапії на основі протеїнових наночастинок (PNP), які самі збираються з власних протеїнів людини<sup>11</sup>.

ШІ впливає не лише на розроблення ліків і вакцин, а й на дослідження в галузі охорони здоров'я та розвиток персоналізованої медицини. Так, було створено програму AlphaMissense на основі ШІ, яка, використовуючи дані про просторову структуру протеїнів, обчислену AlphaFold, може точно передбачити, які мутації в протеїнах здатні спричинити захворювання. Виявилося, що з 70 млн міссенс-мутацій у геномі людини лише 32 % спричиняють захворювання, а решта, здається, не мають шкідливого впливу на здоров'я [29].

З появою методів вимірювання рівнів тисяч протеїнів у крихітних зразках крові з'явилися нові можливості для поглибленого вивчення стану здоров'я людини. Наприклад, компанія SomaLogic розробила засіб для вимірювання понад 10 тис. протеїнів, а прилад Thermo Fisher's Olink аналізує понад 5,4 тис. протеїнів у зразку об'ємом лише 2 мкл. Поява таких можливостей є особливо доречною з огляду на те, що

<sup>11</sup> Grinstein J.D. Nature's needle: Feng Zhang's team re-engineers bacterial "syringes" for programmable protein delivery. *Genetic Engineering & Biotechnology News*. 29.03.2023. <https://surl.li/tmxizg>

кількість протеїнів в організмі людини (понад 100 тис.) значно перевищує кількість генів протеїнів (20 тис.), а в результаті альтернативного сплайсингу утворюються сотні тисяч ізоформ (варіантів) протеїнів. Протягом останніх 10 років завдяки використанню ШІ відбувся значний прогрес у з'ясуванні ознак більшості поширених захворювань і оцінюванні ризиків їх виникнення шляхом аналізу інформації про однонуклеотидні поліморфізми (SNP), мінливість у протеомних, метаболомних та епігеномних профілях, а також у мікробіомі, імунотомі та експосомах, що відображають вплив зовнішнього середовища та способу життя. Це значно поглибило наше розуміння причин виникнення захворювань, процесу старіння, дозволило потенційно прогнозувати майбутні зміни стану здоров'я конкретної людини. Очевидно, що високопродуктивна протеомна оцінка стану організму з часом знайде своє місце в повсякденній персоналізованій медицині, незважаючи на поки що високу вартість (\$500—\$1000 за обстеження однієї людини) [30].

Використання методів глибокого навчання дало можливість комп'ютерам успішно виконувати завдання, з яким завжди краще справлялися люди: розпізнавати конкретні клітинні структури на тлі щільного безладного фону біологічного матеріалу. Спроби навчати навичці сегментації почалися з програми аналізу зображень DeepCell, а у 2015 р. німецькі дослідники створили програму U-Net, яка досі залишається основою архітектури для більшості інструментів сегментації. На базі цієї програми у 2019 р. у Центрі біологічних досліджень у Сегеді (Угорщина) було створено алгоритм pncleAIzer, який перевершив інші подібні інструменти. Зараз вчені працюють над новими викликами, такими як інтерпретація зображень об'ємної електронної мікроскопії, класифікація функціональних характеристик нейронів (гальмівних і збуджувальних) на основі їхньої морфології, профілювання молекулярної патології раку та визначення відповідних шляхів лікування. Можливо, найближчим часом морфологічний простір стане кількісним, і його поєднання з простором omics дасть нові і несподівані позитивні результати [31].

З появою 30 листопада 2022 р. безплатного чат-бота ChatGPT від компанії OpenAI у вечних з'явився інструмент, який дозволив значно швидше впоратися з написанням статей чи заявок на гранти. Вже через два місяці ChatGPT було зазначено як автора кількох дослідницьких статей, хоча він не задовольняв критерії авторства, оскільки не міг нести відповідальності за роботу. Лише за перші п'ять днів у ChatGPT було близько мільйона користувачів, а тепер їх кількість перевищує 180 млн. В опитуванні вечних, проведеному журналом Nature у 2023 р., 30 % респондентів відповіли, що використовували генеративні інструменти ШІ для написання рукописів, а 15 % — для написання заявок на гранти. Вже зараз наслідки гіперпродуктивності відчувають редакції наукових журналів, навантаження на які відчутно зросло. Можливо, це змусить перейти від культури «публікуйся або помри» до системи, яка віддає перевагу якості над кількістю, інакше доведеться зробити редагування статей платним, адже надійного способу контролювати використання технології ШІ поки що немає [32]. У 2023 р. Elsevier запустив пілотну версію програми Scopus AI, яка швидко аналізує та узагальнює наукову інформацію і може стати дуже корисним інструментом для створення оглядів. Найбільший огляд, створений людьми, охоплює близько 1600 статей, але генеративний ШІ здатний на більше, адже для нього це лише мізерна частка всієї наукової літератури [33].

Поява ChatGPT вразила світ і поклала початок гонитві у Big Tech (п'ятірці найбільших американських технологічних компаній: Facebook, Amazon, Apple, Netflix і Alphabet's Google), до якої долучаються і нові гравці. Amazon створює власного конкурента ChatGPT під кодовою назвою Olympus, який буде однією з найбільших будь-коли навчених моделей. Google DeepMind випустила чат-бот Med-PaLM, призначений для відповідей на медичні запитання [34]. Тим часом нова компанія Ілона Маска xAI представила пілотну модель ШІ Grok, яку навчали на даних із соціальної мережі X (Twitter) і програмували відповідати на запитання з допечністю та бунтарською жилкою.

Усі ці чат-боти засновані на великих мовних моделях (LLM) — алгоритмах, навчених на мільйонах прикладів тексту, зібраного з Інтернету. LLM вивчають статистичні зв'язки між словами, а отримавши підказку, наприклад запитання, вони генерують текст, вгадуючи за ймовірністю кожне наступне слово. Створення LLM стало можливим завдяки тому, що у 2017 р. дві дослідницькі групи виявили спосіб перекладу між людськими мовами без необхідності використання Розеттського каменю. Відкриття ґрунтувалося на перетворенні семантичних зв'язків між словами на геометричні. Моделі машинного навчання тепер здатні перекладати між різними людськими мовами, вирівнюючи їхні форми, наприклад використовуючи частоту, з якою такі слова, як «мама» і «дочка», з'являються поруч одне з одним, щоб точно передбачити, що буде далі. У 2020 р. ця сфера досягла ще однієї віхи, коли обробка природної мови здатна «розглядати все як мову». Наприклад, система DALL-E 2 може генерувати реалістичні зображення на основі словесних описів з надзвичайною точністю. Для мовної моделі, що використовує метод отримання статистичних асоціацій, не важливо, якою мовою її навчають. Тоді її можна навчати «мовою» послідовностей ДНК або протеїнів і замість нового речення отримати нові протеїни, що є гарними мішенями для ліків. Крім того, ШІ нарешті може дозволити нам розмовляти з тваринами. Зараз вчені використовують машинне навчання, щоб спробувати зрозуміти вокалізацію кашалотів, крики ворон тощо<sup>12</sup> [34].

ШІ проникає не лише у всі сфери життя, а й у мозок людини. Найбагатша людина світу, бізнесмен і технолог Ілон Маск поставив собі амбітне завдання — поєднати ШІ та людський мозок, для чого у 2016 р. він заснував компанію Neuralink. Нещодавно Маск у соціальній мережі X (Twitter) оголосив, що Neuralink уперше успішно імплантувала в мозок людини бездротовий пристрій Telepathy, який містить комп'ютерний чип і масиви електродів із 1024

надтонких, гнучких провідників, які робот-хірург протягує в кору головного мозку. Електроди, призначені для реєстрації думок, пов'язаних з рухом, мають забезпечувати можливість керувати комп'ютером за допомогою мислення. Це дослідження проводили в межах схвалених FDA у травні 2023 р. клінічних випробувань для людей з тетраплегією (паралічем усіх чотирьох кінцівок). Поки що новина викликала неоднозначну реакцію суспільства через відсутність інформації про особу, яка отримала імплантат, та її медичний стан, а також через те, що можливість керування комп'ютерною мишкою силою думки за допомогою інших пристроїв раніше вже було продемонстровано, і вона не є чимось надзвичайним. Так, у 2004 р. пристрій VCI компанії Synchron було імплантовано у кровоносну судину головного мозку, де він реєстрував усереднене спрацьовування нейронних популяцій. Можливо, у випадку Neuralink єдиною новою проблемою, що вимагала непростого рішення за участі ШІ, була необхідність відведення сигналів від понад 1000 електродів і аналізу великої кількості інформації. Висловлювалися сподівання, що ця технологія допоможе сліпим людям бачити, а глухим чути, але імітування сенсорного стимулювання мозку через електроди — це поки що надзвичайно складне завдання. З іншого боку, природа науки така: ніколи не знаєш, що чекає за рогом. Вже сталося багато речей, які ніхто б ніколи не передбачив<sup>13</sup>.

Можливості програм з використанням ШІ вражають, однак їх усе частіше критикують за те, що вони нерідко розуміють щось відверто неправильно або просто вигадують (так звані галюцинації ШІ), через що часто виникають скандали та/або непорозуміння. Так, чат-боти служби підтримки клієнтів постійно створюють для своїх компаній різні проблеми, а чат Microsoft Bing порадив репортеру New York Times покинути свою дружину. Схильність ШІ до вигадування інформації може призвести до потоку рукописів із помилками, а можливо, навіть з підробками. До того ж моделі ШІ за-

<sup>12</sup> Parshley L. Artificial intelligence could finally let us talk with animals. *Scientific American*. 01.10.2023. <https://surl.li/aoquqk>

<sup>13</sup> Guarino B. Elon Musk's Neuralink has implanted its first chip in a human brain. What's next? *Scientific American*. 30.01.2024. <https://surl.li/caqazx>



лучають інтернет-контент, не турбуючись про упередженість, згоду чи авторські права, їх використання дехто вважає «автоматизованим плагіатом за задумом» [33].

Попри численні докази того, що ШІ важко контролювати і він часто поводить ся непередбачувано, технологічні компанії поспішають із запуском нових продуктів на основі ШІ. Насправді ніхто точно не знає, як працює глибоке навчання, що лежить в основі сучасного ШІ. Незрозуміло, як великі мовні моделі, такі як Gemini чи GPT-4, можуть навчитися робити те, чому їх не вчили. Можна навчити мовну модель на математичних задачах англійською, а потім показати їй французьку літературу, і на основі цього вона зможе навчитися розв'язувати математичні задачі французькою. Навіть сама назва «штучний інтелект» вводять людей в оману. Мовні моделі лише здаються розумними, тому що вони генерують текст, передбачаючи наступне слово в реченні. Проведене дослідження показало, що незначні зміни в підказках, які людина дає ChatGPT (наприклад, додавання пробілу, вибір іншого формату чи привітання на початку тексту), можуть призвести до величезних змін у відповіді чат-бота, що було названо «ефектом метелика»<sup>14</sup>. Отже, не слід переоцінювати ШІ, принаймні поки дослідження механізмів його роботи перебувають на початковій стадії.

Інша проблема, пов'язана з ШІ, стосується галузі біобезпеки. З одного боку, стрімкий розвиток ШІ значно полегшив роботу вчених у галузі молекулярної біології, а з іншого боку, — спричинив певне занепокоєння, адже програми на його основі можуть сприяти розробленню біологічної зброї, зокрема нових токсинів або вірусів, що легко передаються. Неправильно згорнуті протеїни (пріони), які викликають нейродегенеративні захворювання, можна поширювати за допомогою аерозолу і викликати неправильне згортання відповідних протеїнів в організмі. Конвенція про біологічну зброю, яку підписали практично всі країни світу, забороняє розроблення або використання біологічної зброї на основі патогенів, але ніхто

<sup>14</sup> Plumb T. Why LLMs are vulnerable to the 'butterfly effect'. *Venture Beat*. 23.01.2024. <https://surl.li/diazaa>

не додумався поширити її дію на протеїни, що ніколи не існували в природі. Останнім часом лунають заклики до вчених посилити контроль власних досліджень і поліпшити скринінг синтезованої ДНК, яка може бути джерелом інформації для потенційно шкідливих протеїнів. Щоб проаналізувати можливість зловмисного використання дизайнерських протеїнів, Інститут дизайну протеїнів Бейкера при Університеті Вашингтона у 2023 р. організував саміт з безпеки ШІ, на якому висловлювали пропозиції щодо створення експертного комітету для перевірки програмного забезпечення або заходів державного регулювання. Однак досі залишається незрозумілим, як можна регулювати інструменти дизайну протеїнів, зважаючи на швидкий темп їх розвитку. Принаймні зараз потенційні переваги дизайну протеїнів за допомогою ШІ значно перевищують небезпеки [35], та й використання штучних протеїнів для лихих справ — це поки що надто витончено, адже є багато більш доступних для цього засобів. Водночас використання ШІ в науках про життя вже зараз розглядають як можливу технологію «подвійного використання».

Існує ймовірність, що завдяки ШІ вартість інформації та інтелекту знизиться майже до нуля, прискорюючи темп інновацій і глобальних змін, що спричинить значну трансформацію суспільства. З початком бурхливого розвитку ШІ постало безліч актуальних питань, на які поки що немає остаточної відповіді. Як ми можемо контролювати ШІ? Як ми можемо забезпечити справедливий розподіл його благ? Як ми можемо отримувати користь від цієї дивовижної технології та зменшити ризики? І взагалі, що далі?

Минуло 74 роки відтоді, як відомий британський вчений у галузі обчислювальної техніки та математики Алан Тюрінг запропонував гру «імітація» для перевірки мислення комп'ютерів. Його фундаментальна стаття 1950 р. починалася словами: «Пропоную розглянути питання «Чи можуть машини мислити?» [36]. Гра в імітацію полягала в намаганнях виявити комп'ютер за допомогою текстових розмов з ним та іншими людьми. Якщо це не вдавалося, можна було вва-

жати, що машина може мислити. Гра захопила увагу публіки і стала відомою як тест Тюрінга. Всупереч популярності цієї ідеї, вважають, що цей тест є надто розпливчастим і зосередженим на обмані, тому він не може стати серйозним дослідницьким інструментом або метою для ШІ. Однак питання про те, яку роль може відігравати мова в оцінюванні та створенні інтелекту, актуальне сьогодні як ніколи через появу LLM. Завдяки їхньому неперевершеному хисту вести розмову, а також низці інших можливостей, зокрема написанню есе та віршів, програмних кодів, складанню іспитів і узагальненню текстів, ці боти викликали хвилювання щодо перспектив розвитку ШІ та значення його появи для людства. ChatGPT пройшов тест Тюрінга, і тому зараз шукають нові способи оцінити здатність ШІ мислити. Як LLM роблять усі ці неймовірні речі? Як і в інших нейронних мережах, значна частина дій LLM є наслідком їх навчання, а не програмування. Тому механізми, які лежать в основі їхньої поведінки, невідомі навіть їхнім творцям і мають бути терміново досліджені за допомогою нових тестів. Це завдання є надзвичайно важливим, хоча воно схоже на дослідження «інопланетного розуму». Системи ШІ повинні отримувати схвалення для певних застосувань, зокрема стосовно лікарських засобів вони мають повідомити користувачу інформацію про можливі побічні ефекти та небезпечність використання у деяких випадках. Так, вчені з Інституту Санта-Фе в Нью-Мексико створили серію візуальних головоломок ConceptARC для перевірки здатності систем ШІ оперувати абстрактними поняттями (спойлер: GPT-4 впорався з цим погано). Іншою проблемою є відсутність прозорості з боку компаній, які створюють LLM, щодо розкриття того, на яких моделях даних було навчено ШІ. Проте перевірка цих компаній з боку регуляторних органів посилюється, і це може змусити їх у майбутньому розкривати більше таких даних. Зараз важко уявити важливіший виклик для галузі ШІ, ніж розуміння сильних і слабких сторін ШІ та механізмів, які ними керують [37].

Поки тривають дискусії про доцільність використання ШІ та можливості його регулюван-

ня, ця технологія впевнено захоплює всі галузі діяльності людини, передусім науку. Досягнення Девіда Бейкера, Деміса Гассабіса та Джона Джемпера в галузі обчислювального проектування протеїнів і прогнозування їхньої структури відкрили нову еру біохімічних і біологічних досліджень, а їх вплив матиме далекосяжні наслідки для людства. Штучні протеїни, яких ніколи не існувало в природі, входять тепер до складу нових вакцин і ліків. Завдяки розумінню механіки згортання протеїнів відкриваються перспективи дослідження етіології таких захворювань, як муковісцидоз, хвороби Альцгеймера, Паркінсона та Гантінгтона, а також можливості розроблення нових класів ліків, які могли б виправляти або замінювати дефектні протеїни. З нових протеїнів можна побудувати мікроскопічні «наноклітки» для пакування та транспортування ліків в організмі або молекулярні детектори, які спрацюють, коли стикаються зі специфічними протеїновими маркерами, наприклад на поверхні ракових клітин. Розробляють незвичний спосіб виробництва ефективних надпровідників і батарей, що ґрунтується на протеїнах, які самі збираються в мікроскопічні стільникові сітки та стимулюють відкладення мінералів. Інші протеїни поглинають енергію світла, як це відбувається під час фотосинтезу в рослинах, і перетворюють її на електроенергію чи паливо. Можливості використання штучних протеїнів досягають астрономічних висот, а наші найкращі мислителі та суперкомп'ютери не в змозі осягнути ці перспективи. Вже зараз природні протеїни, які ми використовуємо, наприклад інсулін для лікування діабету, здаються такими ж архаїчними, як заточені камені, які колись використовували наші предки, що жили у кам'яному віці. І якими б захопливими не були нинішні дизайнерські протеїни, ймовірно, що в майбутньому вони будуть сприйматися так, як у наш час сприймаються сонячні годинники чи колеса возу. Ніхто зараз не здатен уявити собі повною мірою новий майбутній світ, наповнений унікальними молекулами, створеними людиною [37], адже дуже складно дивитися вперед у нову епоху, не озираючись на можливі

ризики, про які попереджав відомий фізик Стівен Гокінг у своїй книзі «Короткі відповіді на великі запитання», що вийшла у 2018 р. через кілька місяців після його смерті: «Ми можемо зіткнутися з вибухом інтелекту, який зрештою призведе до появи машин, інтелект яких пере-

вищує наш набагато більше, ніж наш — равликів. ...Вам спокусливо відкинути поняття про високоінтелектуальні машини просто як наукову фантастику, але це було б помилкою — і, можливо, нашою найгіршою помилкою в історії».

## REFERENCES

- Huang B. et al. Protein Structure Prediction: Challenges, Advances, and the Shift of Research Paradigms. *Genomics, Proteomics & Bioinformatics*. 2023. **21**(5): 913—925. <https://doi.org/10.1016/j.gpb.2022.11.014>
- Pearce R., Zhang Y. Toward the solution of the protein structure prediction problem. *J. Biol. Chem.* 2021. **297**(1): 100870. <https://doi.org/10.1016/j.jbc.2021.100870>
- Chou P.Y., Fasman G.D. Prediction of protein conformation. *Biochemistry*. 1974. **13**(2): 222—245. <https://doi.org/10.1021/bi00699a002>
- Regan L., DeGrado W.F. Characterization of a helical protein designed from first principles. *Science*. 1988. **241**(4868): 976—978. <https://doi.org/10.1126/science.3043666>
- Dahiyat B.I., Mayo S.L. De novo protein design: fully automated sequence selection. *Science*. 1997. **278**(5335): 82—87. <https://doi.org/10.1126/science.278.5335.82>
- Duan Y., Kollman P.A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*. 1998. **282**(5389): 740—744. <https://doi.org/10.1126/science.282.5389.740>
- Göbel U., Sander C., Schneider R., Valencia A. Correlated mutations and residue contacts in proteins. *Proteins*. 1994. **18**(4): 309—317. <https://doi.org/10.1002/prot.340180402>
- Qian N., Sejnowski T.J. Predicting the secondary structure of globular proteins using neural network models. *J. Mol. Biol.* 1988. **202**(4): 865—884. [https://doi.org/10.1016/0022-2836\(88\)90564-5](https://doi.org/10.1016/0022-2836(88)90564-5)
- Sklar J. QnAs with Demis Hassabis and John M. Jumper: Winners of the 2023 Albert Lasker Basic Medical Research Award. *Proc. Natl. Acad. Sci.* 2023. **120**(39): e2313816120. <https://doi.org/10.1073/pnas.2313816120>
- Tunyasuvunakool K. et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021. **596**(7873): 590—596. <https://doi.org/10.1038/s41586-021-03828-1>
- Simons K.T., Bonneau R., Ruczinski I., Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins*. 1999. **3**: 171—176. [https://doi.org/10.1002/\(sici\)1097-0134\(1999\)37:3+<171::aid-prot21>3.3.co;2-q](https://doi.org/10.1002/(sici)1097-0134(1999)37:3+<171::aid-prot21>3.3.co;2-q)
- Jones T.A., Thirup S. Using known substructures in protein model building and crystallography. *EMBO J.* 1986. **5**(4): 819—822. <https://doi.org/10.1002/j.1460-2075.1986.tb04287.x>
- Cooper S., Khatib F., Treuille A., Barbero J., Lee J., Beenen M., Leaver-Fay A., Baker D., Popović Z., Players F. Predicting protein structures with a multiplayer online game. *Nature*. 2010. **466**(7307): 756—760. <https://doi.org/10.1038/nature09304>
- Kuhlman B., Dantas G., Ireton G.C., Varani G., Stoddard B.L., Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003. **302**(5649): 1364—1368. <https://doi.org/10.1126/science.1089427>
- Röthlisberger D. et al. Kemp elimination catalysts by computational enzyme design. *Nature*. 2008. **453**(7192): 190—195. <https://doi.org/10.1038/nature06879>
- Tinberg C.E. et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*. 2013. **501**(7466): 212—216. <https://doi.org/10.1038/nature12443>
- Langan R.A. et al. De novo design of bioactive protein switches. *Nature*. 2019. **572**(7768): 205—210. <https://doi.org/10.1038/s41586-019-1432-8>
- Bale J.B., Gonen S., Liu Y., Sheffler W., Ellis D., Thomas C., Cascio D., Yeates T.O., Gonen T., King N.P., Baker D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science*. 2016. **353**(6297): 389—394. <https://doi.org/10.1126/science.aaf8818>
- Humphreys I.R. et al. Computed structures of core eukaryotic protein complexes. *Science*. 2021. **374**(6573): eabm4805. <https://doi.org/10.1126/science.abm4805>
- AI's potential to accelerate drug discovery needs a reality check. *Nature*. 2023. **622**(7982): 217. <https://doi.org/10.1038/d41586-023-03172-6>
- Agard D.A., Bowman G.R., DeGrado W., Dokholyan N.V., Zhou H.X. Solution of the protein structure prediction problem at last: crucial innovations and next frontiers. *Fac. Rev.* 2022. **11**: 38. <https://doi.org/10.12703/r-01-0000020>

22. Arnold C. AlphaFold touted as next big thing for drug discovery — but is it? *Nature*. 2023. **622**(7981): 15—17. <https://doi.org/10.1038/d41586-023-02984-w>
23. Ren F. et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat. Biotechnol.* 2024. <https://doi.org/10.1038/s41587-024-02143-0>
24. Mullard A. When can AI deliver the drug discovery hits? *Nat. Rev. Drug Discov.* 2024. **23**(3): 159—161. <https://doi.org/10.1038/d41573-024-00036-0>
25. Callaway E. ‘A landmark moment’: scientists use AI to design antibodies from scratch. *Nature*. 2024. Mar 19. <https://doi.org/10.1038/d41586-024-00846-7>
26. Callaway E. How AlphaFold and other AI tools could help us prepare for the next pandemic. *Nature*. 2023. **622**(7983): 440—441. <https://doi.org/10.1038/d41586-023-03201-4>
27. Dolgin E. ‘Remarkable’ AI tool designs mRNA vaccines that are more potent and stable. *Nature*. 2023. May 2. <https://doi.org/10.1038/d41586-023-01487-y>
28. Ko K.T. et al. Structure of the malaria vaccine candidate Pfs48/45 and its recognition by transmission blocking antibodies. *Nat. Commun.* 2022. **13**(1): 5603. <https://doi.org/10.1038/s41467-022-33379-6>
29. Callaway E. AlphaFold tool pinpoints protein mutations that cause disease. *Nature*. 2023. Sep 19. <https://doi.org/10.1038/d41586-023-02943-5>
30. Topol E.J. The revolution in high-throughput proteomics and AI. *Science*. 2024. **385**(6716): eads5749. <https://doi.org/10.1126/science.ads5749>
31. Eisenstein M. AI under the microscope: the algorithms powering the search for cells. *Nature*. 2023. **623**(7989): 1095—1097. <https://doi.org/10.1038/d41586-023-03722-y>
32. Prillaman M. Is ChatGPT making scientists hyper-productive? The highs and lows of using AI. *Nature*. 2024. **627**(8002): 16—17. <https://doi.org/10.1038/d41586-024-00592-w>
33. Conroy G. How ChatGPT and other AI tools could disrupt scientific publishing. *Nature*. 2023. **622**(7982): 234—236. <https://doi.org/10.1038/d41586-023-03144-w>
34. Savage N. Drug discovery companies are customizing ChatGPT: here’s how. *Nat. Biotechnol.* 2023. **41**(5): 585—586. <https://doi.org/10.1038/s41587-023-01788-7>
35. Callaway E. Could AI-designed proteins be weaponized? Scientists lay out safety guidelines. *Nature*. 2024. **627**(8004): 478. <https://doi.org/10.1038/d41586-024-00699-0>
36. Turing A.M. Computing machinery and intelligence. *Mind*. 1950. **LIX**(59): 433—460. <https://doi.org/10.1093/mind/LIX.236.433>
37. ChatGPT is a black box: how AI research can break it open. *Nature*. 2023. **619**(7971): 671—672. <https://doi.org/10.1038/d41586-023-02366-2>

Svitlana I. Romaniuk

*Palladin Institute of Biochemistry of the National Academy of Sciences of Ukraine, Kyiv, Ukraine*

ORCID: <https://orcid.org/0000-0002-3900-6755>

Serhiy V. Komisarenko

*Palladin Institute of Biochemistry of the National Academy of Sciences of Ukraine, Kyiv, Ukraine*

ORCID: <https://orcid.org/0000-0002-3244-3194>

#### A BREAKTHROUGH IN SPATIAL STRUCTURE PREDICTION AND COMPUTATIONAL DESIGN OF PROTEINS, OR WHETHER WE CAN TRUST ARTIFICIAL INTELLIGENCE PREDICTIONS

The authors of the article analyze the 2024 Nobel Prize in Chemistry, which was awarded to American biochemist and computational biologist David Baker for “computer-aided protein design”, as well as to representatives of Google DeepMind: British artificial intelligence systems specialist Demis Hassabis and American chemist and computer scientist John M. Jumper for “protein structure prediction”. The achievements of the Nobel laureates in the field of computational protein design and structure prediction have opened a new era of biochemical and biological research, which, combined with the use of artificial intelligence tools, will have far-reaching consequences for humanity.

**Keywords:** Nobel Prize in Chemistry 2024, David Baker, Demis Hassabis, John M. Jumper, protein design, protein structure prediction.

**Cite this article:** Romaniuk S.I., Komisarenko S.V. A breakthrough in spatial structure prediction and computational design of proteins, or whether we can trust artificial intelligence predictions? *Visn. Nac. Akad. Nauk Ukr.* 2025. (2): 16—35. <https://doi.org/10.15407/visn2025.02.016>