

AlfaFold și Premiul Nobel în Chimie (2024)

Catalin VRABIE / 21.03.2025

Dacă în articolul precedent am vorbit despre Premiul Nobel în Fizică și legătura acestei științe cu inteligența artificială (AI) [1], acum este momentul să discut despre Premiul Nobel în Chimie și legătura dintre chimie și AI. Acordat la o zi după Premiul Nobel în Fizică, laureații Premiului Nobel în Chimie sunt David Baker de la Universitatea *Washington* din *Seattle*, Statele Unite și Demis Hassabis, alături de John Jumper ambii de la *Google DeepMind*, Londra, Marea Britanie. Fascinant este faptul că deși acest premiu, la drept vorbind, nu implică foarte multă chimie, el o susține dintr-o serie întreagă de motive: înțelegerea vieții, tratarea diferitelor boli și multe alte aspecte.

Un progres esențial adus de AI (și, aș putea spune, chiar revoluționar nu numai pentru chimie, ci și pentru biologie și științele medicale) este capacitatea de a vizualiza rapid structura proteinelor și de a crea proteine noi. Prin dezvoltarea aplicației *AlphaFold*, laureații din acest an au făcut posibilă prezicerea formei proteinelor – molecule biologice prezente în toate organismele și care sunt codificate de ADN [2], facilitând astfel înțelegerea modului în care funcționează elementele fundamentale vieții.

David Baker

“for computational protein design”



© Nobel Prize Outreach. Photo: Clément Morin

Demis Hassabis

“for protein structure prediction”



© Nobel Prize Outreach. Photo: Clément Morin

John Jumper

“for protein structure prediction”



© Nobel Prize Outreach. Photo: Clément Morin

Sursa: nobelprize.org

Problema intelectuală cunoscută sub denumirea de *Protein Folding Problem* [3, 4, 5], supranumită și *The Grand Challenge in Biochemistry*¹, a fost formulată de laureatul Nobel Christian Anfinsen încă din 1972 și se referă la modul în care proteinele, alcătuite din sute de lanțuri de aminoacizi, se pliază într-o structură tridimensională specifică [6, 7]. Anfinsen a emis celebra sa conjectură conform căreia această structură ar trebui să fie teoretic posibil de determinat, spunând că o proteină ar lua în mod unic forma care minimizează energia liberă din sistem. Cu toate acestea, numărul de configurații posibile pe care o proteină le poate adopta este uriaș² [6], depășind chiar numărul estimat al atomilor din univers – o idee formulată în 1969 de biologul american Cyrus Levinthal în lucrarea *How to Fold Graciously*, lucrare cunoscută ulterior sub numele de „Paradoxul Levinthal” [8].

Astăzi, știința a avansat semnificativ în înțelegerea mecanismelor prin care proteinele își dobândesc forma specifică, iar Premiul Nobel a fost acordat pentru două abordări distincte ale acestei probleme fundamentale [2].

¹ Marea provocare a biochimiei.

² Estimarea este de 10^{300} .

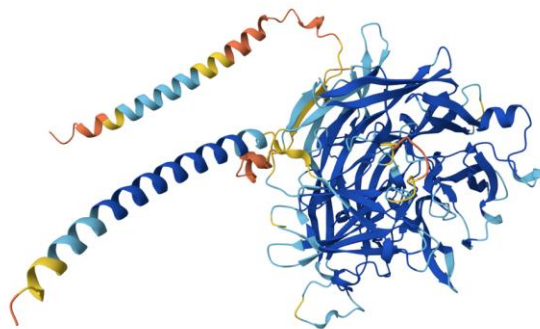
David Baker a utilizat algoritmi *deep learning* – tehnologii care se bazează pe cercetările făcute de Geoffrey Hinton [1]. Ceea ce a făcut el a fost să folosească astfel de tehnologii pentru a permite proiectarea unor proteine care să se plieze într-o anumită formă, în special pentru lanțuri scurte [2].

Echipa de la *DeepMind* a abordat problema din cealaltă direcție: dată fiind o secvență lungă de aminoacizi, ea și-a propus să prezică forma finală în care se va plia proteina. Și aici au fost utilizate tehnologiile *deep learning*, dar printr-o abordare computațională diferită. Un element care a contribuit substanțial la progresul celor două echipe a fost existența unei competiții anuale intitulată *Critical Assessment of Protein Structure Prediction (CASP)*, finanțată și organizată începând cu 1994 de *US National Institute of General Medical Sciences (NIH/NIGMS)*; în cadrul acesteia sunt selectate aproximativ o sută de proteine ale căror structuri sunt cunoscute prin experimente precise cu raze X sau microscopie electronică dar nepublicate încă [2, 9]. Concursul îi provoacă pe informaticienii să dezvolte algoritmi de predicție a acestor structuri pe baza secvenței de aminoacizi, urmând apoi să trimită rezultatele predicțiilor spre evaluare (prin comparare), cel mai bun algoritm fiind declarat câștigător.

Dacă pe informaticieni îi interesează dezvoltarea aplicațiilor, biochimistii vor să înțeleagă cum se pliază proteinele – mai mult decât din ce sunt compuse și ce conțin, deoarece, fără acest aspect nu le pot înțelege nici funcția, nici rolul. Și cum într-un organism există milioane de proteine, chiar dacă le este cunoscută compoziția (practic, secvența de aminoacizi), fără analiza formei, nu le poate fi dedusă funcția din interiorul corpului uman. Astăzi însă, problema a fost rezolvată. Soluția nu a venit din biochimie, dar având în vedere amploarea descoperirii, experții domeniului nu au găsit deloc inadecvată decizia Comitetului Nobe; în definitiv, toți oamenii de știință folosesc computere într-un fel sau altul, iar faptul că aici a fost utilizată inteligența artificială nu a produs îngrijorări. Pentru *DeepMind*, acest proces a fost asemănător cu analizarea unei serii de partide de șah, deducând regulile pe baza observațiilor făcute și aplicându-le ulterior în contexte noi. Descoperirea cu adevărat importantă a avut loc în 2020, în timpul pandemiei, când echipa *DeepMind*, condusă de cei doi laureați, a reușit să prezică cu precizie¹ 90% din structura proteinelor – o realizare extraordinară premiată la *CASP*². De atunci, aproximativ două milioane de cercetători au utilizat algoritmi lor, iar numărul structurilor de proteine înțelese a crescut de aproximativ o mie de ori ajungându-se astăzi la aproximativ două sute de milioane.



Q8W3K0: o potențială proteină rezistentă la bolile plantelor



Q8I3H7: o proteină ce poate proteja sistemul imunitar împotriva parazitului malariei

Sursa: alphafold.ebi.ac.uk

Aplicația, denumită așa cum am precizat, *AlphaFold*³, este o realizare ingenioasă a ingineriei rețelelor neuronale⁴, care a demonstrat o performanță fără precedent în prezicerea structurii proteinelor. Folosind informații bogate de date experimentale, stocate în ceea ce se numește *Protein Data Bank (PDB)* – o bază de date ce conține aproximativ două sute treizeci de mii de structuri și secvențe

¹ Eroarea minimă acceptată trebuia să fie mai mică decât dimensiunea unui atom (<1,00Å).

² Pentru premiul cel mare condiția impusă de organizatorii concursului era ca predicția făcută să aibă o acuratețe de minim 90%, valoare care nu fusese niciodată atinsă până atunci.

³ În realitate, aplicația premiată de *CASP* a fost *AlphaFold2* – dar am considerat acest amănunt mai puțin important pentru scopul prezentului articol.

⁴ În vederea înțelegerii conceptului, recomand spre citire articolul „Inteligența Artificială și Premiul Nobel în Fizică (2024)” publicat anterior în revista *All in on Tech (AloT)* [1].

cunoscute experimental [10], algoritmi sunt antrenați să descopere corelații și modele între secvențele de aminoacizi [11]. Acest lucru îi permite să producă modele structurale de o precizie uimitoare, direct din secvențe. *It is truly no exaggeration to say that AlphaFold has caused a revolution in structural biochemistry*¹ – menționează Prof. Johan Åqvist în cadrul Festivității de Decernare a premiilor Nobel 2024 [5].

Lăsând la o parte premiul primit, important este că de acum, într-un timp relativ scurt, cercetătorii pot avea o idee foarte bună despre structura unei proteine, fără a fi nevoie să realizeze experimente complicate și de durată. Chiar dacă uneori predicțiile nu sunt perfecte, situația este infinit mai bună decât înainte când, de exemplu, determinarea structurii hemoglobinei² de către un alt laureat Nobel, Max F. Perutz (1962), a durat treizeci și trei de ani pentru o singură moleculă [12, 13]. Acum, în doar patru ani, s-au analizat, așa cum am menționat, peste două sute de milioane de proteine [14] – bază de date care, trebuie spus, a fost făcută publică, pe internet, de către *DeepMind*. Este, fără îndoială, un salt uriaș³.

Mai mult, David Baker a realizat un alt progres semnificativ: a proiectat pe calculator o proteină complet nouă – una care nu exista în natură înainte. În vederea testării, colaboratorii săi au creat noul lanț, iar proteina s-a pliat așa cum a fost prezis de aplicația *software* folosită. Nu a fost o potrivire perfectă, dar a fost extraordinar de aproape, ceea ce a dus mai departe la crearea unei varietăți imense de structuri pentru diverse alte funcții decât cea la care s-a lucrat inițial [15] – precum nanomateriale și senzori.

Astfel, datorită muncii acestor trei oameni de știință, există acum o nouă metodă de abordare a problemei plierii proteinelor.

Demis Hassabis, liderul echipei *DeepMind* și fondatorul *Google DeepMind*, a declarat în cadrul discuțiilor *Nobel Minds* că are două mari ambiții: prima este să rezolve problema inteligenței (interese ce se suprapun și celor ale lui Ilya Sutskever despre care am vorbit în articolul „Superinteligența – una dintre cele mai mari provocări tehnice ale momentului” [16]), iar apoi să folosească această cunoaștere pentru a rezolva toate celelalte probleme [17]. Echipa sa lucrează în multe alte domenii – de la schimbările climatice la jocuri precum șahul sau Go [18], mesajul important și subliniat totodată de prestigioasa revistă *Nature*, fiind că noile abordări computaționale sunt capabile de a rezolva probleme la care oamenii de știință au încercat să găsească soluții de mai bine de jumătate de secol [19].

References

- [1] C. Vrabie, „Inteligența Artificială și Premiul Nobel în Fizică (2024),” *All in on Tech (AIoT)*, vol. 1, nr. 1, 2025.
- [2] The Nobel Prize, „The Nobel Prize in Chemistry 2024,” 9 10 2024. [Interactiv]. Available: <https://www.nobelprize.org/prizes/chemistry/2024/popular-information/>. [Accesat 08 03 2025].
- [3] D. Searls, „Grand challenges in computational biology,” în *New Comprehensive Biochemistry*, Elsevier, 1998, pp. 3-10.
- [4] C. Giulivi, „Grand challenges in cellular biochemistry: the “next-gen” biochemistry,” *Frontiers in Chemistry*, vol. 2, 2014.
- [5] The Nobel Prize, „2024 Nobel Prize award ceremony,” YouTube, Stockholm, 2024.
- [6] National Library of Medicine, „The Shape and Structure of Proteins,” National Center for Biotechnology Information, 2002. [Interactiv]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK26830/>. [Accesat 08 03 2025].
- [7] The Nobel Prize, „Nobel Prize in Chemistry 1972, Award ceremony speech,” 1972. [Interactiv]. Available: <https://www.nobelprize.org/prizes/chemistry/1972/ceremony-speech/>. [Accesat 09 03 2025].
- [8] C. Levinthal, „How to Fold Graciously,” în *ossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.*, 1969.
- [9] CASP, „Protein Structure Prediction Center,” 2025. [Interactiv]. Available: <https://predictioncenter.org/>. [Accesat 08 03 2025].
- [10] Protein Data Bank, „A Living Digital Data Resource That Enables Scientific Breakthroughs Across The Biological Sciences,” 2025. [Interactiv]. Available: <https://www.rcsb.org/pages/about-us/index>. [Accesat 09 03 2025].

¹ Nu este deloc o exagerare să spunem că *AlphaFold* a provocat o revoluție în biochimia structurală.

² Una dintre primele proteine caracterizate.

³ Pentru exemplificare menționez faptul că pentru înțelegerea unei singure structuri proteice era înainte nevoie de studii doctorale (care, dacă e să iau în considerare exemplul propriu, durează cinci ani). Înmulțind cinci ani cu cele două sute de milioane de structuri proteice existente, ajungem la uriașa valoare de un miliard de ani de cercetări la nivel doctoral, timp economisit grație algoritmilor *AlphaFold*.

- [11] Demis Hassabis, „Accelerating scientific discovery with AI - Nobel lecture with the Nobel Laureate in Chemistry 2024 Demis Hassabis,” The Royal Swedish Academy of Sciences YouTube Channel, Stockholm, 2024.
- [12] K. Bren, R. Eisenberg și H. Gray, „Discovery of the magnetic behavior of hemoglobin: A beginning of bioinorganic chemistry,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 43, 2015.
- [13] The Nobel Prize, „Nobel Prize in Chemistry 1962,” 1963. [Interactiv]. Available: <https://www.nobelprize.org/prizes/chemistry/1962/summary/>. [Accesat 08 03 2025].
- [14] Google DeepMind, „AlphaFold Protein Structure Database,” 2025. [Interactiv]. Available: <https://alphafold.ebi.ac.uk/>. [Accesat 08 03 2025].
- [15] The Nobel Committee for Chemistry, „Computational Protein Design and Protein Structure Prediction,” 09 10 2024. [Interactiv]. Available: <https://www.nobelprize.org/uploads/2024/10/advanced-chemistryprize2024.pdf>. [Accesat 08 03 2025].
- [16] C. Vrabie, „Superintelența—una dintre cele mai mari provocări tehnice ale momentului,” *All in on Tech (AIoT)*, vol. 1, 2025.
- [17] The Nobel Prize, „Nobel Minds 2024,” YouTube, Stockholm, 2024.
- [18] C. Vrabie, AI de la idee la implementare. Traseul sinuos al Inteligenței Artificiale către maturitate. [AI from idea to implementation. The winding path of Artificial Intelligence to maturity], Bucharest: Pro Universitaria, 2024.
- [19] D. Castelvechi, E. Callaway și D. Kwon, „AI comes to the Nobels: double win sparks debate about scientific fields,” *Nature*, 10 10 2024. [Interactiv]. Available: <https://www.nature.com/articles/d41586-024-03310-8>. [Accesat 08 03 2025].