

## PROGETTARE LE PROTEINE: LA RIVOLUZIONE COMPUTAZIONALE

*Il premio Nobel per la Chimica del 2024 è stato assegnato a David Baker, Demis Hassabis e John Jumper per il loro lavoro trasformativo nella previsione e nella progettazione della struttura proteica. I software Rosetta di Baker e AlphaFold2 di DeepMind hanno rivoluzionato il campo, consentendo previsioni strutturali accurate da sequenze di amminoacidi. Questa sinergia tra intelligenza artificiale e biologia computazionale apre nuove strade per la progettazione di nuove proteine, con implicazioni significative in medicina e biotecnologia.*

Il Premio Nobel per la Chimica 2024 celebra le proteine, molecole fondamentali della vita. Vincitore è David Baker dell'Università di Washington (Seattle, USA), per l'uso del computer nella progettazione di nuove proteine, unitamente a Demis Hassabis e John M. Jumper di Google DeepMind, per l'uso dell'intelligenza artificiale nella modellazione delle loro strutture (Fig. 1).

### Le proteine: molecole essenziali per la vita

Le proteine sono molecole indispensabili per una vasta gamma di processi biologici essenziali del nostro corpo, dal supporto strutturale delle cellule ad un esteso repertorio di vitali reazioni biochimiche. Queste attività fondamentali sono intrinsecamente connesse alla forma tridimensionale specifica di ogni proteina che, a sua volta, è determinata dalla sua esatta composizione, ovvero da numero, tipo e sequenza di amminoacidi che la costituiscono [1, 2]. È quindi la specifica sequenza degli amminoacidi a determinare la struttura tridimensionale (3D) della proteina nello spazio e quest'ultima, a sua volta, a conferirle la sua funzione specifica [3]. Detta struttura 3D si organizza in elementi distintivi chiamati alfa eliche, foglietti beta e loops. Il ripiegamento 3D finale è, dunque, il risultato dell'interazione tra tutti gli amminoacidi della proteina, i quali si dispongono nello spazio in una configurazione ottimale ed unica. Se la forma viene alterata, per esempio a causa di mutazioni genetiche

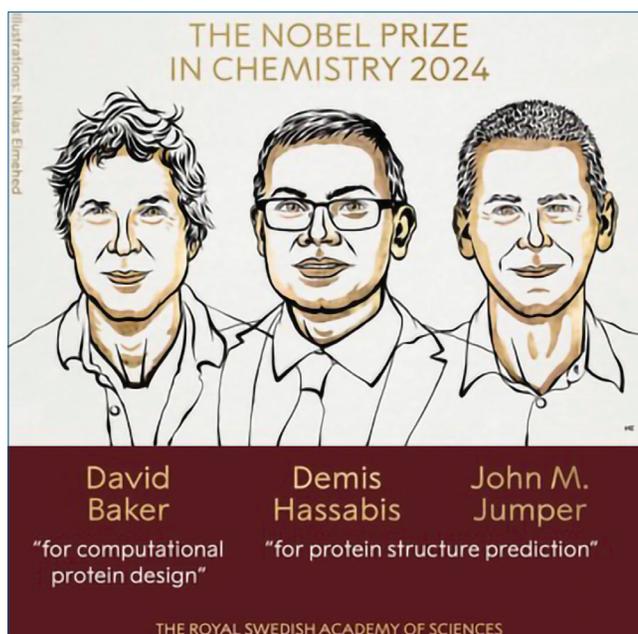


Fig. 1 - I vincitori del premio Nobel per la Chimica 2024: David Baker per la progettazione computazionale della struttura tridimensionale delle proteine e Demis Hassabis e John M. Jumper per la previsione della struttura tridimensionale delle proteine

di amminoacidi di una proteina, anche la sua funzione può risultare compromessa [4, 5]. Questa è, infatti, la causa di molte malattie in cui alterazioni strutturali delle proteine, provocate spesso da mutazioni genetiche della loro composizione amminoacidica, portano ad un malfunzionamento cellulare [3].

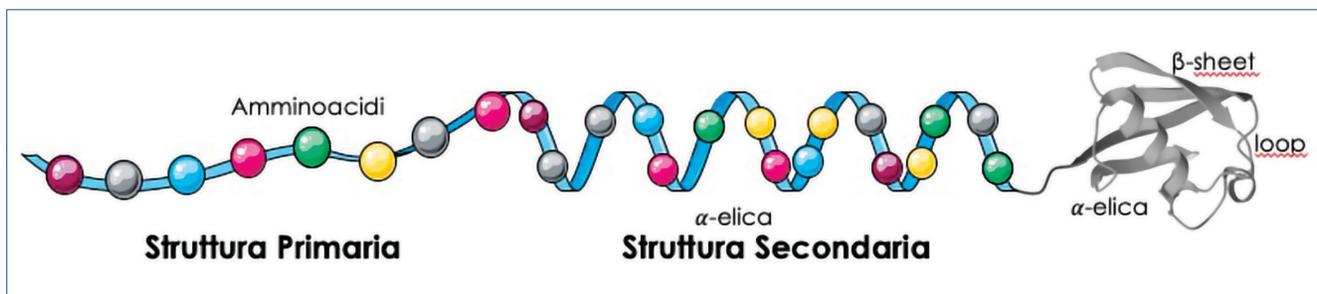


Fig. 2 - Dalla sequenza di amminoacidi primaria fino ad arrivare alla struttura tridimensionale di una proteina composta da diverse strutture secondarie

Quindi la forma 3D di ogni proteina è la chiave per la sua funzione: ogni proteina è un “oggetto” unico, che ha una sua specifica struttura 3D ed una sua dinamica nello spazio che le permette di eseguire precisi ed essenziali funzioni per la vita cellulare. La forma è, dunque, sostanza (Fig. 2).

### La sfida: comprendere per progettare

Decifrare e modellare la struttura 3D delle proteine ha immense implicazioni in scienza, medicina e tecnologia. Questo spiega la lunga corsa scientifica degli ultimi cinquant'anni che, nel tentativo di progettare e regolarne la funzione, ha aperto la strada ad innovazioni rivoluzionarie in numerosi campi.

Per risalire alla struttura di una proteina, i ricercatori spesso utilizzano diverse tecniche sperimentali avanzate che permettono di “risolvere” (*i.e.*, vedere) questa forma con grande precisione, o “risoluzione”, nel senso di osservare finemente la disposizione dei singoli atomi di ogni amminoacido. Sviluppare tecniche sempre più avanzate è fondamentale quindi per decifrare come la struttura di una proteina contribuisca alla sua funzione. Sebbene l'importanza delle proteine in biologia sia nota fin dal diciannovesimo secolo, è stato solo dalla metà del Novecento che i progressi della biochimica e della biologia molecolare hanno consentito di analizzarne la struttura in modo dettagliato, aprendo la strada alle moderne applicazioni nel campo del design proteico.

La prima tecnologia per scoprire la struttura tridimensionale delle proteine è conosciuta come cristallografia a raggi X [1, 2], fruttata il premio Nobel per la Chimica del 1962 a John Kendrew e Max Perutz. Ed è proprio grazie a questa tecnica che il premio Nobel per la Chimica del 1972 Christian Anfinsen ha fatto la sua grande scoperta: se si osserva dispiegare una proteina dal suo stato 3D riportandola ad una sem-

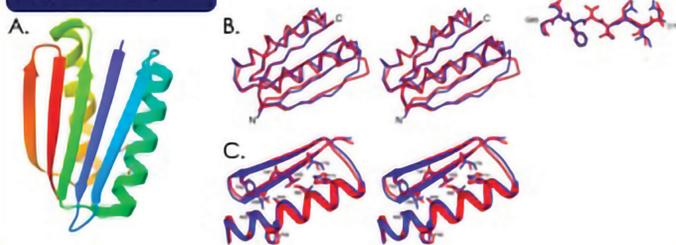
plice catena lineare di amminoacidi e la si lascia poi libera di adottare una forma qualsiasi, questa catena lineare di amminoacidi torna a ripiegarsi esattamente nella stessa struttura 3D che aveva in origine. Un po' come sciogliere un gomitolo di lana, rendendolo un filo lineare e vedere che, al rilascio, esso torna spontaneamente a formare l'esatto gomitolo di partenza. Un fenomeno affascinante che dimostrò, quindi, che la struttura 3D di ogni proteina dipende unicamente dalla sequenza di amminoacidi che la compone [3]. Si aprì così quella che un altro premio Nobel per la Chimica (tanta densità di Nobel riflette l'importanza di questo campo di studi) - l'esperto di ribosomi Venkatraman Ramakrishnan - ha definito “la grande sfida cinquantennale della biologia”: scoprire come determinare la struttura che assumerà una proteina, a partire dalla sequenza di amminoacidi da cui è composta.

La cristallografia a raggi X per determinare la struttura 3D delle proteine è una tecnica di analisi molto precisa e tuttora utilissima, ma presenta alcuni svantaggi: è lenta, costosa e non applicabile a tutte le proteine. Basti pensare che, delle oltre 200 milioni di proteine con sequenze di amminoacidi note, è stata determinata la struttura 3D attraverso la cristallografia di solo circa 200 mila (0,1%). Ne deriva quindi che la capacità di prevedere la struttura di una proteina a partire semplicemente dalla sua composizione di amminoacidi, ovvero senza l'uso della cristallografia a raggi X ma attraverso l'uso della modellistica computazionale, comporterebbe quindi un enorme vantaggio per la ricerca scientifica. Proprio per questo, nel 1994 è stato avviato il progetto “Critical Assessment of Protein Structure Prediction” (CASP) (<https://prediction-center.org/>) che si è trasformato in una competizione scientifica biennale. Da allora, gruppi di ricerca da tutto il mondo si sono affrontati nel tentativo di prevedere-

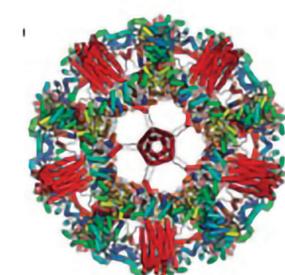
## David Baker e il software ROSETTA

Il mondo della **progettazione** computazionale di nuove proteine

### Top7: la prima



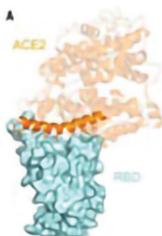
### Negli anni



**2016:** nuovi nanomateriali (Nature **2016**)



**2017:** sensori ambientali per il potente oppioide fentanil. (eLife **2017**)



**2020:** miniproteine inibitrici di SARS-CoV-2. (Science **2020**)



**2022:** macchine molecolari. (Science **2022**)

Fig. 3 -A) Struttura cristallina di Top7; B, C) confronto tra il modello computazionale progettato (blu) con il software Rosetta e la struttura a raggi X risolta (rosso) di Top7. Proteine progettate da Rosetta nel corso degli anni, tra cui la creazione di nuovi materiali, sensori, vaccini e macchine molecolari

re con l'uso di tecniche computazionali la struttura 3D di proteine senza conoscerne la vera struttura sperimentale, già risolta ma tenuta segreta solo per poi verificare l'accuratezza del modello proposto.

David Baker ha partecipato a questa competizione sin dalla sua prima edizione. Grazie all'approccio innovativo del suo team, composto da esperti di biochimica e biologia computazionale, Baker ha apportato un contributo decisivo alla modellazione della struttura della proteina attraverso l'ideazione, alla fine degli anni Novanta, del software *Rosetta*, auspicandone evidentemente la stessa abilità di decodifica consentita dalla celebre Stele [6] (Fig. 3). Questo approccio ha permesso di ottenere previsioni sempre più accurate negli anni, aprendo nuove strade nella comprensione delle strutture proteiche. I risultati ottenuti hanno anche messo in luce l'importanza della bioinformatica nel campo della biologia strutturale, dimostrando come la combinazione di analisi computazionali e dati sperimentali possa portare a scoperte significative nella scienza delle proteine. Il computer iniziò quindi ad essere usato come un acceleratore per lo studio della struttura di proteine.

### David Baker: il Nobel e il futuro delle proteine

Il debutto di Baker al CASP, nel 1998, fu promettente e segnò l'inizio di un lungo percorso di ricerca scientifica. Baker e il suo gruppo ebbero infatti un'intuizione brillante: perché non invertire il processo di analisi? Invece che partire dalla sequenza di amminoacidi per prevedere la struttura 3D delle proteine, perché non utilizzare Rosetta per determinare la sequenza di amminoacidi necessaria ad ottenere una specifica struttura 3D? Questa idea aprì nuove prospettive, non solo per una comprensione più profonda della biologia delle proteine, ma anche per la progettazione di strutture 3D di proteine fatte su misura. Una sorta di alta sartoria per proteine, ottenute disegnando nuove strutture 3D e nuove funzioni per applicazioni in ambito medico e biotecnologico. Fu un passo fondamentale nel campo del *protein design*, ovvero la progettazione di proteine artificiali (*i.e.*, non presenti in natura) con funzioni inedite. Baker, con una metafora, spiegò l'importanza del suo approccio: "Se vuoi costruire un aeroplano, non inizi modificando un uccello, piuttosto cerchi di capire i principi dell'aerodinamica, e con questi costruisci una macchina volante". Ed



è esattamente ciò che Baker e il suo team dell'Università di Washington hanno fatto in questi decenni. Un esempio pionieristico di questo approccio è stato la progettazione della proteina *Top7*, ideata per assumere una struttura 3D completamente originale. Rosetta fu in grado di prevedere quale sequenza di amminoacidi sarebbe stata necessaria per realizzarla e, per farlo, analizzò un database contenente tutte le strutture proteiche conosciute fino a quel momento, identificando frammenti strutturali simili alla struttura 3D desiderata. Utilizzando le regole del ripiegamento delle proteine, Rosetta propose la sequenza di amminoacidi produttiva della struttura 3D desiderata. A questo punto a Baker non restava che realizzare in laboratorio la sequenza di amminoacidi, lasciare si ripiegasse nella propria struttura 3D e risolverne la struttura a raggi X per constatare che la proteina *Top7* presentava esattamente la struttura 3D desiderata e progettata al computer. Era il 2003, e *Top7* divenne così la prima proteina artificiale progettata al computer, con una struttura 3D mai vista in natura [7]. Era nato un nuovo approccio per fare *protein design* basato sulla computazione. Da allora, Baker ha creato molte altre proteine artificiali, tra cui miniproteine in grado di neutralizzare virus come il SARS-CoV-2 responsabile del COVID-19 [8], capaci di colpire le cellule cancerose, ed anche velocizzare (*i.e.*, catalizzare) complesse reazioni (bio)chimiche o assemblarsi in nuovi nanomateriali con proprietà inedite (Fig. 3). Utilizzando la progettazione computazionale di proteine, è oggi possibile creare molecole proteiche altamente specifiche, riducendo i costi e migliorando la stabilità e l'efficacia del processo di sviluppo di nuove tecnologie. Oggi, David Baker è un professore in diverse discipline scientifiche, ha pubblicato oltre 640 articoli di ricerca e più di 100 brevetti. È inoltre il fondatore della Institute for Protein Design (IPD), una struttura di ricerca presso l'Università di Washington (<https://www.ipd.uw.edu/>). L'IPD si concentra sulla progettazione di proteine innovative attraverso approcci computazionali avanzati, con l'obiettivo di sviluppare applicazioni in diversi settori, dalla medicina alla biotecnologia. Infine, è anche co-fondatore di 21 aziende biotecnologiche, tra cui Prospect Genomics, acquisita da una sussidiaria di Eli Lilly nel 2001, e Icosavax, che è stata acquisita da AstraZeneca nel 2023. Queste iniziative dimostrano il suo impegno a tradurre le scoperte scientifiche in applicazioni pratiche, contribuendo all'innovazione nel campo delle biotecnologie ([\[ipd.uw.edu/baker-technology-transfer-roles/\]\(https://www.ipd.uw.edu/baker-technology-transfer-roles/\)\). Una scienza, quella originata da Baker, che ha dimostrato il suo grande valore.](https://www.</a></p></div><div data-bbox=)

### L'impatto dell'intelligenza artificiale in biologia strutturale

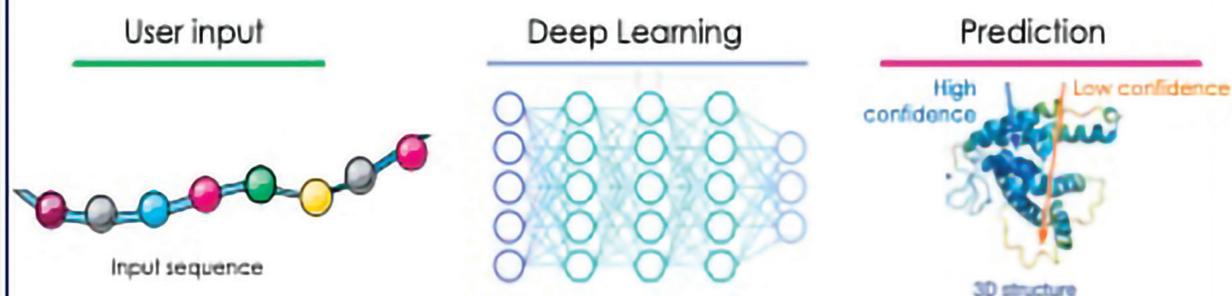
Per quasi vent'anni, la competizione CASP ha visto scarsi progressi, subendo una battuta d'arresto nel tentativo di prevedere le strutture proteiche. Questo quantomeno fino all'arrivo di AlphaFold, un modello di intelligenza artificiale (IA) sviluppato da Demis Hassabis, creatore di DeepMind, una società specializzata nello sviluppo di IA per videogiochi, acquistata da Google nel 2014 (<https://deepmind.google/technologies/alphafold/>). Hassabis è una figura eclettica, con un passato da prodigioso scacchista ed una vasta esperienza nello sviluppo di IA. Appassionato di sfide, scelse la previsione computazionale della struttura 3D delle proteine come il suo nuovo campo di prova. Così, arrivò il primo modello del software AlphaFold, che partecipò al CASP nel 2018, ottenendo un'accuratezza predittiva del 60%, molto superiore alla media della competizione che si attestava intorno al 40%. Questo risultato era, tuttavia, ancora inferiore all'obiettivo del 90% fissato dagli organizzatori. La storia si sarebbe potuta fermare lì, con un parziale successo di AlphaFold. Non fu così. Entrò in scena John Jumper, un giovane fisico appena dottorato in chimica con esperienza nel campo computazionale delle simulazioni di proteine. Jumper si unì a DeepMind nel 2017 e collaborò con Hassabis per sviluppare AlphaFold2, una seconda versione del software iniziale che fu migliorata grazie ad un utilizzo di IA che si basa sul *deep learning*, noto come "trasformatore", addestrato su un vasto dataset contenente tutte le strutture proteiche e le sequenze di amminoacidi conosciute. Quando AlphaFold2 partecipò al CASP nel 2020, segnò una svolta rivoluzionaria: era capace di prevedere la struttura delle proteine con una precisione paragonabile a quella della cristallografia a raggi X [9] (Fig. 4). Il risultato fu che, dopo oltre cinquant'anni, la sfida della previsione delle strutture proteiche era stata finalmente vinta grazie all'utilizzo di IA, accendendo di entusiasmo il mondo scientifico.

La storia di AlphaFold è, dunque, straordinaria ed è stata seguita nel 2021 da un'importante collaborazione tra DeepMind e l'European Molecular Biology Laboratory (EMBL) allo scopo di creare una banca dati pubblica contenente le strutture 3D - predette

## Demis Hassabis e John M. Jumper e il software AlphaFold

IA per la predizione di strutture 3D

### AlphaFold



### Risultati

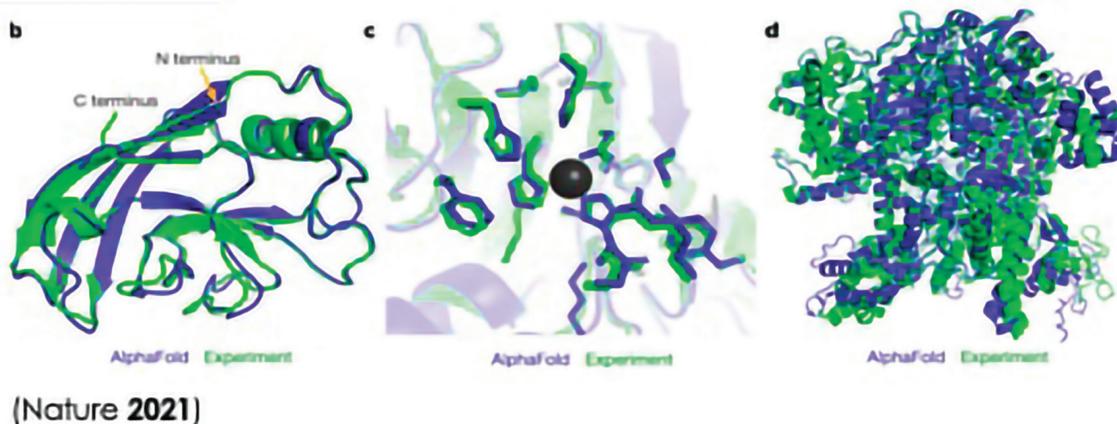


Fig. 4 - AlphaFold: dall'input dell'utente (la sequenza di aminoacidi), al metodo di apprendimento basato sul *deep learning* fino alla previsione della struttura 3D. Risultato: confronti tra modelli AlphaFold (blu) e strutture sperimentali risolte ai raggi X (verde)

da AlphaFold - per quasi tutte le proteine umane e per molte altre proteine di interesse biologico [10]. Questo database è diventato una risorsa preziosissima per la comunità scientifica globale. Inoltre, recentemente è stato riportato Alphafold3, una nuova versione che promette di prevedere la struttura non solo delle proteine ma anche di strutture 3D degli acidi nucleici, RNA e DNA [11], in tal modo elevando il livello di complessità della sfida vista la flessibilità tipica degli acidi nucleici. Prevedere le strutture 3D di RNA e DNA sarà un'ulteriore conferma delle potenzialità del computer e dell'IA in campo scientifico.

### Computer, IA e dati sperimentali per un unico obiettivo

Sicuramente i vincitori del premio Nobel del 2024 sono accumulati dalla volontà di comprendere come le proteine formino la loro struttura 3D e come questa regoli la loro funzione. Tuttavia, ad unirli è anche l'impiego di computer, algoritmi, fisica e chimica applicati alla biologia, un approccio multidisciplinare che ha portato a risultati straordinari, evidenziando l'importanza di tecnologie come l'High Performance Computing (HPC) e l'intelligenza artificiale (IA) nel campo della biologia strutturale. L'uso di strumenti



avanzati ha reso possibile l'analisi di enormi quantità di dati: nel 2022 Alphafold2 ha modellato le strutture tridimensionali di oltre 200 milioni di proteine e, dalla sua pubblicazione nel 2021, è stato utilizzato da oltre due milioni di ricercatori di 190 Paesi nel mondo. Ha anche generato una nuova sfida in questo campo: infatti, nel 2022, il software Meta AI, rivale di AlphaFold, ha predetto la struttura di 600 milioni di proteine, spingendo il campo verso un rapido progresso attraverso l'utilizzo di strumenti computazionali che hanno non solo rivoluzionato la previsione delle strutture 3D di proteine, ma anche permesso di integrare dati sperimentali con modelli computazionali, rendendo il processo più efficiente.

Tuttavia, il successo così rapido di AlphaFold2, culminato nel Nobel dopo soli tre anni dalla sua pubblicazione nel 2021, è un caso quasi unico nella storia di questo riconoscimento, che spesso viene concesso solo dopo decenni di verifica dell'impatto scientifico della scoperta. Sarà interessante, dunque, negli anni futuri, verificare appieno l'impatto di questa tecnologia sulla ricerca scientifica e tecnologica. Un contributo, quello di Alphafold, riconosciuto rapidamente anche forse grazie all'entusiasmo diffuso per tutto ciò che riguarda attualmente la rivoluzione della IA ed all'attenzione al suo potenziale, che ha dimostrato di saper dare risposte rapide e spesso corrette, anche in campo scientifico. Tutto questo, però, con il limite intrinseco della IA, ovvero quello di permetterci una comprensione solo ridotta dell'elaborazione dati che conduce a tali rapide soluzioni. La ricerca scientifica, quindi, potrà sempre trovare stimolo da questo limite della IA per continuare a studiare i principi primi e fondamentali che guidano i complessi fenomeni fisici e chimici che spesso si trovano in natura e tecnologia. Ci sarà dunque modo di insegnare un po' di scienza ai **nostri computer e algoritmi IA**, mentre impareremo da loro nuove possibili soluzioni generate da complesse analisi di grandi quantità di dati. Una sinergia, questa, oltremodo stimolante.

### Nuove proteine nel nostro futuro

È a questo punto importante sottolineare il potenziale trasformativo di questi studi scientifici. Le applicazioni vanno dalla medicina personalizzata, dove le nuove proteine possono essere progettate per trattamenti specifici, fino a tecnologie emergenti che potrebbero migliorare sensibilmente il nostro approccio alla salute ed alla sostenibilità, spaziando dalla medicina alle tecnologie green per l'inquinamento, fino alla creazio-

ne di nuovi materiali. Le aziende co-fondate da David Baker, ad esempio, hanno attirato finanziamenti significativi, con investimenti che superano il miliardo di dollari, confermando la fiducia del settore finanziario nella ricerca applicata alla biologia delle proteine. Questi successi non solo pongono le basi per ulteriori scoperte, ma ispirano una nuova generazione di scienziati ad esplorare le infinite possibilità offerte dall'integrazione di biologia e tecnologia. Il Nobel per la Chimica di quest'anno premia, dunque, non soltanto gli straordinari sforzi del passato, ma incoraggia con fiducia un viaggio esplorativo che è solo agli inizi. Il futuro è dunque luminoso, ed il potenziale innovativo è praticamente illimitato.

*Ad maiora!*

### Bibliografia

- [1] J.C. Kendrew, R.E. Dickerson *et al.*, *Nature*, 1958, **181**(4610), 662.
- [2] M.F. Perutz, *Nature*, 1956, **177**, 307.
- [3] C.B. Anfinsen, *Science*, 1973, **181**, 223.
- [4] L. Pauling, H.A. Itano, *Science*, 1949, **110**(2865), 543.
- [5] D. Baker, A. Sali, *Science*, 2001, **294**(5540), 93.
- [6] K.T. Simons C. Kooperberg *et al.*, *Mol. Biol.*, 1997, **268**(1), 209.
- [7] D. Baker, *Science*, 2003, **302**, 1364.
- [8] D. Baker, M.S. Diamond, *Cell Host & Microbe*, 2021, **29**(7), 1151.
- [9] A.W. Senior, R. Evans *et al.*, *Nature*, 2020, **577**, 706.
- [10] J. Jumper, R. Evans *et al.*, *Nature*, 2021, **596**, 583.
- [11] J. Abramson, J. Adler *et al.*, *Nature*, 2024, **630**, 493.

### Protein Design: The Computational Breakthrough

The 2024 Nobel Prize in Chemistry was awarded to David Baker, Demis Hassabis, and John Jumper for their transformative work in protein structure prediction and design. Baker's Rosetta software and DeepMind's AlphaFold2 have revolutionized the field, allowing for accurate structural predictions from amino acid sequences. This synergy between AI and computational biology opens new avenues for designing novel proteins with significant implications in medicine and biotechnology.