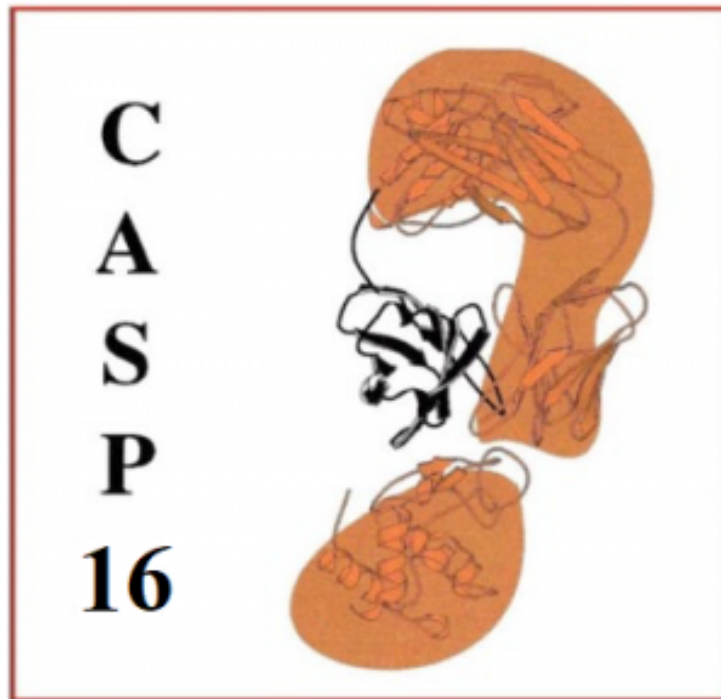


Le “Olimpiadi Proteiche” e l’oro di AlphaFold

Dalla nascita del CASP alla rivoluzione dell'intelligenza artificiale: come abbiamo risolto l'enigma del ripiegamento delle proteine

di **ALESSIA TERZANO, ROBERTA SERRA**

CASP: DOVE LA SCIENZA SI METTE ALLA PROVA



Logo dell'ultima edizione del CASP tenuta da maggio ad agosto 2024.

Le proteine sono alla base della vita, dalle difese immunitarie alla catalisi di reazioni metaboliche, sono le responsabili della maggior parte delle attività cellulari. Comprendere la loro struttura è fondamentale per capire a fondo le loro funzioni e creare nuove terapie. Determinare la struttura proteica per via sperimentale, però, non è solo un lavoro lunghissimo, ma anche molto costoso e complesso. Considerando la quantità enorme di proteine esistenti, abbiamo bisogno di un supporto che ci aiuti ad accelerare questa ricerca. Infatti, se oggi siamo a conoscenza di miliardi di sequenze di proteine (la struttura primaria), conosciamo il modo in cui si ripiegano (la struttura tridimensionale) solo di un numero limitato di queste. Ed è qui che entrano in gioco gli approcci computazionali. Questi sono in grado di colmare il "gap" andando a prevedere la struttura delle proteine sulla base della loro sequenza amminoacidica. Se all'inizio questi modelli risultavano poco affidabili, negli ultimi anni, grazie all'integrazione del deep learning, hanno raggiunto un livello di accuratezza straordinario, in grado di competere a tutti gli effetti con gli approcci sperimentali.

I modelli di previsione computazionale della struttura delle proteine vengono valutati dal CASP (Critical Assessment of Protein Structure Prediction), che potremmo chiamare le "Olimpiadi Proteiche": esperimento comunitario fondato nel 1994 che riunisce, con cadenza biennale, scienziate e scienziati esperti di struttura e ripiegamento proteico.

Il CASP si propone, quindi, come un modo per mettere alla prova scienziat* e i loro modelli computazionali per capire a che punto siamo arrivati, dove stiamo andando e dove potremmo arrivare. Ogni due anni, i partecipanti sono chiamati a creare i modelli strutturali di un set di molecole e complessi di macromolecole (proteine, RNA, ligandi). Il tutto si svolge come un vero 'test al buio': le strutture sono già state determinate sperimentalmente in laboratorio, ma non sono ancora state rese pubbliche. In questo modo è impossibile 'barare' e la competizione valuta la reale capacità predittiva degli algoritmi.

Nelle recenti edizioni abbiamo assistito a un enorme aumento dell'accuratezza: prima nel 2020 (CASP14), in cui i modelli ottenuti erano molto simili a quelli sperimentali; poi nel 2022 (CASP15), con un grande incremento nella previsione dei complessi proteici. Il motore di questa rivoluzione è stato l'ingresso del Deep Learning, che ha trovato la sua massima espressione in AlphaFold.

Ma come è riuscita l'intelligenza artificiale a risolvere una sfida che tormentava gli scienziati da oltre mezzo secolo?

LA SVOLTA DI ALPHAFOLD: QUANDO L'IA IMPARA DALLA NATURA



Visualizzazione tridimensionale della conformazione spaziale di una proteina

Per oltre cinquant'anni il problema del ripiegamento delle proteine era affrontato soprattutto dal punto di vista fisico-chimico: si cercava di simulare tutte le forze che agiscono sugli amminoacidi (legami, interazioni elettrostatiche, idrofobicità) e calcolare quale configurazione fosse la più stabile. Il problema è che le possibili conformazioni di una proteina sono tante: esplorarle tutte è praticamente impossibile.

AlphaFold, invece di simulare ogni possibile movimento, ha imparato dai dati. Ha analizzato milioni di sequenze proteiche e le strutture già note, individuando schemi ricorrenti tra sequenza e organizzazione tridimensionale.

La prima versione di AlphaFold, presentata nel 2018, utilizzava reti neurali profonde per analizzare grandi quantità di dati evolutivi e individuare correlazioni tra amminoacidi anche molto distanti nella sequenza. Queste correlazioni fornivano indizi su quali parti della proteina potessero trovarsi vicine nello spazio, permettendo di ricostruire una struttura tridimensionale plausibile. Si trattava di un grande passo avanti rispetto ai metodi precedenti, anche se richiedeva ancora dei miglioramenti.

La vera svolta arrivò con AlphaFold 2 (2020–2021). A differenza della prima versione, il nuovo modello non si limitava a stimare distanze o contatti tra amminoacidi, ma costruiva direttamente la struttura tridimensionale della proteina all'interno della rete neurale. Inoltre, la struttura veniva progressivamente migliorata attraverso un processo chiamato "recycling", in cui il modello rielaborava più volte la propria previsione per renderla sempre più coerente e precisa.

Grazie a questo nuovo approccio, proprio al CASP 14, AlphaFold 2 ha raggiunto un livello di accuratezza senza precedenti, in molti casi paragonabile a quello ottenuto con metodi sperimentali. In questo modo aprì la strada alla creazione dell'AlphaFold Protein Structure Database, una risorsa ad accesso libero che oggi contiene milioni di modelli strutturali. Questo database ha permesso di conoscere molte più strutture tridimensionali rispetto a prima, accelerando la ricerca in ambiti che vanno dalla biologia di base allo sviluppo di nuovi farmaci.

Più recentemente, AlphaFold 3 ha esteso le capacità del sistema alla modellazione di complessi biomolecolari, includendo ligandi, ioni e acidi nucleici. Questo passaggio è importante perché sposta l'attenzione dalla singola proteina isolata alle interazioni molecolari che ne determinano la funzione biologica. Restano però alcune sfide aperte: le proteine non sono strutture statiche ma dinamiche: possono assumere conformazioni diverse, ossia piegarsi in modi diversi a seconda delle condizioni o delle interazioni con altre molecole, ma il modello tende a predirne una sola; inoltre, la previsione dell'RNA resta ancora meno accurata rispetto ai casi proteici più semplici. Anche l'impatto delle mutazioni o delle varianti genetiche non è sempre stimabile con precisione. Nonostante questi limiti, le prestazioni raggiunte mostrano che l'intelligenza artificiale può affrontare con successo uno dei problemi più difficili della biologia moderna, aprendo al tempo stesso nuove domande per la ricerca futura.

ALPHAFOLD E CASP: UNA RELAZIONE CHE HA CAMBIATO IL MODO DI STUDIARE LE PROTEINE

Il rapporto tra CASP e AlphaFold è stato fondamentale e ha influenzato il destino di entrambi. CASP ha fornito la prova concreta che AlphaFold fosse davvero capace di predire con grande accuratezza la struttura delle proteine. Allo stesso tempo, il successo di AlphaFold ha cambiato la natura stessa della competizione: una volta raggiunta un'accuratezza vicina a quella sperimentale per molte proteine singole, la sfida si è spostata verso problemi più complessi, come i complessi multiproteici e le interazioni con altre molecole. Così, mentre CASP ha permesso di dimostrare l'efficacia di AlphaFold, AlphaFold ha ridefinito gli obiettivi di CASP, segnando insieme una nuova fase per la biologia computazionale.

Rubrica a cura di **Generazione Stem**

BIOGRAFIA AUTRICE

Alessia Terzano è una studentessa magistrale in Biotecnologie Avanzate. Ama raccontare la scienza in modo accessibile e creativo anche attraverso la divulgazione online. Collabora con progetti come Generazione STEM per rendere la scienza più accessibile e inclusiva.

Roberta Maria Serra è una laureanda in Scienze Biologiche presso l'Università degli Studi di Palermo. Unisce la passione per la scienza con la scrittura per rendere temi complessi più accessibili. Collabora con diverse realtà come Generazione Stem per avvicinare sempre più persone alla conoscenza scientifica e per contribuire ad abbattere gli stereotipi di genere.

FONTI

Sito ufficiale del CASP <https://predictioncenter.org>

Fatima Ali, N., Khan, S., & Zahid, S. (2025). *A critical address to advancements and challenges in computational strategies for structural prediction of protein in recent past. Computational biology and chemistry*, 117, 108430.

<https://doi.org/10.1016/j.compbiolchem.2025.108430>

<https://alphafold.ebi.ac.uk/>

Skolnick, J., Gao, M., Zhou, H., & Singh, S. (2021). *AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. Journal of chemical information and modeling*, 61(10), 4827–4831. <https://doi.org/10.1021/acs.jcim.1c01114>.

Jennifer Fleming et Al. (2025). *AlphaFold Protein Structure Database and 3D-Beacons: New Data and Capabilities*, Journal of Molecular Biology, Volume 437, Issue 15, 168967, ISSN 0022-2836. <https://doi.org/10.1016/j.jmb.2025.168967>.

Immagine: foto di **National Institute of Allergy and Infectious Diseases** su **Unsplash**

Logo tratto dal sito ufficiale del CASP <https://predictioncenter.org/casp16/index.cgi>