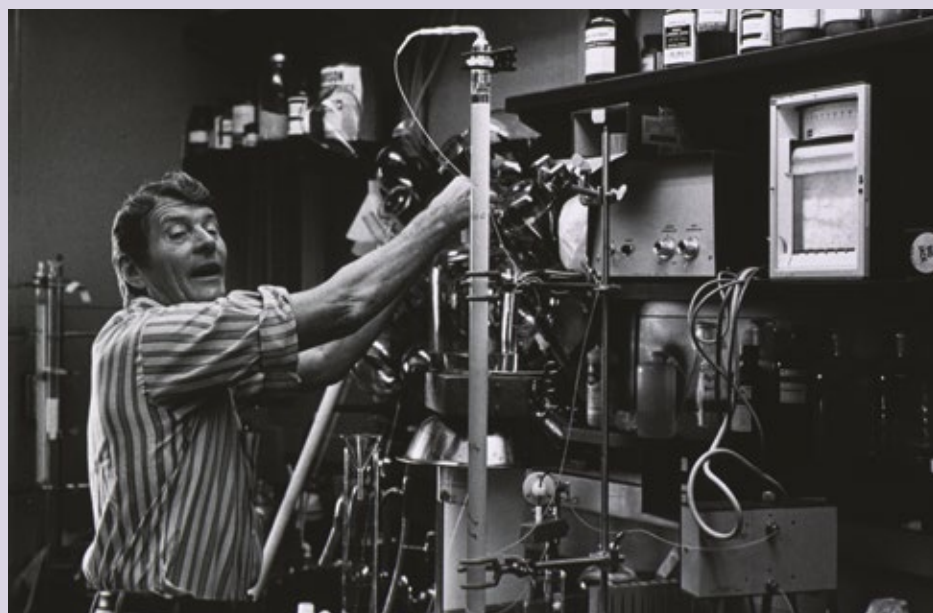


Facile per la natura, difficile per l'uomo

Il problema del *protein folding*

di Piero Fariselli



a.
Christian Anfinsen, premio Nobel per la chimica del 1972. Dimostrò che l'informazione contenuta nella sequenza di amminoacidi è sufficiente a determinare la struttura tridimensionale di una proteina, e quindi la sua funzione biologica.

Se il DNA è la molecola che contiene le istruzioni per la costruzione e il funzionamento degli organismi viventi, sono le proteine a svolgere la maggior parte delle funzioni biologiche. Codificate nei geni, le proteine vengono tradotte come catene di amminoacidi, formate da una combinazione di 20 unità diverse. In ambiente fisiologico, queste catene si ripiegano spontaneamente in una forma tridimensionale specifica (la struttura "terziaria"), essenziale per la loro funzione. Nel 1972, Christian Anfinsen vinse il premio Nobel per la chimica per aver dimostrato che l'informazione contenuta nella sequenza degli amminoacidi è sufficiente a determinare la struttura tridimensionale di una proteina, la quale assume una geometria spaziale adatta a svolgere le sue funzioni, interagendo con le molecole dell'ambiente circostante. Da allora, ricercatori di tutto il mondo hanno cercato di risolvere il cosiddetto *protein folding problem*: predire la struttura tridimensionale di una proteina a partire dalla sua sequenza, facilmente ottenibile, quest'ultima, dal sequenziamento del DNA. I fisici hanno modellato il problema come una ricerca del minimo di energia libera ma, nonostante la teoria sia formalmente descrivibile

mediante la meccanica quantistica, le soluzioni pratiche sono impossibili da calcolare anche solo approssimativamente, dato l'enorme numero di atomi coinvolti.

Questa difficoltà ha portato alla formulazione del famoso "paradosso di Levinthal": se una proteina esplorasse casualmente tutte le possibili conformazioni, impiegherebbe un tempo astronomico per trovare quella corretta. Eppure, in natura, il *folding* avviene in modo rapido ed efficiente (nell'ordine dei micro- o millisecondi). Arthur Lesk ha riassunto perfettamente il dilemma con la frase: "Facile per la natura, difficile per l'uomo".

Per affrontare questo problema, la comunità scientifica ha sviluppato numerosi approcci computazionali e sperimentali. Nel 1994 è nato il CASP (Critical Assessment of protein Structure Prediction), una competizione biennale che valuta i progressi nella predizione delle strutture proteiche. Durante il CASP, vengono fornite le sequenze di proteine la cui struttura tridimensionale è stata determinata sperimentalmente, ma non ancora resa pubblica. I partecipanti devono predirne la struttura; i risultati vengono poi confrontati con le strutture



b.
David Baker, Demis Hassabis e John Jumper
all'Accademia Reale Svedese delle Scienze durante
la conferenza dei premi Nobel 2024.

reali, una volta rese note. Questa iniziativa ha permesso di monitorare nel tempo l'evoluzione delle tecniche di predizione, mostrando come l'integrazione tra dati evolutivi, modelli fisico-chimici e, più recentemente, intelligenza artificiale abbia progressivamente migliorato le prestazioni.

Il punto di svolta è arrivato nel 2020, quando l'azienda di intelligenza artificiale DeepMind ha presentato AlphaFold2, un sistema basato sul *deep learning*, che ha superato ogni aspettativa. I risultati presentati al CASP14 hanno mostrato una precisione paragonabile, e in molti casi equivalente, a quella di tecniche sperimentali (come la cristallografia a raggi X e la risonanza magnetica nucleare).

Il successo si basa su tre elementi fondamentali: l'utilizzo dell'informazione evolutiva, l'"autoconsistenza" geometrica e il *deep learning* come collante.

AlphaFold2 sfrutta enormi database di sequenze proteiche per costruire allineamenti multipli di sequenze (MSA). Le mutazioni naturali che si sono accumulate nel tempo conservano, attraverso la selezione, le strutture tridimensionali funzionali. Analizzando *pattern* di co-variazione tra residui conservati in famiglie evolutivamente correlate, il modello può dedurre quali amminoacidi tendono a essere vicini nello spazio, anche se lontani nella sequenza. Questa informazione è molto più potente di qualsiasi predizione basata solo sulla fisica.

Una delle intuizioni chiave è stata far "chiudere il cerchio" al modello: AlphaFold2 non predice solo distanze tra residui, ma genera una struttura tridimensionale completa, che viene poi reinterpretata internamente per verificare se è coerente con le informazioni iniziali. Questo processo, chiamato "auto-consistenza", consente al modello di correggersi autonomamente: la struttura generata viene reinserita nel

sistema per un nuovo ciclo di raffinamento, migliorando ogni volta la coerenza globale. Non serve alcun confronto con dati sperimentali durante la predizione. Infine, AlphaFold2 sfrutta un'architettura chiamata Evoformer, che consiste in una rete neurale trasformativa progettata per trattare in parallelo l'MSA e le relazioni tra residui. Questo modello è in grado di integrare contesto evolutivo e geometrico, imparando rappresentazioni complesse che legano la sequenza alla struttura.

Con AlphaFold2, DeepMind ha dimostrato che è possibile predire con elevata precisione la struttura tridimensionale di una proteina partendo dalla sola sequenza degli amminoacidi, sfruttando informazioni evolutive e un sofisticato meccanismo di raffinamento iterativo delle predizioni. Questo straordinario risultato non è passato inosservato: nel 2024 il premio Nobel per la chimica è stato assegnato a Demis Hassabis e John Jumper (AlphaFold) e David Baker (RoseTTAFold), proprio per aver rivoluzionato la predizione strutturale delle proteine attraverso l'intelligenza artificiale.

Abbiamo dunque ottenuto una soluzione efficace al problema, ma possiamo affermare di aver compreso realmente come avvenga il *folding*? Il modello predice accuratamente le strutture, eppure i meccanismi da esso appresi rimangono in gran parte oscuri. Il paradosso è che la comprensione scientifica umana sembra aver ceduto il passo alla straordinaria capacità predittiva delle reti neurali. Forse non è la natura a essere intrinsecamente complessa, ma piuttosto lo è il modo in cui la mente umana cerca di interpretarla. Possiamo dunque concludere che il problema del *folding*, così essenziale per la vita, oggi risulti semplice per la natura, calcolabile per l'intelligenza artificiale, ma ancora irrimediabilmente difficile per l'uomo.

Biografia

Piero Fariselli è professore all'Università di Torino e responsabile dell'unità di intelligenza artificiale e biomedicina computazionale presso il Dipartimento di Scienze Mediche. Da oltre trent'anni si occupa dello sviluppo di modelli di *machine learning* per applicazioni biofisiche e biomediche.