

[https://www.huffingtonpost.it/entry/struttura-delle-proteine-da-google-una-rivoluzione-epocale-di-m-brunori-e-s-gianni\\_it\\_5fe620acc5b6acb53457ec4c?utm\\_hp\\_ref=it-homepage](https://www.huffingtonpost.it/entry/struttura-delle-proteine-da-google-una-rivoluzione-epocale-di-m-brunori-e-s-gianni_it_5fe620acc5b6acb53457ec4c?utm_hp_ref=it-homepage)

# Struttura delle proteine, da Google una rivoluzione epocale (di M. Brunori e S. Gianni)

Un risultato straordinario ottenuto grazie all'intelligenza artificiale con l'algoritmo DeepMind's AlphaFold

25/12/2020 19:18

(di Maurizio Brunori, Professore Emerito, Sapienza Università di Roma; e Presidente Emerito della Classe di Scienze FMN dell'Accademia Nazionale dei Lincei; Stefano Gianni, Professore Ordinario di Biologia Molecolare, Dipartimento di Scienze Biochimiche, Sapienza Università di Roma)

Michael Levitt, Professore all'Università di Stanford e premio Nobel per la Chimica, riporta in un lavoro pubblicato su *Current Opinion in Structural Biology* il frammento di una sua conversazione con Francis Crick, scienziato universalmente noto in quanto scopritore con James Watson della struttura a doppia elica del DNA. Nel 1975 nel corso di quella conversazione, Francis Crick dichiarò: "It is very difficult to conceive of a scientific problem that would not be solved in the next twenty years ... except for a model of brain function and protein folding", individuando due enormi problemi scientifici che sarebbero rimasti irrisolti per decine di anni. Cosa è il protein folding e perché è così importante?

L'unità elementare di ogni essere vivente è la cellula, un oggetto minuscolo ma molto complesso, le cui funzioni e attività biochimiche sono dettate dai suoi componenti più piccoli, le molecole. Tra queste, le più affascinanti e versatili sono le proteine che, capaci di svolgere una impressionante molteplicità di ruoli, garantiscono alla cellula la possibilità di sopravvivere egregiamente e di moltiplicarsi. Il sequenziamento del genoma umano ha permesso di identificare circa 25.000 geni che codificano per altrettante proteine costituite da amino acidi uniti da legami chimici forti. La struttura covalente di questa catena (detta polipeptidica) assomiglia a una collana costituita da perline di forma e colore differenti, che rappresentano i venti amino acidi canonici. La sequenza che compone queste grandi molecole (contenenti un numero geneticamente determinato di amino acidi, da circa 50 a più di mille unità) si chiama sequenza amino acidica o struttura primaria della proteina.

In analogia con tutti gli oggetti che popolano la nostra quotidianità, la specificità e l'efficienza di ogni proteina dipendono dalla loro forma o, più propriamente, dalla loro struttura tridimensionale. In base alla struttura primaria, ogni proteina va incontro al processo di folding che conduce a una caratteristica struttura tridimensionale detta 'struttura nativa', una forma peculiare e cruciale per la funzione in condizioni fisiologiche. La FIGURA 1 illustra il modello in scala di alcune proteine rappresentative del panorama complessivo delle circa 25.000 macromolecole proteiche di *Homo sapiens*; la struttura nativa della proteina è definita dalle coordinate delle migliaia di atomi che la costituiscono. La mioglobina di balena, la prima proteina di cui fu determinata da John Kendrew la struttura 3D, è formata da 153 amino acidi e circa 10.000 atomi che nel modello tridimensionale hanno coordinate atomiche precise.

Perché la morfogenesi delle proteine è così complessa e tanto importante per la Biologia e la Medicina?

Il processo attraverso il quale una catena proteica disordinata (detta anche denaturata) raggiunge la forma nativa funzionalmente competente e strutturalmente definita, è una reazione chimica spontanea e molto complessa che comporta la formazione di migliaia di legami chimici deboli. La FIGURA 2 mostra uno schema semplificato di tale transizione, e riassume l'esperimento fondamentale di Christian Anfinsen e coll. pubblicato nel 1961, con il quale si dimostrò che la denaturazione delle proteine è un processo reversibile, esperimento che gli valse il Premio Nobel. Si osservò, infatti, che l'enzima ribonucleasi denaturato chimicamente, se trasferito in soluzione fisiologica tende ad assumere spontaneamente la struttura tridimensionale funzionalmente competente (forma nativa o native form). La teoria emersa da questo esperimento implica che l'informazione necessaria e sufficiente per l'acquisizione della struttura nativa di una proteina è contenuta nella sua stessa sequenza amino acidica; la decifrazione del codice della morfogenesi o folding, dovrebbe permettere di predire la struttura 3D di ogni proteina presente in Natura a partire dalla sola struttura primaria.

In buona sostanza, la decifrazione di questo codice stereochimico sarebbe di fondamentale importanza perché la conoscenza della struttura 3D delle proteine è la chiave per risolvere importanti problemi della Biologia e della Medicina. Inoltre gli enormi progressi della genomica consentono di prevedere che saranno disponibili le sequenze di moltissimi genomi, e di conseguenza si potranno dedurre le sequenze di amino acidi di un enorme numero di proteine. D'altro canto la determinazione della loro struttura 3D con metodi sperimentali sarebbe un'impresa ciclopica, costosissima e a volte impossibile; basterebbe ricordare che per risolvere la struttura dell'emoglobina con la diffrazione di raggi X il grandissimo Max Perutz, considerato il padre della Biologia strutturale, lavorò più di 20 anni. Adesso i tempi sono enormemente abbreviati, ma comunque le proteine devono essere cristallizzate e non sempre è possibile.

Il paradigma di Anfinsen ha stimolato negli ultimi 50 anni la curiosità e l'impegno di innumerevoli scienziati di grandissimo talento, con risultati molto importanti ma non definitivi. Il problema del folding si è dimostrato estremamente complesso, nonostante l'enorme progresso delle tecnologie sperimentali e dei sistemi di calcolo. Nel 1994 John Moult, professore all'Università del Maryland (USA), diede vita con altri biofisici ed esperti computazionali a un programma innovativo definito CASP (Critical Assessment of protein Structural Prediction). Lo scopo del progetto era la valutazione rigorosa, obiettiva e indipendente di nuovi metodi volti a prevedere la struttura 3D delle proteine a partire dalla sequenza. Agli scienziati coinvolti in questo "esperimento" viene assegnato un insieme di sequenze di proteine la cui struttura 3D era stata già risolta con metodi sperimentali ma mantenuta segreta. Il delicato compito di paragonare i dati sperimentali con le previsioni prodotte dai teorici è delegato a valutatori affidabili e competenti; la nostra cara amica Anna Tramontano, professoressa di Biochimica alla Sapienza purtroppo prematuramente scomparsa, ha svolto con entusiasmo e riconosciuta competenza questo importante compito per molte sessioni di CASP.

Il 30 novembre, a conclusione dell'ultimo CASP14, è stato comunicato un risultato straordinario ottenuto grazie a un algoritmo di Google detto DeepMind's AlphaFold. "This is a big deal, (and we could conclude) that in some sense the problem is solved" ha dichiarato John Moult. La qualità della previsione viene valutata quantitativamente dal parametro GDT (Global Distance Test) che esprime la percentuale di amino acidi della proteina in esame che sono a una distanza uguale o inferiore rispetto a un valore di riferimento deciso dal Comitato di CASP. Secondo John Moult un valore di GDT di circa 90 definisce una struttura che si può già considerare uguale alla corrispondente struttura sperimentale di miglior qualità. Il sistema AlphaFold2 ha raggiunto un valore di GDT di circa 90, molto al di sopra di tutti gli altri concorrenti (FIGURA 3). Le previsioni

di struttura 3D così ottenute hanno dimostrato un livello di accuratezza eccellente. Il risultato implica che potremo conoscere la posizione di ogni singolo atomo della proteina con una precisione di circa 1,6 Å, un risultato straordinario.

Venki Ramakrishnan, laureato Nobel e Presidente della Royal Society, ha dichiarato che la performance di AlphaFold rappresenta un enorme passo avanti nella comprensione del problema del folding delle proteine 70 anni dopo il fondamentale esperimento di Anfinsen. Janet Thornton, Direttore emerito del European Bioinformatics Institute del EMBL a Cambridge (UK), che temeva di non arrivare a vedere la soluzione del problema, ora prevede un futuro ricco di risultati per la scienza delle proteine e più in generale per le scienze della vita. Basta considerare che la rivoluzione genomica consentirà di decifrare la sequenza amino acidica di tutte le proteine presenti in Natura, circa 180 milioni; a fronte della attuale disponibilità di circa 170.000 strutture tridimensionali oggi depositate nella Protein Data Bank.

I risultati straordinari ottenuti con AlphaFold e i miglioramenti metodologici che seguiranno nei prossimi mesi dimostrano le enormi potenzialità dell'Intelligenza Artificiale come metodologia di fondamentale impatto nella ricerca scientifica del XXI secolo. Demis Hassabis, cofondatore di DeepMind, ha dichiarato che oltre all'eccezionale potenza dimostrata nel prevedere la struttura 3D delle proteine AlphaFold potrebbe generare importanti contributi nello sviluppo di programmi di ampio respiro e di grande impatto generale.

Ma qual è la novità e la forza dell'algoritmo AlphaFold rispetto alle tecniche computazionali sviluppate in precedenza dalla comunità scientifica? AlphaFold si basa su una complessa metodologia computazionale nota come rete neurale artificiale, che permette l'apprendimento automatico della macchina. In analogia con il funzionamento del cervello, questi modelli computazionali sono composti di unità chiamate "neuroni" organizzati in strati successivi. Ciascun neurone risulta interconnesso a tutti gli altri dello strato successivo tramite connessioni pesate. Il processo di apprendimento delle reti neurali artificiali è rappresentato dall'aggiornamento di una matrice di pesi che può spegnere una percentuale di questi neuroni in base all'esperienza che la rete acquisisce. L'elegante disegno dell'algoritmo AlphaFold, insieme all'apprendimento automatico basato sulle strutture 3D attualmente depositate nella Protein Data Bank, ha permesso ai ricercatori della DeepMind di raggiungere un risultato eccezionale.

E' interessante rileggere la previsione di Francis Crick riportata in apertura, riguardante il funzionamento del cervello e la morfogenesi di una proteina ("...except for a model of brain function and protein folding"); e prendere atto che le tecniche di apprendimento automatico sono state originariamente introdotte per tentare di simulare il funzionamento del nostro cervello. E' particolarmente affascinante osservare come lo studio di quest'ultimo problema abbia contribuito in maniera determinante alla soluzione del protein folding, un mistero di natura biochimico-molecolare che, grazie a DeepMind, è stato chiarito e che potrebbe in futuro produrre ulteriori straordinari risultati di interesse del tutto generale.