



Come l'intelligenza artificiale sta rivoluzionando la ricerca scientifica

**Dario Cardamone
Sara Joubbi**

**Data Science for Health (DASCH) Lab
Fondazione Toscana Life Sciences**



Glossario, parole che ci servono

Modelli di linguaggio: I modelli di linguaggio sono un tipo di programma che aiuta i computer a comprendere e generare il testo scritto in diverse lingue in maniera intelligente, che risulta estremamente naturale.

Proteine: Le proteine sono piccole molecole che svolgono un ruolo enorme nella biologia. Si trovano all'interno di ogni essere vivente e agiscono come componenti strutturali, trasportatori, molecole segnali e molto altro.

Amminoacidi: Gli amminoacidi sono i mattoni fondamentali delle proteine, svolgendo ruoli cruciali nella struttura e nella funzione delle cellule e dei tessuti.



Glossario, parole che ci servono

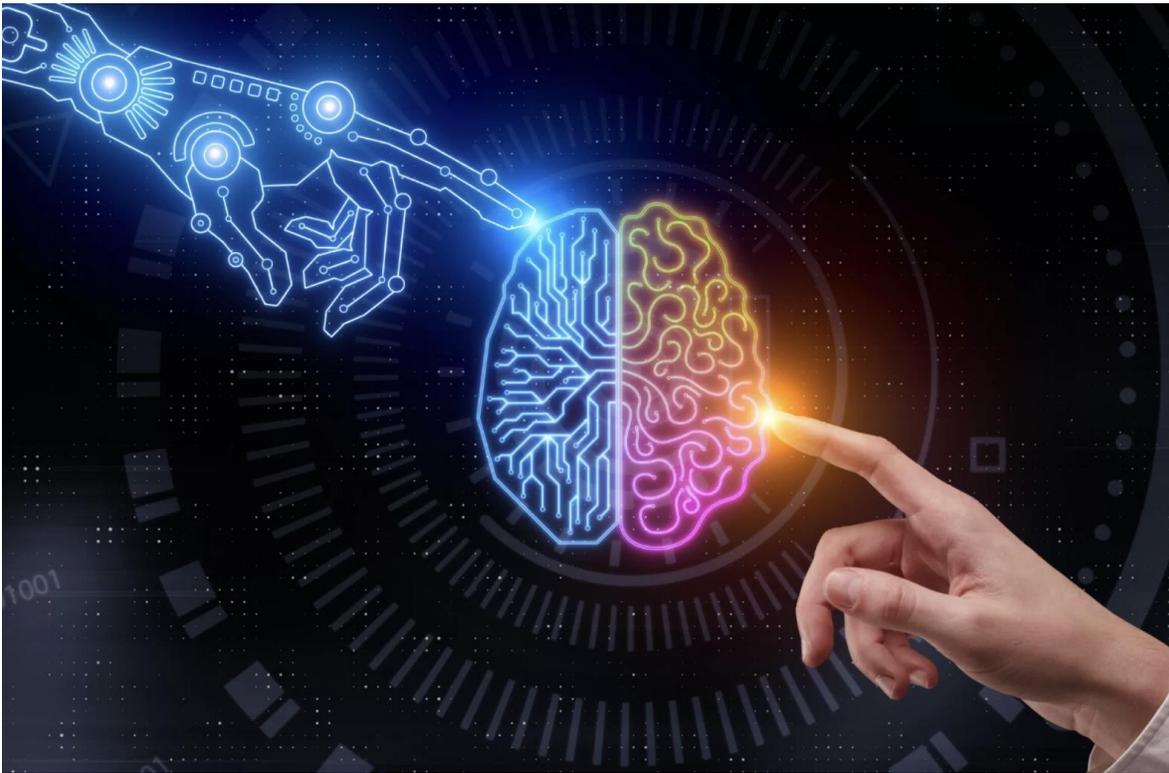
Sequenza proteica: Una sequenza proteica è l'ordine specifico in cui gli amminoacidi sono collegati, uno dopo l'altro, per formare una proteina. Un pò come le lettere in una parola.

Struttura proteica: La struttura proteica si riferisce alla forma tridimensionale precisa in cui una proteina si ripiega per raggiungere la sua forma finale, nella quale svolge la sua funzione biologica.

Design di proteine: Il design di proteine è il processo di progettazione e creazione di nuove protein sintetiche con funzioni specifiche o caratteristiche desiderate, utilizzando principi di biologia molecolare e chimica.



Cosa è l'intelligenza artificiale?



“L'intelligenza artificiale (IA) è l'abilità di una macchina di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività.”
– europa.eu

L'intelligenza artificiale sta accompagnando le nostre vite di tutti i giorni

Assistente Virtuale



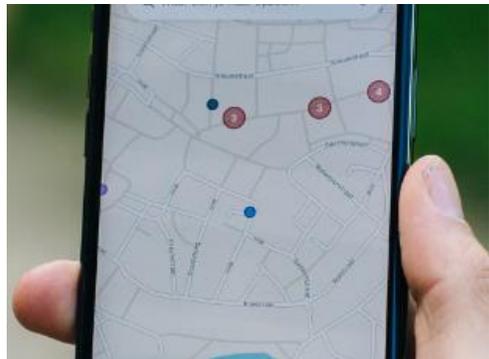
Image credit: unsplash.com

Intrattenimento



L'intelligenza artificiale sta accompagnando le nostre vite di tutti i giorni

Mappe e delivery



Self-driving cars



L'intelligenza artificiale sta accompagnando le nostre vite di tutti i giorni

Generazione testo

Default (GPT-3.5)

DU

Can you please translate the following paragraph in Italian?
Proteins are tiny molecules that play an enormous role in biology. They are found inside every living thing and act as structural components, transporters, signaling molecules, and much more.



Le proteine sono piccole molecole che svolgono un ruolo enorme nella biologia. Si trovano all'interno di ogni essere vivente e agiscono come componenti strutturali, trasportatori, molecole segnali e molto altro.



Generazione immagini



AI nelle scienze

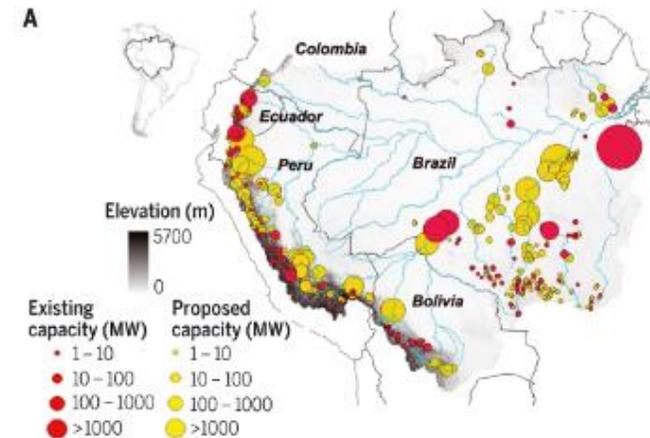


L'AI sta rivoluzionando il modo in cui i ricercatori affrontano le sfide scientifiche

Previsioni meteo



Selezione di siti per la produzione energetica in zone ad alto rischio ambientale

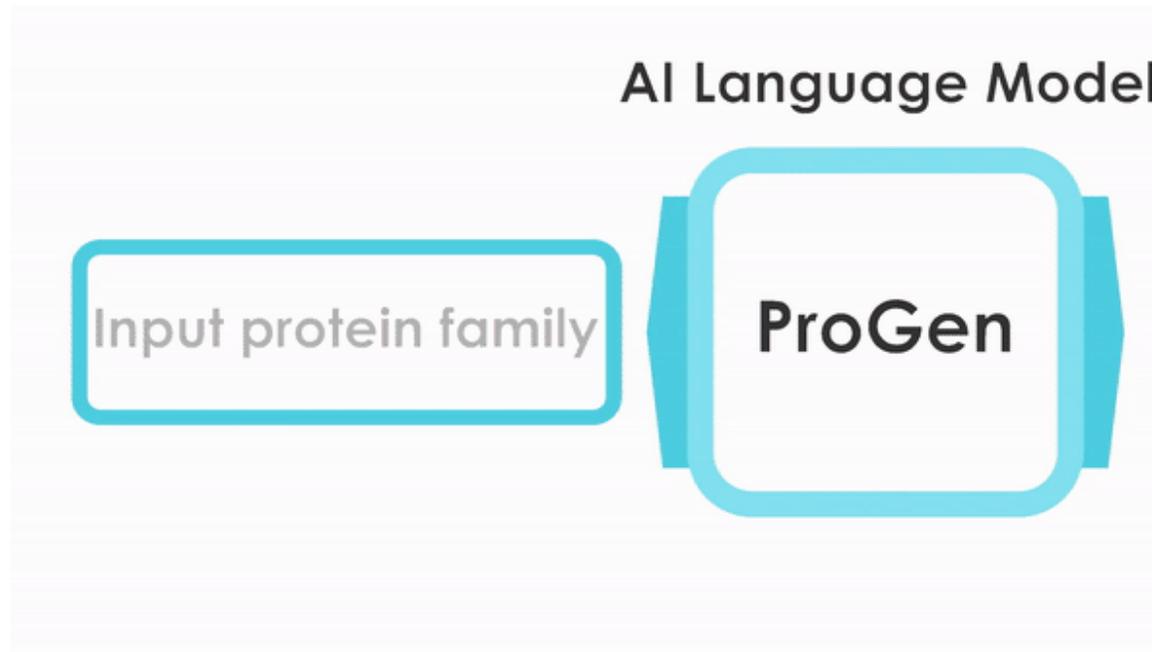


Left: <https://www.deepmind.com/blog/nowcasting-the-next-hour-of-rain>

Right: <https://www.science.org/doi/10.1126/science.abj4017>

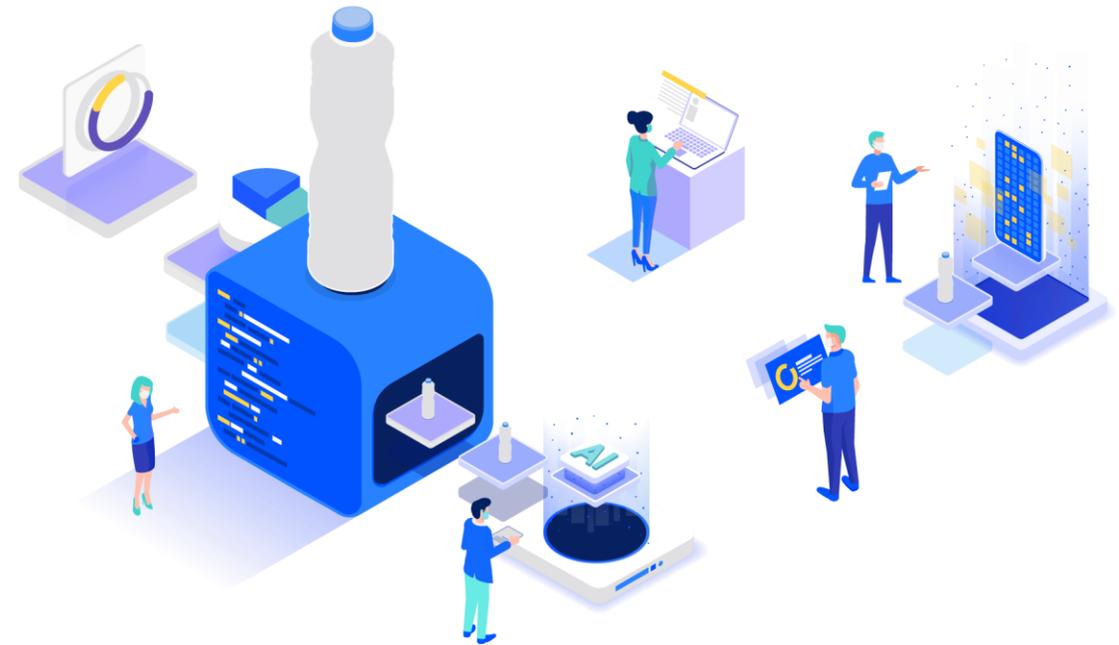
L'AI sta rivoluzionando il modo in cui i ricercatori affrontano le sfide scientifiche

Modelli di linguaggio per sequenze proteiche ed Identificazione di nuovi farmaci su larga scala



Come riesce l'AI ad aiutare gli scienziati a fare meglio il proprio lavoro?

- 1. Analisi dei dati:** L'AI può analizzare enormi quantità di dati scientifici in tempi molto più brevi rispetto agli esseri umani. Questo è cruciale in campi come la genomica, l'astronomia e la ricerca climatica, dove la mole di dati è vastissima
- 2. Ricerca di modelli:** L'AI può individuare modelli e correlazioni nei dati che potrebbero non essere immediatamente evidenti agli esseri umani. Questa capacità è fondamentale per scoprire nuove conoscenze scientifiche.



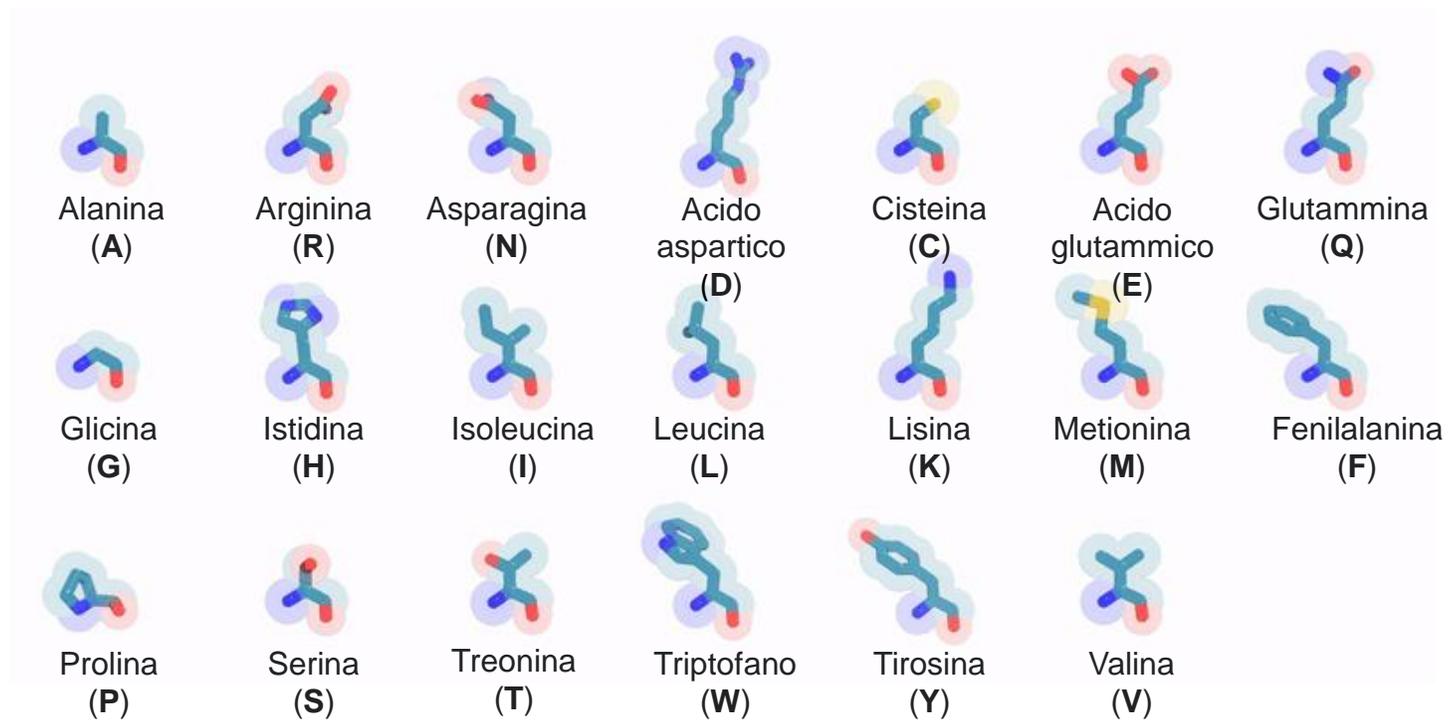
Applicazione pratica al mondo delle proteine



Proteine 101

Le proteine sono macromolecole biologiche fondamentali che svolgono una vasta gamma di funzioni essenziali nella cellula e nell'organismo.

Sono costituite da catene di amminoacidi, che sono gli "ingredienti" di base delle proteine.



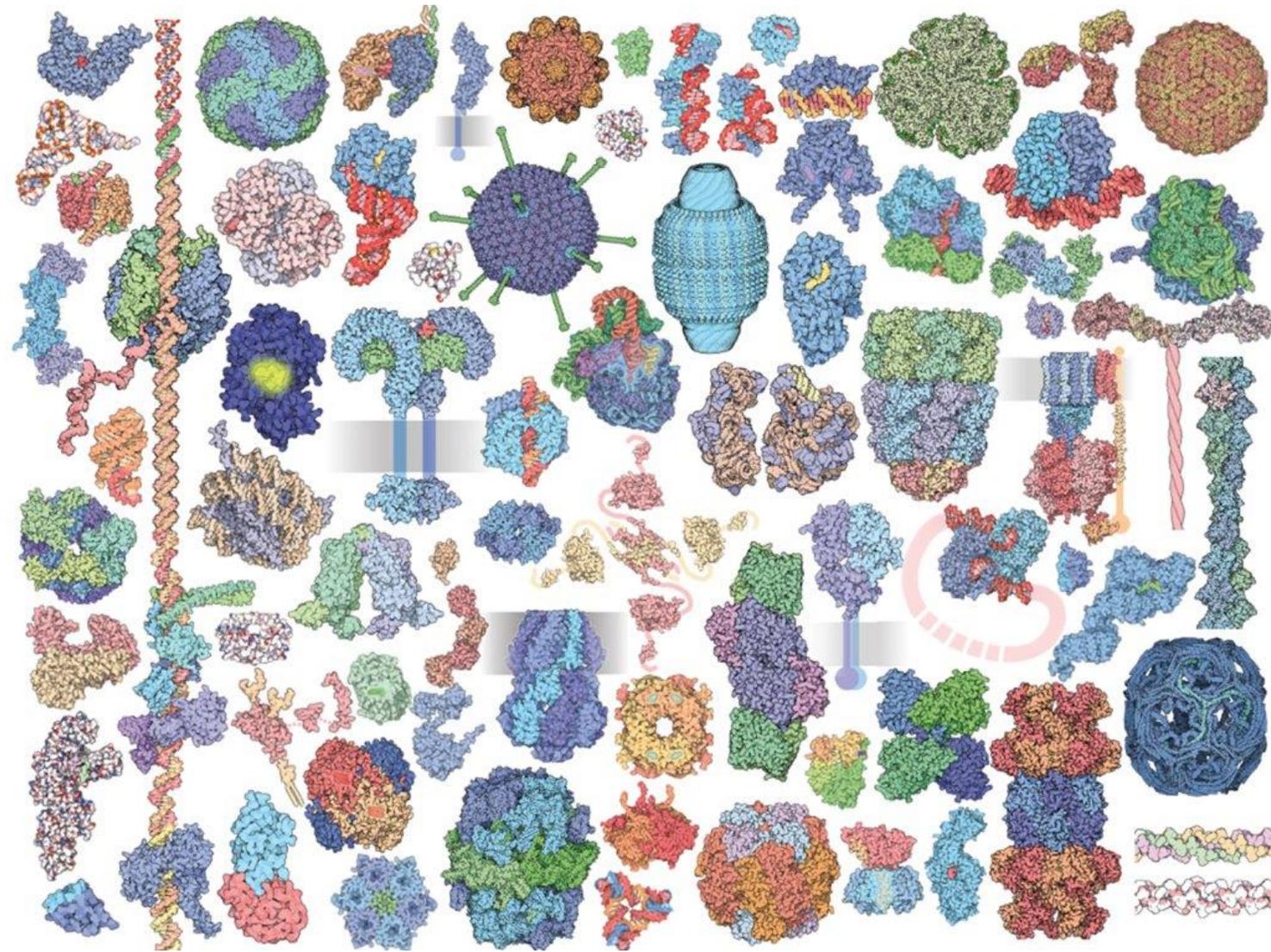
La struttura, forma, di una proteina determina la sua funzione



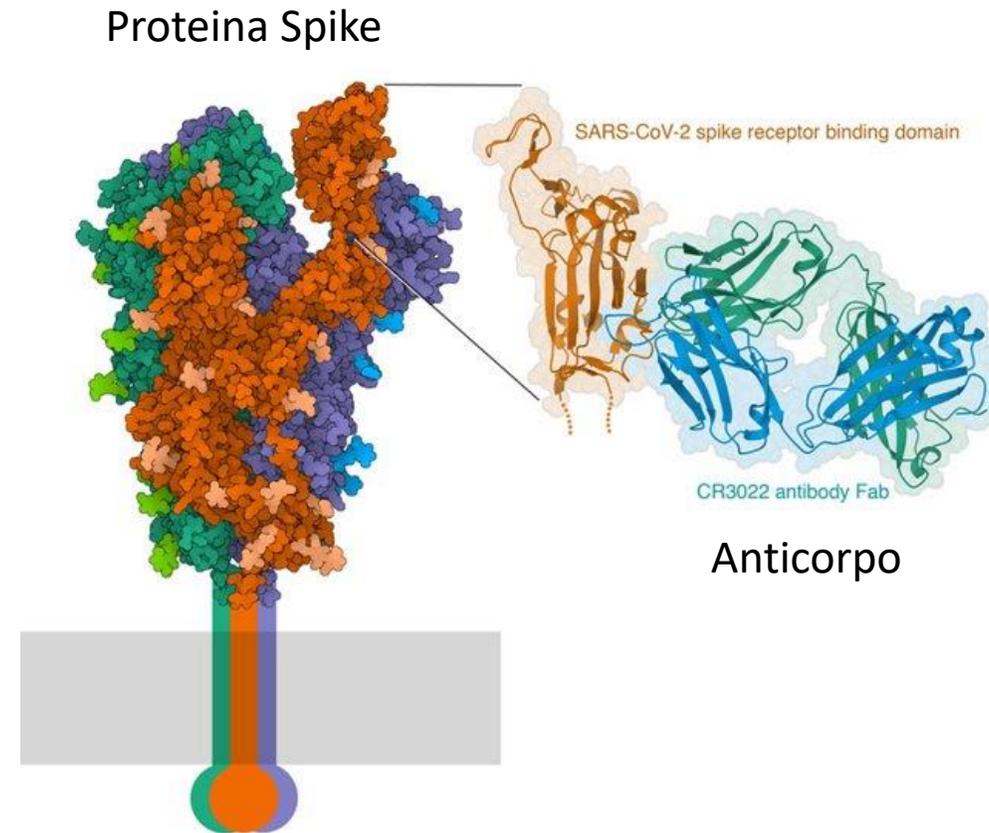
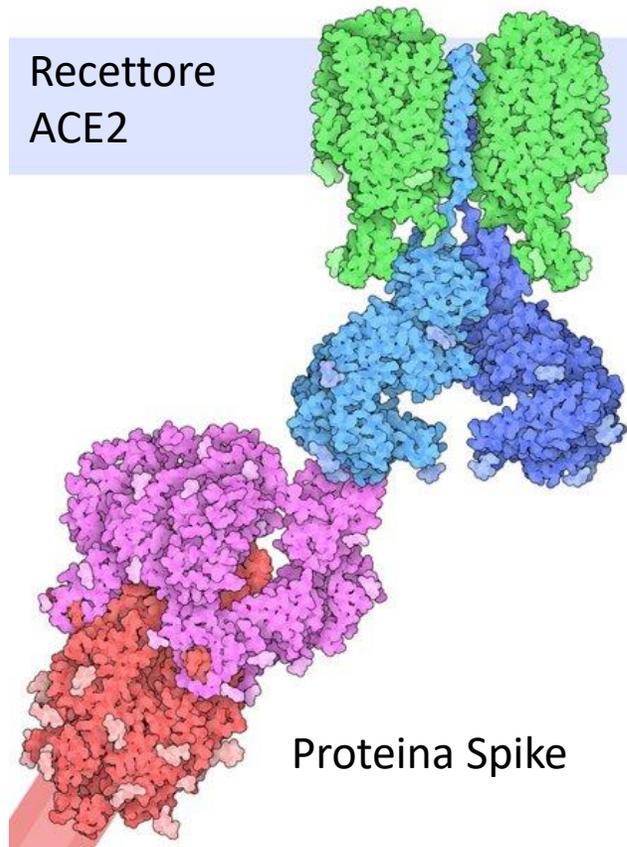
Photocase
Good for your eyes.

margie
photocase.com/1402331

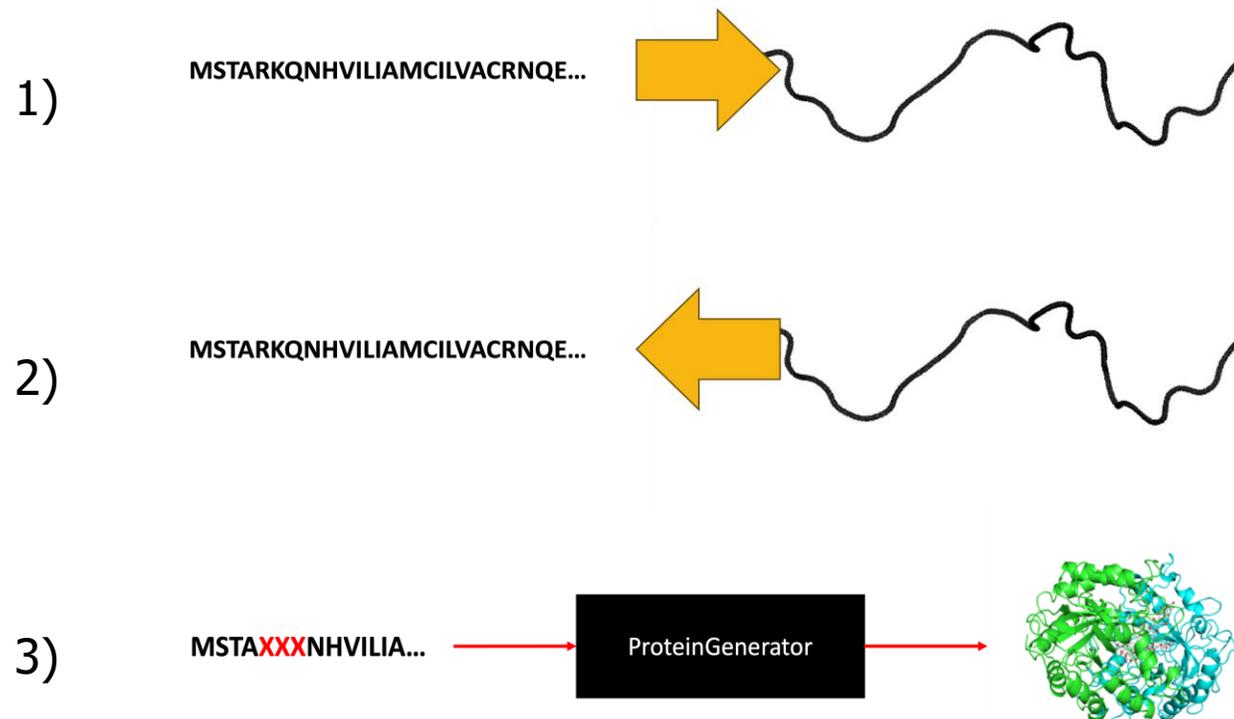
**Sono note
190.000
strutture di
proteine
diverse
presenti in
natura**



Un esempio del binomio struttura e funzione



Studiare la struttura per comprendere la funzione, 3 esempi



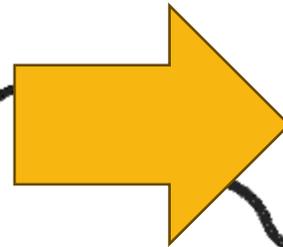
1) Design di proteine: Protein folding

Il problema del folding proteico consiste nel comprendere come una sequenza di amminoacidi si piega per raggiungere la conformazione nativa

*Sequenza
proteina*

Struttura proteina

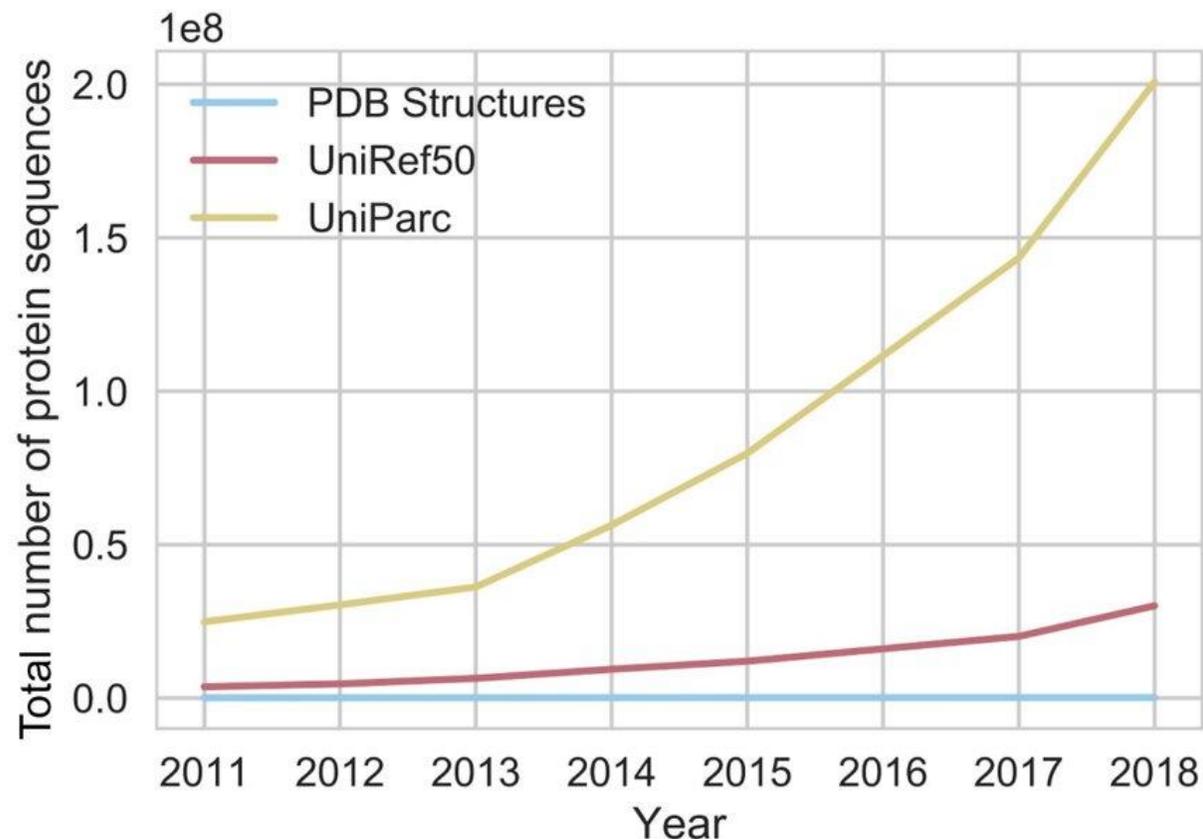
MSTARKQNHVILIAMCILVACRNQE..



Si conoscono molte sequenze e pochissime strutture proteiche

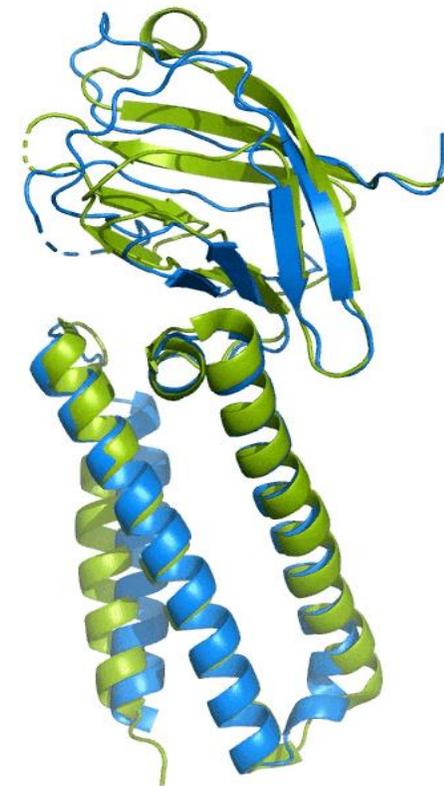
- Il processo sperimentale con il quale si ricava la struttura di una proteina è sia molto costoso che molto lungo
- Fino al 2020 si conosceva la struttura di solo il 1% di tutte le proteine di cui era nota la sequenza.

In pratica partendo dal 1900 venivano scoperte 10 strutture ogni anno.



La prima rivoluzione: AlphaFold

- In soli 18 mesi, [AlphaFold](#) ha previsto la struttura di quasi tutte le proteine finora catalogate dalla scienza.
- Ad oggi il numero di proteine umane senza copertura strutturale è passata da 5027 a 29.
- Tutte le strutture predette da AlphaFold sono presenti su un [database](#) .
 - Questo lavoro ha già portato a progressi nel combattere la malaria, la resistenza agli antibiotici e i rifiuti di plastica, e potrebbe accelerare la scoperta di nuovi farmaci.



- Struttura sperimentale
- Struttura AlphaFold

1) Protein folding – sessione pratica

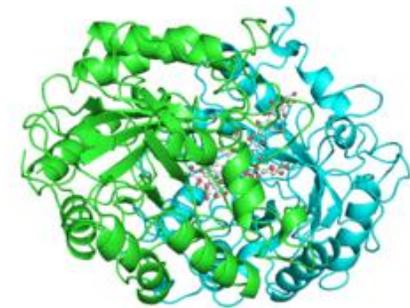
Colab-fold:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=KK7X9T44pWb7>

MSTARKQNHVILIA...



AlphaFold



2) Design di proteine: Inverse folding

L'Inverse Folding, o reverse engineering proteico, è il processo di progettazione di nuove sequenze di amminoacidi che si piegheranno in una struttura proteica specifica con funzioni desiderate.

*Sequenza
proteina*

Struttura proteina

MSTARQNHVILIAMCILVACRNQE...

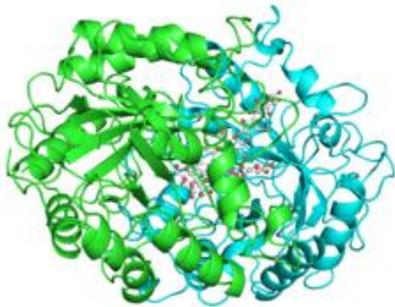


2) Inverse folding – sessione pratica

ProteinMPNN:

https://colab.research.google.com/github/dauparas/ProteinMPNN/blob/main/colab_notebooks/quickdemo_w_AF2.ipynb#scrollTo=0TNhcwok8d_w

MSTARKQNHVIL
IA...



ProteinMPNN

MSAARKQNHVILIA...

MSTATKPNHVILIA...

MSTATKPNHVICIA...

MHTATKPNHVICIA...

Il design di proteine aiuterà la medicina, lo sviluppo di vaccini e la sostenibilità ambientale

Medicina: Utilizziamo la progettazione per sviluppare terapie con alta specificità e bassa tossicità; creare proteine che si legano selettivamente alle cellule tumorali e altre che inibiscono l'infezione virale, aprendo nuove strade per lo sviluppo di farmaci.

Vaccini: creare nanoparticelle proteiche auto-assemblanti che agiscono come vaccini altamente efficaci. Queste proteine sintetiche imitano la struttura delle proteine virali, stimolando una robusta risposta immunitaria senza necessità di virus vivi.

Sostenibilità ambientale: le proteine possano essere utilizzate per creare nuovi materiali biodegradabili, catturare il carbonio e decomporre inquinanti ambientali, offrendo soluzioni innovative alle sfide ambientali.

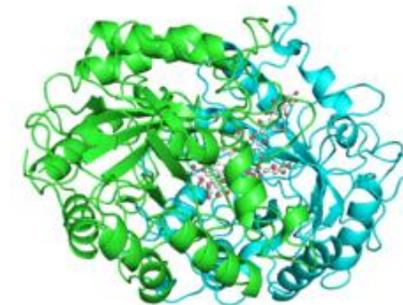
3) Design di proteine – sessione pratica

- Video: <https://www.youtube.com/watch?v=bS71K2U0amA>
- Hugging Face: https://huggingface.co/spaces/merle/PROTEIN_GENERATOR

MSTAXXXNHVILI
A...



ProteinGenerator





Fondazione Toscana Life Sciences
Via Fiorentina, 1 - 53100 Siena ITALY
www.toscanalifesciences.org

