UNIVERSITÀ DEGLI STUDI FIRENZE

Da un secolo, oltre.

HR EXCELLENCE IN RESEARCH

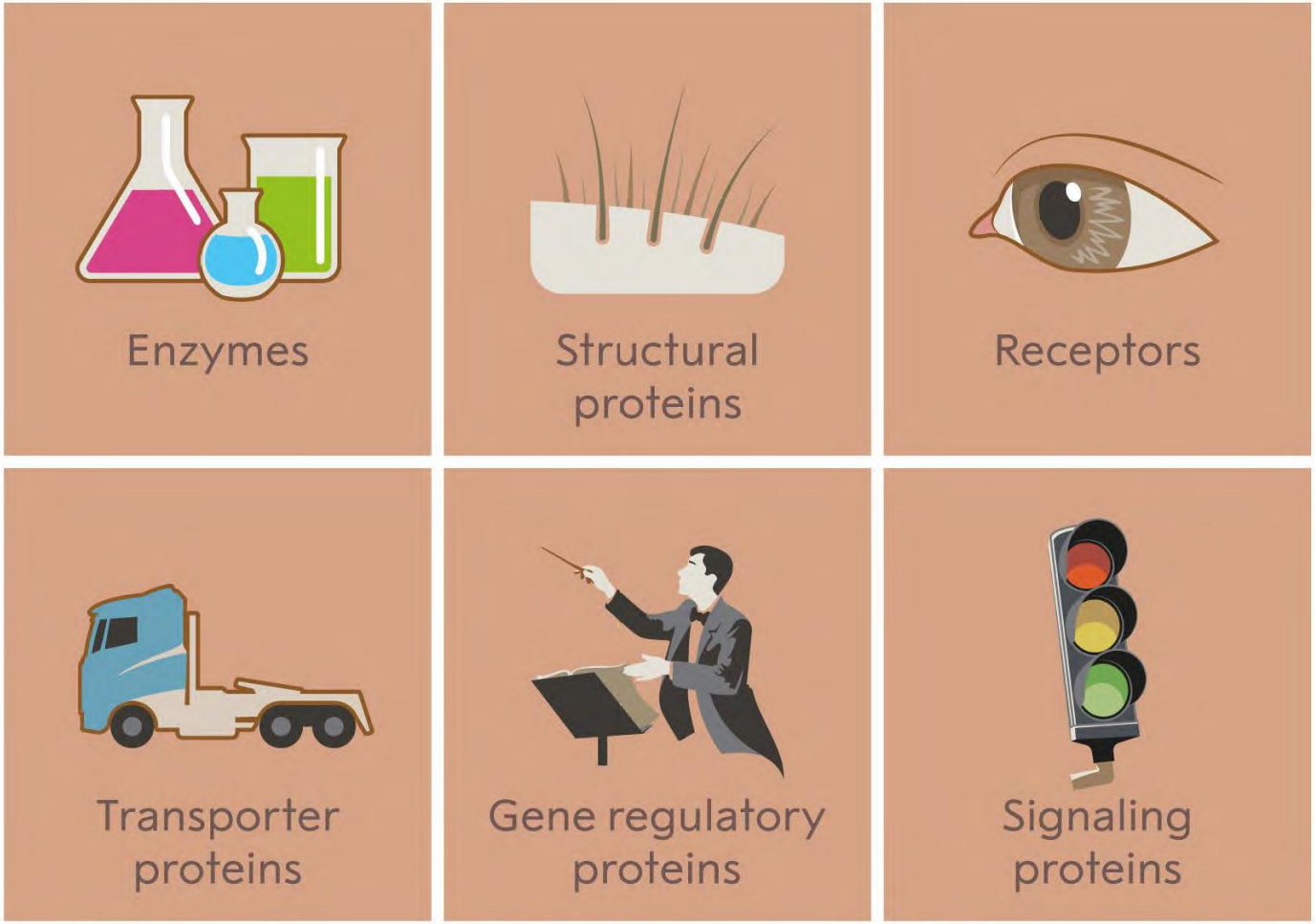
I segreti delle proteine svelati dall'intelligenza artificiale

La soluzione a un problema cinquantennale

Antonio Rosato

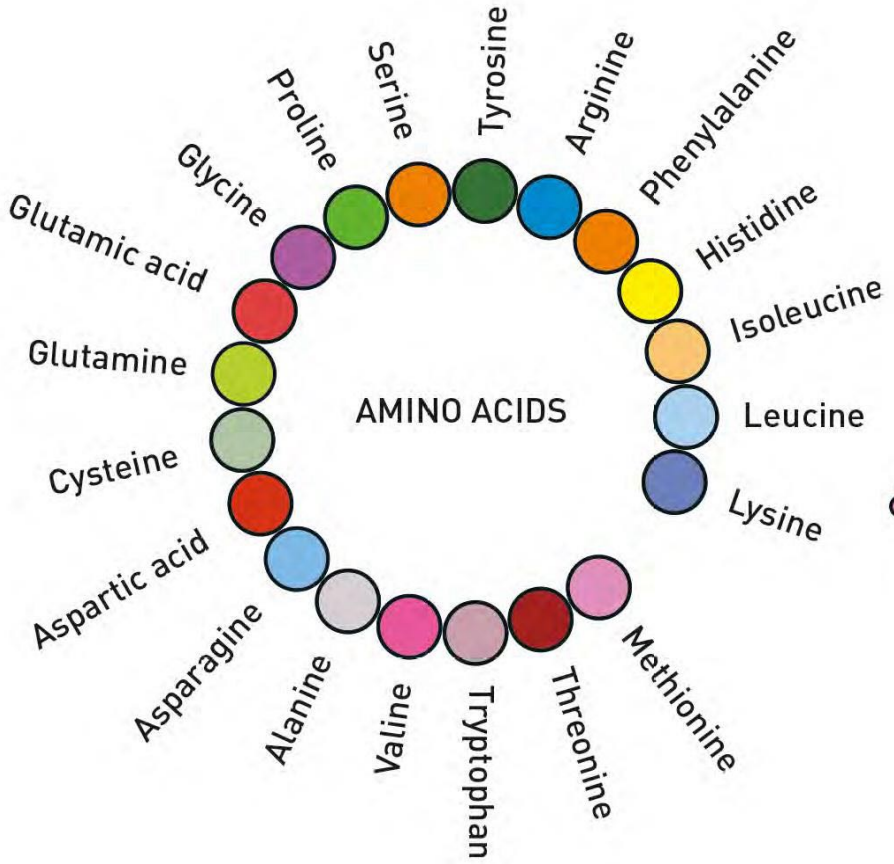
antonio.rosato@unifi.it

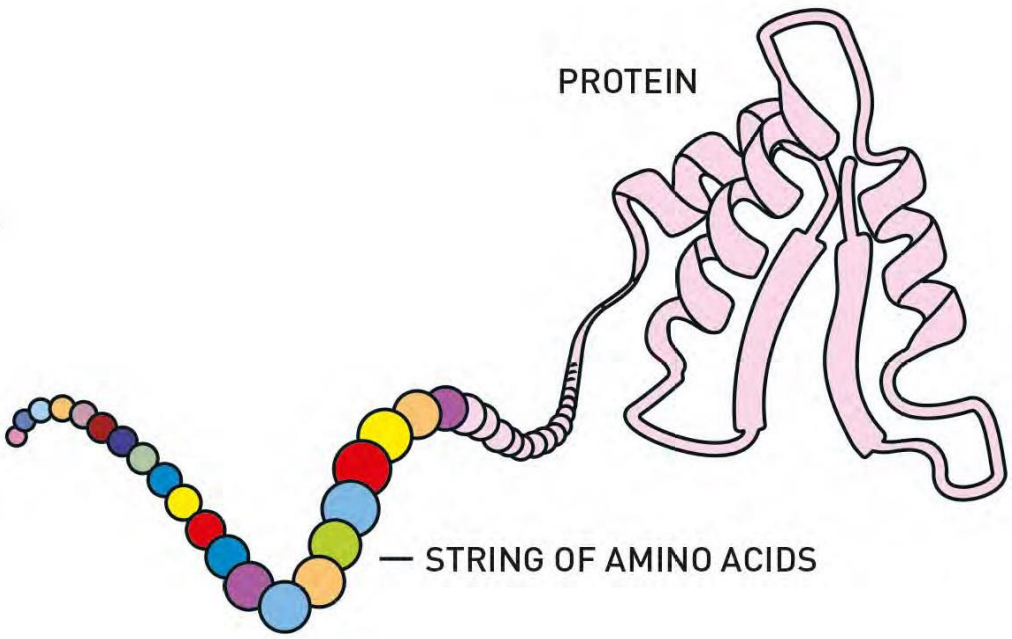
Le proteine sono le molecole della vita



Le proteine si trovano in tutti gli esseri viventi, inclusi i virus, e svolgono un grande numero di funzioni diverse

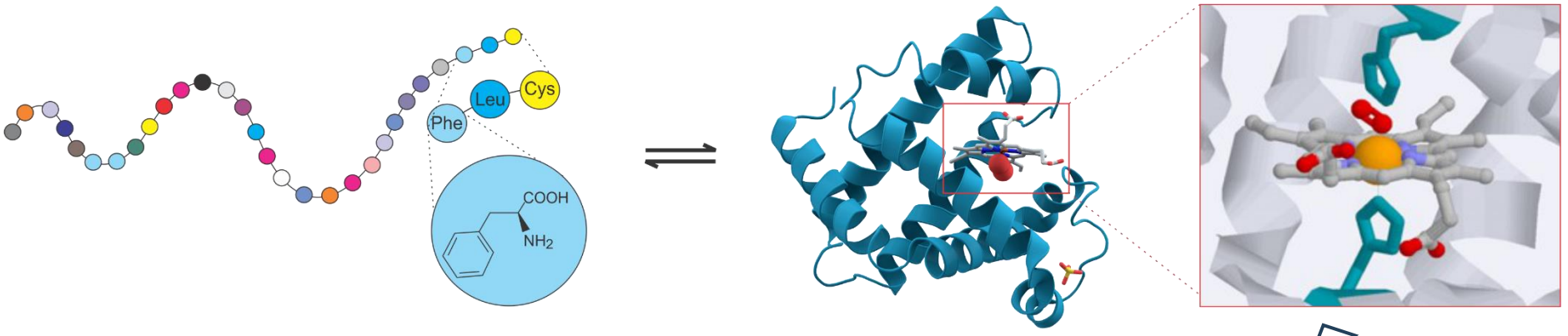
Le proteine sono costituite da aminoacidi





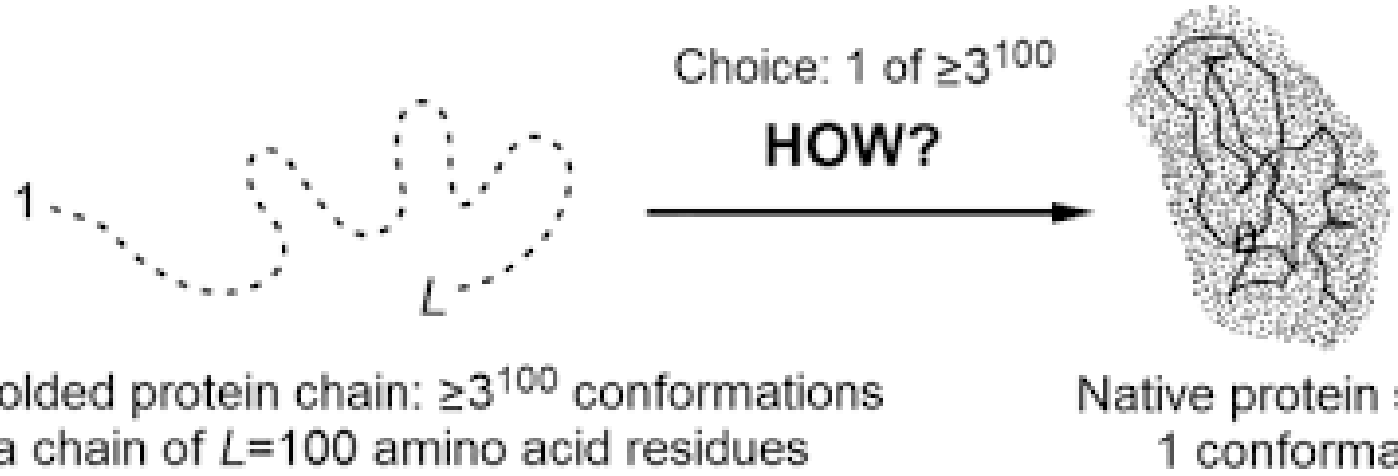
Le proteine sono formate combinando i 20 aminoacidi in catene contenenti varie centinaia di questi blocchi chimici

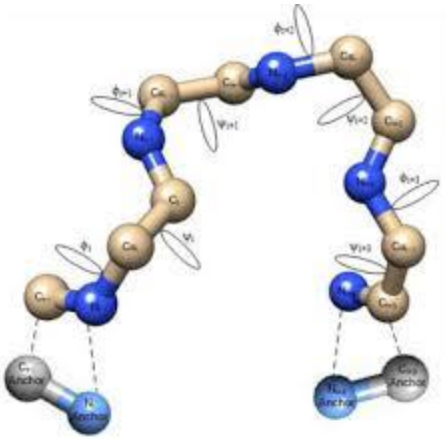
Il paradigma sequenza-struttura-funzione



La catena degli aminoacidi si avvolge (**struttura 3D**) in una maniera unicamente determinata dalla sua sequenza.
La struttura 3D determina la reattività (bio)chimica della proteina, p.es. formando **siti di legame** per altre molecole.

Il paradosso di Levinthal

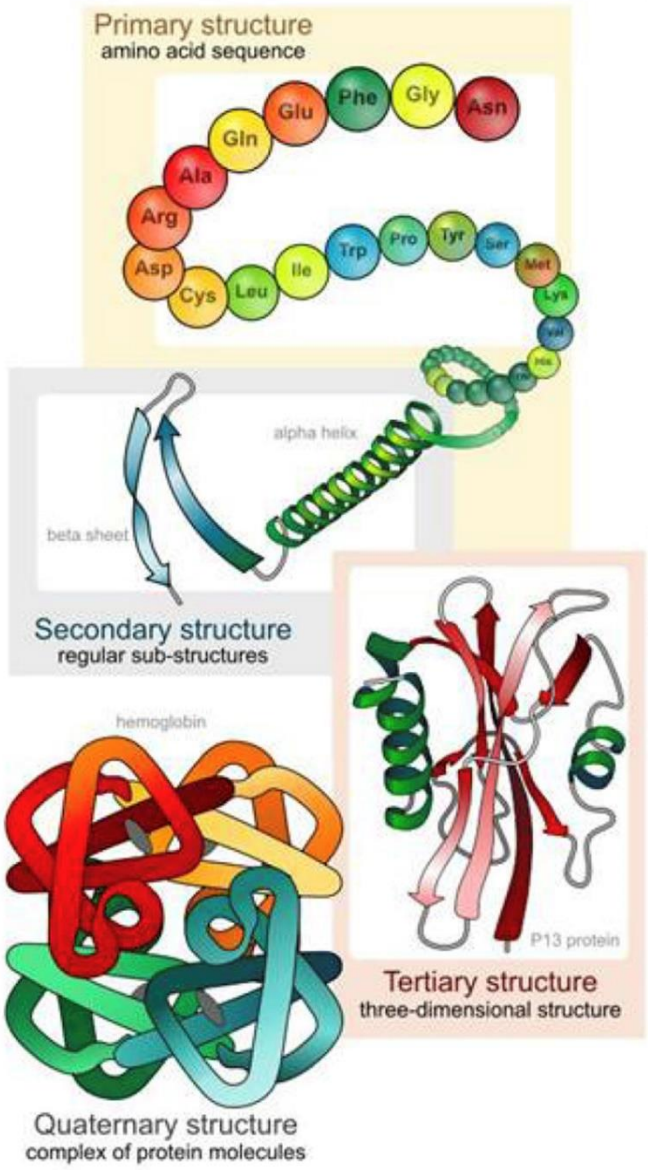


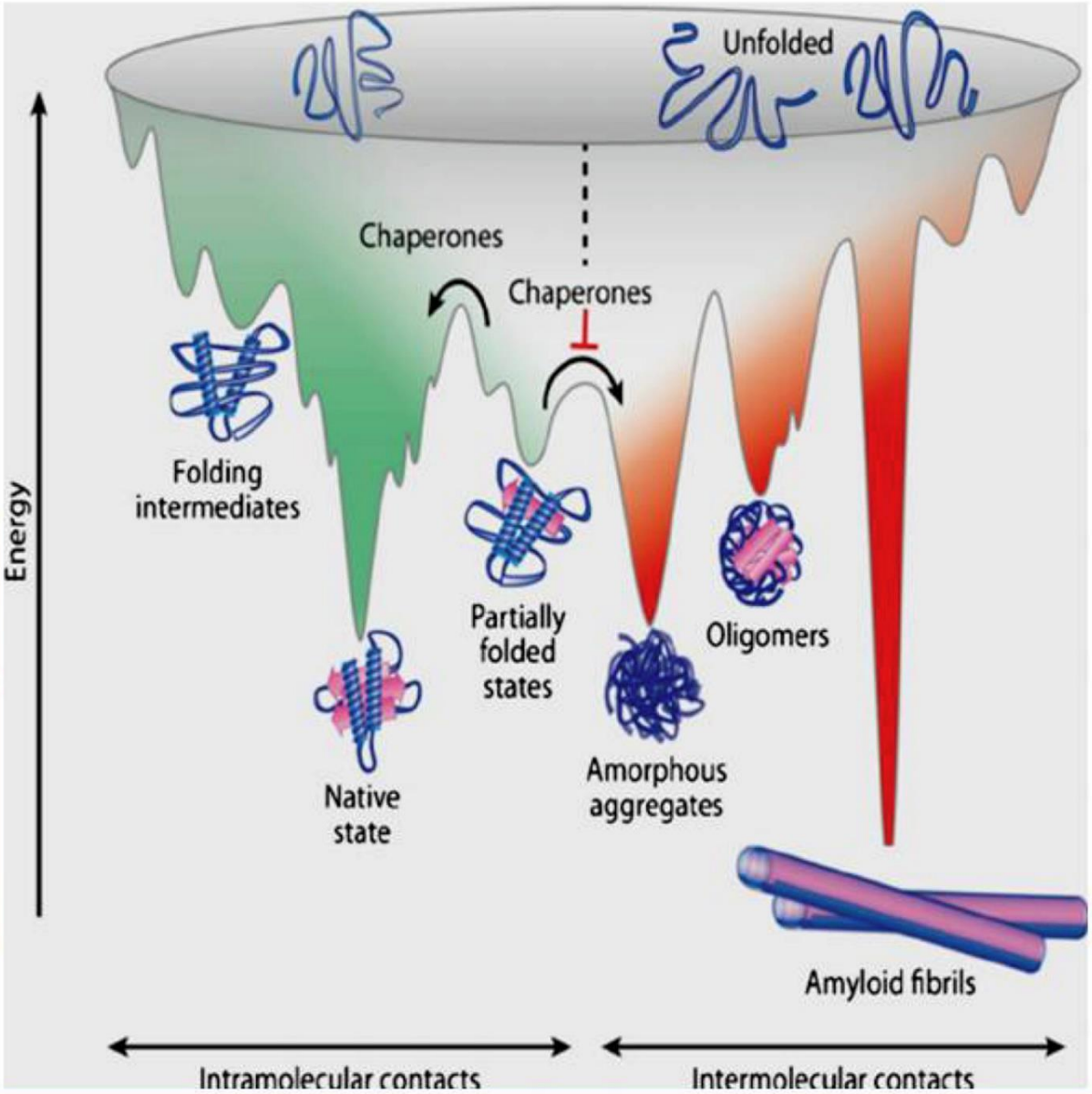


Tenendo conto dei gradi di libertà disponibili, una proteina di 100 aminoacidi può assumere $5 \cdot 10^{47}$ conformazioni. Se dovesse campionarle tutte per trovare la sua conformazione nativa, anche alla velocità di una ogni attosecondo, impiegherebbe 10^{22} anni per avvolgersi.

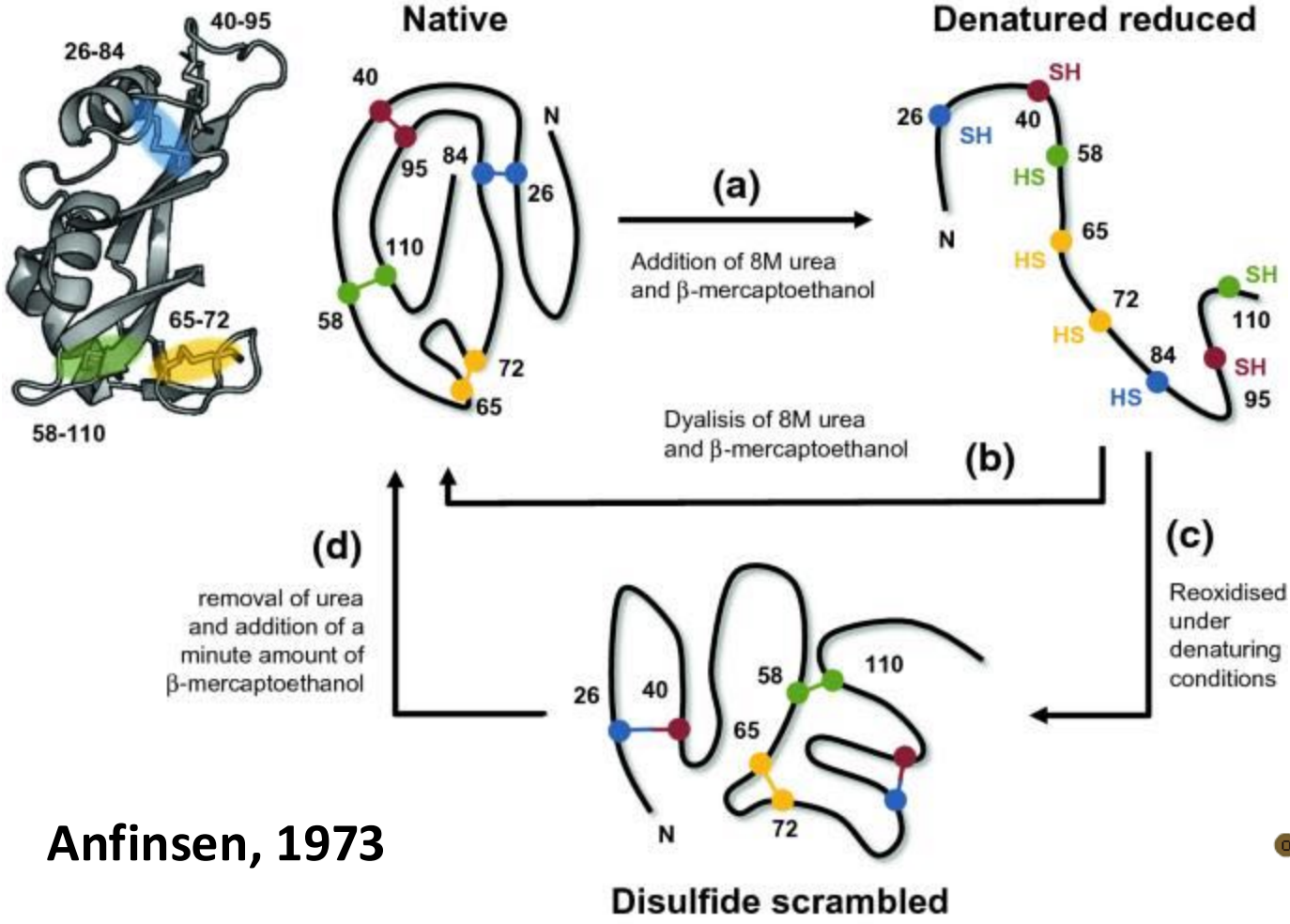
Per contro, una cellula di *E. coli* impiega alcune decine di minuti per replicarsi.

The dynamics of protein folding and misfolding



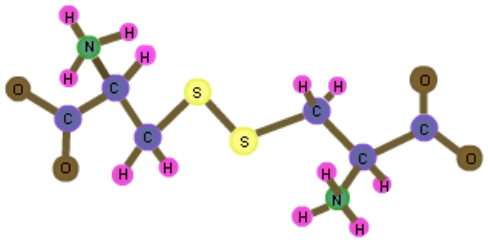


La sequenza determina la struttura

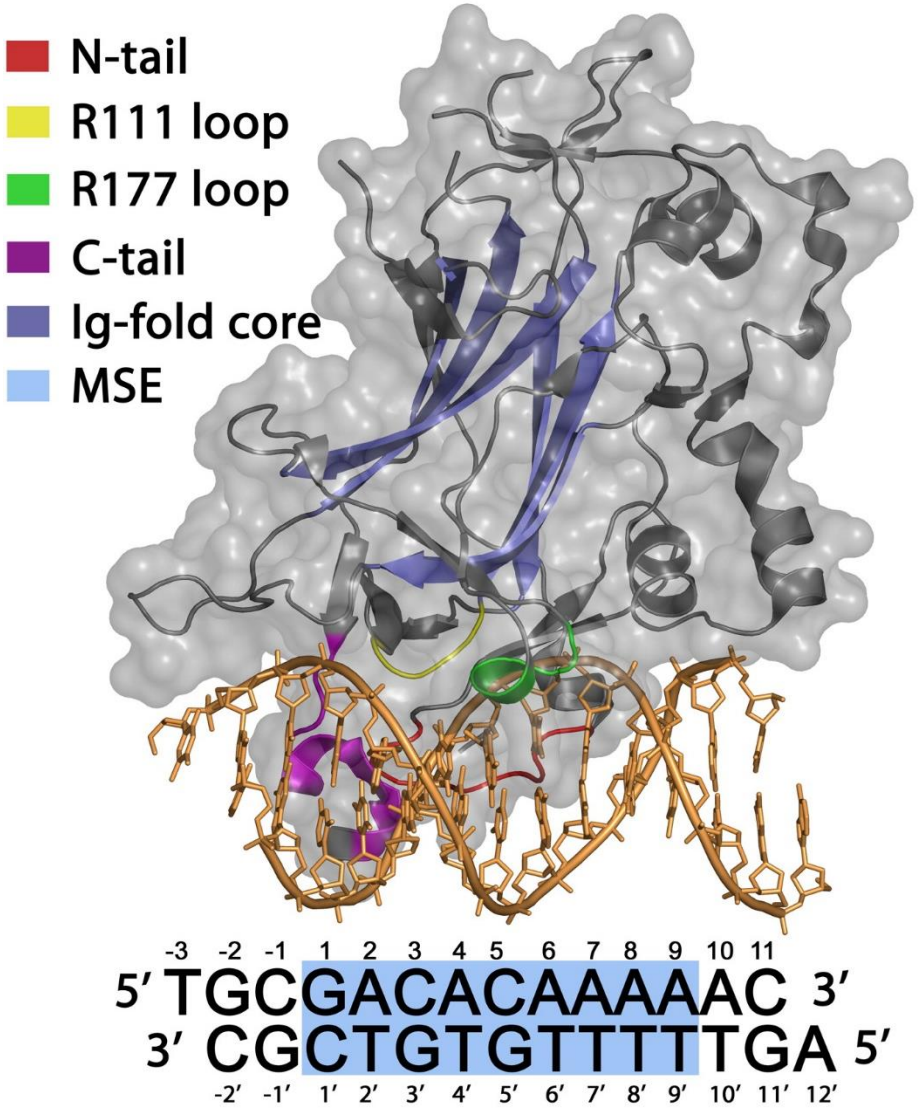


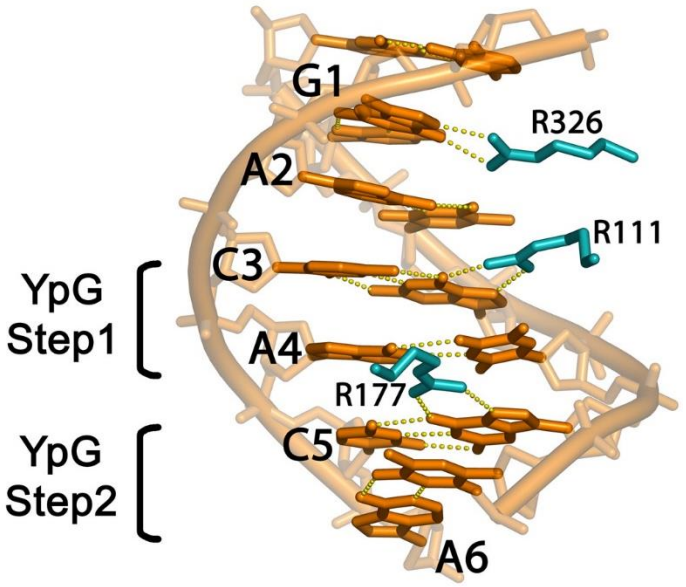
a) +8M urea +βME
b) -urea -βME
c) +8M urea -βME
d) -urea +βME (in minima quantità)

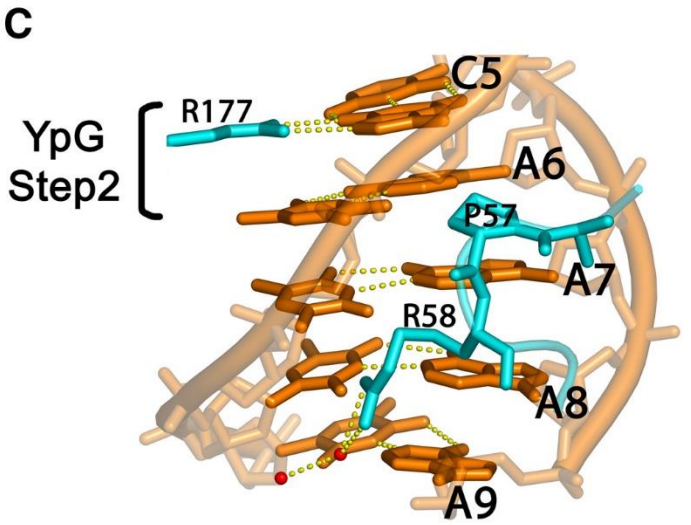
Anfinsen, 1973



La struttura determina la funzione







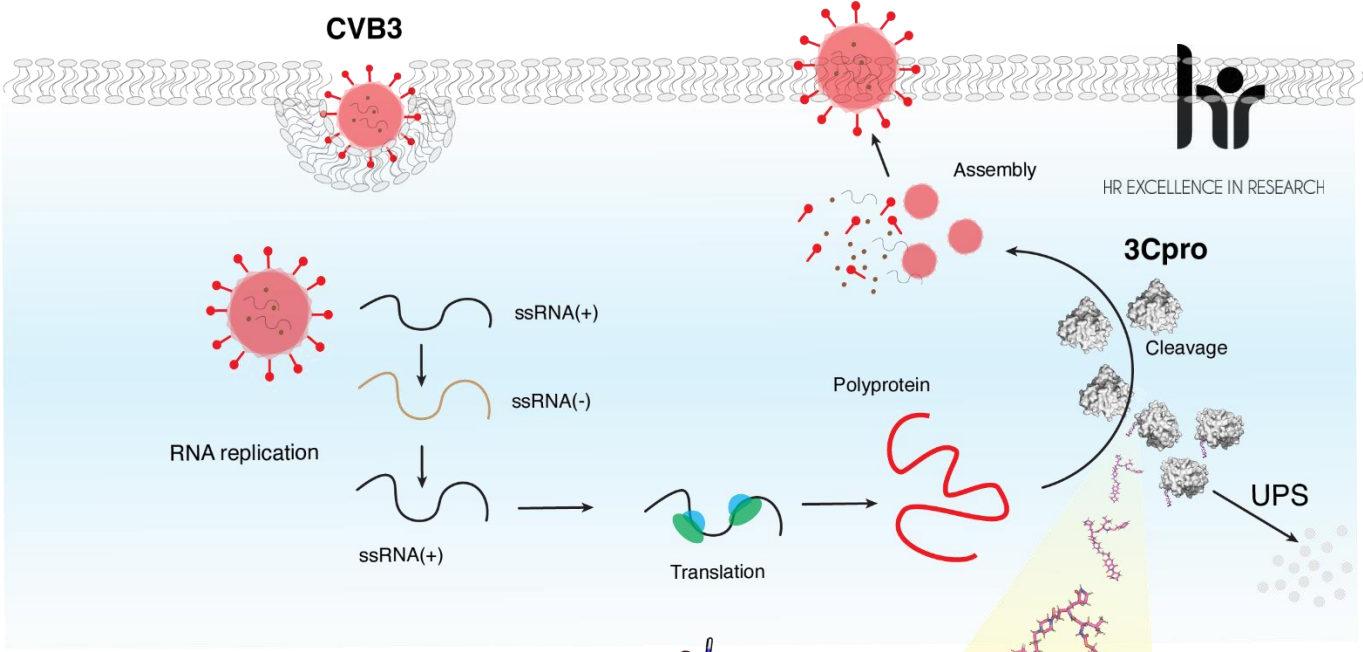
Principles of Protein-DNA Recognition … Lamoureux & Glover, *Structure*, 2006

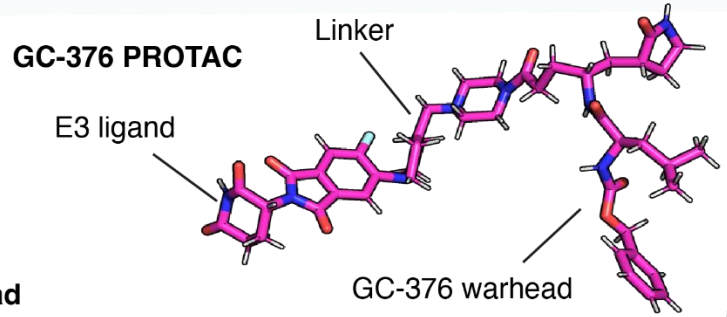
UNIVERSITÀ DEGLI STUDI FIRENZE

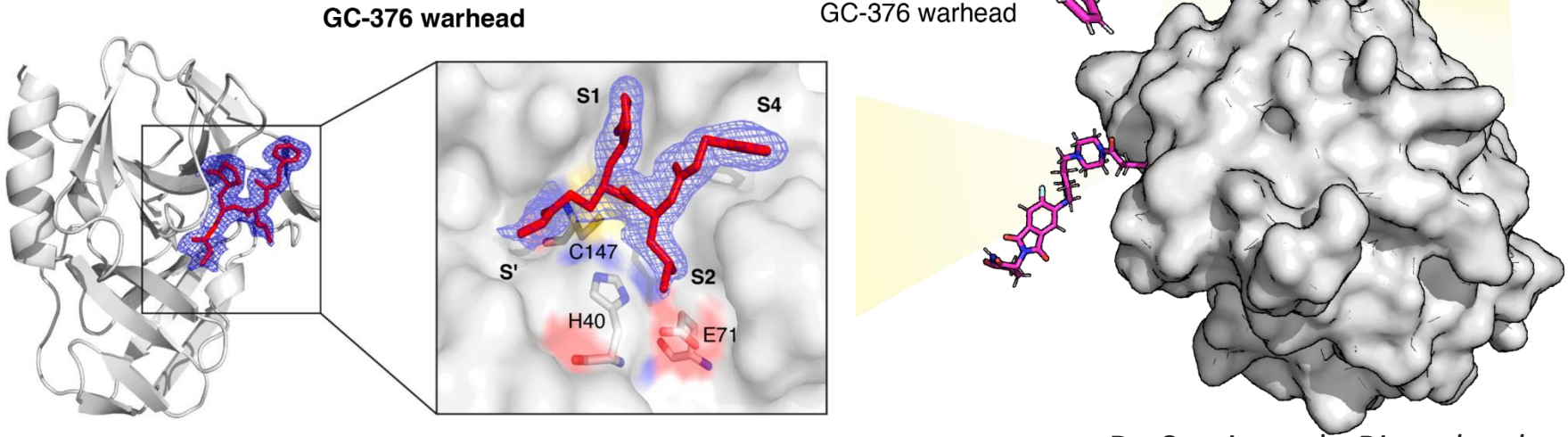
Da un secolo, oltre.

Conoscere la struttura è utile

Image courtesy of Andrea Orsetti







De Santis et al., *Biomolecules*, 2024

Riassumendo …

- Per capire pienamente il ruolo biologico di una proteina bisogna conoscerne sia la sequenza sia la struttura 3D
- Ogni proteina (globulare) raggiunge una conformazione 3D stabile in un tempo relativamente breve, attraverso un meccanismo di progressiva stabilizzazione energetica
- La conformazione 3D stabile dipende solo dalla sequenza aminoacidica (dimostrato nel 1973)

✓ Pertanto, nota la sequenza aminoacidica p.es. tramite sequenziamento genico, dovrebbe essere possibile predirne la struttura 3D sulla base di calcoli di energia e imparare qualcosa sulla funzione della proteina

UNIVERSITÀ DEGLI STUDI FIRENZE
Da un secolo, oltre.

Determinare sperimentalmente la struttura 3D di una proteina richiede tempo e investimenti

HR EXCELLENCE IN RESEARCH

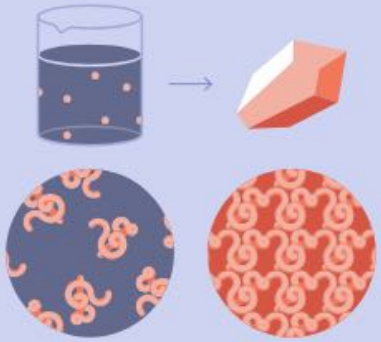
How X-Ray Crystallography Works

For decades, structural biologists primarily used X-ray crystallography to deduce protein structures. This technique reconstructs the shape of a protein from how its crystallized form scatters X-rays.

1

Protein in solution

Crystallized protein



Purified proteins in solution are grown into a crystal, locking the structures into a fixed pattern. Because not all proteins crystallize easily, this step can take years of trial and error.

2

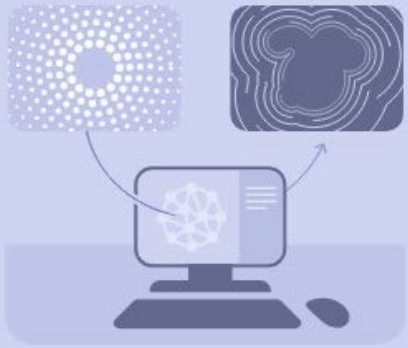
Detector

The protein crystal is bombarded with X-rays from many angles. The X-rays bounce off the atoms' electrons and scatter in different directions. A detector determines the "diffraction pattern" of scattering.

3

Diffraction pattern

Electron density map

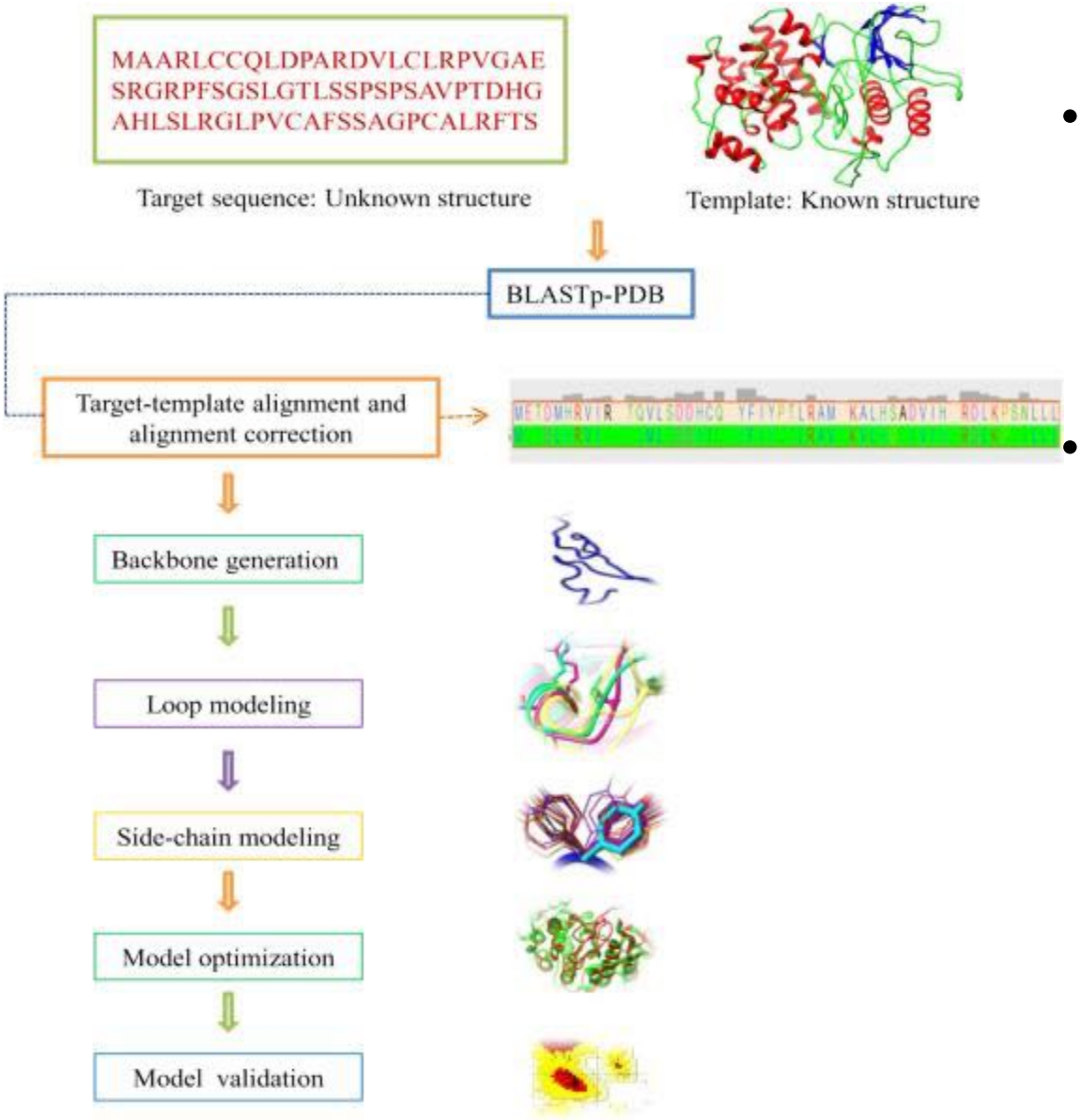


A computer processes the diffraction patterns to build an electron density map that shows where electrons congregate in the crystal.

4

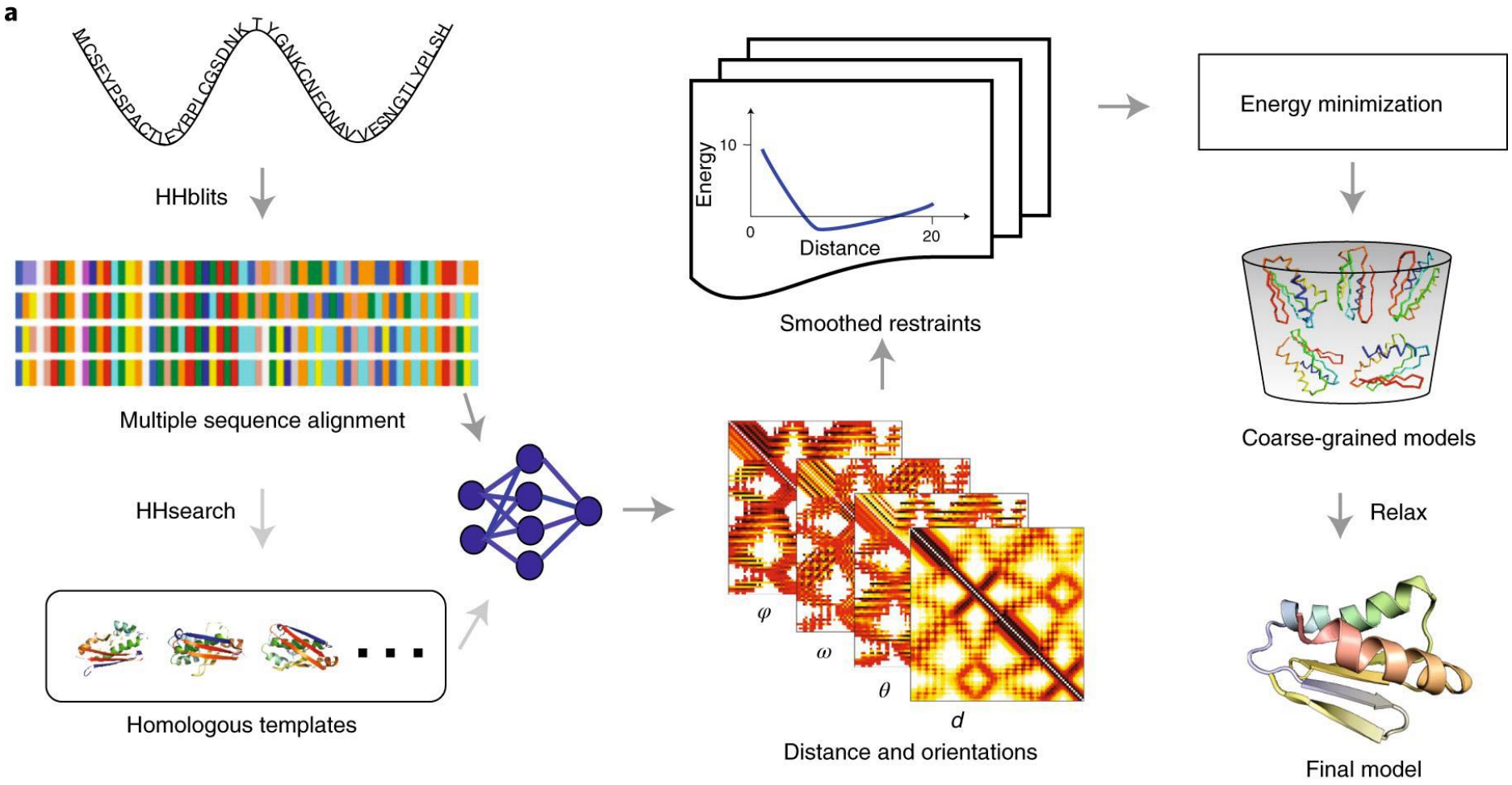
Constructed protein

The electron density map is used to build a model of the protein. Where electrons congregate, an atom is likely to be present.



- Predire la struttura 3D di una proteina sulla base del calcolo dell'energia delle diverse conformazioni è troppo lungo e costoso
- Si è fatto quindi ricorso a metodi alternativi, quali il **modeling per omologia**, basato sul riconoscere somiglianze fra la mia proteina d'interesse e una già caratterizzata

Il programma trRosetta di David Baker & Co.



Proc. Natl Acad. Sci. USA **117**, 1496–1503 (2020)

Una competizione internazionale per la predizione strutturale delle proteine

Critical assessment of structure prediction (CASP) is a biennial community experiment to advance methods of computing three-dimensional protein structure from amino acid sequence

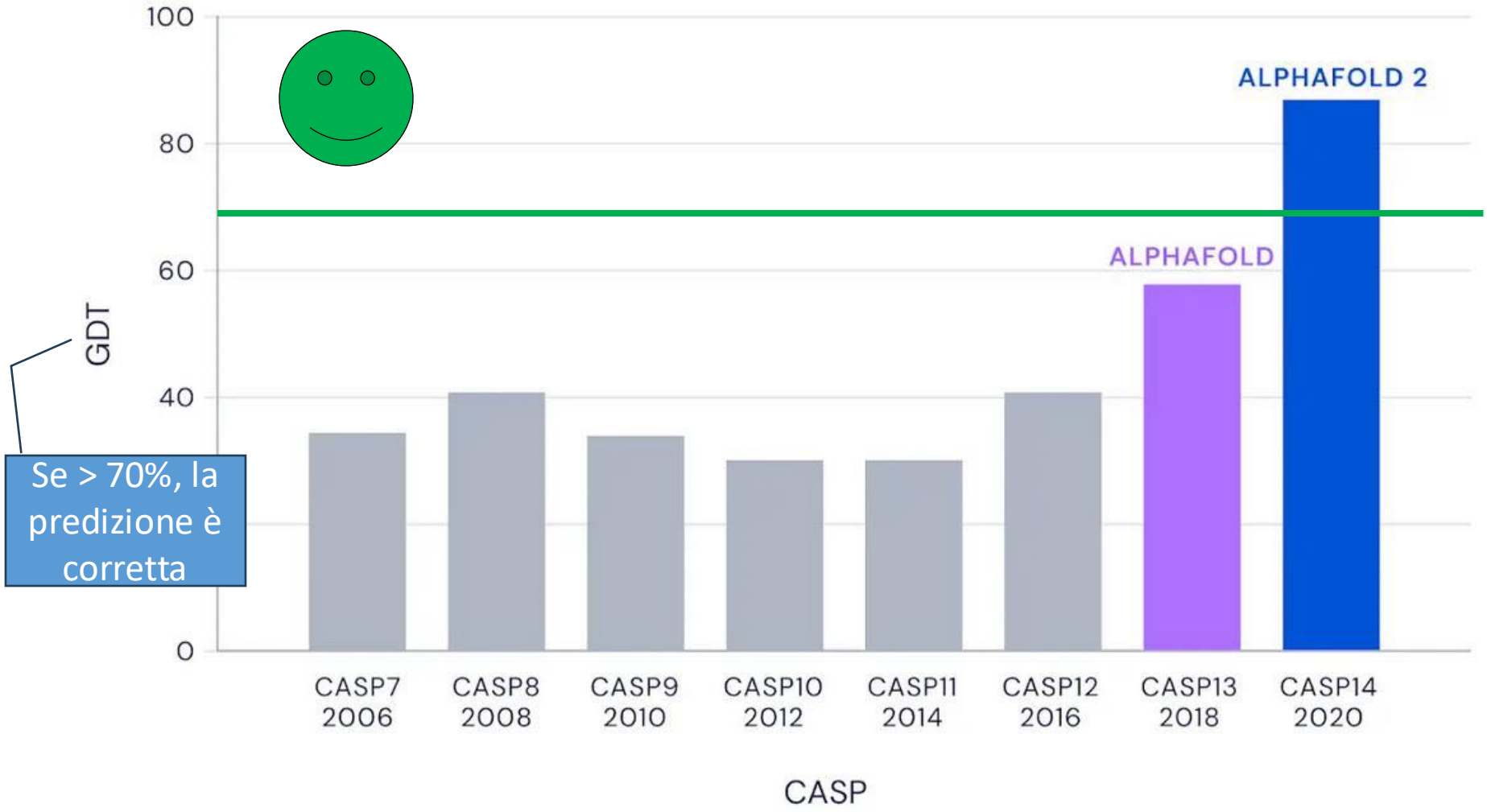
Gli organizzatori di CASP raccolgono decine di nuove strutture 3D di proteine determinate da ricercatori in tutto il mondo e ne sottopongono la sequenza aminoacidica ai partecipanti alla competizione **prima** che le strutture sperimentali siano rese pubbliche

Ogni due anni si effettua un confronto sistematico fra predizioni e strutture sperimentali

Prima edizione nel 1994, tuttora attivo

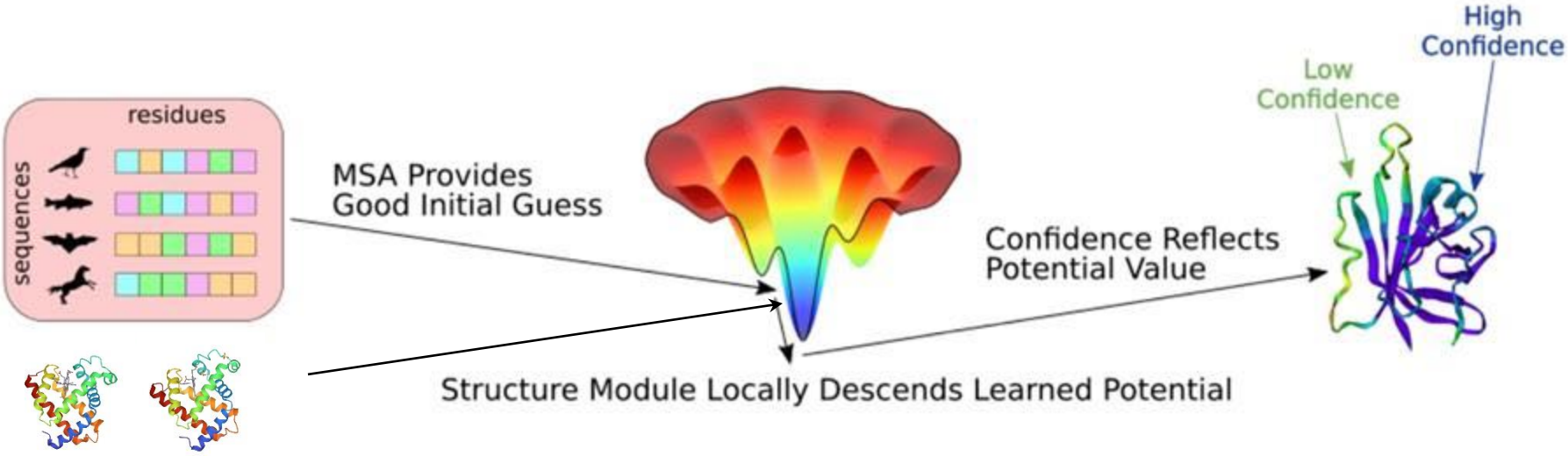
Nel 2020 AlphaFold 2 vince

Prestazione media su tutte le predizioni effettuate dal migliore programma in ogni edizione



Se > 70%, la predizione è corretta

Dall'analisi statistica dei database di strutture di proteine (Protein Data Bank), AlphaFold ha imparato a capire se **l'energia** di un possibile avvolgimento della catena polipeptidica **è soddisfacente oppure no**



AlphaFold è composto da due parti: la prima stima una conformazione iniziale, la seconda ne valuta l'energia e la ottimizza

A partire dal 2020, le predizioni di AlphaFold sono in media corrette

One of biology's biggest mysteries 'largely solved' by AI

By Helen Briggs
BBC science correspondent

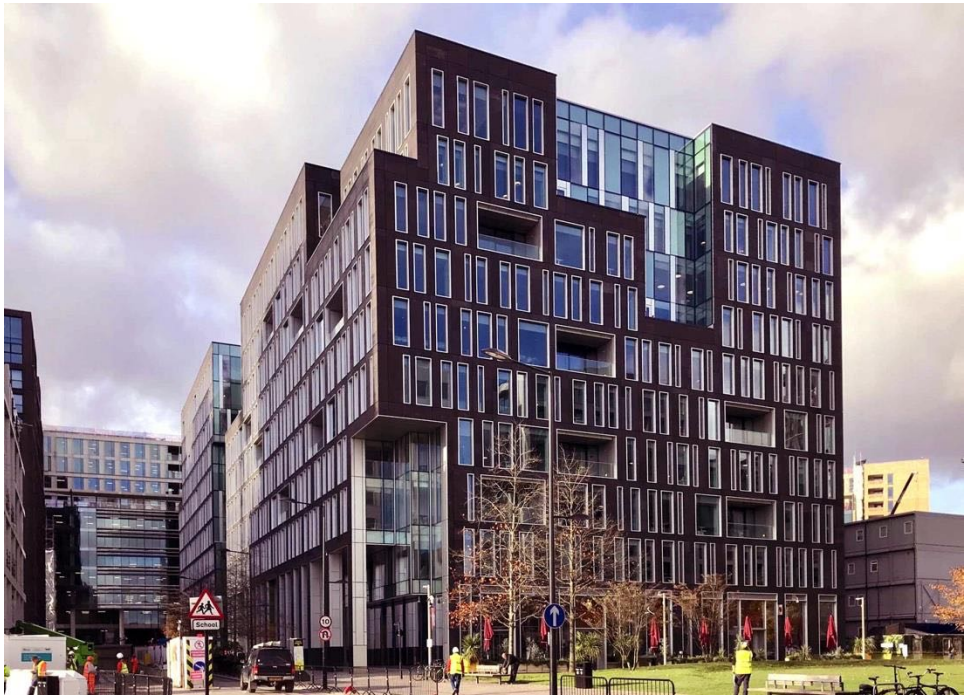
NEWS | 30 November 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

A.I. Predicts the Shapes of Molecules to Come

DeepMind has given 3-D structure to 350,000 proteins, including every one made by humans, promising a boon for medicine and drug design.

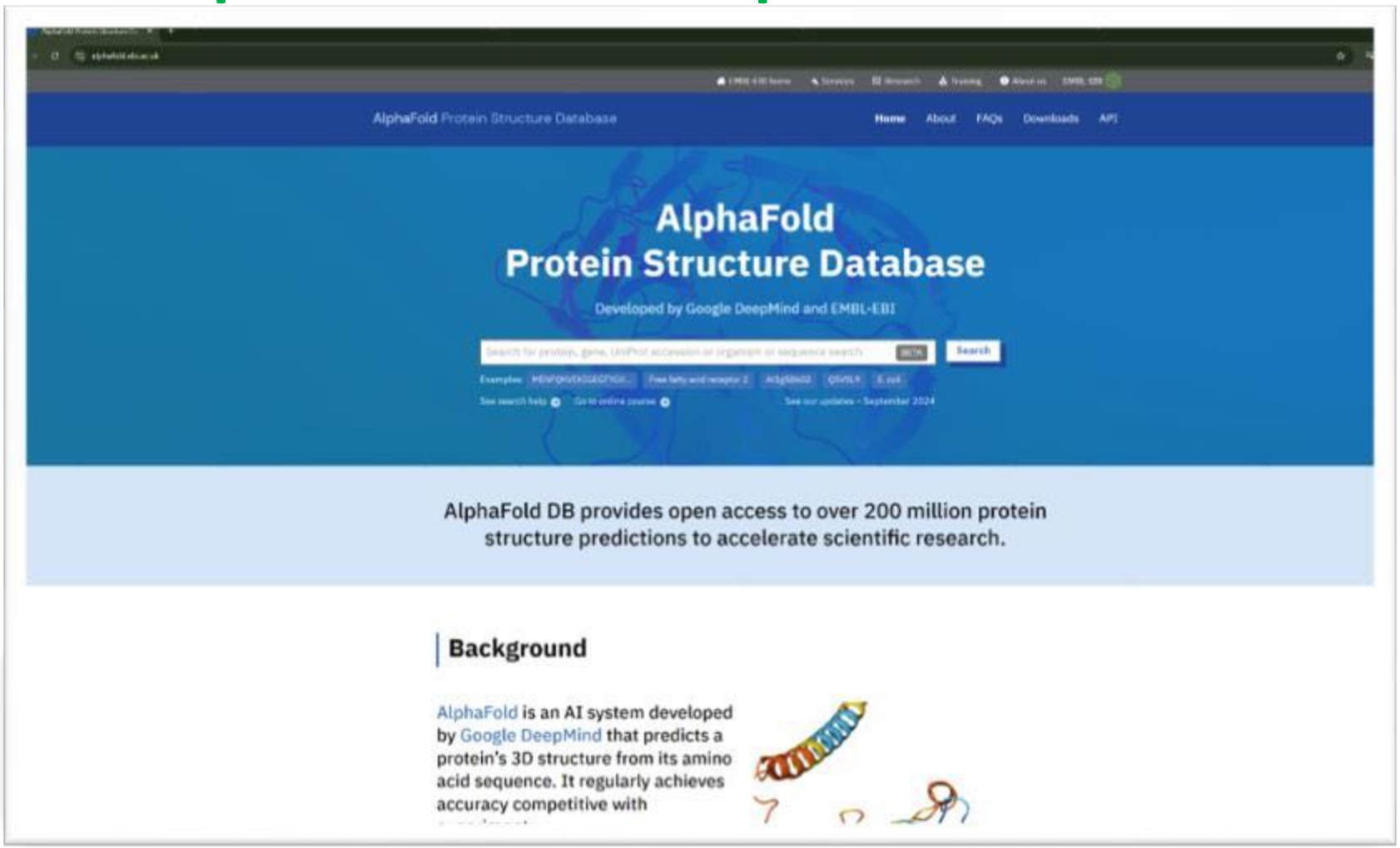


La sede di DeepMind a Londra

UNIVERSITÀ DEGLI STUDI FIRENZE
Da un secolo, oltre.

Le predizioni di AlphaFold sono pubblicamente disponibili

HR EXCELLENCE IN RESEARCH



Varadi, M et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Research (2024).

UNIVERSITÀ DEGLI STUDI FIRENZE

Da un secolo, oltre.

THE NOBEL PRIZE

HR EXCELLENCE IN RESEARCH

The 2024 chemistry laureates

David Baker is awarded *"for computational protein design"* and Demis Hassabis and John Jumper *"for protein structure prediction"*.

David Baker
Born: 1962, USA

Demis Hassabis
Born: 1976, UK

John Jumper
Born: 1985, USA

ILL. NIKLAS ELMEHED © NOBEL PRIZE OUTREACH

In parallelo…

David Baker oltre ad aver lavorato intensamente sul problema della predizione strutturale (il suo algoritmo Rosetta è stato fra i vincitori più frequenti prima di AlphaFold), ha cercato di sfruttare quello che imparava da queste applicazione per poter disegnare nuove proteine, mai esplorate dalla selezione naturale, che avessero forme e/o funzioni scelte *a priori*

→ **PROTEIN DESIGN**

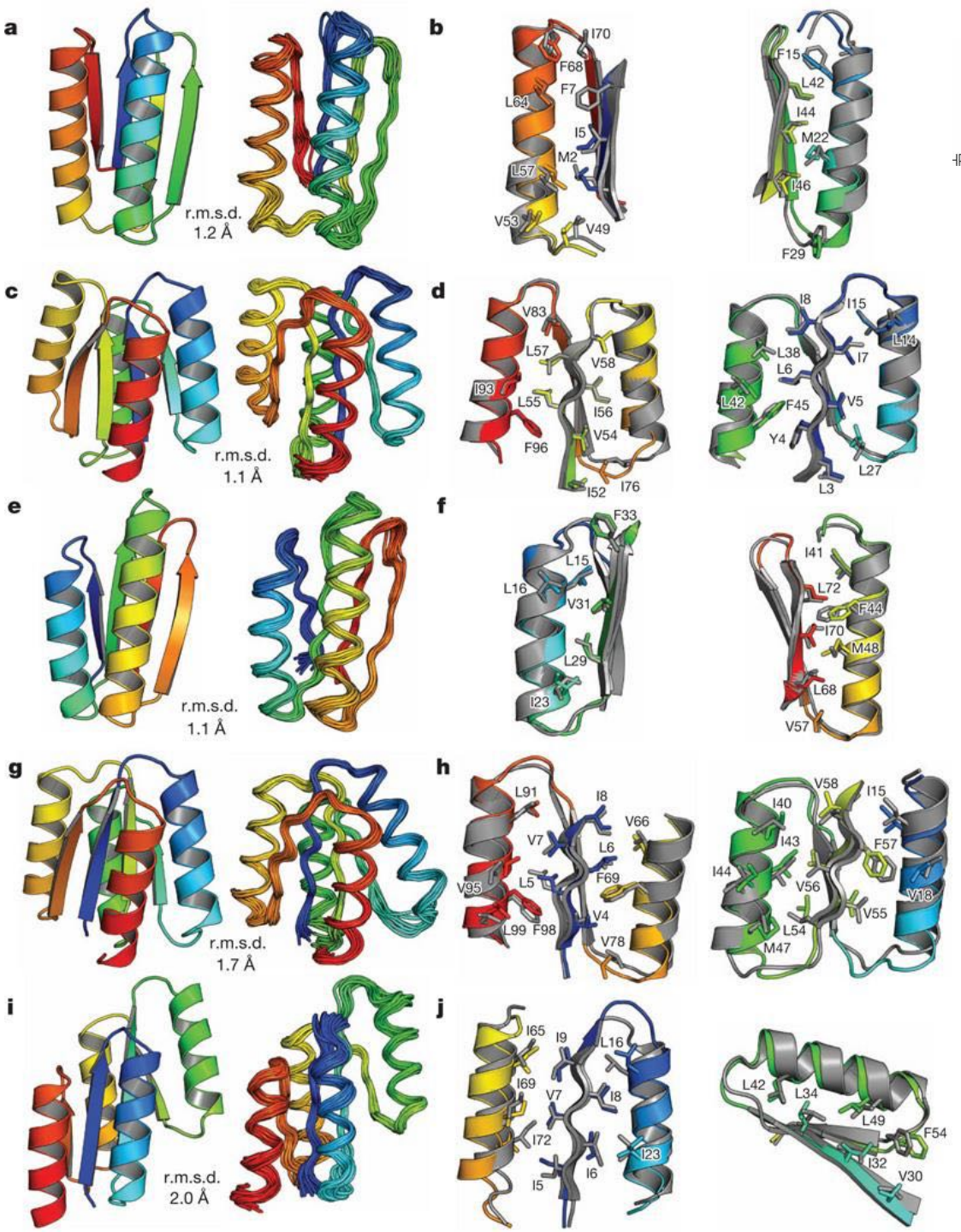
Per esempio, stimando l'energia nel modo che usa AlphaFold (o simile) si può prevedere se la nuova proteina disegnata sarà **stabile** quando prodotta realmente e quindi **se sarà possibile usarla in laboratorio**

UNIVERSITÀ DEGLI STUDI FIRENZE

Da un secolo, oltre.

HR EXCELLENCE IN RESEARCH

Proteine disegnate con topologie complesse sono state prodotte e la loro struttura determinata sperimentalmente.
Il disegno riesce a definire anche i principali contatti fra catene laterali.



Nature 491:222–227 (2012)

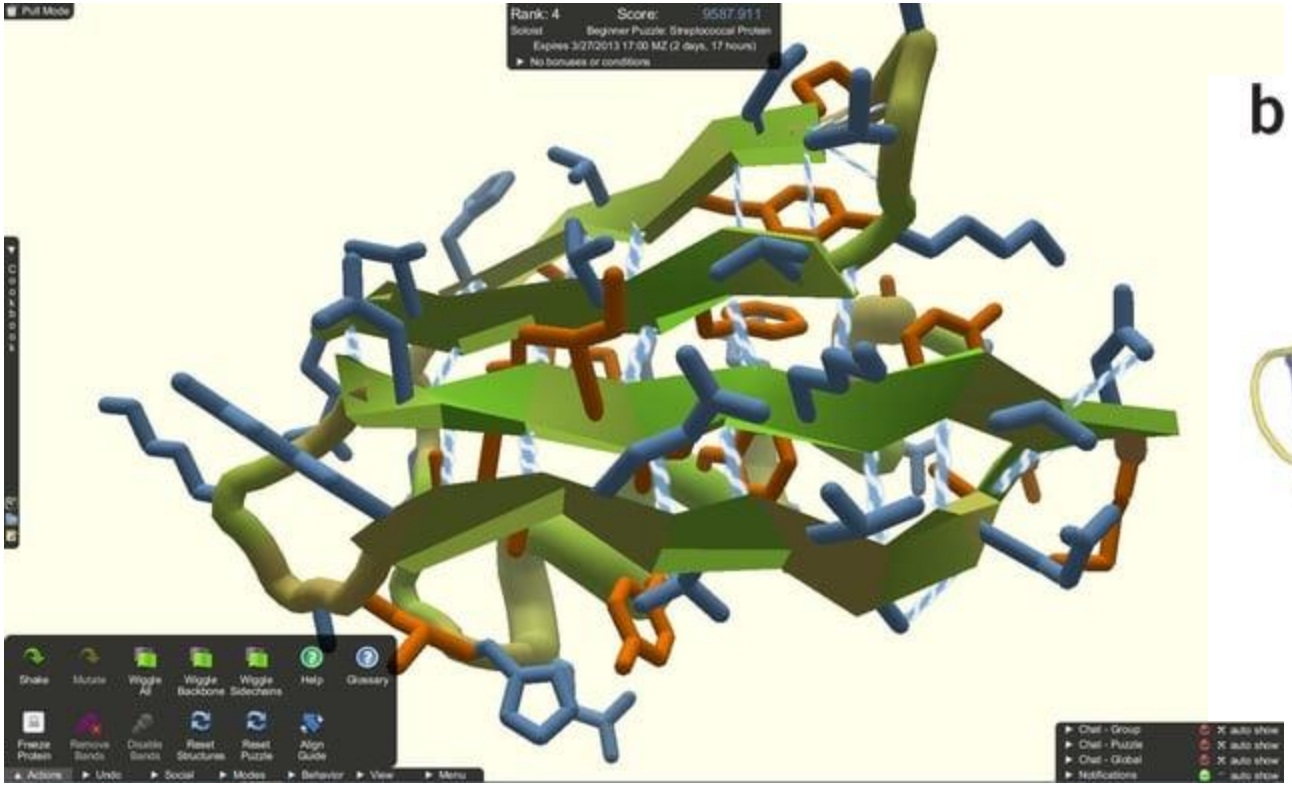
. 2011

UNIVERSITÀ DEGLI STUDI FIRENZE

Da un secolo, oltre.

Foldit, un gioco online per la predizione strutturale e il protein design

HR EXCELLENCE IN RESEARCH



High-resolution structure of a retroviral protease folded as a monomer

Nature Struct Mol Biol 2011

Acta Crystallogr D Biol Crystallogr. 2011

Struttura predetta automaticamente

Struttura predetta dai giocatori

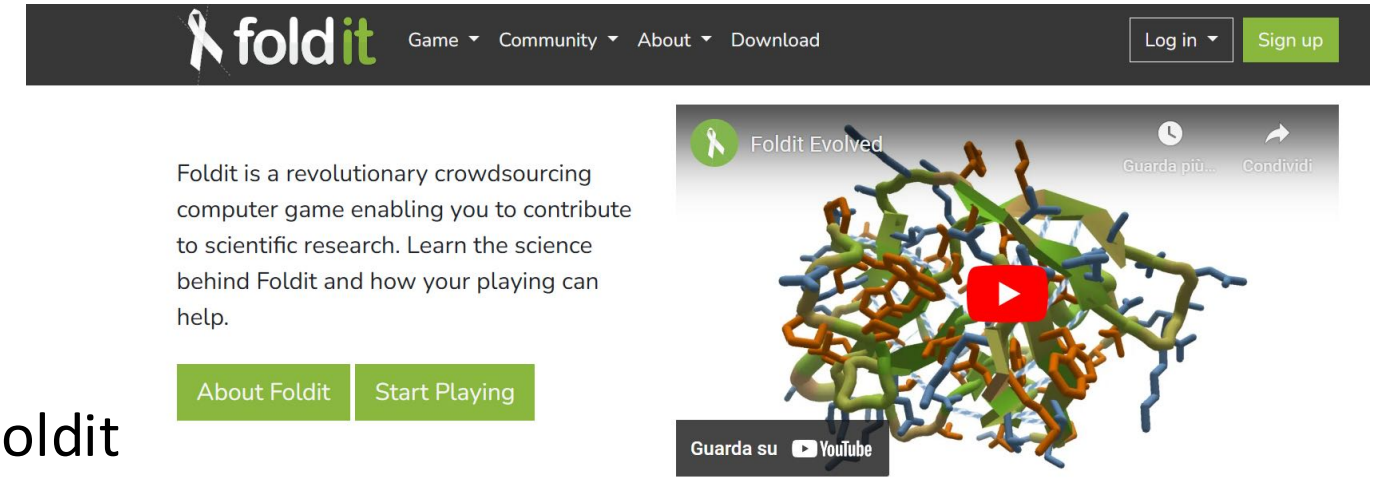
Struttura sperimentale, risolta successivamente

UNIVERSITÀ DEGLI STUDI FIRENZE

Da un secolo, oltre.

Fra le alter cose, Foldit è usato anche per il disegno di nuovi inibitori/candidati farmaci

https://fold.it/

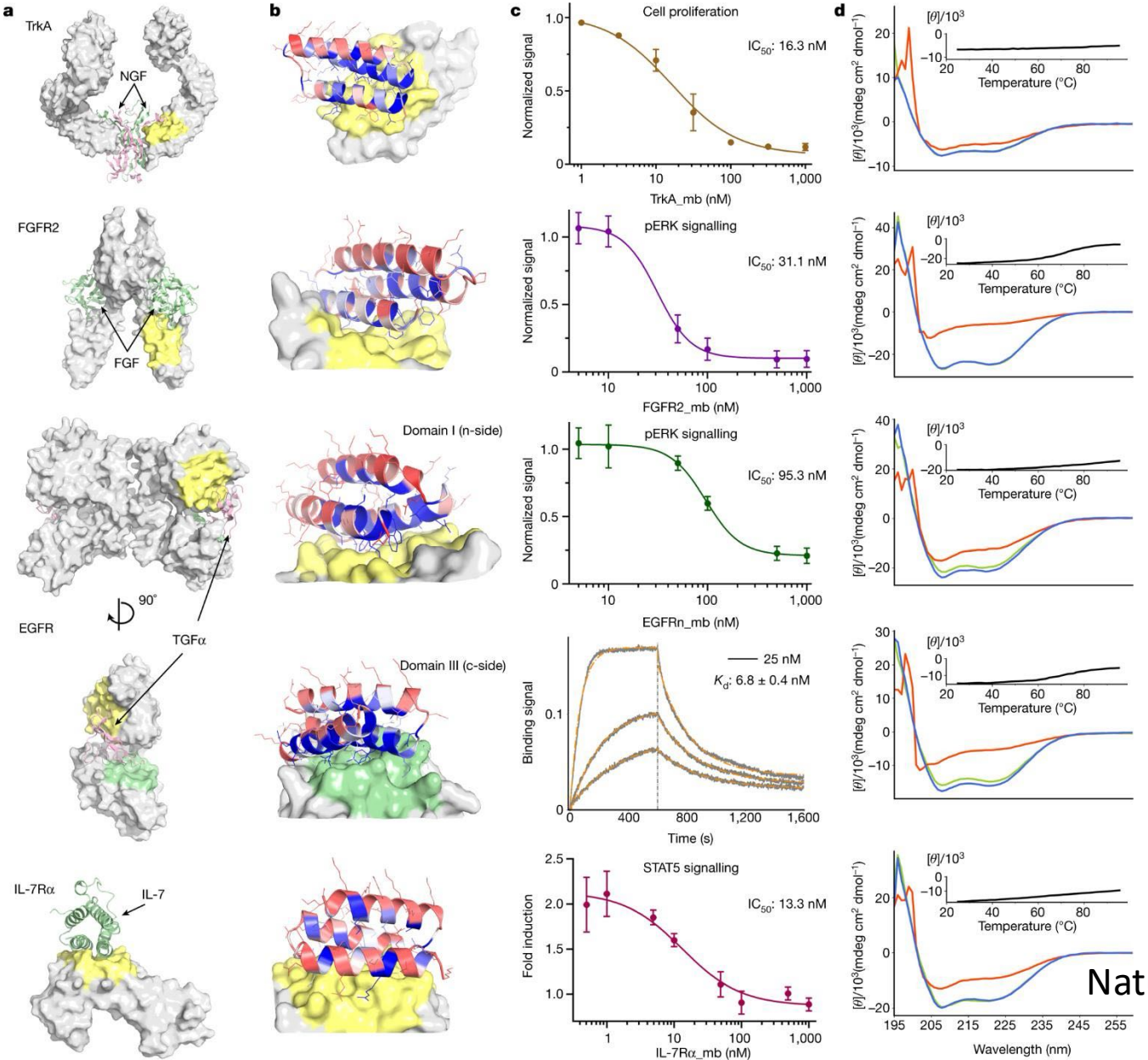


UNIVERSITÀ DEGLI STUDI FIRENZE

Da un secolo, oltre.

HR EXCELLENCE IN RESEARCH

Dati su colture cellulari



Proteine disegnate per interagire con specifici recettori (minibinders, mb) inducono effetti misurabili *in vivo*

Nature 605:551–560 (2022)

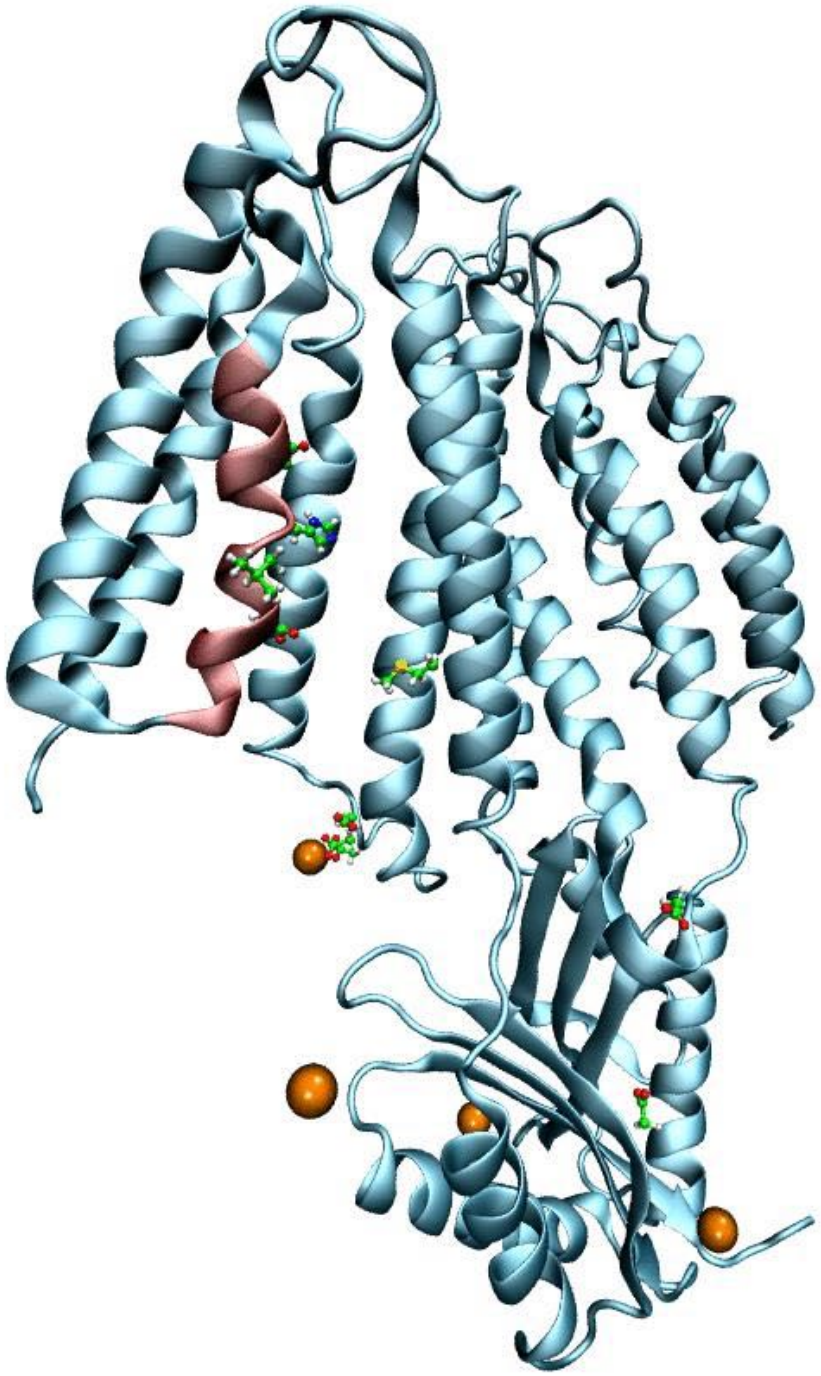
In conclusione

- La struttura 3D di proteine semplici è ora prevedibile computazionalmente con ottima affidabilità – **il paradosso di Levinthal è risolto**
- **Cosa resta da studiare**: interazioni fra macromolecule (proteine, acidi nucleici) e con piccoli ligandi (inibitori, cofattori, …); proteine multi-dominio e proteine altamente flessibili
- Il disegno di proteine permette di **generare forme e architetture desiderate**, con un buon controllo anche sulle proprietà biologiche (ma la necessità di verifica sperimentale resta!)
- **Cosa resta da studiare**: nuovi catalizzatori enzimatici per processi biochimici mirati (biotecnologie e chimica verde); sistemi reattivi agli stimoli

UNIVERSITÀ
DEGLI STUDI
FIRENZE
Da un secolo, oltre.

HR EXCELLENCE IN RESEARCH

Grazie per l'attenzione!



HR EXCELLENCE IN RESEARCH

We used MetalPDB stats MFSs to derive parameters for a **non-bonded** MD force field

Macchiagodena M, Pagliai M, Andreini C, Rosato A, Procacci P.
J. Chem. Inf. Model. 2019; ACS Omega 2020

An atomistic view of structural changes in bacterial and human YiiP upon zinc(II) binding

D. Sala, A. Giachetti and A. Rosato
Biochim. Biophys. Acta - General Subjects (2019)
J. Chem. Inf. Model. 2021