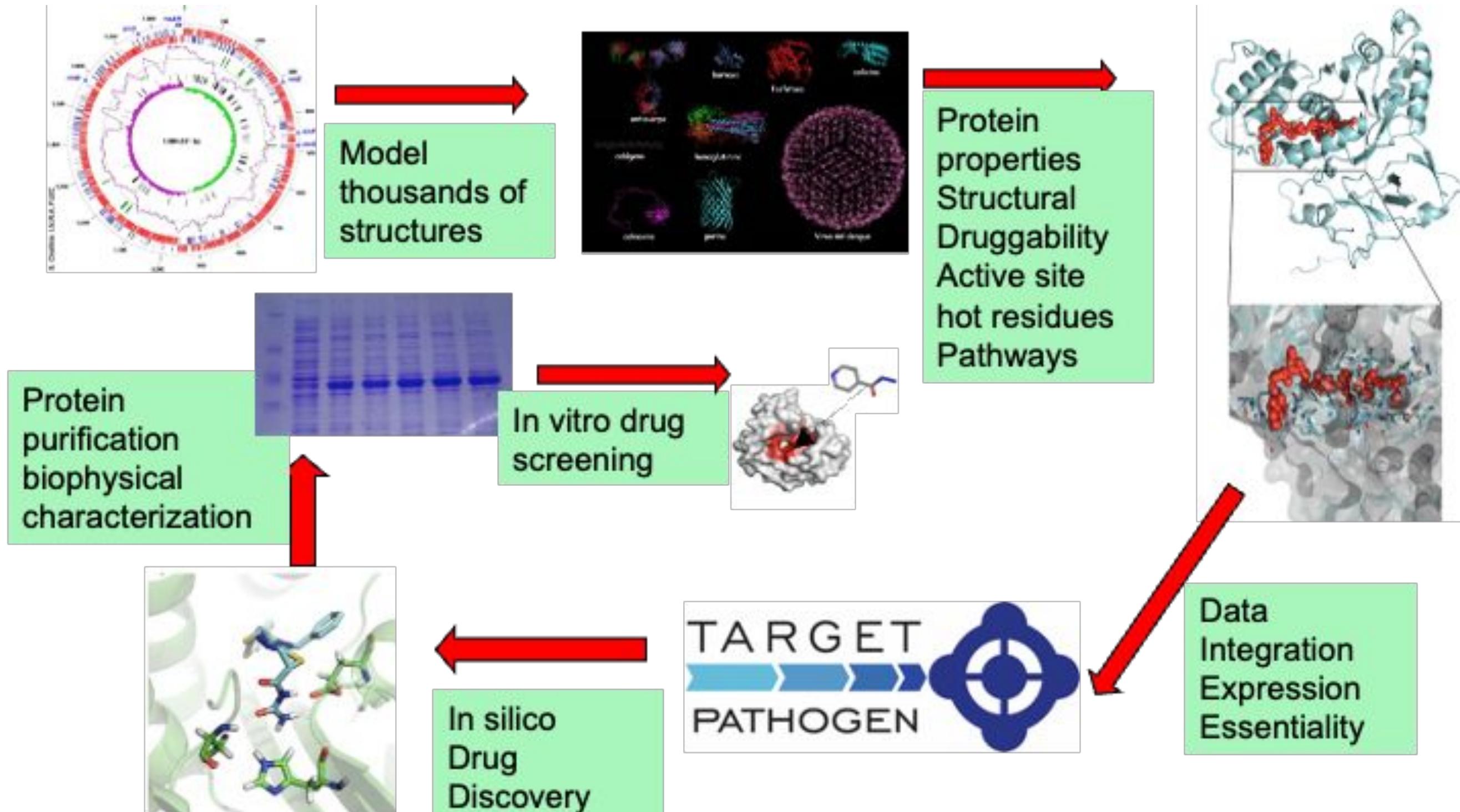


“Enfoques estructurales para el análisis de la drogabilidad del proteoma”

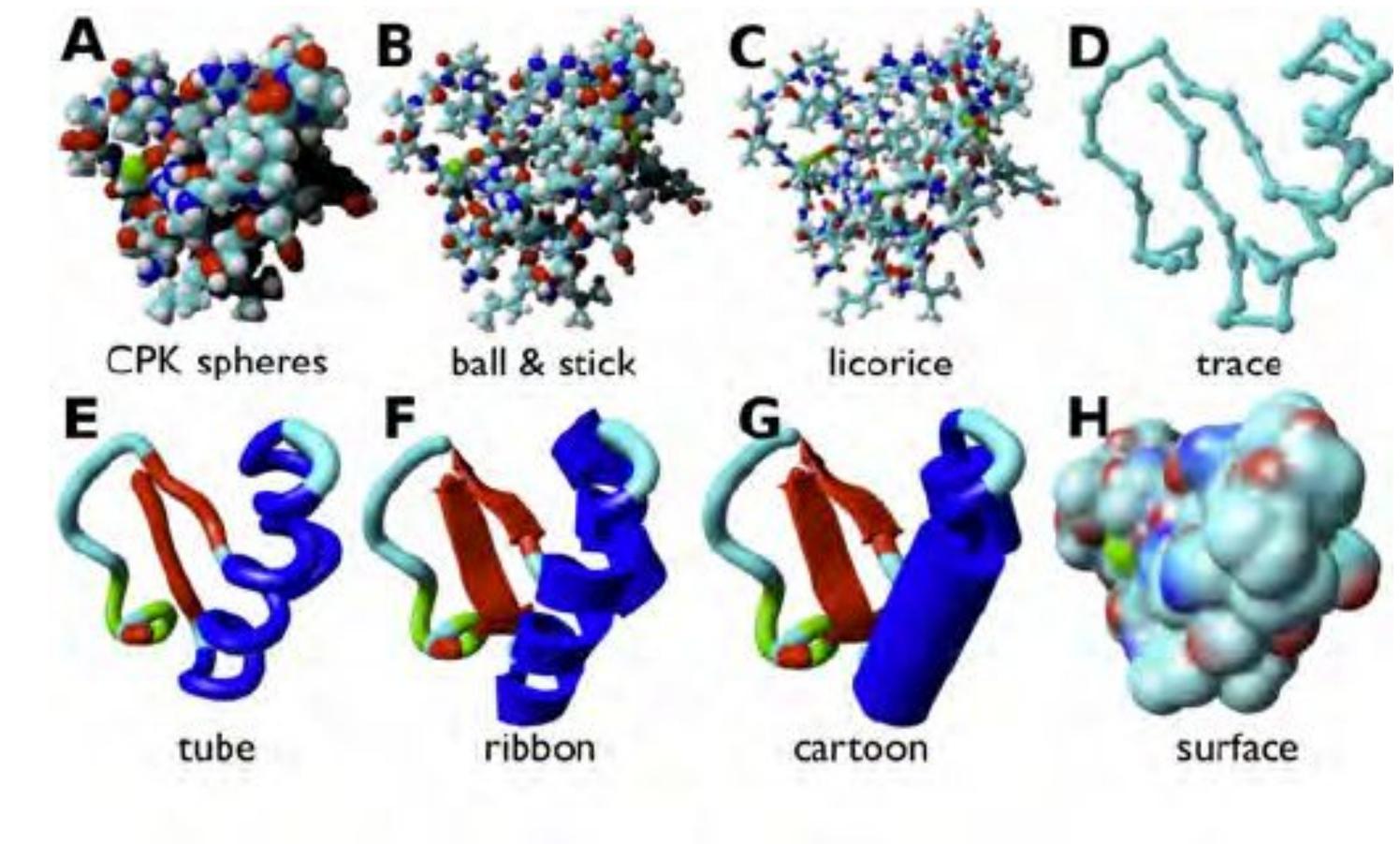
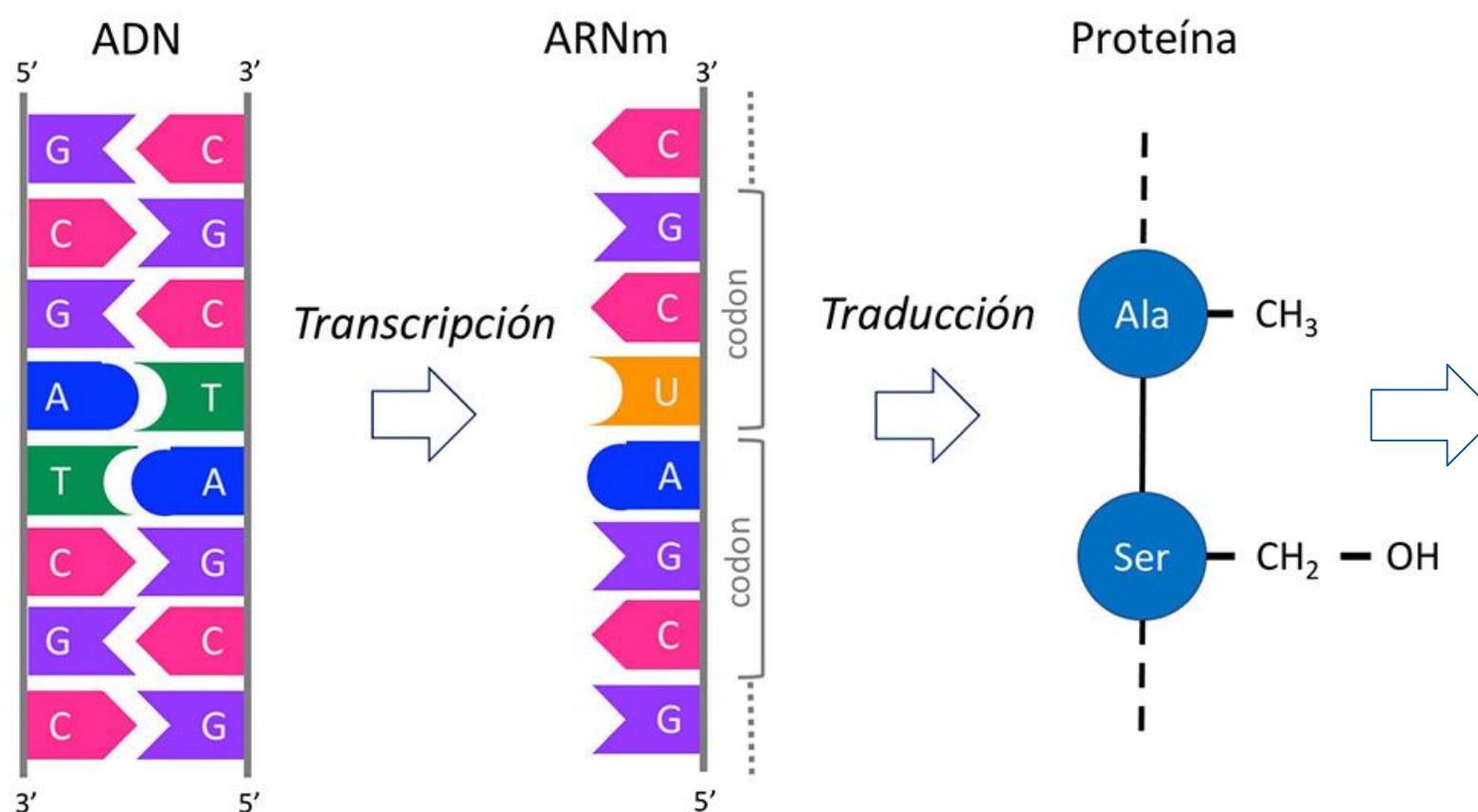


Adrian Turjanski
Structural Bioinformatics Lab
Departamento de Química Biológica
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

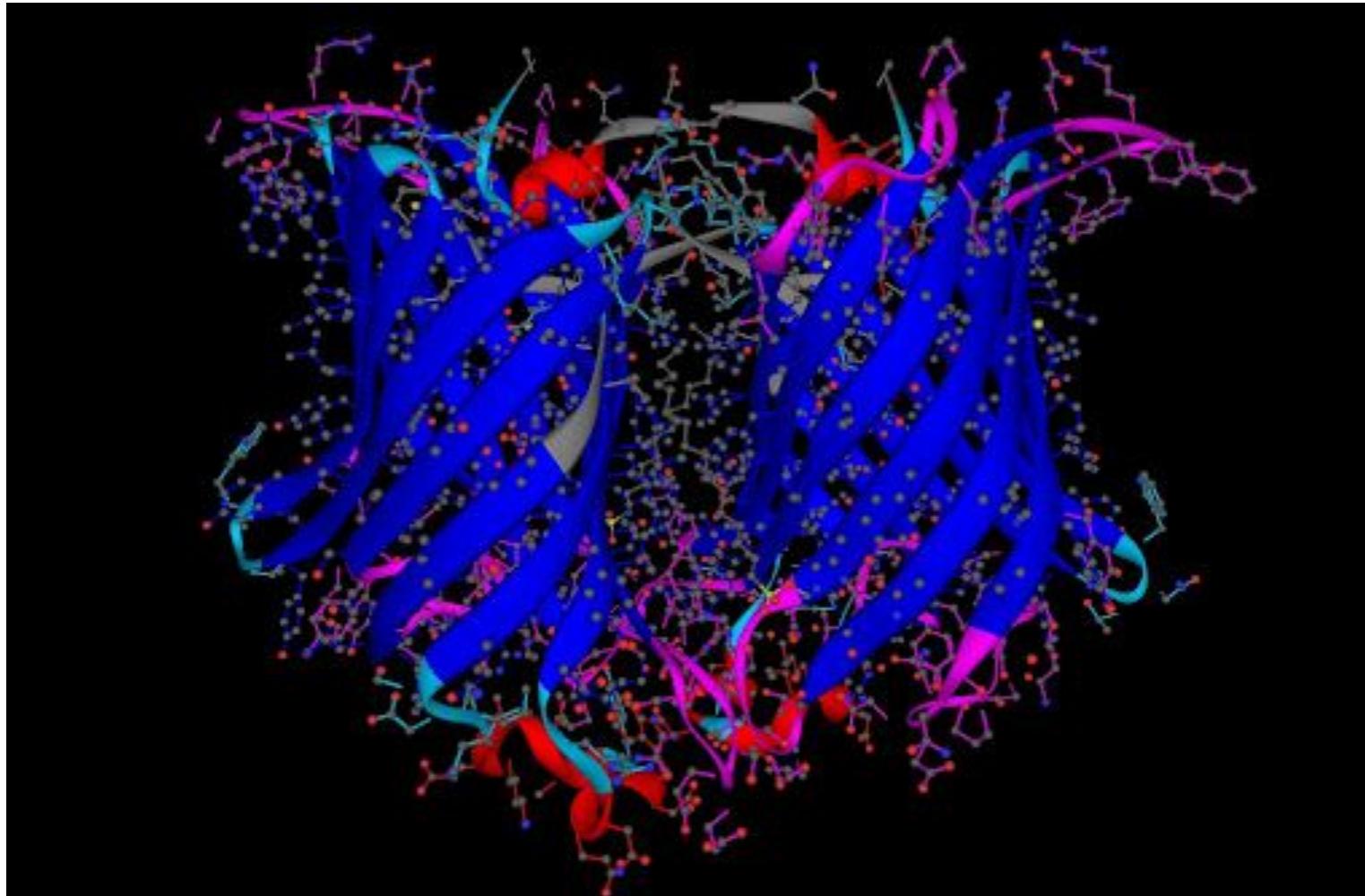
Del genoma a la droga



La estructura de macromoléculas



Para qué una estructura?



- Enzimas funcionales en biotecnología
- Vacunas
- Medicamentos biológicos
- Desarrollo de Drogas
- Diagnóstico: Estudios de variaciones genéticas
- Nanotecnología: Sensores, Delivery, etc...

Estructura de proteínas: Ramachandran

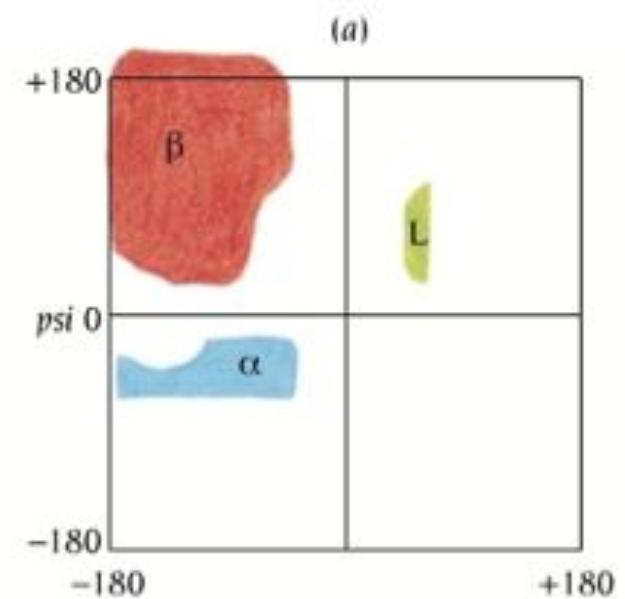
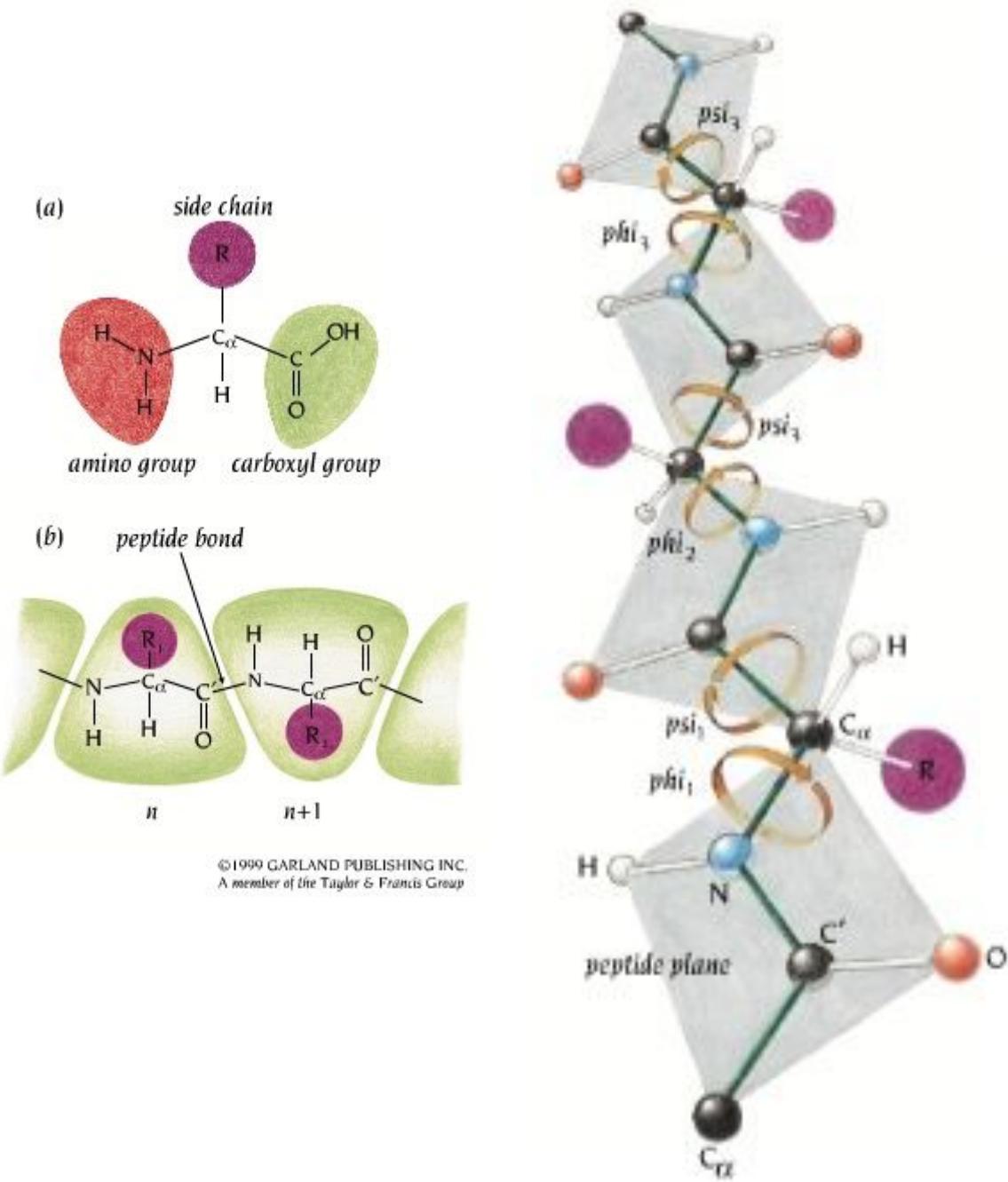
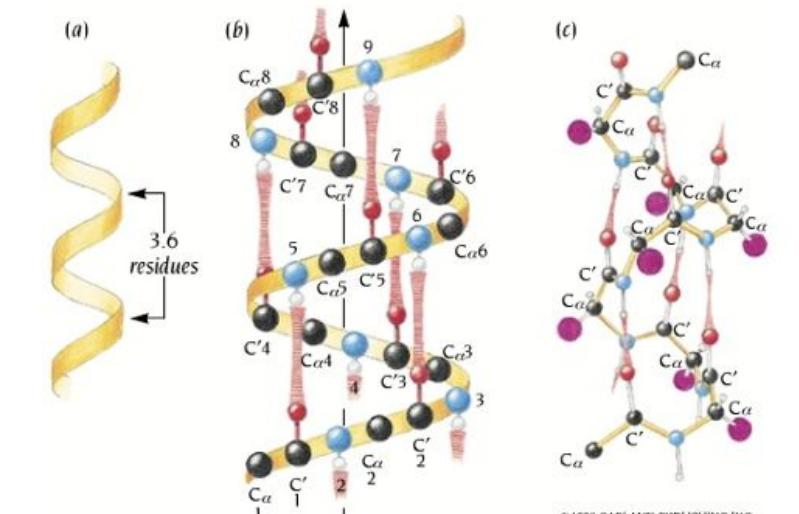
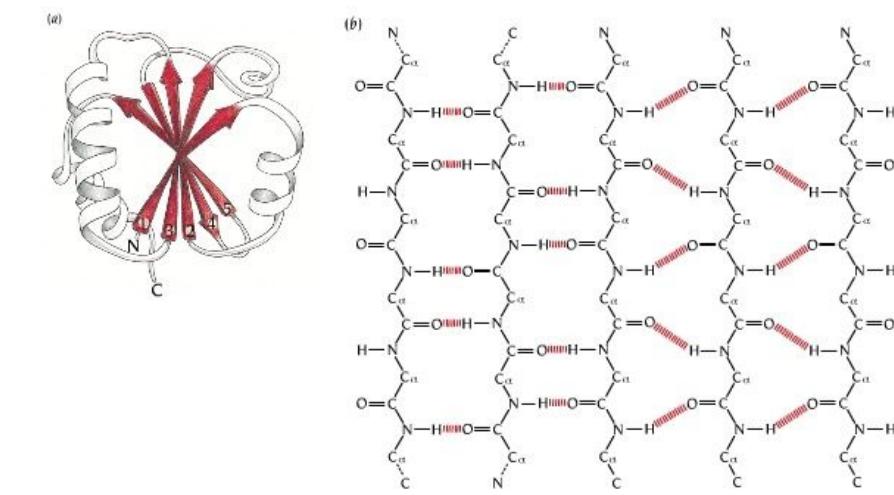


Gráfico de Ramachandran



Hélices

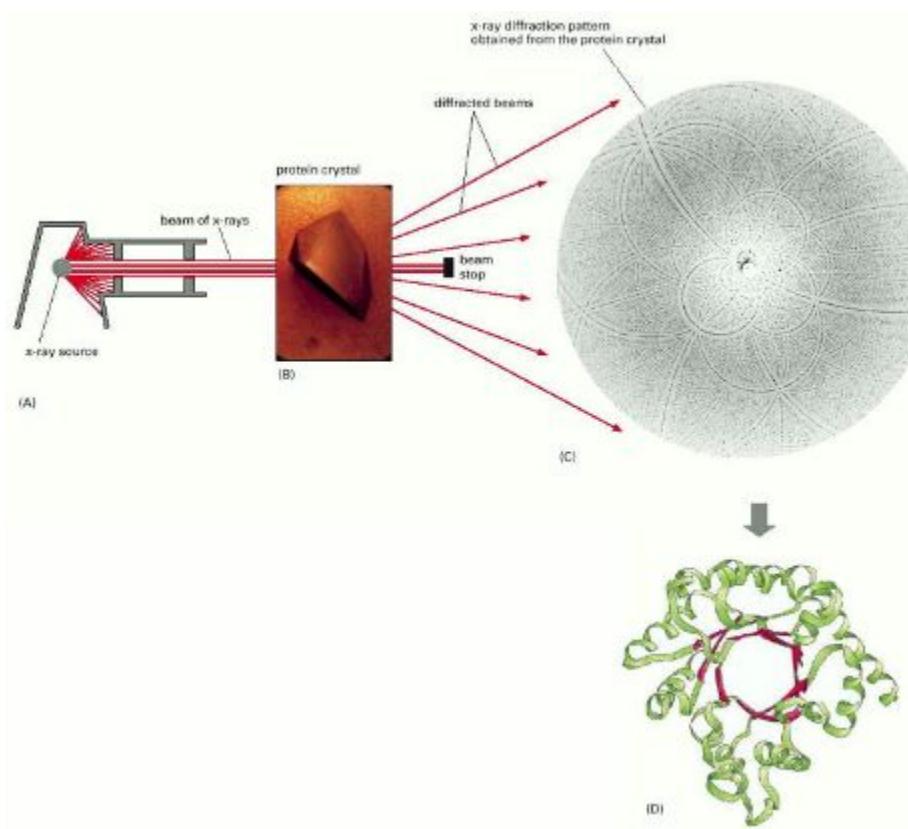


Hojas Beta



Loops

Revolución en las técnicas experimentales

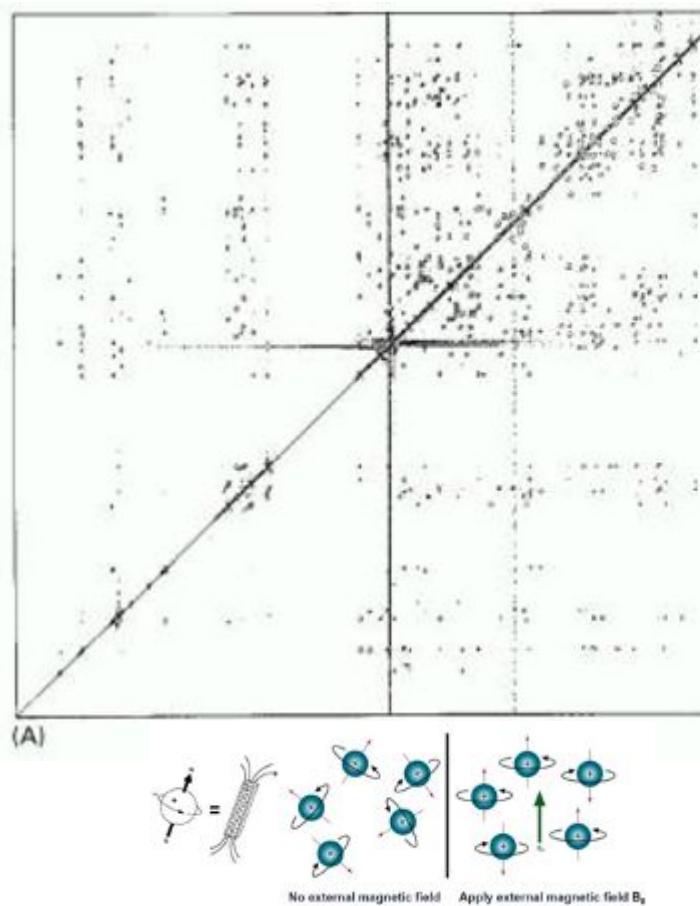


-Resolución Atómica

-En un cristal

-Difícil obtener cristales

-Limites en tamaño en general miles de aminoácidos

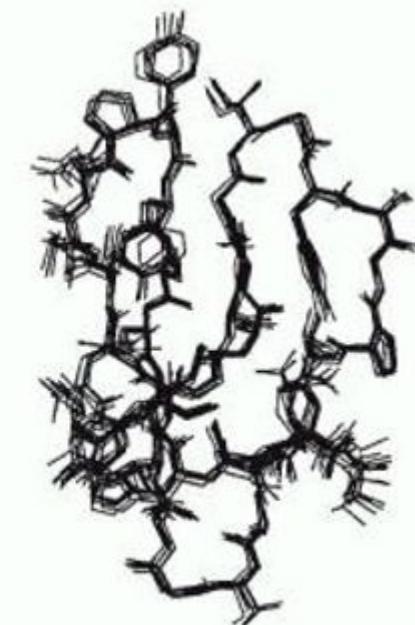


-Resolución Atómica

-En solución

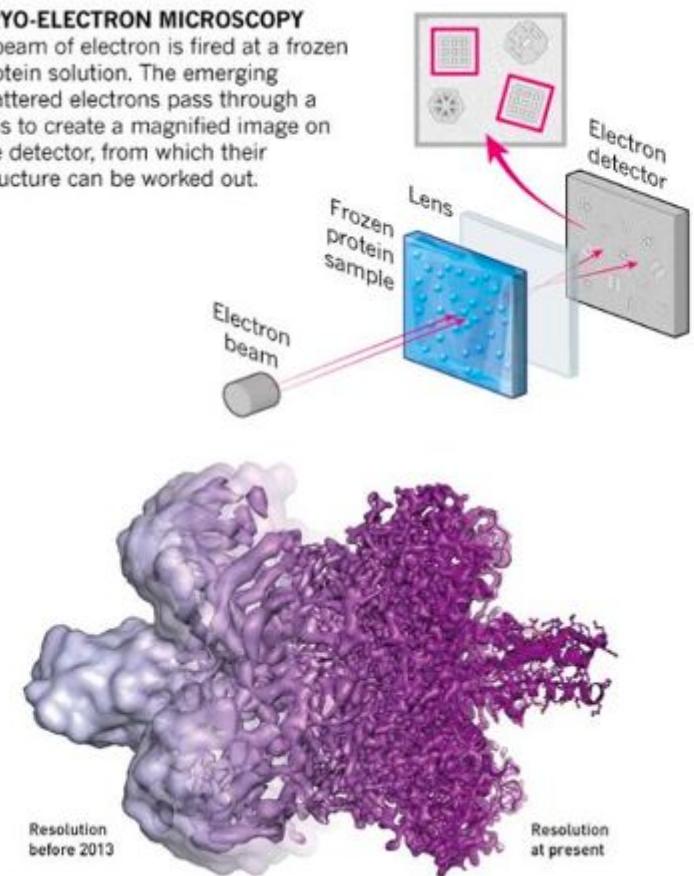
-Difícil obtener grandes cantidades

-Limites en tamaño en general cientos de aminoácidos (<300)



(B)

CRYO-ELECTRON MICROSCOPY
A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.



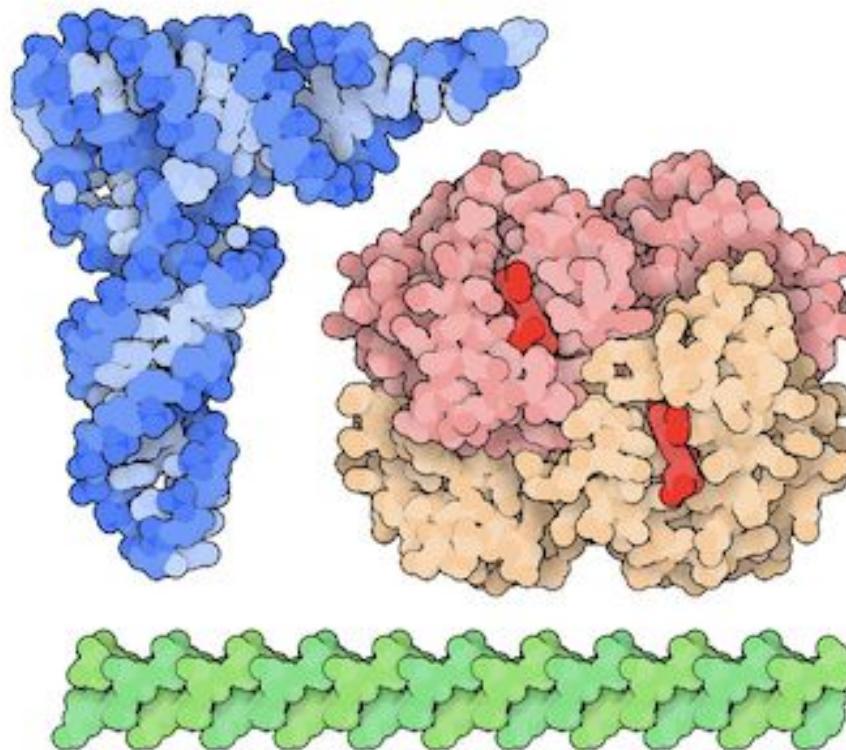
-Resolución Atómica

-En solución

-Limites en tamaño a más grandes, miles de aminoácidos

-Fácil de preparar

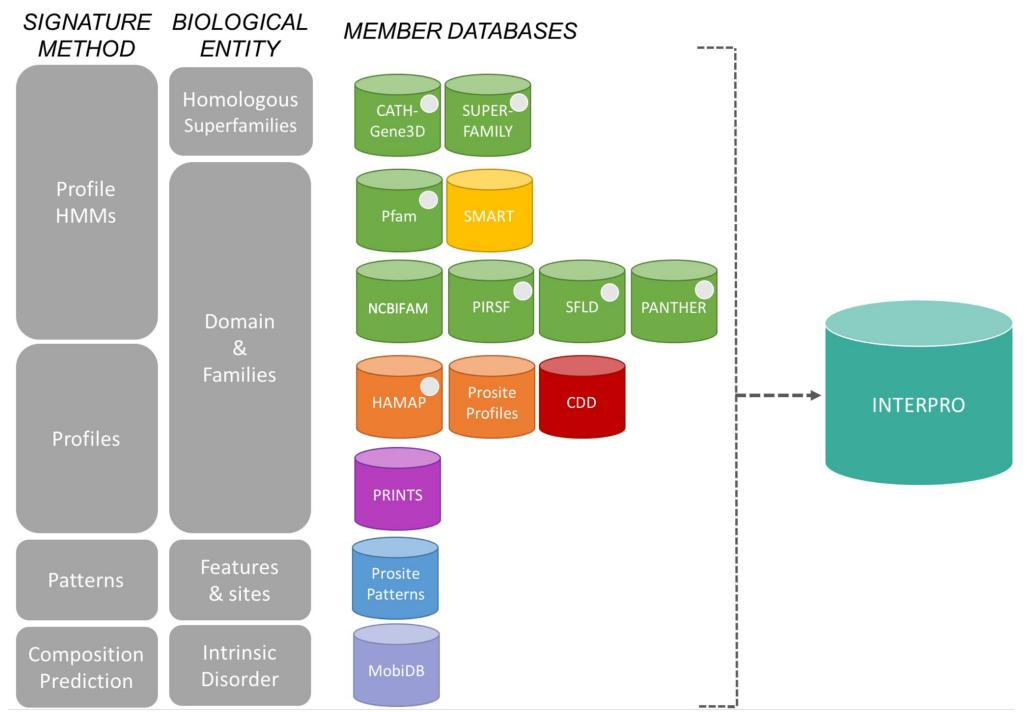
BASES DE DATOS BIOLÓGICAS: PDB



Molecular Type	X-ray	NMR	EM	Multiple methods	Neutron	Other	Total
Protein (only)	146871	11954	7471	186	72	32	166586
Protein/Oligosaccharide	8676	31	1306	5	0	0	10018
Protein/NA	7750	277	2369	3	0	0	10399
Nucleic acid (only)	2445	1408	62	11	2	1	3929
Other	154	31	5	0	0	0	190
Oligosaccharide (only)	11	6	0	1	0	4	22
Total	165907	13707	11213	206	74	37	191144

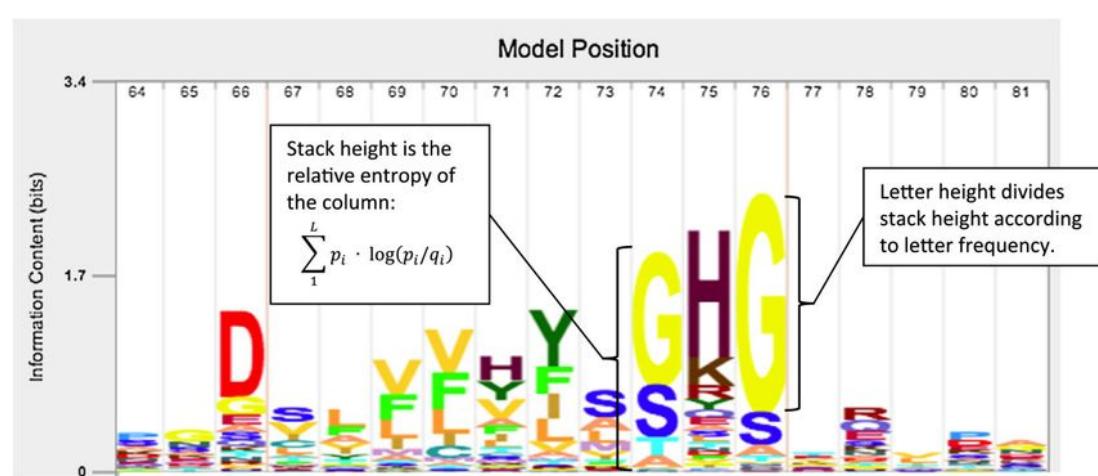
Base de datos curada, internacional, integrada

BASES DE DATOS BIOLÓGICAS: Interpro



D Kringle 7k 605 702 114 Kringles are autonomous structural domains, found throughout the blood clotting and fibrinolytic domain IPR000001	F Retinoid X receptor/HNF4 3k 25 640 112 Steroid or nuclear hormone receptors (NRs) constitute an important superfamily of transcription regulators that family IPR000003	F Metallothionein, vertebrate 1k 3 281 15 Metallothioneins (MT) are small proteins that bind heavy metals, such as zinc, copper, cadmium, nickel, family IPR000006	D Tubby, C-terminal 6k 91 916 5 Tubby, an autosomal recessive mutation, mapping to mouse chromosome 7, was domain IPR000007
D C2 domain 110k 3k 2k 162 The C2 domain is a Ca ²⁺ -dependent membrane-targeting module found in many cellular proteins domain IPR000008	F Protein phosphatase 2A regulatory subunit PR55 6k 76 2k 1 Protein phosphatase 2A (PP2A) is a serine/threonine phosphatase implicated in many cellular processes, family IPR000009	D Cystatin domain 9k 132 1k 59 Cystatins are a family of cysteine protease inhibitors belonging to MEROPS inhibitor family I25, clan IH domain IPR000010	F Ubiquitin/SUMO-activating enzyme E1 5k 154 2k 24 This entry also contains the SUMO activating enzyme E1. family IPR000011

Integra bases de datos de secuencia y permite clasificar dominios

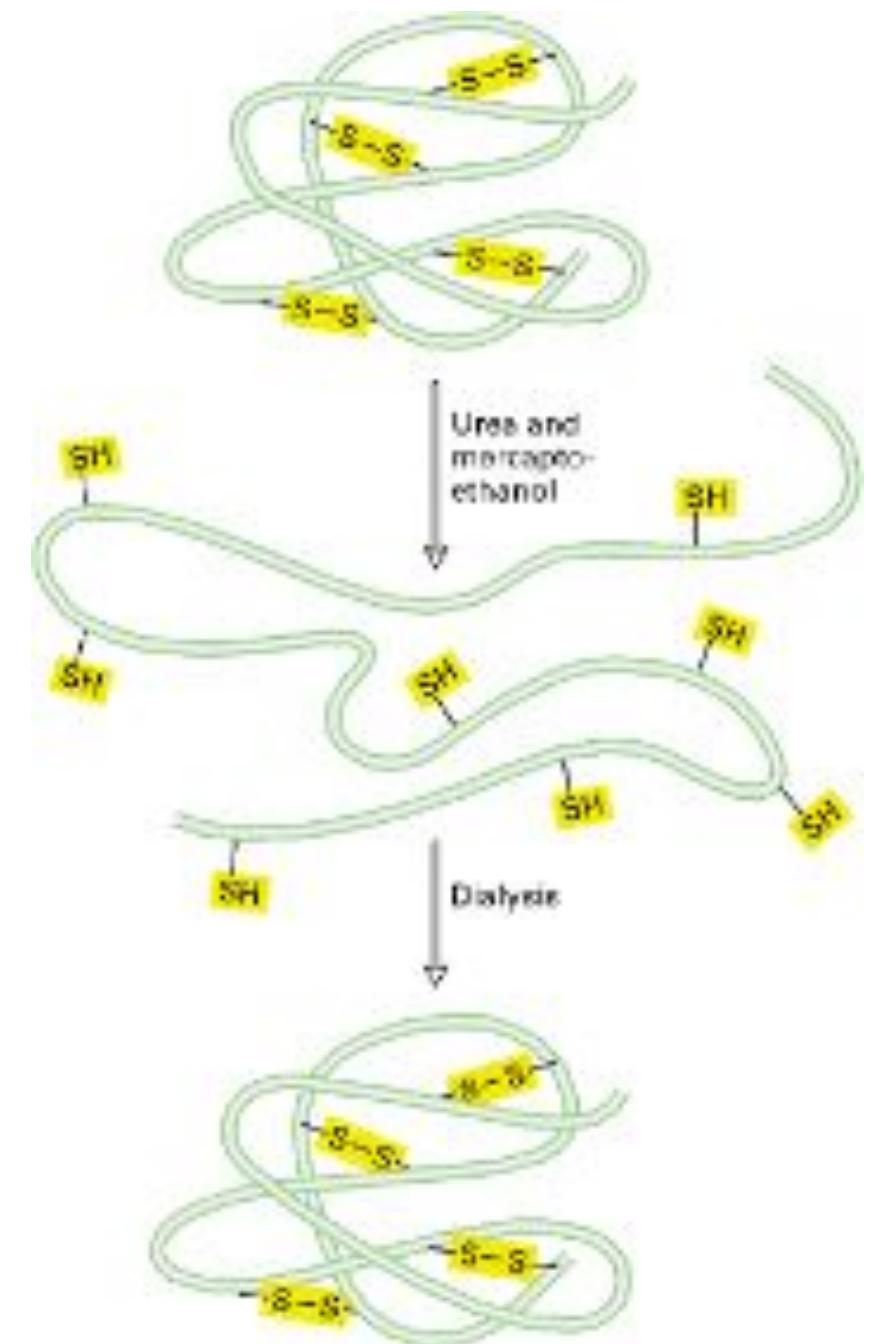


Pregunta básica: Se puede predecir la estructura a partir de la secuencia?

Toda la información de la estructura nativa de la proteína esta codificada en la secuencia + la solución o entorno en el que se encuentra.

Anfinsen wanted to show that the information for protein folding resided entirely within the amino acid sequence of the protein. He chose ribonuclease A as his model for folding but he couldn't completely denature the protein unless he treated it with the denaturant urea **plus** 2ME to break the disulfide bridges.

Under those conditions, the protein unfolded. It would refold spontaneously once he removed urea and 2ME from the folding solution. Ribonuclease A regained biological activity under those conditions. This demonstrated that refolding could take place *in vitro*.



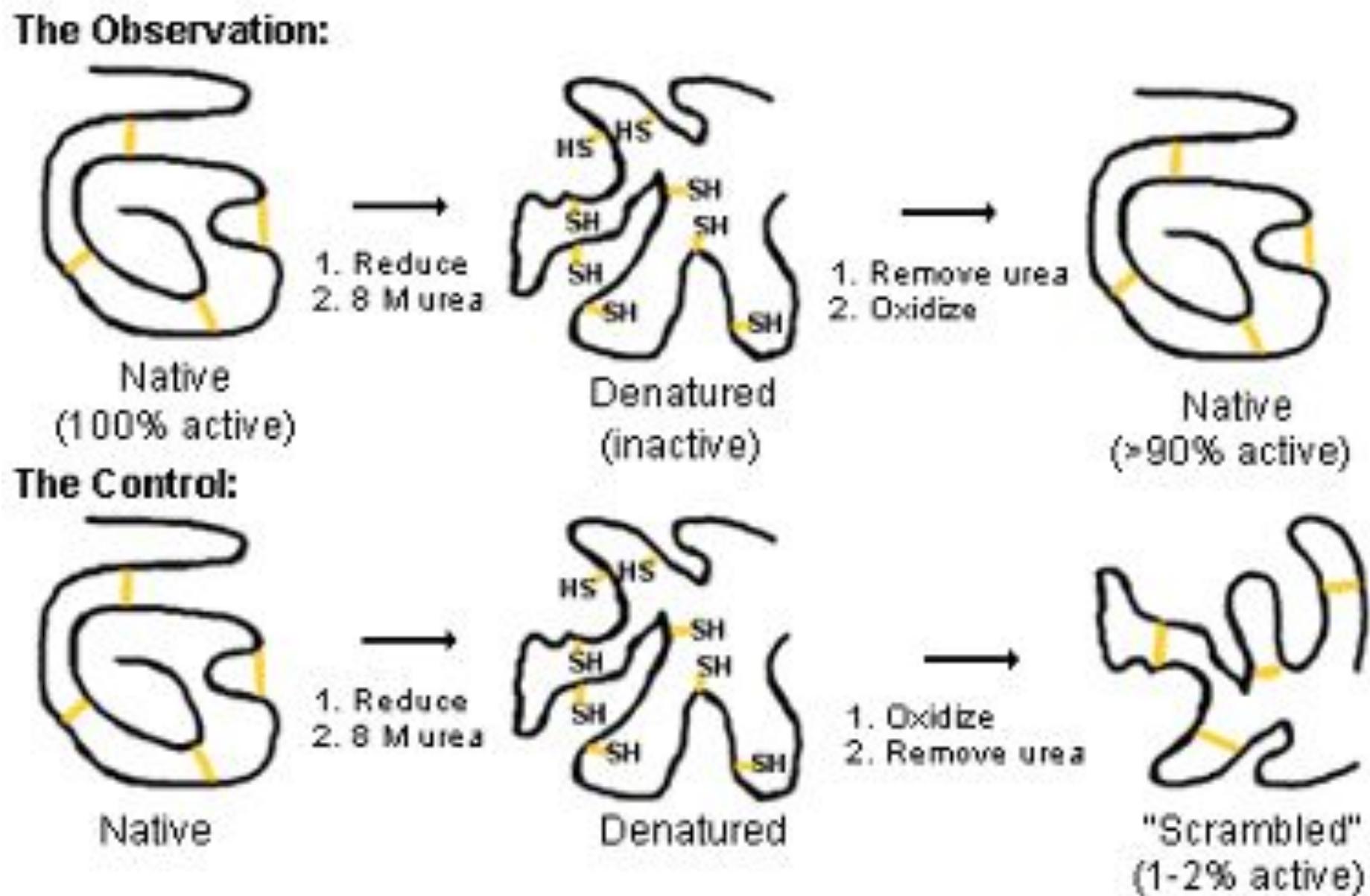
Anfinsen, 1973

Pregunta básica: Se puede predecir la estructura a partir de la secuencia?

Control

Anfinsen discovered that removing 2ME but not urea led to recovery of 1% of the activity. This is attributed to the formation of random disulfide bridges between the 8 cysteines present in the protein. There are 105 different possibilities ($7 \times 5 \times 3 \times 1$) so the 1% recovery makes sense.

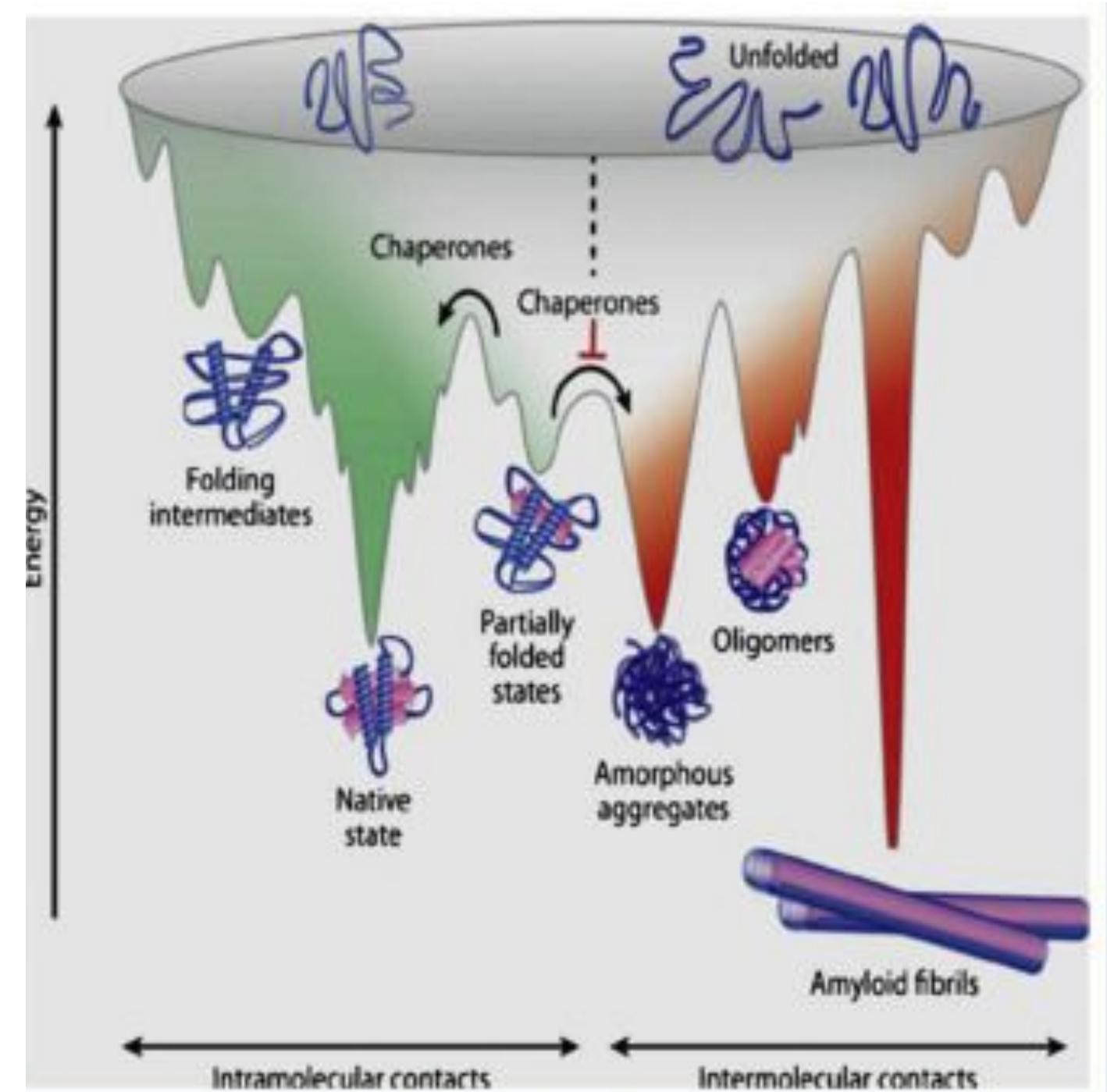
It also shows that the correct three-dimensional conformation must be achieved fairly rapidly when urea is removed since most of the protein under those conditions becomes active.



Paradoja de Levinthal

1969 Cyrus Levinthal :

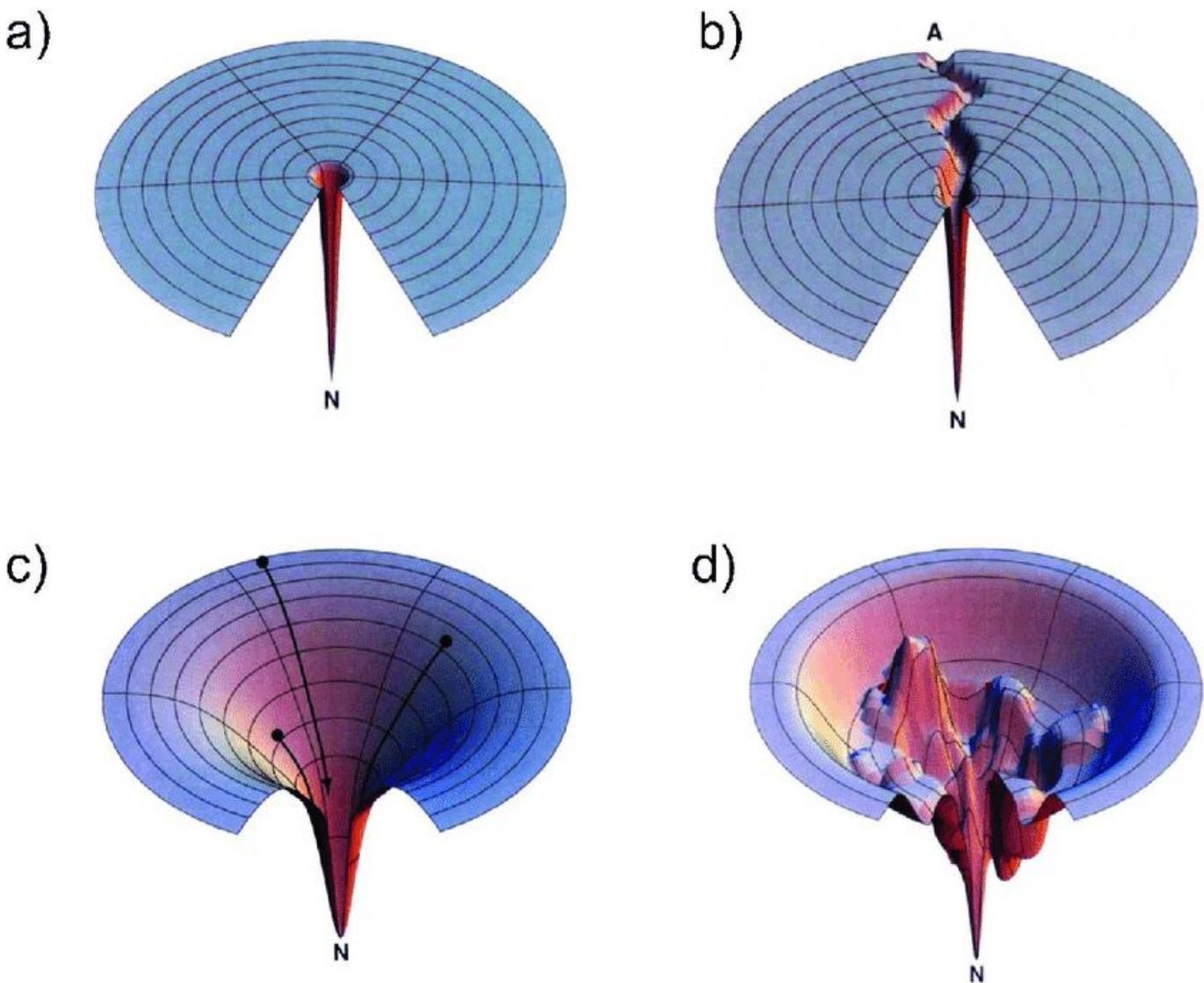
- Encontrar el Estado nativo visitando conformaciones al azar NO ES POSIBLE.
- Cadena polipeptídica corta tiene 10^{143} conformaciones!!
 - Conformaciones son probadas en nanosegundos o picosegundos.
 - Tiempo mayor a la edad del universo!!!!



Paradoja de Levinthal

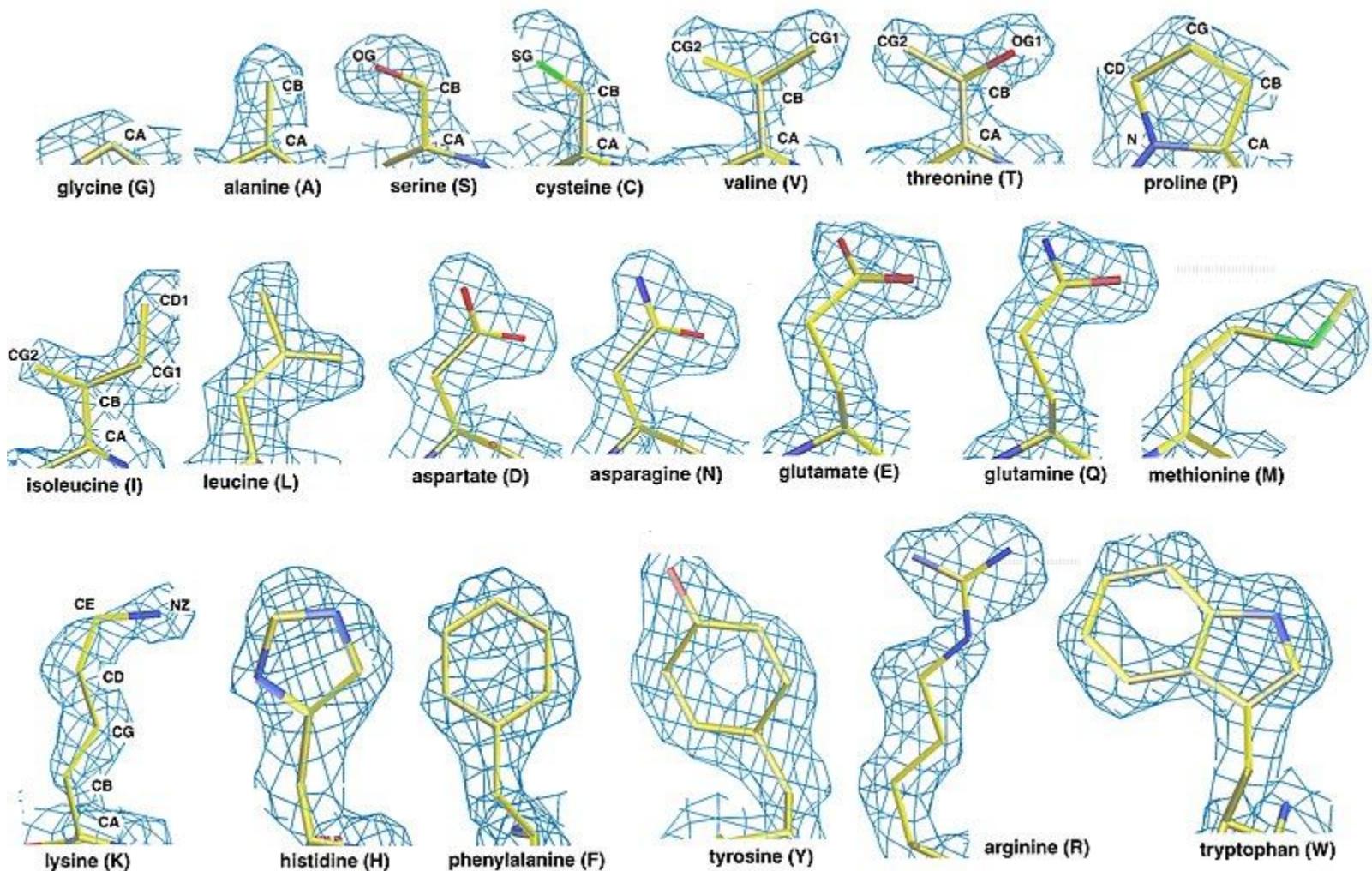
1969 Cyrus Levinthal :

- Encontrar el Estado nativo visitando conformaciones al azar NO ES POSIBLE.
- Cadena polipeptídica corta tiene 10^{143} conformaciones!!
 - Conformaciones son probadas en nanosegundos o picosegundos.
 - Tiempo mayor a la edad del universo!!!!



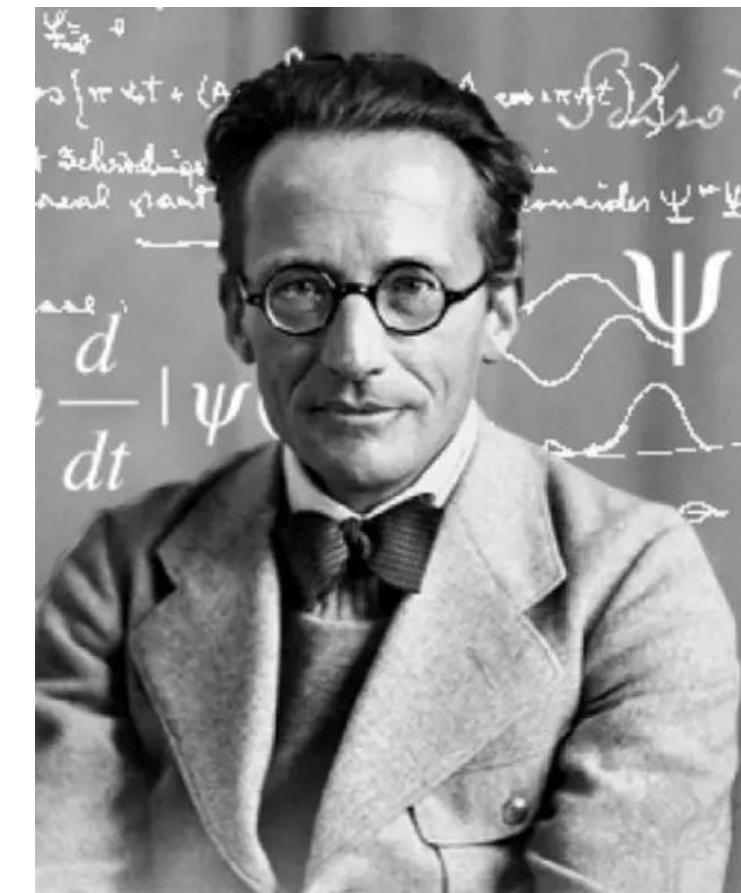
Hay algo en la evolución que determina el plegado de proteínas

Cómo estudiar el plegado?



Estructura Electrónica

Mecánica cuántica



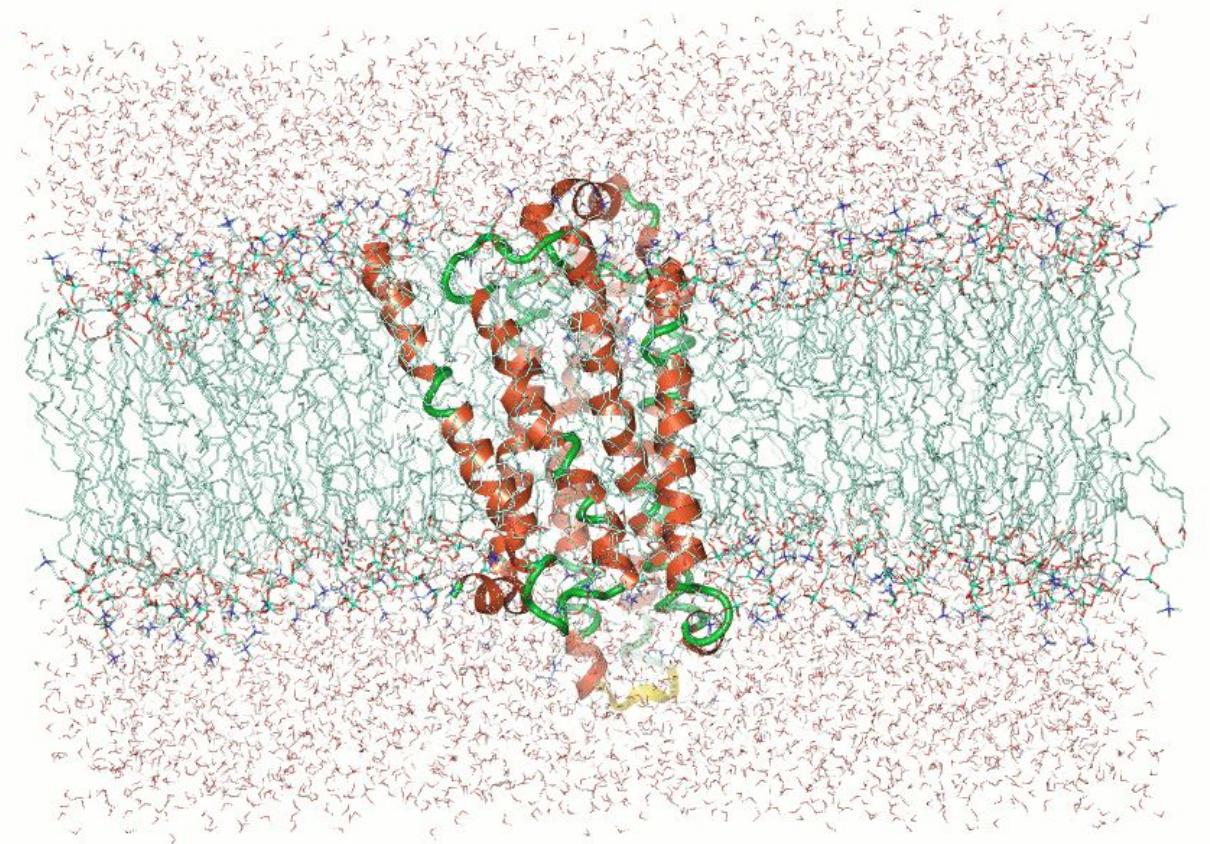
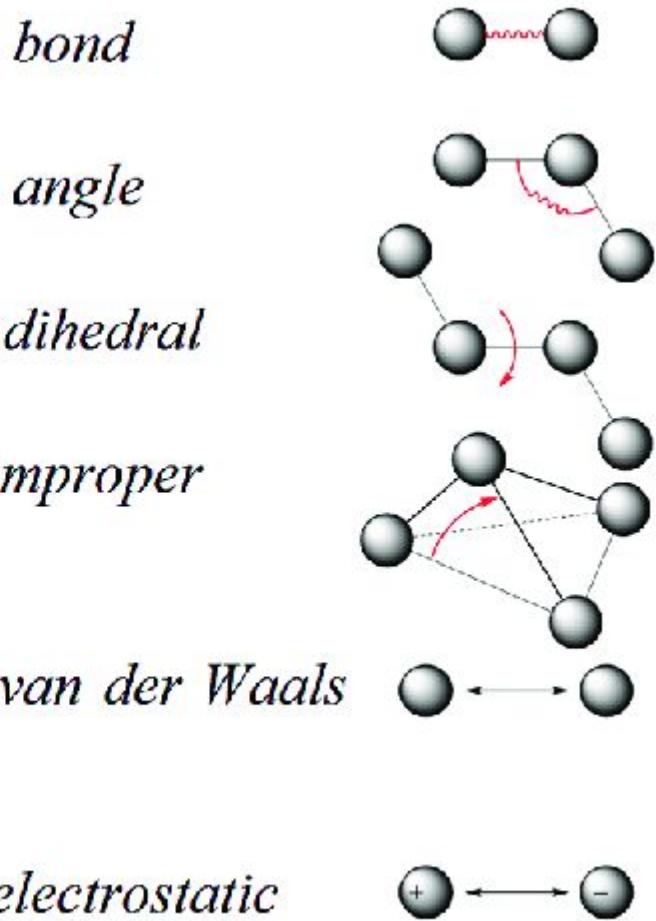
$$-\frac{\hbar}{i} \frac{\partial \Psi(r,t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi(r,t) + V(r,t) \Psi(r,t)$$

Schrodinger: Premio Nobel de Física en 1933 por haber desarrollado la ecuación de Schrödinger.

Cómo estudiar el plegado?

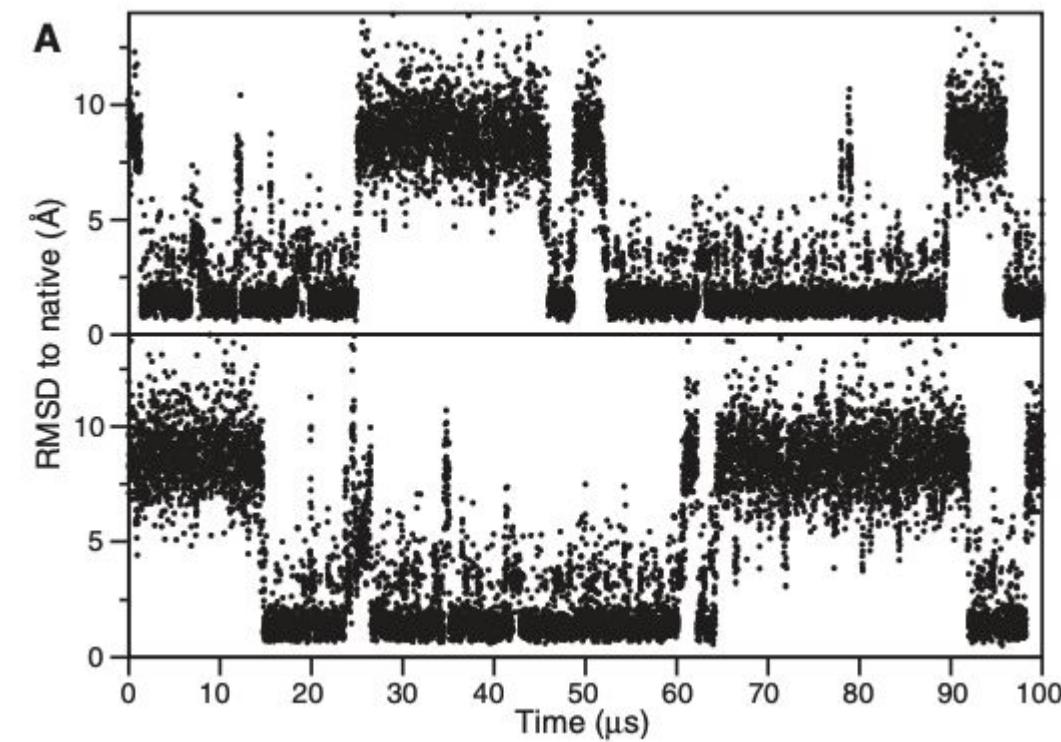
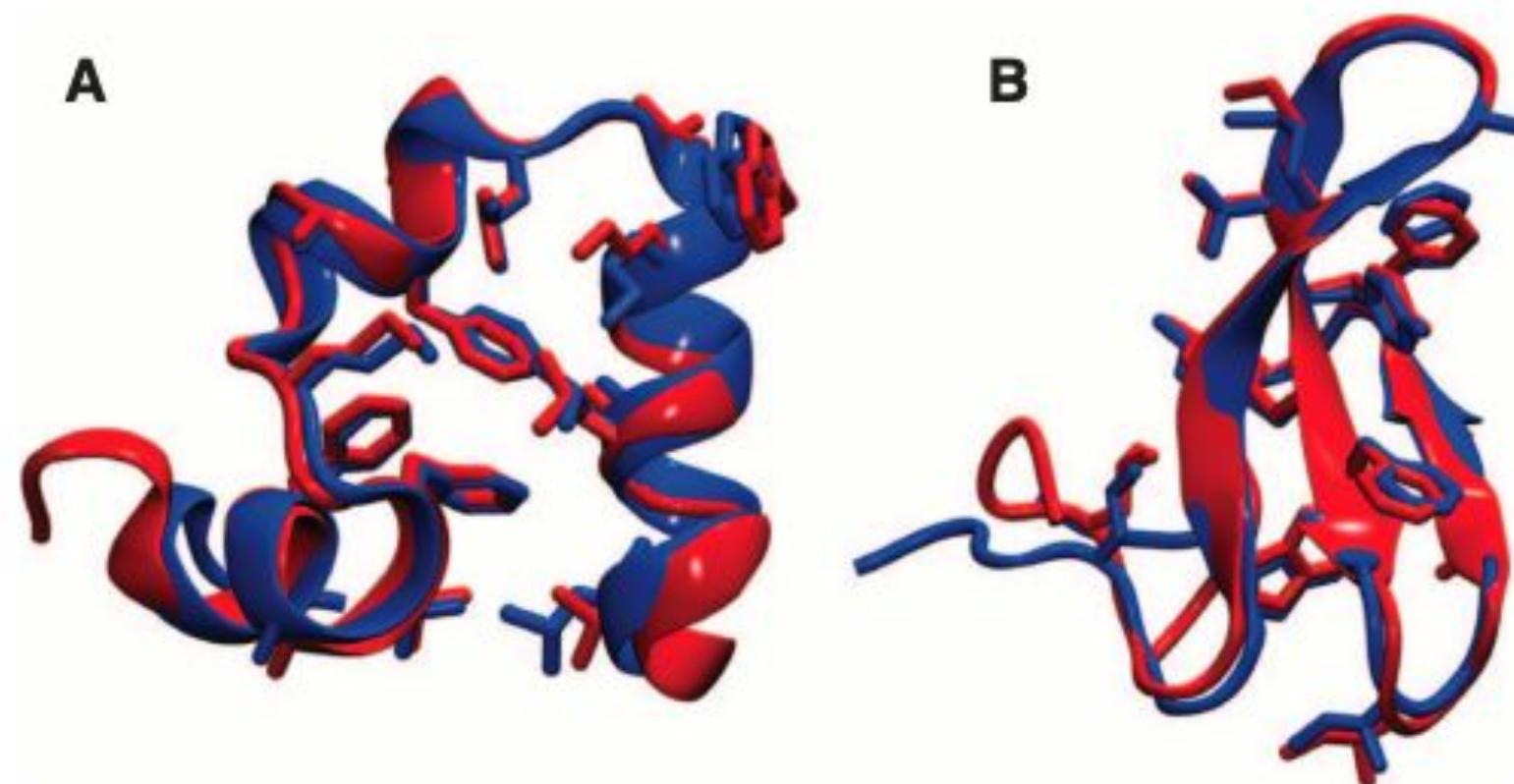
Mecánica clásica

$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) + \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 + \sum_{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] + \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$



Cómo estudiar el plegado?

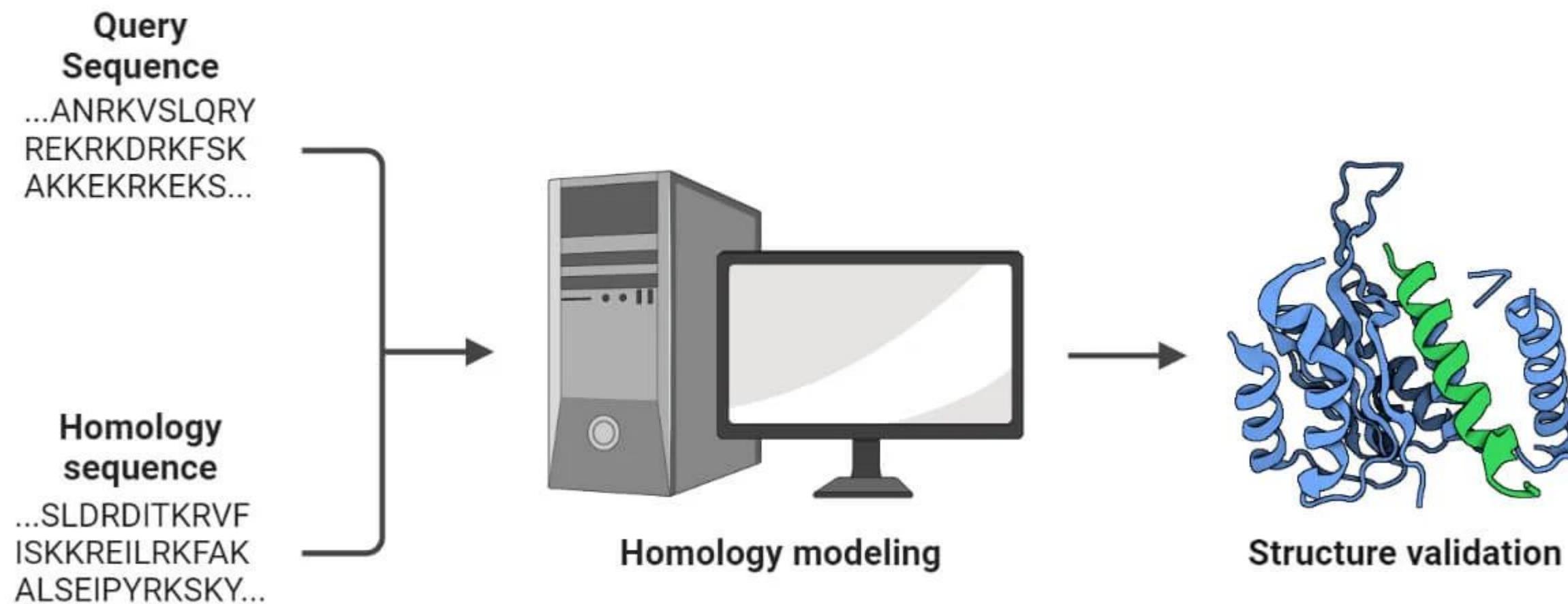
Mecánica clásica



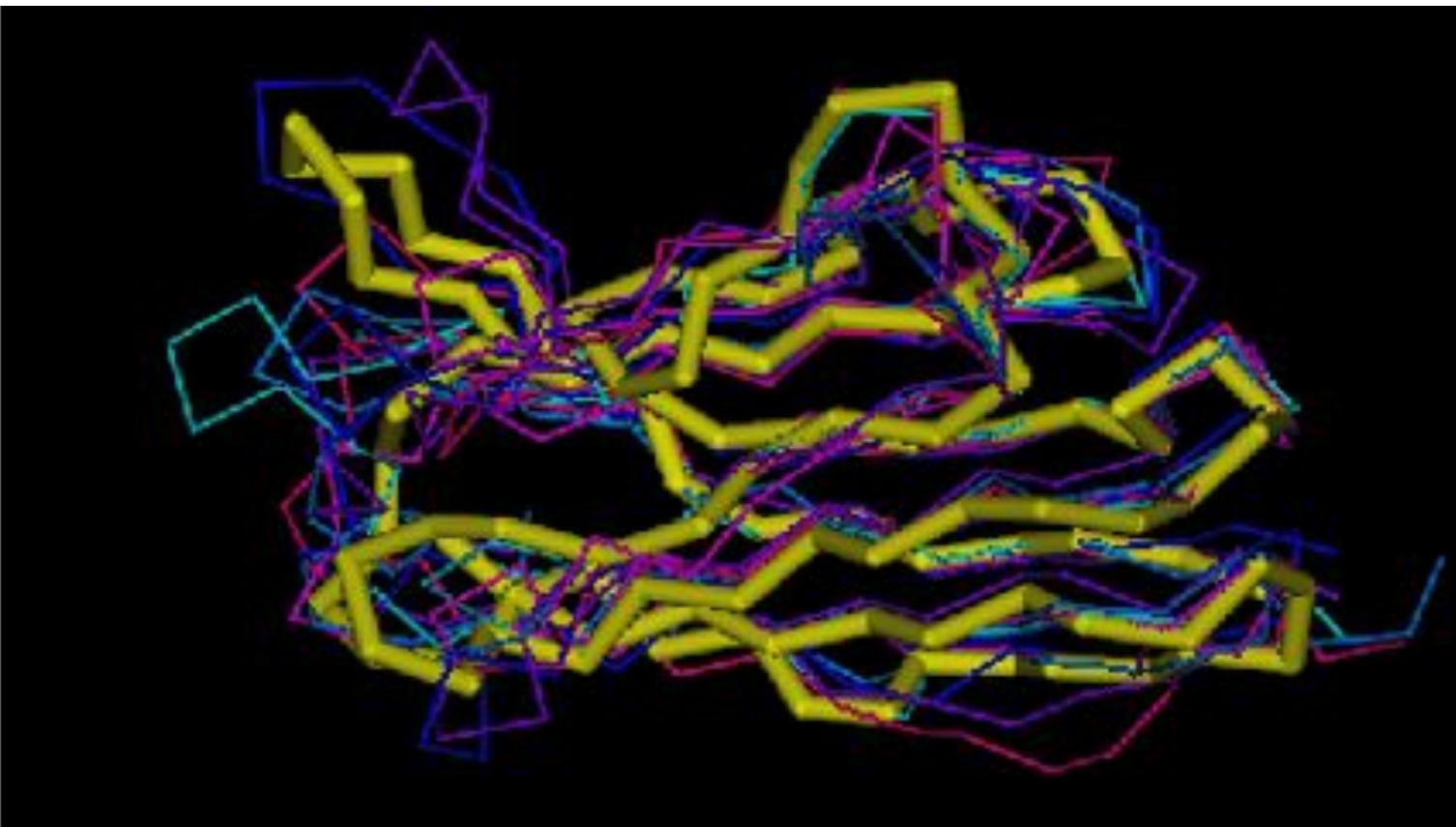
Atomic-Level Characterization of the Structural Dynamics of Proteins
David E. Shaw *et al.*
Science 330, 341 (2010);
DOI: 10.1126/science.1187409

Cómo encontrar la estructura?

Modelado Comparativo o Modelado Por Homología



Modelado Comparativo o Modelado Por Homología



ACDEFGHIKLMNPQRST--FGHQWERT-----TYREWYEGHADS
ASDEYAHLRILDPQRSTVAYAYE--KSFAPPGSFKWEYEAHADS
MCDEYAHIRLMNPERSTVAGGHQWERT---GSFKEWYAAHADD

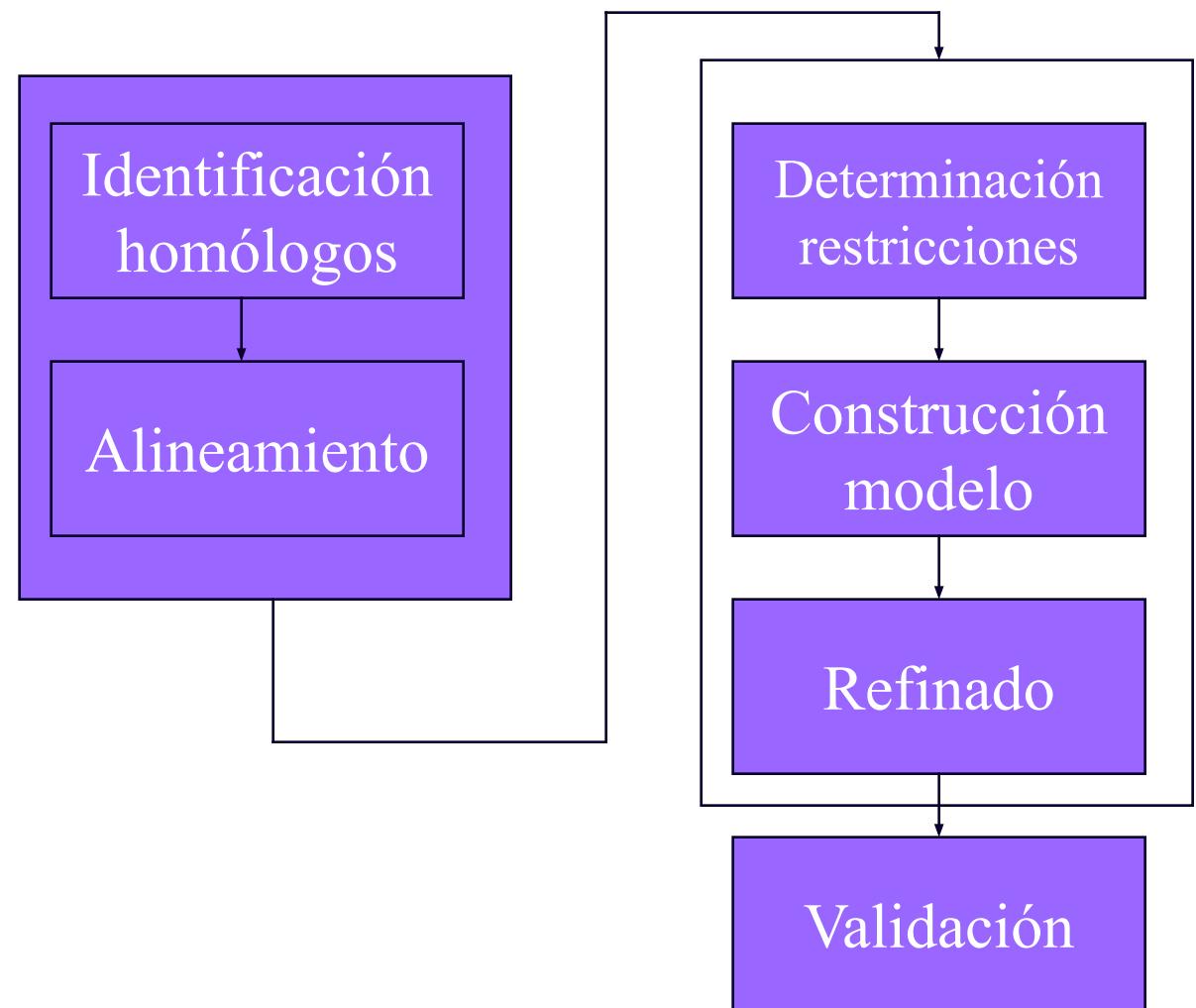
El universo de plegados o “folds”?

2.5 Mill millones
secuencias
(no redundantes)



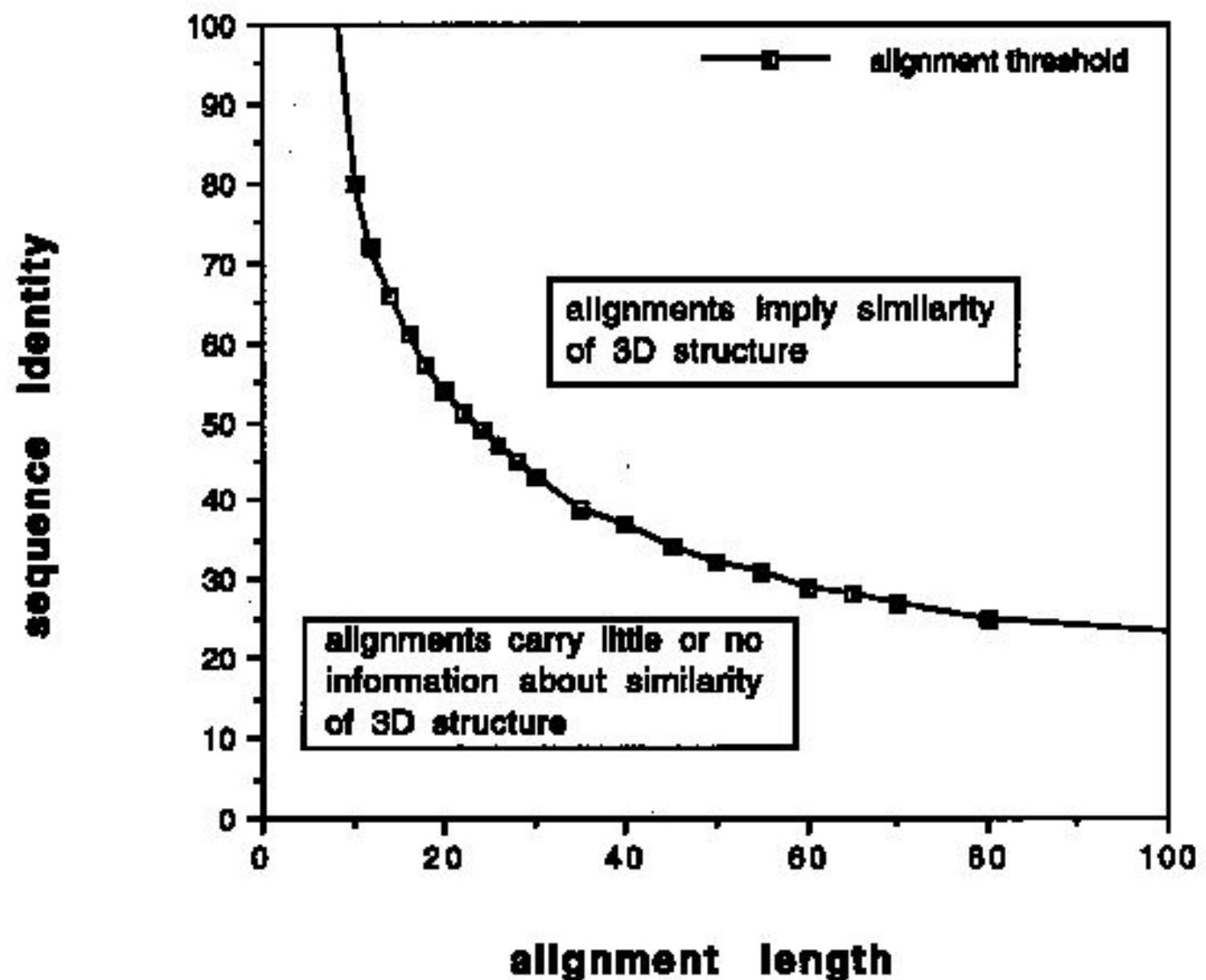
5000
plegados?

Etapas del modelado comparativo



- ▶ 1-Encontrar un templado adecuado (Ej: BLAST pdb).
- ▶ 2-Alineamiento Target-Template (Es el paso más importante).
 - Corregir el alineamiento
- ▶ 3-Construcción del modelo.
 - Modelo de la cadena carbonada
 - Modelado de los loops
 - Cadenas Laterales
- ▶ 4-Evaluación
- ▶ 5-Refinamiento del modelo

Alineamiento



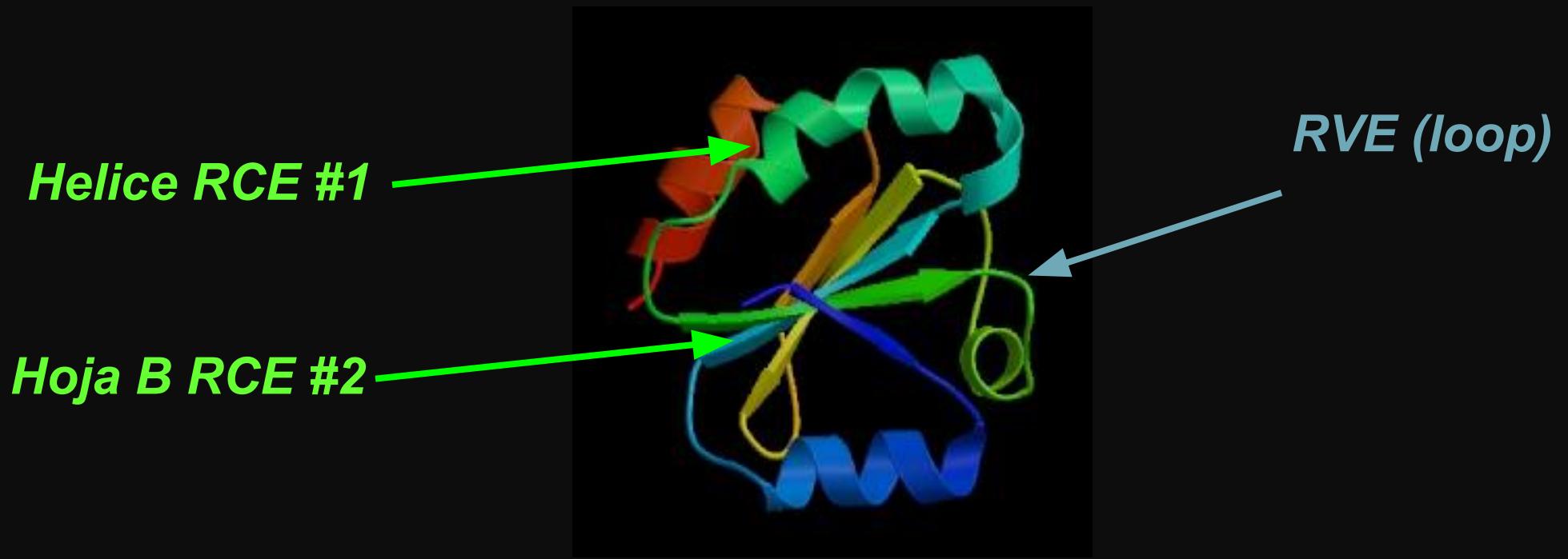
¿Cuándo identidad secuencia implica similitud estructural?

- Idealmente identidad > 40%
- Entre 25-40% zona gris
- alinear a lo largo de toda la secuencia “target”
- Pequeños errores en alineamiento grandes errores en el modelo

Búsqueda de regiones conservadas y variables estructuralmente

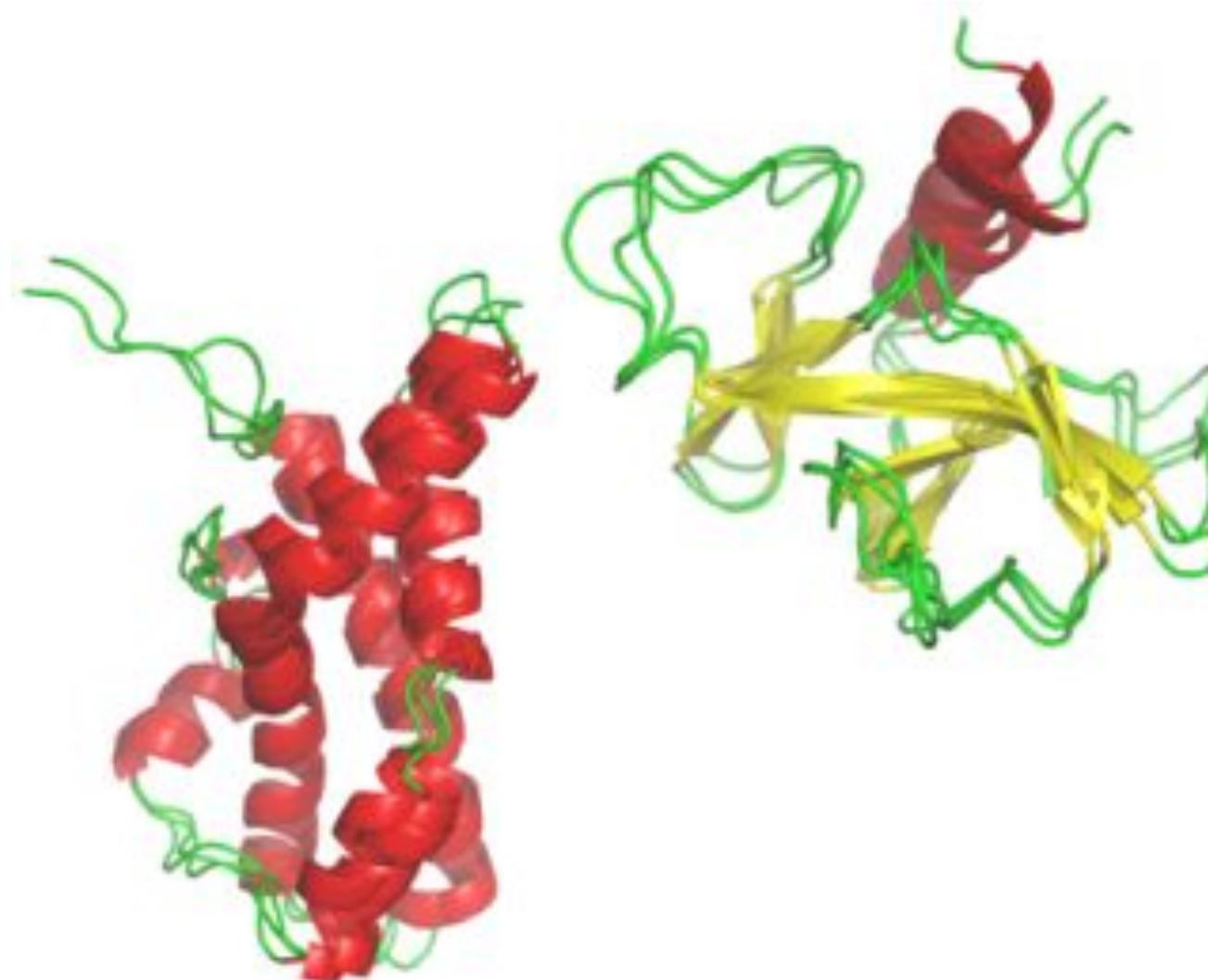
Target	ACDEF GH I KL MNPQR S T--FG HQ WERT----TYREWYEG
Tmpl.1	ASDE YA H L R I LDPQR S T V A Y A YE --KS FAE FGS FK WEYEA
Tmpl.2	MCDE YA H I RLMN P ER S T V A G G HQ WERT----GS FKEWYAA
	HHHHHHHHHHHHCCCCCCCCCCCCCCCCCBBBBBBBBBBB

RCE #1 *RVE (loop)* *RCE #2*



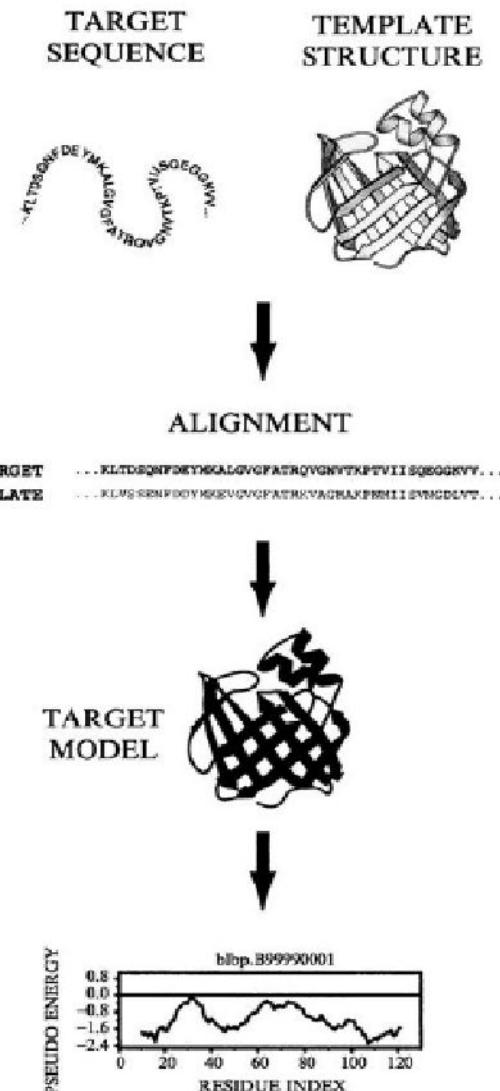
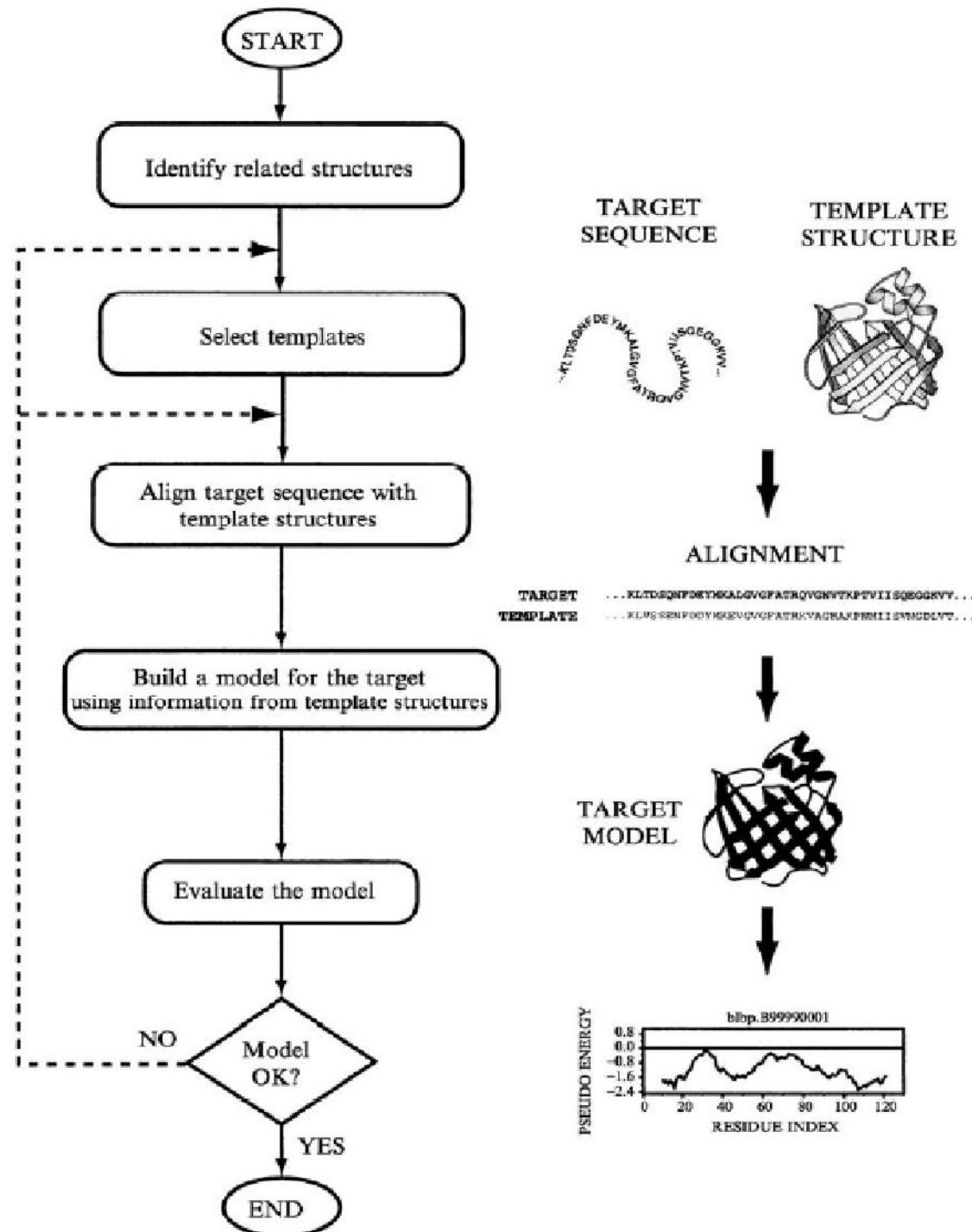
La calidad del modelo va a depender de la conservación

Modelado de los loops



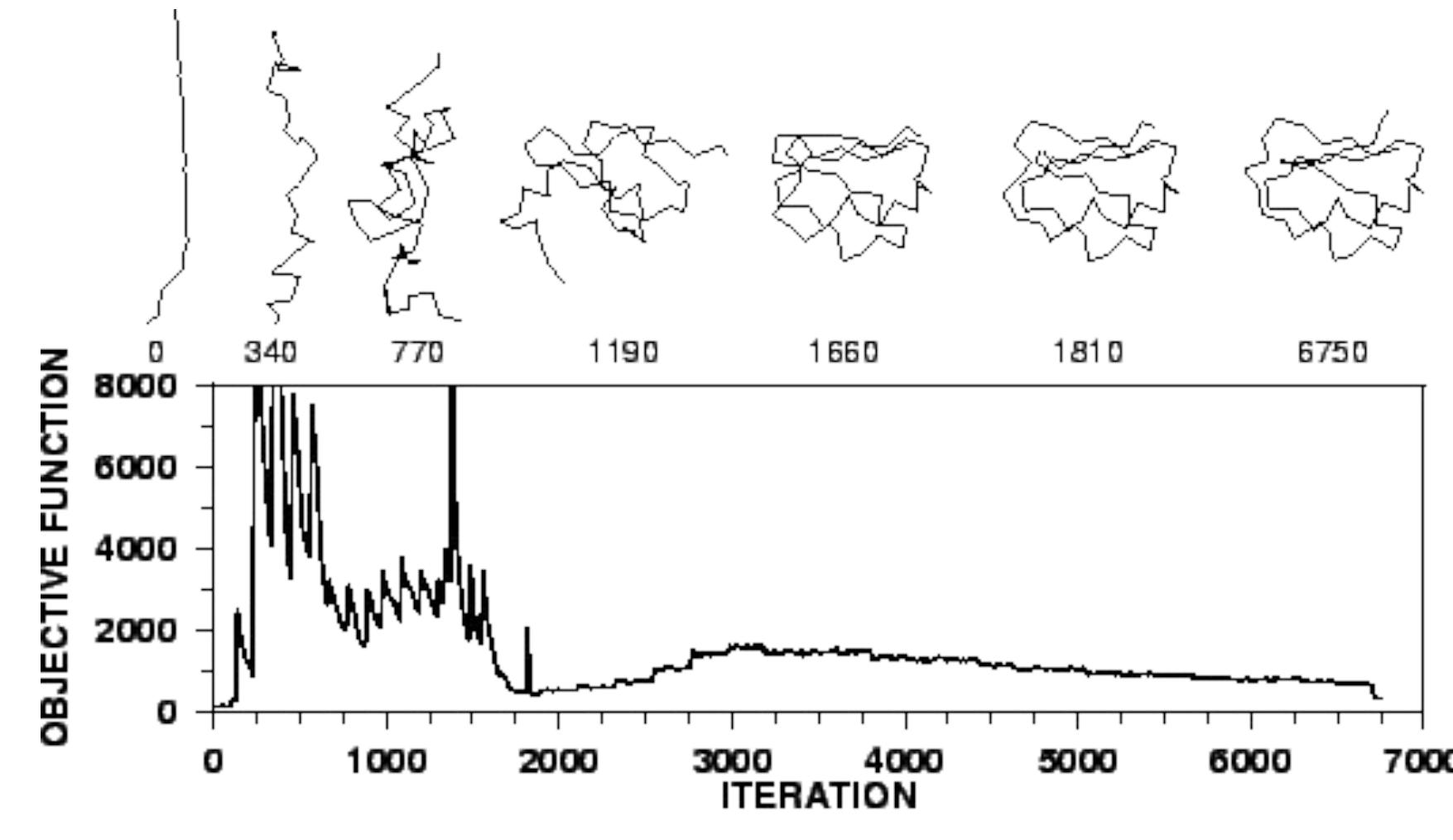
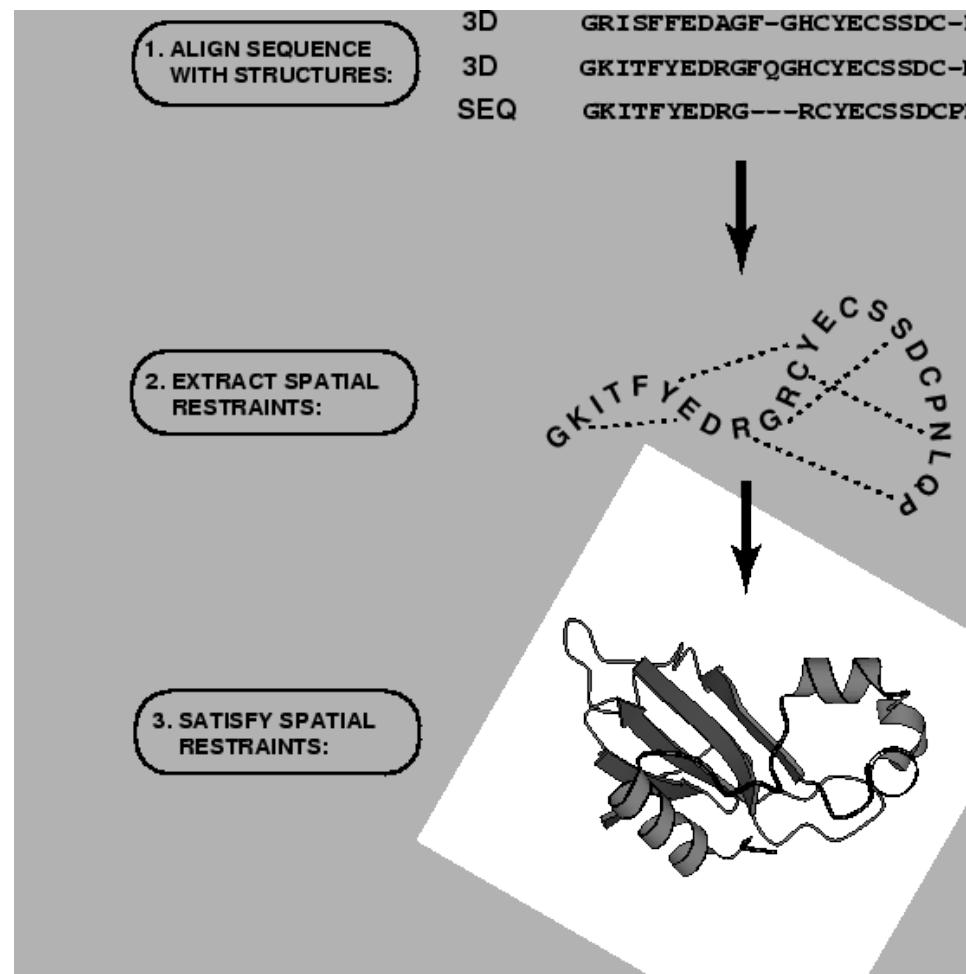
- Son flexibles!
- Cambian de conformación en solución!
- Cambian de conformación al interactuar!

Modeler un programa ejemplo



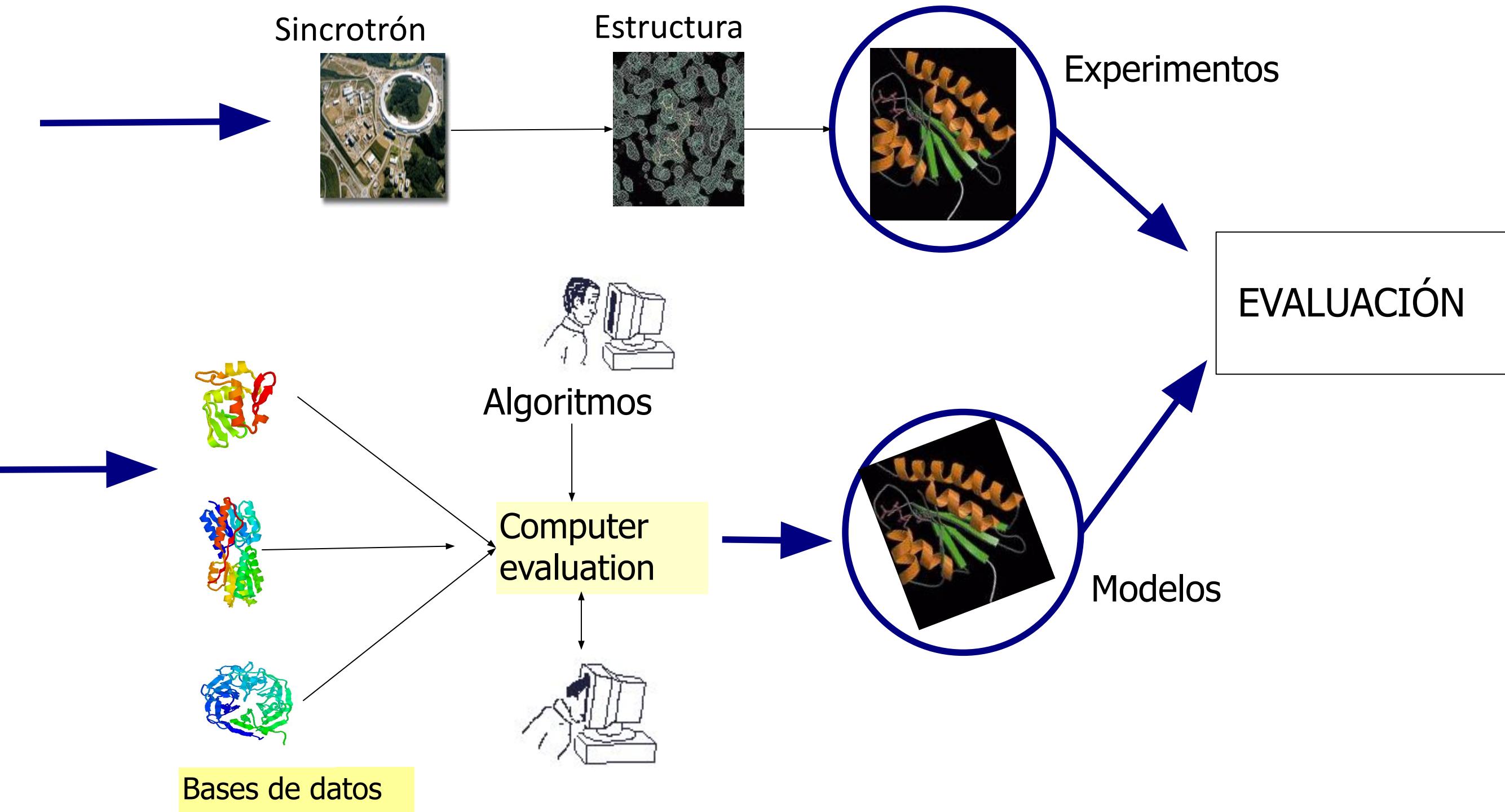
- Fue funcional para copiar estructuras
- Basado en restraints parecido a NMR
- Funcionó bien con identidad arriba de 30%

Modeler un programa ejemplo

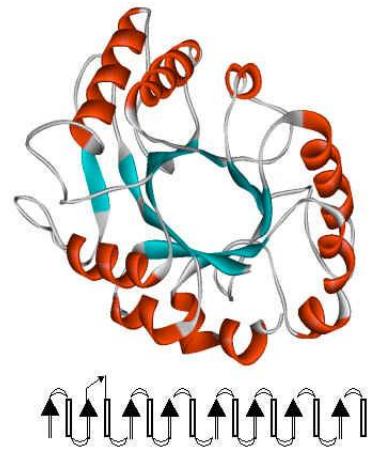


Evaluación de los modelos

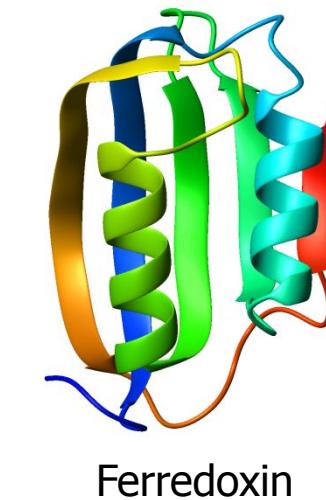
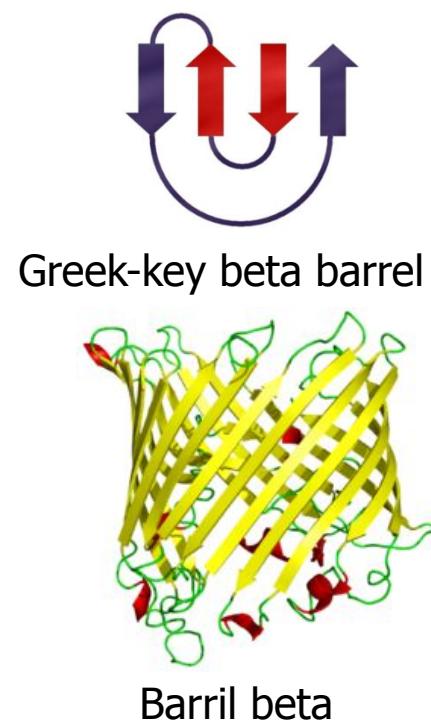
MAKEFGIPAAVAGTVLNVVEAGGWVTTIVSILTAGSG
GLSLLAAGGRESIKAYLKKEI KKGKRAVIAW



Threading o enebrado



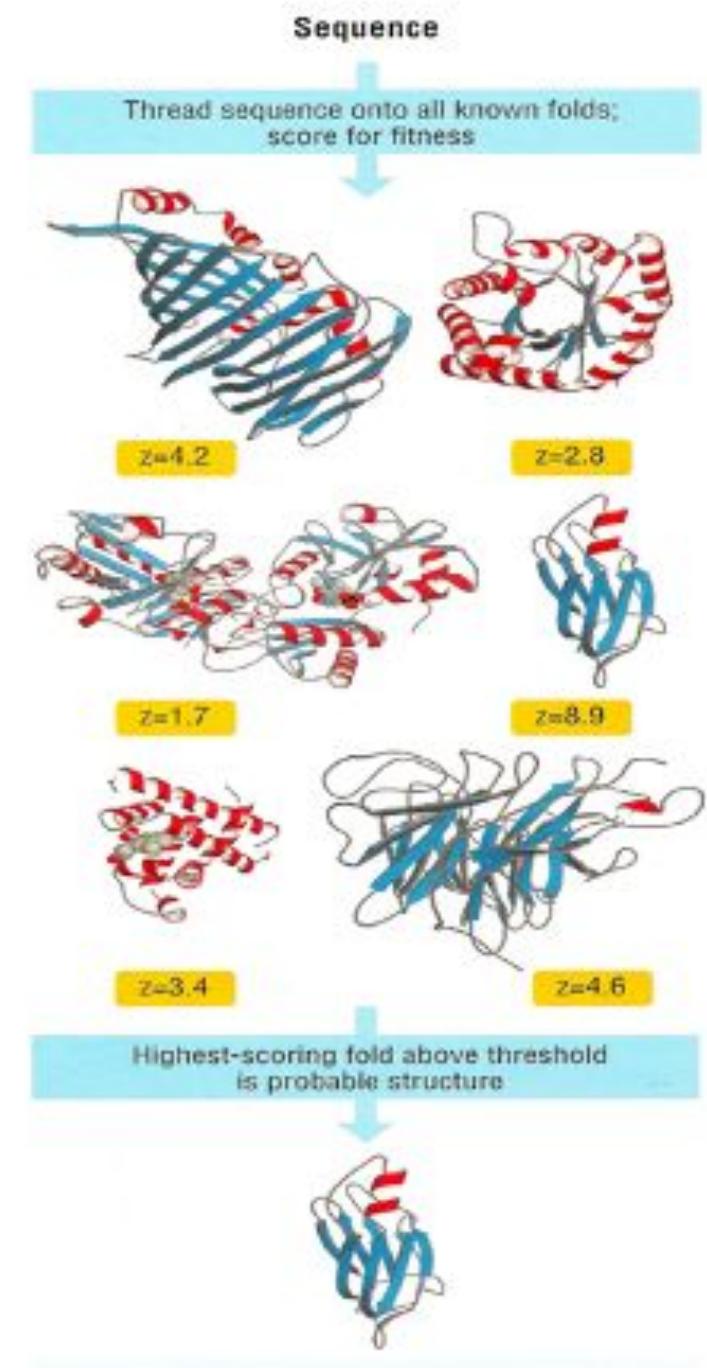
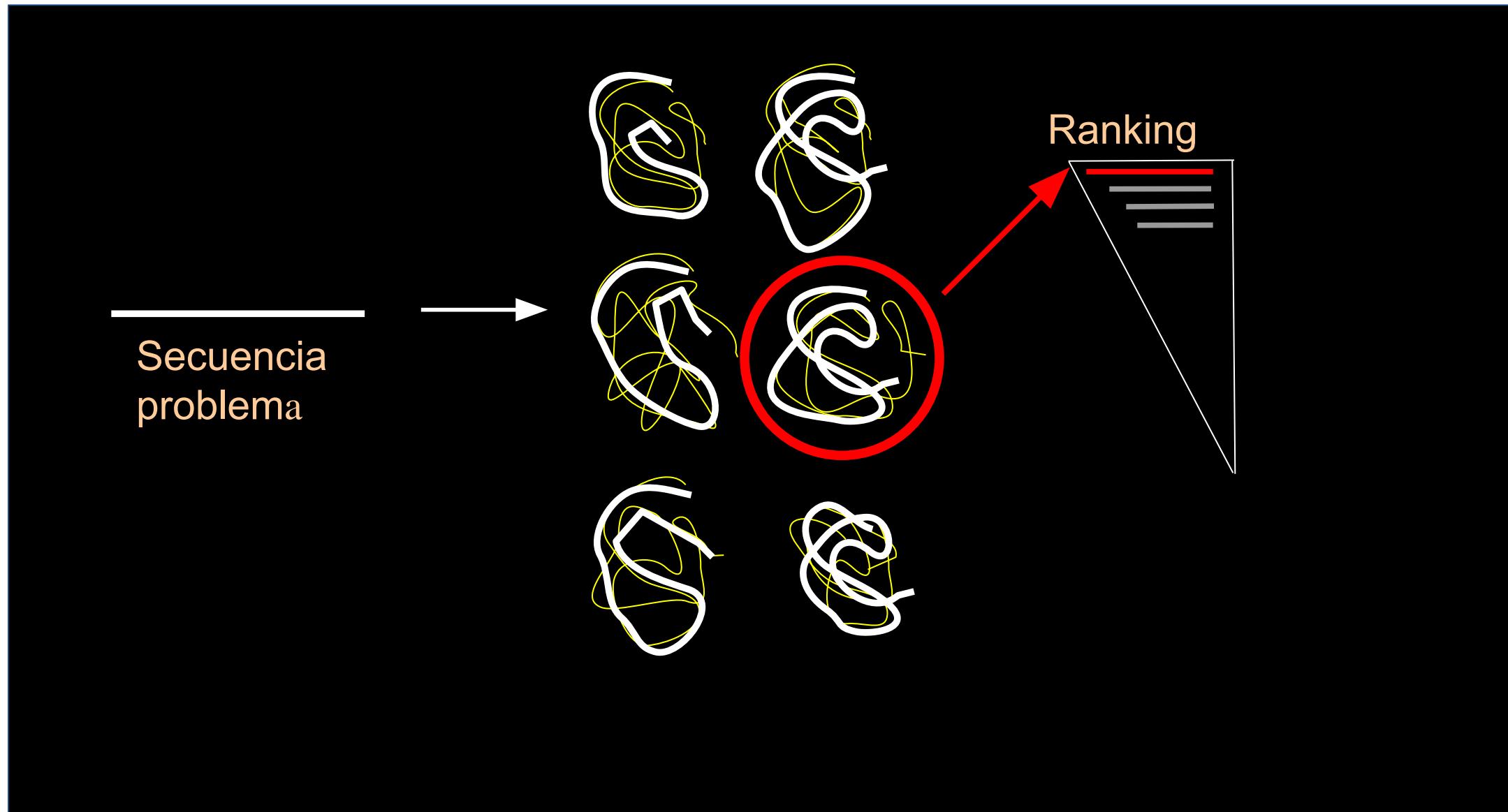
(TIM barrel)
aprox 10% de las
enzimas



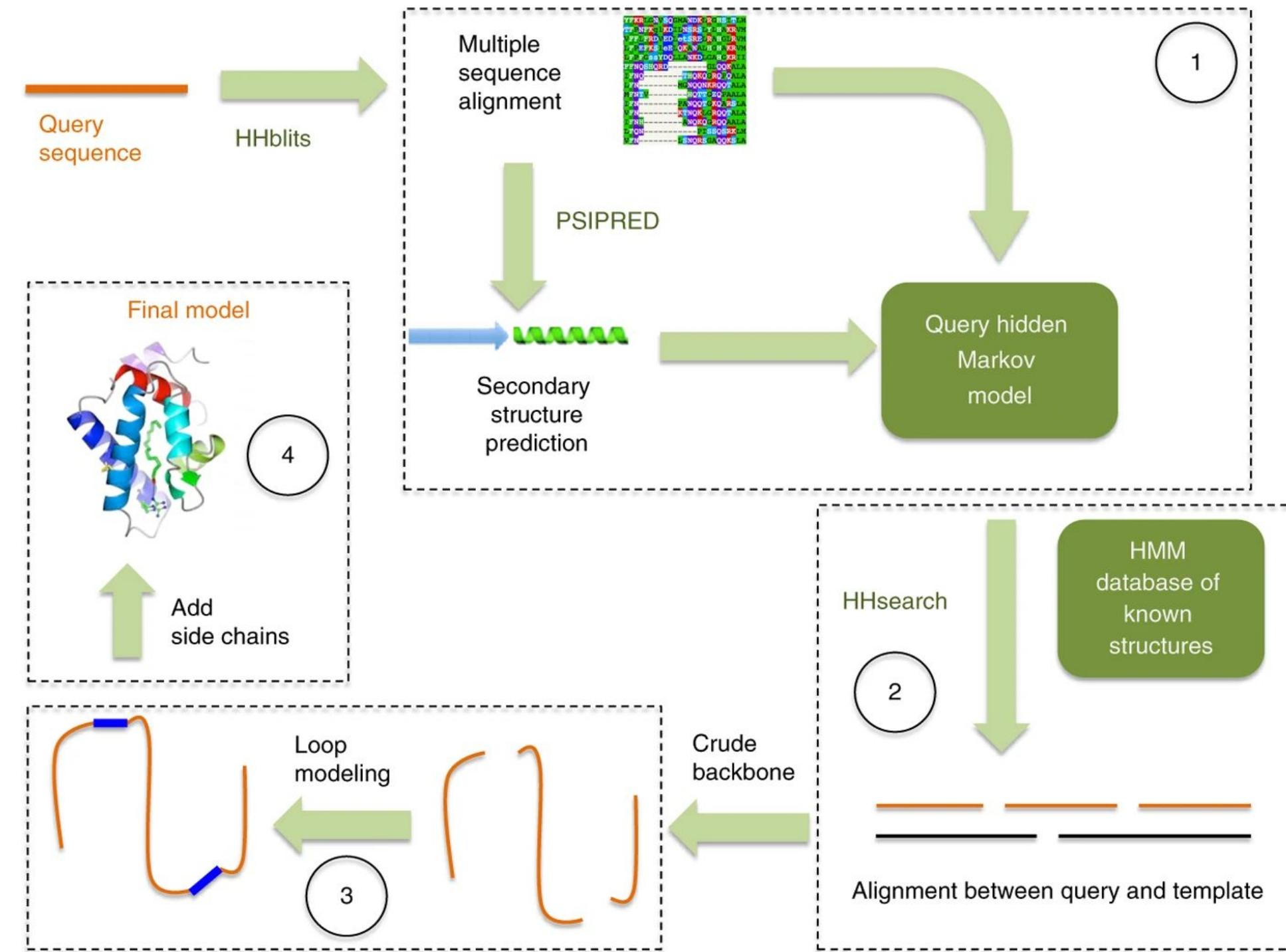
- **Evolución divergente:**
Estas proteínas están relacionadas solo que se modificaron a lo largo tiempo.
- **Evolución Convergente:**
Requerimientos comunes llegan al mismo plegado: Por ejemplo reconocimiento de los mismos blancos proteicos. Poco común.

Proteínas disímiles en secuencia
adoptan plegamientos similares.

Threading o enebrado

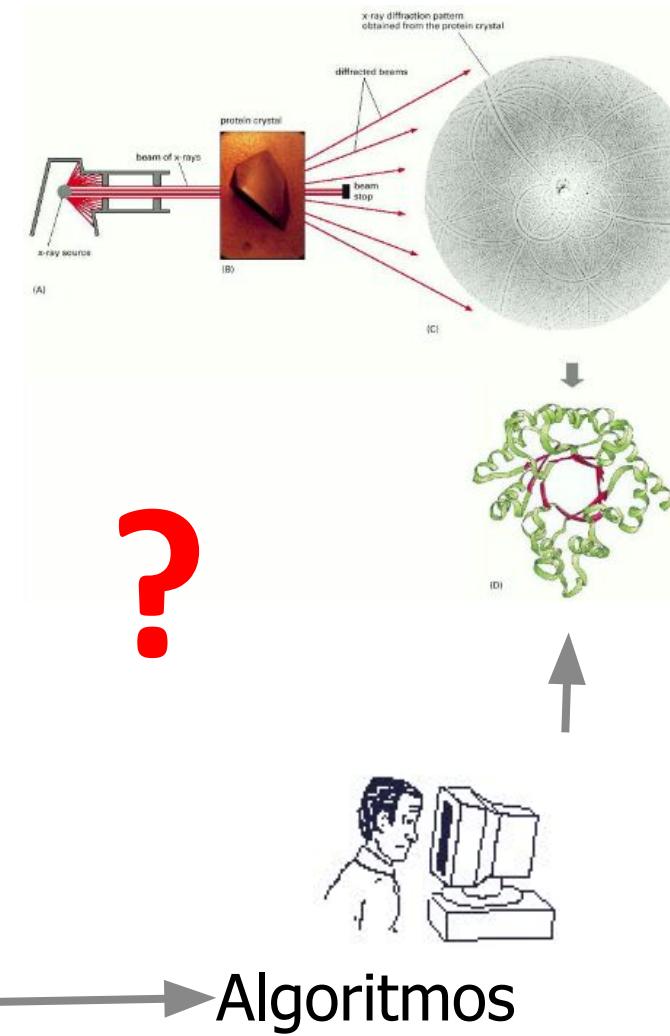


Phyre: Método que fue exitoso



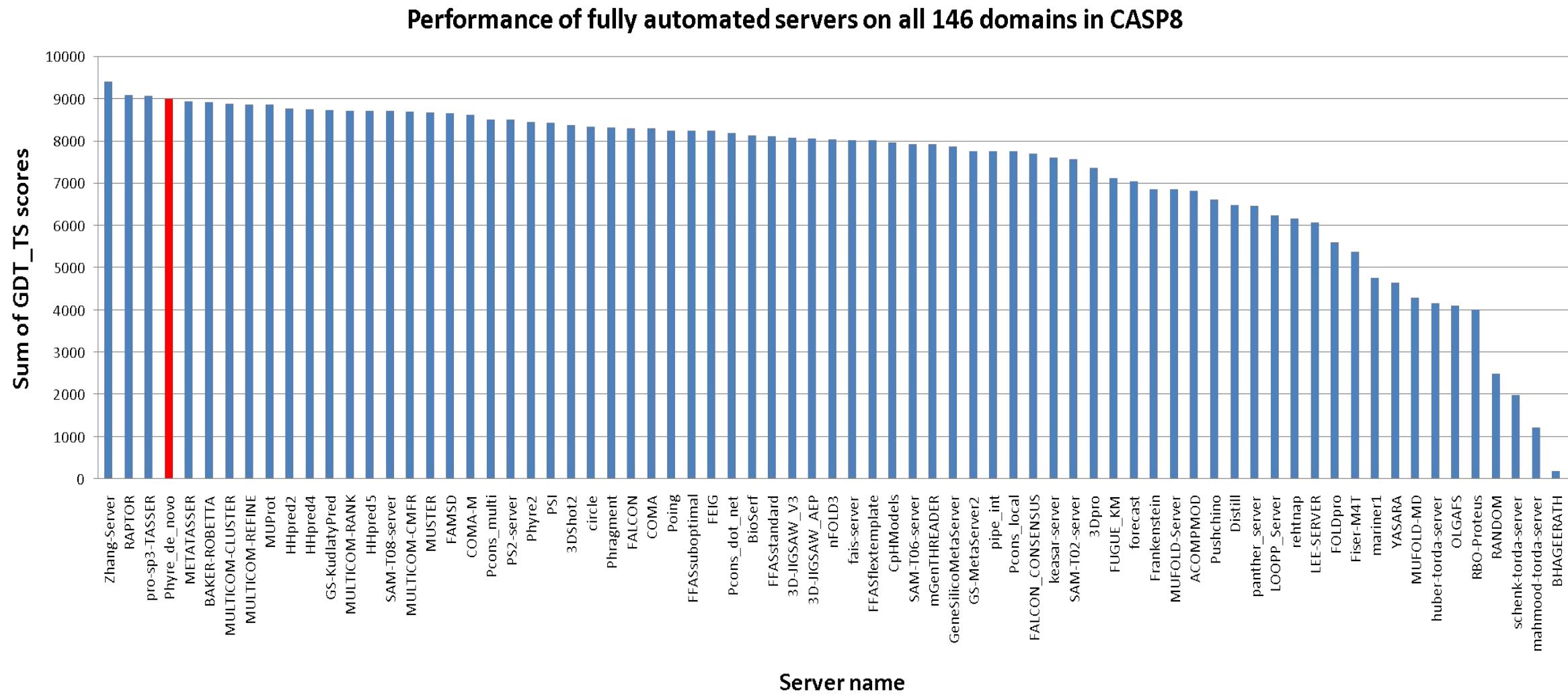
CASP: Critical Assessment of Protein Interactions

MAKEFGIPAAWAGTVLNVV
EAGGGWVTTIVSILTAVGSG
GLSLLAAAGRESIKAYLKK
EI KKGKRAVIAW

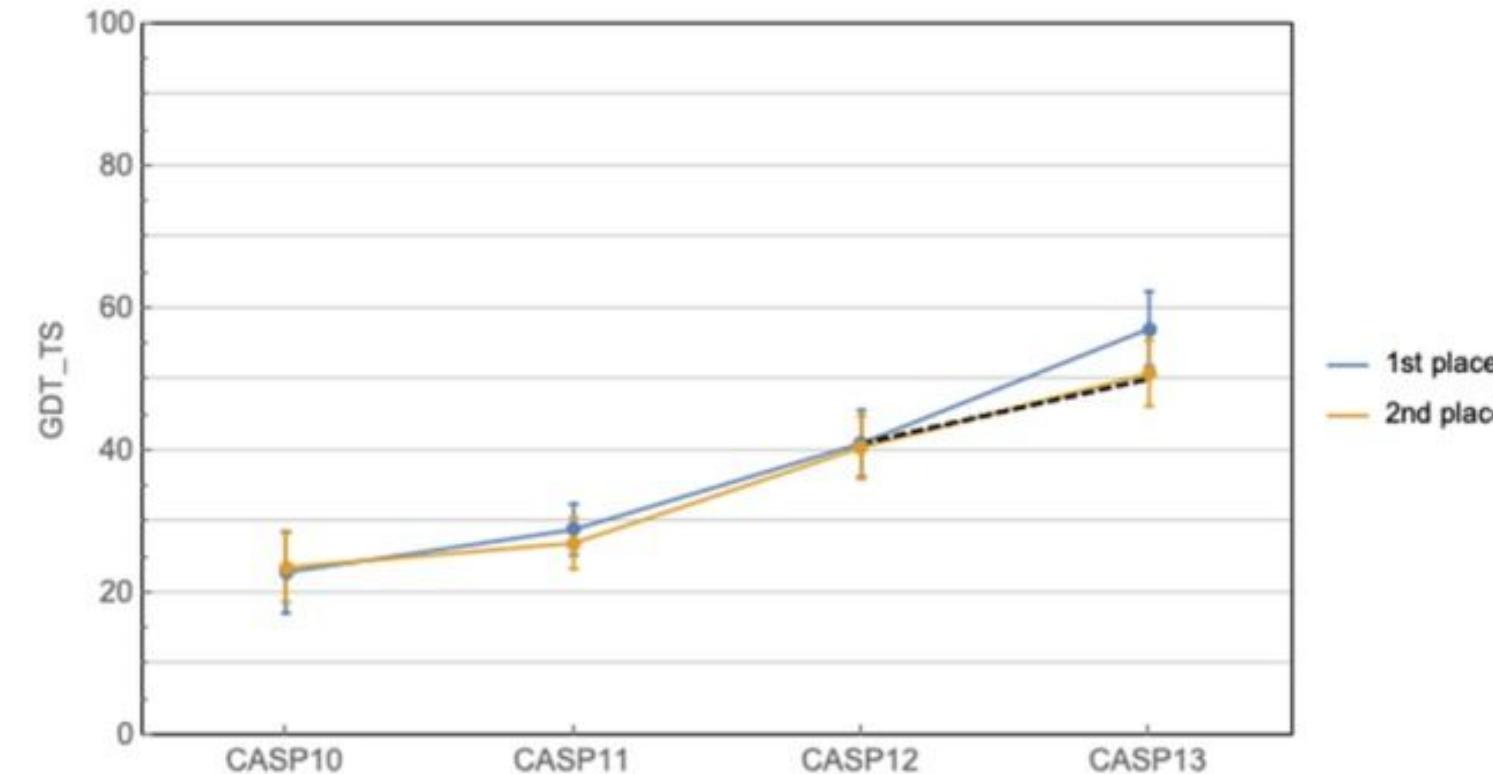
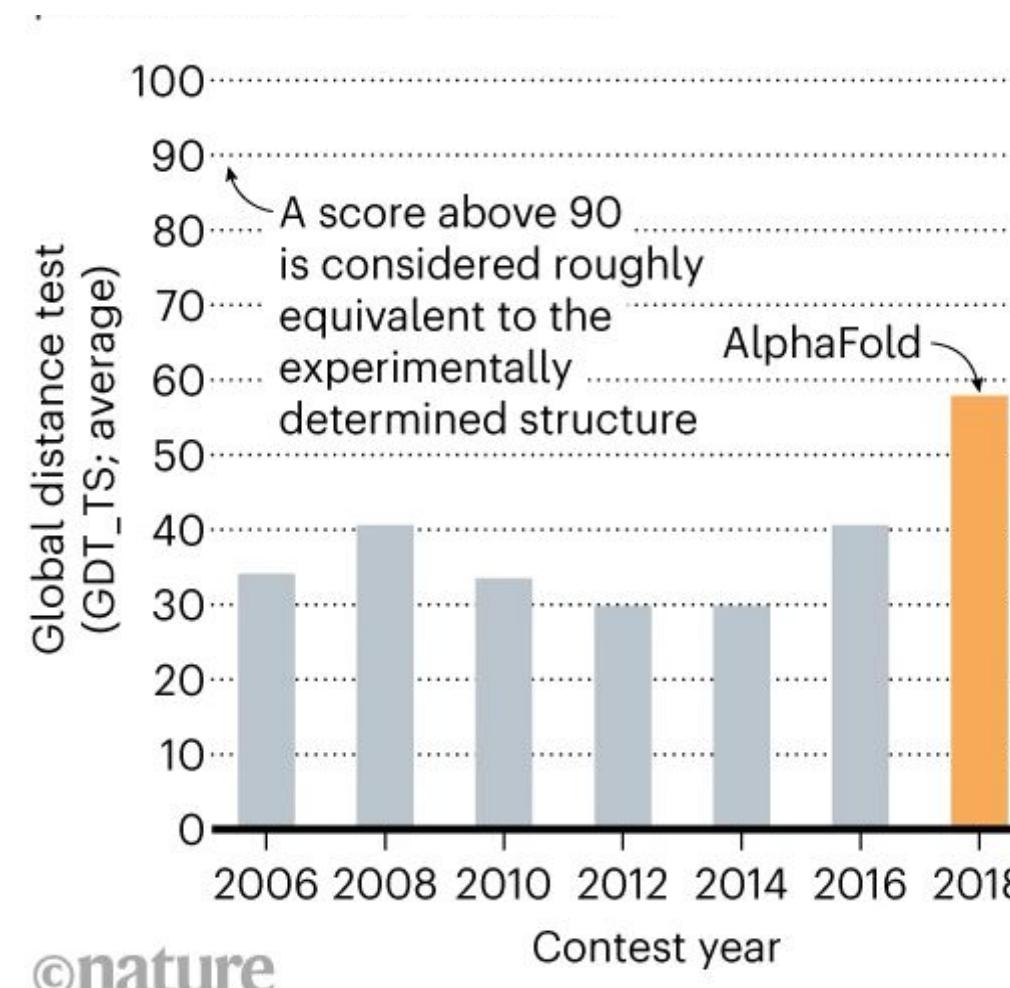


- The main goal of CASP is to obtain an in-depth and objective assessment of our current abilities and inabilities in the area of **protein structure prediction**. To this end, participants will predict as much as possible about a set of **soon to be known structures**. These will be true predictions, not 'post-dictions' made on already known structures.

CASP: Critical Assessment of Protein Interactions



La primera participación de AlphaFold en CASP: Un avance significativo, pero con margen de mejora



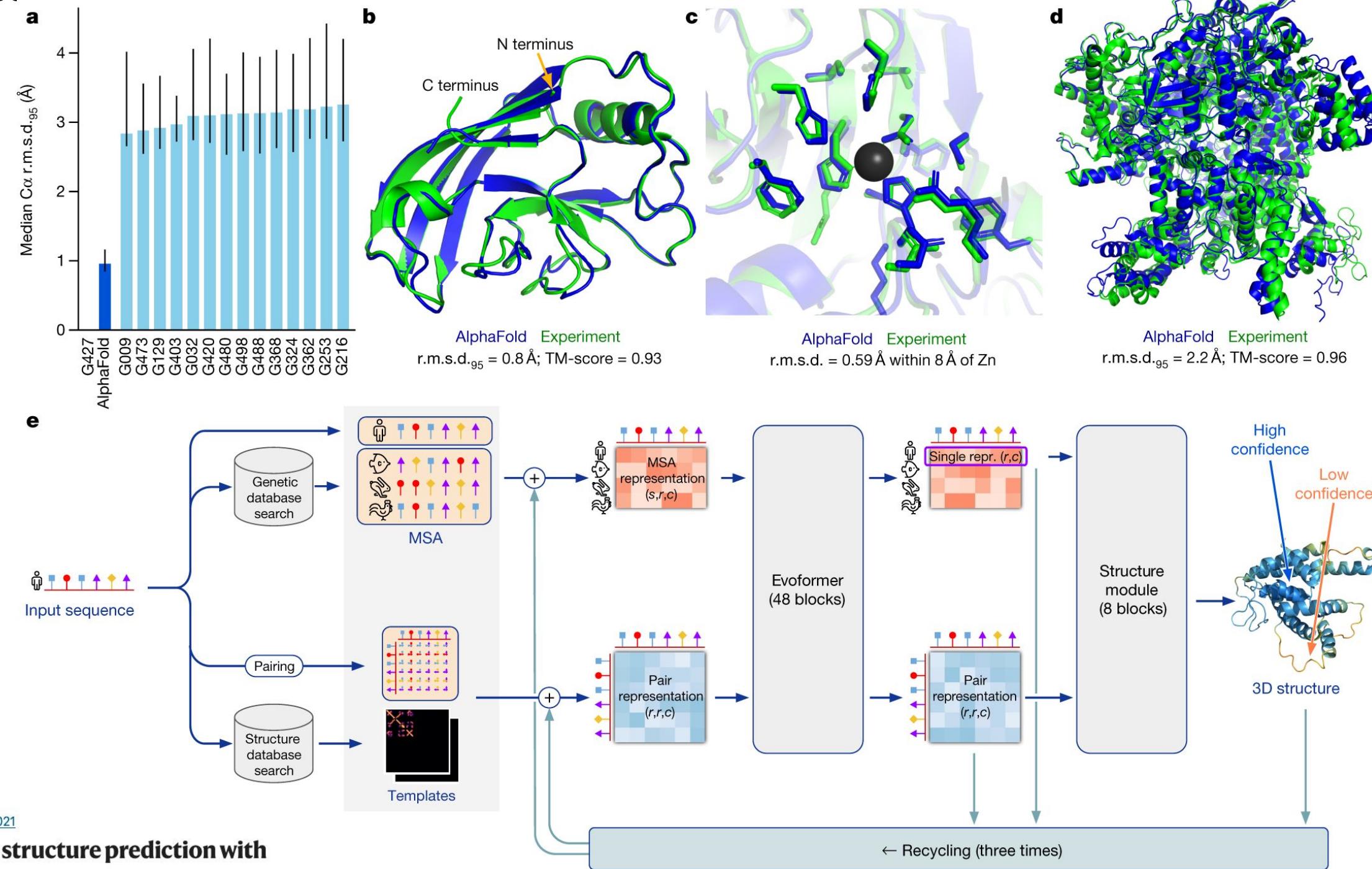
Año: 2018

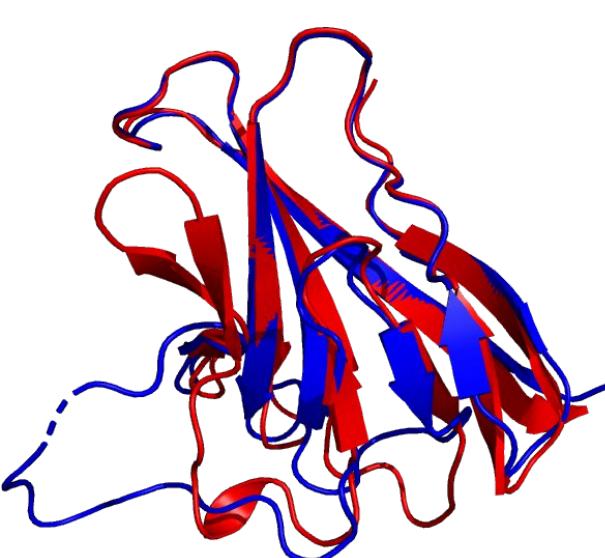
Concurso: Critical Assessment of Protein Structure Prediction (CASP) 13.

Logro: AlphaFold predijo la estructura de proteínas con una precisión sin precedentes, superando a otros métodos.

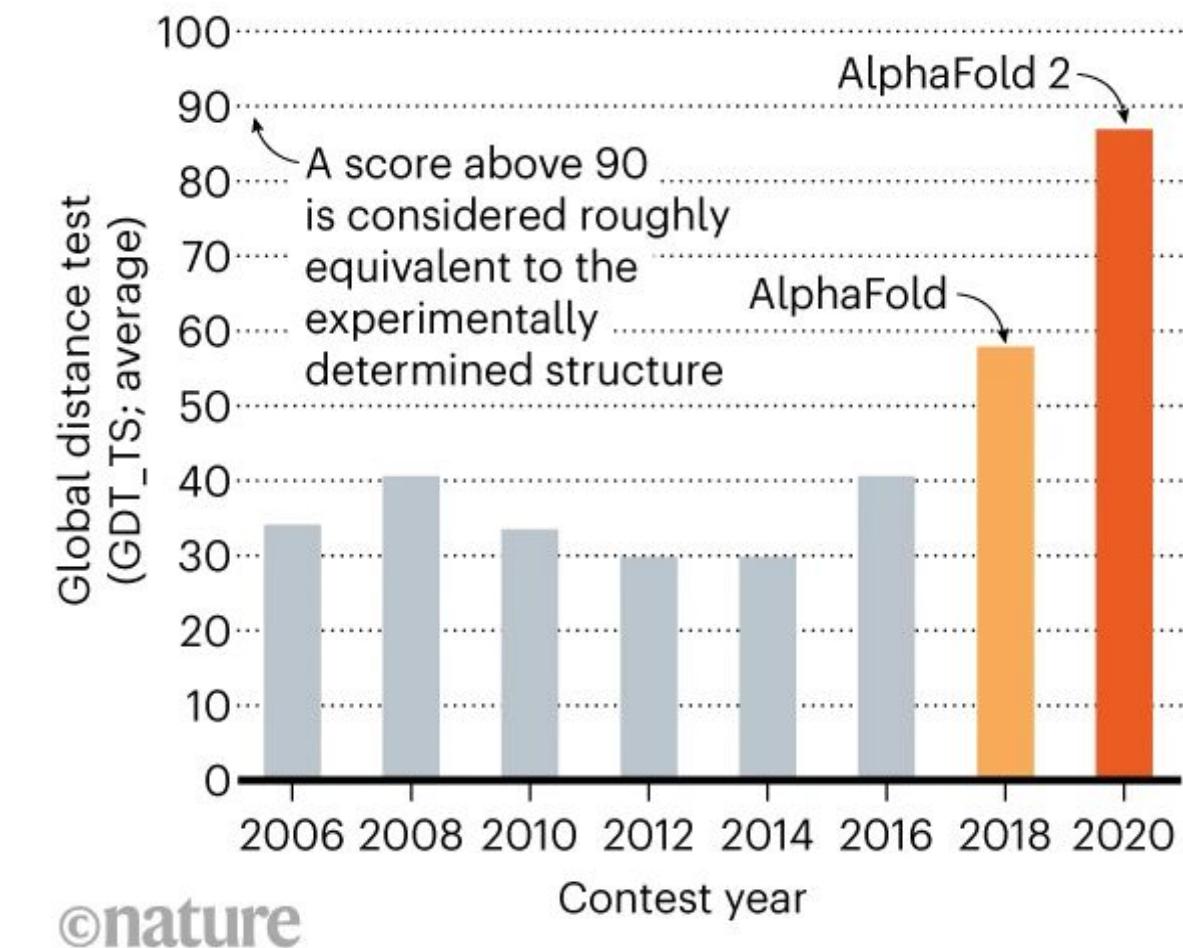
Limitaciones: A pesar del avance significativo, AlphaFold no alcanzó la precisión deseada en algunas estructuras de proteínas complejas.

ALPHAFOLD: De la secuencia a la estructura





La incorporación de biólogos impulsa el rendimiento de AlphaFold



©nature

Año: 2020

Concurso: Critical Assessment of Protein Structure Prediction (CASP) 14.

Colaboración: Tras la participación inicial en CASP, DeepMind incorporó biólogos al equipo de desarrollo de AlphaFold.

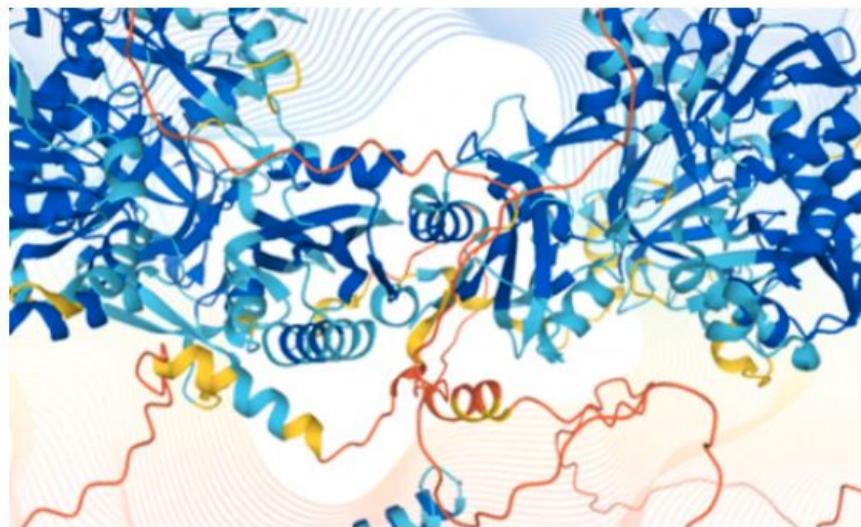
Logro: AlphaFold-2 predijo la estructura de proteínas la cual logró una mejora significativa en la precisión casi comparable con la experimental.

ALPHAFOLD: De la secuencia a la estructura

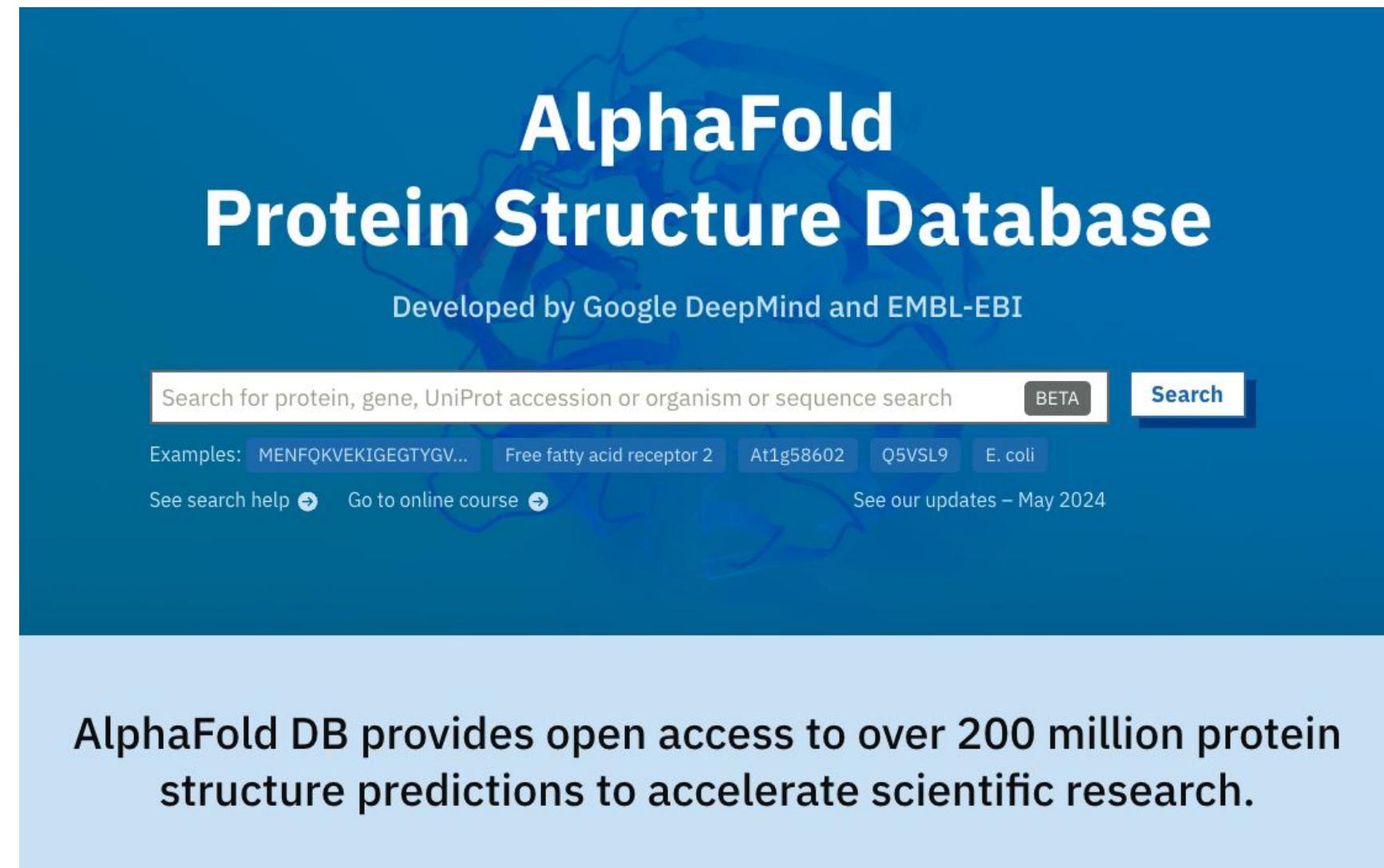
- Here we provide the first computational method that can regularly predict protein structures **with atomic accuracy** even in cases in which no similar structure is known. We validated an **entirely redesigned version of our neural network-based model**, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)¹⁵, demonstrating **accuracy competitive with experimental structures** in a majority of cases and greatly outperforming other methods.
- Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates **physical and biological knowledge about protein structure**, **leveraging multi-sequence alignments**, into the design of the deep learning algorithm.

ALPHAFOLD Protein Structure Database

DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins



Protein structures representing the data obtained via AlphaFold. Source image: AlphaFold. Design credit: Karen Arnott/EMBL-EBI



AlphaFold Protein Structure Database

Developed by Google DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism or sequence search BETA Search

Examples: MENFKQKVEKIGEGTYGV... Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli

See search help Go to online course See our updates – May 2024

AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.

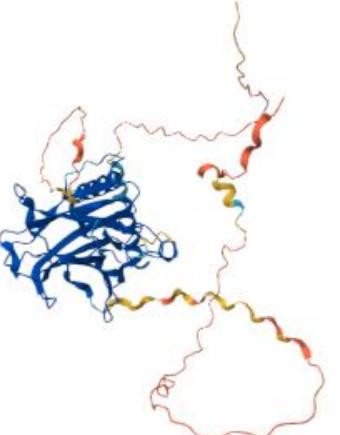
<https://alphafold.ebi.ac.uk/>

ALPHAFOLD Protein Structure Database

Structure viewer

Sequence of AF-P04637-F1 Chain 1: Cellular tu... A

MEEPQSDPSVEPLSQETFSDLWKLLPENNVLSPQLPSQAMDDLMLSPDIEQWFTEDPGPDEAPR
MPEAAPPVAPAPAAPTAAAPAPAPSWLSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPAL
NKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAITYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLI



Structure Tools

Structure AF-P04637-F1 Type Model Nothing Focused

Quick Styles Default Stylized Illustrative

Components AF-P04637-F1 Preset + Add Polymer Cartoon

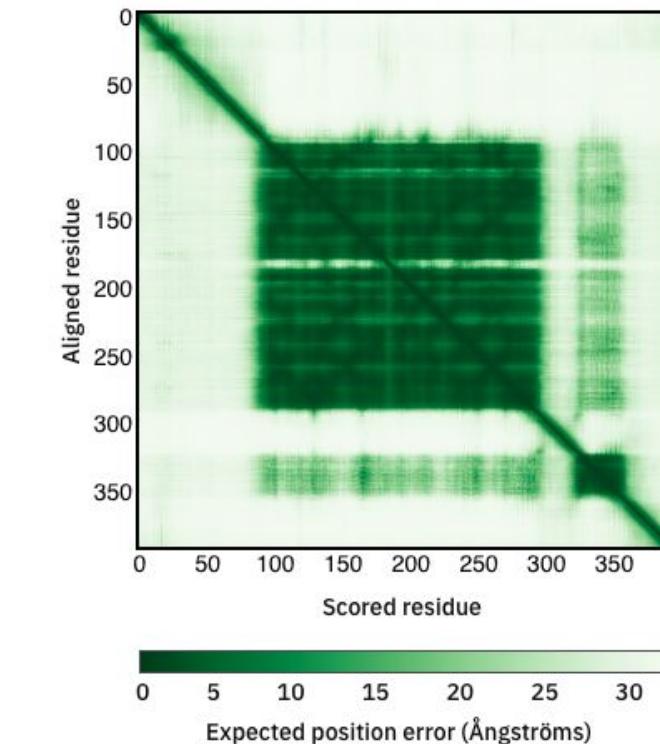
Measurements + Add

Model Confidence AlphaMissense Pathogenicity

Very high (pLDDT > 90) High (90 > pLDDT > 70) Low (70 > pLDDT > 50) Very low (pLDDT < 50)

AlphaFold produces a per-residue model confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Hide colour legend ^

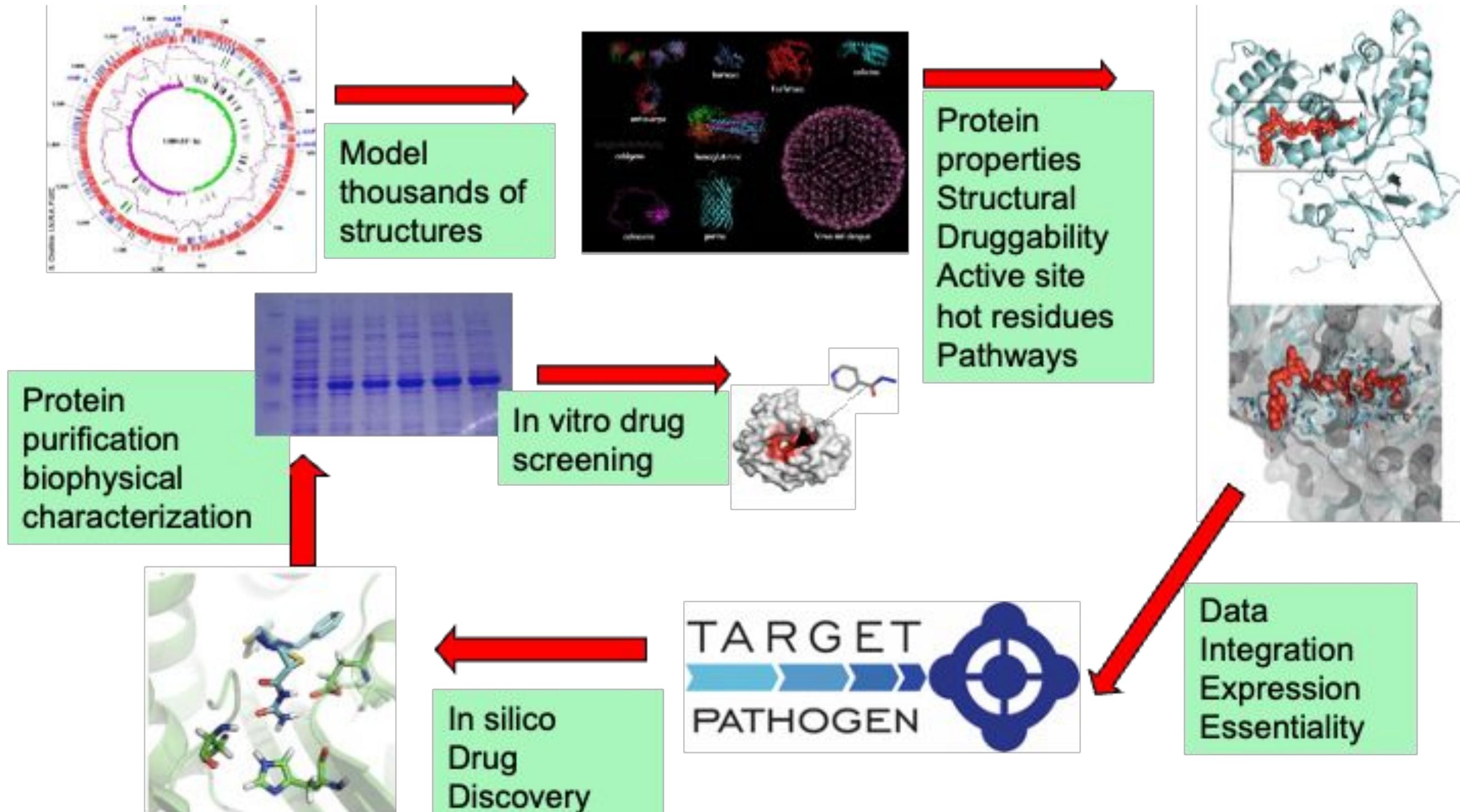


Predicted aligned error (PAE)

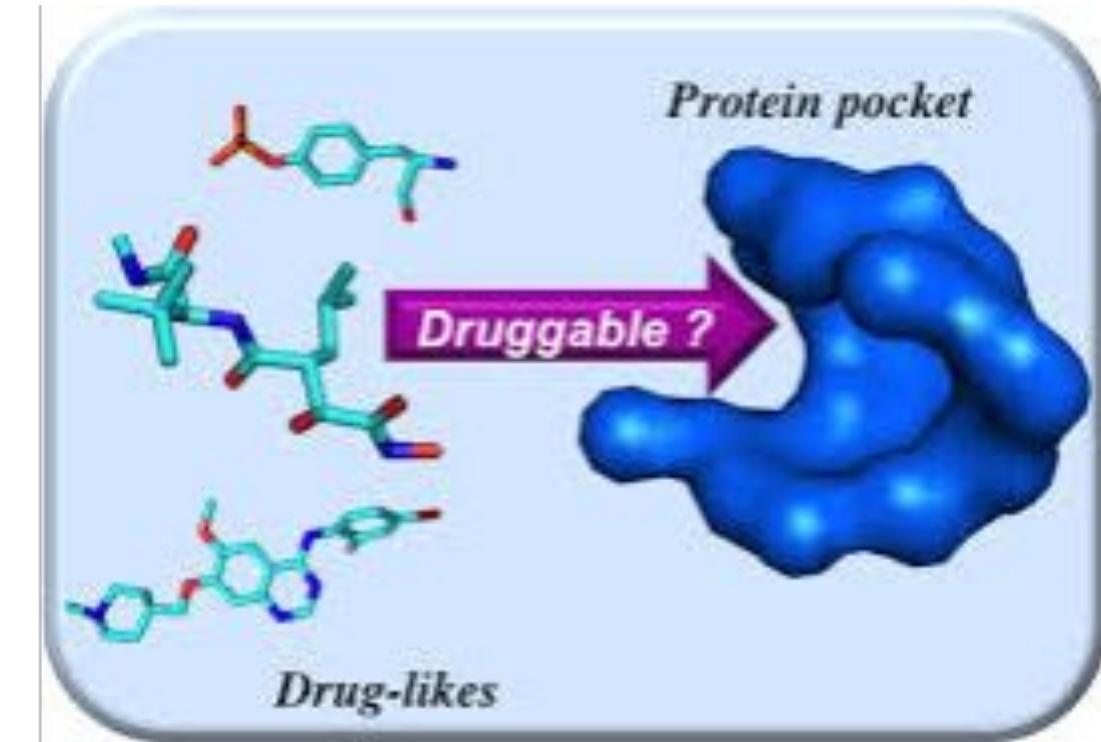
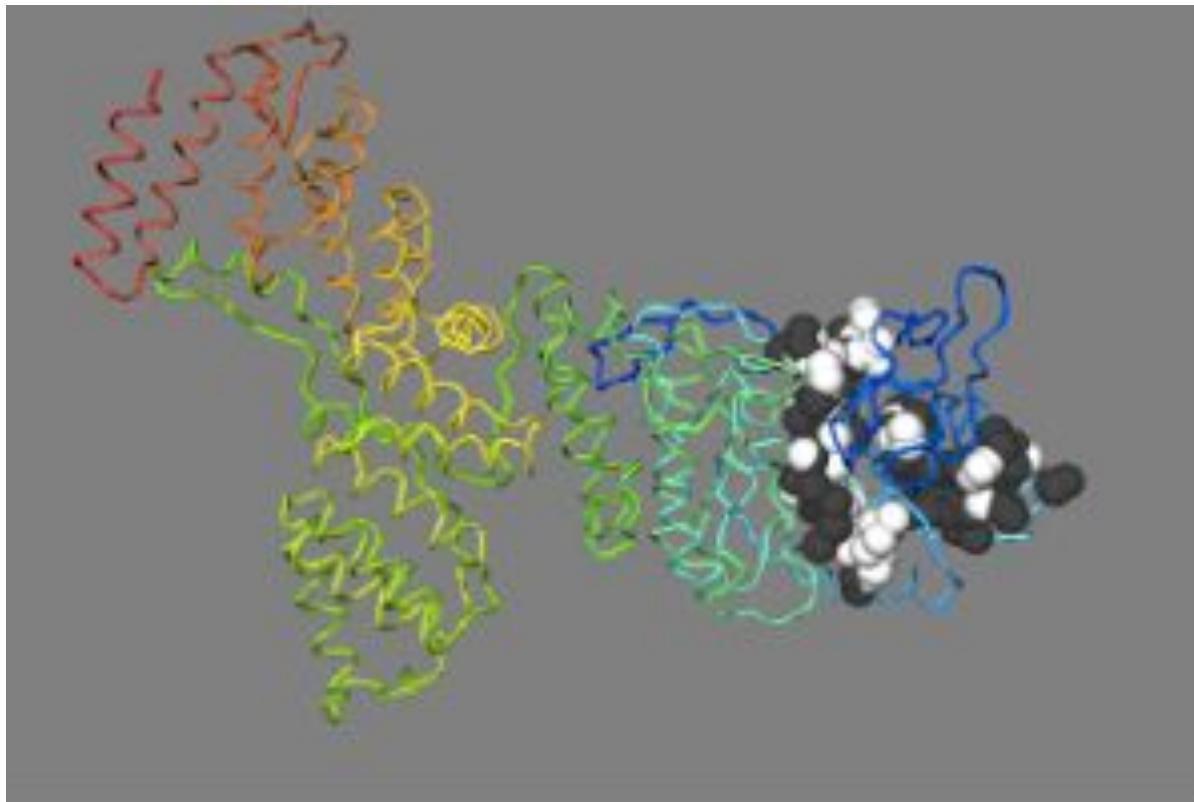
Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – go to [Help section below](#) for more information.

Del genoma a la droga



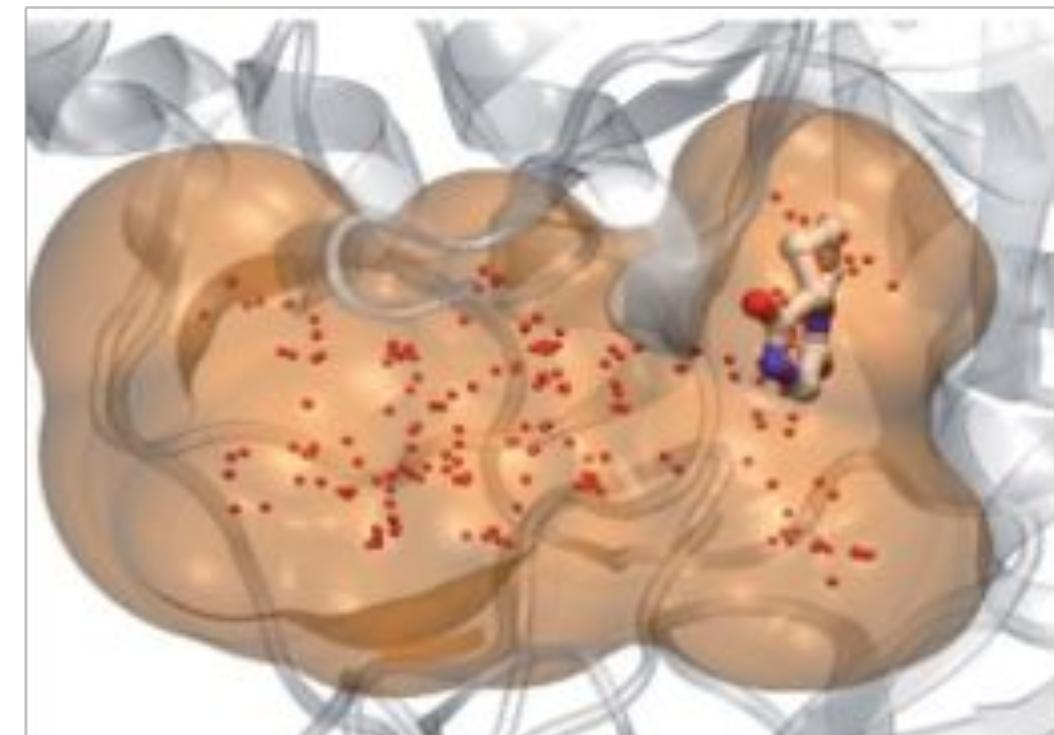
El pocket y la drogabilidad



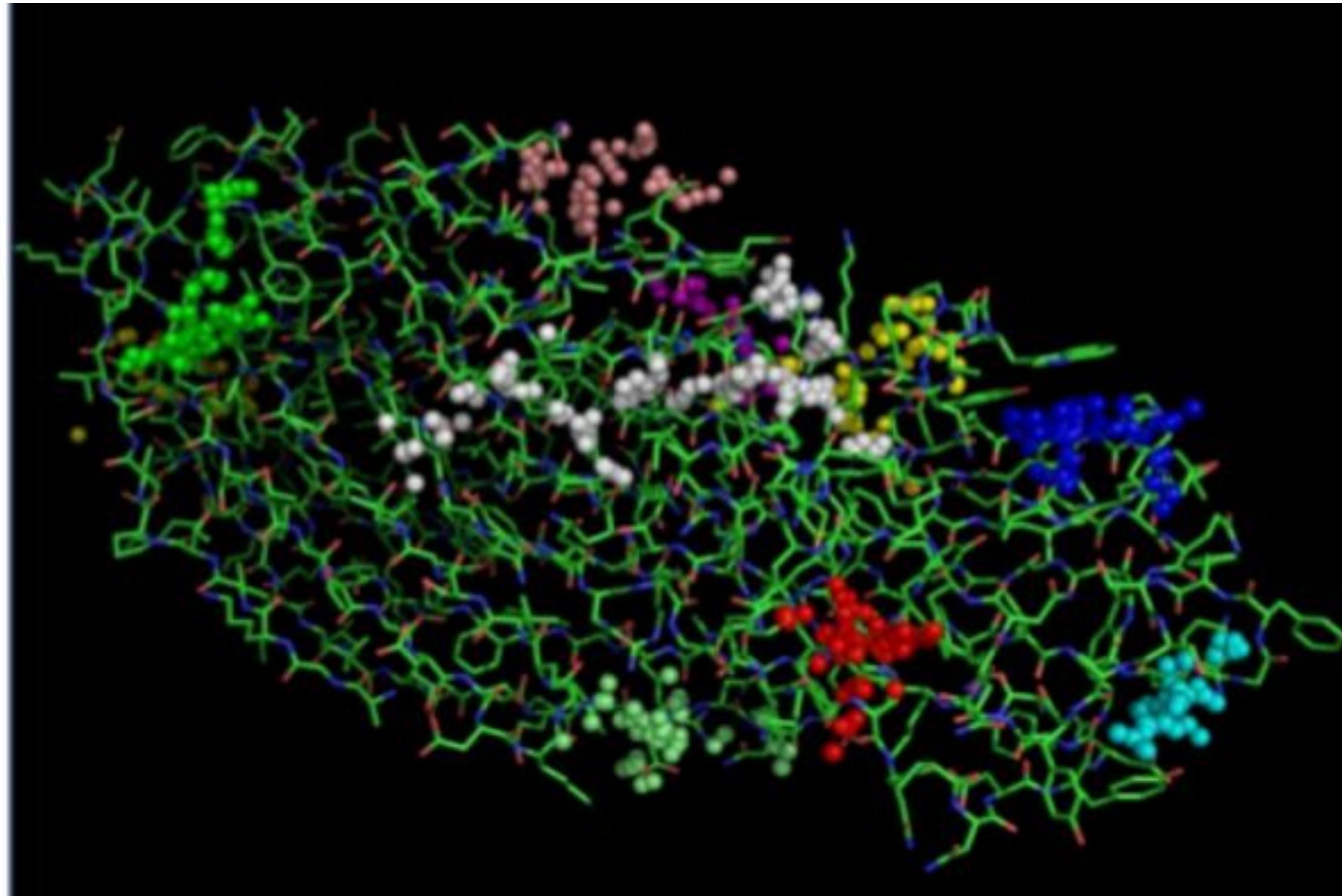
To identify a POCKET!

Fpocket: We implemented a pocket detector program
We estimated pocket properties

and Determine druggability

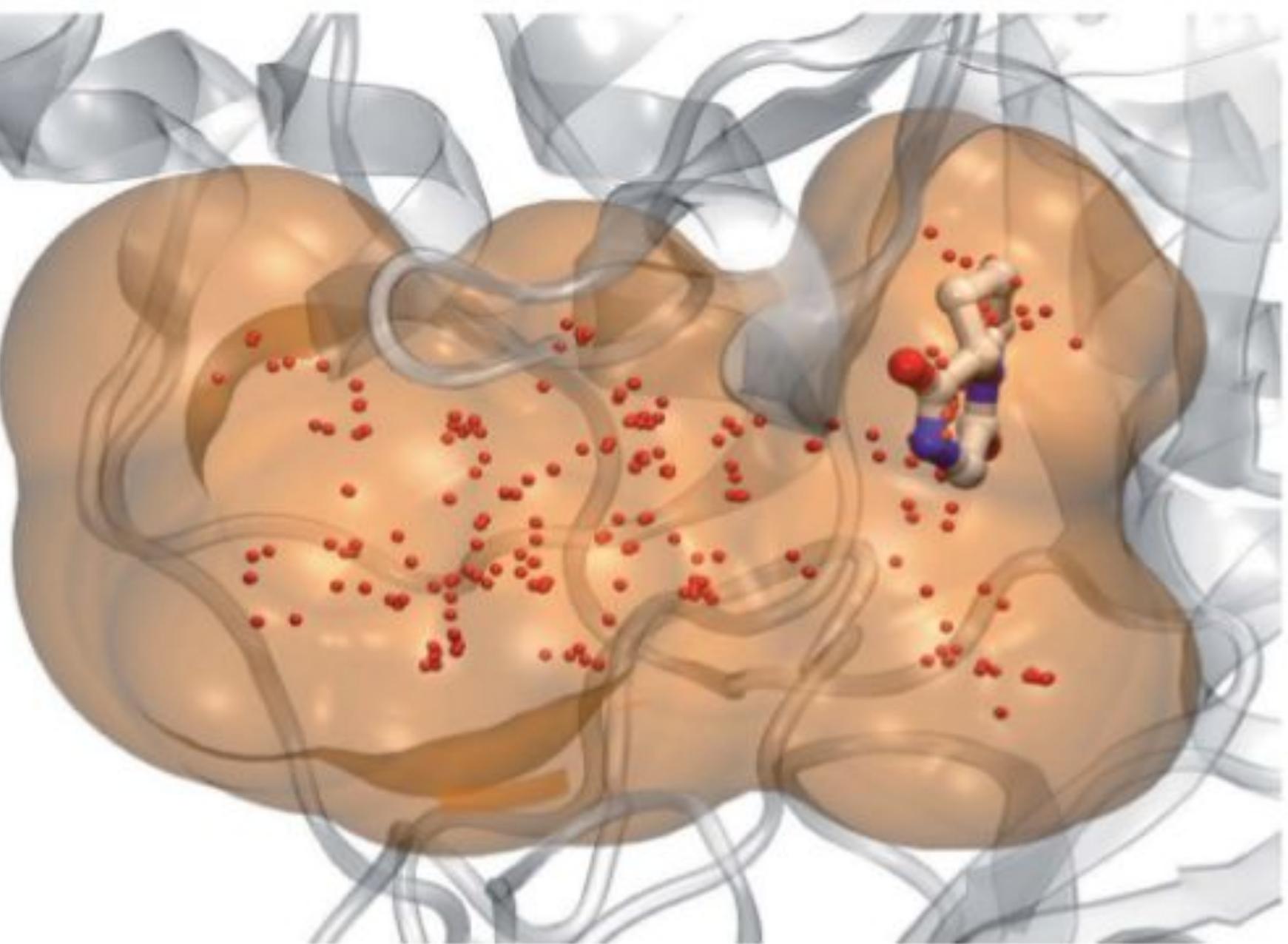


Una proteína varios pockets



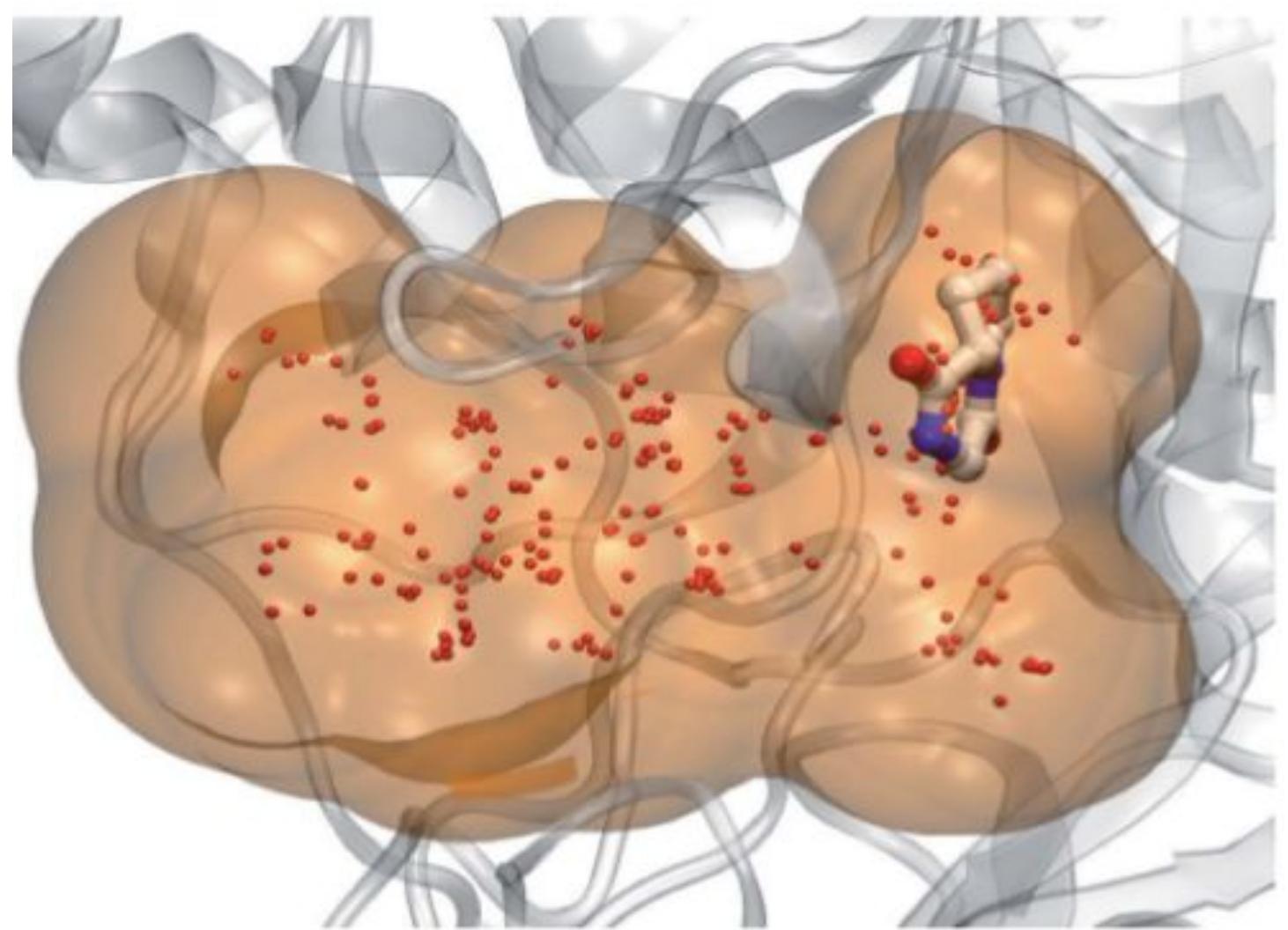
Características del pocket

- Druggability Score : 0.788
- Number of Alpha Spheres : 247
- Total SASA : 844.370
- Polar SASA : 322.358
- Apolar SASA : 522.012
- Volume : 1799.399
- Mean local hydrophobic density : 67.902
- Mean alpha sphere radius : 3.947
- Mean alp. sph. solvent access : 0.479
- Apolar alpha sphere proportion : 0.660
- Hydrophobicity score: 29.833
- Amino Acid Composition
- Distances between Aminocids

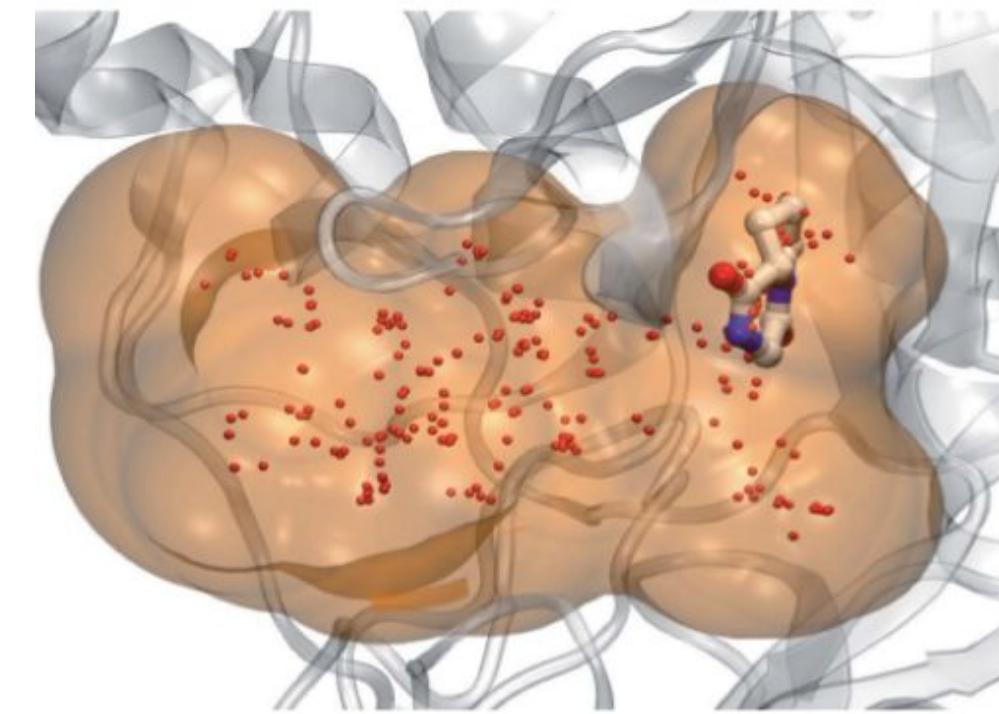
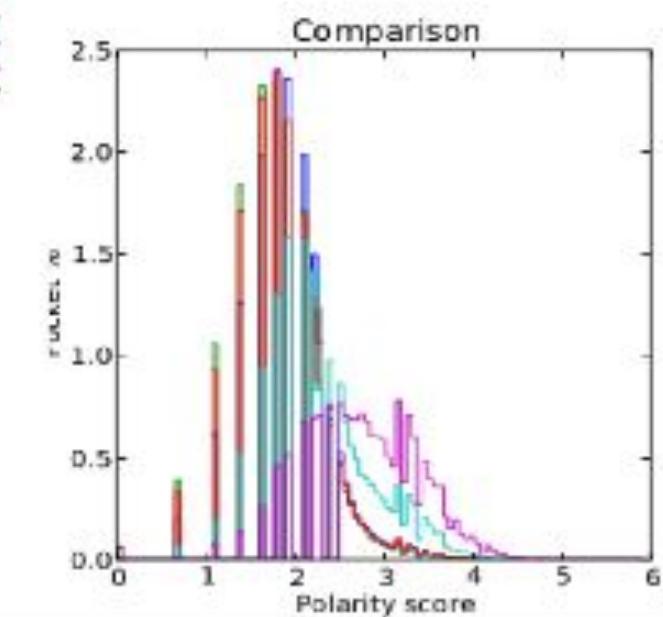
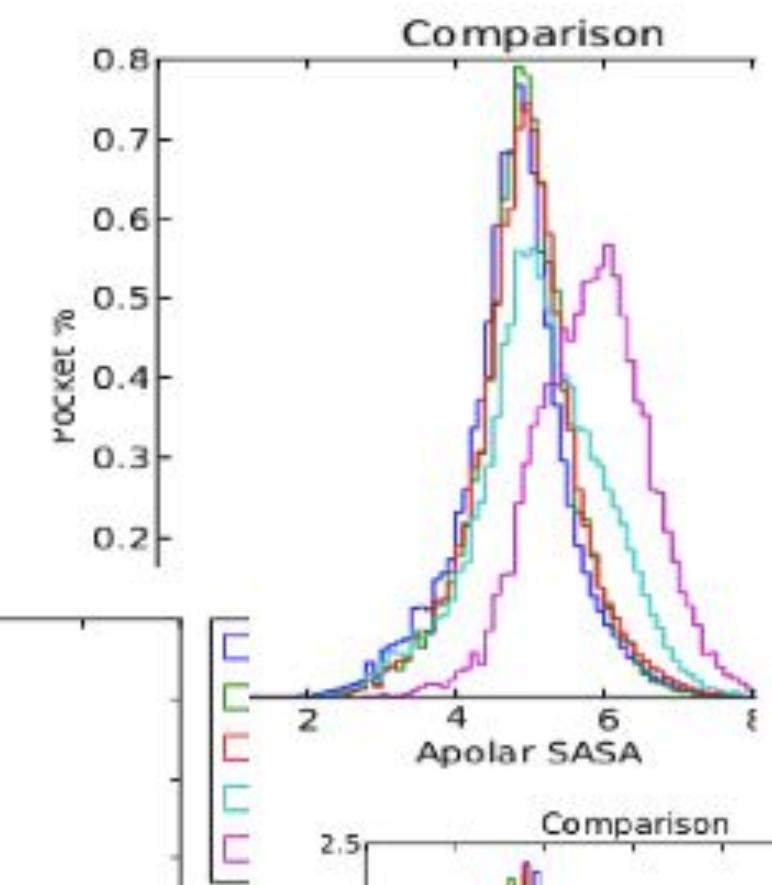
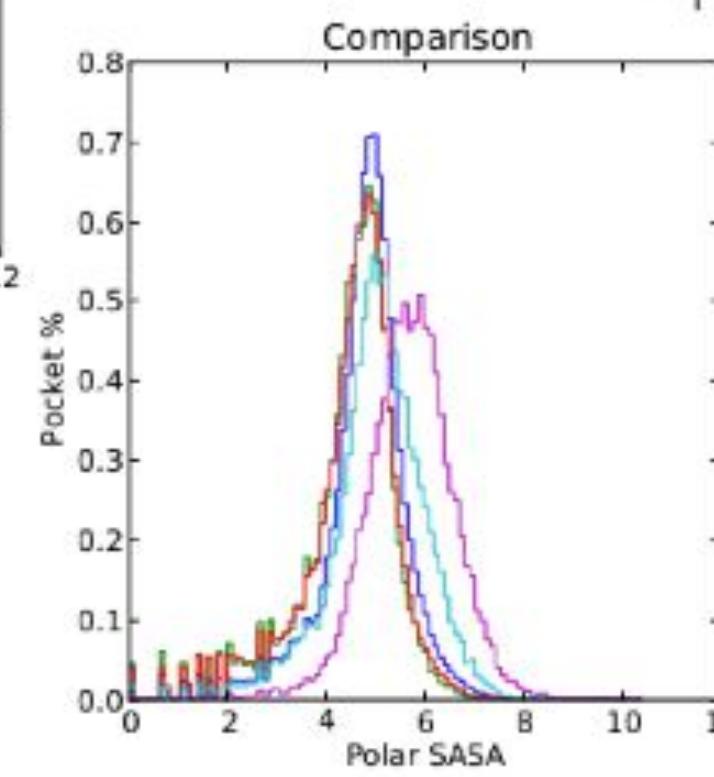
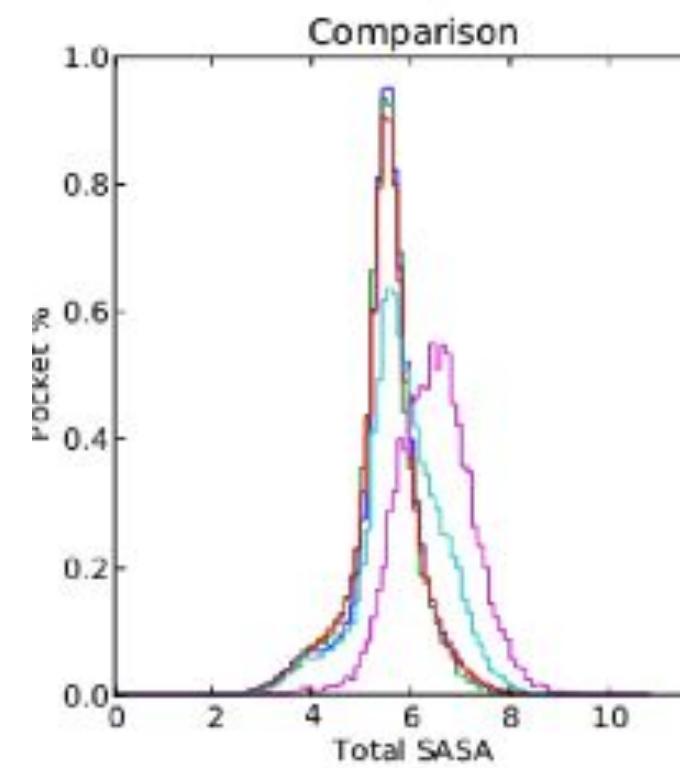


Elegir un set de pockets

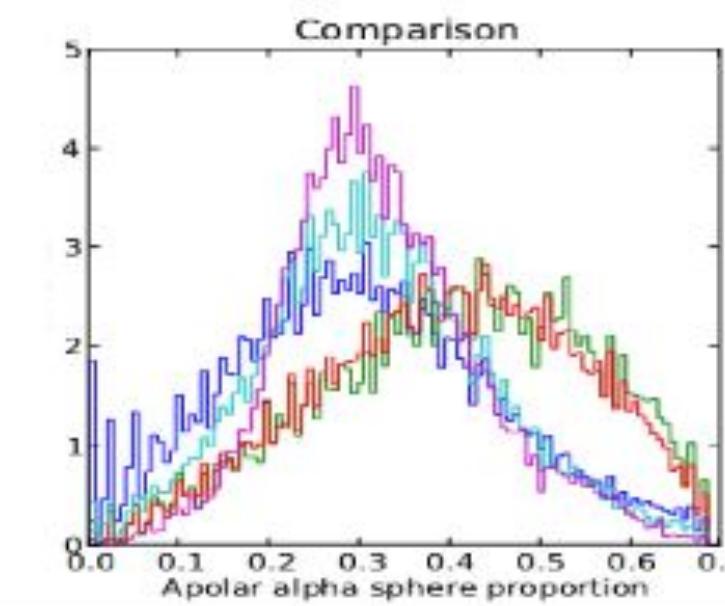
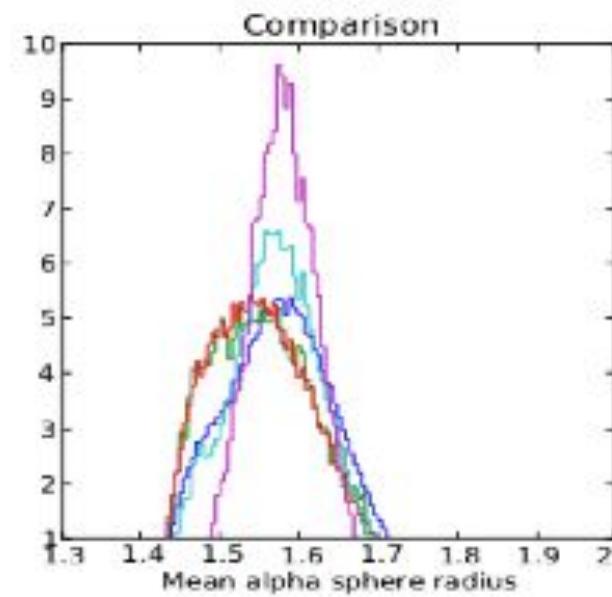
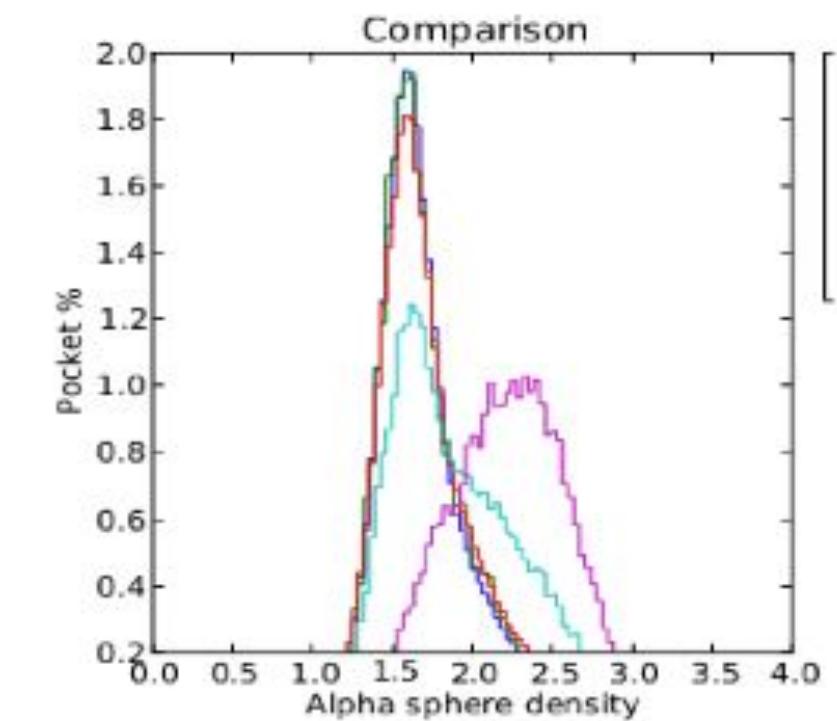
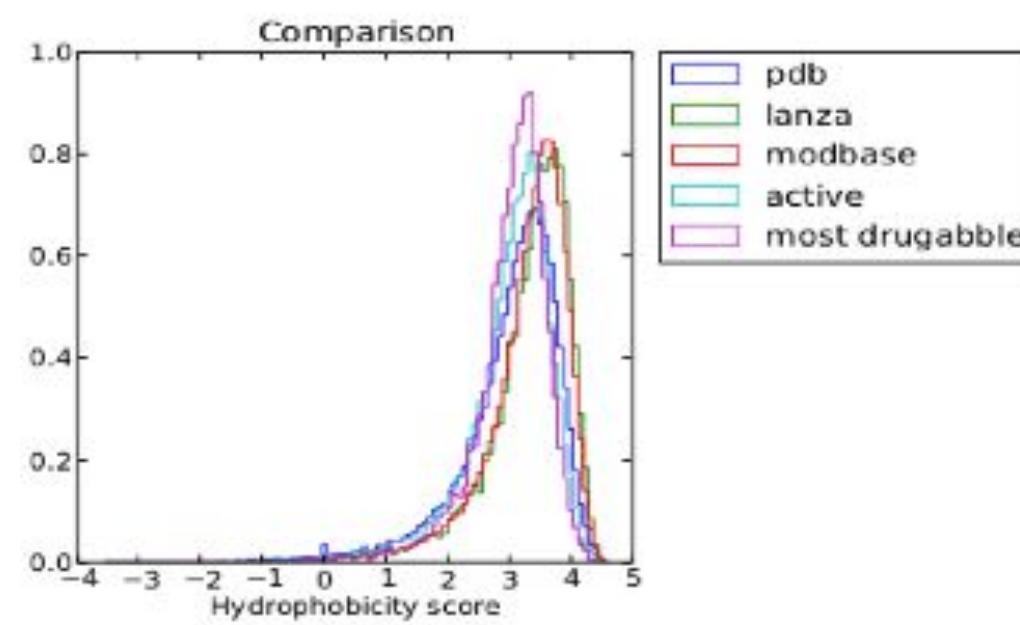
- Comparar pockets:
 - Elegir pockets que se sabe unen droga
 - Elegir pockets que no han sido reportados que unen drogas
 - Elegir pockets con características específicas:
 - De superficie
 - De proteínas no drogables
 - De modelos
 -



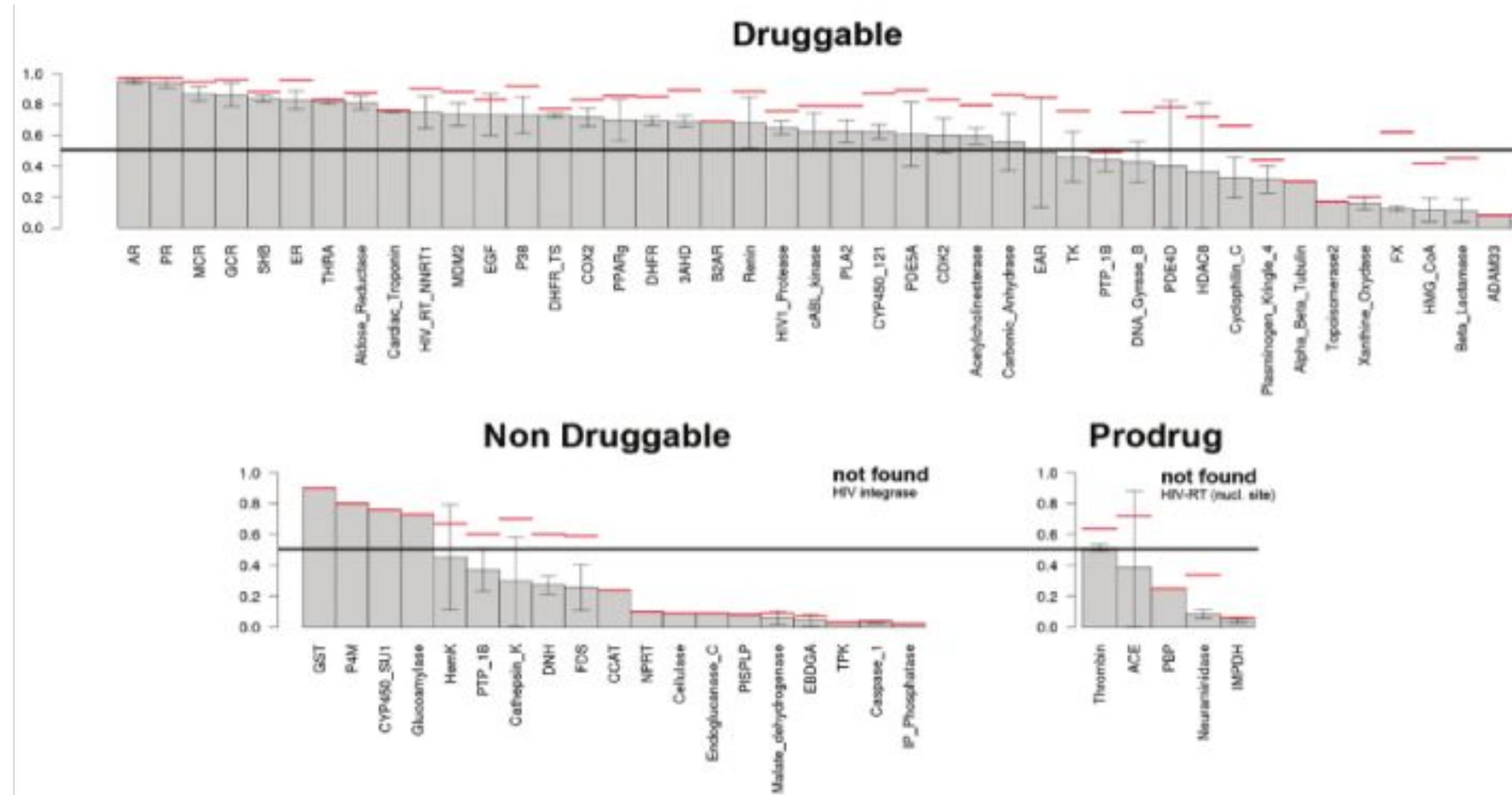
Comparar pockets



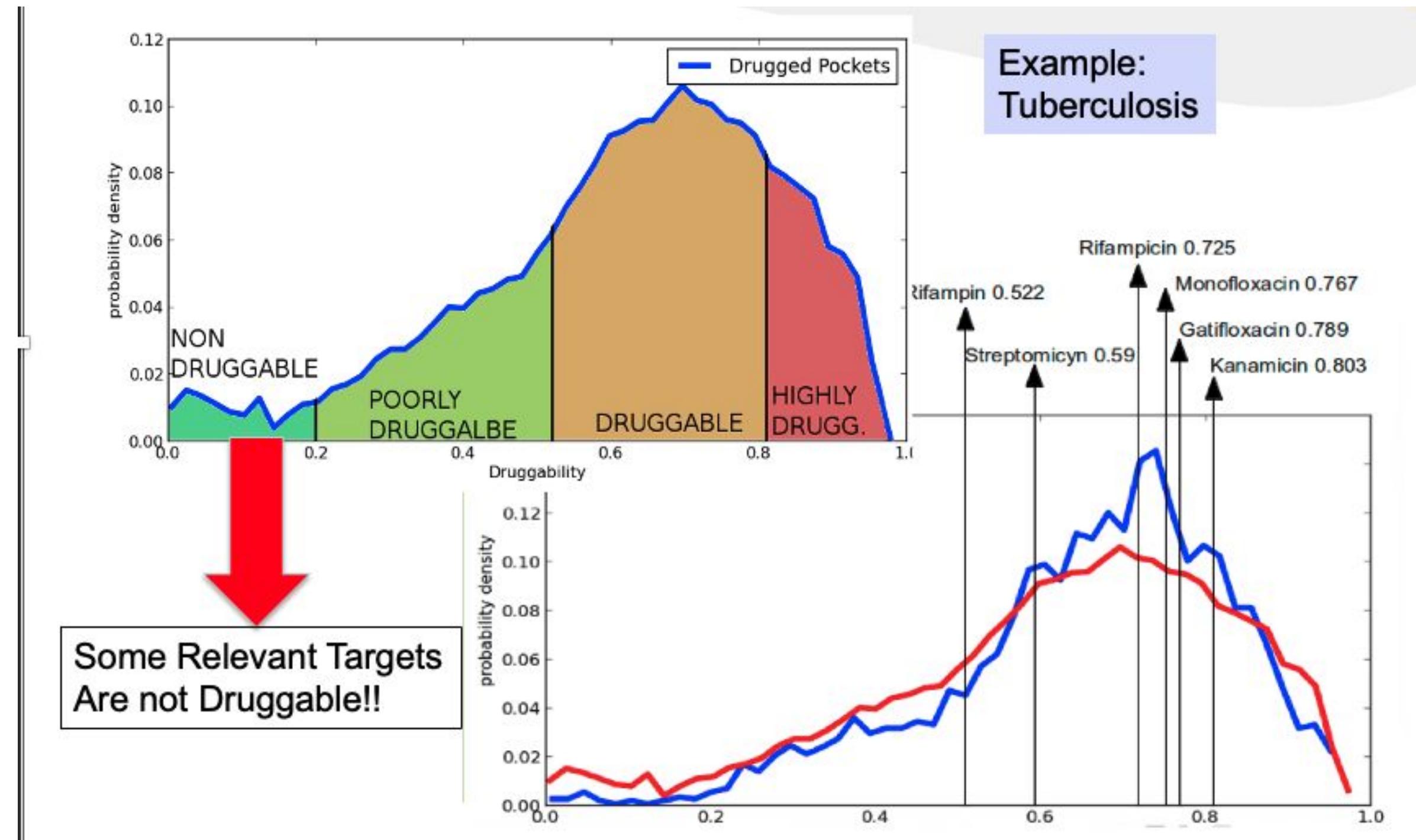
Comparar pockets



Set original fpocket: Machine Learning

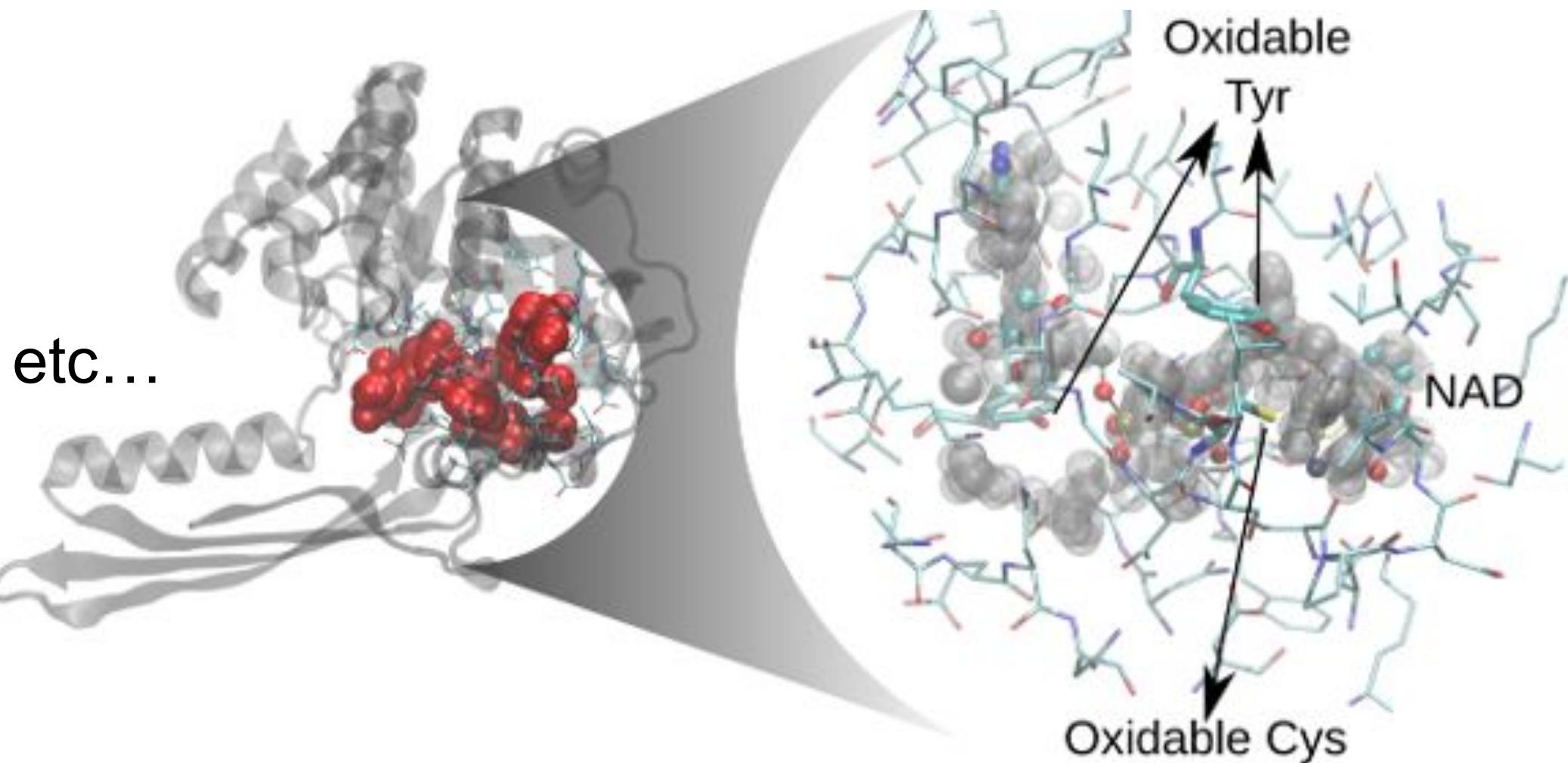


Genoma Completo



Pocket Relevante

- Catalytic Residues: CSA: Catalytic Site Atlas
- Binding Residues
- Conserved Residus
- Reactive Residues Metals, Cys, etc...



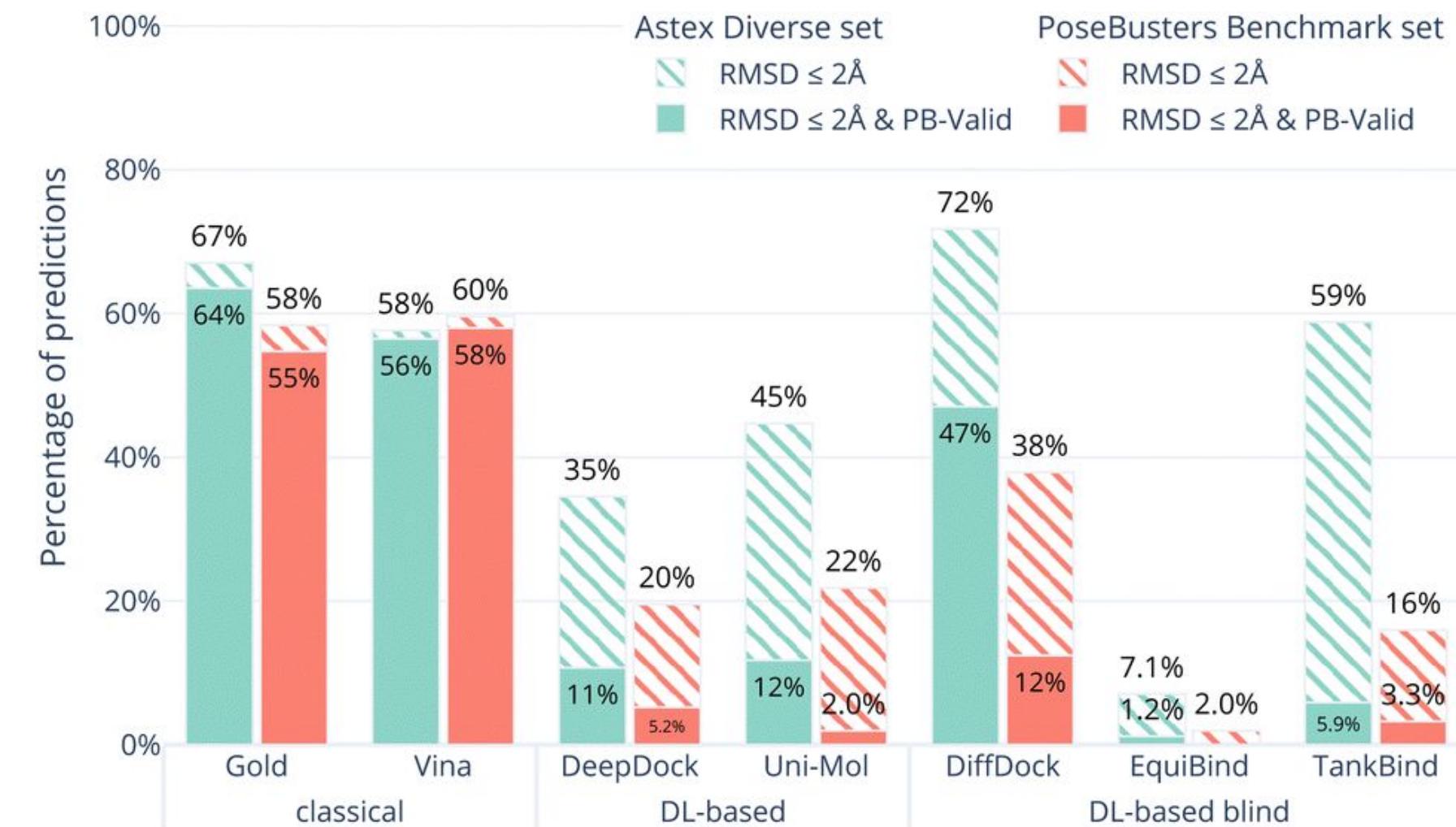
PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences^t

Martin Buttenschoen , Garrett M. Morris  and Charlotte M. Deane  *

Department of Statistics, 24-29 St Giles', Oxford OX1 3LB, UK. E-mail: deane@stats.ox.ac.uk

Received 10th August 2023, Accepted 17th November 2023

First published on 13th December 2023



AlphaFold 3 predicts the structure and interactions of all of life's molecules

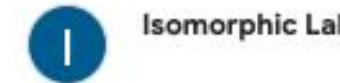
May 08, 2024

6 min read

Introducing AlphaFold 3, a new AI model developed by Google DeepMind and Isomorphic Labs. By accurately predicting the structure of proteins, DNA, RNA, ligands and more, and how they interact, we hope it will transform our understanding of the biological world and drug discovery.

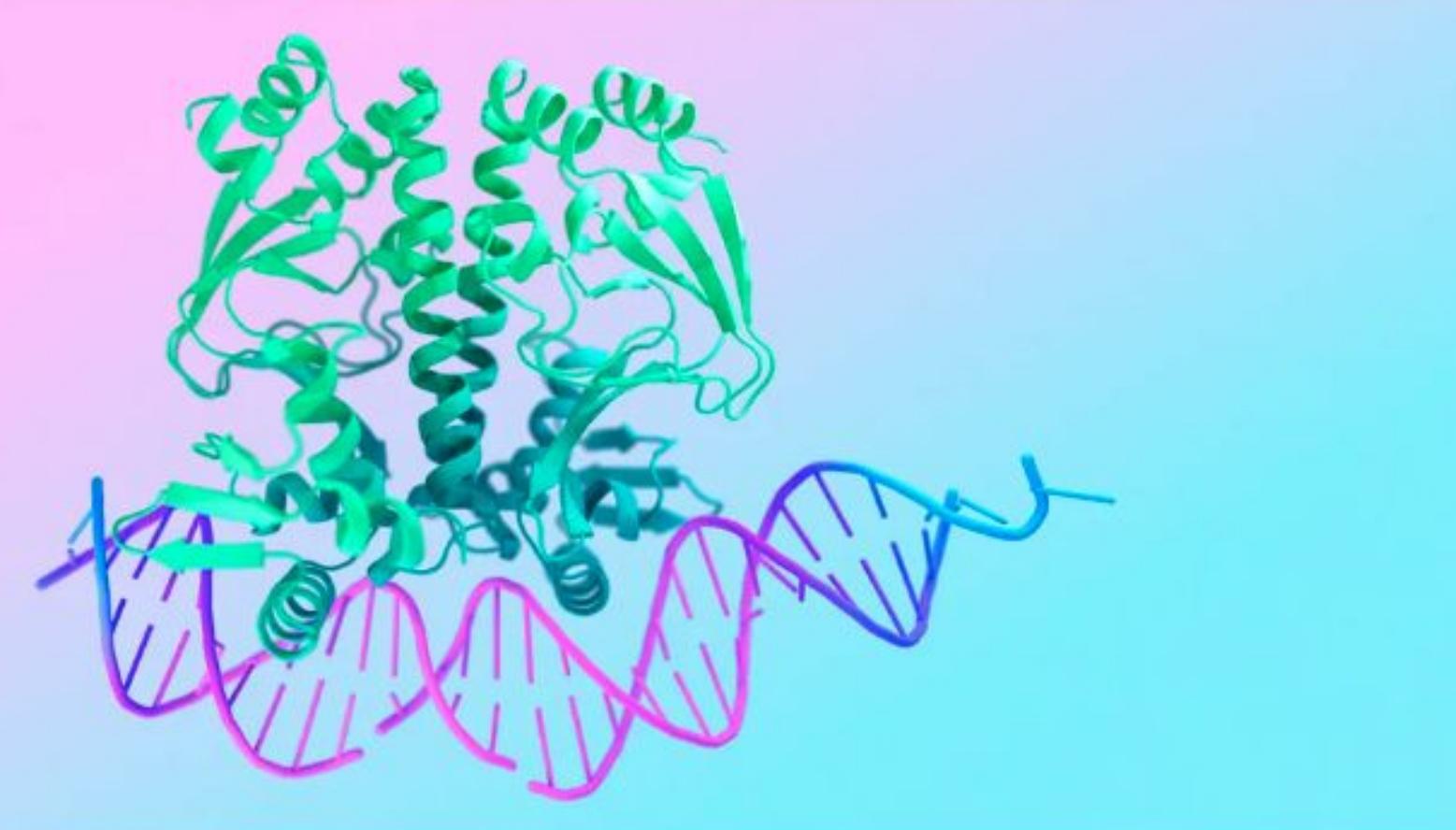


Google DeepMind
AlphaFold team



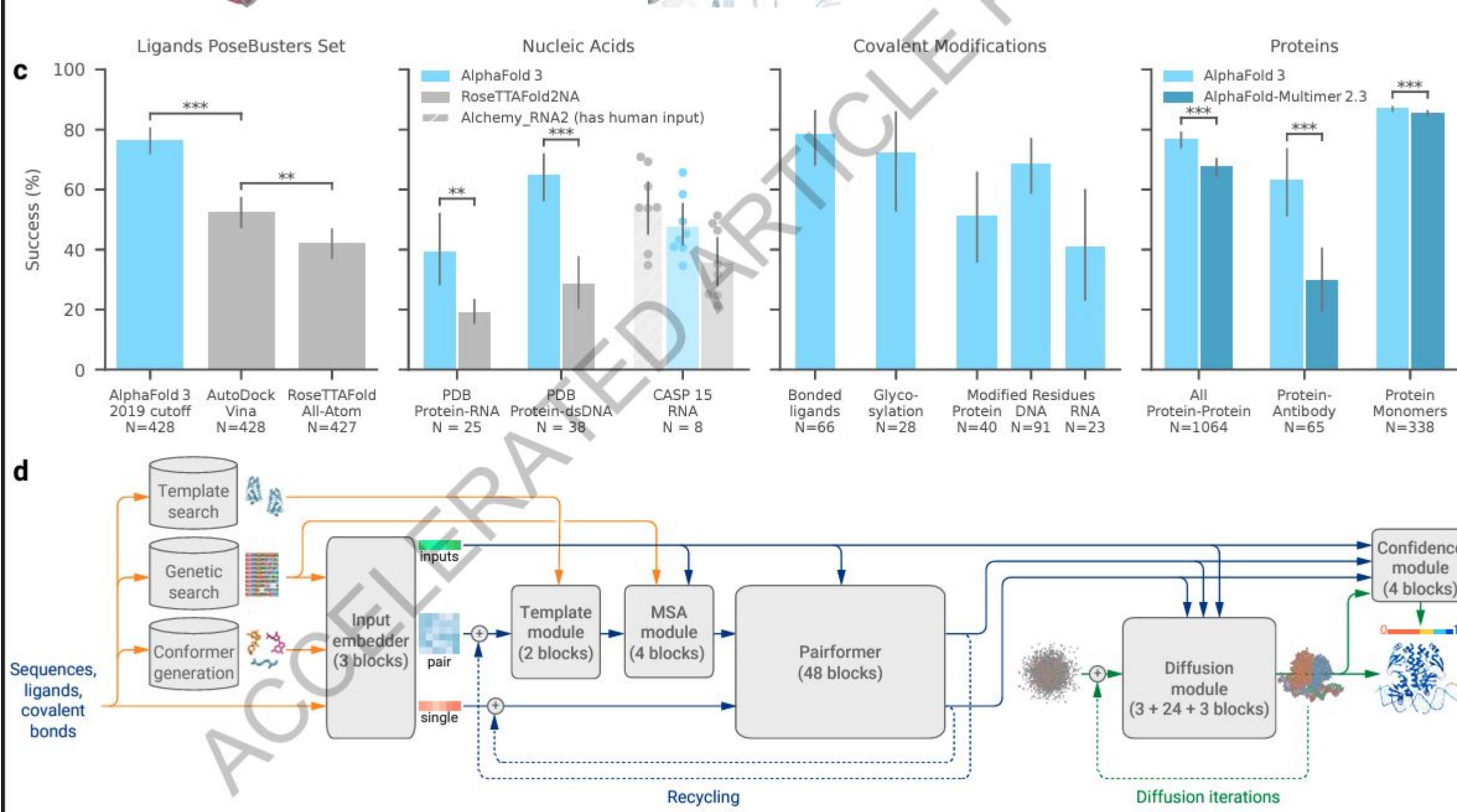
Isomorphic Labs

Share



<https://blog.google/technology/ai/google-deep-mind-isomorphic-alphafold-3-ai-model/#life-molecules>

AlphaFold3: El camino



MUCHAS GRACIAS