

Máster Universitario en Bioinformática

Proteómica y Bioinformática Estructural

Curso académico 2024-2025



Universidad
Internacional
de Valencia

Dra. Magdalena Nikolaeva Koleva

magdalena.nikolaeva@professor.universidadviu.com

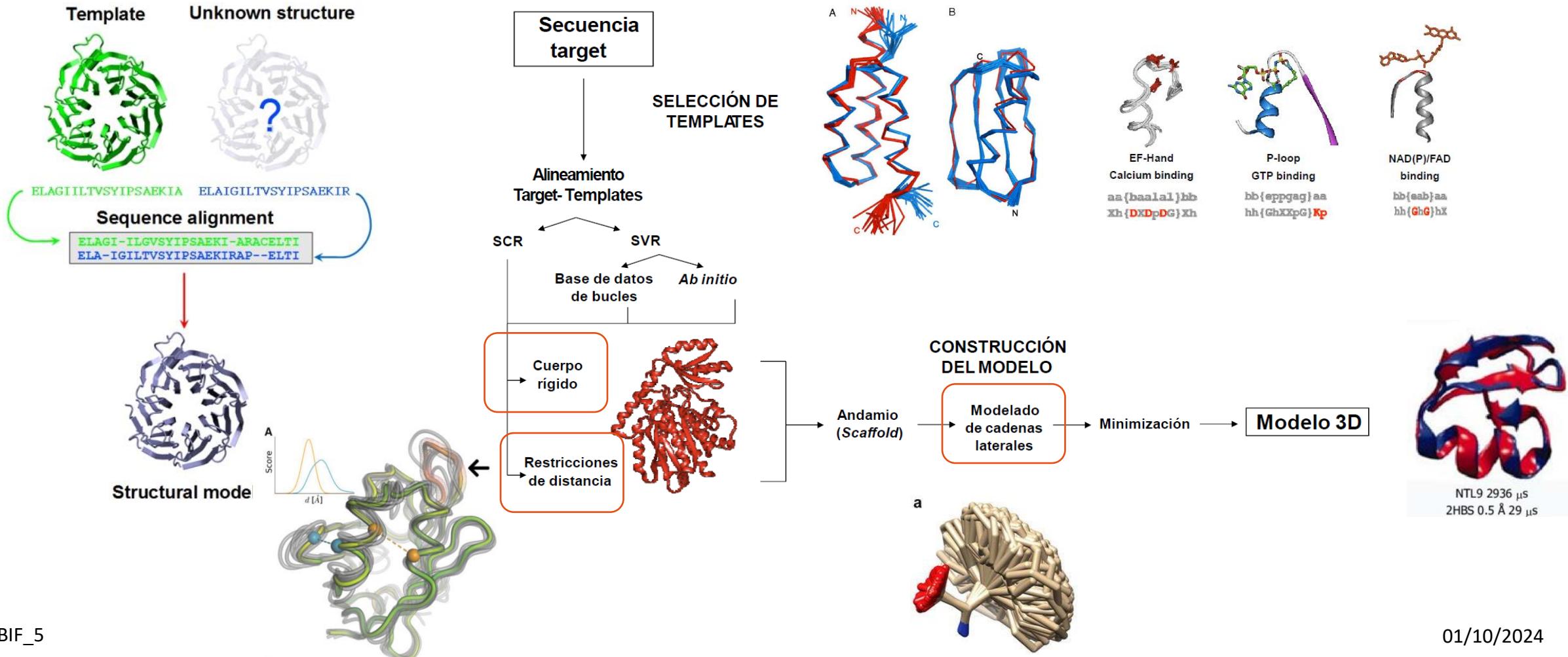
Sesión 5

Modelado estructural de proteínas (II):
Modelado por reconocimiento del plegamiento y *ab initio*



Universidad
Internacional
de Valencia

¿Qué vimos en la sesión anterior?



Temario - Contenidos

Tema 4. Modelización estructural de proteínas

- 4.1. Modelización por homología
- 4.2. Modelización por reconocimiento de plegamiento
- 4.3. Modelización *ab initio*

Temario - Contenidos

Tema 4. Modelización estructural de proteínas

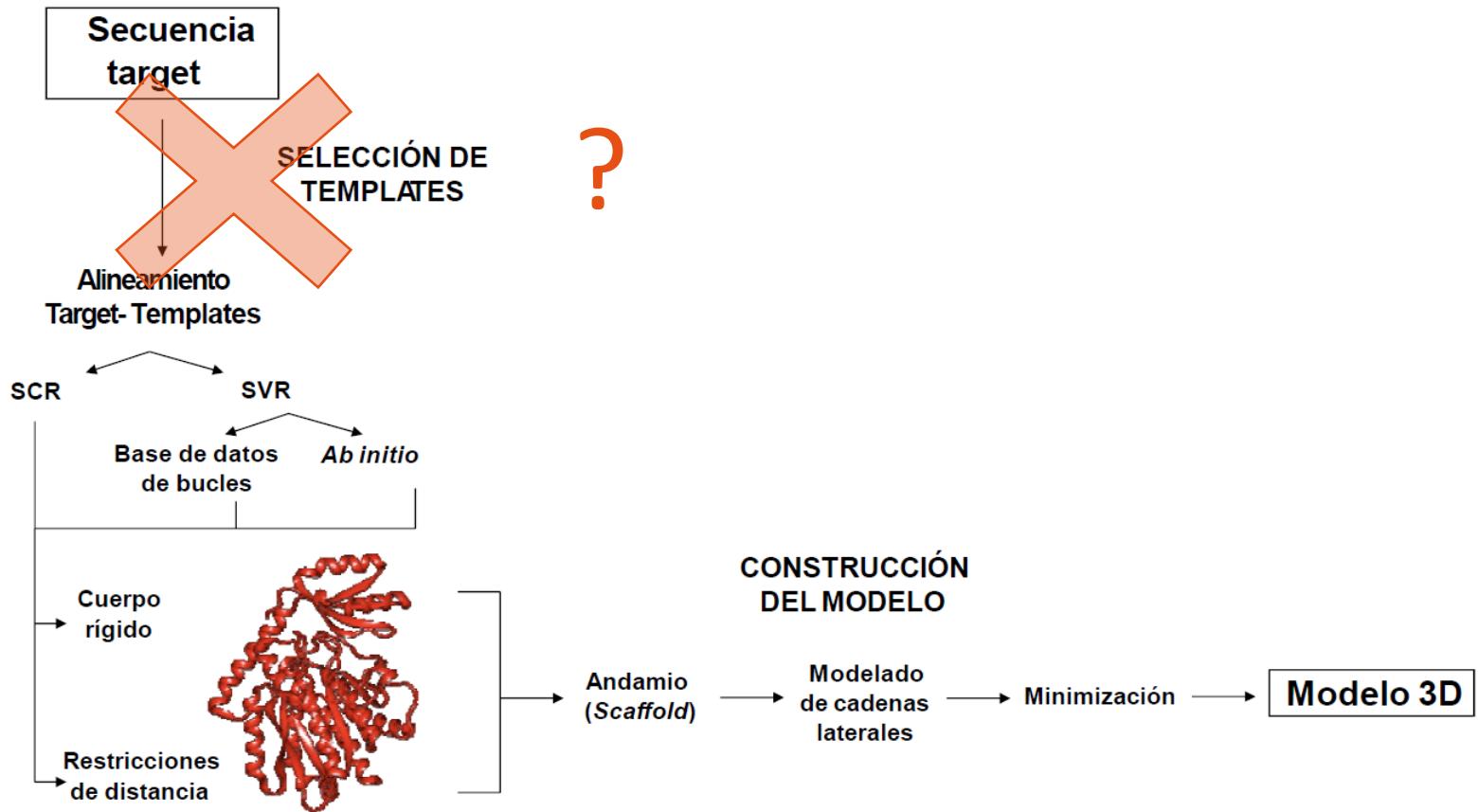
4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización *ab initio*

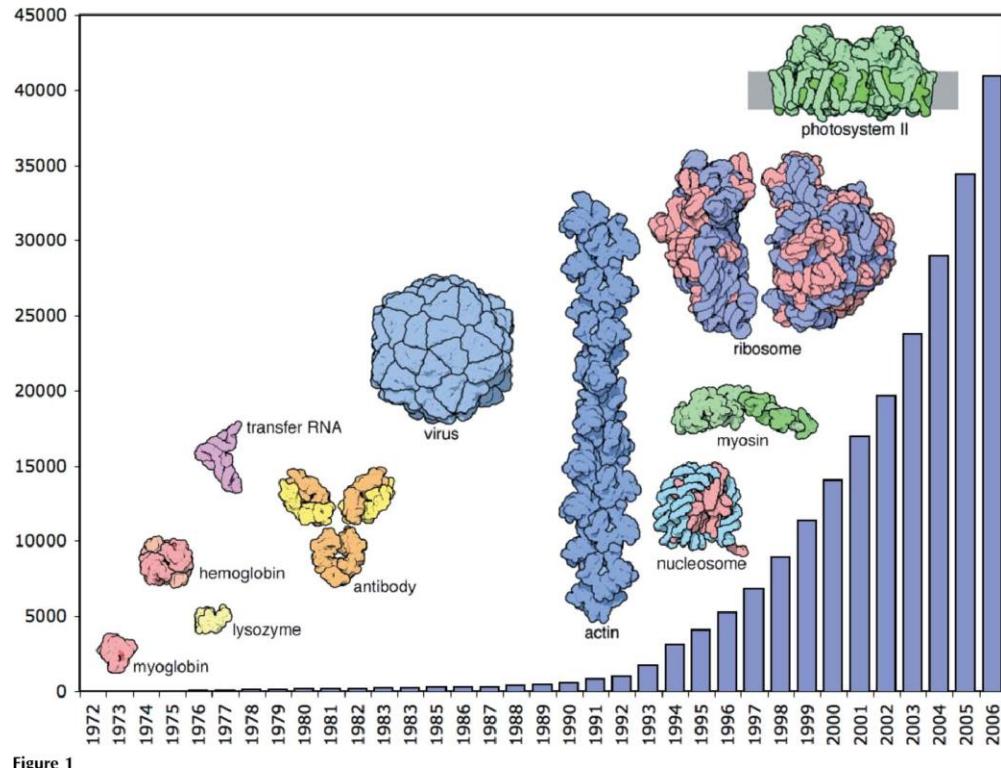
- Reconocimiento del plegamiento: aspectos básicos
- Potenciales estadísticos
- Reconocimiento de homología remota
- Predicción de estructura secundaria
- Modelización *ab initio*: principios físicos
- Modelización *ab initio*: métodos empíricos

4.2. Modelización por reconocimiento del plegamiento



4.2. Modelización por reconocimiento del plegamiento

- Publicación de nuevas estructuras vs nuevos *folds*.
- Número de plegamientos diferentes es limitado.
- A pesar de diferencias en secuencia, mismo *fold*.

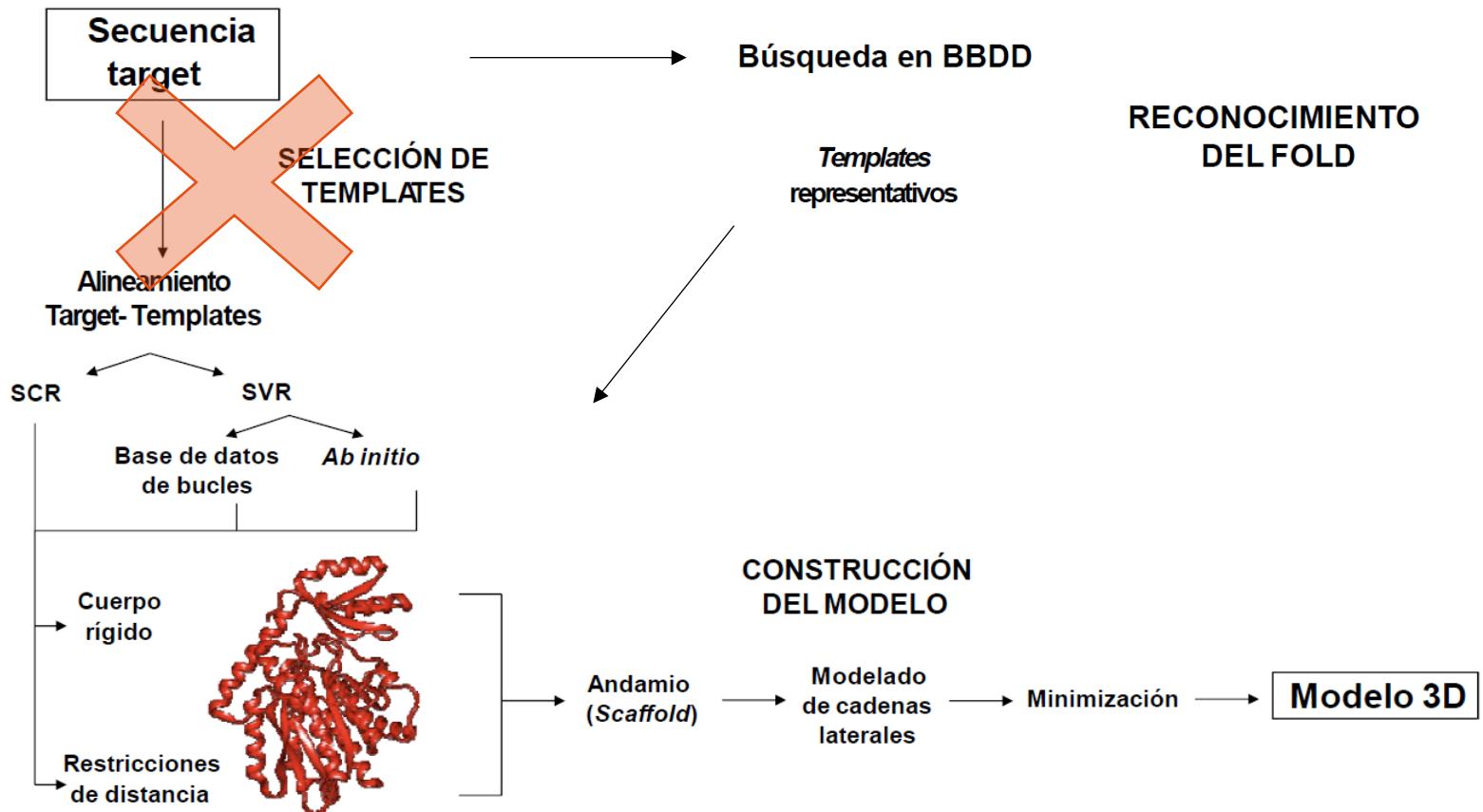


The screenshot shows the CATH-Gene3D Hierarchy browser. At the top, there is a navigation bar with links for Home, Search, Browse, Download, About, and Support, along with a search bar. Below the navigation bar, the title "Browse CATH-Gene3D Hierarchy" is displayed. There are two visualization options: "Tree" and "Sunburst". The "Tree" option is selected. On the left, there is a sidebar with links for "Browse Hierarchy", "Highly Diverse Superfamilies", and "Superfamily Comparison". Below the sidebar, there is a text input field labeled "Select a CATH node...". The main content area displays a bar chart titled "Top of CATH Hierarchy (4 Classes)" with the text "~1500 folds" in red. The chart shows the following data:

Clase	Arquitecturas	Plegamientos	SUPERFAMILIAS	DOMENOS
1 Mainly Alpha	5	404	2033	103788
2 Mainly Beta	21	244	1250	124032
3 Alpha Beta	14	634	2337	262275
4 Few Secondary Structures	1	108	181	5716
6 Special	2	82	790	4427

A red box highlights the "Special" class entry in the list.

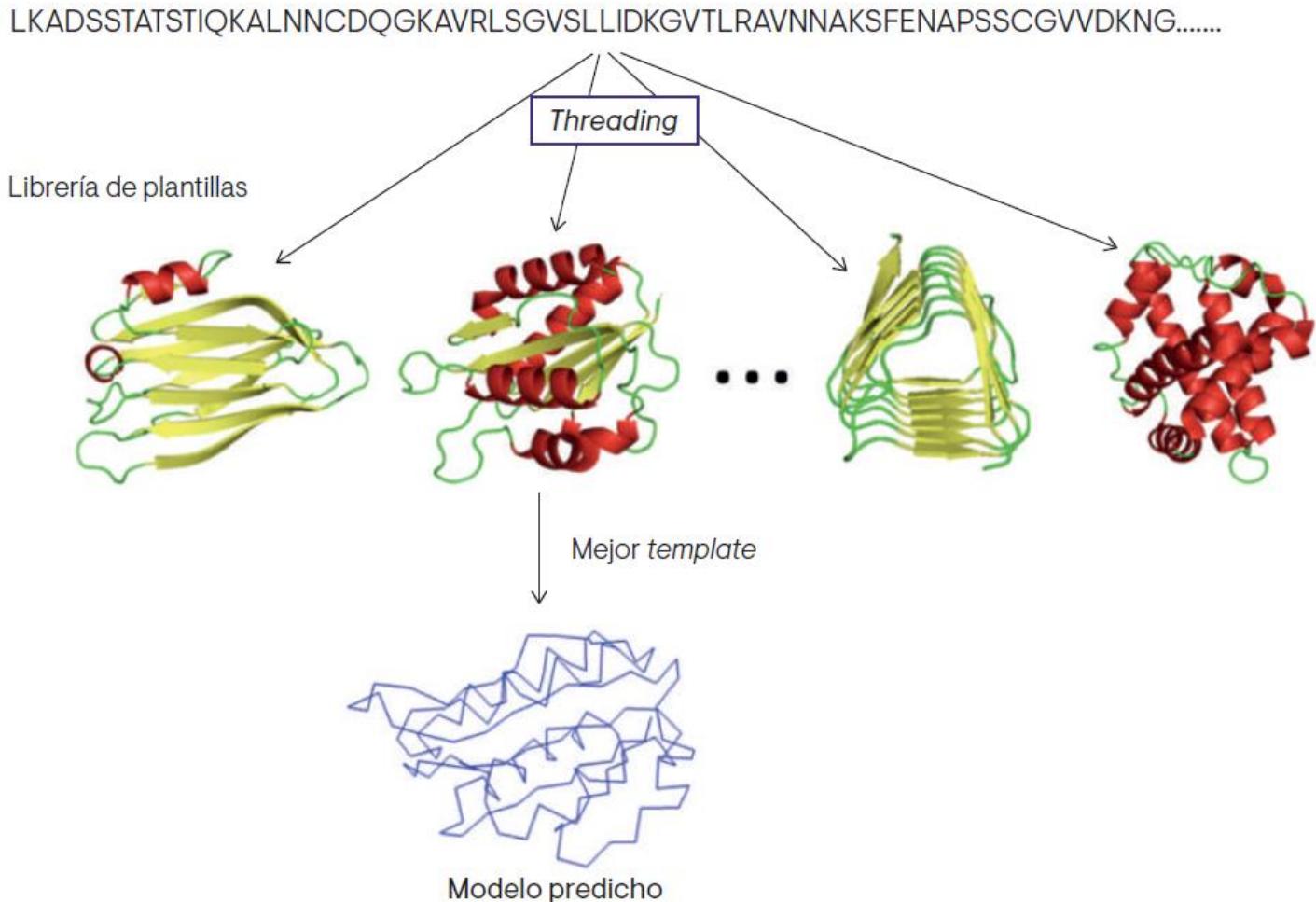
4.2. Modelización por reconocimiento del plegamiento



4.2. Modelización por reconocimiento del plegamiento

Threading

- Generación de distintos modelos mediante el **desplazamiento en la secuencia** de la proteína diana.
- Conocer la **posición óptima** de los residuos dentro de la secuencia para un determinado plegamiento.
- Evaluación de los modelos mediante **funciones de evaluación**.



Temario - Contenidos

Tema 4. Modelización estructural de proteínas

4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización *ab initio*

- Reconocimiento del plegamiento: aspectos básicos
- **Potenciales estadísticos**
- Reconocimiento de homología remota
- Predicción de estructura secundaria
- Modelización *ab initio*: principios físicos
- Modelización *ab initio*: métodos empíricos

4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

Probabilidad de pertenecer a una población
cuando se encuentra en equilibrio

De acuerdo a la ley de Boltzmann:

$$P(x) = \frac{1}{Z} e^{-E(x)/k_B T}$$

Energía

Temperatura

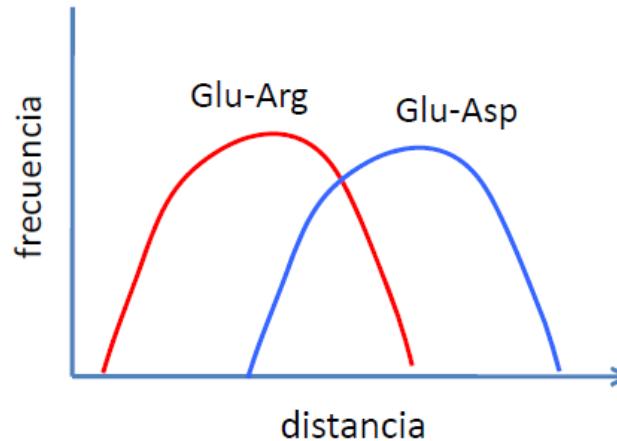
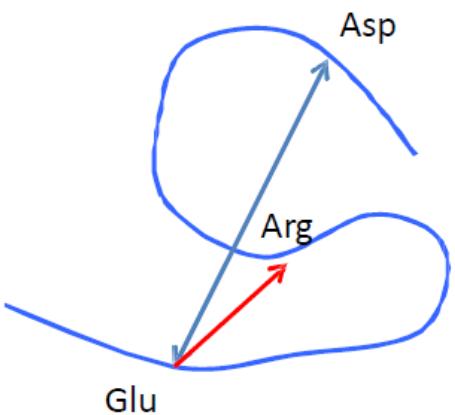
Es decir, la energía está relacionada con la probabilidad:
(estabilidad)

$$P(\text{Glu-Asp}, d=10\text{\AA}) \rightarrow E(\text{Glu-Asp}, d=10\text{\AA})$$

4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

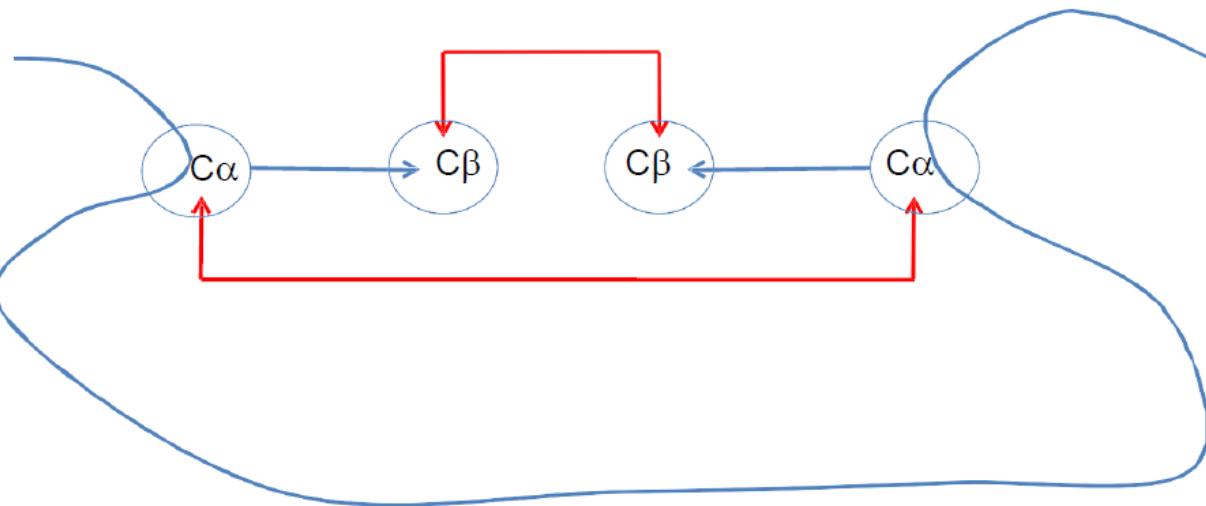
- Inferir la **energía de interacción** entre residuos (r_i, r_{ii}) a una distancia (d) según frecuencia con la que aparecen a esa (d) en estructuras proteicas conocidas.
- Interacción favorable = pares (r_i y r_{ii}) a (d) corta.
- Los potenciales se calculan a nivel de **residuo**.



4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

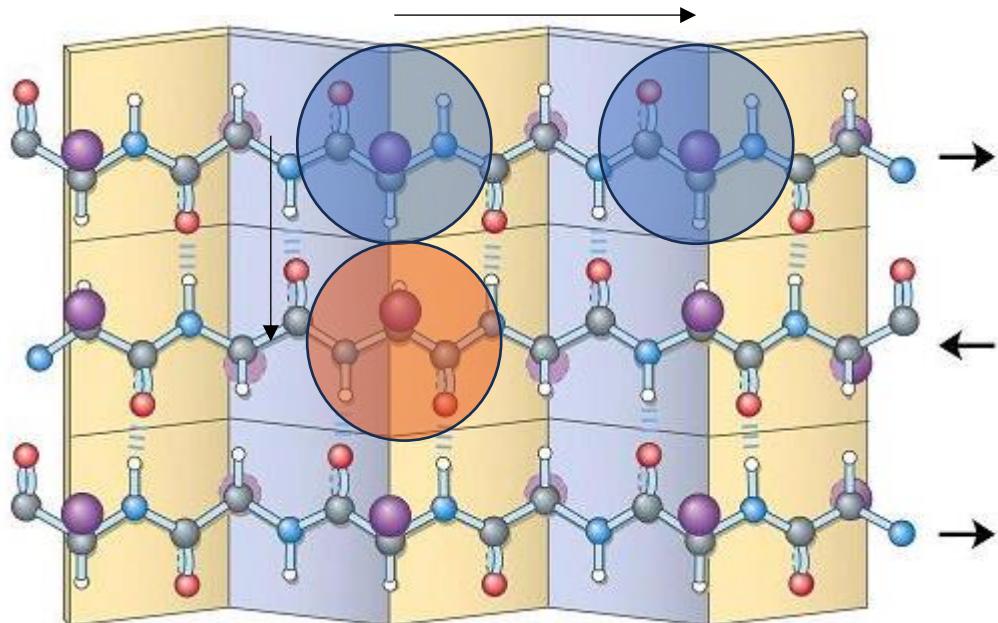
- Las distancias se calculan a nivel de átomo.
- Dependiendo de la orientación de la cadena lateral, puede dar lugar a diferencias.
- Usar $C\beta$ porque indica la dirección de la cadena lateral.
- Otros: distancia promedio, mínima, entre $C\alpha$.



4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

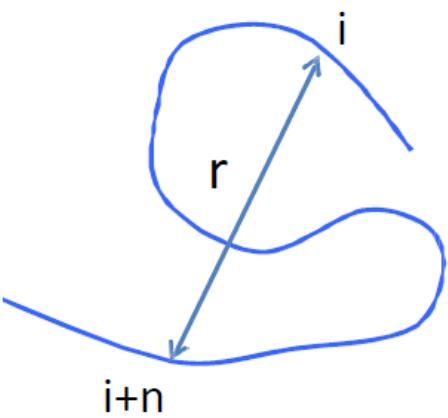
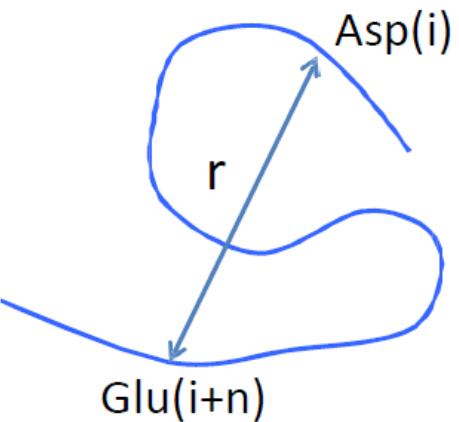
- Base de datos usada para extraer las distancias:
 - Evitar estructuras redundantes – homólogos, miembros de la misma familia/superfamilia.
 - Evitar tendencias parciales (*bias*) – elegir estructuras con similitud de secuencia <40%.
- Frecuencias de pares de residuos deben tener en cuenta si distancia de los residuos en la secuencia.
 - Normalizar según distancia en secuencia.



4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

- Entorno de los residuos. Dependencia de solvente, número de residuos, estructura proteica alrededor del residuo.
 - Definir estado de referencia normalizado.
 - Usar conjunto entero de pares de residuos y calcular energías relativas en vez de absolutas.



$$\Delta E(r/(Glu, Asp, C\beta, C\beta, n)) = -kT \ln \left(\frac{N(r/ED, n)}{N(r/n)} \right)$$

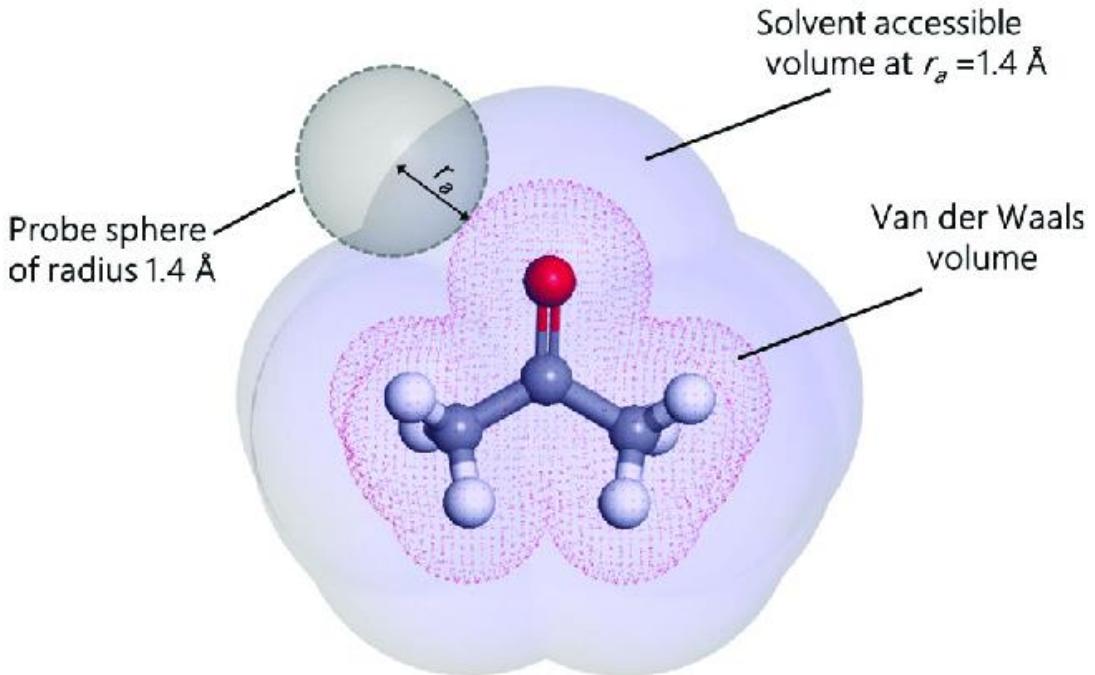
Número de pares *ED* con distancia *r* entre sus átomos $C\beta$ y a distancia *n* de secuencia

Número total de pares de residuos a distancia *n* de secuencia y distancia *r* entre sus átomos $C\beta$

4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

- Solvatación: tendencia de un residuo de interaccionar con el solvente
 - En potenciales estadísticos las moléculas de agua no son explícitas, **aproximaciones**.
 1. **ASA (Accessible Surface Area)** - la energía de solvatación de un residuo puede ser calculada en función de su **área accesible al solvente**, corregida por un **factor de tendencia intrínseca** de ese tipo de residuo a ser solvatado.
 2. Frecuencia de un tipo de residuo de aparecer en las **superficies proteicas**.



4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

- ¿Cómo calcular la energía total del modelo para su comparación con otros?

$$E = \sum_j E_i + \beta \sum_j E_{solv}(i)$$

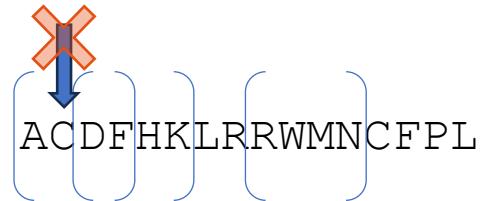
Factor de combinación

$$E_i = \sum_{j \neq i} E_{ij}(r, n = |j - i|)$$
$$E_{ij}(r, n = |j - i|) = \Delta E \left(\frac{r}{Glu(j), Asp(i), C\beta, C\beta, n} \right)$$
$$E_{solv}(i) = \sigma_{Asp} ASA(i)$$

4.2. Modelización por reconocimiento del plegamiento

Potenciales estadísticos

- ¿Cómo calcular la energía total del modelo para su comparación con otros?



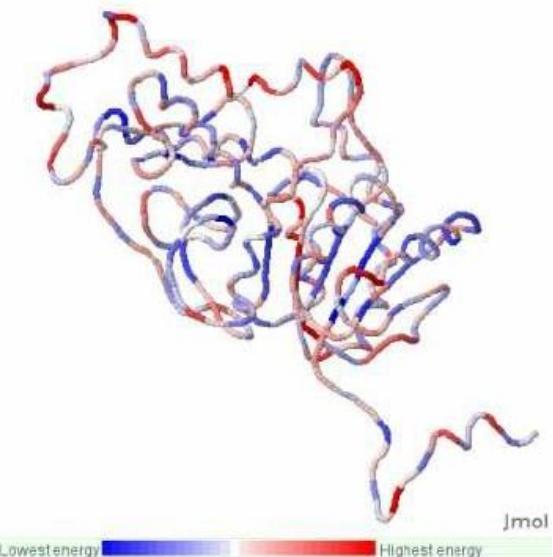
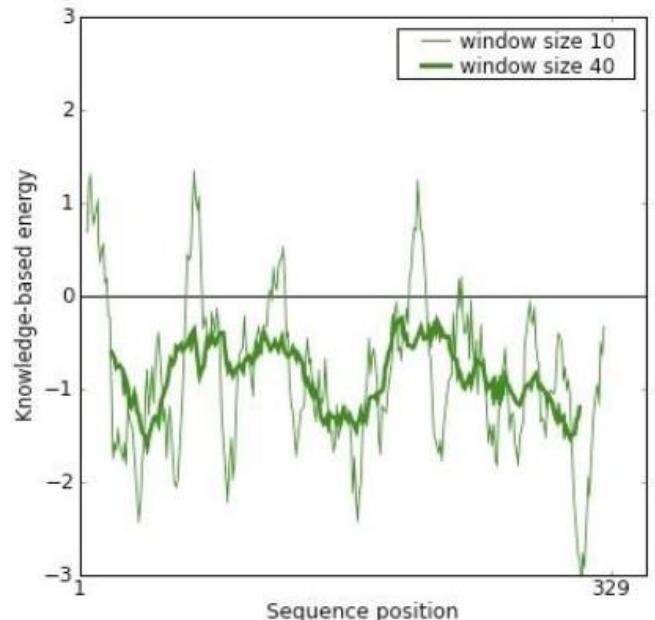
- Perfiles energéticos:
 - Evaluar el modelo
 - Adecuación del *fold*
 - Detectar zonas mal modeladas

○ **Z-score:** valor de energía normalizado con respecto a la distribución normal de valores de energía de folds conocidos.

- ProSa

<https://prosa.services.came.sbg.ac.at/prosa.php>

Identificar errores y zonas de alta energía



Temario - Contenidos

Tema 4. Modelización estructural de proteínas

4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización *ab initio*

- Reconocimiento del plegamiento: aspectos básicos
- Potenciales estadísticos
- **Reconocimiento de homología remota**
- Predicción de estructura secundaria
- Modelización *ab initio*: principios físicos
- Modelización *ab initio*: métodos empíricos

4.2. Modelización por reconocimiento del plegamiento

Reconocimiento de homología remota

- Homólogos con relación evolutiva
- Similitud de secuencia muy baja
- PSI-BLAST (*Position Specific Iterated BLAST*)
 - Usar PSSM
 - Refinar con función proteica (PHYRE)
 - Asociaciones funcionales (interacción proteínas).

Búsqueda de **palabras clave funcionales** de homólogos en UniProt. Para secuencias totalmente nuevas (sin información) se extraen **patrones** de proteínas **interaccionantes** para seleccionar los homólogos con estructura más idóneos (ModLink).

R06098	T	C	A	C	A	C	G	T	G	G	G	A
R06099	G	G	C	C	A	C	G	T	G	C	A	G
R06100	T	G	A	C	A	C	G	T	G	G	G	T
R06102	C	A	G	C	A	C	G	T	G	G	G	G
R06103	T	T	C	C	A	C	G	T	G	C	G	A
R06104	A	C	G	C	A	C	G	T	T	G	G	T
R06097	C	A	G	C	A	C	G	T	T	T	T	C
R06101	T	A	C	C	A	C	G	T	T	T	T	C

Count matrix

Residue\position	1	2	3	4	5	6	7	8	9	10	11	12
A	1	3	2	0	8	0	0	0	0	0	1	2
C	2	2	3	8	0	8	0	0	0	2	0	2
G	1	2	3	0	0	0	8	0	5	4	5	2
T	4	1	0	0	0	0	0	8	3	2	2	2
Sum	8	8	8	8	8	8	8	8	8	8	8	8

Tom Schneider's sequence logo

(generated with Web Logo <http://weblogo.berkeley.edu/logo.cgi>)



Temario - Contenidos

Tema 4. Modelización estructural de proteínas

4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización *ab initio*

- Reconocimiento del plegamiento: aspectos básicos
- Potenciales estadísticos
- Reconocimiento de homología remota
- **Predicción de estructura secundaria**
- Modelización *ab initio*: principios físicos
- Modelización *ab initio*: métodos empíricos

4.2. Modelización por reconocimiento del plegamiento

Predicción de estructura secundaria

- Frecuencia relativa de aparición de aminoácidos en estructuras secundarias.
- *Machine learning* y redes neuronales.
 - Probabilidad según aparición en estructuras resueltas.
 - *Training set, test set* para validar modelo.
 - Mucha fiabilidad

Protein Sequence	D	M	N	W	H	I	G	M	C	R	C	N	N
P(alpha)	101	145	67	108	100	108	57	145	70	98	70	67	67
Running total for P(alpha)	101	246	313	421	521	629	686	831					
P(beta)	54	105	89	137	87	160	75	105	831 is greater than 812, so conclusion: The sequence is an alpha helix!				
Running total for P(beta)	54	159	248	385	472	632	707	812					
Evaluation:	Structure is an ALPHA HELIX												

Temario - Contenidos

Tema 4. Modelización estructural de proteínas

4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización *ab initio*

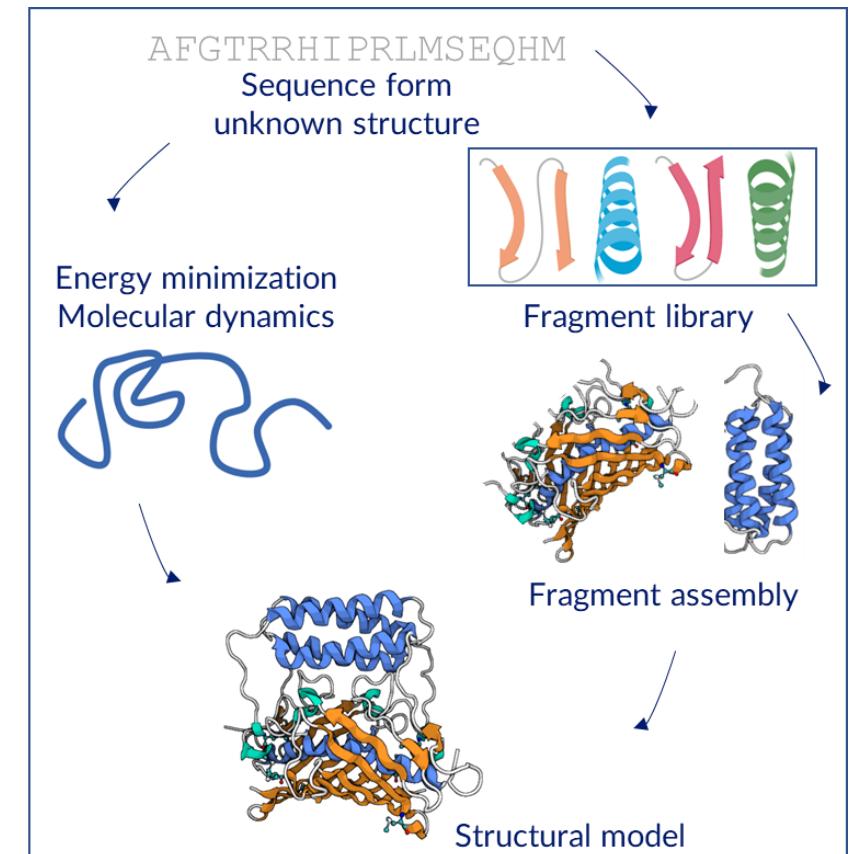
- Reconocimiento del plegamiento: aspectos básicos
- Potenciales estadísticos
- Reconocimiento de homología remota
- Predicción de estructura secundaria
- **Modelización *ab initio*: principios físicos**
- Modelización *ab initio*: métodos empíricos

4.3. Modelización *ab initio*

Imposibilidad de encontrar *templates* estructurales apropiados.

Recurrir a métodos *template-free*. Métodos basados en:

- Principios físicos
- Principios empíricos



4.3. Modelización *ab initio*

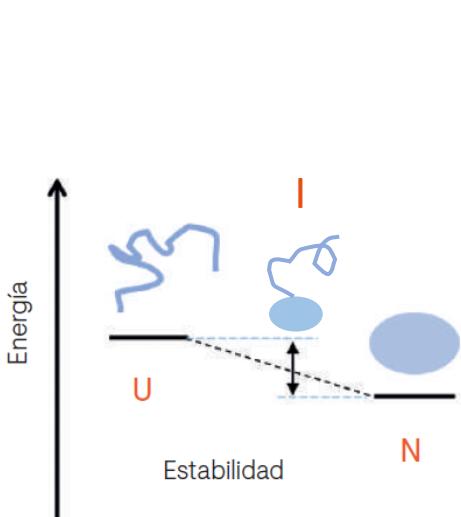
Principios físicos

- Principios físico-químicos determinantes de estructura 3D.
- Cadenas polipeptídicas adoptan de forma espontánea un plegamiento estable y único.
 - Exceptuando algunas proteínas multidominio (chaperonas).

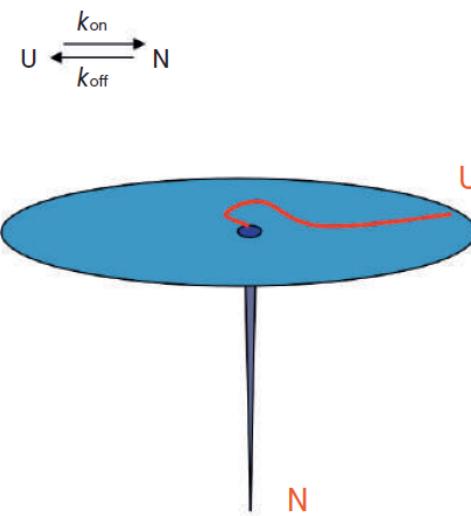
↓
Entender los aspectos termodinámicos y cinéticos

Desde punto de vista termodinámico:

- Espontáneo
- Reversible
- Estabilidad - ΔE



b)



4.3. Modelización *ab initio*

Principios físicos

Estabilidad – ΔE : Métodos para calcular E:

- Mecánica cuántica: muy costosa y no aplicable a sistemas complejos
- Mecánica molecular o clásica: menos costosa y aplicable a sistemas
- Métodos semiempíricos, potenciales estadísticos
- Métodos empíricos: conjunto de parámetros simples encajados en un ensayo

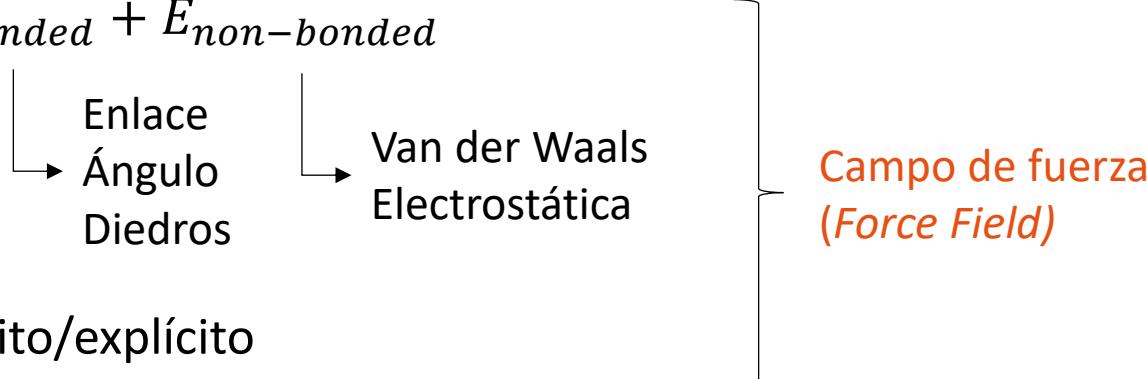
calidad



4.3. Modelización *ab initio*

Principios físicos

Mecánica clásica: $\sum E_{bonded} + E_{non-bonded}$



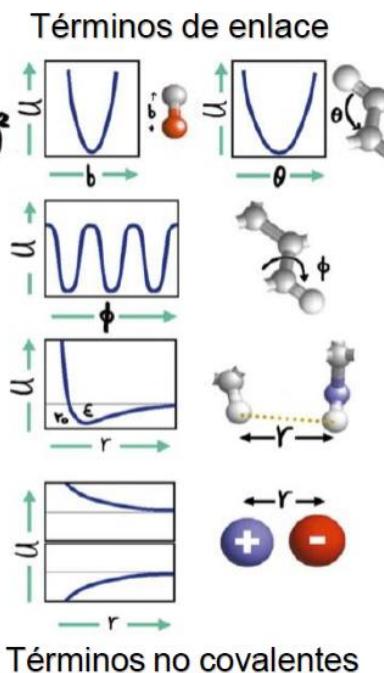
- Solvente: implícito/explícito

$$U = \sum_{\text{All Bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{All Angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{All Torsion Angles}} K_\phi [1 - \cos(n\phi + \delta)]$$

$$+ \sum_{\text{All nonbonded pairs}} \epsilon \left[\left(\frac{r_0}{r} \right)^12 - 2 \left(\frac{r_0}{r} \right)^6 \right]$$

$$+ \sum_{\text{All partial charges}} 332 q_i q_j / r$$



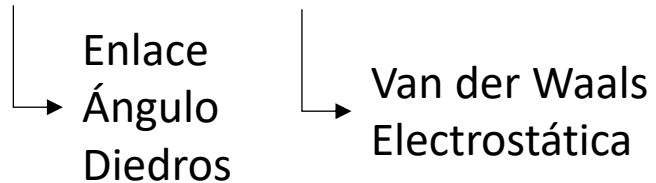
Campos de fuerzas
o forcefields

- AMBER
- CHARMM
- GROMACS
- (...)

4.3. Modelización *ab initio*

Principios físicos

Mecánica clásica: $\sum E_{bonded} + E_{non-bonded}$

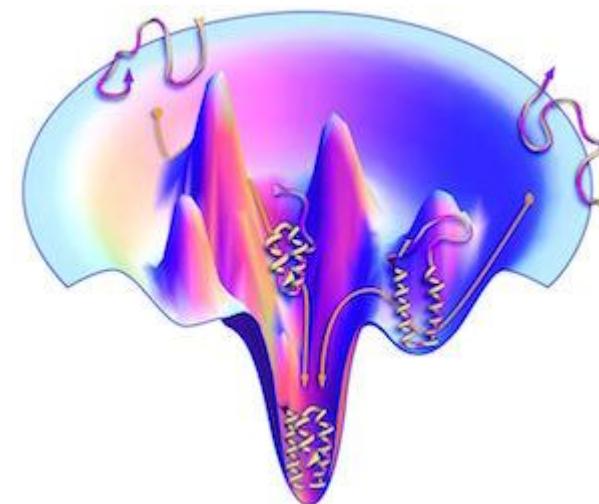
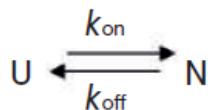


Campo de fuerza
(Force Field)

- Solvente: implícito/explícito

Energía de estado nativo difícil de calcular

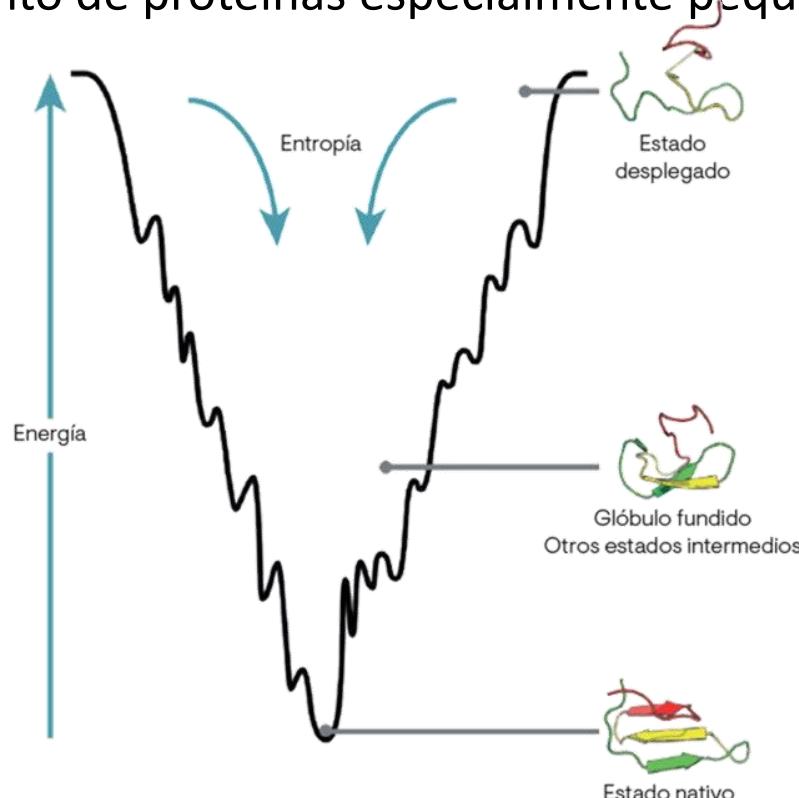
- Métodos Monte-Carlo
 - Steepest descent
 - Conjugate gradient
 - Simulated annealing
- Dinámica molecular
 - Cinética de plegamiento



4.3. Modelización *ab initio*

Principios físicos

- Dinámica molecular
 - Cinética de plegamiento
 - Paradoja de Levinthal – imposibilidad de exploración de conformaciones al azar
 - Descripción del plegamiento de proteínas especialmente pequeñas.



Temario - Contenidos

Tema 4. Modelización estructural de proteínas

4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización *ab initio*

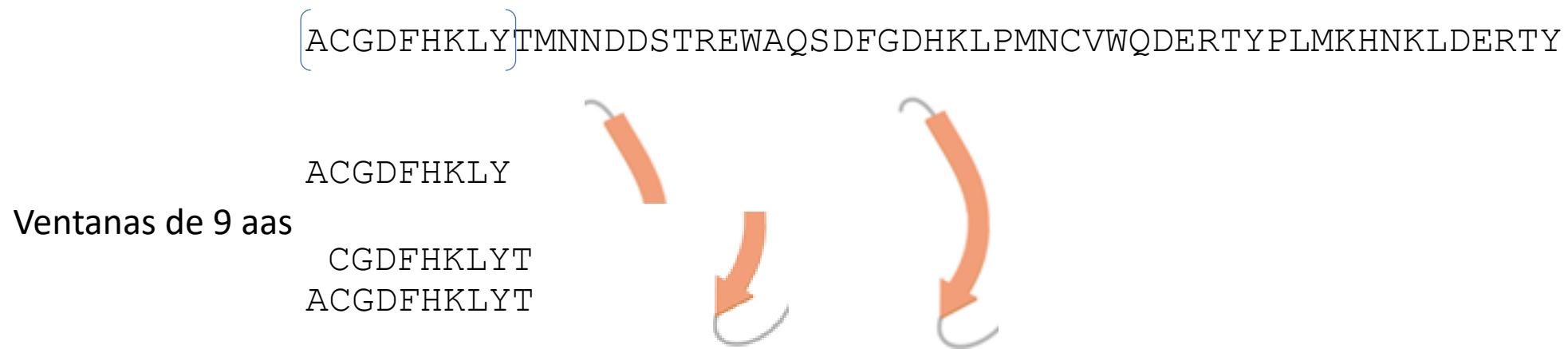
- Reconocimiento del plegamiento: aspectos básicos
- Potenciales estadísticos
- Reconocimiento de homología remota
- Predicción de estructura secundaria
- Modelización *ab initio*: principios físicos
- Modelización *ab initio*: métodos empíricos

4.3. Modelización *ab initio*

Principios empíricos

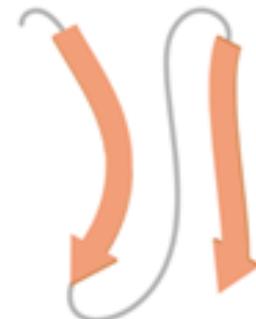
Basados en información contenida en datos estructurales.

Rosetta



Ventanas de 9 aas

- Puntuaciones de potenciales estadísticos sin considerar entorno.
- Cuando se llega a cierta longitud, se recalcula la energía incluyendo el entorno.
- Cambios de ángulos de torsión para unir fragmentos.
- *Simulated annealing* para optimizar



4.3. Modelización *ab initio*

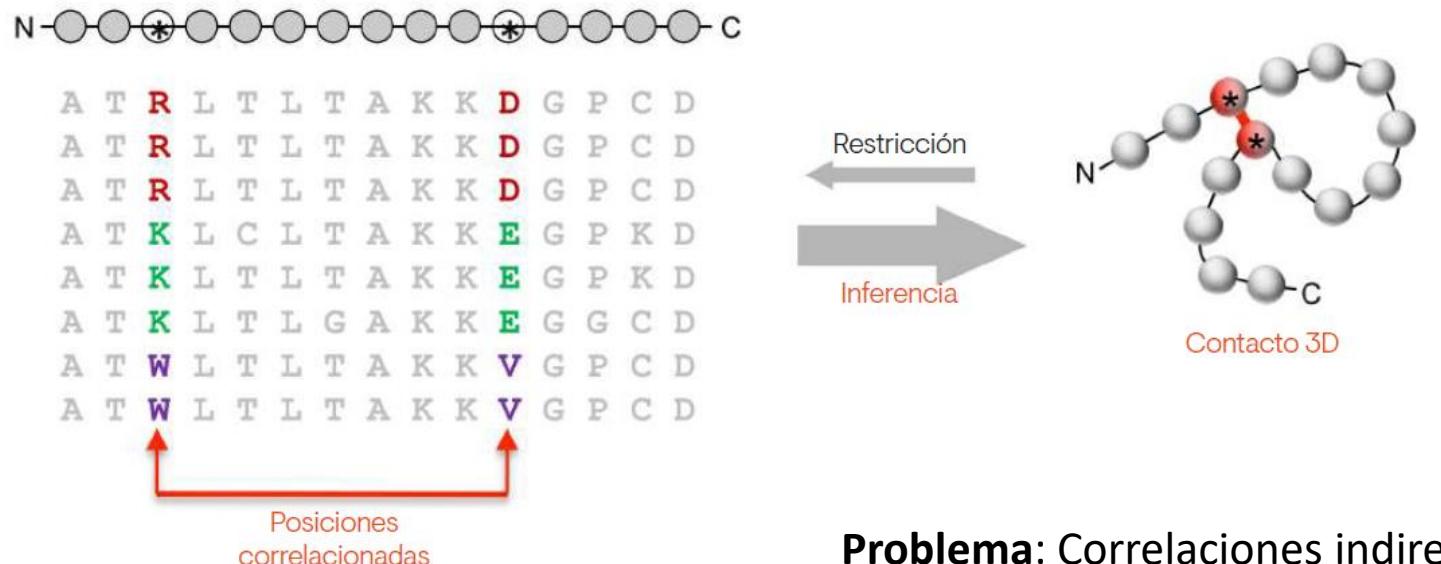
Principios empíricos

Basados en información contenida en datos estructurales.

Información mutua

Residuos que interaccionan mutan de forma simultánea para conservar la interacción.

Punto de contacto y restricción en construcción de modelos.

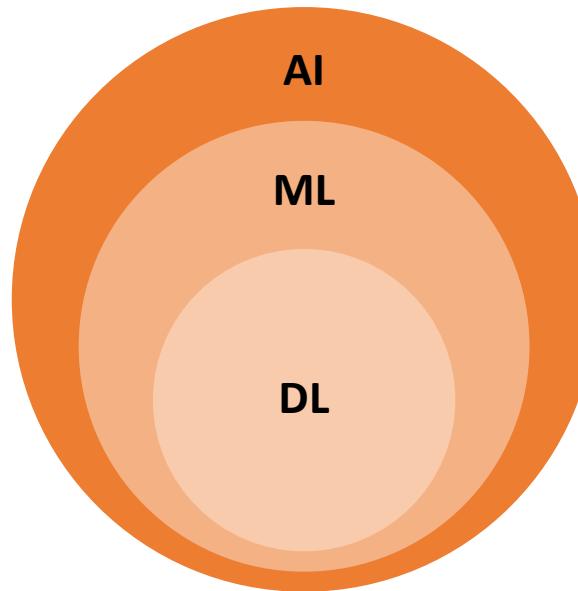


Problema: Correlaciones indirectas

4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

Aprendizaje automático mediante redes neuronales a partir de estructuras conocidas.



Inteligencia Artificial

Ingeniería para crear programas o máquinas inteligentes

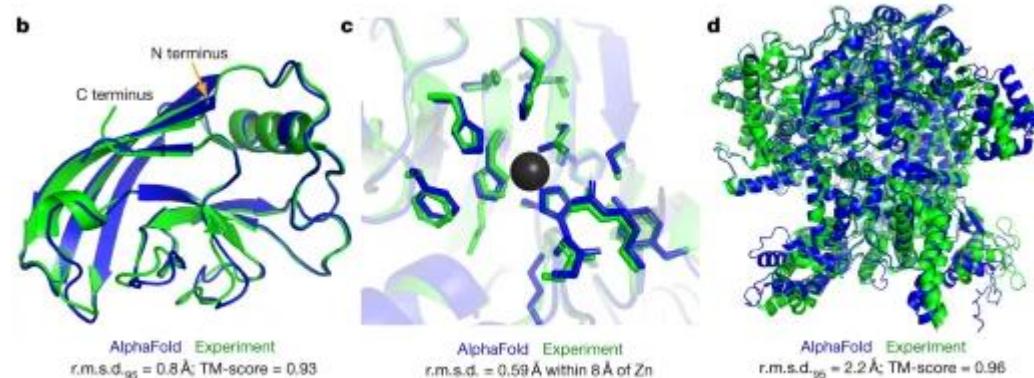
Aprendizaje Automático

Habilidad de aprendizaje sin programación explícita para ello

Aprendizaje Profundo

Aprendizaje basado en redes neuronales profundas

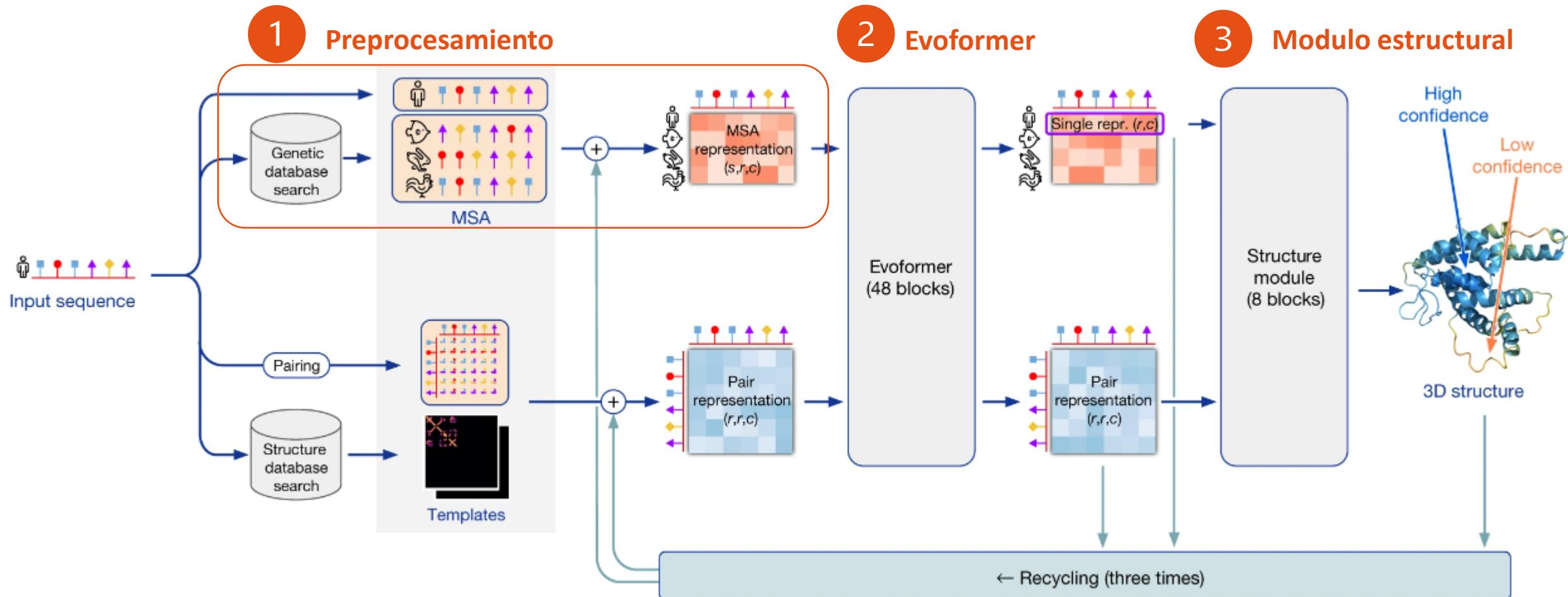
Fig. 1: AlphaFold produces highly accurate structures.



4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

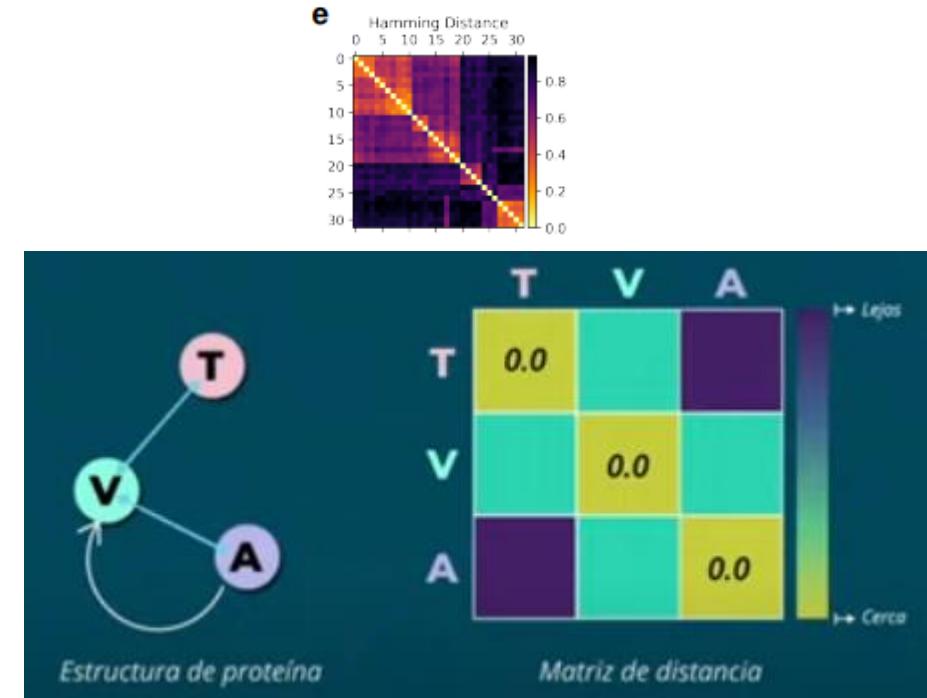
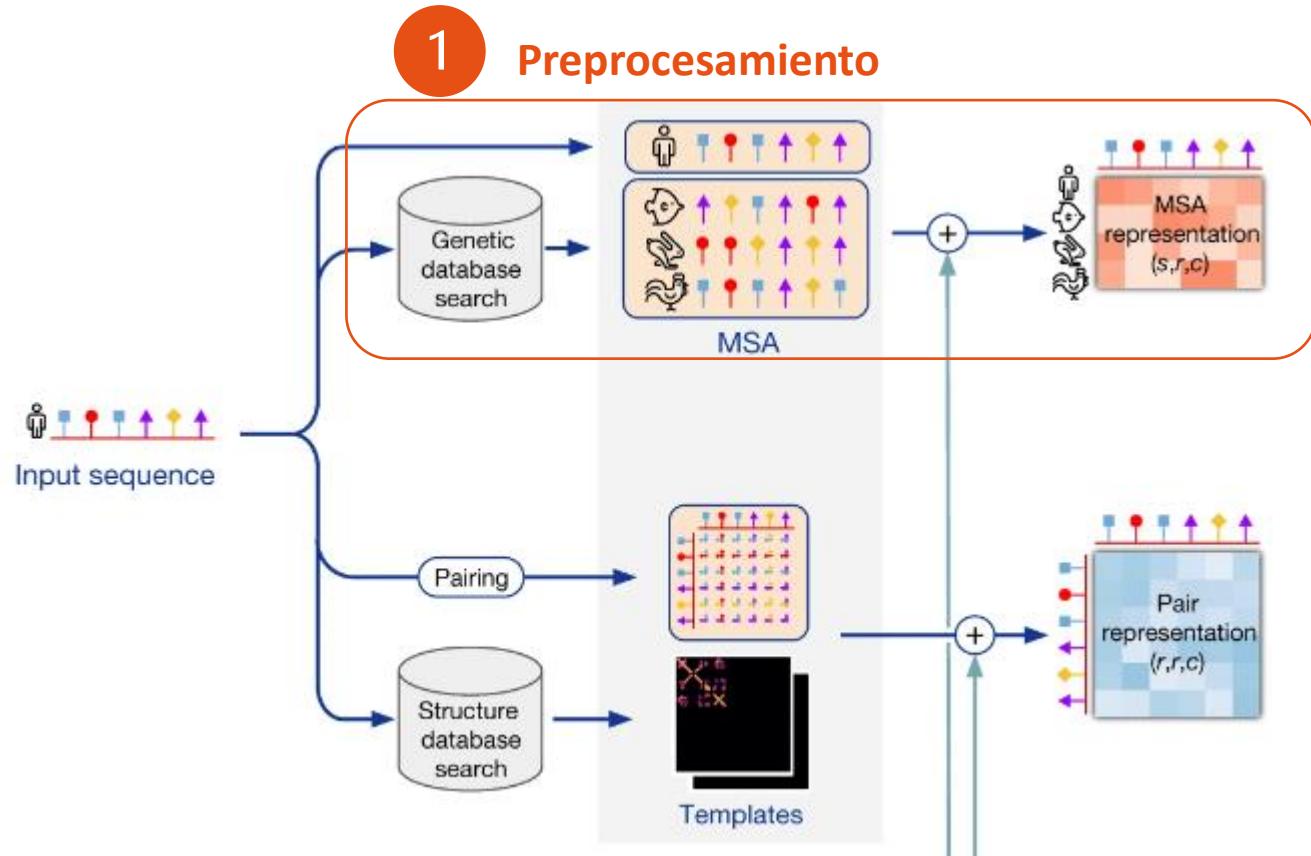
AlphaFold produces highly accurate structures.



4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

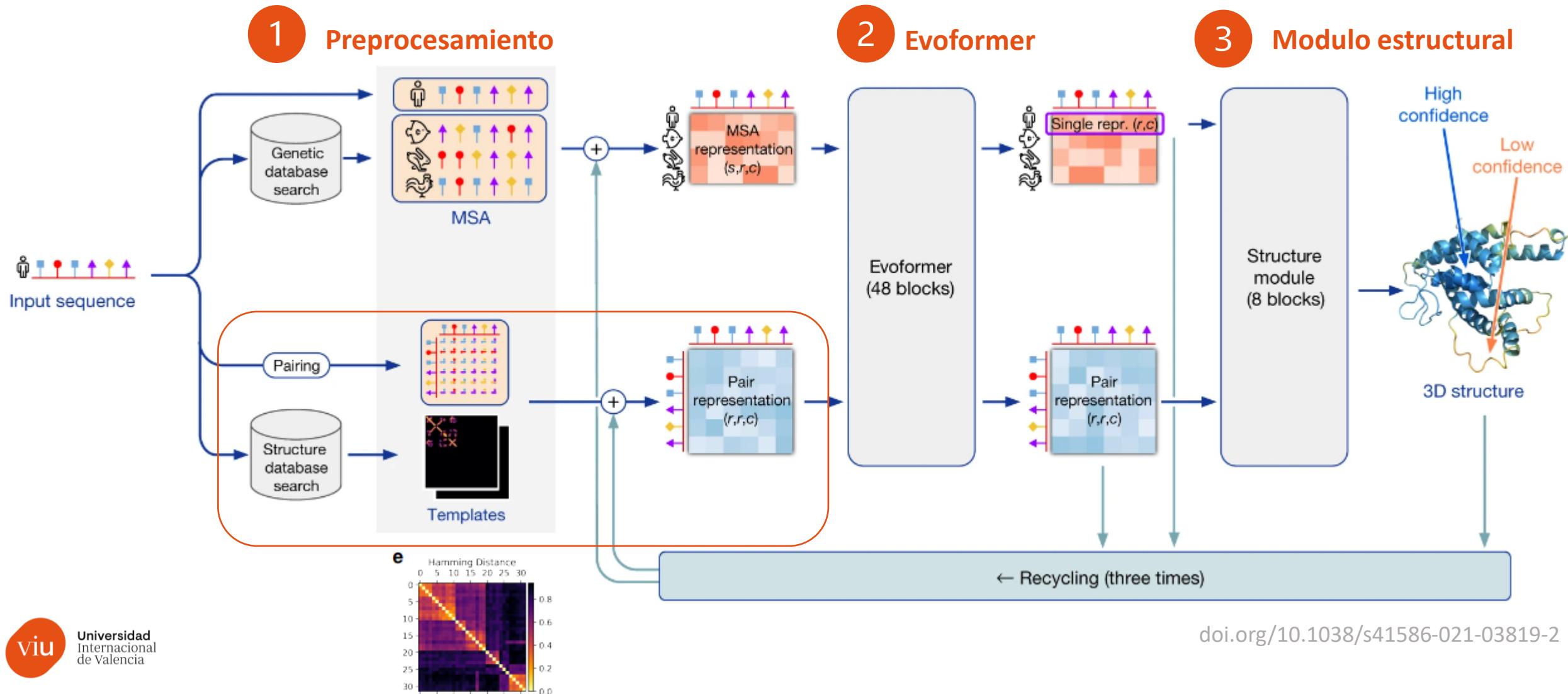
AlphaFold produces highly accurate structures.



4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

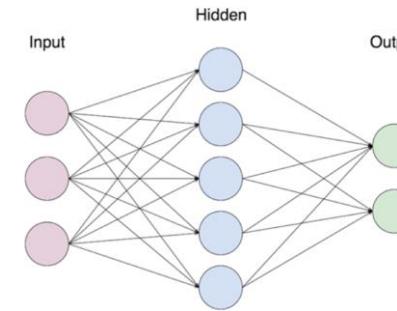
AlphaFold produces highly accurate structures.



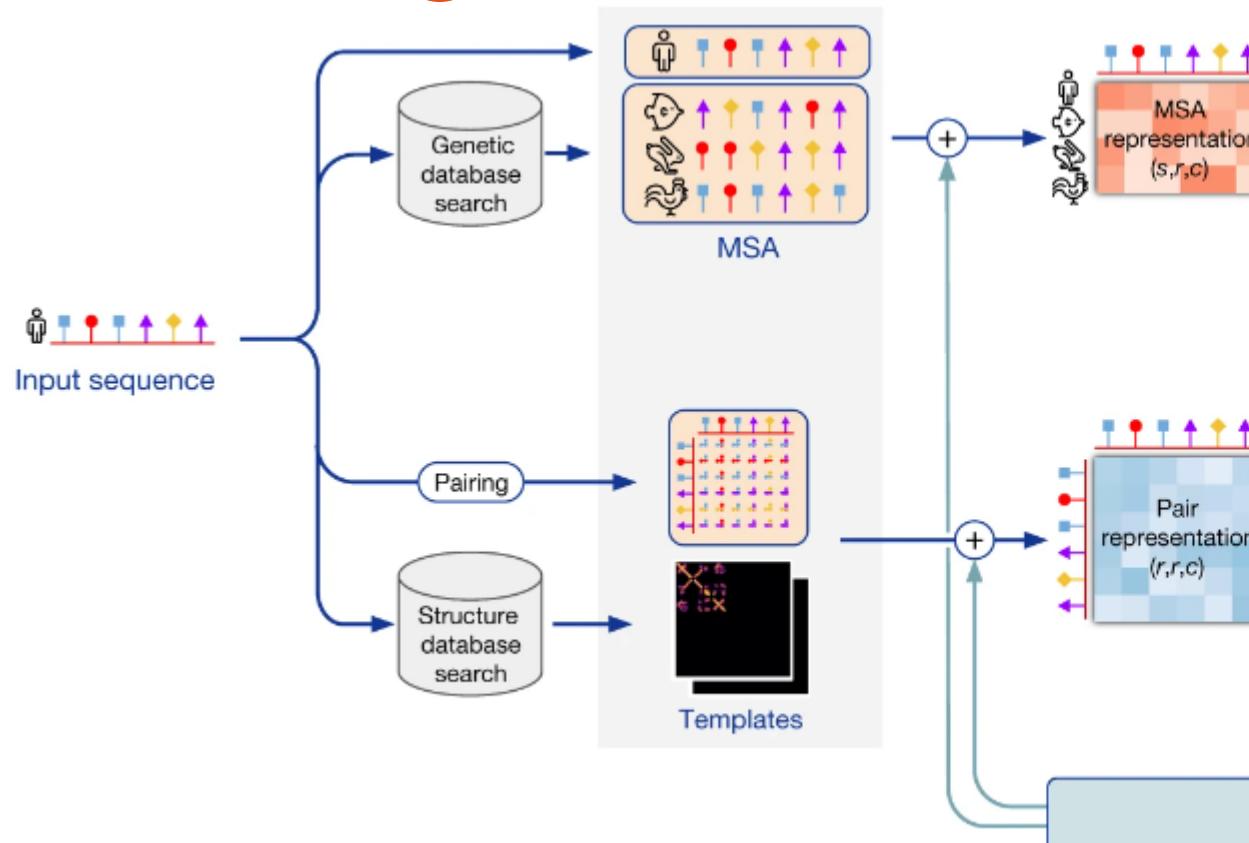
4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

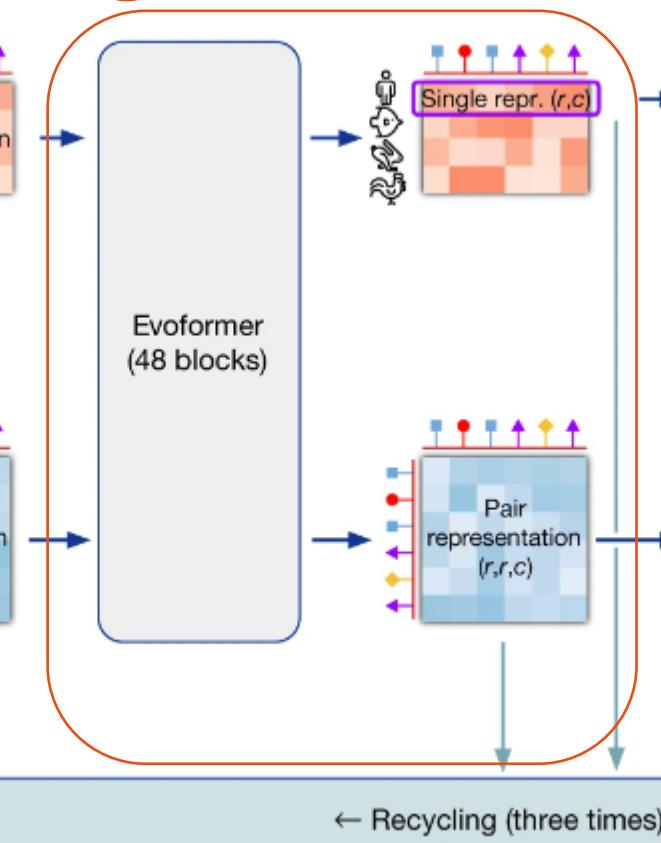
AlphaFold produces highly accurate structures.



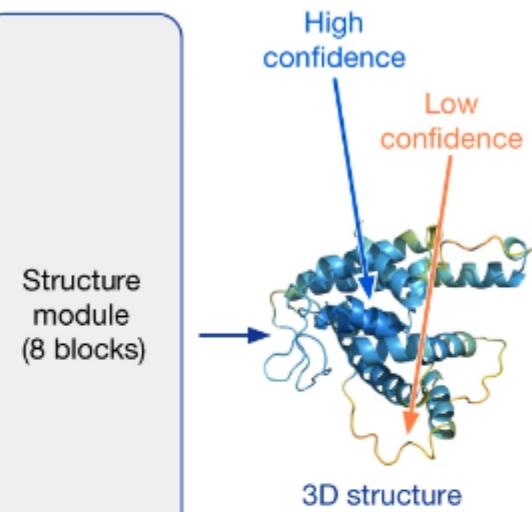
1 Preprocesamiento



2 Evoformer



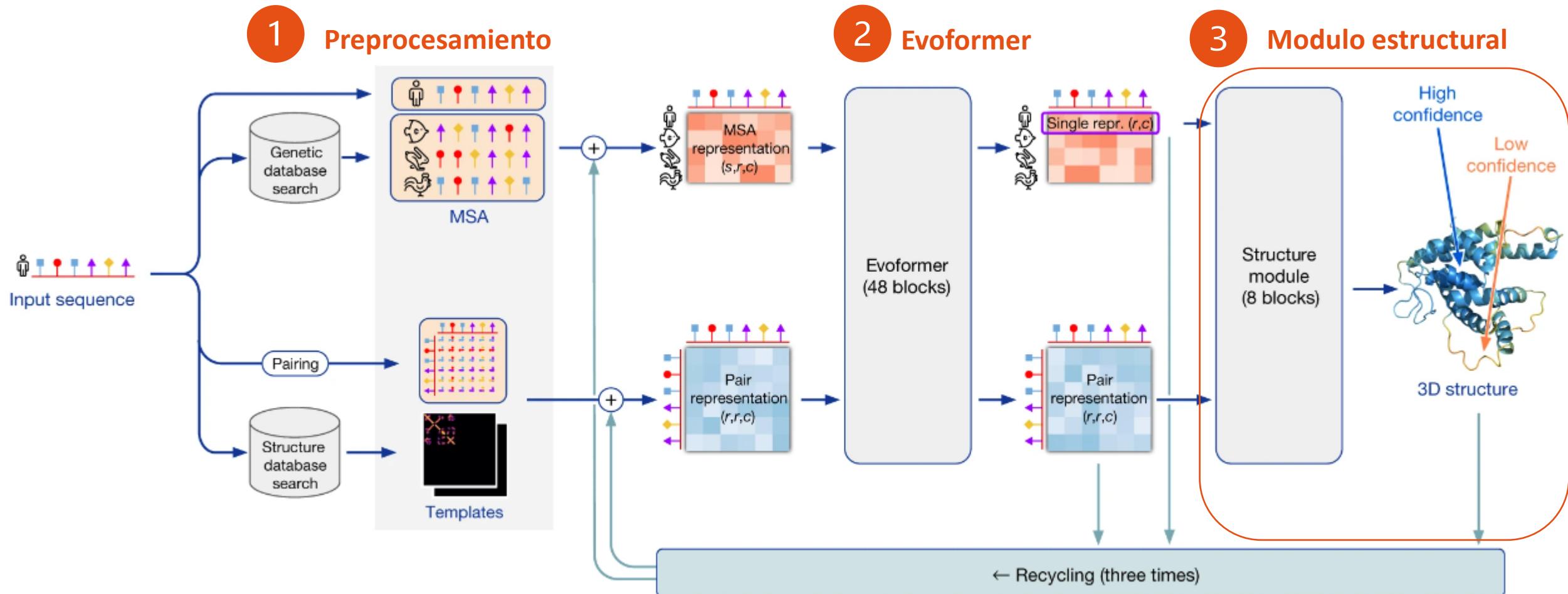
3 Modulo estructural



4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

AlphaFold produces highly accurate structures.



4.3. Modelización *ab initio*

Principios basados en inteligencia artificial

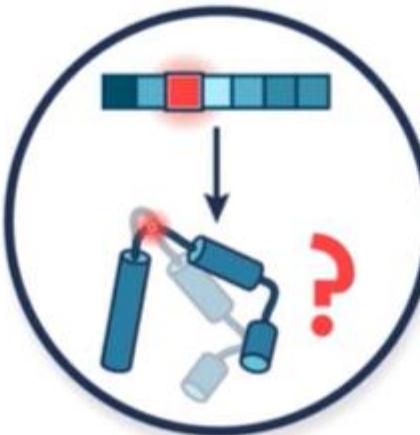
AlphaFold produces highly accurate structures.

Aplicaciones

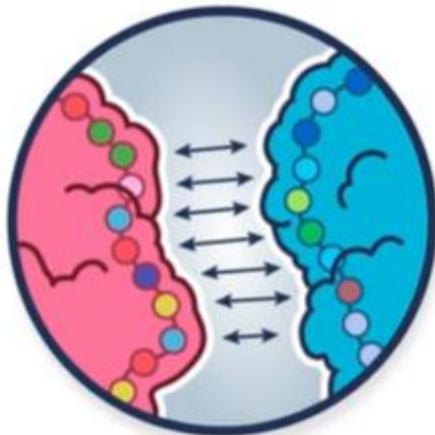
Discovering drugs



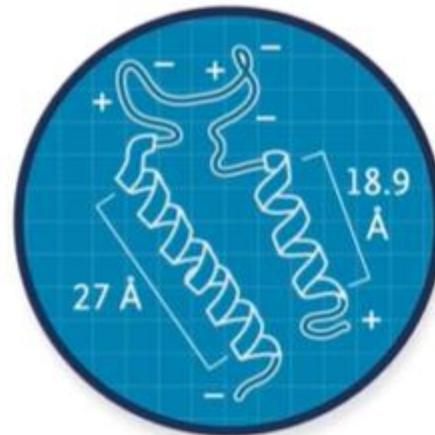
Effect of genetic variants



Modeling protein–protein interactions



Engineering artificial proteins





viu

Universidad
Internacional
de Valencia

universidadviu.com

De:
 Planeta Formación y Universidades