

# Introducción al *docking* molecular



Cristina Coscolín Galán

[cristinacoscolin@gmail.com](mailto:cristinacoscolin@gmail.com)

[www.linkedin.com/in/cristinacoscolin](https://www.linkedin.com/in/cristinacoscolin)



# Índice

## Introducción

- Qué es el docking molecular y qué busca predecir
- Por qué es importante
- Ejemplos en biología y biotecnología

## Fundamentos

- Concepto de ligando y receptor
- Tipos de docking: rígido vs flexible
- Etapas del proceso

## Scoring

- Qué mide la función de scoring
- Cómo se interpreta
- Limitaciones: flexibilidad, solvente, calidad estructural

## Herramientas y flujo de trabajo

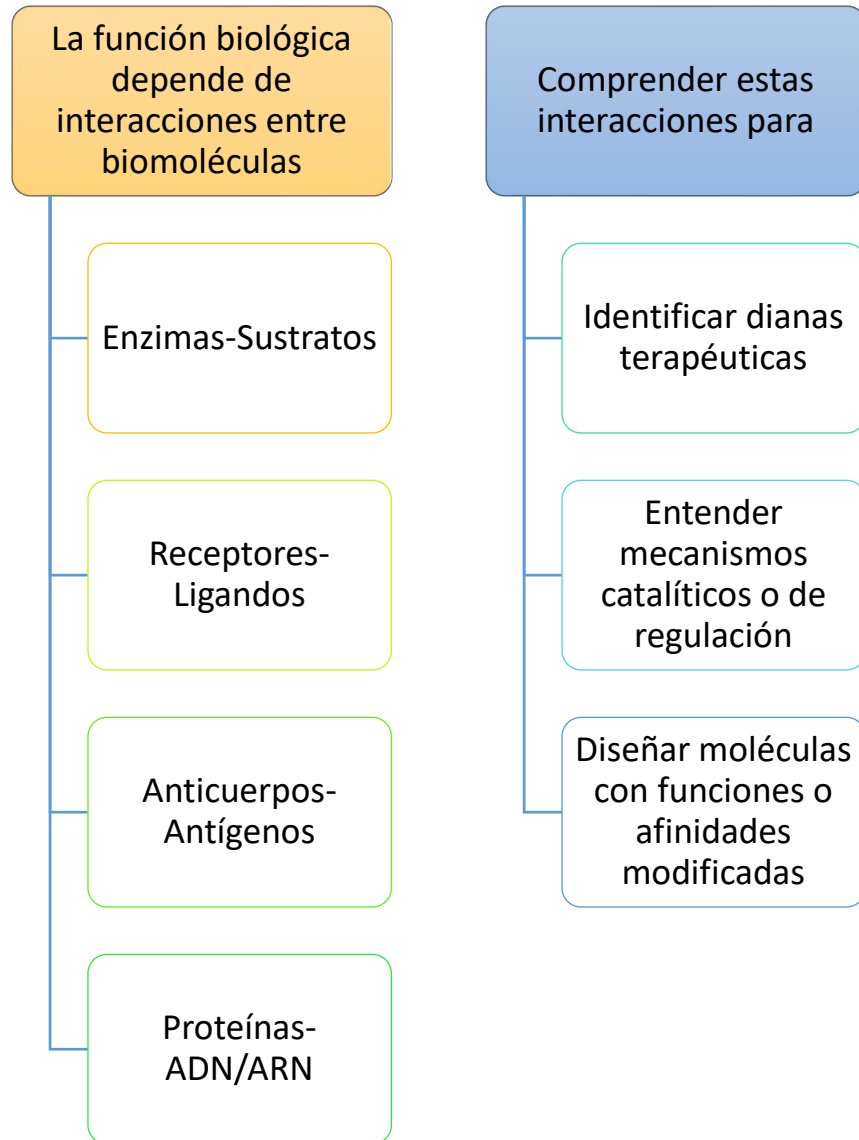
- Principales programas (AutoDock, Vina, SwissDock)
- Visualización de resultados (Chimera, PyMol, Maestro)
- Pipeline típico

## Ejemplo práctico con SwissDock



## Dudas

# Contexto biológico



<https://doi.org/10.2210/pdb4EY7/pdb>

La unión de la enzima **acetilcolinesterasa** con su inhibidor (*donepezilo*) se estudia mediante docking para comprender la base estructural de la inhibición en Alzheimer.

# ¿Qué es el docking molecular?

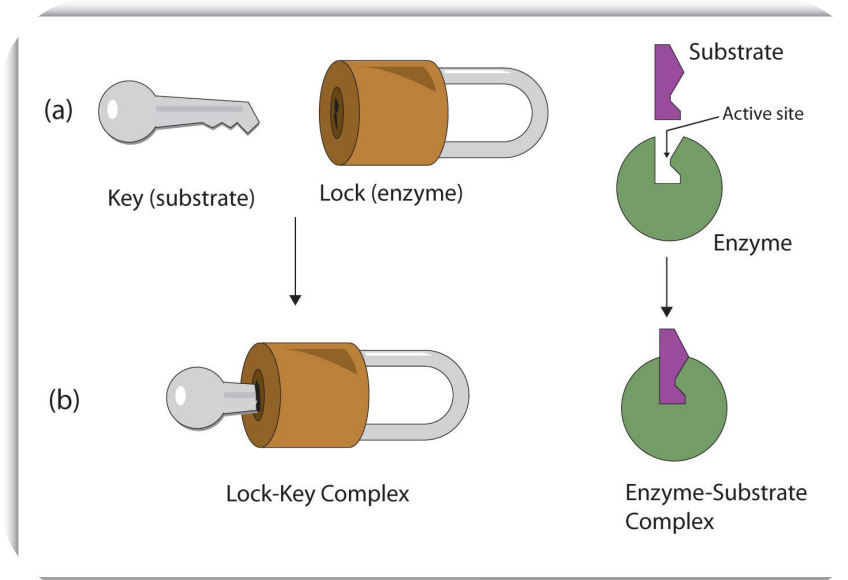
- El *docking molecular* es un método computacional que predice la **orientación óptima (pose)** y la **afinidad de unión** de una molécula pequeña (*ligando*) al unirse con una macromolécula (*receptor*), generalmente una proteína.

Identificar la mejor pose del ligando en el sitio activo

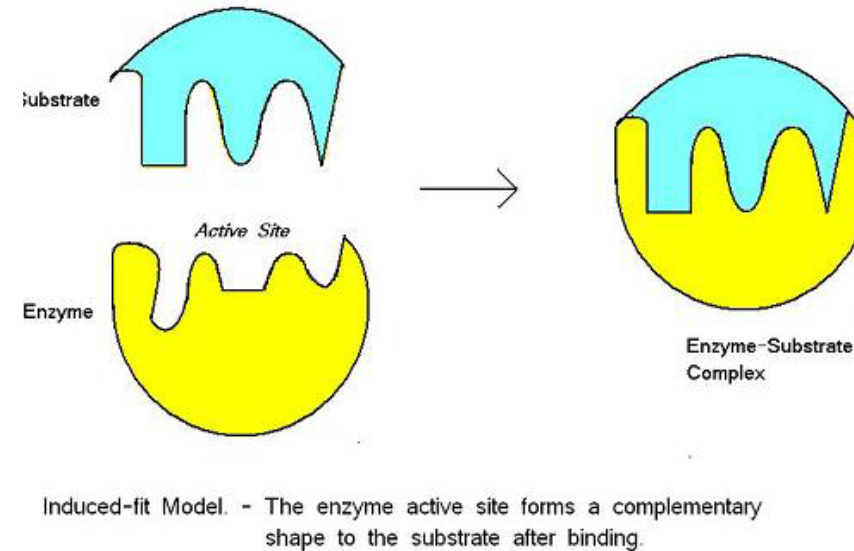
Estimar la energía libre de unión o puntuación (scoring)

Ranking de ligandos según su afinidad teórica

# Llave cerradura vs induced fit



Llave-cerradura  
Emile Fischer (1894)



Induced fit Daniel  
Koshland (1958)

El docking moderno intenta modelar este **ajuste dinámico** con algoritmos de exploración conformacional.



# Relevancia del docking molecular



**Mutagénesis dirigida:** evaluación *in silico* del efecto de mutaciones sobre la unión del sustrato

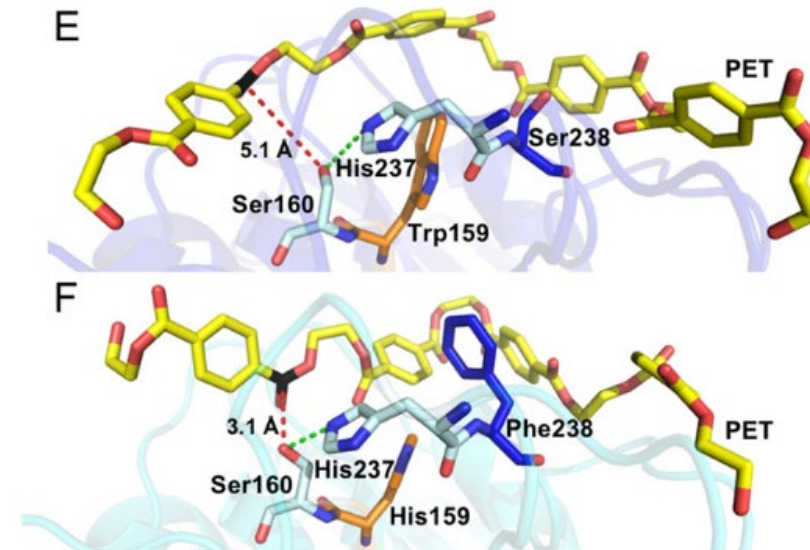
**Biología estructural:** apoyo a la interpretación de datos experimentales (cristalografía, NMR, cryo-EM)

**Biocatálisis:** predicción del modo de unión en enzimas industriales: PETasas, esterasas, lipasas...

**Descubrimiento de fármacos:** cribado virtual (virtual screening) de miles de compuestos contra una proteína diana (SARS-CoV-2 Mpro)

## Characterization and engineering of a plastic-degrading aromatic polyesterase

Harry P. Austin<sup>a,1</sup>, Mark D. Allen<sup>a,1</sup>, Bryon S. Donohoe<sup>b,1</sup>, Nicholas A. Rorrer<sup>c,1</sup>, Fiona L. Kearns<sup>d,1</sup>, Rodrigo L. Silveira<sup>b,e</sup>, Benjamin C. Pollard<sup>d</sup>, Graham Dominick<sup>c</sup>, Ramona Duman<sup>f</sup>, Kamel El Omari<sup>f</sup>, Vitaliy Mykhaylyk<sup>f</sup>, Armin Wagner<sup>f</sup>, William E. Michener<sup>c</sup>, Antonella Amore<sup>b</sup>, Munir S. Skaf<sup>e</sup>, Michael F. Crowley<sup>b</sup>, Alan W. Thorne<sup>a</sup>, Christopher W. Johnson<sup>c</sup>, H. Lee Woodcock<sup>d,2</sup>, John E. McGeehan<sup>a,2</sup>, and Gregg T. Beckham<sup>c,2</sup>

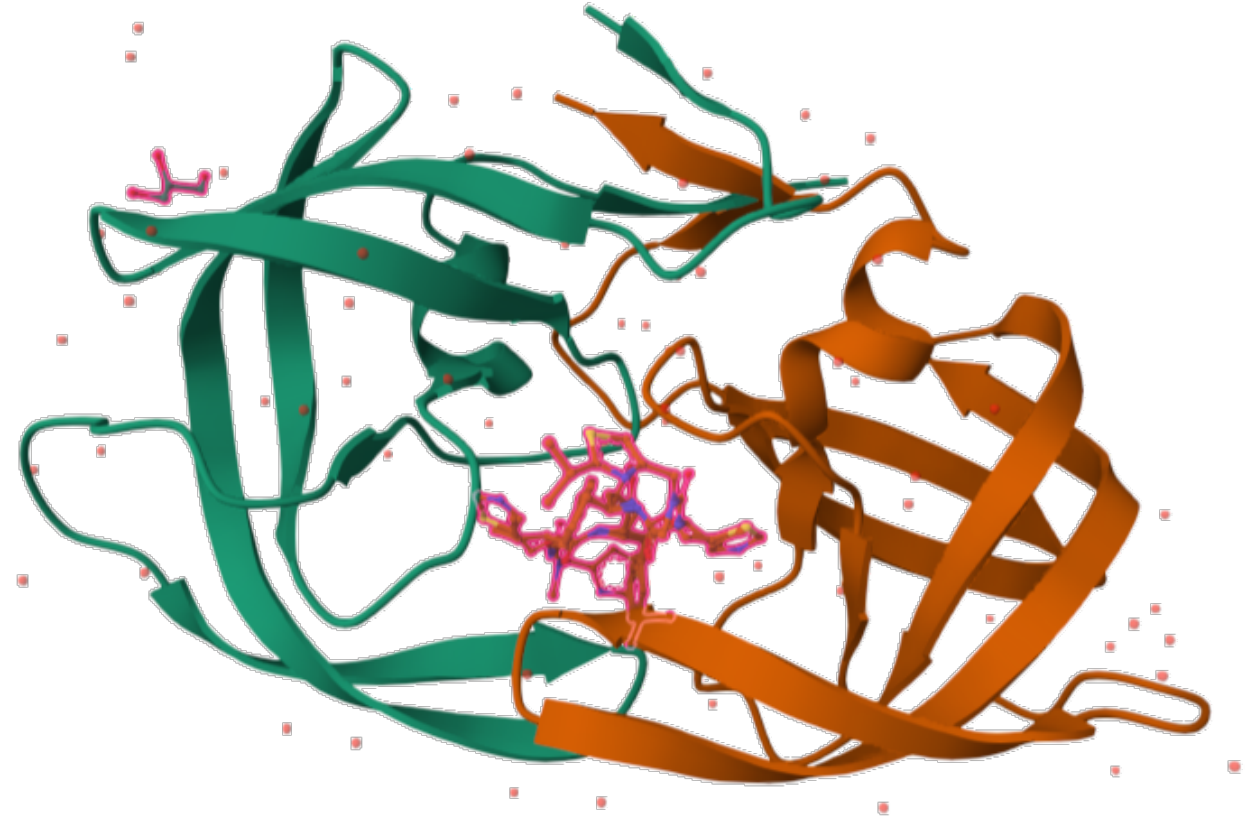


El docking de sustratos plásticos en una **PETasa** permitió identificar residuos críticos para la degradación de PET, orientando el diseño de mutantes más eficientes.

# Componentes principales

**Ligando:**  
molécula  
pequeña,  
péptido o  
metabolito  
que se une al  
receptor

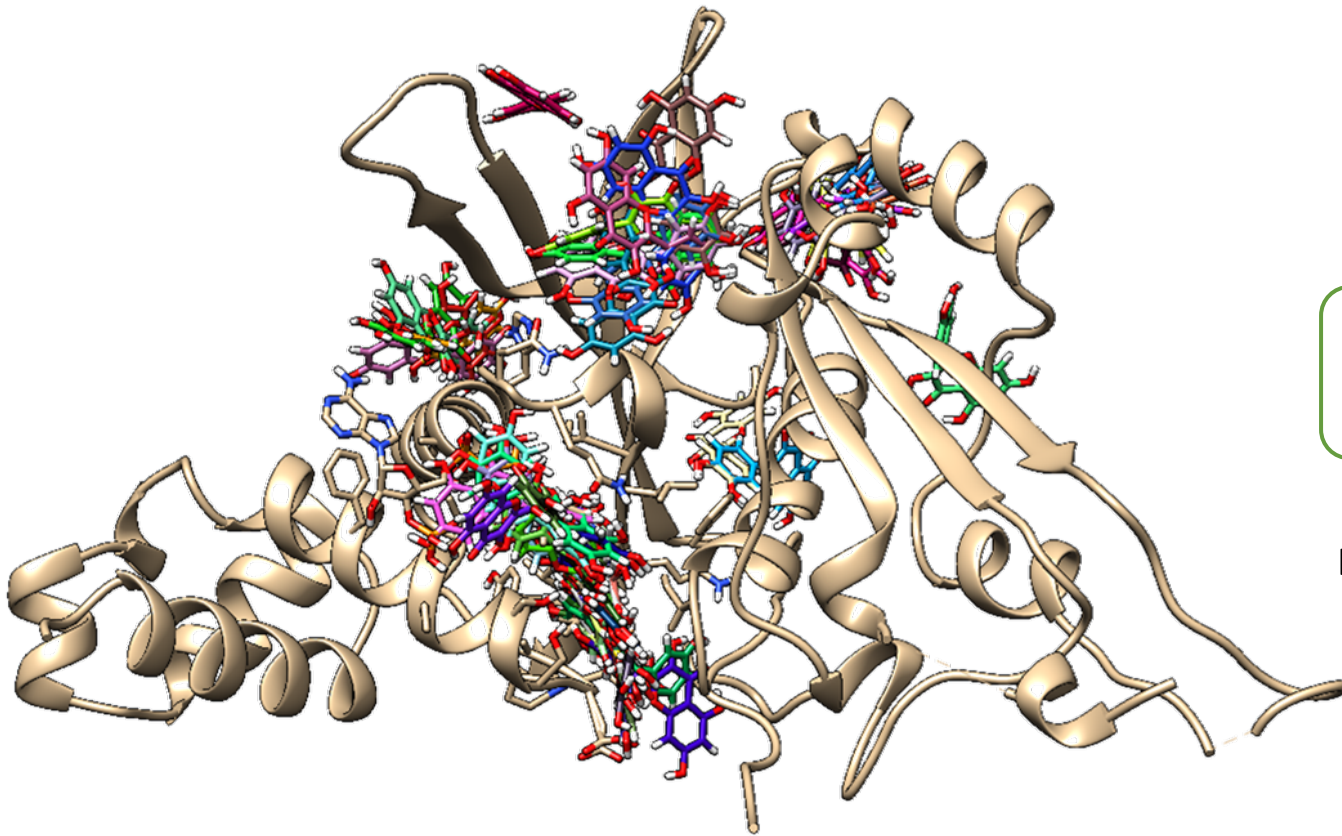
**Receptor:**  
normalmente  
es un  
proteína  
aunque  
también  
puede ser  
ADN, ARN o  
una  
membrana



Inhibidor **ritonavir** con la **proteasa del VIH-1**, el receptor es la enzima viral y el ligando es el fármaco.

<https://doi.org/10.2210/pdb2B60/pdb>

# Clasificación del docking



## Rígido:

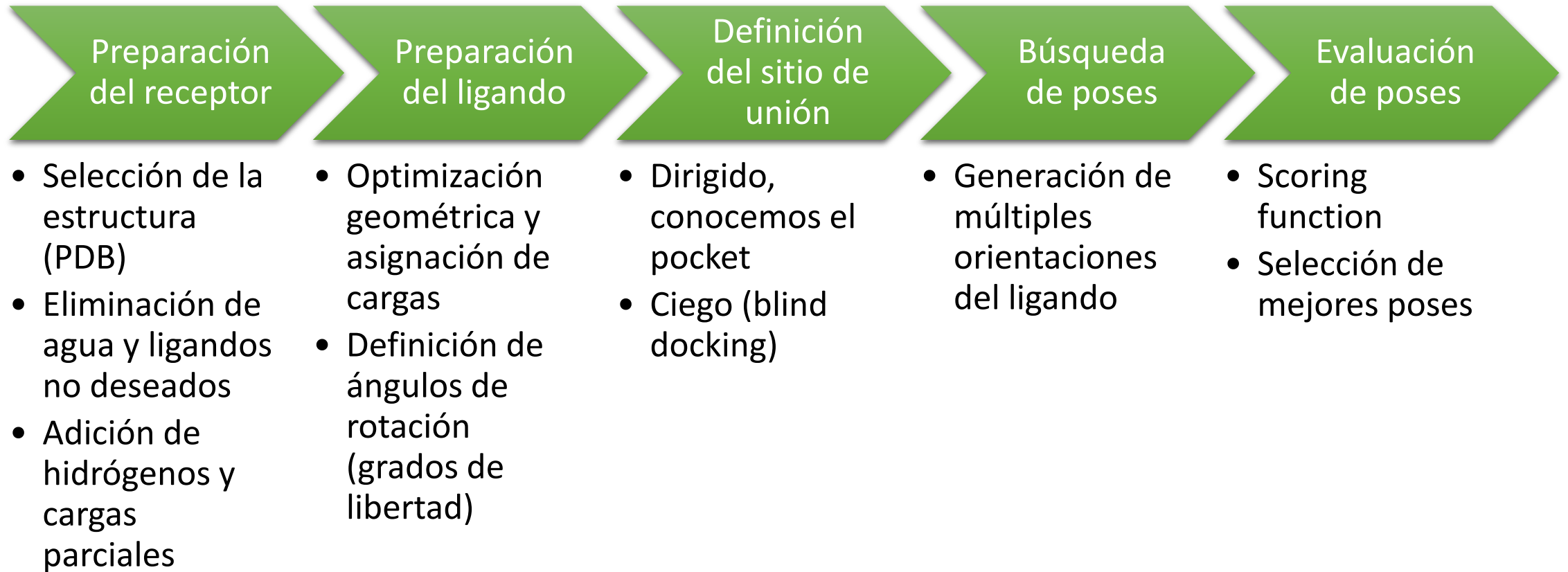
- Tanto el receptor como el ligando se consideran estructuras rígidas
- Rápido, pero menos realista
- Útil para cribados iniciales o estructuras bien definidas

## Flexible:

- Se permite que el ligando (y, a veces, el receptor) adopte distintas conformaciones
- Modela mejor el ajuste inducido
- Mayor coste computacional



# Etapas fundamentales del docking molecular



# Algoritmos de exploración conformacional

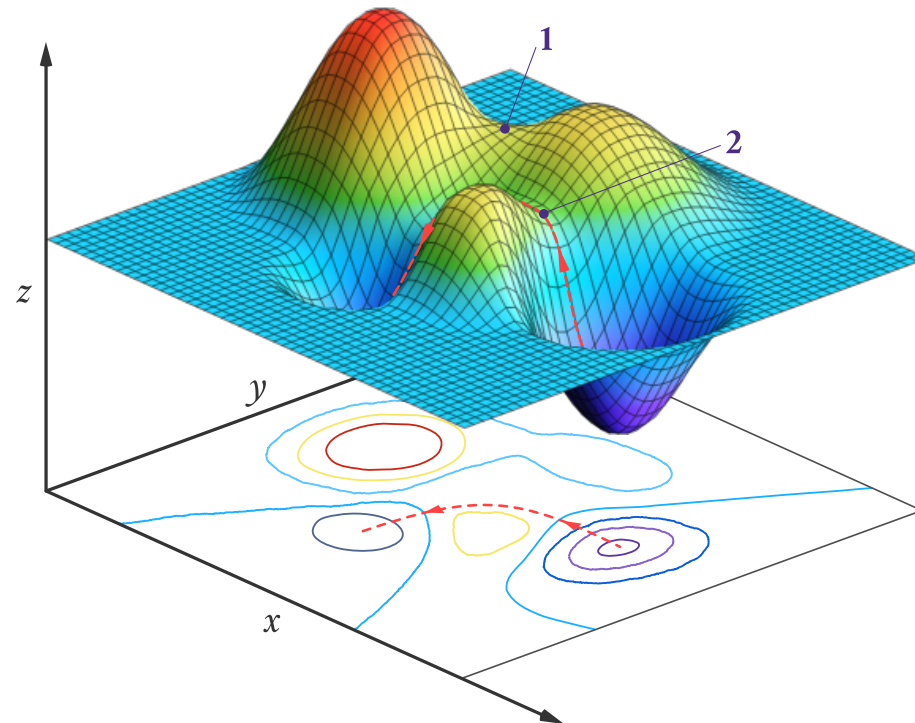
- En *docking molecular*, el objetivo es **encontrar la conformación (pose)** del ligando que **minimiza la energía de interacción** con el receptor.
- Esto equivale a buscar el **mínimo global** en un “paisaje energético” muy complejo (con muchos mínimos locales).

## Métodos deterministas:

- Grid search
- Incremental construction

## Métodos estocásticos

- Algoritmos genéticos (AutoDock)
- Monte Carlo simulated annealing
- Swarm optimization



# Métodos deterministas

- Siguen **rutas predefinidas y sistemáticas** para explorar el espacio conformacional.
- Dado el mismo punto de partida y parámetros, **siempre devuelven el mismo resultado**.
- Buscan el **mínimo local o global** mediante un recorrido exhaustivo o un gradiente.



Reproducibles y más predecibles. Requieren menos tiempo si el espacio de búsqueda es pequeño.

Escalan muy mal cuando hay muchos grados de libertad. Tienden a quedarse atrapados en mínimos locales.

## Grid Search


- Se exploran posiciones y orientaciones discretas en una malla

## Incremental construction (FlexX)

- El ligando se construye pieza a pieza evaluando cada posición

# Métodos estocásticos

- Incorporan **aleatoriedad controlada** para explorar el espacio conformacional.
- No siguen un camino fijo: **pueden saltar entre mínimos locales**.
- Cada ejecución puede producir un resultado diferente.



Mayor capacidad de encontrar un mínimo global. Escalan mejor con ligandos flexibles.

Requieren mas tiempo de cálculo. Resultados algo menos reproducibles (aunque se puede fijar una “semilla”).

## Algoritmos genéticos (AutoDock)

- Generan “poblaciones” de soluciones que evolucionan

## Monte Carlo simulated annealing

- Acepta situaciones peores temporalmente para escapar de mínimos locales

## Particle swarm optimization

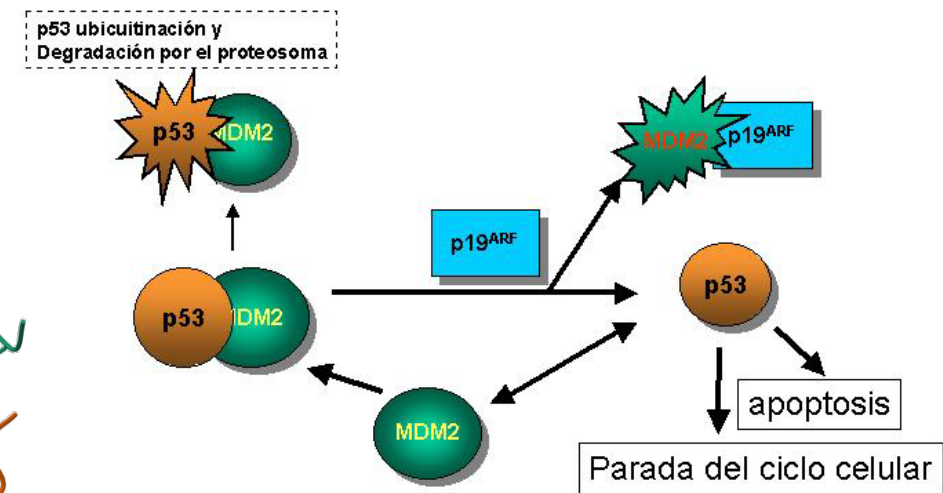
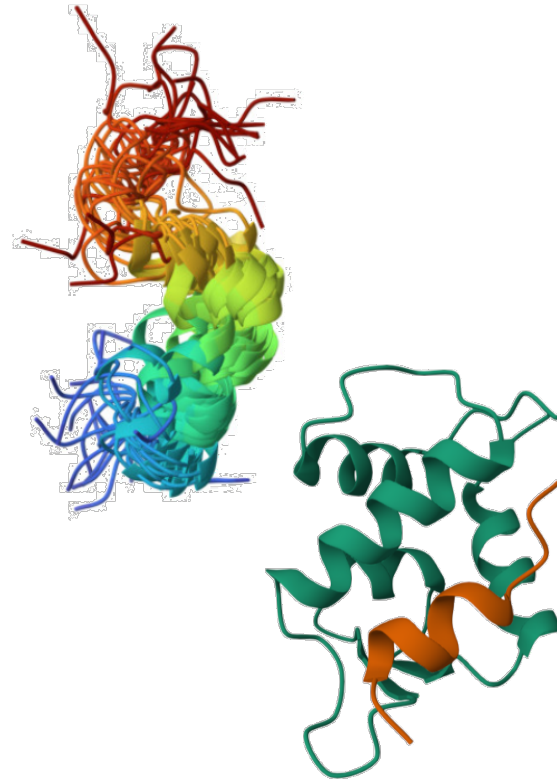
- Simula “enjambres” de partículas buscando colectivamente el mejor punto

# Limitaciones y fuentes de error

- **Flexibilidad limitada del receptor** → el sitio activo real puede adoptar otras conformaciones.
- **Efectos del solvente y la entropía** suelen simplificarse o ignorarse.
- **Calidad de la estructura PDB:** errores o resoluciones bajas afectan la precisión.
- **Simplificación del scoring:** no siempre correlaciona bien con la afinidad experimental.

p53 es un supresor tumoral que detiene el ciclo celular y promueve la apoptosis ante daño en el ADN. MDM2 regula negativamente a p53 mediante unión directa y degradación proteasomal.

El péptido N-terminal de p53 que se une a MDM2 es muy flexible y solo adopta estructura al acoplarse, lo que impide modelarlo con precisión y estudiarlo mediante docking rígido.



**MDM2 es una ubiquitina-ligasa de p53**



# Función de scoring: cuantificación de la afinidad de unión

- El *docking* genera muchas posibles **poses** del ligando en el sitio activo.
- Cada pose se evalúa mediante una **función de scoring**, que estima su **energía libre de unión ( $\Delta G_{bind}$ )**.
- El objetivo es identificar la **pose más estable** (energía mínima) y **hacer un “ranking”** los ligandos según su afinidad teórica.

Cuanto menor es la energía, mayor es la afinidad predicha

El score no es un valor absoluto, debe interpretarse comparativamente entre ligandos o poses

# Tipos de funciones de Scoring

## Physics-based

DOCK: 
$$E_{bind} = \sum_{i=1}^L \sum_{j=1}^R \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right)$$

**Physics\_based:** basadas en potenciales físicos (electrostática, Wan der Waals). Calculan interacciones átomo a átomo mediante campos de fuerza clásicos. Ej: DOCK.

## Empirical

X-Score: 
$$E_{bind} = w_0 + w_1 \Delta G_{vdW} + w_2 \Delta G_{Hbond} + w_3 \Delta G_{rot} + w_4 \Delta G_{hydro}$$

**Empirical:** Combinan términos energéticos ponderados por regresión (hidrofobicidad, puentes de H, desolvación. Ajustados a datos experimentales de afinidad. Ej: X-Score, Chem-Score.

## Knowledge-based

PMF: 
$$E_{bind} = \sum_{i=1}^L \sum_{j=1}^R -k_B T \ln[g(r)]$$

**Knowledge-based:** derivan potenciales estadísticos de estructuras cristalográficas. Capturan tendencias observadas en complejos reales. Ej: Potencial Of Mean Force.

## Machine learning-based

RF-Score: 
$$E_{bind} = f_{RF}(x_m)$$
$$x_{ij} = \sum_{i=1}^L \sum_{j=1}^R \theta[d_{cutoff} - d_{ij}]$$

**Machine-learning-based:** Aprenden relaciones entre descriptores estructurales y afinidad experimental mediante algoritmos de ML. Permiten modelar interacciones no lineales y combinaciones complejas de términos. EJ: RF-Score.

# Limitaciones de las funciones de scoring

Las funciones son aproximaciones a la energía libre de unión real.

No capturan bien:

- Efectos del solvente
- Flexibilidad del receptor
- Entropía conformacional

Diferentes programas suelen dar resultados distintos para el mismo sistema

# Flujo de trabajo y herramientas



Estructura



Preparación



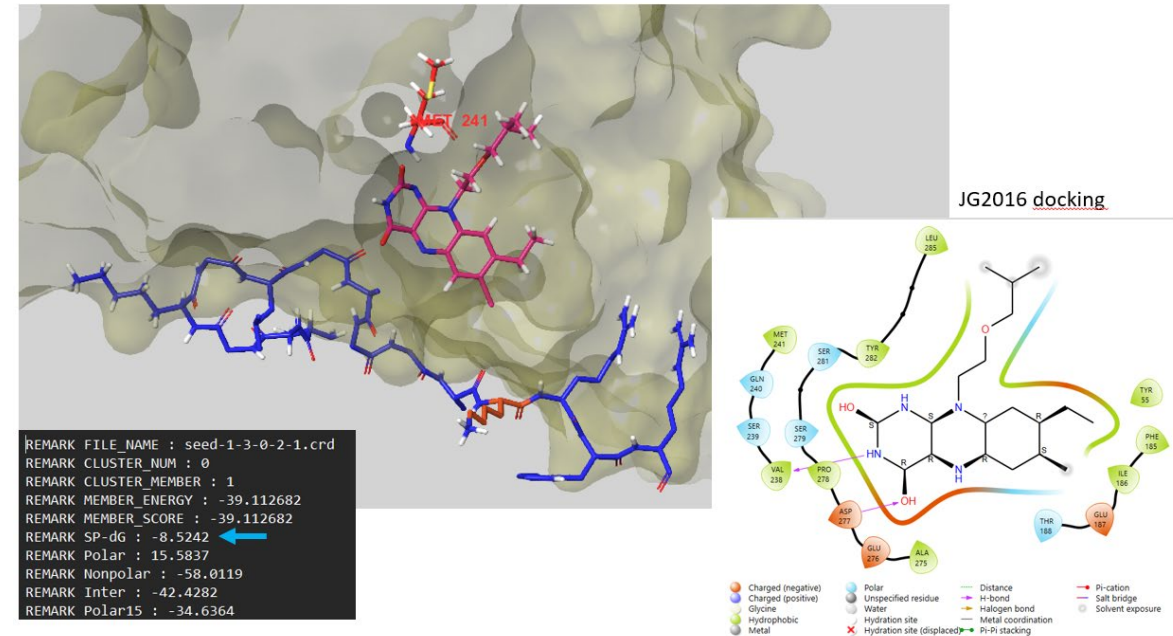
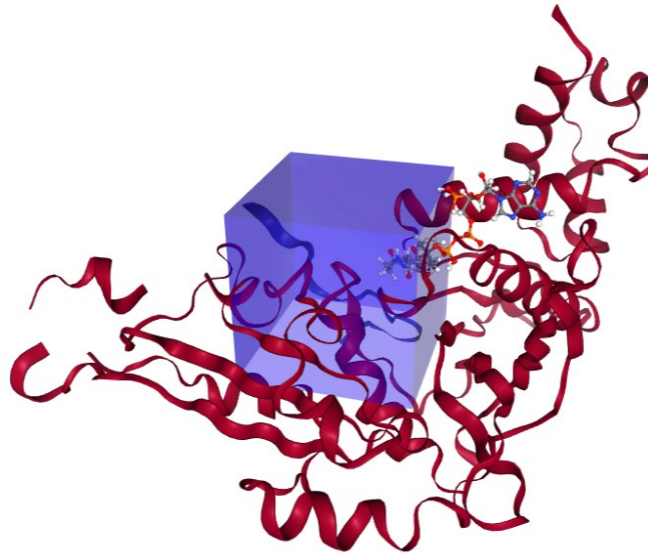
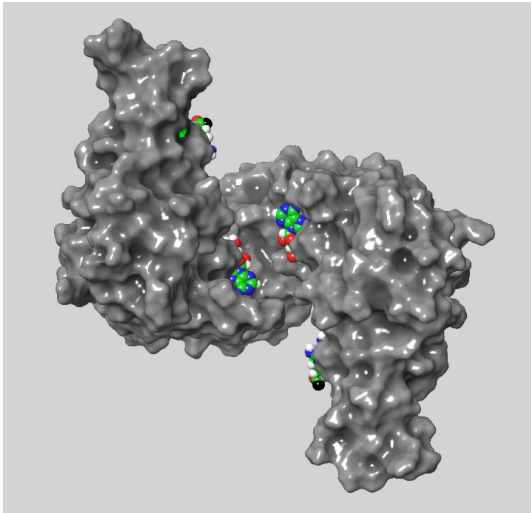
Docking



Scoring



Análisis



Prácticamente todos los programas de docking, siguen este flujo de trabajo. La diferencia radica en el algoritmo de búsqueda, la función de scoring y la interfaz de usuario.

# Preparación del receptor

<https://www.rcsb.org/>

The screenshot shows the RSCB Protein Data Bank (PDB) homepage. The header includes navigation links: RCSB PDB, Deposit, Search, Visualize, Analyze, Download, Learn, About, Careers, and COVID-19. Below the header, statistics are displayed: 245,074 Structures from the PDB archive and 1,068,577 Computed Structure Models (CSM). A search bar is present with the text 'Enter search term(s), Ligand ID or sequence'. The main content area features a 'Welcome' message, a list of structure types (Experimentally-determined 3D structures, Integrative 3D Structures, and Computed Structure Models), and a 'November Molecule of the Month' section highlighting GLP-1 Receptor Agonists. A sidebar on the left contains links for Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn.



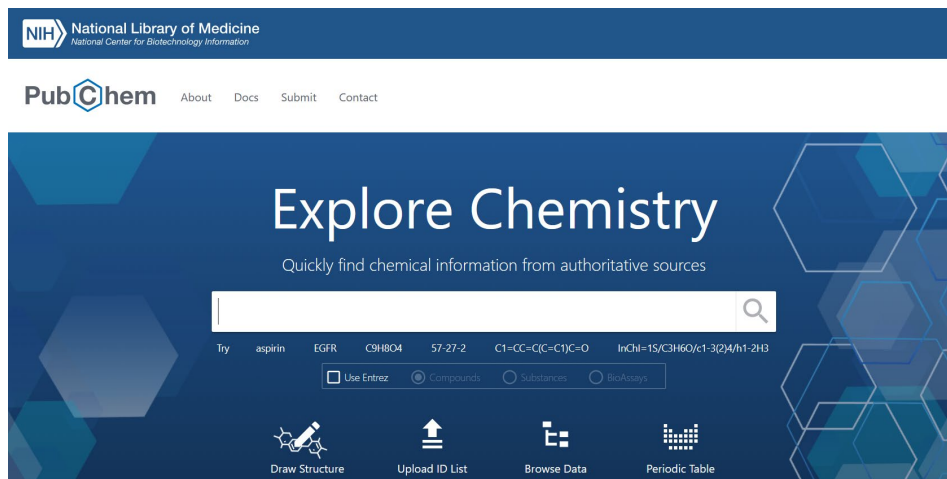
<https://alphafoldserver.com/about>

The screenshot displays the 'Protein Preparation Workflow' interface. It features three tabs: 'Preparation Workflow', 'Diagnostics', and 'Substructures'. The 'Preparation Workflow' tab is active, showing a series of steps: 'Specify Protein', 'Preprocess', 'Optimize H-bond Assignments', and 'Minimize and Delete Waters'. The 'Specify Protein' step includes a dropdown menu for 'Use structures from:' (set to 'Workspace (included entry)') and a 'Get PDB...' button. The 'Preprocess' step is checked and shows '6 actions selected' with a 'Reset' button. The 'Optimize H-bond Assignments' and 'Minimize and Delete Waters' steps also have 'Settings' buttons. At the bottom, there is a 'Job name:' field with the value 'proteinprep\_56' and a 'Run' button.

<https://www.schrodinger.com/platform/products/maestro/>



# Preparación del ligando

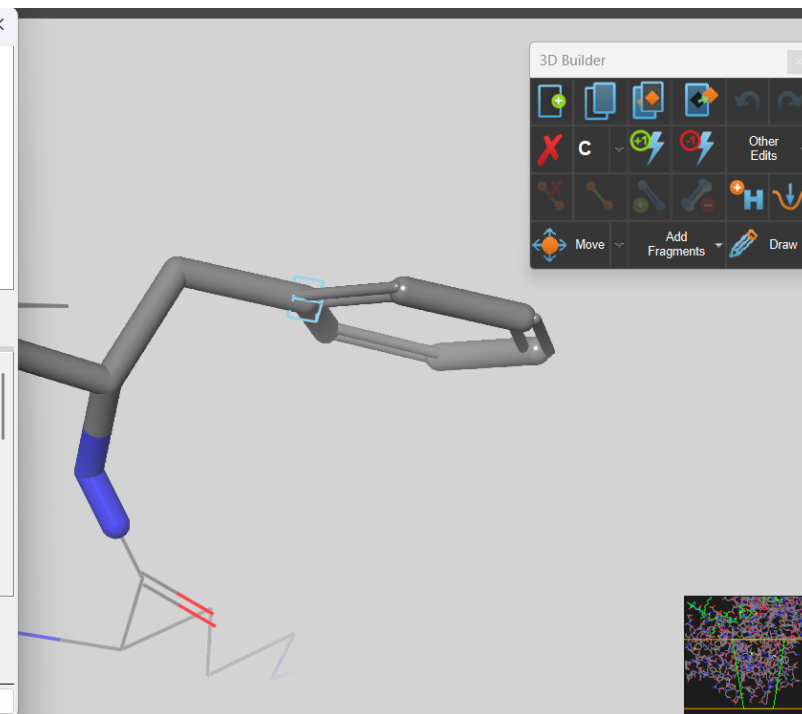


- Se puede obtener de bases de datos como **PubChem**, **ZINC**, o diseñarse manualmente.

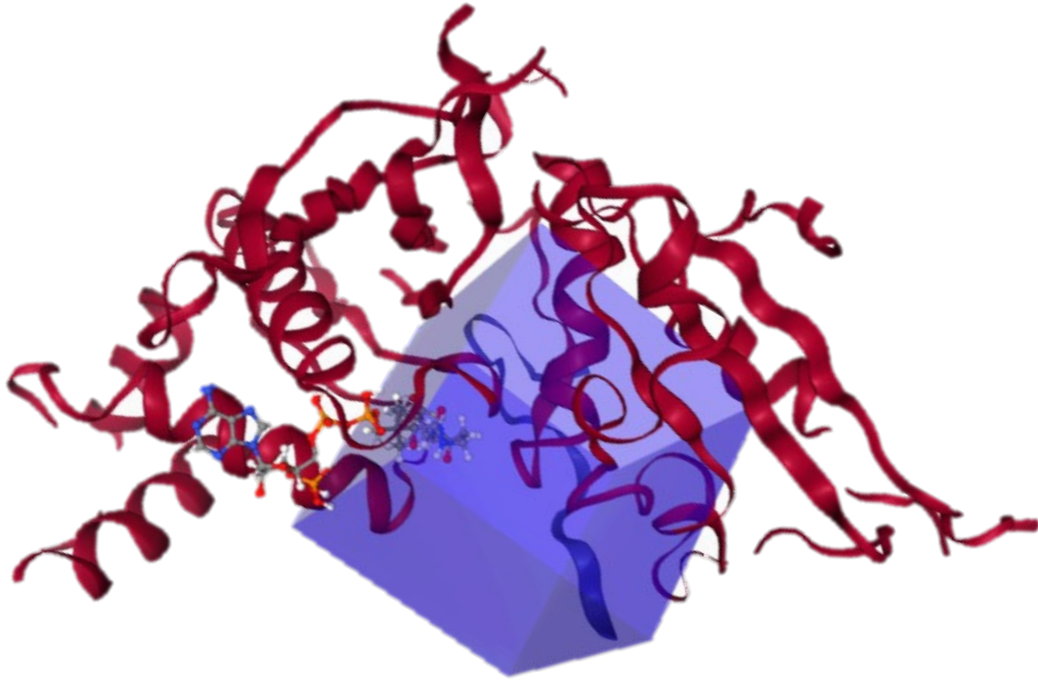
- Minimización de energía y cálculo de cargas (por ejemplo, con Open Babel).
- Definir **enlaces rotables** (grados de libertad).
- Guardar en formato PDBQT (para AutoDock).

The image shows a "Rotamers" window in a molecular modeling software. It has a "Residues" list with "A:PHE:40" selected. Below the list are "Next" and "Previous" buttons. A checkbox "Fit selected residue to Workspace" is checked. A table shows rotamer data for the selected residue. The table has columns: Rotamer, %, Chi1, Chi2, Chi3, Chi4, Chi5, Fit, and Clash. The data rows are: original, 1, 2, 3, 4, 5, 6, and 7. The "original" row is highlighted. The "Fit" column has a "+" sign for rotamers 1, 2, 3, and 4. Below the table are "Next", "Previous", and "Restore to Original Conformation" buttons. There is also an unchecked checkbox "Animate fit to density" and a text label "Choose Best Fit [Requires a single electron density map.]". At the bottom are "OK", "Cancel", and "Help" buttons.

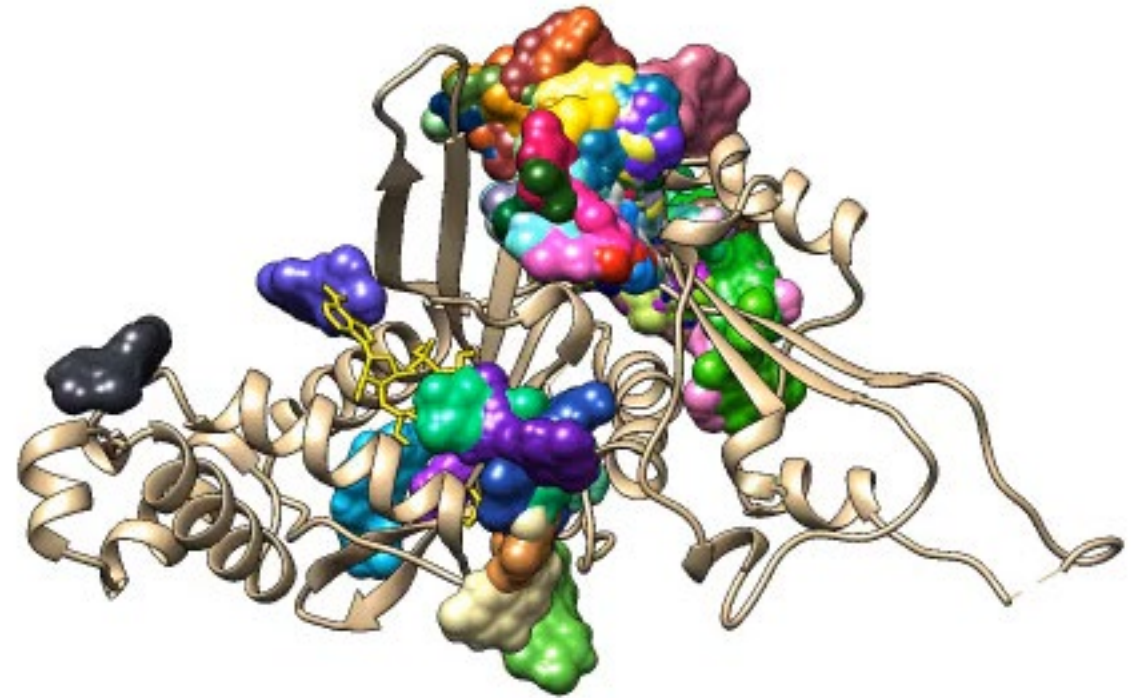
Rotamer	%	Chi1	Chi2	Chi3	Chi4	Chi5	Fit	Clash
original		-66	-88					
1	30.8	-67	98					+
2	20.7	-179	78					+
3	7.5	180	60					+
4	7.2	63	90					+
5	6.0	-60	-60					
6	5.7	-60	-90					
7	5.3	-68	-15					



# Definición del sitio de *docking*



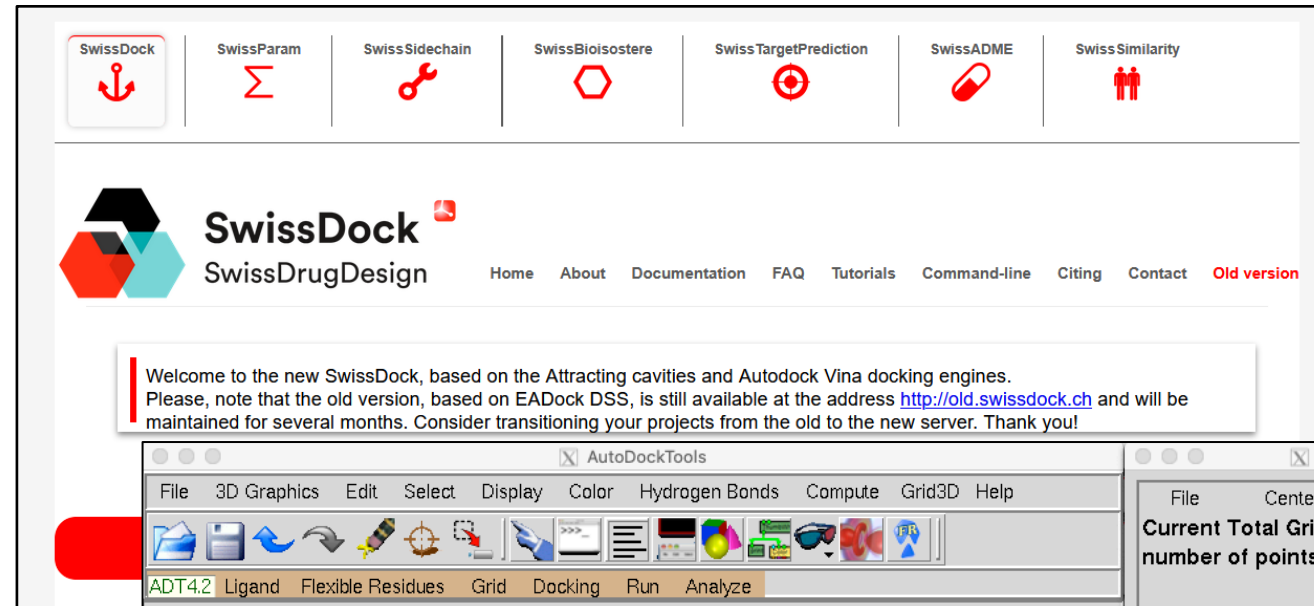
Dirigido -> Grid Box



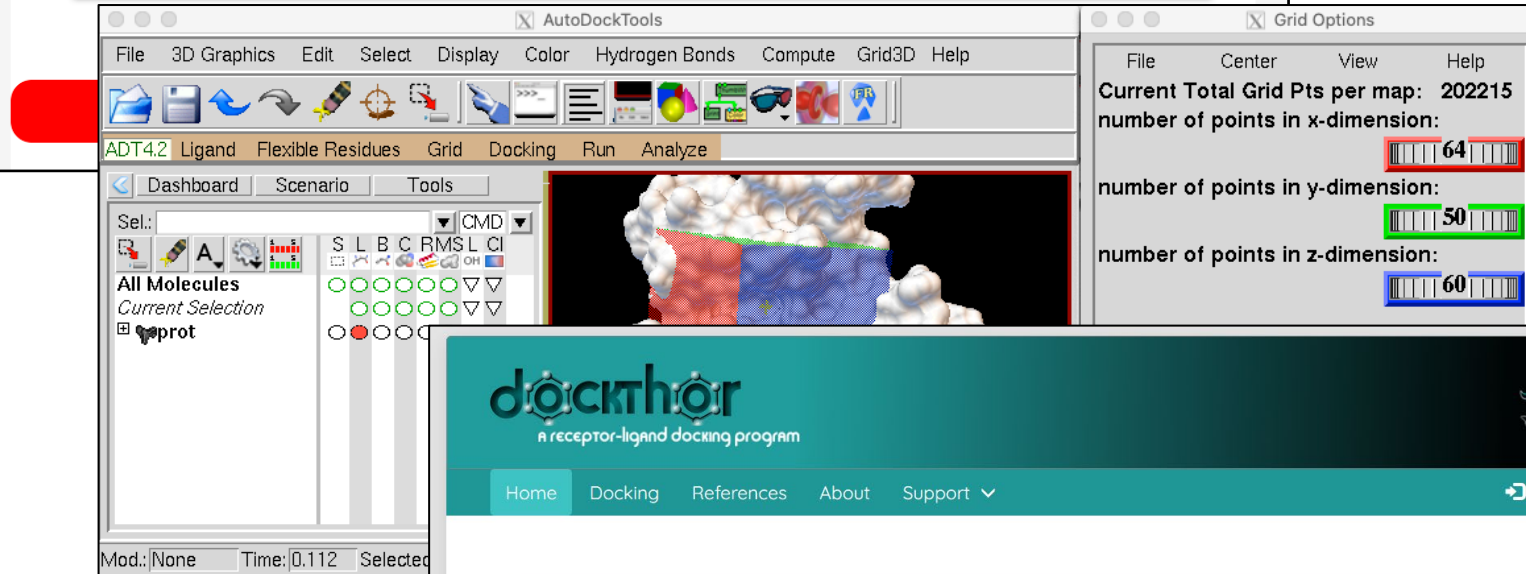
Blind docking

# Ejecución

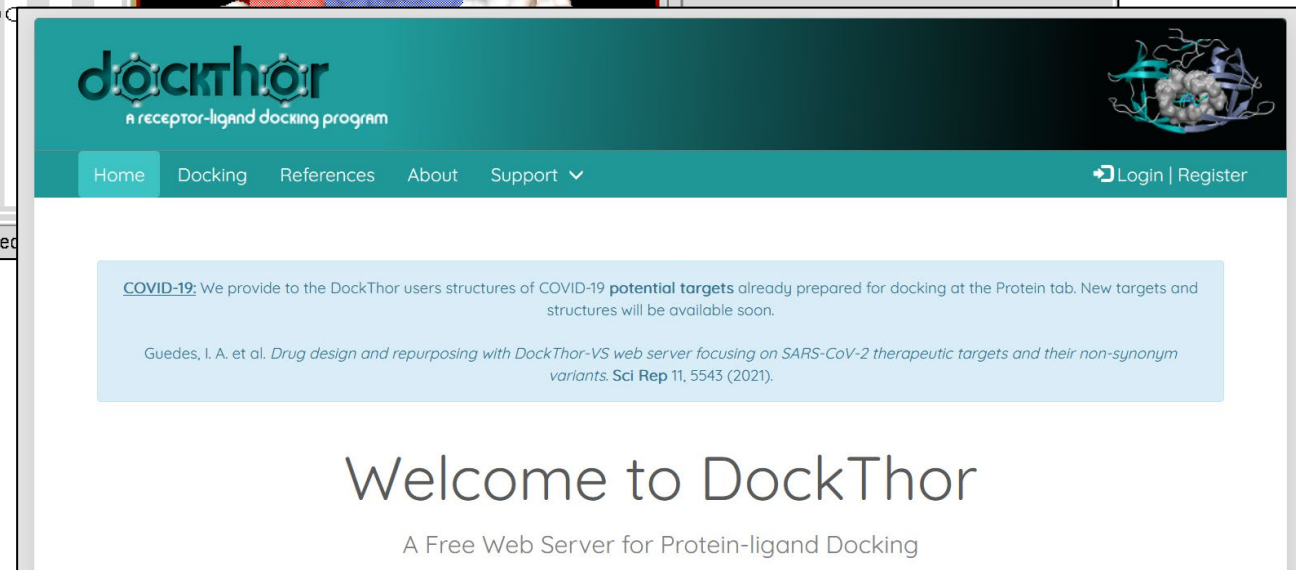
<https://www.swissdock.ch/>



AutoDockTools



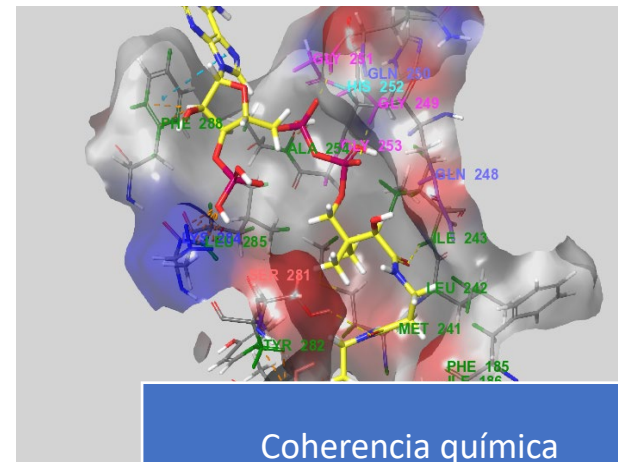
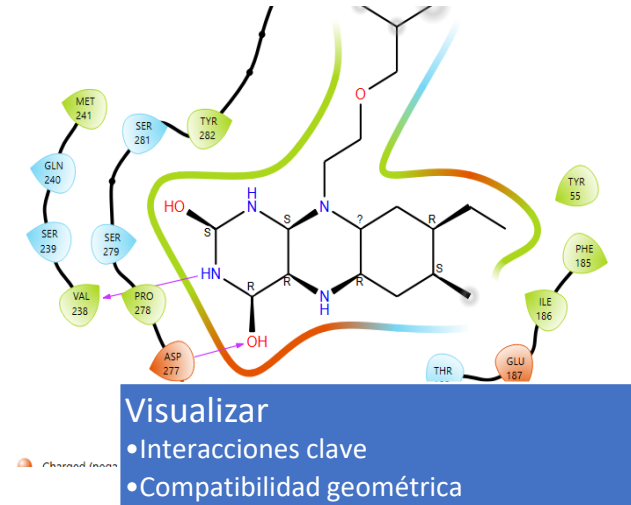
<https://dockthor.lncc.br/v2/>



# Interpretación de resultados

cluster	Average Member Score	Average SP-dG	Member Count	representar
1	15,56455	-6,5188	48	pose_0_1
0	14,83741	-6,34828	21	pose_2_1
2	19,36506	-6,39935	15	pose_23_1
3	15,61145	-6,37746	10	pose_8_1
6	14,77426	-6,45042	9	pose_6_1
5	16,69382	-6,26173	4	pose_5_1
7	18			
4	14			

Seleccionar las mejores poses



# Validación del protocolo

## Redocking

- Volver a acoplar el ligando original del cristal y comprobar  $\text{RMSD} < 2 \text{ \AA}$ .

## Cross-docking

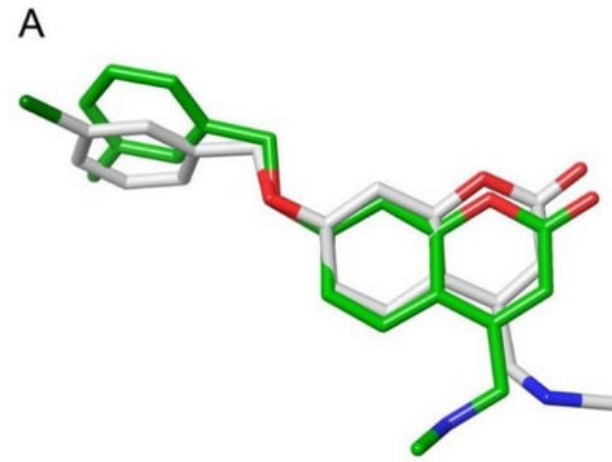
- Usar varios complejos del mismo receptor. Usar distintos algoritmos.

## Benchmarking

- Comparar con datos experimentales de afinidad ( $\text{IC}_{50}$ ,  $\text{Kd}$ ).

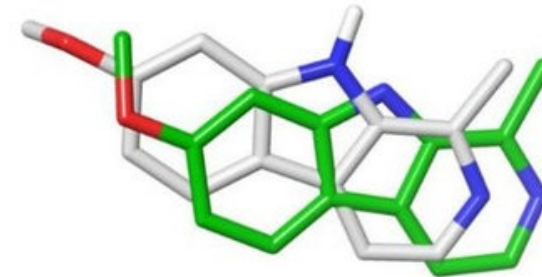
## Validación visual

- Comprobar que las interacciones son plausibles.



$\text{RMSD} = 1.81 \text{ \AA}$

B



$\text{RMSD} = 1.47 \text{ \AA}$

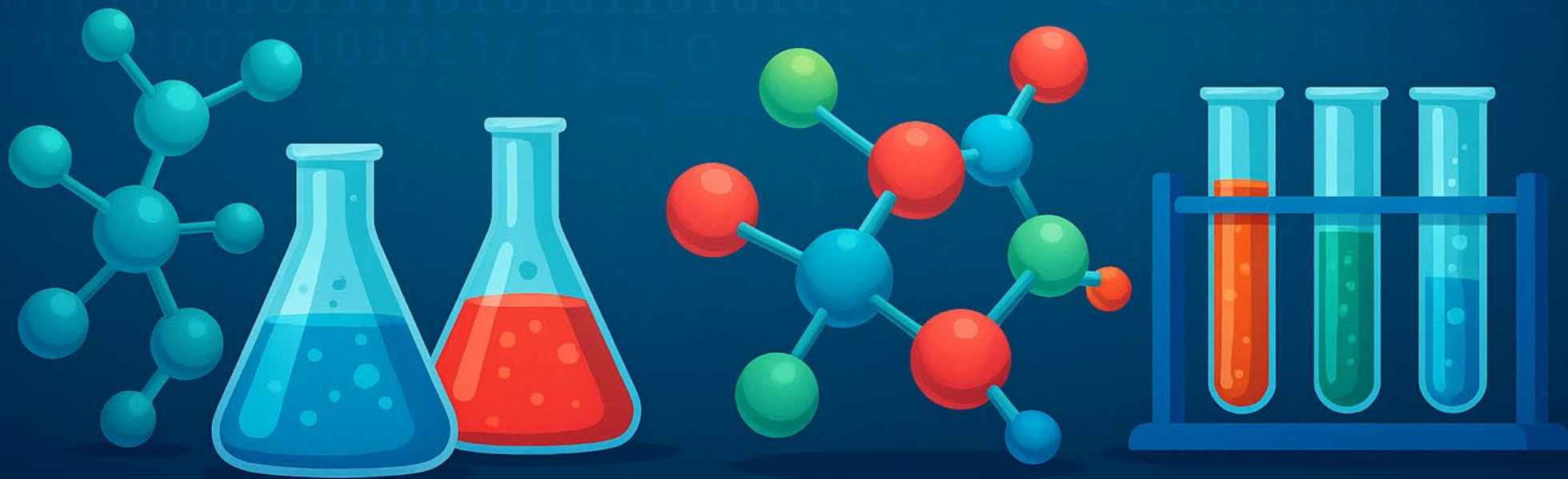
**RMSD:** Root Mean Square Deviation o Desviación Cuadrática Media de la Raíz. Mide la distancia promedio entre los átomos correspondientes de dos estructuras superpuestas.



Gracias por su atención



# EJEMPLO PRÁCTICO DE DOCKING MOLECULAR



# Dudas

