

Introducción al *docking* molecular



Cristina Coscolín Galán

cristinacoscolin@gmail.com

www.linkedin.com/in/cristinacoscolin

010010101



Universidad
de Alcalá

0101010

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

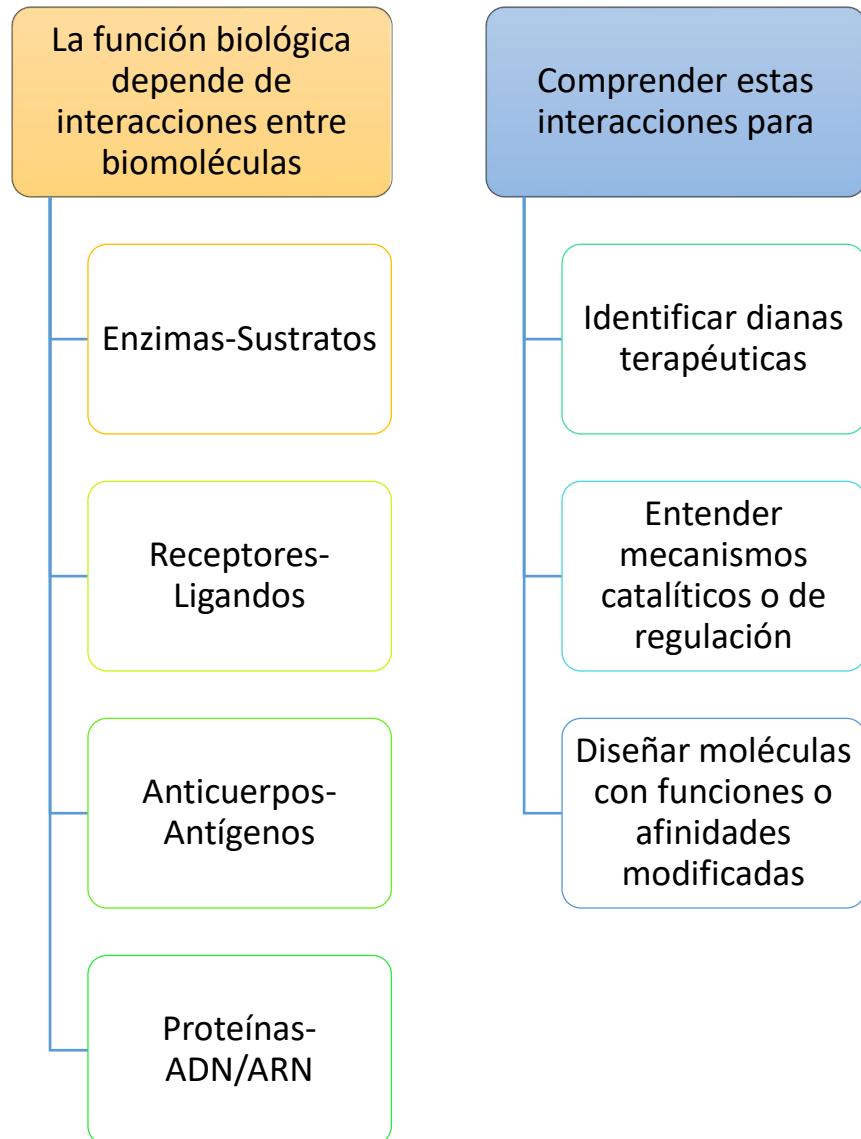
- Principales programas (AutoDock, Vina, SwissDock)
- Visualización de resultados (Chimera, PyMol, Maestro)
- Pipeline típico

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?

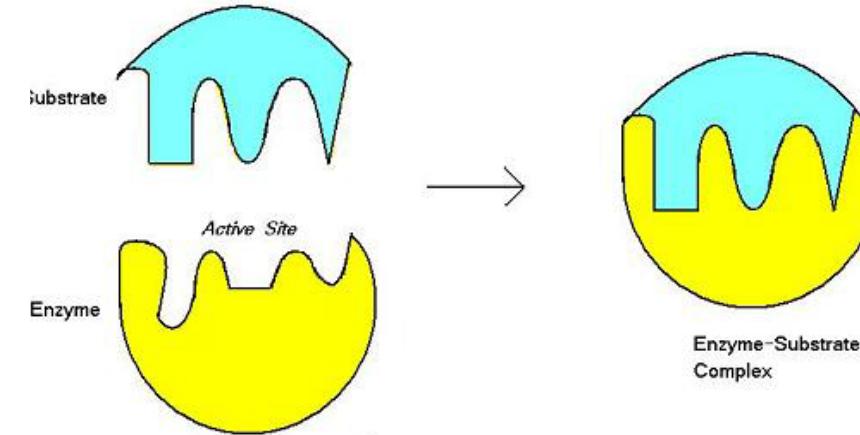
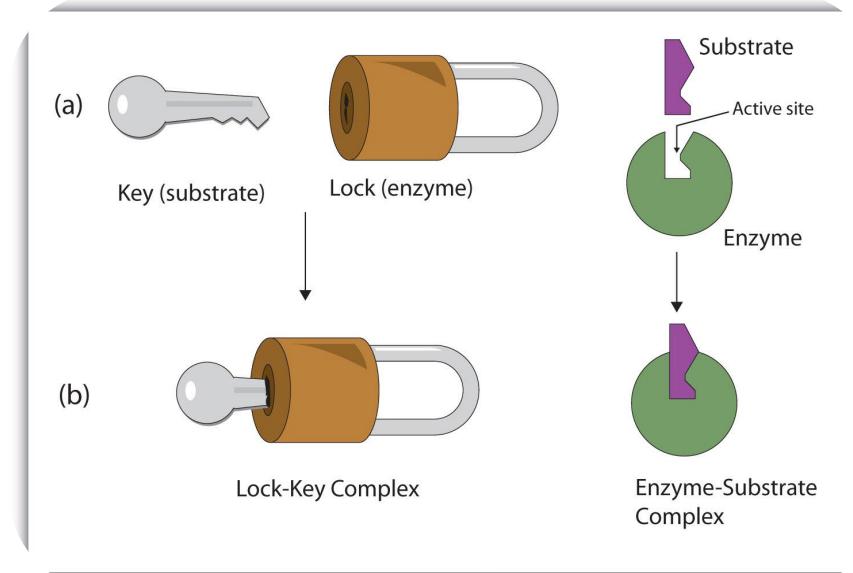
- El *docking molecular* es un método computacional que predice la **orientación óptima (pose)** y la **afinidad de unión** de una molécula pequeña (*ligando*) al unirse con una macromolécula (*receptor*), generalmente una proteína.

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



Induced-fit Model. - The enzyme active site forms a complementary shape to the substrate after binding.

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

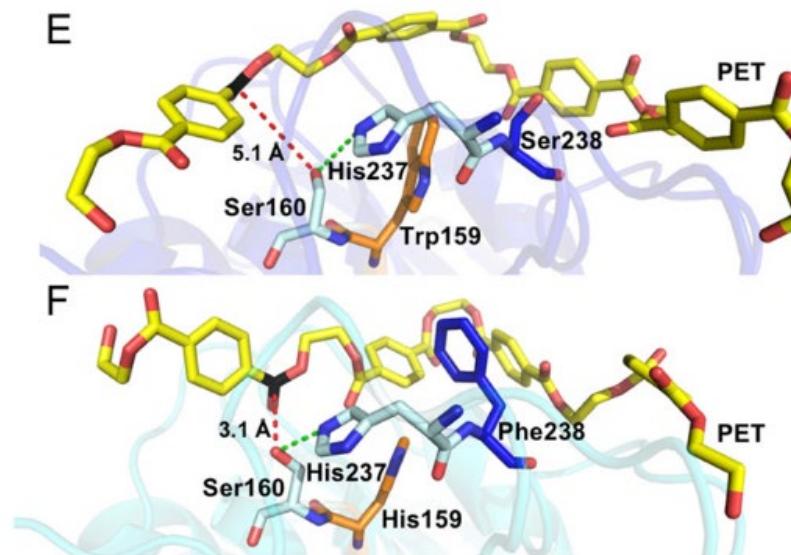
Biología estructural: apoyo a la interpretación de datos experimentales (cristalografía, NMR, cryo-EM)

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



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

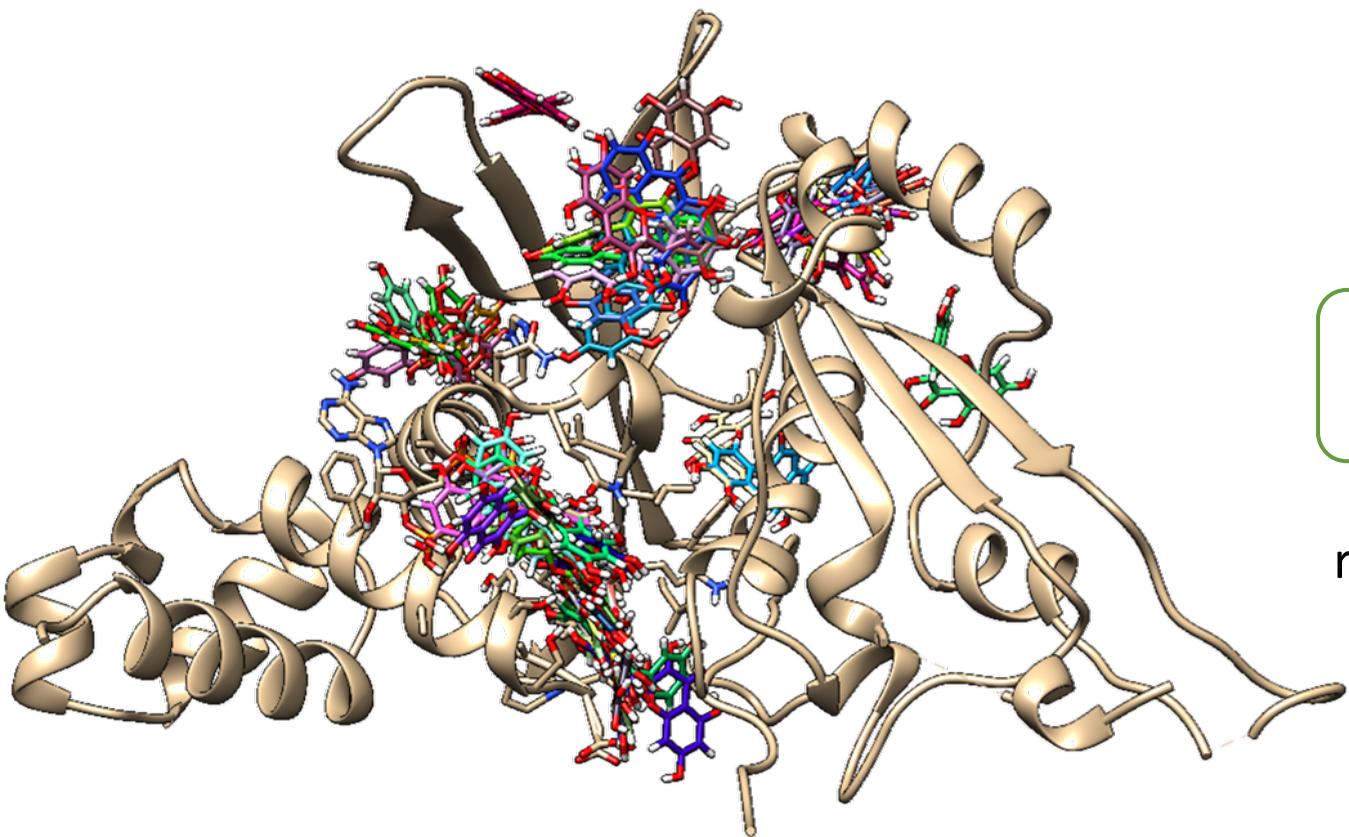
Ligando:
molécula pequeña, péptido o metabolito que se une al receptor

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



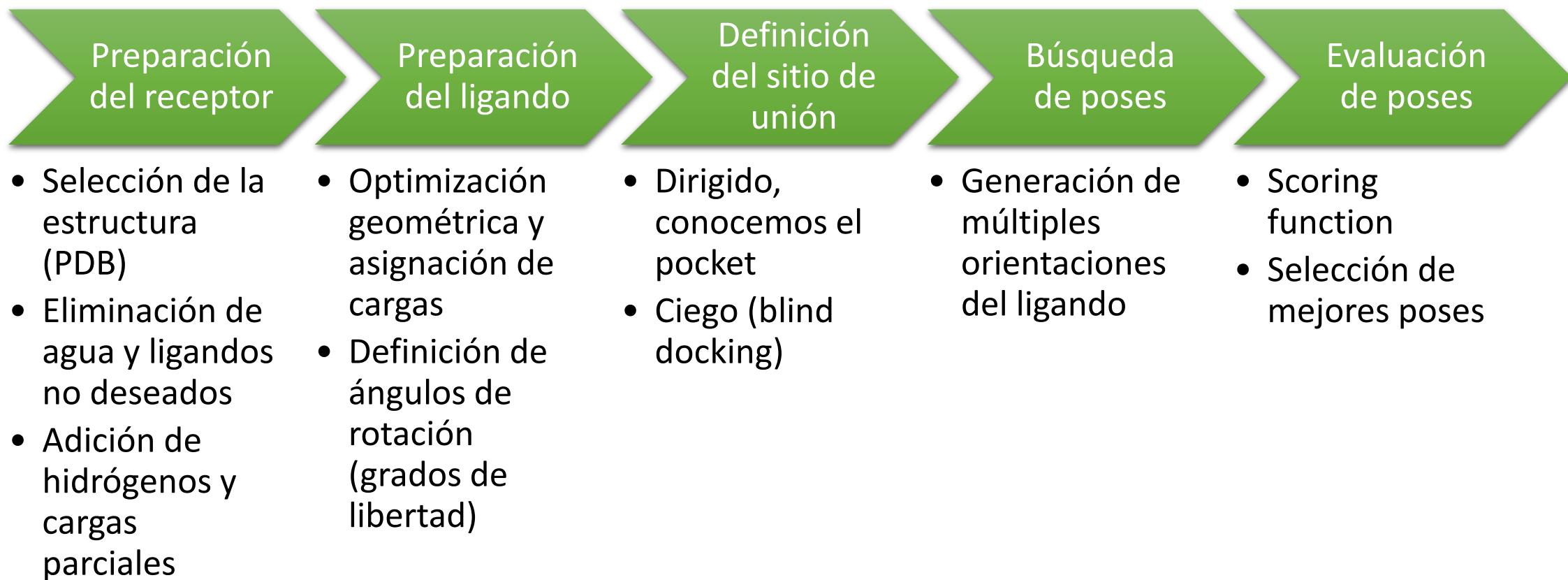
Rígido:

- 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

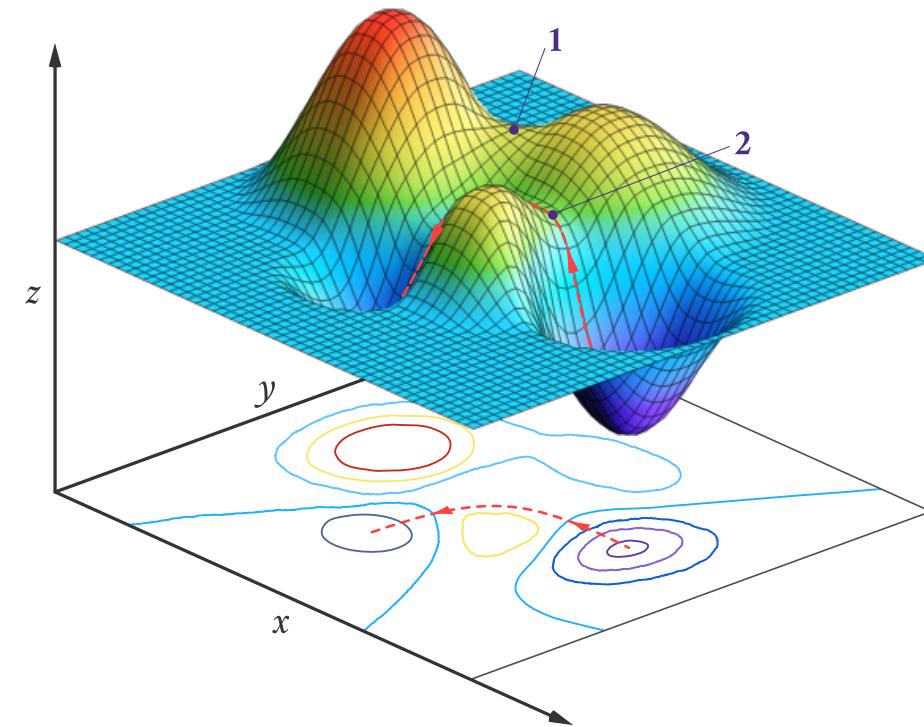
- En *docking molecular*, el objetivo es **encontrar la conformación (pose)** del ligando que **minimiza la energía de interacción** con el receptor.
- Esto equivale a buscar el **mínimo global** en un “paisaje energético” muy complejo (con muchos mínimos locales).

Métodos deterministas:

- Grid search
- Incremental construction

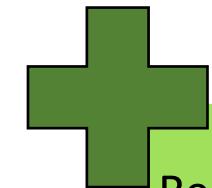
Métodos estocásticos

- 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

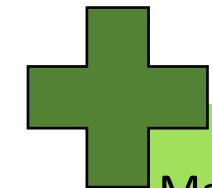
- Se exploran posiciones y orientaciones discretas en una malla

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

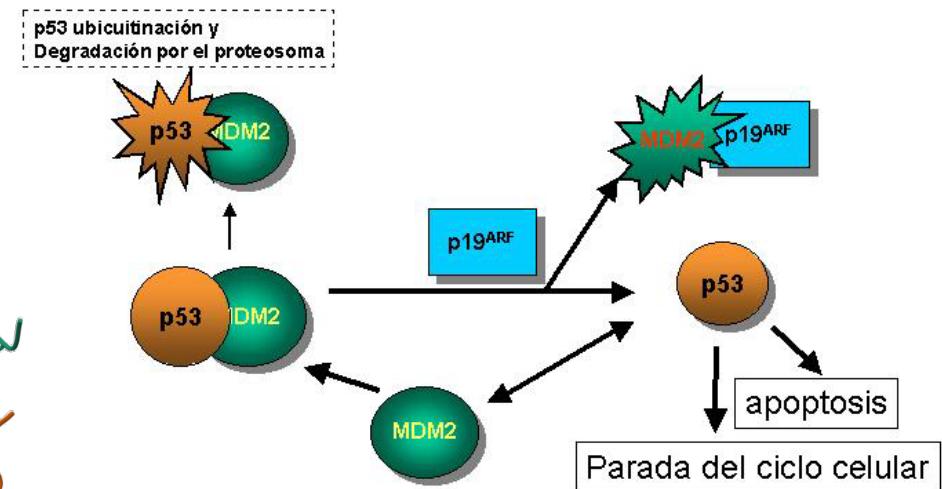
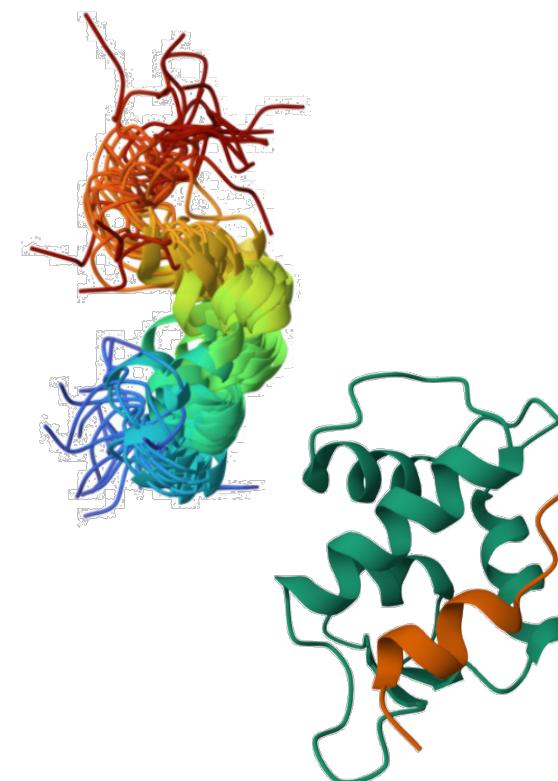
- Simula “enjambres” de partículas buscando colectivamente el mejor punto

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 (ΔG_{bind})**.
- El objetivo es identificar la **pose más estable** (energía mínima) y **hacer un “ranking”** los ligandos según su afinidad teórica.

Cuánto 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



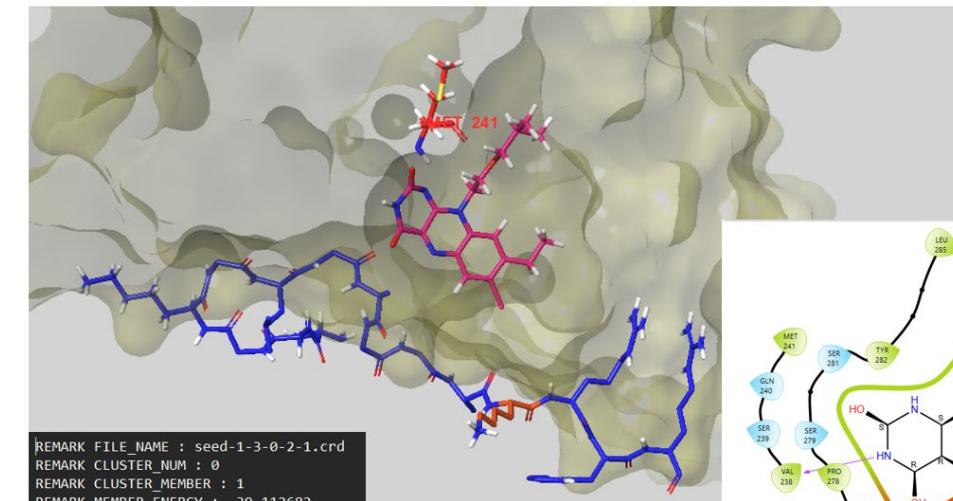
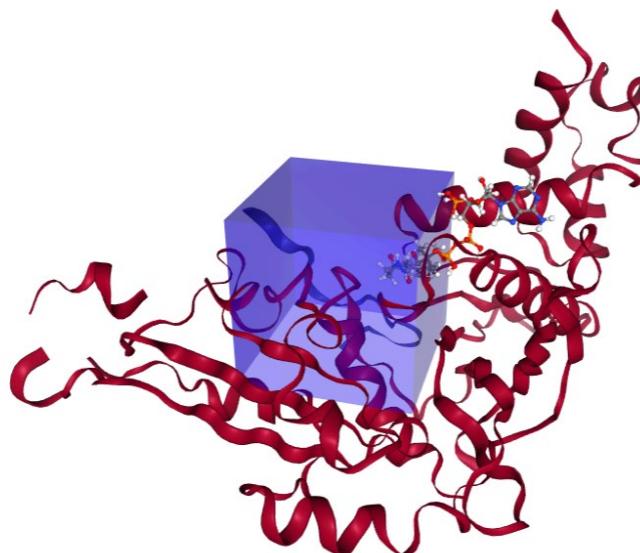
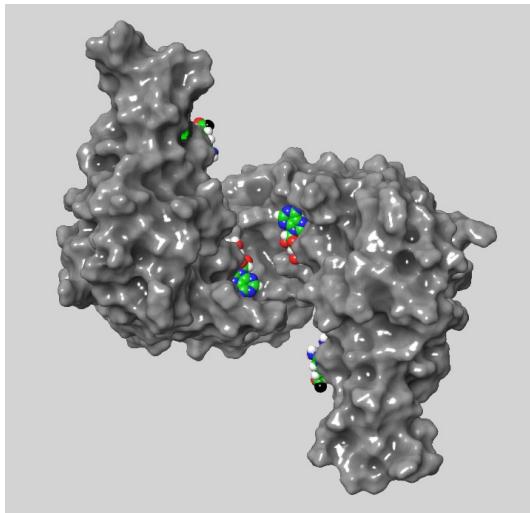
Docking



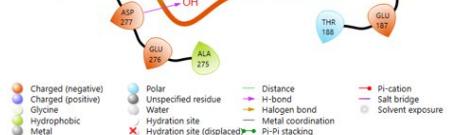
Scoring



Análisis



JG2016 docking



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 RCSB PDB homepage. The top navigation bar includes links for RCSB PDB, Deposit, Search, Visualize, Analyze, Download, Learn, About, Careers, and COVID-19. The main search bar allows users to enter search terms, Ligand ID or sequence, and includes an 'Include CSM' toggle and a search button. Below the search bar, there are links for PDB-101, wwPDB, EMDDataResource, NAKB, wwPDB Foundation, and PDB-IHM. A sidebar on the left provides links for Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. A central banner highlights 'Redesigned PDB Statistics Support Enhanced Functionality' and features sections for 'RCSB Protein Data Bank (RCSB PDB)', 'Experimentally-determined 3D structures', 'Integrative 3D Structures', and 'Computed Structure Models (CSM)'. A 'Molecule of the Month' section for November features four 3D molecular models of GLP-1 Receptor Agonists.

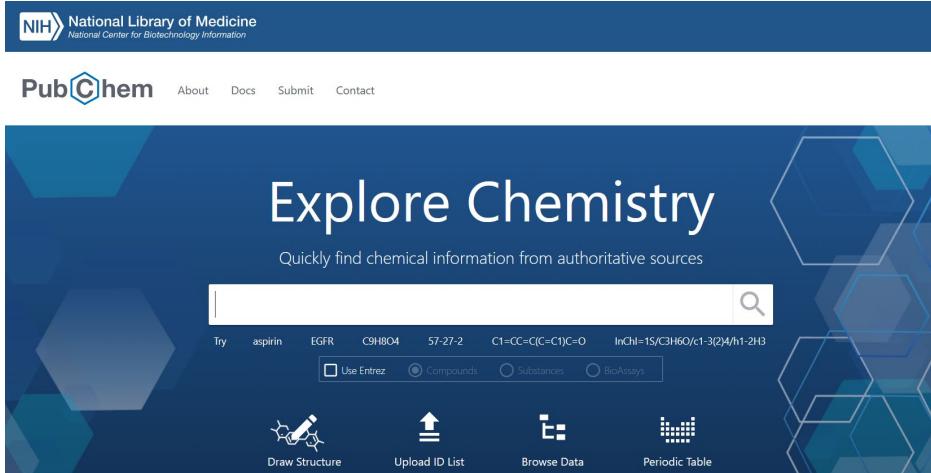


<https://alphafoldserver.com/about>

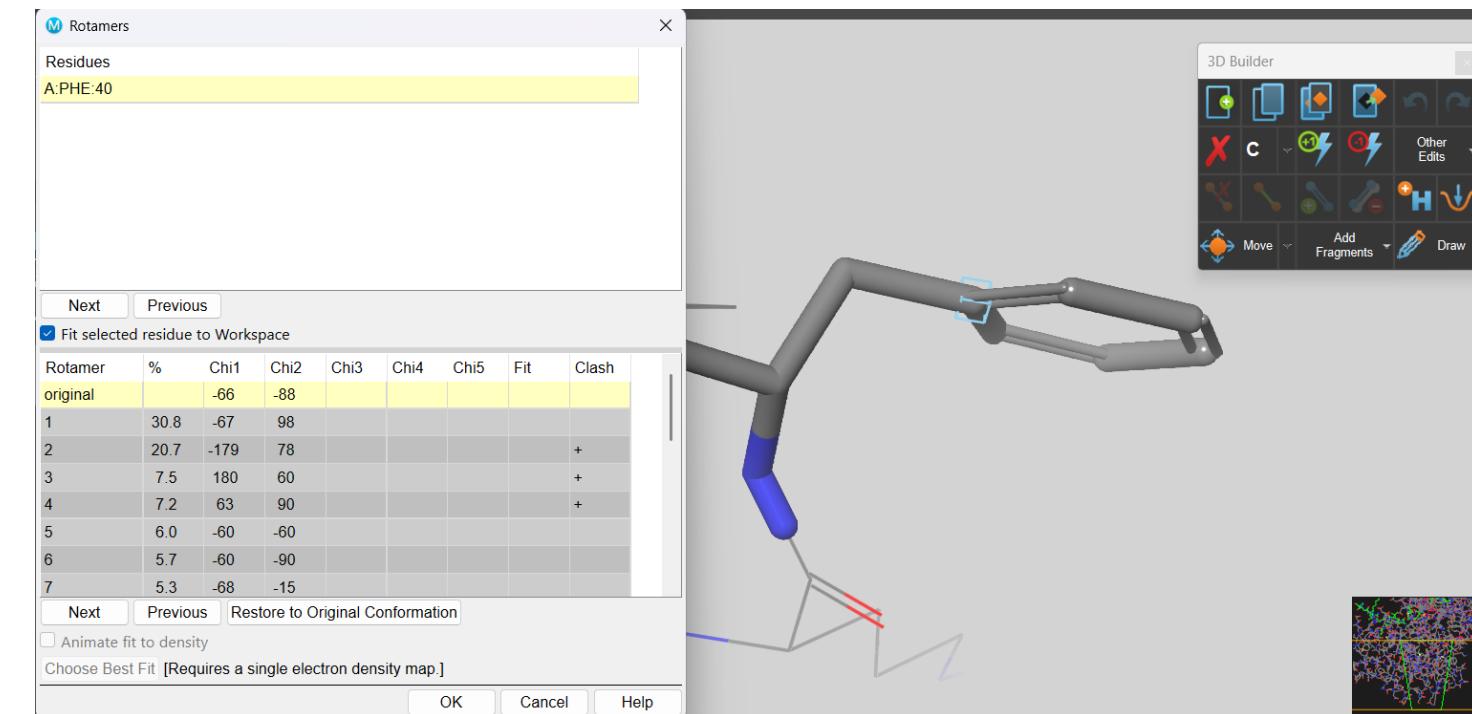
The screenshot shows the 'Protein Preparation Workflow' interface. It includes tabs for Preparation Workflow, Diagnostics, and Substructures, with 'Preparation Workflow' selected. A note says 'Toggle steps on to include them in the job.' and a 'INTERACTIVE' button. The 'Specify Protein' section shows a dropdown for 'Workspace (included entry)' and a 'Get PDB...' button. An 'Entry: (more than one entry)' message is displayed. The 'Review Structure' and 'Global Settings' buttons are also present. The 'Preprocess' section is expanded, showing options for 'Cap termini' (unchecked) and '6 actions selected' with a 'Reset' button. The 'More Options' button is available. The 'Optimize H-bond Assignments' and 'Minimize and Delete Waters' sections are also shown with their respective descriptions and 'Settings' buttons.

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

Preparación del ligando

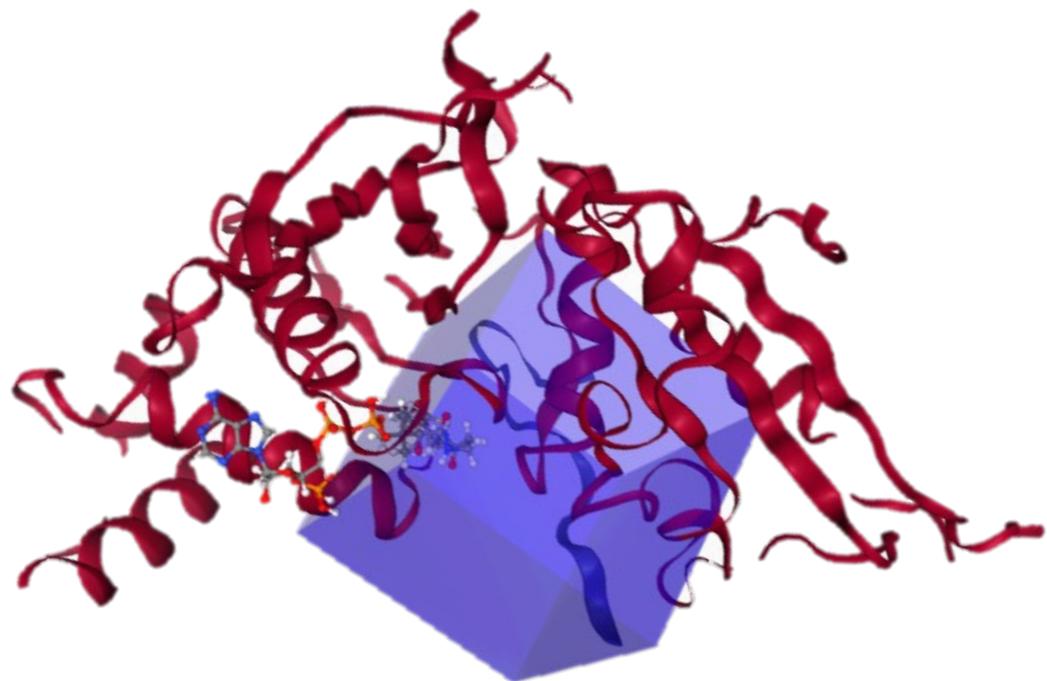


The image shows the homepage of PubChem, a chemical information database. The header includes the NIH National Library of Medicine logo and the PubChem logo. The main title is "Explore Chemistry" with the subtitle "Quickly find chemical information from authoritative sources". Below the title is a search bar with the placeholder "Try aspirin, EGFR, C9H8O4, 57-27-2, C1=CC=C(C=C1)C=O, InChI=1S/C3H6O/c1-3(2)4/h1-2H3". Navigation links include "About", "Docs", "Submit", and "Contact". Below the search bar are buttons for "Draw Structure", "Upload ID List", "Browse Data", and "Periodic Table". The background features a blue hexagonal pattern.

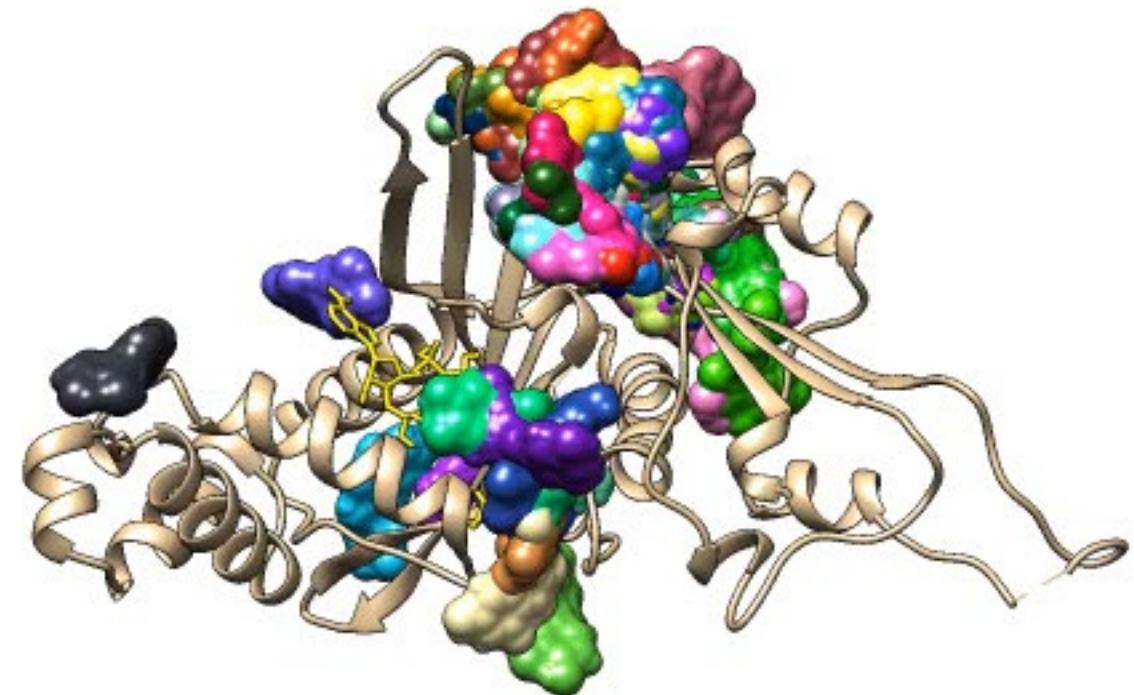


- Se puede obtener de bases de datos como **PubChem**, **ZINC**, o diseñarse manualmente.
 - Minimización de energía y cálculo de cargas (por ejemplo, con Open Babel).
 - Definir **enlaces rotables** (grados de libertad).
 - Guardar en formato PDBQT (para AutoDock).

Definición del sitio de *docking*



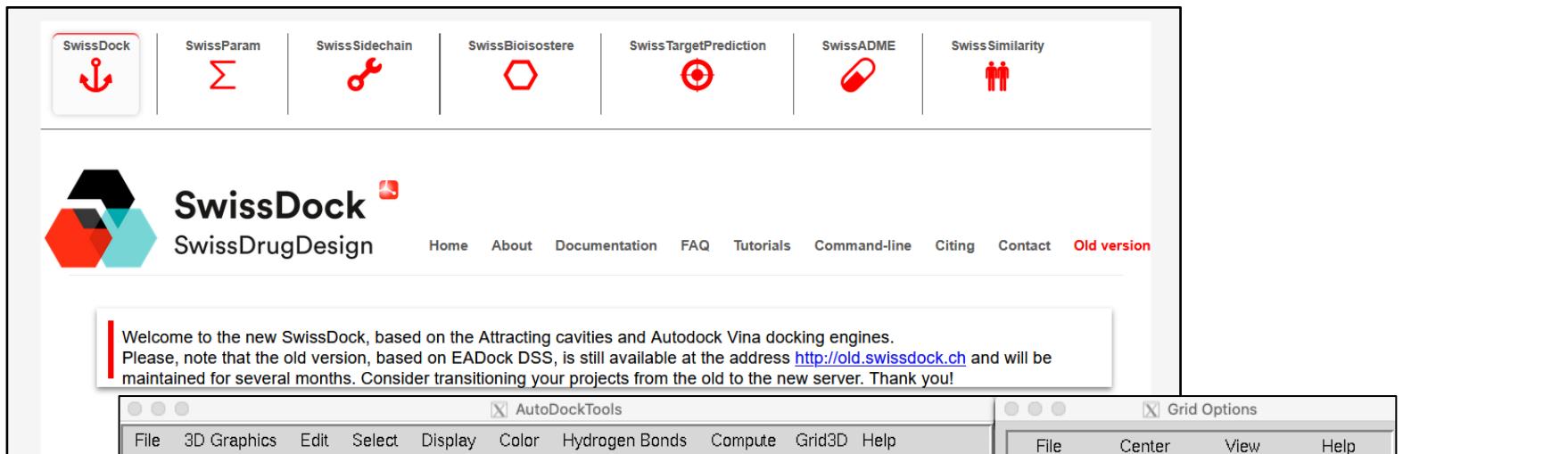
Dirigido -> Grid Box



Blind docking

Ejecución

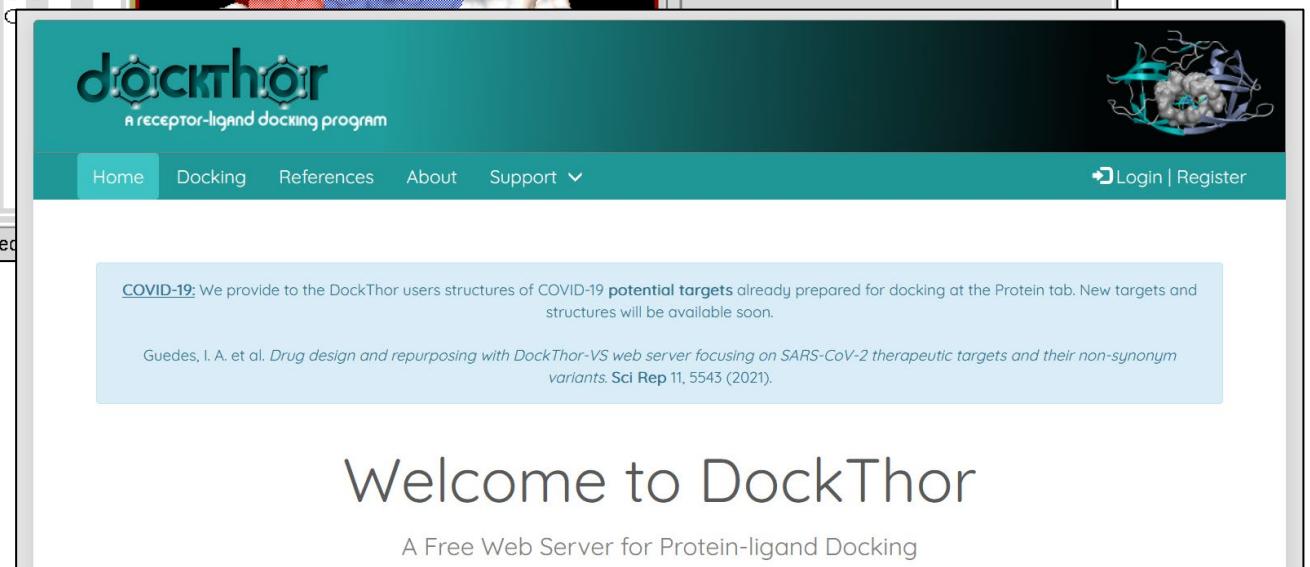
<https://www.swissdock.ch/>



AutoDockTools

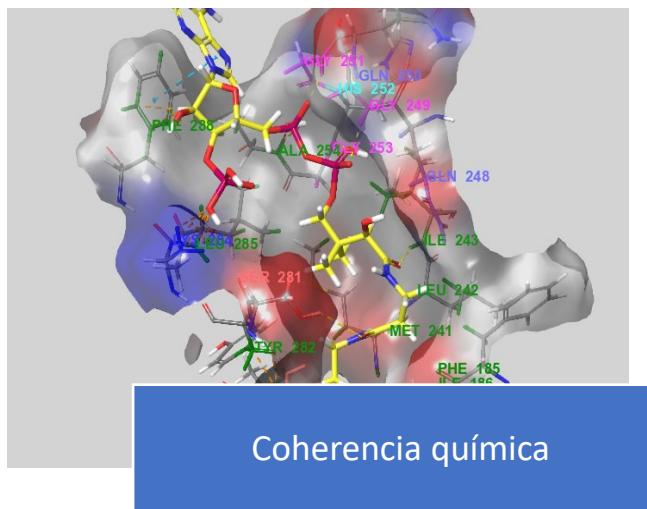
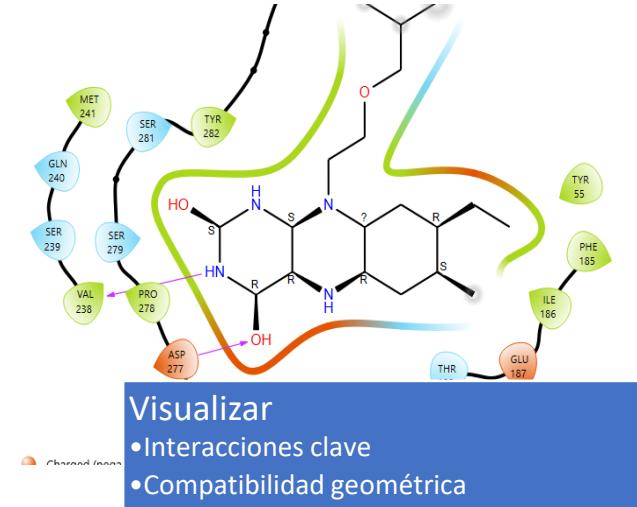


<https://dockthor.lncc.br/v2/>



Interpretación de resultados

luster	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			



Validación del protocolo

Redocking

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

Cross-docking

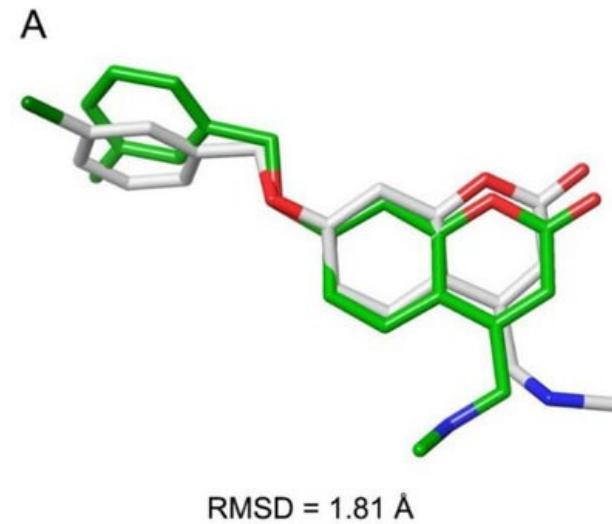
- Usar varios complejos del mismo receptor. Usar distintos algoritmos.

Benchmarking

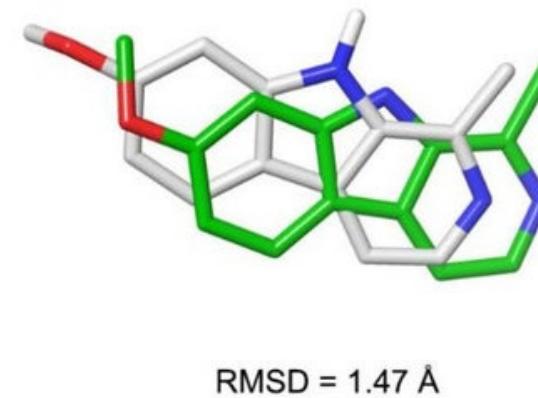
- Comparar con datos experimentales de afinidad (IC_{50} , Kd).

Validación visual

- Comprobar que las interacciones son plausibles.



B



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

010010101

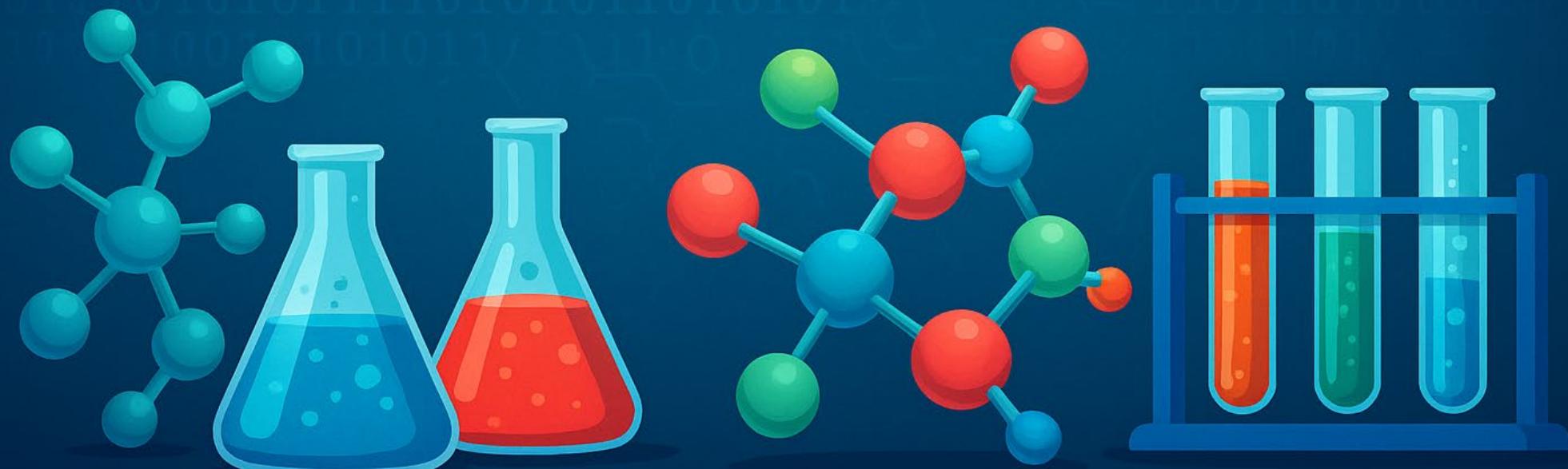
Gracias por su atención



Universidad
de Alcalá

001010

EJEMPLO PRÁCTICO DE DOCKING MOLECULAR



Dudas

