

Molecular Dynamics Simulations: Protein-Ligand Systems

PhD. Conrado Pedebos
PhD. Pablo Ricardo Arantes

Porto Alegre, July 15th 2025

Proteína - Ligante

Summary of Course Flow:

Day 1 - Introduction to Molecular Dynamics

Day 2 - Protein-Ligand Systems

Day 3 - Coarse-Grained Molecular Dynamics and Membrane Proteins

Day 4 - Alternatives to Classical Molecular Dynamics for Conformational Sampling

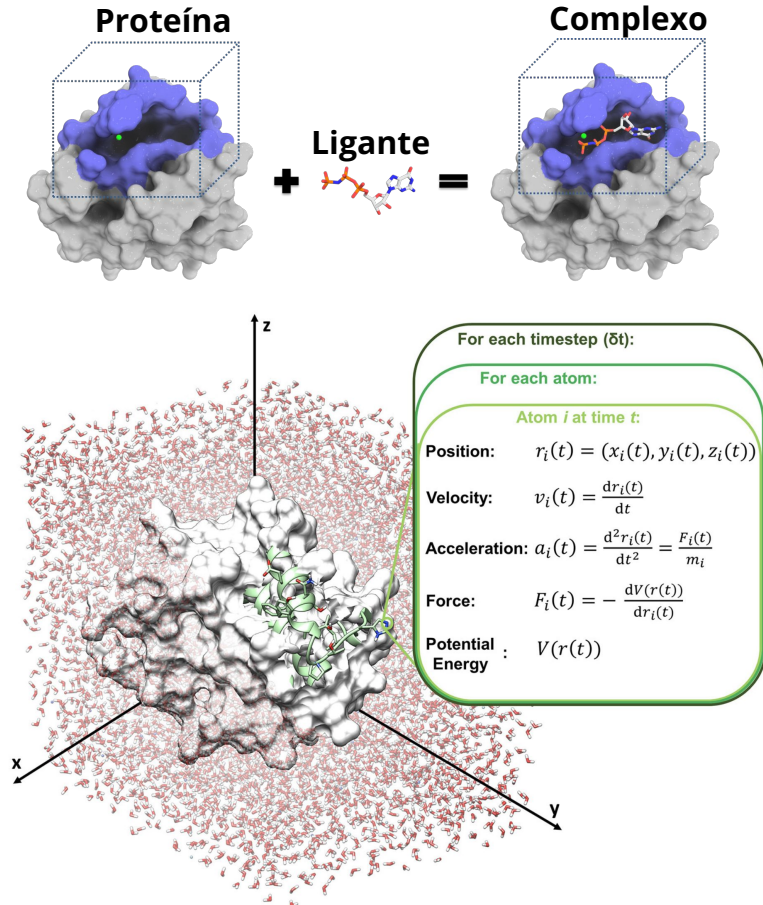
Day 5 - Enhanced Sampling Methods

Slides + Notebooks



Proteína - Ligante

- Técnicas computacionais têm sido aplicadas na descoberta de fármacos desde a década de 1980.
- Primeiras estratégias de modelagem molecular baseavam-se em uma visão rígida do processo de ligação entre o ligante e o alvo.
- Neste módulo do curso, apresentaremos duas técnicas muito utilizadas para modelagem molecular, atracamento (*docking*) molecular e dinâmica molecular do complexo (proteína - ligante).



Proteína - Ligante

Busca tentar **predizer a posição e orientação** de moléculas com seus receptores. A predição da interação é baseada nas possíveis posições, orientações e conformações que uma molécula pode assumir dentro do sítio ativo, ou cavidade de ligação, de um determinado alvo.

- Atracamento;
- *Docking*;
- Ancoramento;

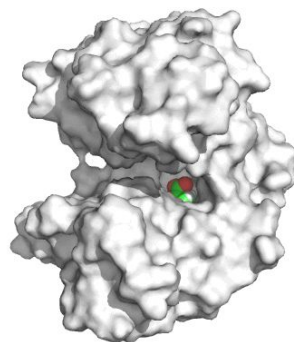
**Algoritmo de
Busca**

Predizer as diversas
possibilidades de conformações
para um ligante.

+

**Função de
Avaliação**

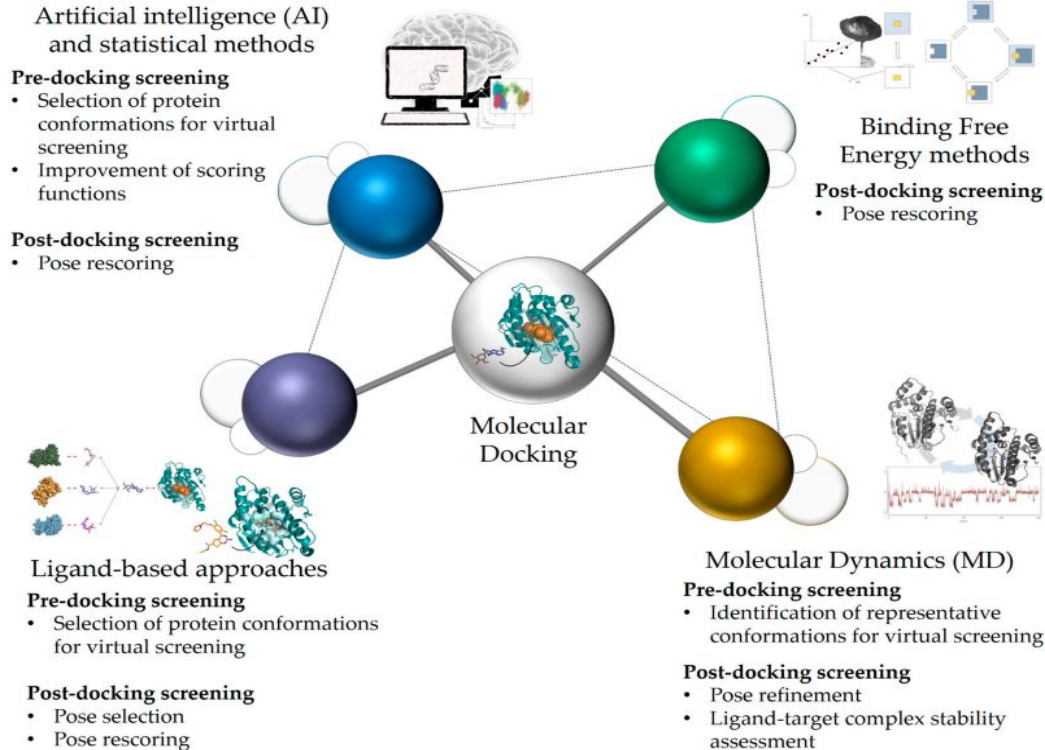
Quantifica a **qualidade**
das poses predita.



Proteína - Ligante

- O *Docking Molecular* (ou atracamento molecular) define-se como uma técnica computacional que visa auxiliar na compreensão dos mecanismos de interação entre macromoléculas como proteínas ou delas com pequenas moléculas, os ligantes.

Proteína - Ligante



Proteína - Ligante

Avaliando a Qualidade das Poses de Docking

No docking molecular, a geração de uma pose de ligação é apenas o primeiro passo. O desafio crucial é avaliar a qualidade e a relevância biológica dessas poses.

- Por que avaliar?
 - Identificar ligantes de alta afinidade (*hit identification*).
 - Otimizar a afinidade de ligação na otimização de *hits* e *leads*.
 - Filtrar estruturas não realísticas e prever o resultado de pequenas modificações químicas.

Proteína - Ligante

- Estrutura inicial: proveniente de dados experimentais (NMR, cristalografia de raio-X crio-microscopia eletrônica) ou proveniente de modelos (AlphaFold2, Modeller, *Docking*).
- No presente curso, utilizaremos um primeiro notebook para obtermos o modelo do complexo proteína-ligante, que posteriormente será utilizado no próximo notebook para a simulação de DM do complexo.

Docking Notebook



MD simulations Notebook



Proteína - Ligante

1º modelo proposto para explicar
reações catalisadas por enzima

- Hermann Emil Fischer 1894;
- Especificidade enzimática (ação enzimática);
- Complementaridade geométrica e físico-química;
- Formato correto para servir;
- Considera a estrutura e a conformação como rígidas;

**1902 - Nobel Prize in Chemistry for his
work on sugar and purine synthesis.**

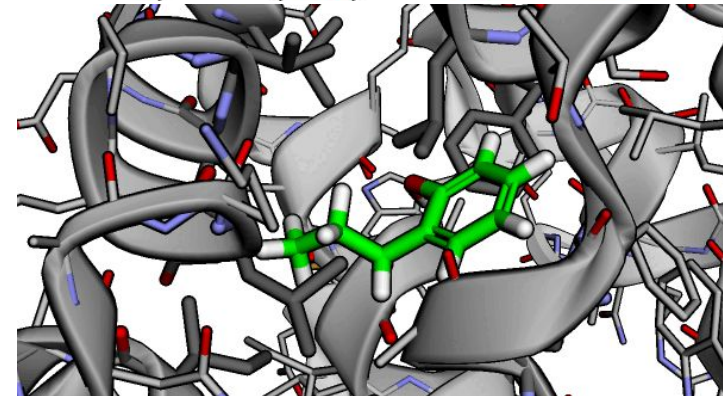
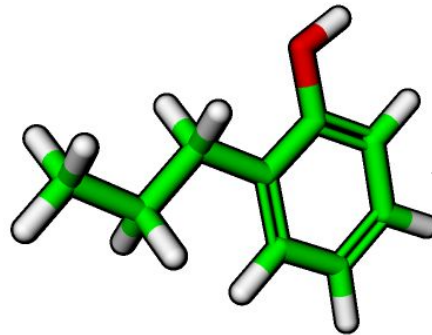
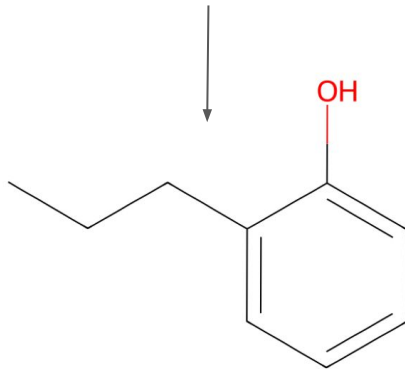


Emil Fischer – Nobel Lecture.
NobelPrize.org. Nobel Prize
Outreach AB 2022. Sun. 2 Jan
2022.

[https://www.nobelprize.org/
prizes/chemistry/1902/fischer/
lecture/](https://www.nobelprize.org/prizes/chemistry/1902/fischer/lecture/)

Proteína - Ligante

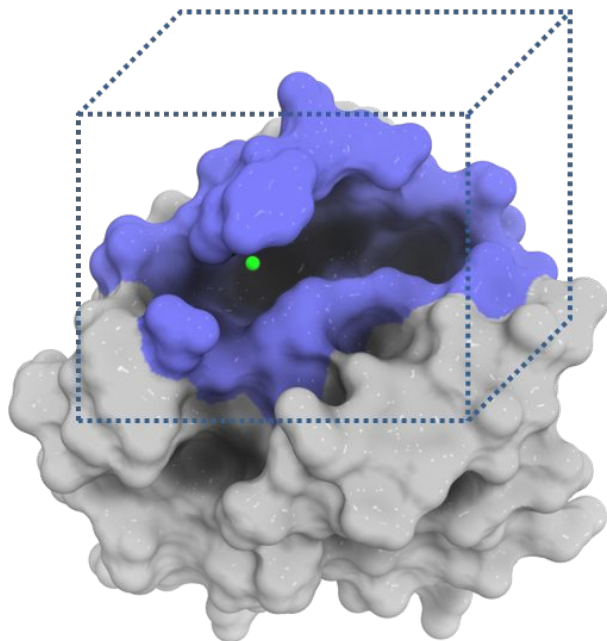
Smiles: CCCC1=CC=CC=C1O



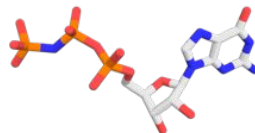
mode	affinity (kcal/mol)	dist from best mode	
		rmsd l.b.	rmsd u.b.
1	-6.814	0	0
2	-6.636	1.568	2.176
3	-5.583	1.249	1.661
4	-5.47	1.383	4.291
5	-5.241	1.603	2.251
6	-5.147	0.9939	1.095
7	-5.074	1.583	3.932
8	-4.991	1.507	4.28
9	-4.618	1.725	4.057
10	-3.997	1.438	2.255

Proteína - Ligante

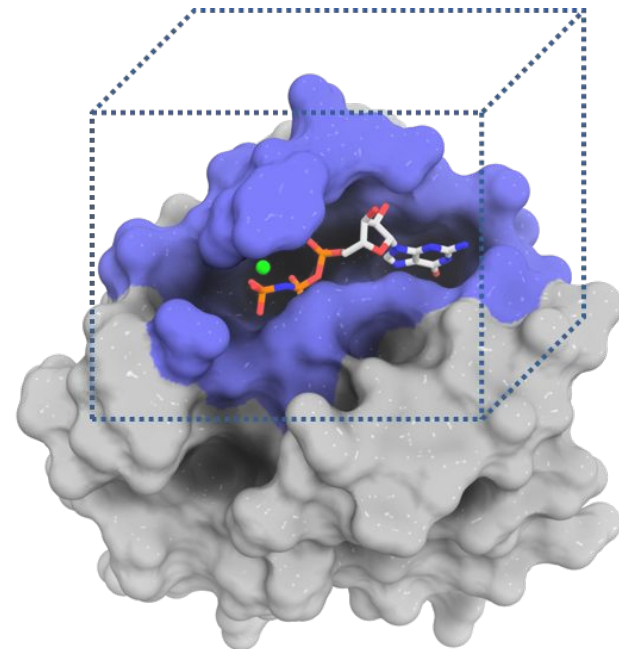
Proteína



Ligante



Complexo



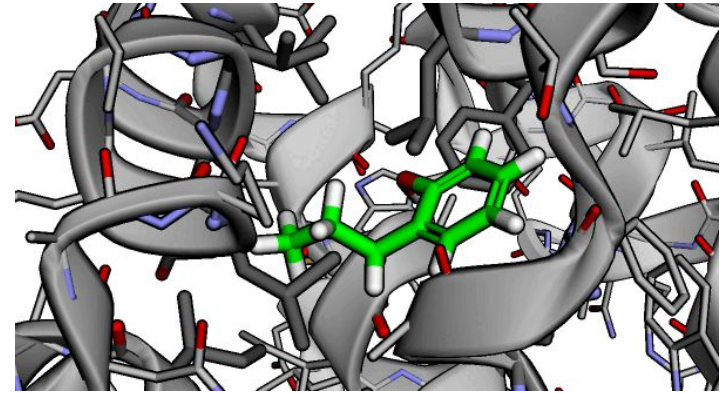
Proteína - Ligante

A alta afinidade entre um ligante e uma proteína é o resultado de uma combinação de fatores-chave:

- Encaixe Perfeito (Complementaridade Estérica):
 - As formas do ligante e da proteína se correspondem.
- Atração Química (Complementaridade Físico-Química):
 - Superfícies com propriedades opostas/compatíveis se atraem.
- Estabilidade Conformacional:
 - O ligante adota uma forma estável e de baixa energia ao se ligar.
- Ausência de Repulsão:
 - Minimização de choques atômicos e repulsão de cargas iguais.

Proteína - Ligante

- Atracamento molecular (*molecular docking*, em inglês), dedicada à previsão do modo de ligação e dos detalhes do reconhecimento molecular proteína-proteína e receptor-ligante.
- As metodologias computacionais de atracamento proteína-ligante estão baseadas no modelo chave-fechadura, proposto por Emil Fischer em 1894.
- Uma visão mais adequada deste processo é denominada de encaixe induzido, onde tanto o ligante quanto a proteína se adaptam um ao outro durante o processo de reconhecimento molecular.



Proteína - Ligante

- Os principais tipos de interações intermoleculares envolvidas no reconhecimento molecular proteína-ligante incluem:

i) ligações de hidrogênio;

ii) interações de van der Waals;

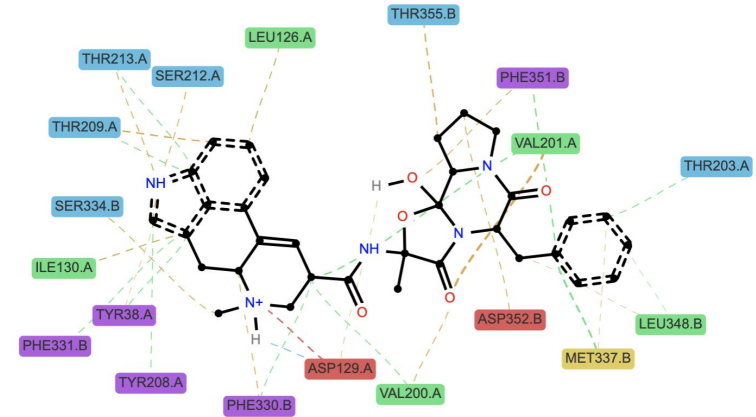
iii) interações iônicas;

iv) interações hidrofóbicas;

v) interações do tipo cátion- π ;

vi) interações envolvendo anéis aromáticos do tipo π - π ;

vii) coordenação com íons metálicos.



Proteína - Ligante

- O processo de reconhecimento molecular proteína-ligante é dirigido por uma combinação de efeitos entálpicos e entrópicos. Estes efeitos podem ser estimados através da energia livre de ligação de Gibbs.
- A energia livre de ligação de Gibbs pode ser obtida através de métodos teóricos:
 - Energia de Interação Linear (LIE, do inglês *Linear Interaction Energy*), a qual trata de estimar as energias livres a partir de simulações de dinâmica molecular utilizando um campo de força molecular clássico.
 - MM-PBSA (*Molecular Mechanics Poisson-Boltzmann Surface Area*) e MM-GBSA (*Molecular Mechanics Generalized-Born Surface Area*). Estes métodos utilizam simulações de dinâmica molecular do ligante/proteína livres e do complexo como base para os cálculos da energia potencial média e de solvatação.

$$\Delta G = \Delta H - T \cdot \Delta S$$

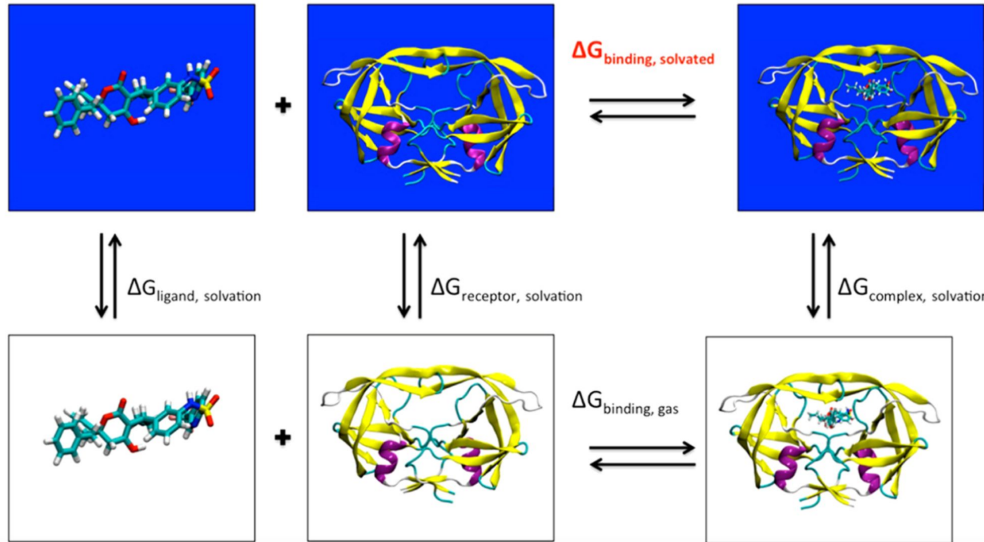
ΔG = Gibbs free energy

ΔH = Change in enthalpy

ΔS = Change in entropy

T = Temperature in K

Proteína - Ligante



$$\Delta G_{\text{binding, solvated}} = \Delta G_{\text{complex, solvated}} - [\Delta G_{\text{receptor, solvated}} + \Delta G_{\text{ligand, solvated}}]$$

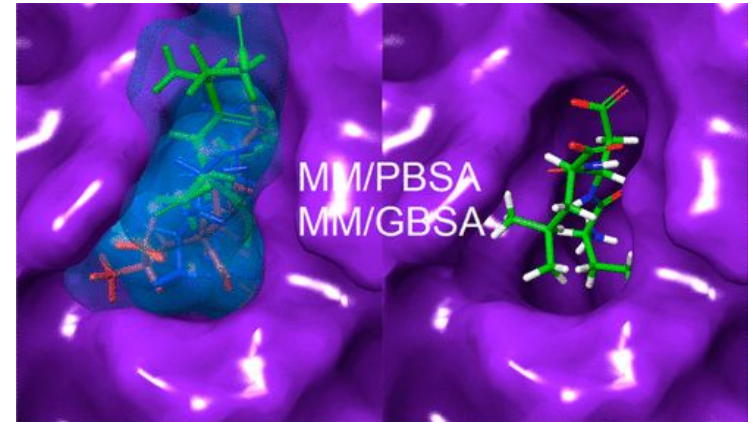
$$\Delta G_{\text{solvated}} = E_{\text{gas}} + \Delta G_{\text{solvation}} - TS_{\text{solute}}$$

The gas-phase energies (E_{gas}) are often the molecular mechanical (MM) energies from the forcefield, while the solvation free energies ($\Delta G_{\text{solvation}}$) are calculated using an implicit solvent model, and the entropic contribution (S) is estimated using known approximations.

Proteína - Ligante

Diferenças principais entre MMGBSA e MMPBSA:

- Modelo Solvente Implícito:
 - MM-GBSA: Usa o modelo Generalized Born (GB) para calcular a energia de solvatação. O modelo GB é um método semi-empírico baseado em uma aproximação do contínuo para representar o solvente (geralmente água).
 - MM-PBSA: Usa o modelo Poisson-Boltzmann (PB) para calcular a energia de solvatação. O modelo PB resolve a equação de Poisson-Boltzmann para obter o campo elétrico e a energia de solvatação em uma proteína e seu ligante.
- Termos de Energia:
 - Em ambos os métodos, a energia de ligação é geralmente decomposta em contribuições de energia eletrostática, energia de van der Waals, energia de solvatação e energia de área de superfície solvente acessível (SASA).

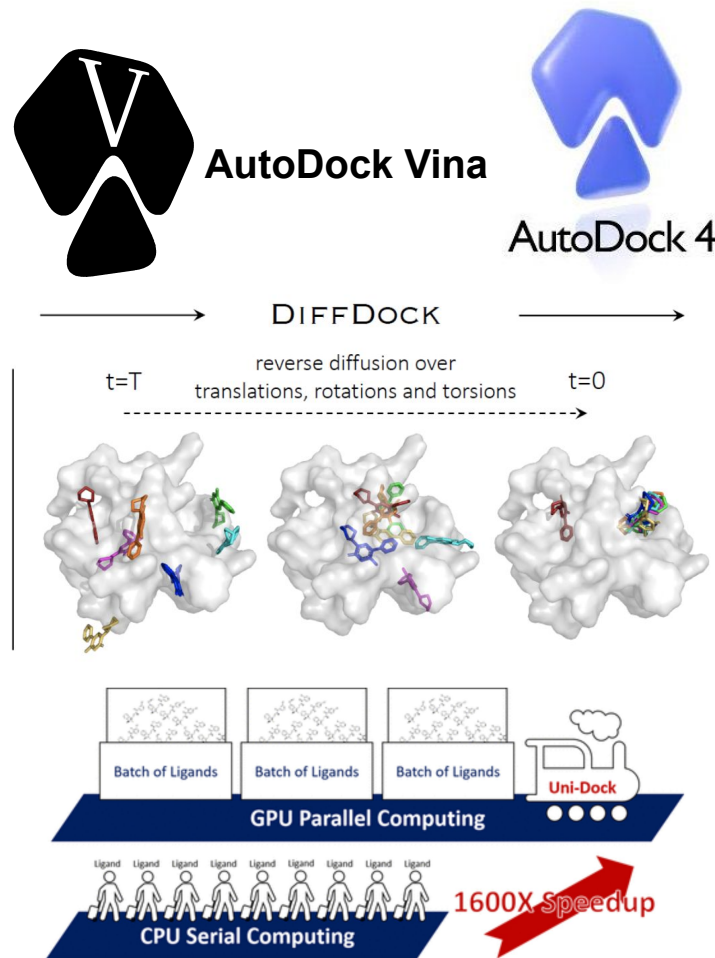


Proteína - Ligante

- O problema de atracamento molecular pode ser dividido em duas partes principais:
 - i) investigação e predição da conformação e orientação de uma molécula ligante no seu sítio de complexação;
 - ii) predição da afinidade em um complexo receptor-ligante, isto é, a energia livre de ligação (normalmente chamado na literatura de função scoring).
- Diversos programas de atracamento molecular disponíveis, distinguindo-se principalmente pelo método de busca e pela função de avaliação de afinidade empregada.
 - **EADock, DSS, AutoDock, DockThor, AutoDock Vina, DOCK, Método de Monte Carlo, PatchDock, MEDock, Smina, GNINA, DiffDock, Uni-Dock.**

Miller et al. (2012) J. Chem. Theory Comput. 8, 3314

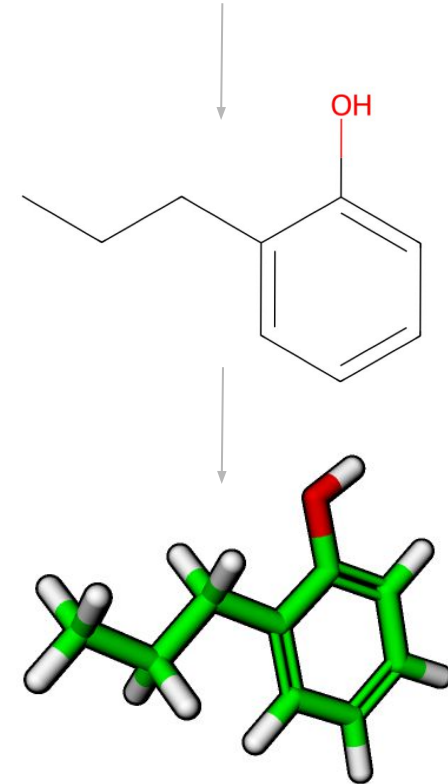
Verli, H (org). (2014). Bioinformática : da biologia à flexibilidade molecular



Proteína - Ligante

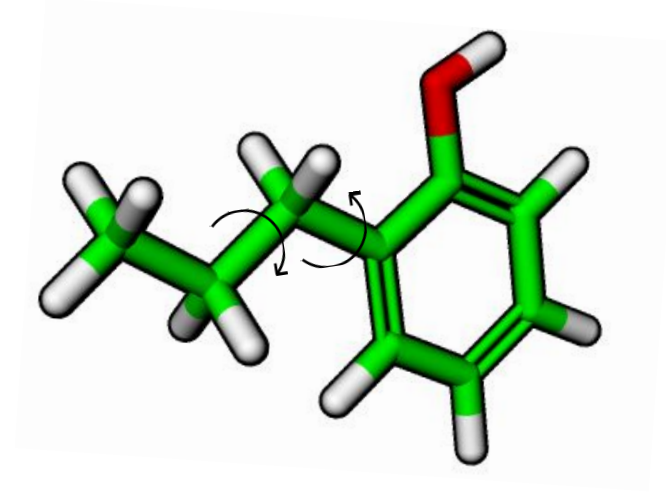
- Preparação do sistema. O primeiro passo nesta etapa é a obtenção das coordenadas das estruturas tridimensionais das moléculas interagentes.
 - Proteína: Protein Data Bank é atualmente a maior fonte pública de estruturas de proteínas e ácidos nucleicos resolvidos experimentalmente.
 - Ligante: As estruturas de ligantes podem ser obtidas de vários bancos de dados contendo milhares a milhões de ligantes no formato 1D (smi, simplified-molecular input-entry system, também chamado de formato SMILES) ou 2D (sdf, structure-data file format, também suporta formato 3D).
- A geração de uma estrutura 3D de um ligante a partir de uma representação 1D ou 2D pode ser feita através de vários programas tais como, CORINA, CONCORD, OMEGA, Balloon, RDKit.

Smiles: CCCC1=CC=CC=C1O



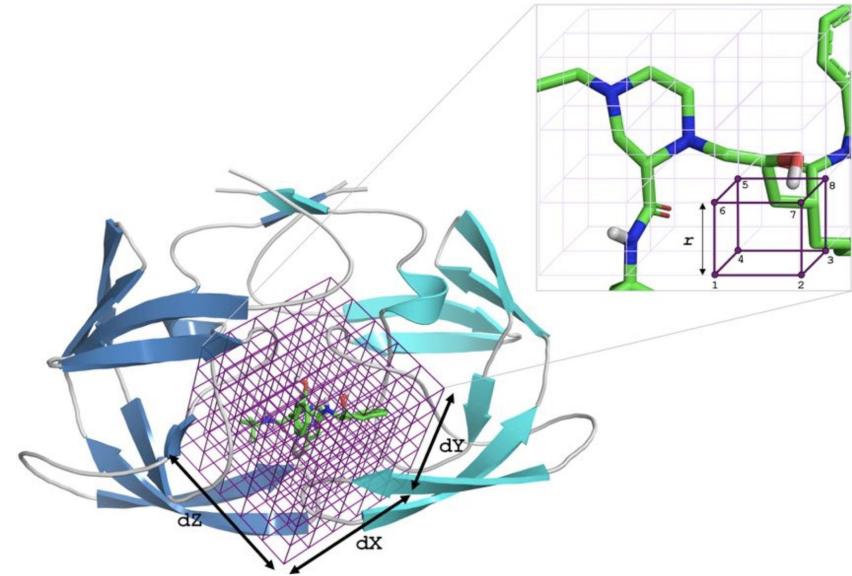
Proteína - Ligante

- A flexibilidade das moléculas interagentes é considerada de maneira variada pelos diversos métodos de atracamento molecular. Três principais estratégias são utilizadas:
 - i) a proteína é considerada rígida, e apenas os graus de liberdade translacionais e rotacionais do ligante são considerados, ou seja, o ligante é fixado em uma conformação rígida;
 - ii) a proteína é considerada rígida, mas todos os graus de liberdade do ligante (translacionais, rotacionais e conformacionais) são levados em conta;
 - iii) a proteína é considerada totalmente ou parcialmente flexível, e todos os graus de liberdade do ligante também são considerados.



Proteína - Ligante

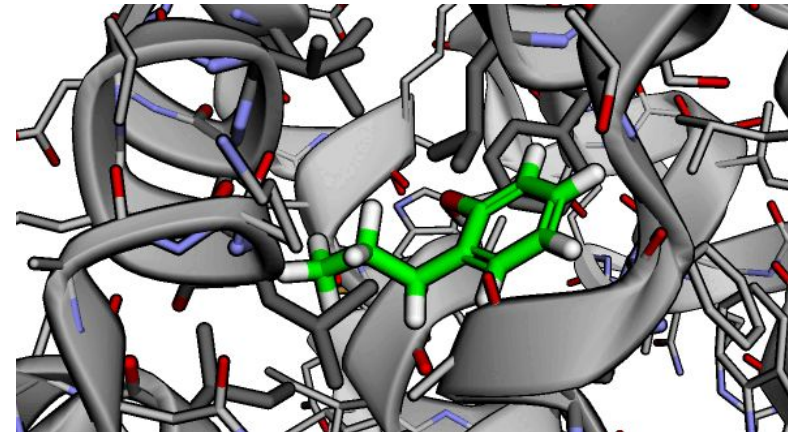
- Definindo a grade de energia (*grid box*):
 - A metodologia de grade é uma estratégia utilizada para aproximar o cálculo das energias eletrostáticas e de van der Waals (outros termos da função energia também podem ser utilizados), reduzindo drasticamente o custo computacional do cálculo da energia de interação intermolecular proteína-ligante.



Proteína - Ligante

- Definindo os parâmetros para o docking:
 - Método de busca: algoritmo utilizado pelo atracamento molecular para encontrar os modos de ligação do ligante no sítio receptor. Explora os graus de liberdade translacionais, rotacionais e conformacionais.
 - Função de score: função de pontuação que tem por objetivo quantificar a qualidade das soluções obtidas no atracamento molecular.
 - Número de modos gerados.

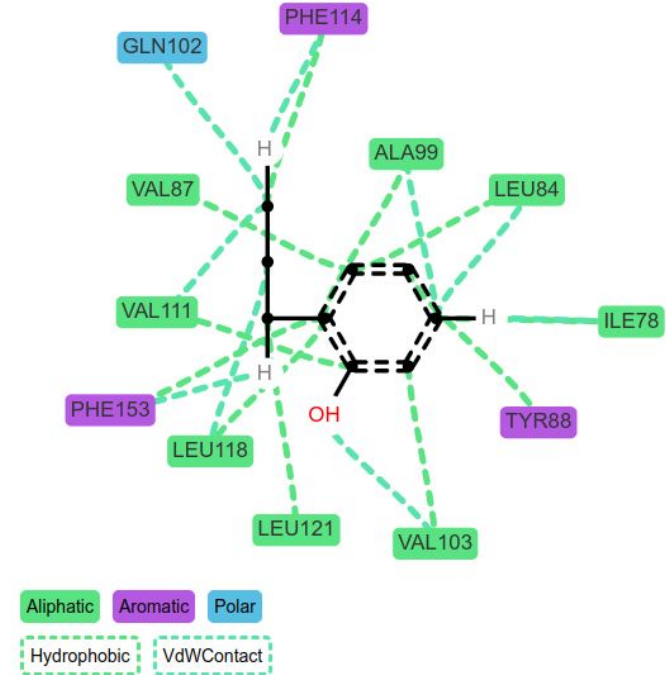
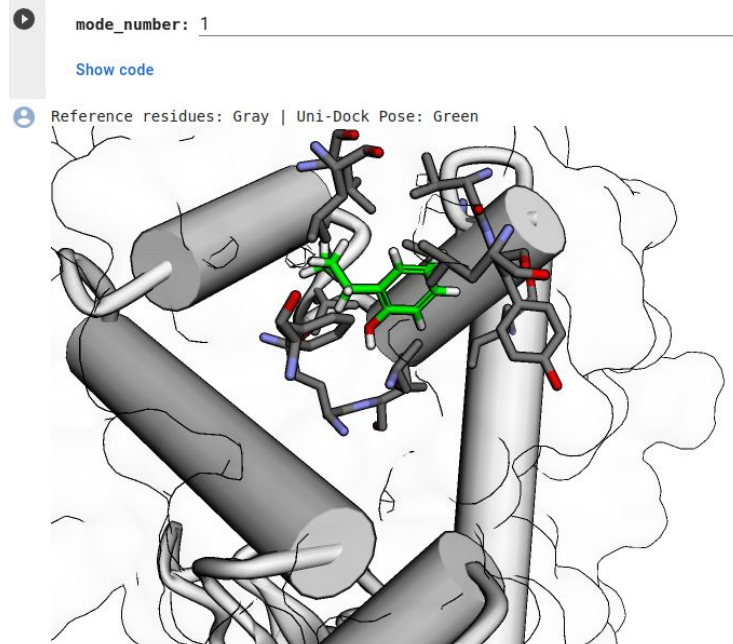
mode	affinity (kcal/mol)	dist from best mode	
		rmsd l.b.	rmsd u.b.
1	-6.814	0	0
2	-6.636	1.568	2.176
3	-5.583	1.249	1.661
4	-5.47	1.383	4.291
5	-5.241	1.603	2.251
6	-5.147	0.9939	1.095
7	-5.074	1.583	3.932
8	-4.991	1.507	4.28
9	-4.618	1.725	4.057
10	-3.997	1.438	2.255



Proteína - Ligante

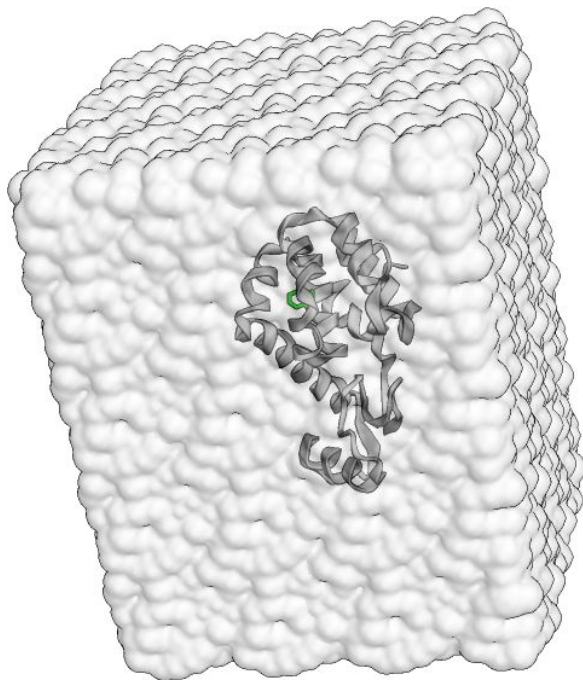
- Seleção da melhor pose:

Best pose selection:

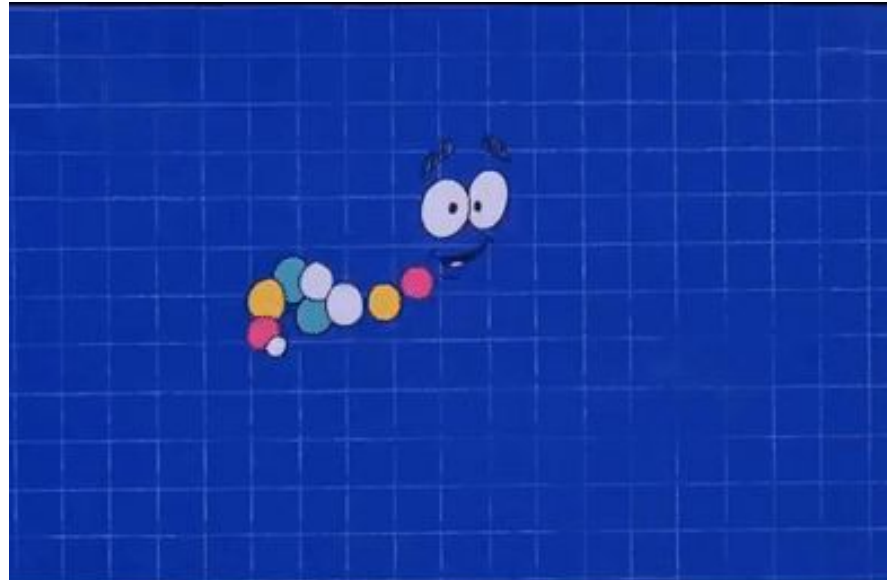


Proteína - Ligante

- 1) Solute a ser estudado (complexo obtido no docking)
- 2) Escolher o **campo de força** para proteína e ligante
- 3) Construir **topologia** do soluto (arquivo contendo os parâmetros do campo de força a serem utilizados para aquele sistema, permitindo o cálculo de Força e Aceleração).
- 4) **Solvatar** e **neutralizar** o sistema
- 5) **Minimizar** a energia do sistema
- 6) **Equilibrar** a caixa de simulação (temperatura, pressão e camadas de solvatação)
- 7) Simular livremente (**etapa de produção**)
- 8) Análises (Energia de Ligação, distâncias, RMSD...)



pablitoarantes@gmail.com
pablo.arantes@ems.com.br
<https://pablo-arantes.github.io>



Obrigado!

