ESCOLA GAÚCHA DE BIOINFORMÁTICA

Molecular Docking in the Cloud:

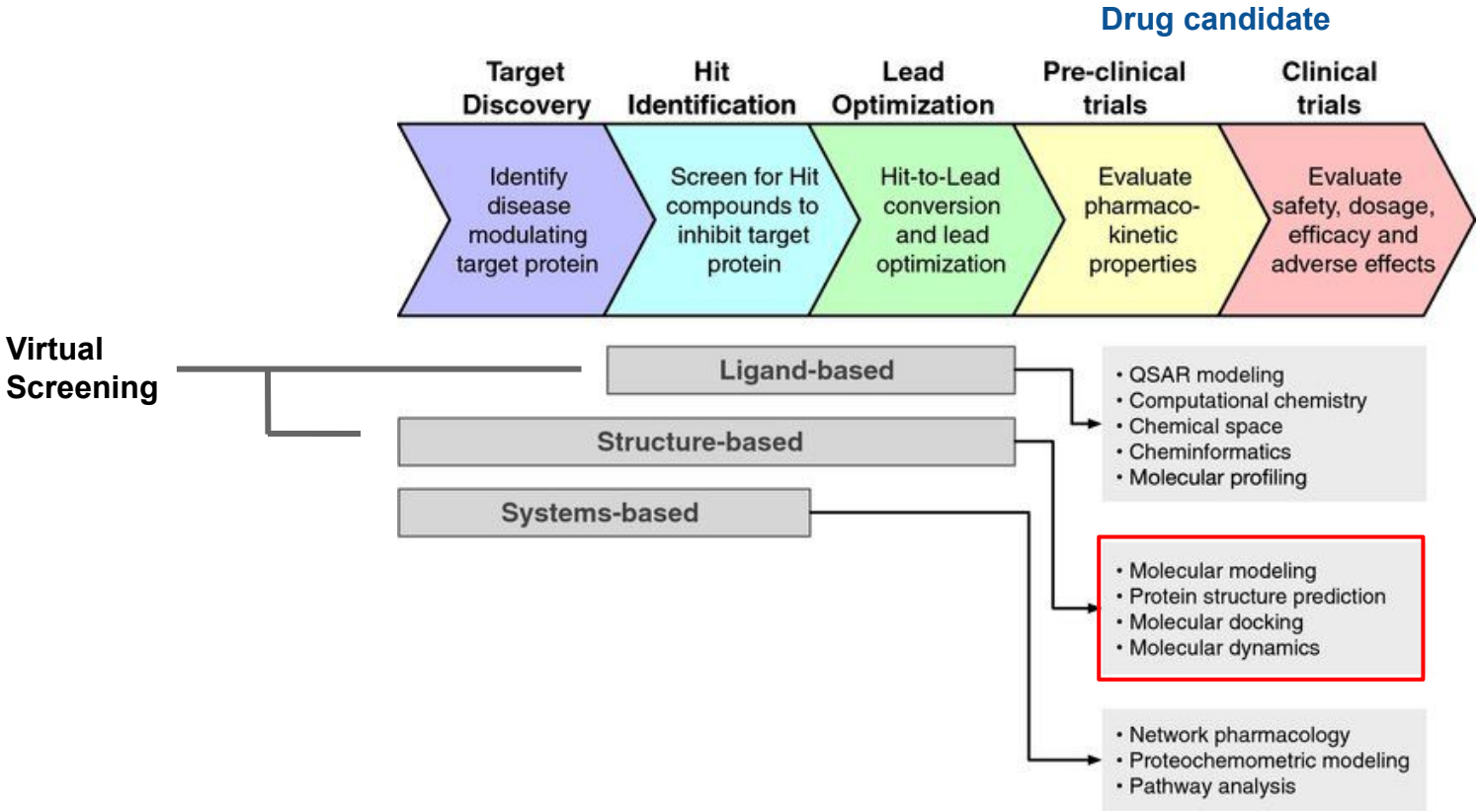
Introduction to Molecular Docking

PhD. Pablo Ricardo Arantes
PhD. Conrado Pedebos

Porto Alegre, July 14th 2025

CBiot Centro de Biotecnologia UFRGS · PPGBCM CBiot/UFRGS · InsCer Instituto do Cérebro · UFRGS Universidade Federal do Rio Grande do Sul · UFCSPA · PUCRS · PPGBio UFCSPA · PPGBM Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS

Drug Discovery Cycle



QSAR = Quantitative Structure-Activity Relationship

Schaduangrat *et al*. (2020) *J. Cheminform*. 12, 9

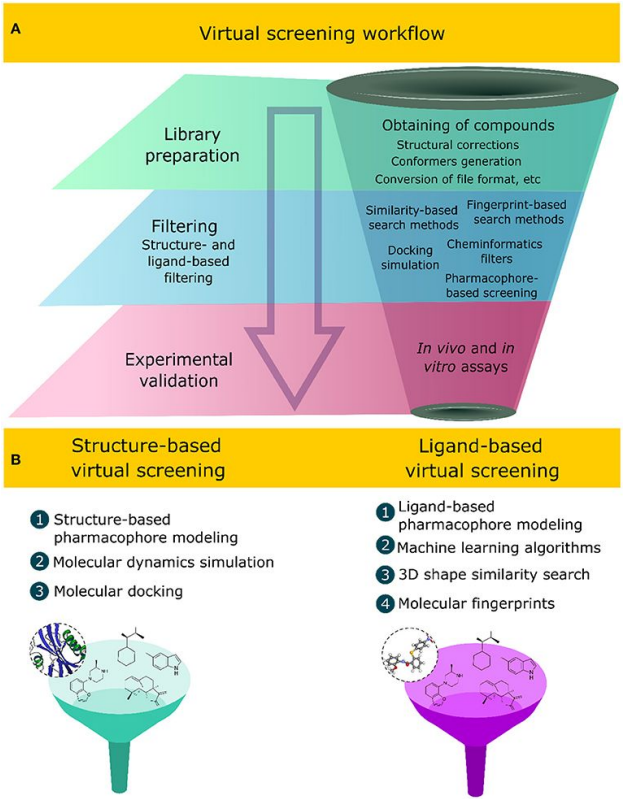
Drug Discovery Cycle

Schaduangrat *et al*. (2020) *J. Cheminform*. 12, 9

Drug Discovery Cycle - Virtual Screening

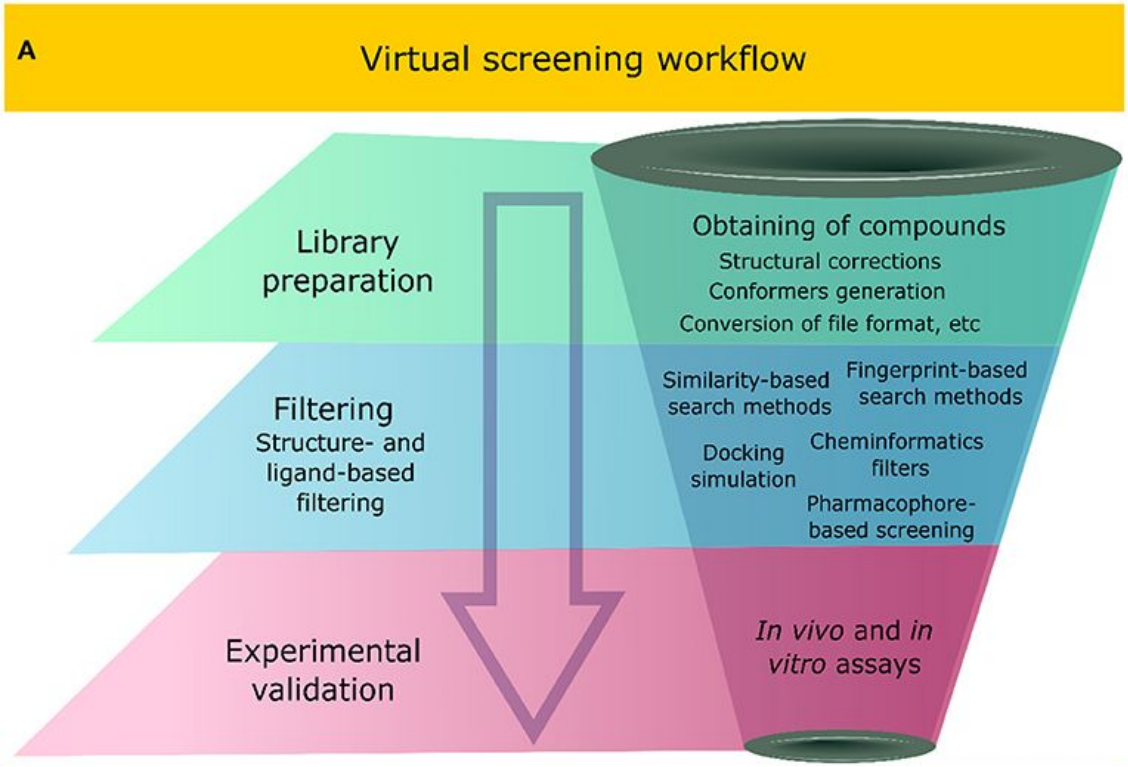
ESCOLA GAÚCHA DE BIOINFORMÁTICA

- Rapidly screen large libraries of small molecules (e.g., millions of compounds) to identify "hits" that bind to a target protein (e.g., an enzyme or receptor involved in a disease).
- Cost-effective alternative to High-Throughput Screening (HTS)
- Part of iterative **Design–Make–Test–Analyze** loop
- **Main types:** Structure-Based (SBVS) and Ligand-Based (LBVS)
- **Advantage:** Reduces wet-lab screening costs by focusing on computationally selected candidates.



Santana *et al*. (2021) *Front Chem*. 9, 662688

Virtual Screening

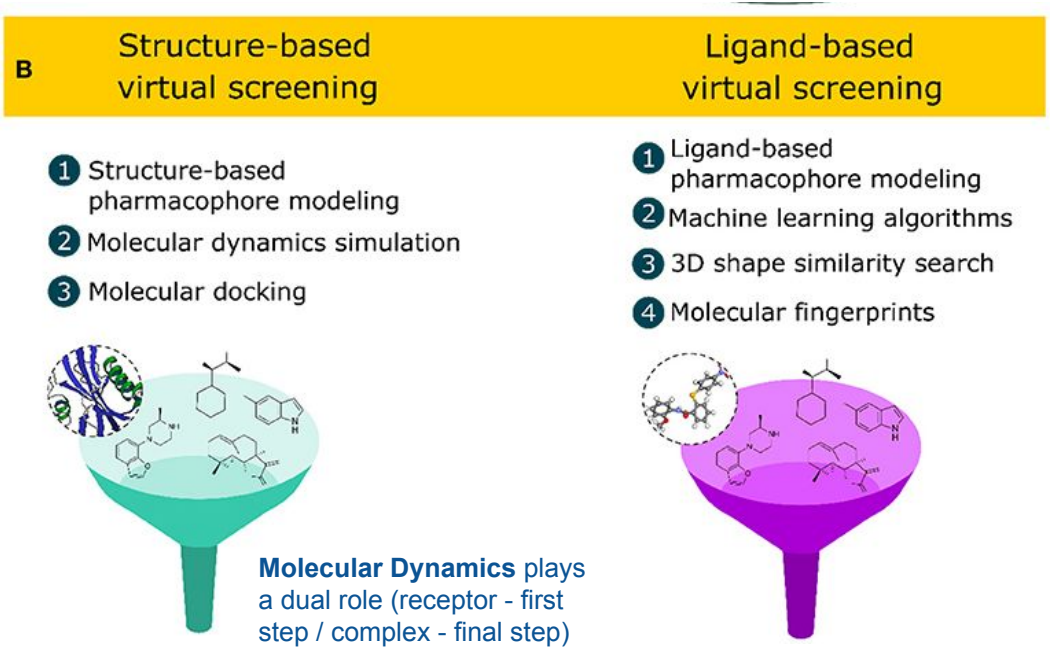


Santana *et al*. (2021) *Front Chem*. 9, 662688

Virtual Screening

ESCOLA GAÚCHA DE BIOINFORMÁTICA

Requires 3D structure of the target
Steps:
- Receptor & ligand preparation
- Docking using a scoring function
- Post-processing & hit selection



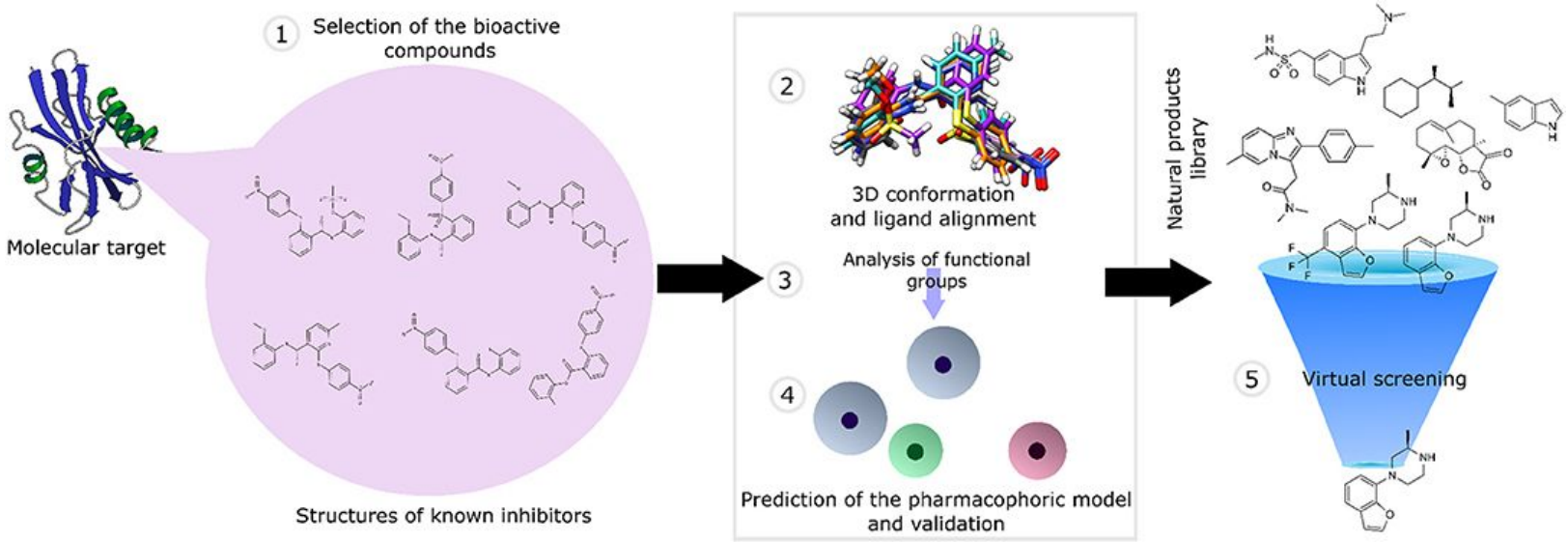
Used when target structure is unknown. Relies on known actives to find similar compounds
Methods:
- 2D/3D similarity searches
- Pharmacophore modeling
- Machine learning classifiers

Santana *et al*. (2021) *Front Chem*. 9, 662688

Virtual Screening - Ligand-Based

- Pharmacophore-based VS

Pharmacophore = a set of steric and electronic characteristics required to ensure better interactions with a particular biological target

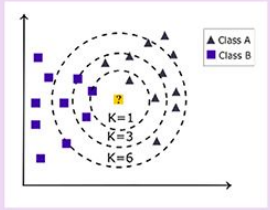


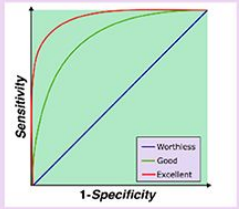
Santana *et al*. (2021) *Front Chem*. 9, 662688

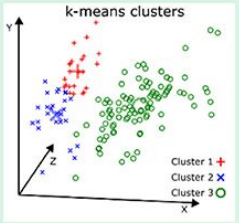
Virtual Screening - Ligand-Based

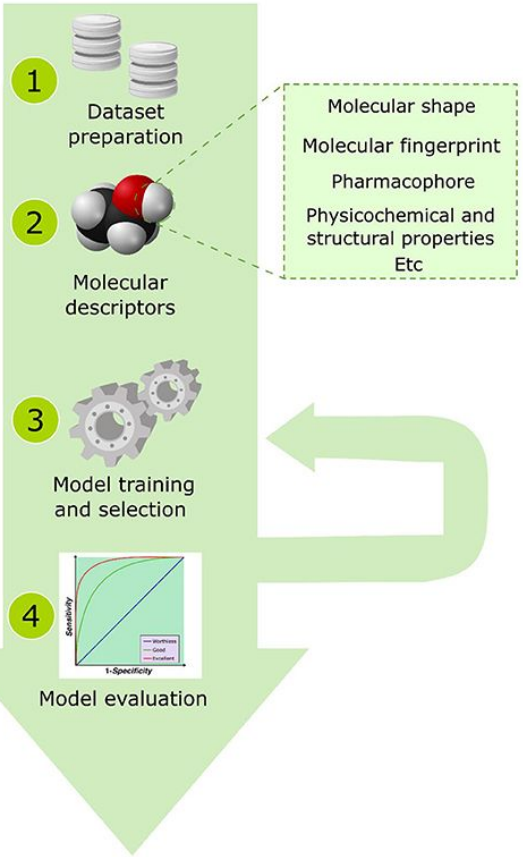
- Machine Learning approaches

Supervised learning techniques	Unsupervised learning techniques
Classifier methods Support vector machine k-Nearest neighbor Naive Bayes classifier	**Clustering methods** K-means clustering Hierarchical clustering Hidden Markov model
Regression analysis methods Decision trees Artificial neural networks Random forest Principal component regression	





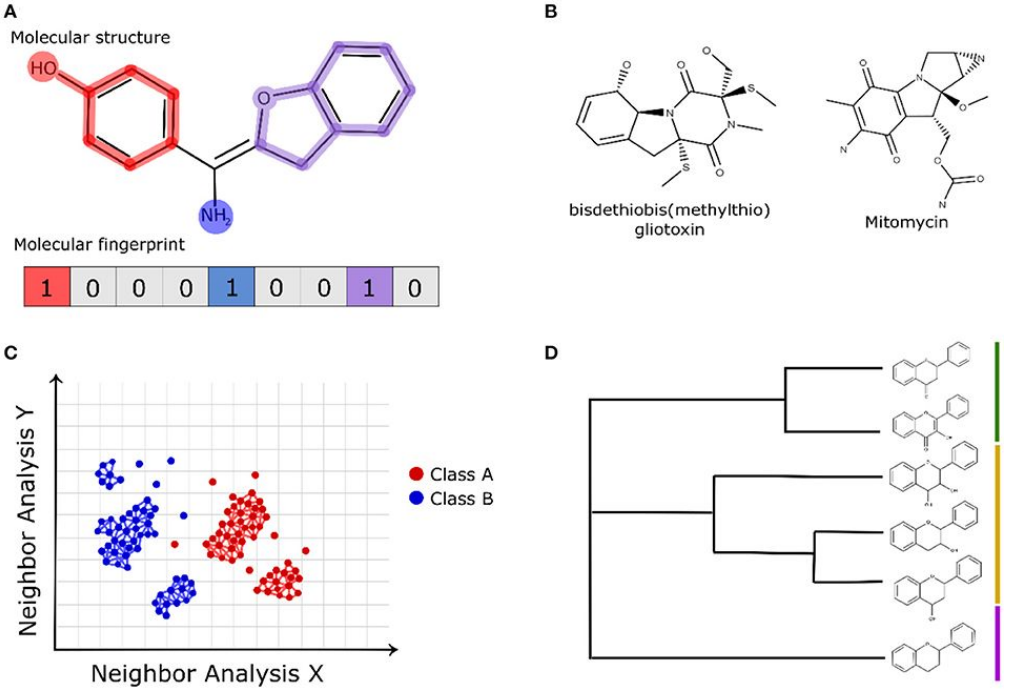




Santana *et al*. (2021) *Front Chem*. 9, 662688

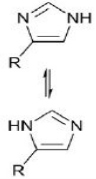
Virtual Screening - Ligand-Based

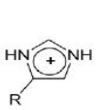
- Fingerprints:
  - Mathematical or vectorial representations = chemical and structural representations of a molecule
  - Similarity between molecules (Tanimoto more common)
  - Scalable, fast to calculate and compare, dimensionality reduction

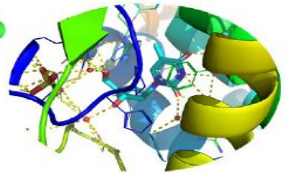


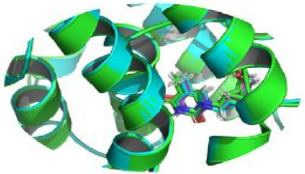
Santana *et al*. (2021) *Front Chem*. 9, 662688

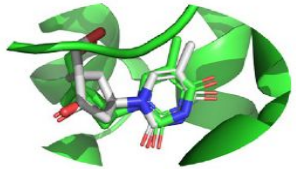
Virtual Screening - Structure-Based











Protein and ligand selection

- Structure of the protein target (3D coordinates)
- 3D coordinates of ligand(s)

Protein ligand preparation

- Add hydrogens
- Check protonation state(s)
- Select proper tautomer/protomer /stereoisomer

Binding site definition

- Experimentally (3D structure of protein-ligand complex is available)
- Cavity detection

Structural water

- Waters that coordinate H-bonds between protein and ligand

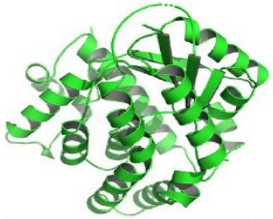
Molecular docking

- Exploring the conformational space (searching algorithm)
- Ranking candidate solutions (scoring function)

Evaluation

- Are all the H-bond donors and acceptors in the ligand satisfied?
- If the complex is known, is the binding mode of a ligand to protein reproduced?

Virtual Screening - Structure-Based



Protein and ligand selection

- Structure of the protein target (3D coordinates)
- 3D coordinates of ligand(s)

Receptor Structure

RCSB PDB
PROTEIN DATA BANK

AlphaFold

Accelerating breakthroughs in biology with AI

Explore the AlphaFold Database ›

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints

Compound Library

ZINC

Enamine

ChEMBL

DRUGBANK

Compound Library (Natural Products)

SuperNatural 3.0

npatlas

COCONUT

Latin American Natural Product Database (LANaPDB)

Virtual Screening - Structure-Based

ESCOLA GAÚCHA DE BIOINFORMÁTICA

Ligand preparation

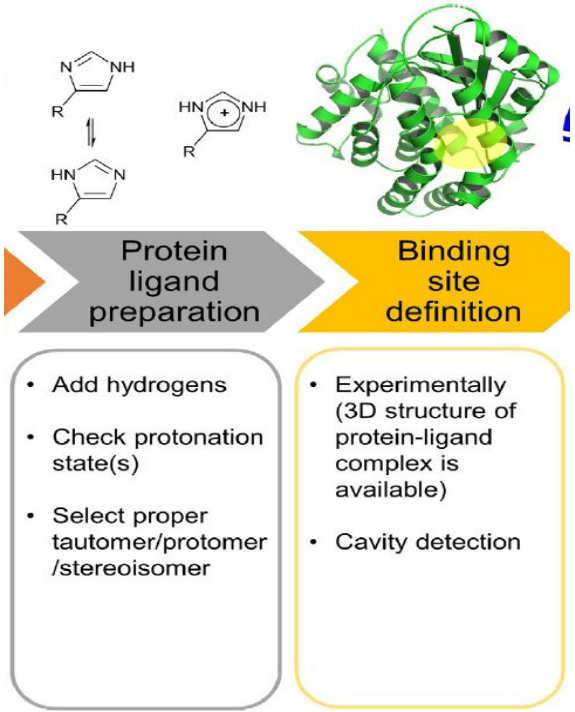
RDKit
Open-Source Cheminformatics and Machine Learning

Protein preparation

PDBFixer

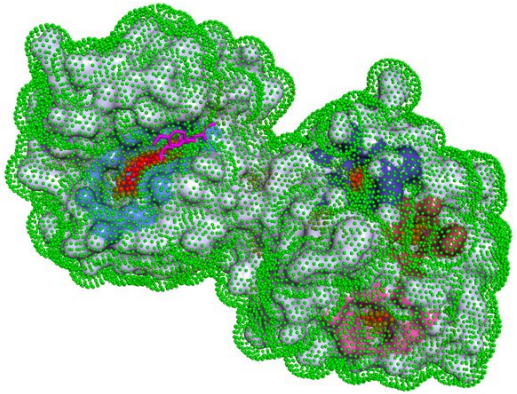
https://github.com/openmm/pdbfixer

AutoDock Tools



P2RANK

Ligand-binding site prediction based on machine learning.



Virtual Screening - Compound Filtering

ESCOLA GAÚCHA DE BIOINFORMÁTICA

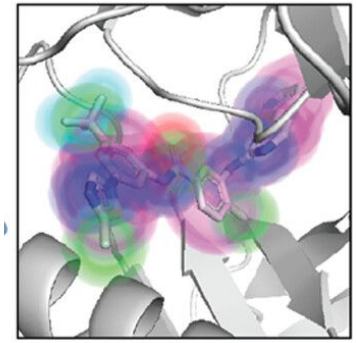
- Property-based filtering - A set of conditions that must be met by a compound to be used in the screening

	MW (Da)	PSA (A^2)	HBA	HBD	cLog*P*/cLog*D*	RTB	NAR	Formal charge	References
Lipinski's rule (RO5)	≤500	–	0–10	0–5	≤5	–	–	–	Lipinski et al., 1997
Ghose's rule	160–480	–	–	–	−0.4 to +5.6	–	20–70	–	Ghose et al., 1999
Oprea's drug-like rule	–	–	2–9	0–2	–	2–8	–	–	Oprea, 2000
Walters	200–500	≤120	0–10	0–5	–	0–8	–	–	Walters and Murcko, 2002
Veber's rule	–	≤140	–	–	–	0–10	–	–	Veber et al., 2002
REOS	200–500	–		0–5	−5.0 to 5.0	0–8		−2 to +2	Walters and Namchuk, 2003
Beyond rule of five (bRO5)	≤1,000	<250	<15	≤6	−2 to 10	≤20	–	–	Doak et al., 2014
Congreve's rule (RO3)	<300	–	≤ 6	≤3	≤3	–	–	–	Congreve et al., 2003
Herbicide-likeness	150–500	–	2–12	< 3	≤3.5	<12	–	–	Tice, 2001
Insecticide-likeness	150–500	–	1–18	≤ 2	0–5	<12	–	–	Tice, 2001
Hao's rule (pesticide-likeness)	≤435	–	≤6	≤ 2	≤6	≤9	≤17	–	Hao et al., 2011

MW, molecular weight; PSA, polar surface area; HBD, hydrogen bond donor; HBA, hydrogen bond acceptor; RTB, rotatable bonds; NAR, number of aromatic rings.

Santana *et al*. (2021) *Front Chem*. 9, 662688

Virtual Screening - Structure-Based

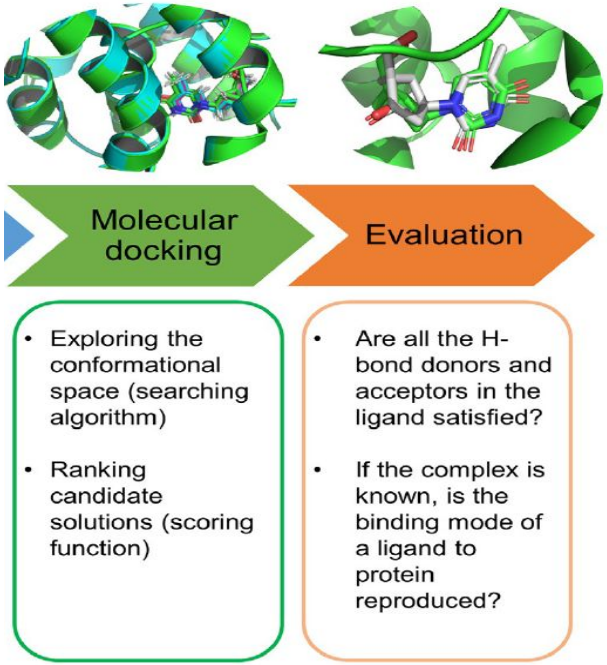


GNINA Docking

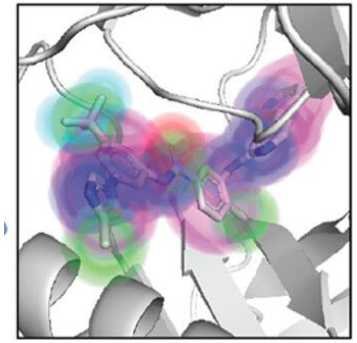
https://github.com/gnina/gnina

Uni-Dock

https://github.com/dptech-corp/Uni-Dock



Virtual Screening - Structure-Based

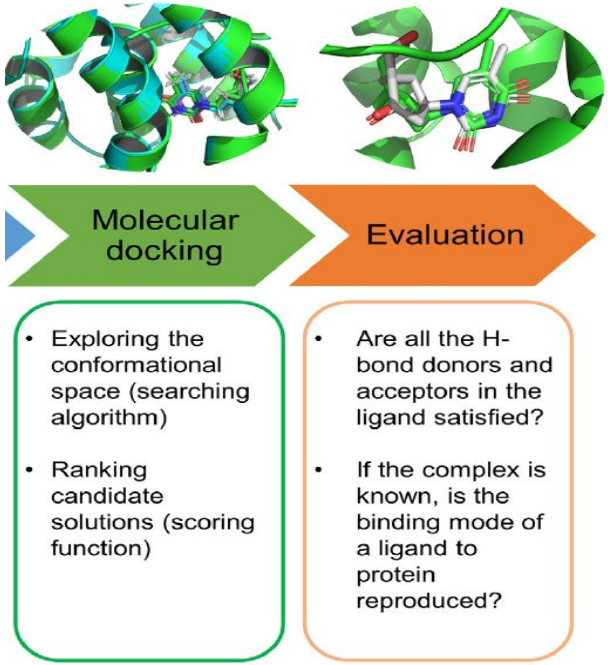


GNINA Docking

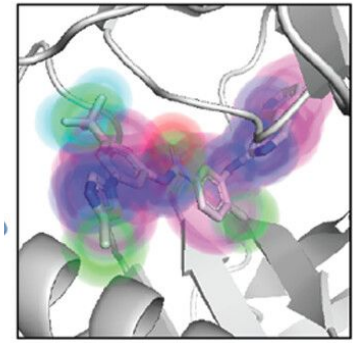
https://github.com/gnina/gnina

Uni-Dock

https://github.com/dptech-corp/Uni-Dock



Virtual Screening - Structure-Based



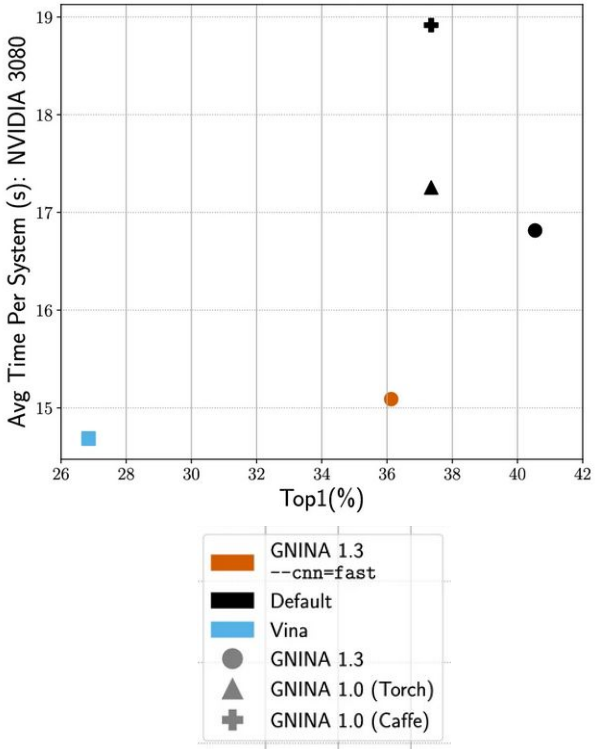
GNINA Docking

https://github.com/gnina/gnina

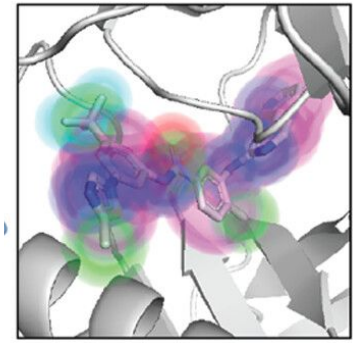
Uni-Dock

https://github.com/dptech-corp/Uni-Dock

10.000 molecules in <42 hours
1.000.000 molecules in 173 days :(



Virtual Screening - Structure-Based



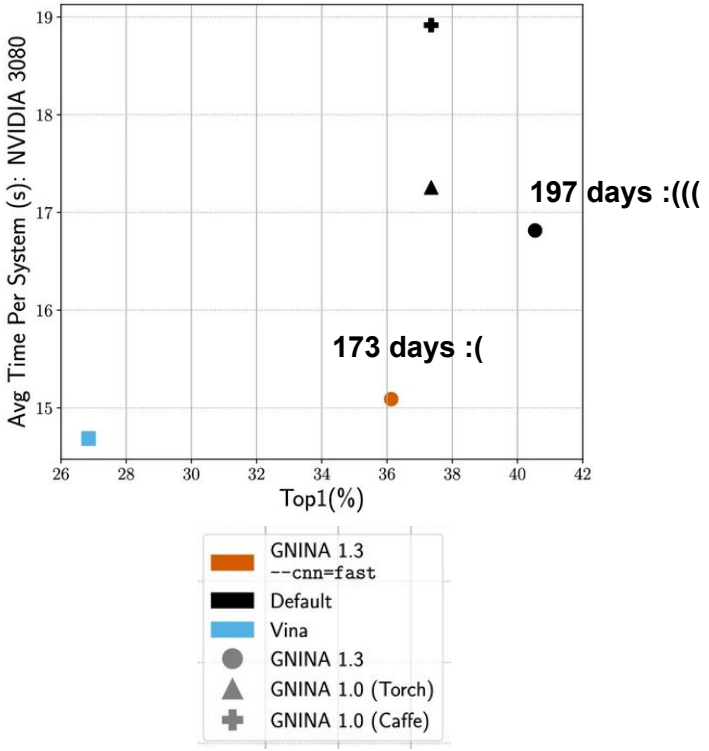
GNINA Docking

https://github.com/gnina/gnina

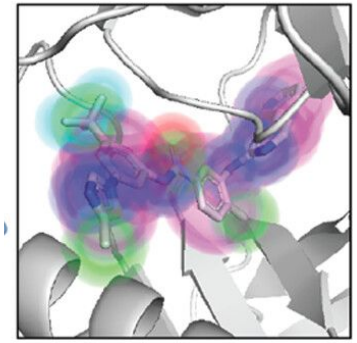
Uni-Dock

https://github.com/dptech-corp/Uni-Dock

10.000 molecules in <42 hours
1.000.000 molecules in 173 days :(



Virtual Screening - Structure-Based



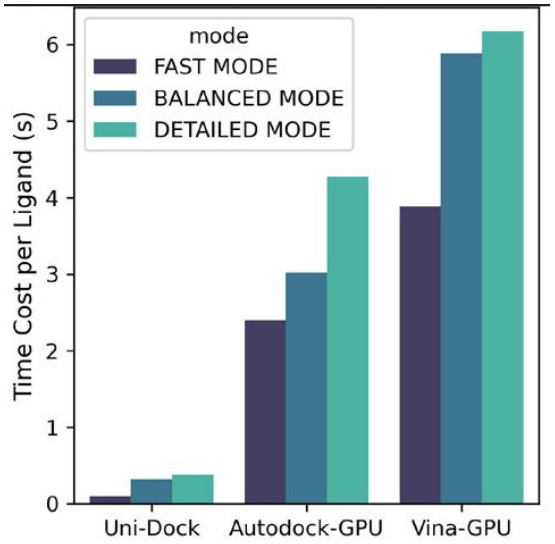
GNINA Docking

https://github.com/gnina/gnina

Uni-Dock

https://github.com/dptech-corp/Uni-Dock

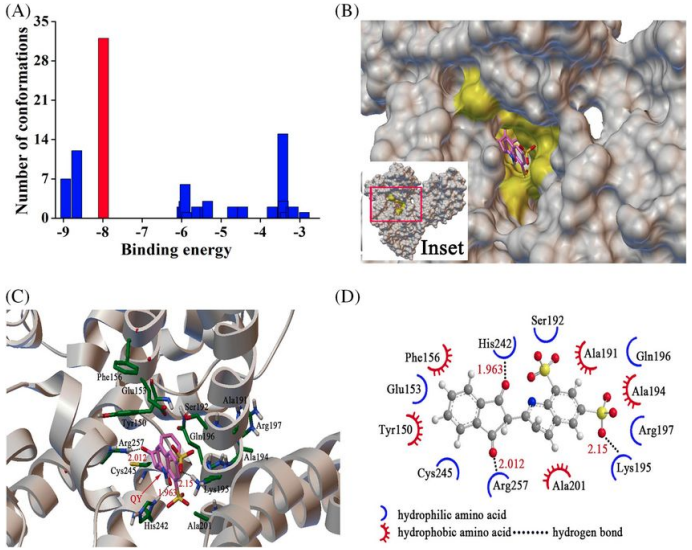
10.000 molecules in less than 1.5 hour
1.000.000 molecules in less than 6 days!!!



Virtual Screening - Structure-Based

ESCOLA GAÚCHA DE BIOINFORMÁTICA

- Evaluation
  - Clusters
  - Molecule ranking (scoring methods)



Product of pose classification and predicted binding affinity

	Molecule_Solution	minimizedAffinity	CNNscore	CNNaffinity	CNN_VS
0	mol_6087_1	-7.93925	0.911983	6.780704	6.183887
1	mol_5986_1	-8.11932	0.915682	6.498751	5.950791
2	mol_1600_1	-7.45823	0.881309	6.533484	5.758017
3	mol_3631_1	-7.96031	0.909520	6.195273	5.634723
4	mol_193_1	-8.73742	0.917535	6.130170	5.624646
...	...	...	...	...	...
62675	mol_1514_8	-6.83488	0.110981	3.370108	0.374019
62676	mol_2437_9	-6.42252	0.100460	3.688916	0.370587
62677	mol_771_9	-5.51312	0.130097	2.767925	0.360099
62678	mol_771_10	-5.47916	0.121043	2.877078	0.348249
62679	mol_2437_10	-6.92232	0.100410	3.424139	0.343818

Wang *et al*. (2018) *J. Sci. Food Agric*. 99, 73-82

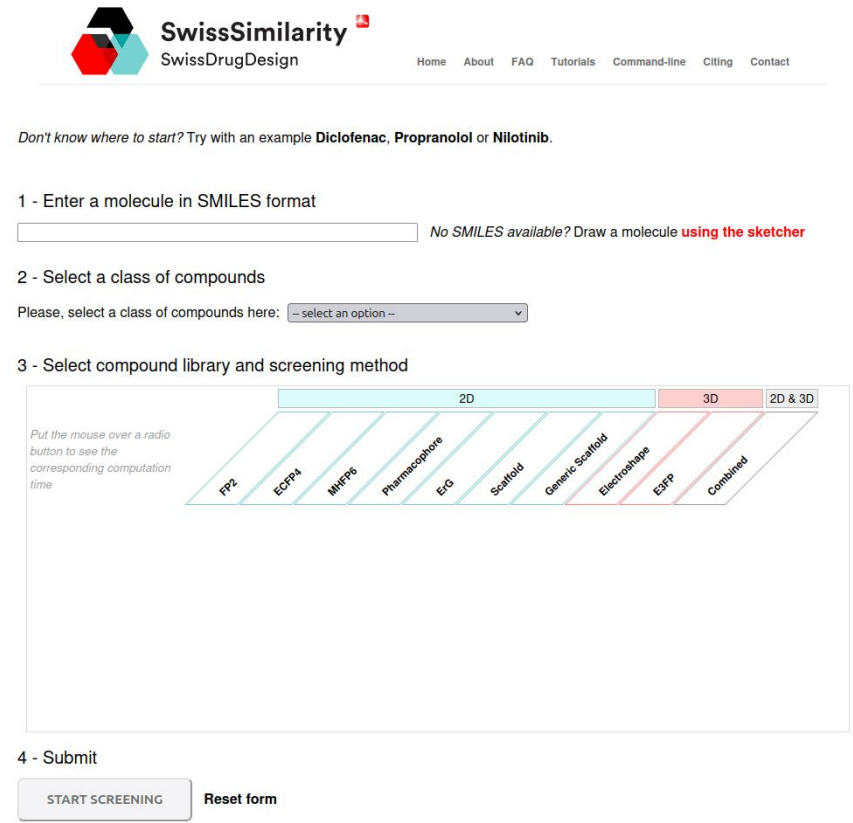
Virtual Screening - Post-processing

ESCOLA GAÚCHA DE BIOINFORMÁTICA

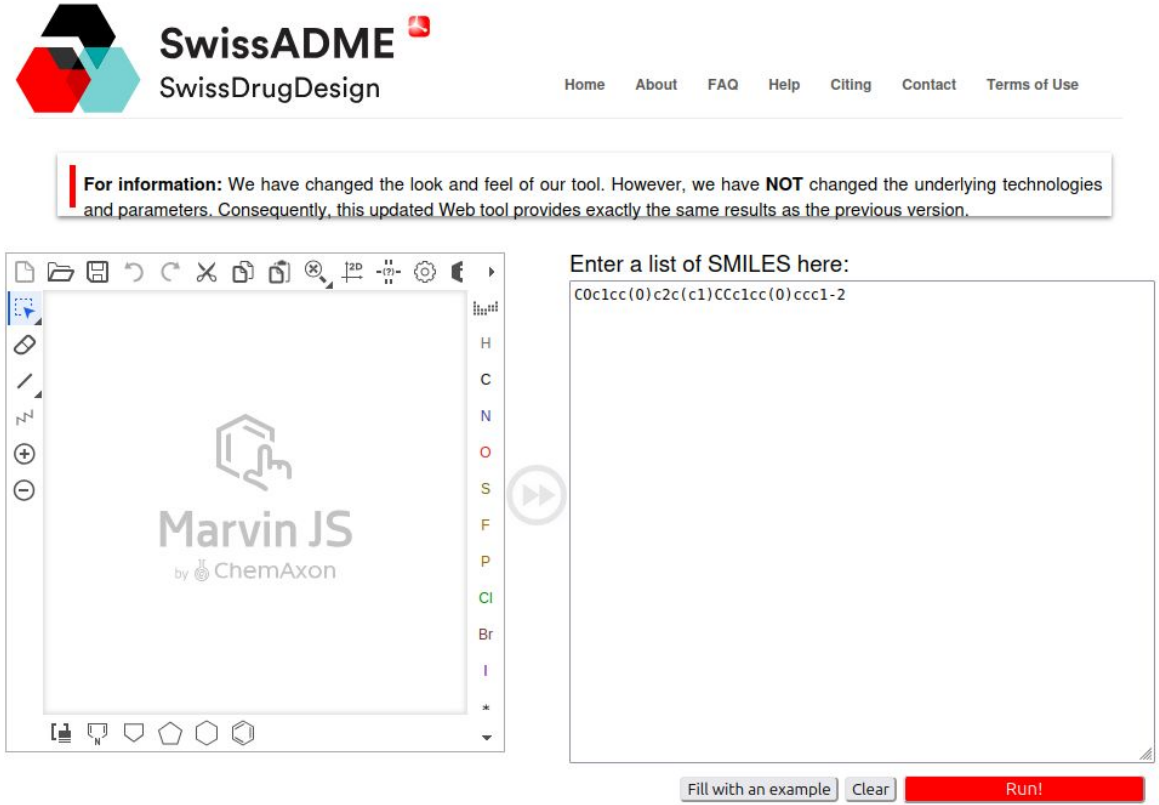
- What's next?
  - Redo the VS using different conformations of the same receptor
    - Copies of the same receptor from the RCSB PDB
    - Ensemble obtained from Molecular Dynamics simulations
  - Molecular Dynamics to refine the top poses
  - Other binding energy predictors
    - MM-PBSA (frames from the MD simulations) - shown to greatly reduce the rate of false positives
    - AEV-PLIG
  - PLACER
  - **Visual Inspection by a medicinal chemist**
    - Reliable interactions?
    - Synthetic feasibility
    - Intellectual Property?
  - **Predictions of ADMETox properties**

Virtual Screening - Post-processing

ESCOLA GAÚCHA DE BIOINFORMÁTICA



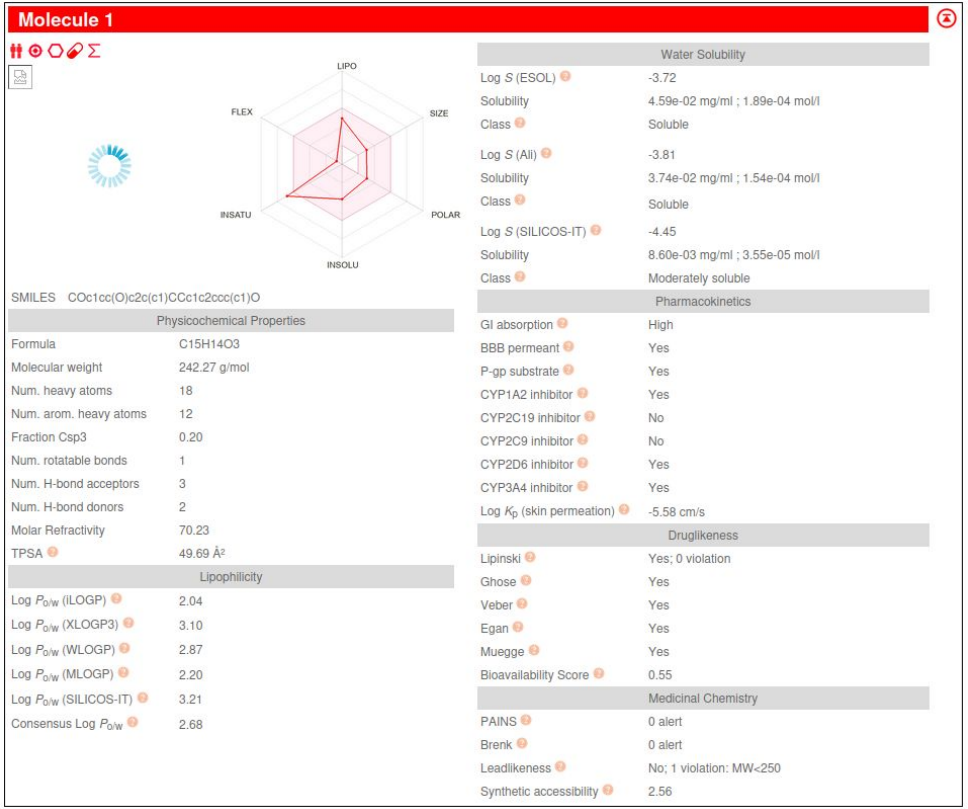
Virtual Screening - Post-processing



Virtual Screening - Post-processing

Arom.: if high, might be less soluble
Fraction Csp3: sp3 carbons; values > 0.3 indicate more complex structures (higher drug likeness); low values -> planar structures.
Rot. bonds: ideal < 10 (Veber), if >, less oral bioavail.
HBA: ideal <10; **HBD**: <5
Molar Refrac.: 40-130 (Ghose); too low = low affinity; too high = low solub.
TPSA: ideal < 140 for oral absorp.; < 90 cross BBB

Ideally: -0.5 a 5.0 (**Lipinski**). < -1: too hydrophilic; > 5: risk having low solubility and high toxicity.



Water solub: ideal > -4.0 and < 0.5 for oral absorp.

Pharmacok:
GI abs. = oral absorp.
BBB = good for SNC drugs, bad for side-effects
P-gp = if Yes, might have less bioavail. and tissue accum.
CYP inh. = relevant for drug-drug interact.
Log Kp = -4 to -8 typically (more -, more perm.)

Virtual Screening - Post-processing