Infraestrutura Nacional de **Computação Distribuída**

FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

NOVA UNIVERSIDADE NOVA DE LISBOA · NOVA SCHOOL OF SCIENCE & TECHNOLOGY · UCIBIO · i4HB Institute for Health and Bioeconomy

Biomolecular Engineering Lab · bioSIM

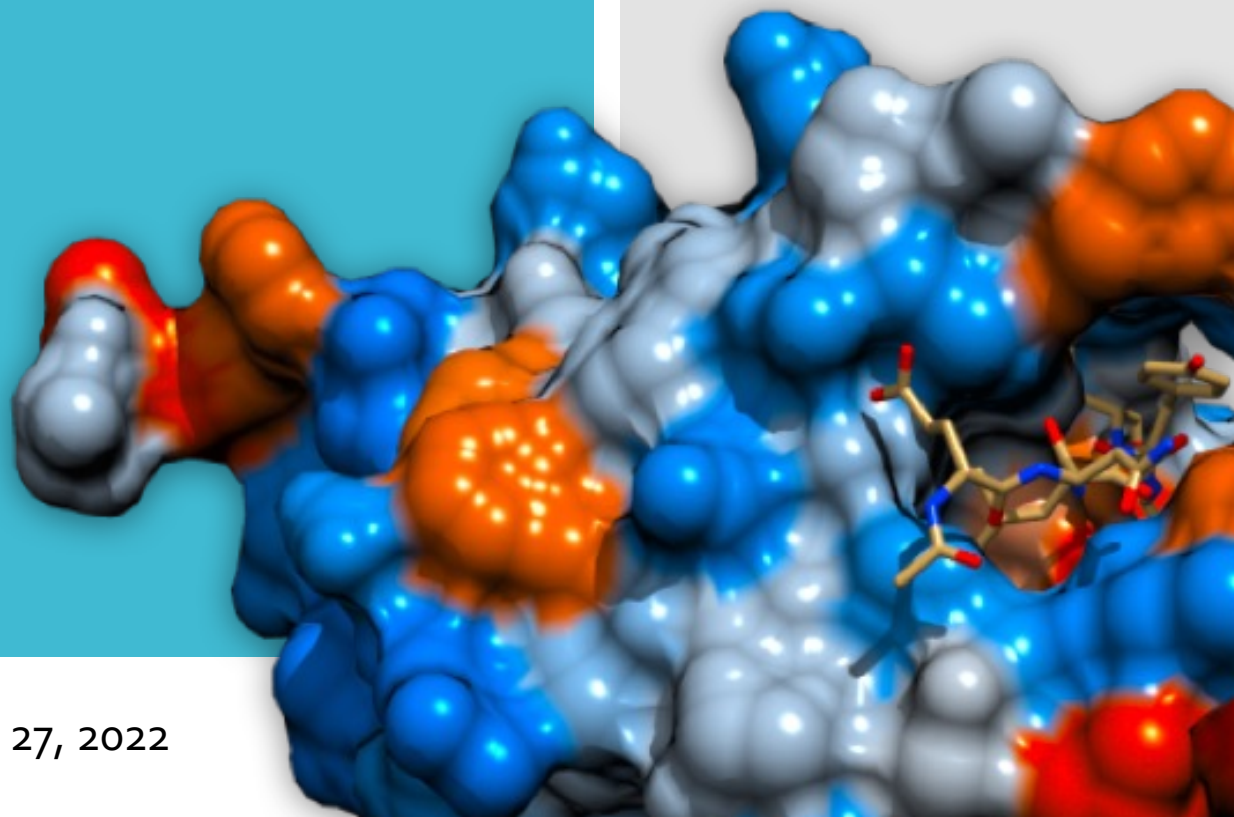
Computational Biosciences Using HPC Systems – Module 4

Structured-Based Virtual Screening

Sérgio F. Sousa & Arménio Barbosa

Email: sergiofsousa@med.up.pt
UCIBIO/REQUIMTE, BioSIM – Departamento de Biomedicina
Faculdade de Medicina da Universidade do Porto, Portugal

Email: aj.Barbosa@fct.unl.pt
UCIBIO-i4HB, Biomolecular Engineering Lab – Departamento de Química
School of Science and Technology, NOVA University Lisbon, Portugal



www.biosim.pt

https://sites.fct.unl.pt/biomolecular_eng/

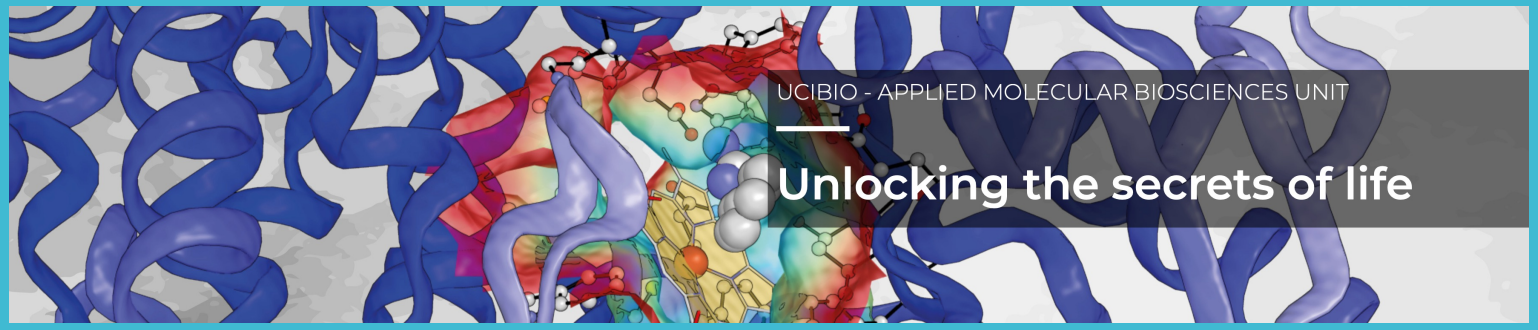
September 27, 2022

UCIBIO
Applied Molecular Biosciences Unit
@REQUIMTE

UCIBIO is formed by the collaborative efforts of researchers from the **University of Porto** and **University NOVA of Lisbon**.

UCIBIO´s research activities occur mainly at the campus of Faculty of Sciences and Technology of the NOVA University of Lisbon (**FCT-NOVA**), and at the Faculty of Sciences (**FCUP**), the Faculty of Pharmacy (**FFUP**), the Faculty of Medicine (**FMUP**) and Instituto de Ciências Biomédicas Abel Salazar (**ICBAS-UP**) from the University of Porto.

UCIBIO combines key expertise in Chemistry and Biological Sciences with an ambitious strategic plan to maximize its national and international impact in terms of scientific productivity, advanced training and translation to society. In the national context, UCIBIO's key strength lies on its **broad scope of fundamental and applied research, standing at the interface of Chemistry, Biology and Engineering** to address pertinent questions at **atomic, molecular, sub-cellular and cellular levels, including cell-to-cell interactions and population evolutionary dynamics**.



BioSIM Research Group

Location

BioSIM – Biomolecular SIMulations Research Group
UCIBIO/REQUIMTE - Departamento de Biomedicina
Faculdade de Medicina da Universidade do Porto
Portugal
www.biosim.pt

U.PORTO
FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

BioSIM
Research Group

Research

BioSIM – Biomolecular SIMulations Research Group

Our research team bridges the gap between theory and experiment applying and developing state-of-the-art computational tools focusing on ***Enzymatic Catalysis, Drug Discovery and Molecular Recognition***

For that we combine: ***QM/MM Methods, Quantum Mechanics, Molecular Dynamics, Docking, Virtual Screening,*** *and* ***Free Energy Perturbation methods****, always in close linking with experiment.*

Several ***Software Applications and Databases*** *have also been developed and made available to the scientific community.*

Website: www.biosim.pt

Biomolecular Engineering Lab

Location

Biomolecular Engineering Lab
UCIBIO-i4HB,– Departamento de Química
School of Science and Technology,
NOVA University Lisbon, Caparica, Portugal
https://sites.fct.unl.pt/biomolecular_eng/

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

Biomolecular Engineering Lab

Research

Biomolecular Engineering Lab

*Composed by a **multidisciplinary team** with expertise in **Applied and Computational Chemistry, Biotechnology, Biomedical Engineering, Chemical Engineering and Physics of Materials**.*

*The group is dedicated to **MINIMAL BIOMIMETIC SYSTEMS**, combining designed molecular recognition agents with functional materials, for **Bioseparation, Biocatalysis, Sensing & Diagnostics, and Nanomedicine**.*

*The main modeling tools to acheive these goals are: **Molecular Dynamics, Docking, Virtual Screening, Protein Structure Prediction, Biomaterials Simulations.***

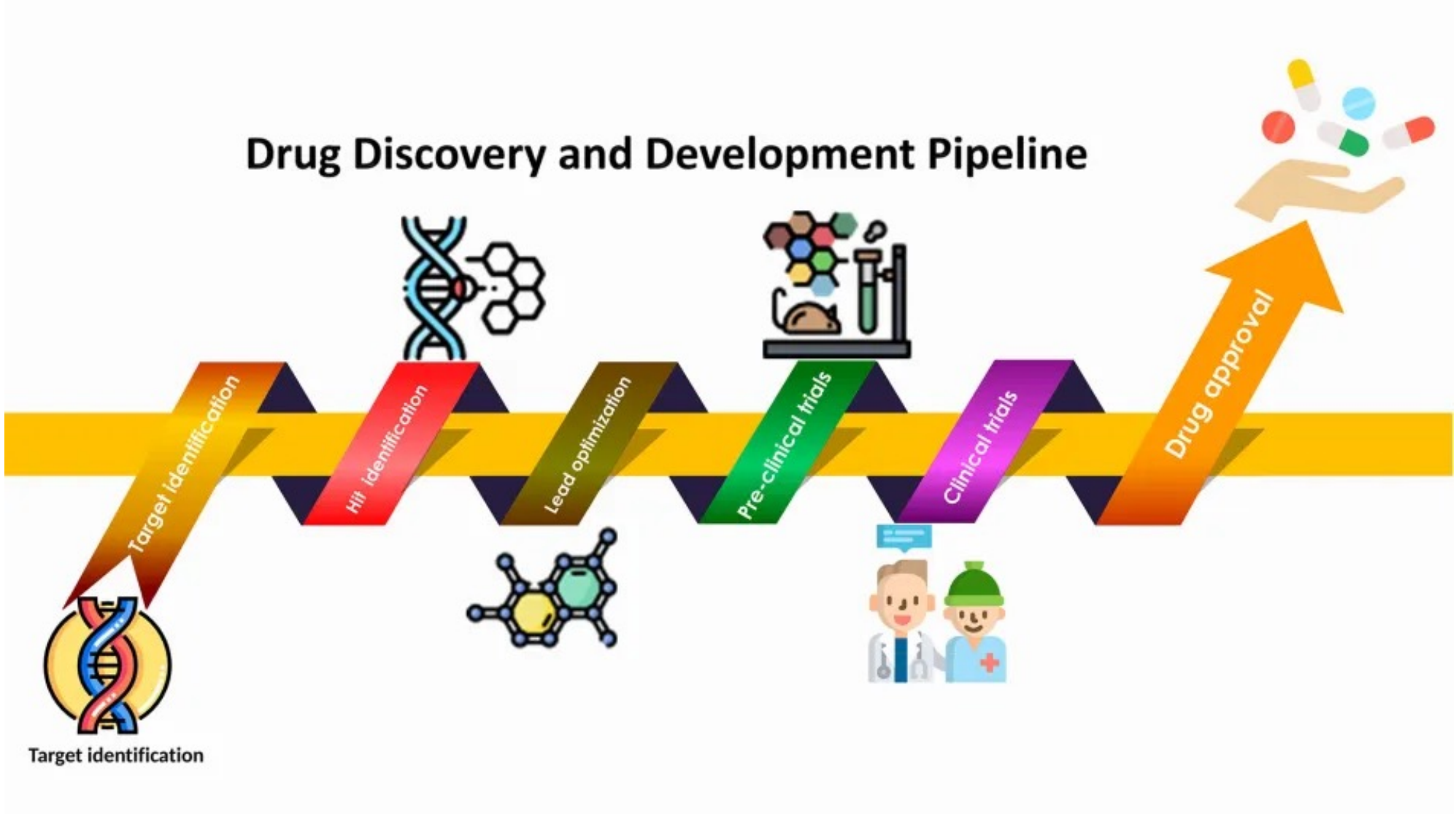
*We have outstanding benefit for the **interconnected in silico and experimental** in impactful research.*

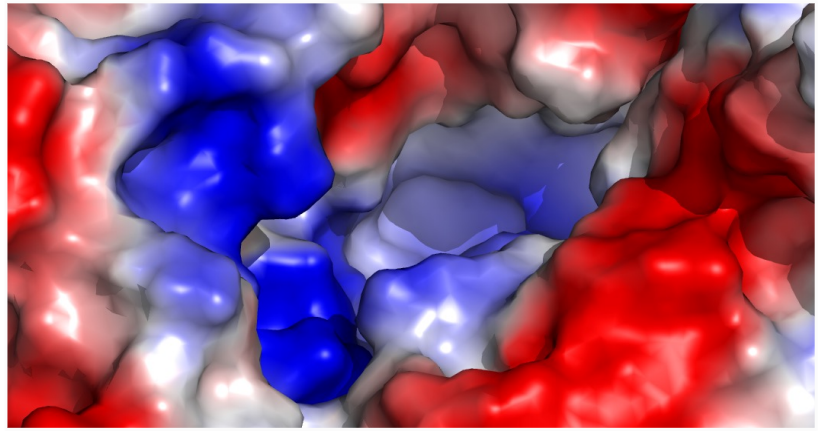
Website: https://sites.fct.unl.pt/biomolecular_eng/home

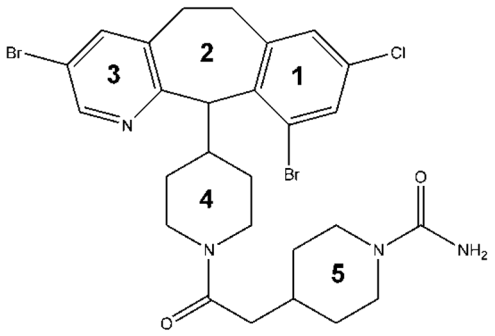
Process of Drug Discovery

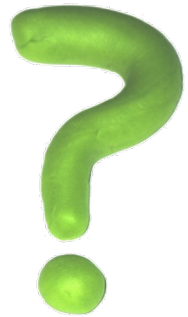
Presently

- Based on the Knowledge of the Biological Target
- Rational Drug Discovery vs Serendipity
- Strategic decisions
- Strong Interdisciplinary Nature
- Involves different areas of research
- Specialization / Dialogue


Drug Discovery and Development Pipeline
Target identification
Hit identification
Lead optimization
Pre-clinical trials
Clinical trials
Drug approval
Target identification



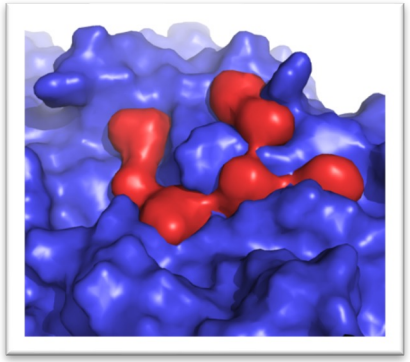


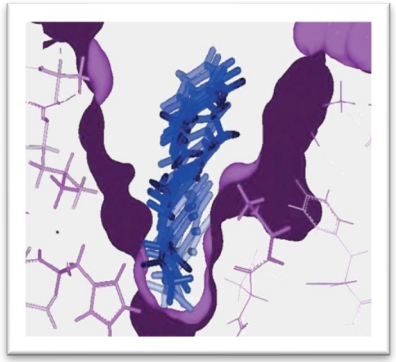


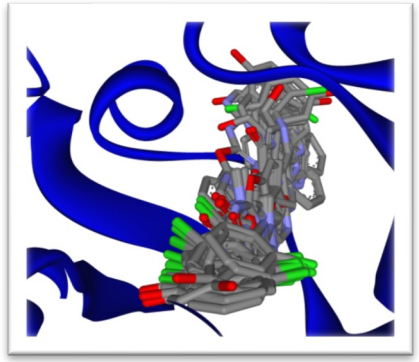
Protein-Ligand Docking

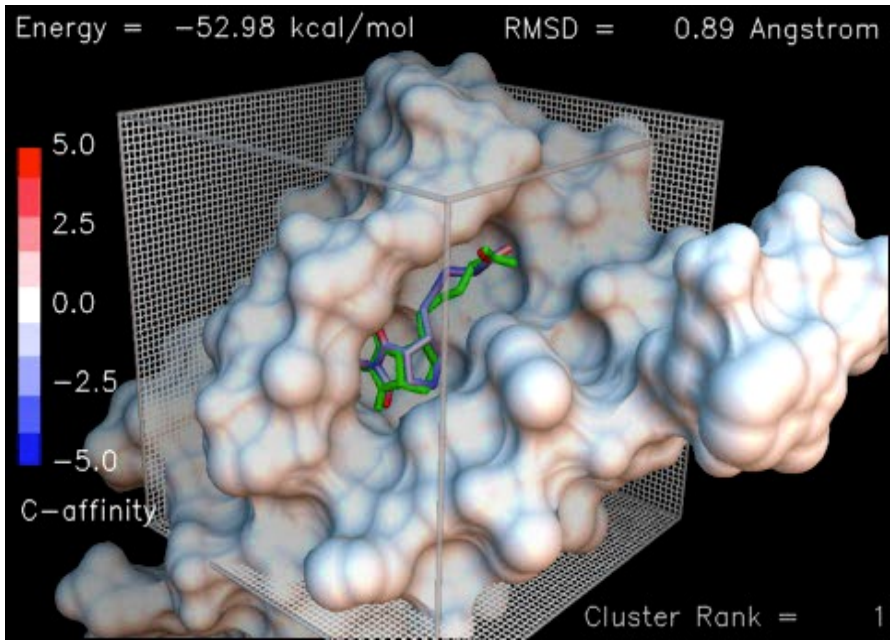
*"Protein-Ligand Docking is computational method that aims to **predict** and **rank** the structure(s) arising from the association between a given ligand and a target protein of known 3D structure"*

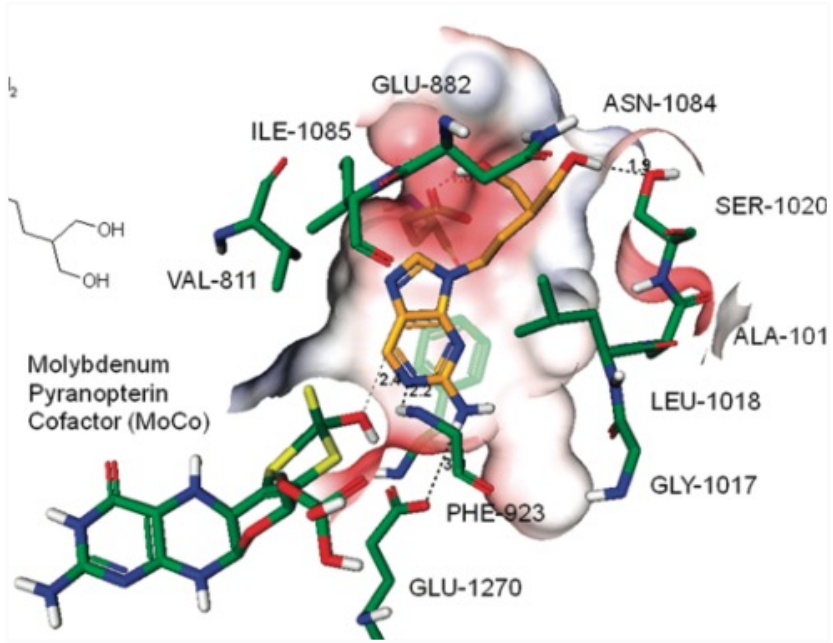
in Protein-Ligand Docking: Current Status and Future Challenges
Sérgio F. Sousa et al.
Proteins, Structure, Function and Bioinformatics, Vol. 65, pp. 15-26, (2006).

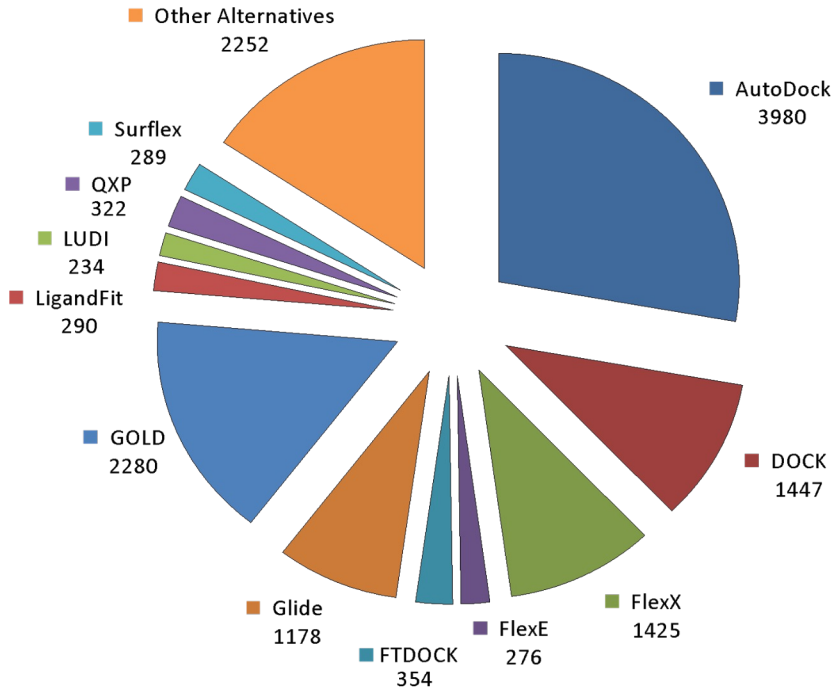




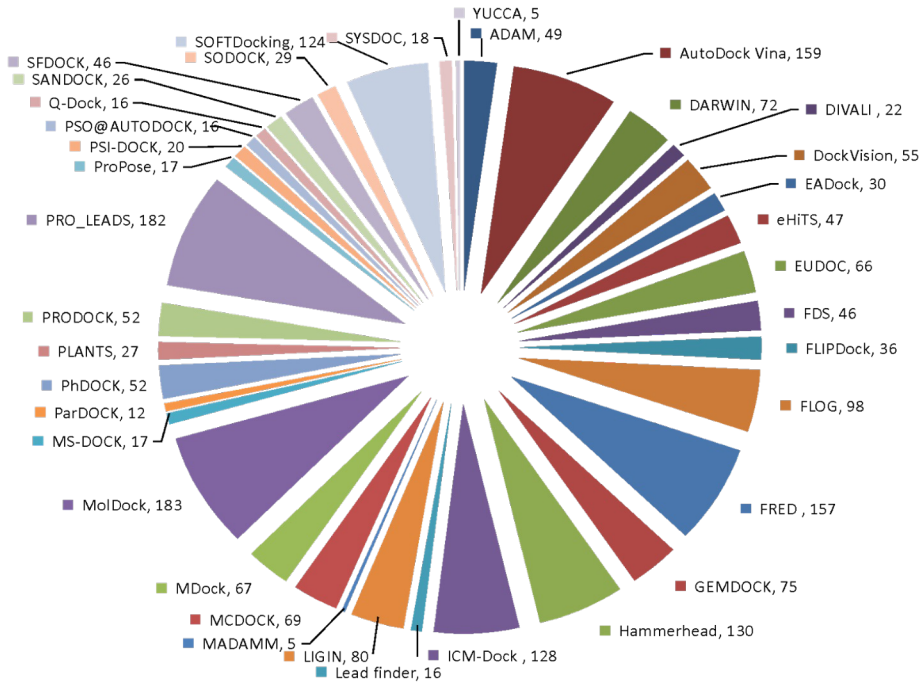



Energy = −52.98 kcal/mol
RMSD = 0.89 Angstrom
5.0
2.5
0.0
−2.5
−5.0
C−affinity
Cluster Rank = 1


GLU-882
ASN-1084
ILE-1085
SER-1020
OH
OH
VAL-811
ALA-101
Molybdenum
Pyranopterin
Cofactor (MoCo)
LEU-1018
GLY-1017
PHE-923
GLU-1270



Global

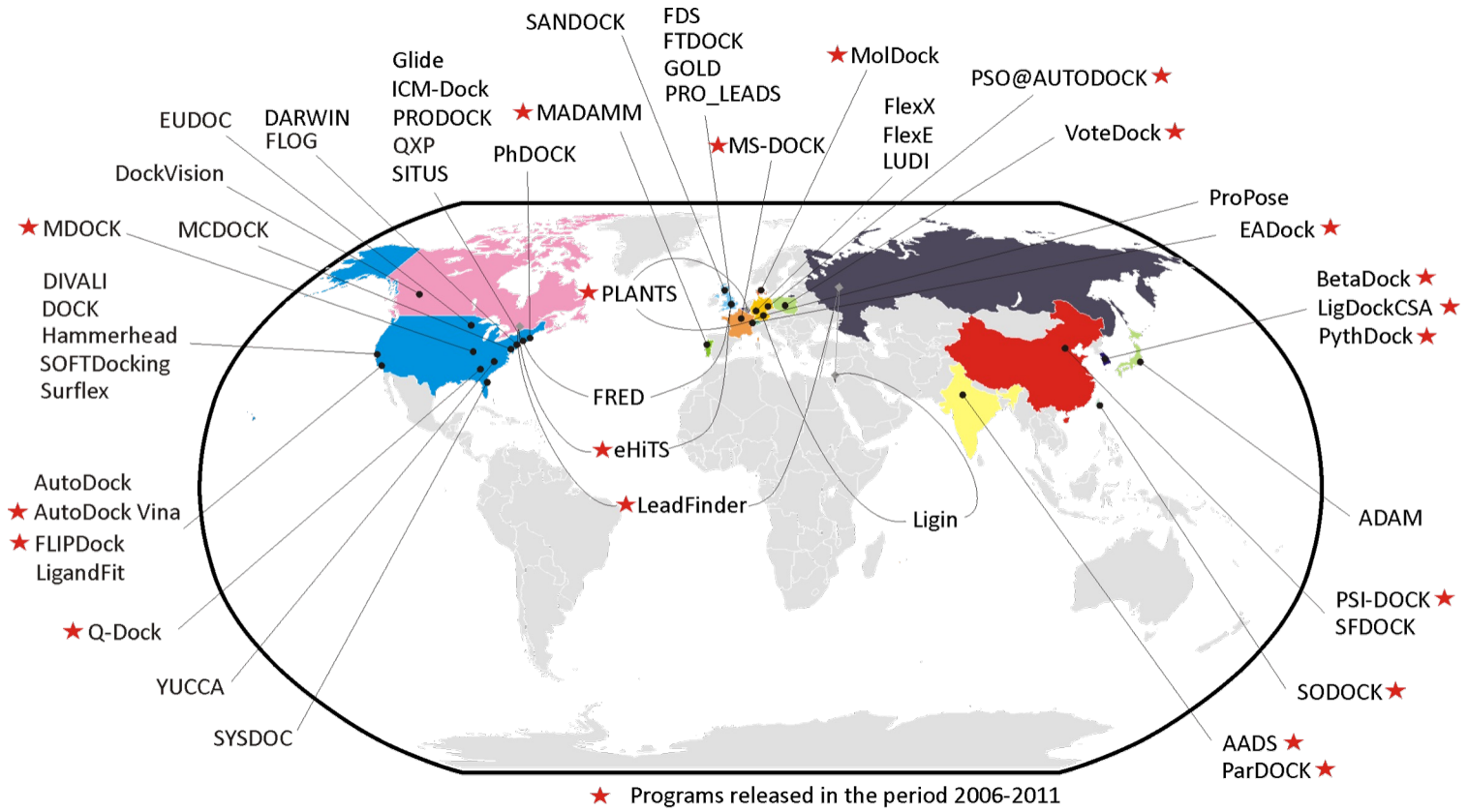


Other Alternatives in detail

Docking

Many Alternatives Available

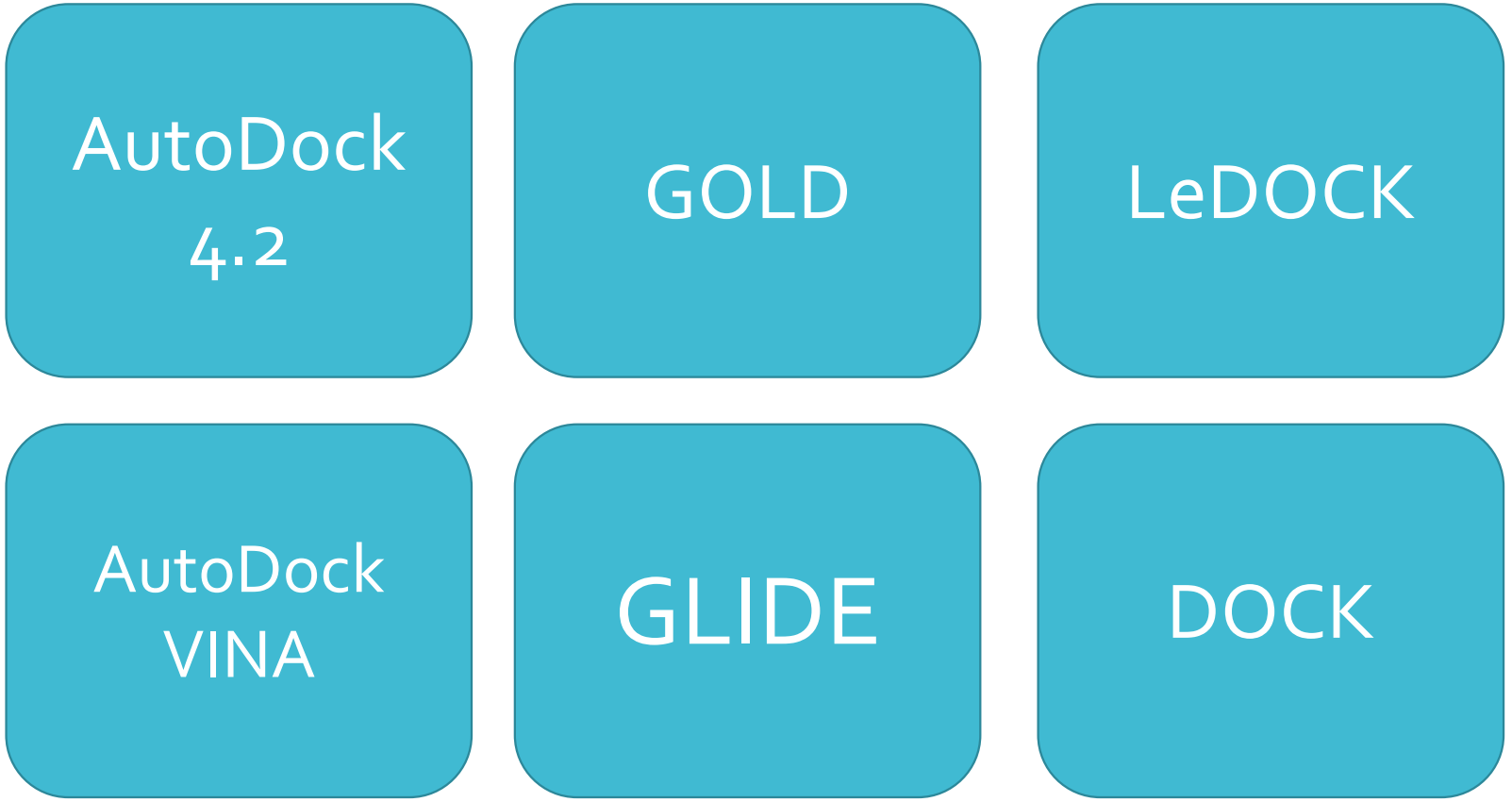
in Protein-Ligand Docking in the New Millenium – A Retrospective of 10 Years in the Field
Sérgio F. Sousa et al
Current Medicinal Chemistry, Vol. 20, pp. 2296-2314, (2013).



Docking

Many Alternatives Available

in Protein-Ligand Docking in the New Millenium – A Retrospective of 10 Years in the Field
Sérgio F. Sousa et al
Current Medicinal Chemistry, Vol. 20, pp. 2296-2314, (2013).



Docking

Many Alternatives Available

	Protein–ligand	Internal ligand
Gold	$E_{vdW} + E_{electrostatic} =$ $\sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$	$E_{vdW} + E_{electrostatic} =$ $\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$ + optional E_{H-bond}
AutoDock	$E_{vdW} + E_{H-bond} + E_{electrostatic} =$ $\sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}}\right) + E(t)\times\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$ $E(t) =$ angular weight factor	$E_{vdW} + E_{H-bond} + E_{electrostatic} =$ $\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}}\right) + E(t)\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + 332.0\frac{q_i q_j}{4(d_{ij}) d_{ij}}\right]$ $E(t) =$ angular weight factor
DOCK (v4.0)	$E_{vdW} + E_{electrostatic} =$ $\sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$	

- ChemScore:

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{hbond} \sum_{il} g_1(\Delta r) g_2(\Delta\alpha) + \Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{iL} f(r_{iL}) + \Delta G_{rot} H_{rot}$$

- GlideScore:

$$\Delta G_{bind} = C_{hbond-neut-neut} \sum g(\Delta r) h(\Delta\alpha) + C_{hbond-neut-charged} \sum g(\Delta r) h(\Delta\alpha) + C_{hbond-charged-charged} \sum g(\Delta r) h(\Delta\alpha) + C_{max-metal-ion} \sum f(r_{lm}) h(\Delta\alpha) + C_{lipo-lipo} \sum f(r_{lr}) + C_{rotb} H_{rotb} + C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} + solvation terms$$

- DrugScore:

$$\Delta W = \gamma \left[\sum_{k_i} \sum_{l_j} \Delta W_{i,j}(r) \right] + (1-\gamma) \times \left[\sum_{k_i} \Delta W_i(SAS, SAS_0) + \sum_{l_j} \Delta W j(SAS, SAS_0) \right]$$

Protein-Ligand Atom Distance Term

Solvent Accessible Surface Terms

- Parametrized Pairwise Potential (PMF) score:

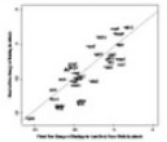
$$PMF = \sum_{prot} \sum_{lig} A_{ij}(d_{ij}) \quad A_{ij}(d_{ij}) = -k_B T \ln \left[f^{j}_{Vol\_corr}(r) \frac{\rho^{ij}_{seg}(r)}{\rho^{ij}_{bulk}} \right]$$

Boltzmann constant

Ligand volume correction factor

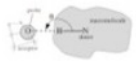
Radial distribution function for a protein atom i and a ligand atom j

- Autodock 4.0



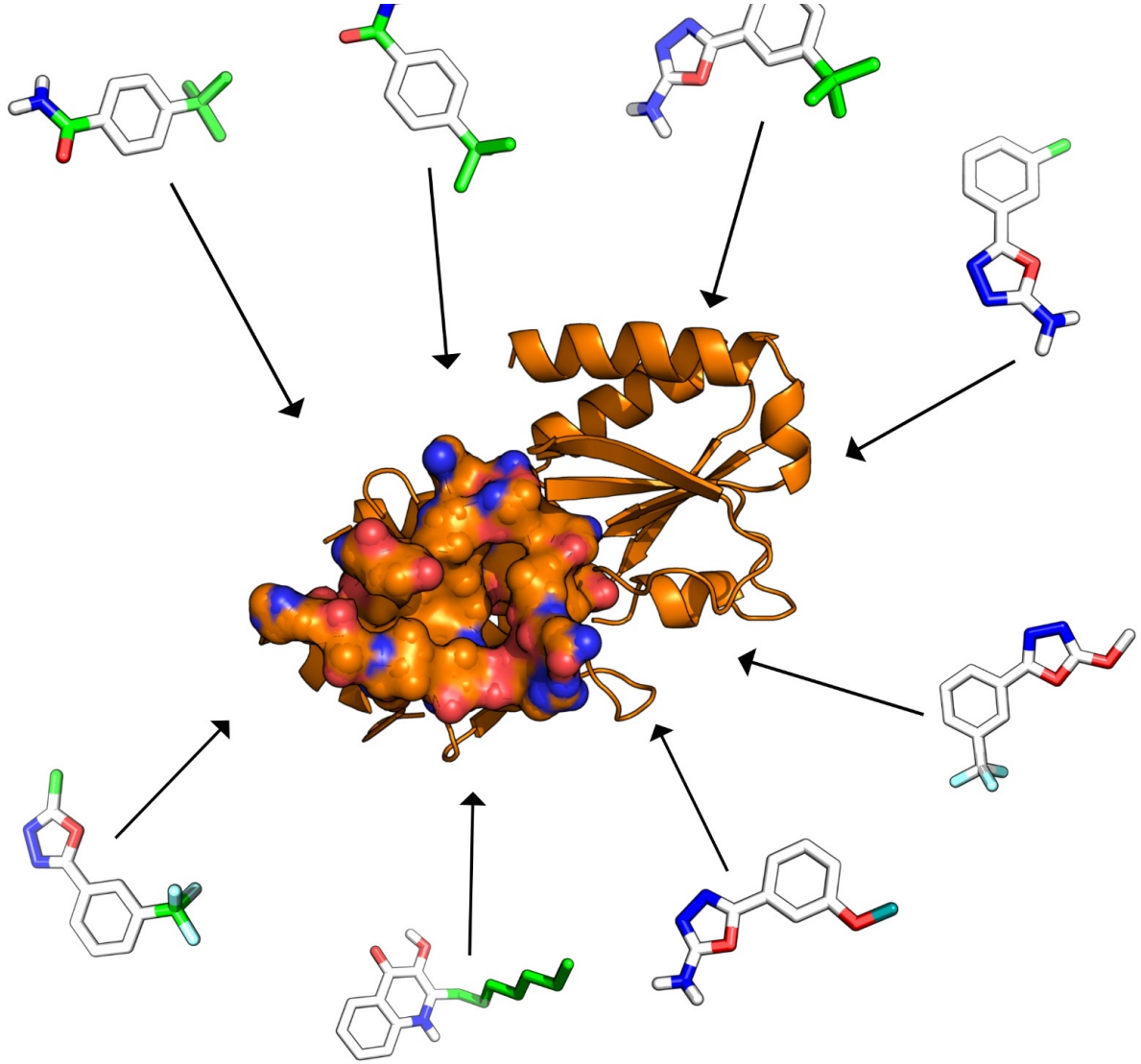
$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta g_{tors}$$

$$\varepsilon(r) = A + \frac{B}{1 + ke^{-\lambda B r}}$$



- $\Delta G_{vdW} = \Delta G_{vdW}$
  12-6 Lennard-Jones potential (with 0.5 Å smoothing)
- ΔG_{elec}
  with Solmajer & Mehler distance-dependent dielectric
- ΔG_{hbond}
  12-10 H-bonding Potential with Goodford Directionality
- ΔG_{desolv}
  Charge-dependent variant of Stouten Pairwise Atomic Solvation Parameters
- ΔG_{tors}
  Number of rotatable bonds

Structured Based Virtual Screening



Virtual Screening

Chimiothèque Nationale

MuTaLig Chemoteca

ZINC Database

PPi

...

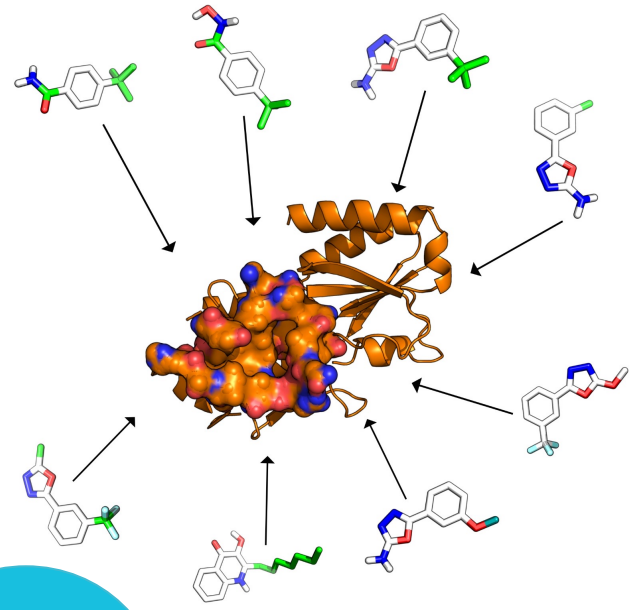
Virtual Libraries of Molecules

Virtual Screening

Hits

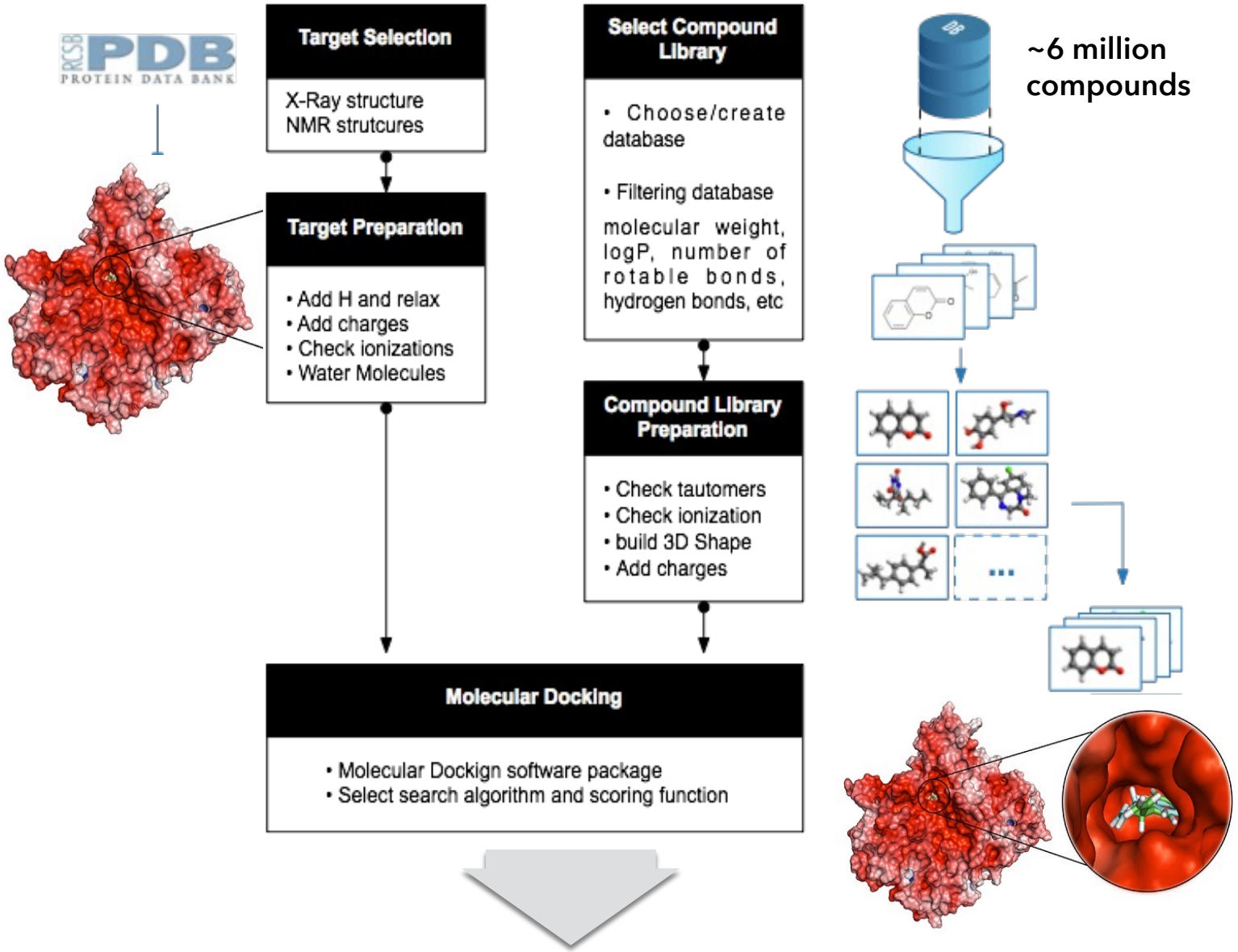
20-100 molecules

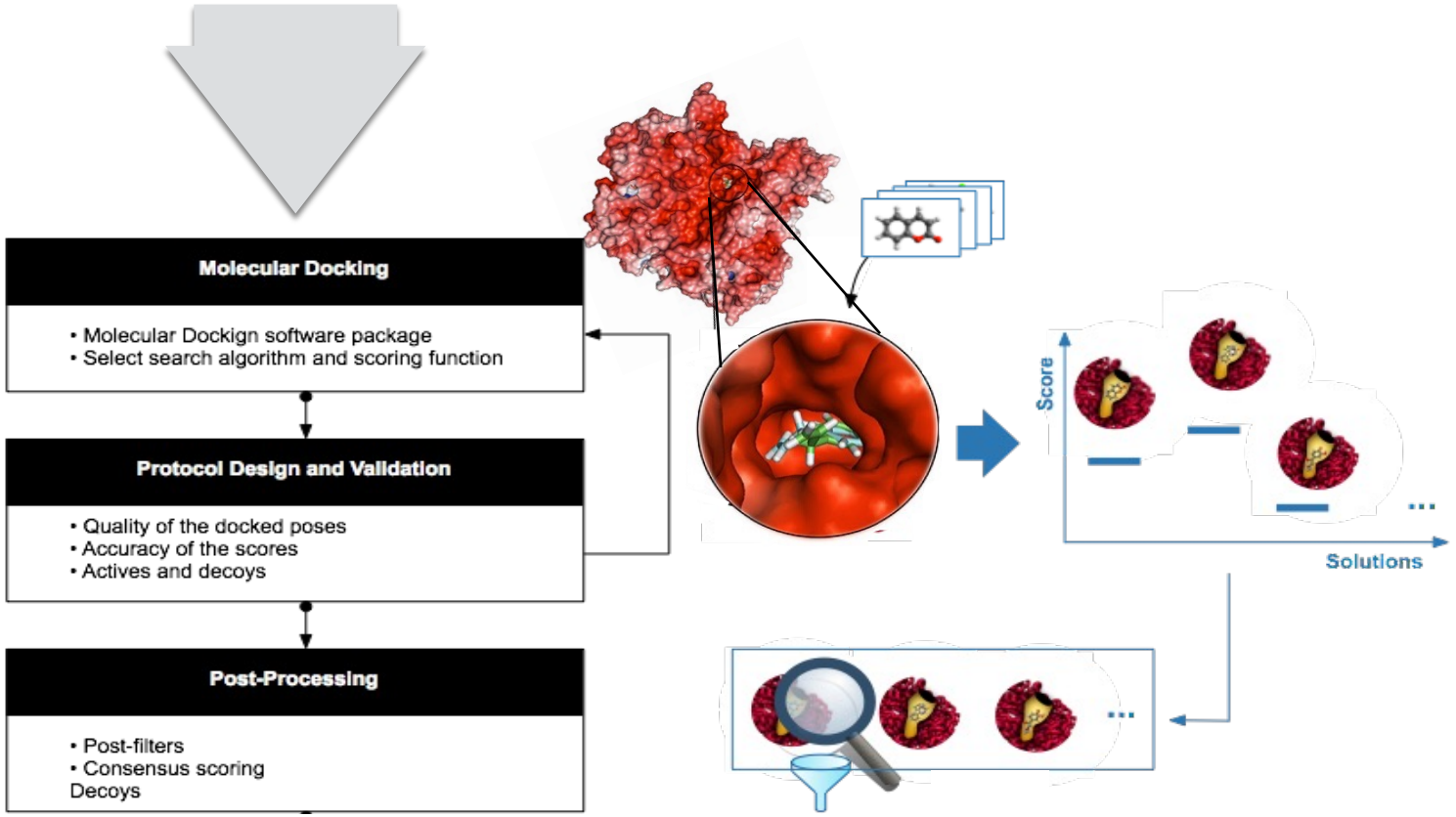
Experimental Evaluation



2-3 molecules (Leads)

Drug Development & Optimization


RCSB PDB
PROTEIN DATA BANK
Target Selection
X-Ray structure
NMR strutcures
Target Preparation
• Add H and relax
• Add charges
• Check ionizations
• Water Molecules
Select Compound Library
• Choose/create database
• Filtering database
molecular weight, logP, number of rotable bonds, hydrogen bonds, etc
Compound Library Preparation
• Check tautomers
• Check ionization
• build 3D Shape
• Add charges
Molecular Docking
• Molecular Dockign software package
• Select search algorithm and scoring function
~6 million compounds


Molecular Docking
• Molecular Dockign software package
• Select search algorithm and scoring function
Protocol Design and Validation
• Quality of the docked poses
• Accuracy of the scores
• Actives and decoys
Post-Processing
• Post-filters
• Consensus scoring
Decoys
Score
Solutions

Docking & Virtual Screening

Problems

False Positives:

Molecules that are erroneously suggested to bind strongly to the target

Can be easily discarded in the preliminary experimental studies with a relatively small cost

False Negatives:

Molecules that VS fails to identify as strong ligands, despite their high affinity

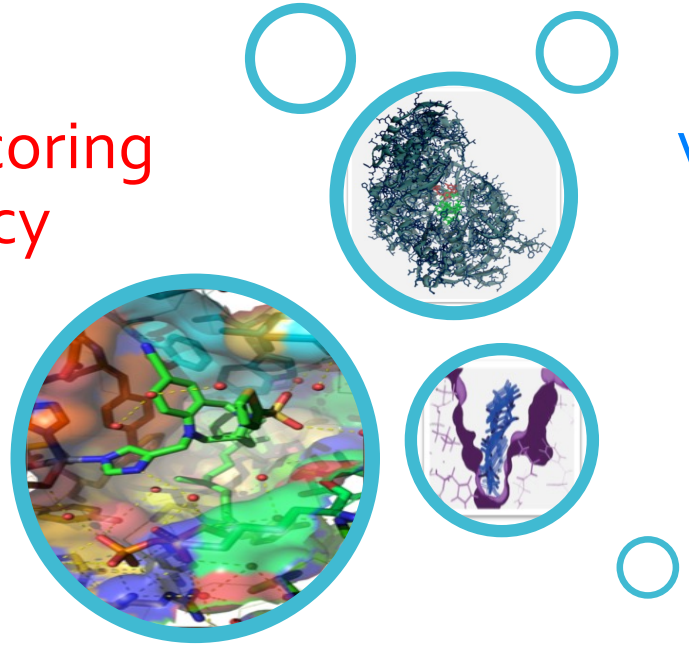
Never reach that stage remaining incognito among millions of compounds, despite their sometimes high potential pharmacological, social and economical value

Both Problems arise from the Imperfections in the Currently Available Scoring Functions!

Docking & Virtual Screening

Problems

Improve Scoring Accuracy

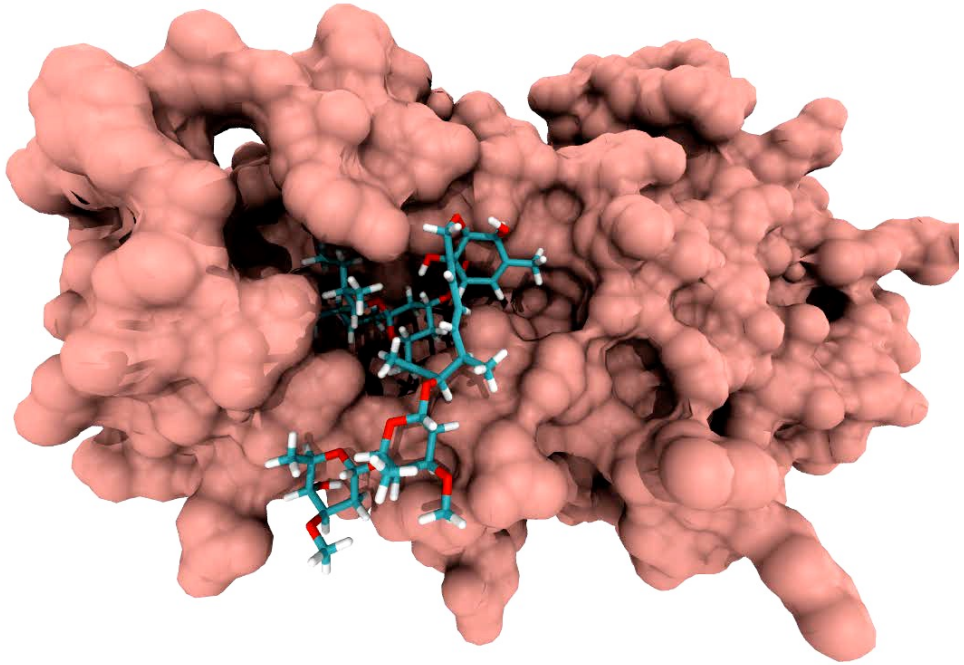


Varies with the Target

Varies with the Ligand

Both Problems arise from the Imperfections in the Currently Available Scoring Functions!

Virtual Screening Libraries



- Commercial Compounds (e.g. ZINC Db)
- Natural Compounds
- Compounds Extracted from Plants
- New Marine Compounds