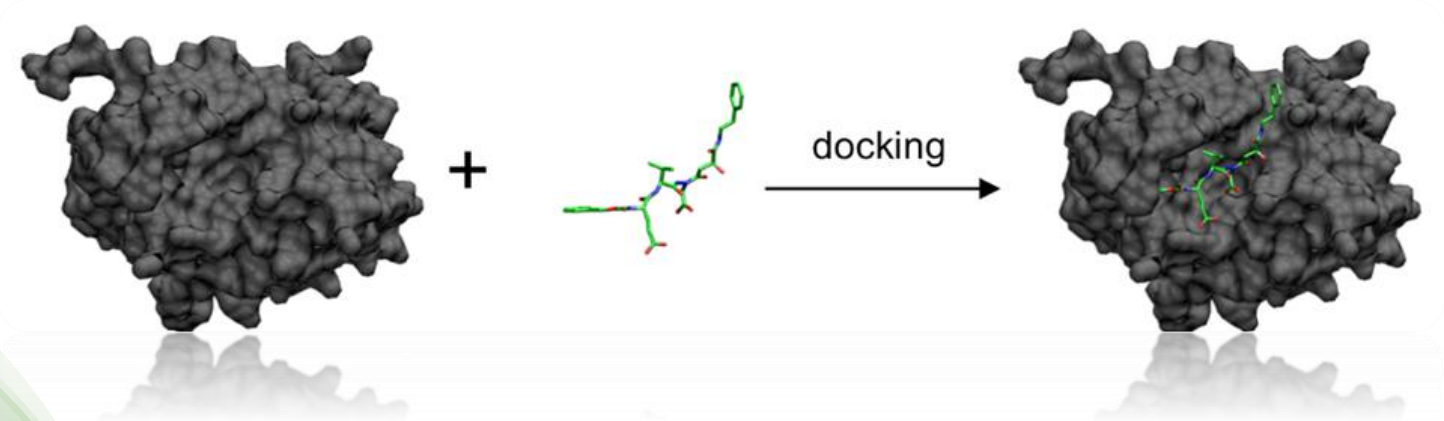
8th Advanced in silico Drug Design workshop 2025

Molecular Docking Lecture



Dr. Federica Moraca
Department of Pharmacy, University "Federico II" of Naples, Italy

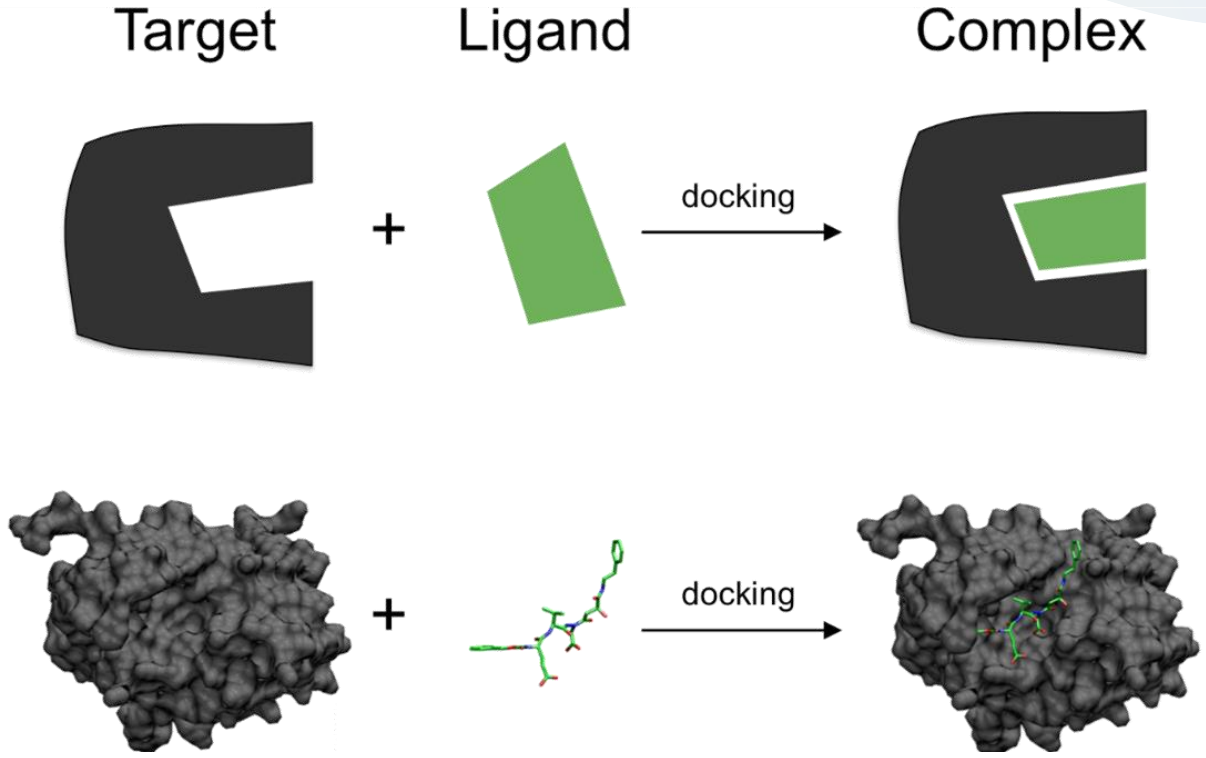
UP Olomouc 27.01. -31.01.2025

What is docking?

Some basic principles

Molecular docking is an important step of the drug discovery process, which aims at calculating the preferred position and shape of one **molecule (ligand) to its target (receptor), predicting its binding affinity.**

This step helps researches to study the behavior of small molecules, within the binding site of a target protein and understand the fundamental biochemical process underlying this interaction



Potential uses

- Drug discovery (Virtual Screening)
- Drug optimization/design
- Peptides optimization/design
- Nutraceutical research

Docking methodologies

Types of docking:

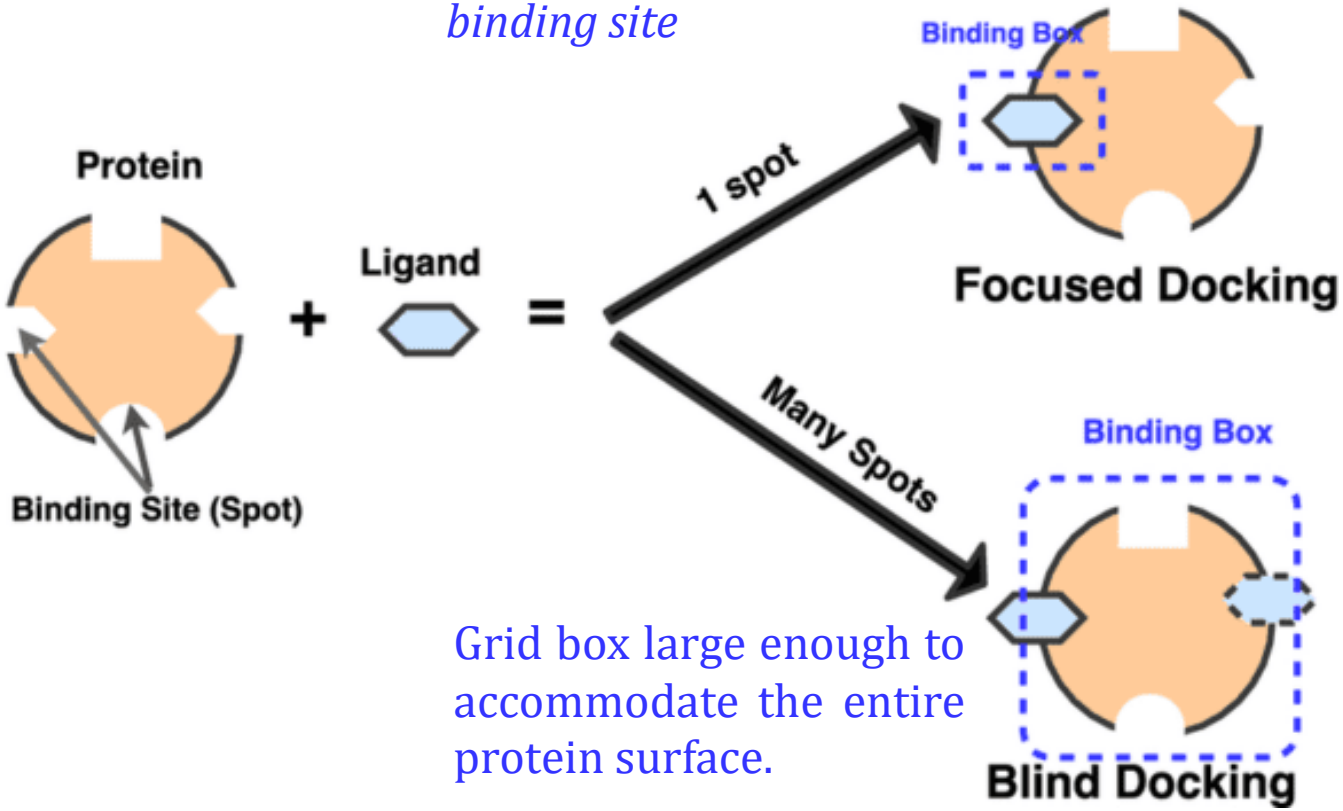
Blind (or Global) docking: the binding site is unknown. Docking is performed on the whole protein surface without any prior knowledge of the binding pocket.

Several trials/runs and several energy calculations are needed before a favorable protein-ligand complex pose is found.

Focused (or Local) docking: the binding site of the receptor is defined within a known binding pocket, so docking aims to find the geometry/position of the ligand in that binding site.

Docking accuracy is based on the input information of the binding site

Grid box need to be set up. *Tips: the center of the grid box in the center of the binding site*

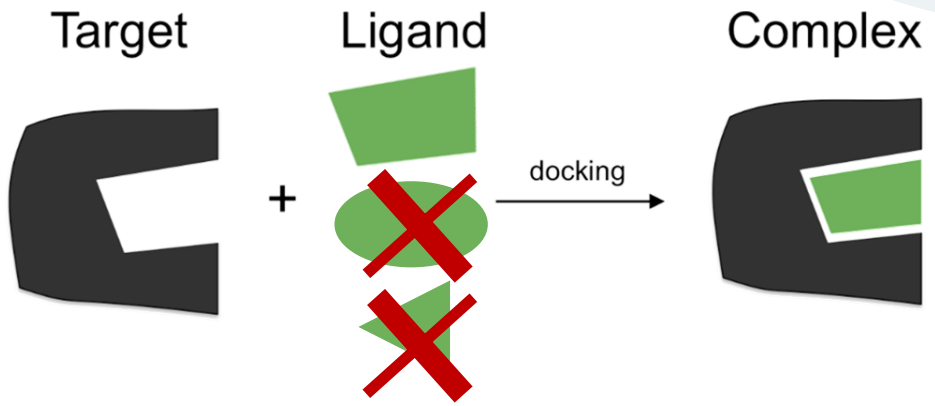


Grid box large enough to accommodate the entire protein surface.

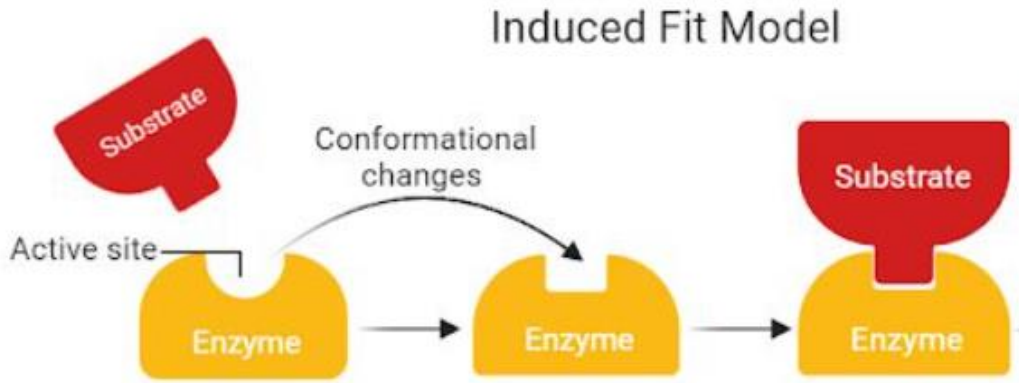
Docking methodologies

Types of docking:

Rigid docking (lock and the key):
complementary geometric shapes that fit perfectly like a 'key in a lock' (Emil Fisher, 1894).



Flexible docking (induced fit): both the ligand and target, adapt to one another by modest conformational changes until an ideal match is reached. (Daniel Koshland, 1958)

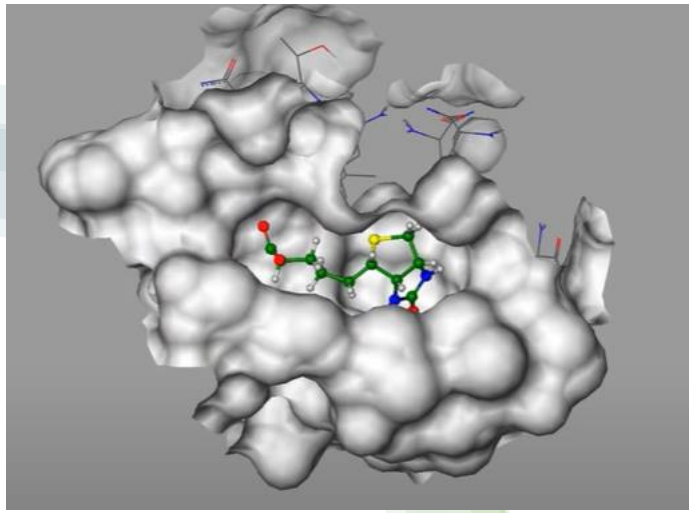


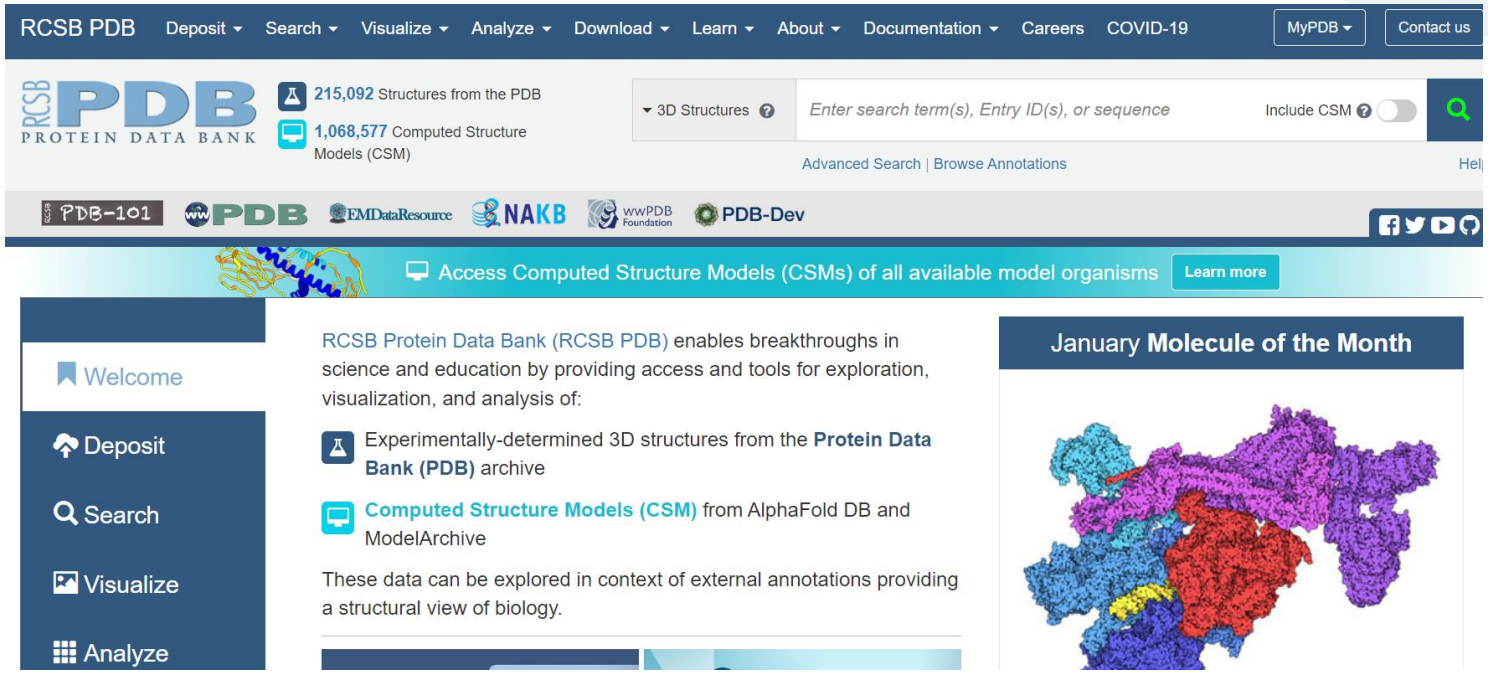
Where to start?

- **Ideally a GOOD Crystal structure**

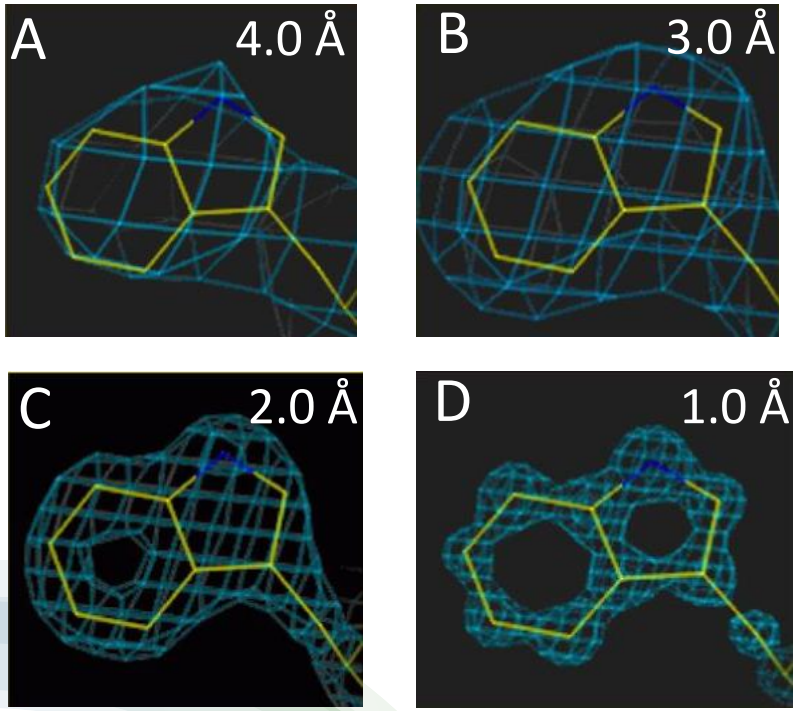
- **Usually worse: a homology model**

- **A ligand: usually the co-crystallized molecule**



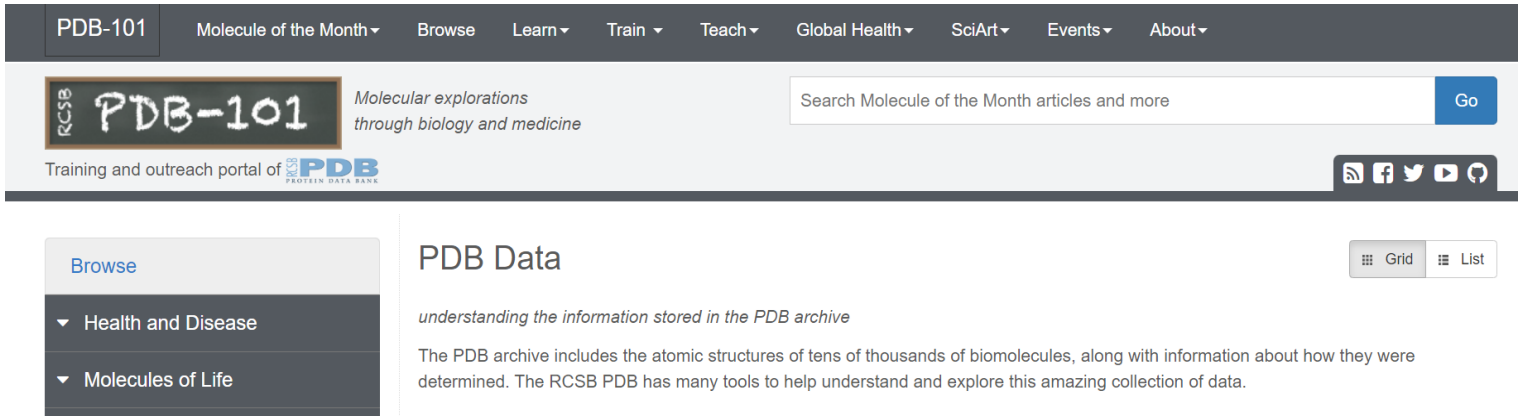


A word about the quality of the PDB structure



X-Ray Electron Density Maps

“PDB Educational Portal”



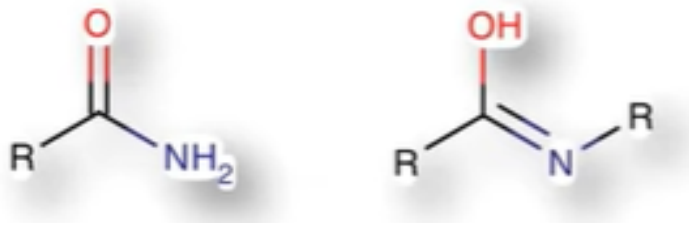
pdb101.rcsb.org/browse/pdb-data

How to treat protons?

X-Ray structures do not have H+ information.

We must “predict” them as good as possible

Not only H+ states but also tautomers are required!

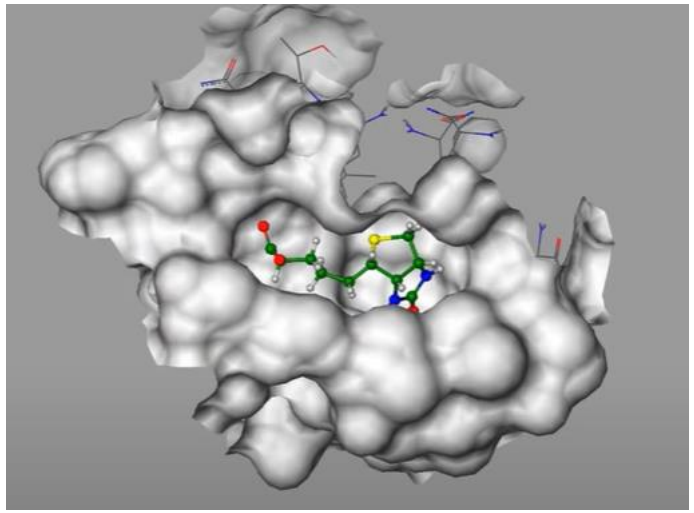


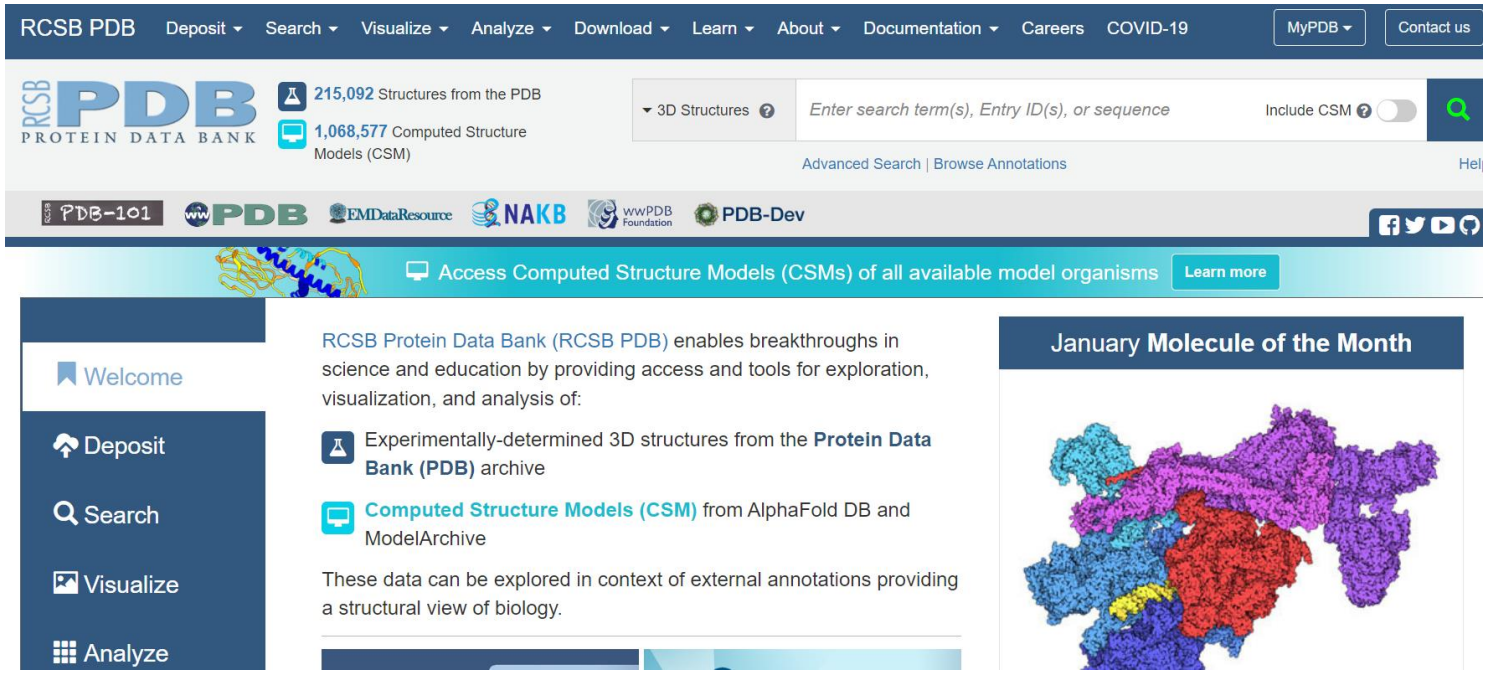
Where to start?

- **Ideally a GOOD Crystal structure**

- **Usually worse: a homology model**

- **A ligand: usually the co-crystallized molecule**





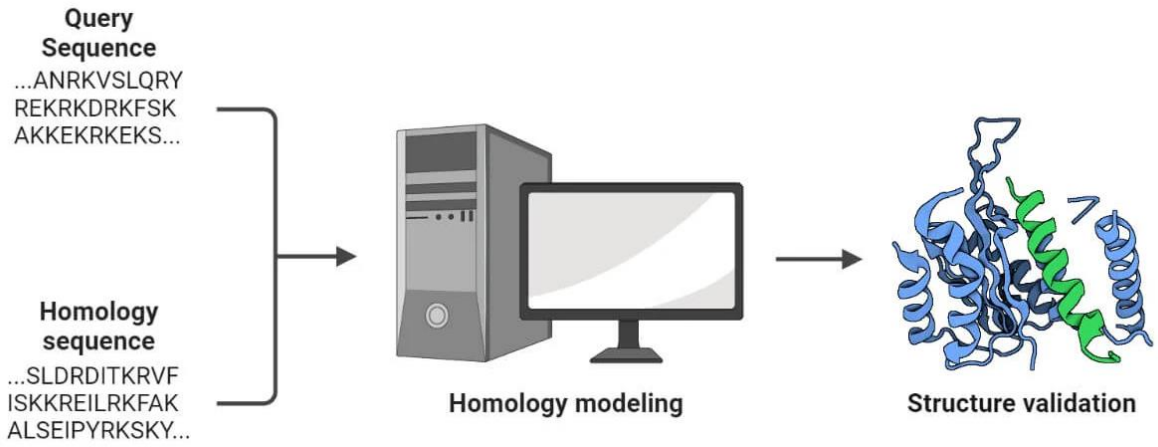
Where to start?

Homology modeling

Predict the three-dimensional (3D) structure of a protein or other macromolecule based on its amino acid sequence and the known structure of a related protein.

Assumptions

proteins with similar sequences frequently assume similar structures and functions.



Phyre2.2

I-TASSER
Protein Structure & Functi ons

PSIPRED

SWISS-MODEL

3dRNA/DNA

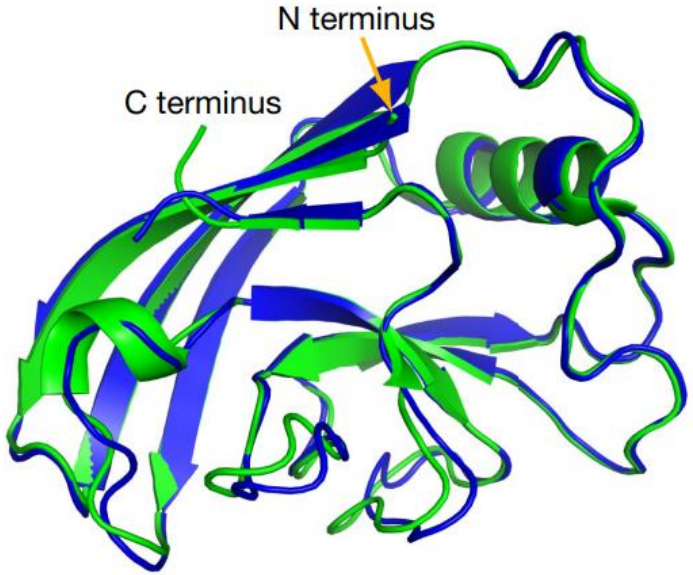
RNAstructure
Web Servers for RNA Secondary Structure Prediction

Kelley LA *et al. Nature Protocols* 10, 845-858 (2015)

Where to start?

Homology modeling limitations

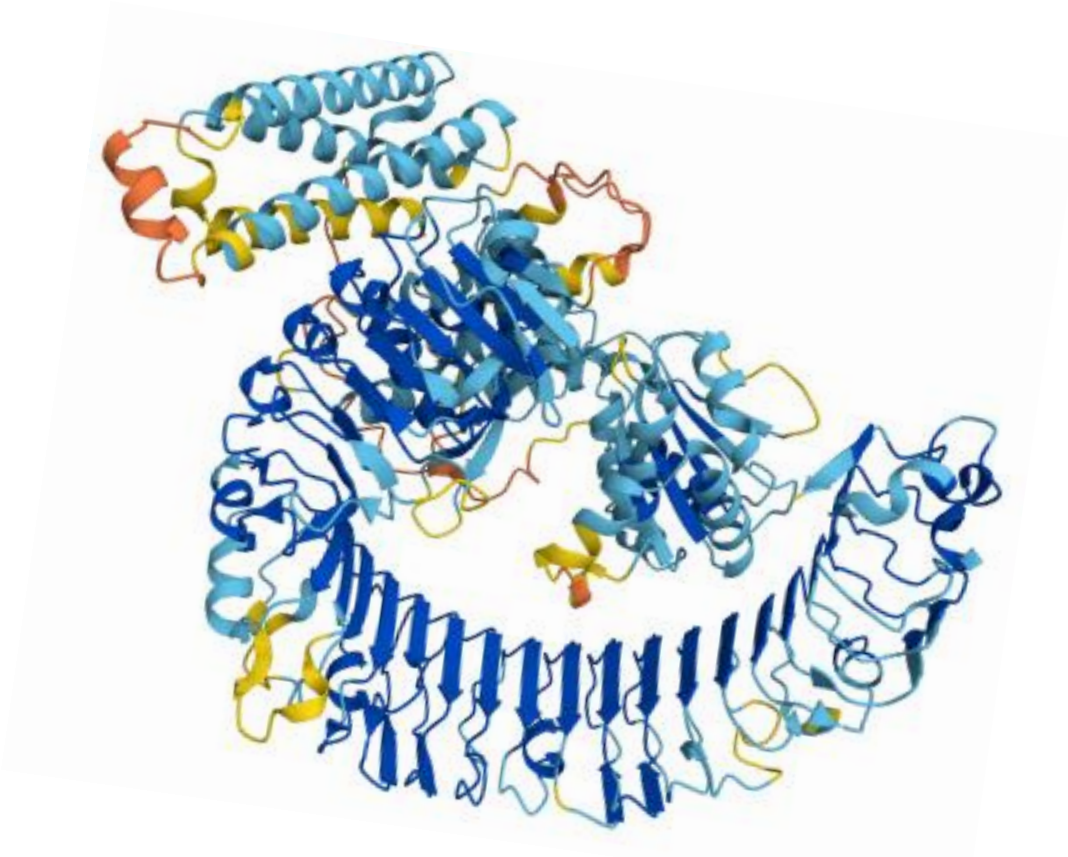
The sequences of two homologous proteins may diverge significantly and consequently may be undetectable by standard approaches.



AlphaFold Experiment

r.m.s.d.$_{95}$ = 0.8 Å; TM-score = 0.93

AlphaFold Protein Structure Database

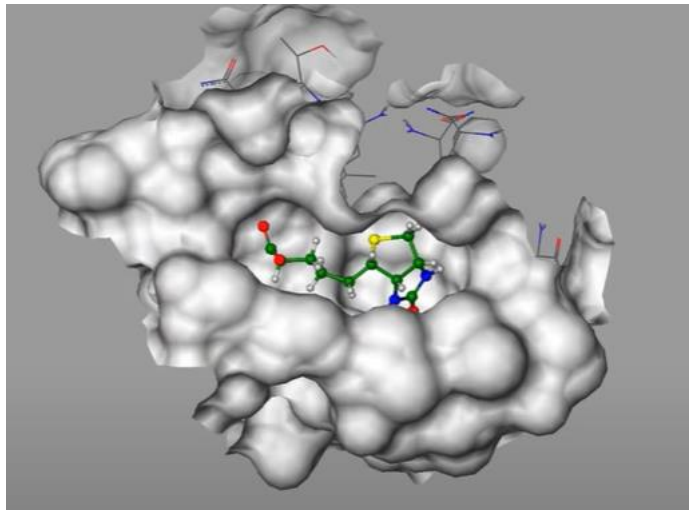


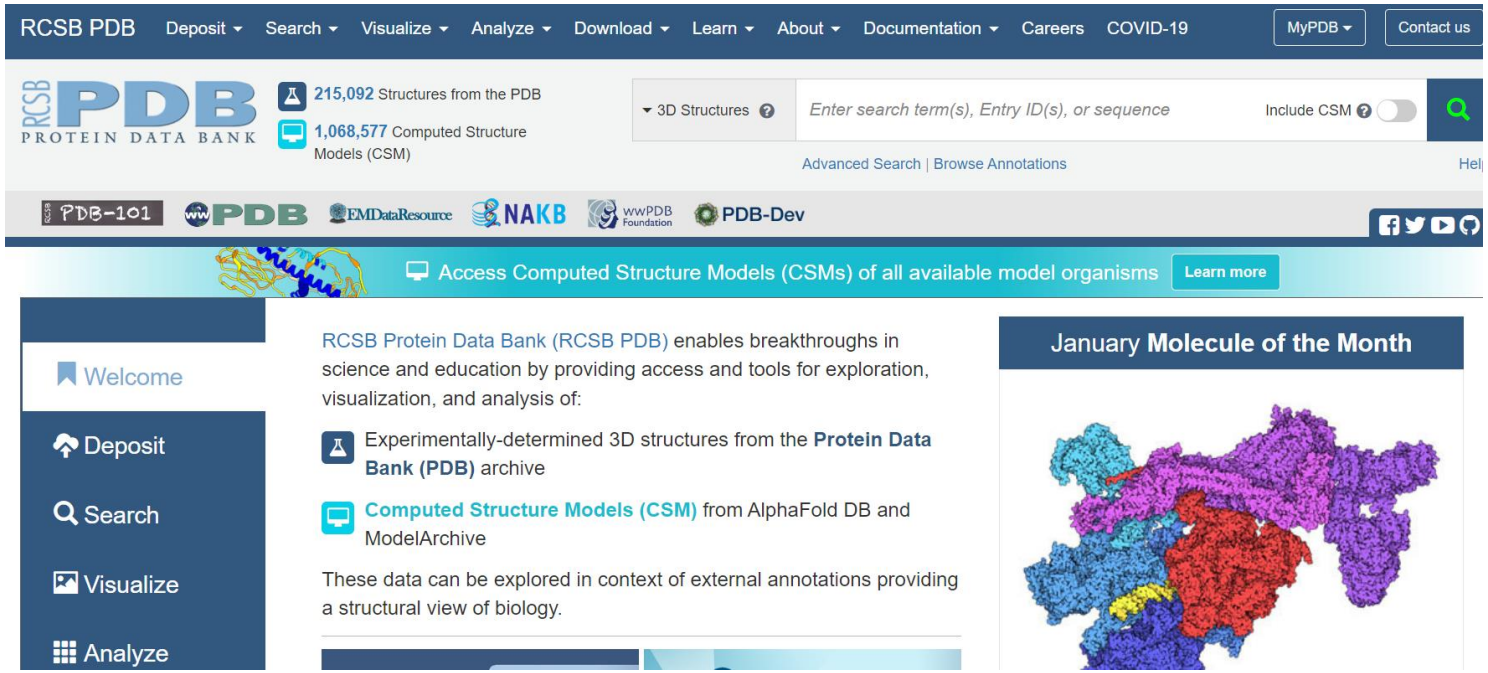
Where to start?

- **Ideally a GOOD Crystal structure**

- **Usually worse: a homology model**

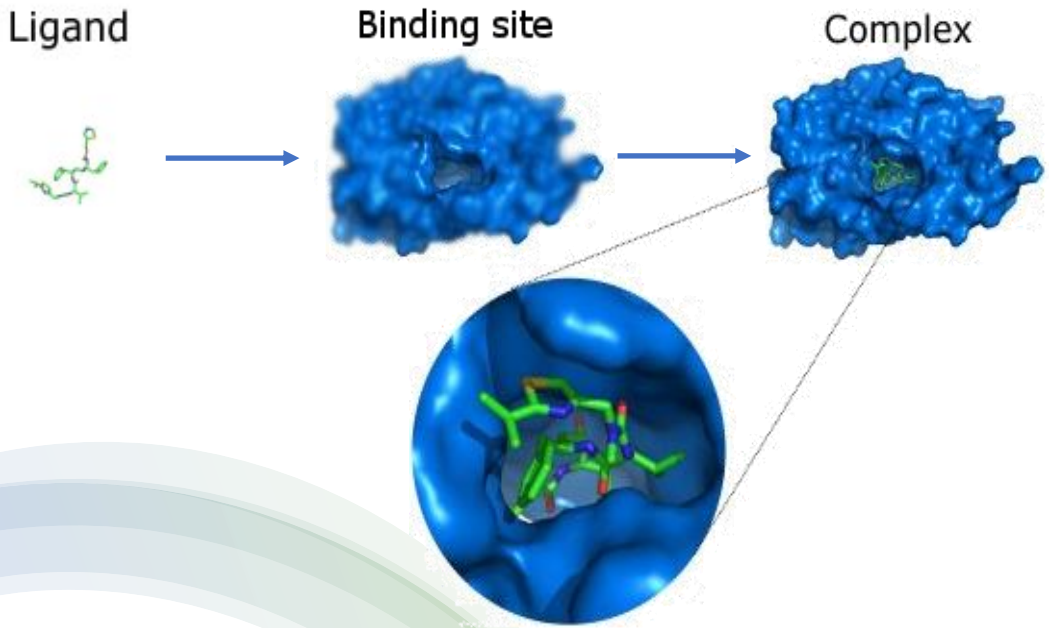
- **A ligand: usually the co-crystallized molecule**





Getting the Ligand into the Pocket

What are the questions/problems?



$$[L] + [P] \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} [LP] \qquad K_b = \frac{k_{on}}{k_{off}}$$

- Translation *(T)* & Rotation *(R)* of ligand needs to be performed

  => An optimization problem in T and R space

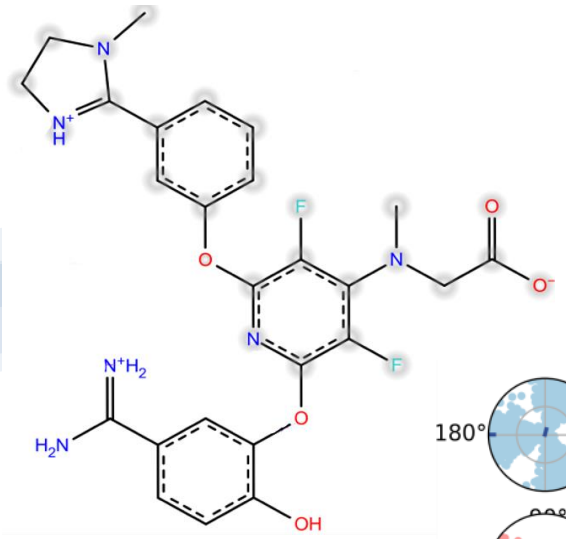
- Torsions will have to adopt to put the ligand into the pocket

  => An optimization problem in "Φ space"

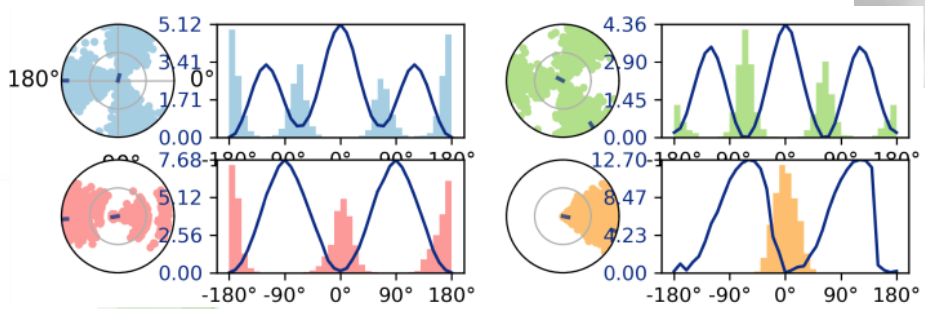
Ligand conformational sampling

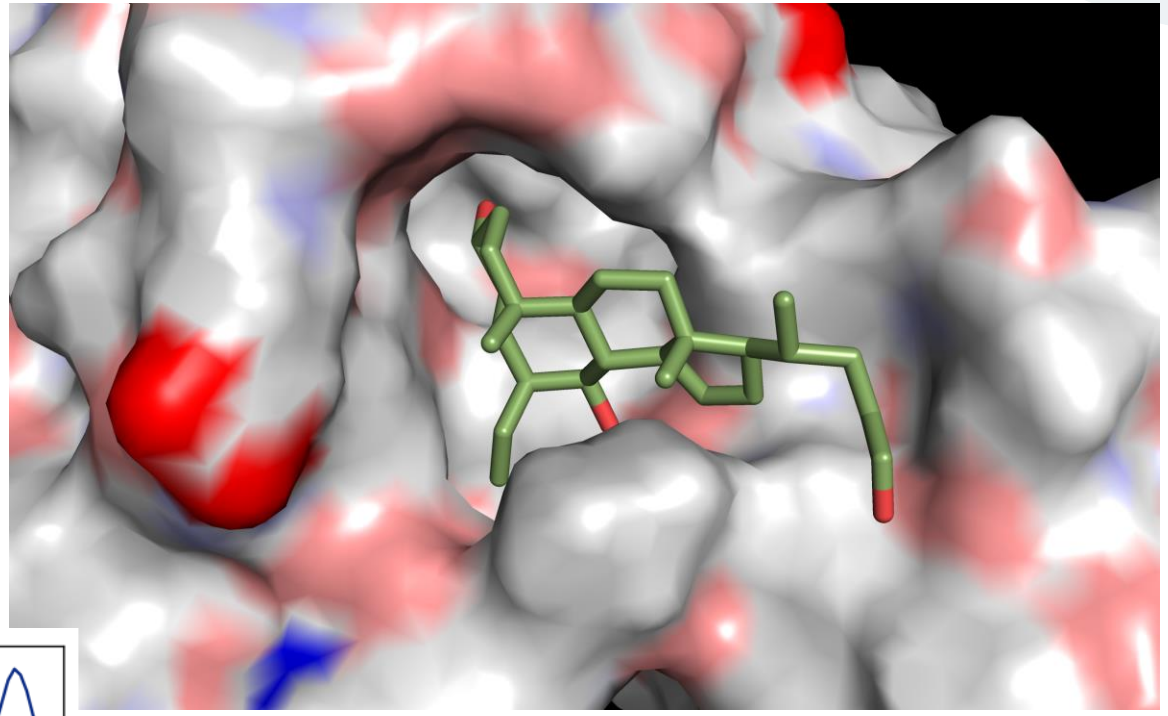
Ligand sampling is the most basic element in protein-ligand docking.

the sampling algorithm generates putative ligand orientations/conformations (*i.e.*, poses) around the chosen binding site of the protein.



Ligand Torsion Profile

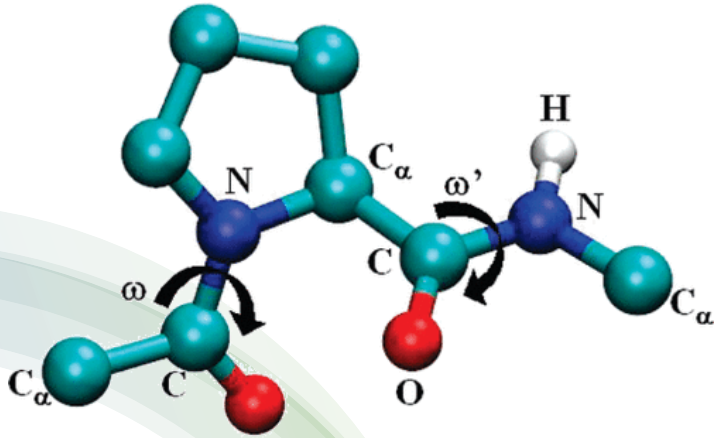




Coleman RG et al. PLoS One. 2013

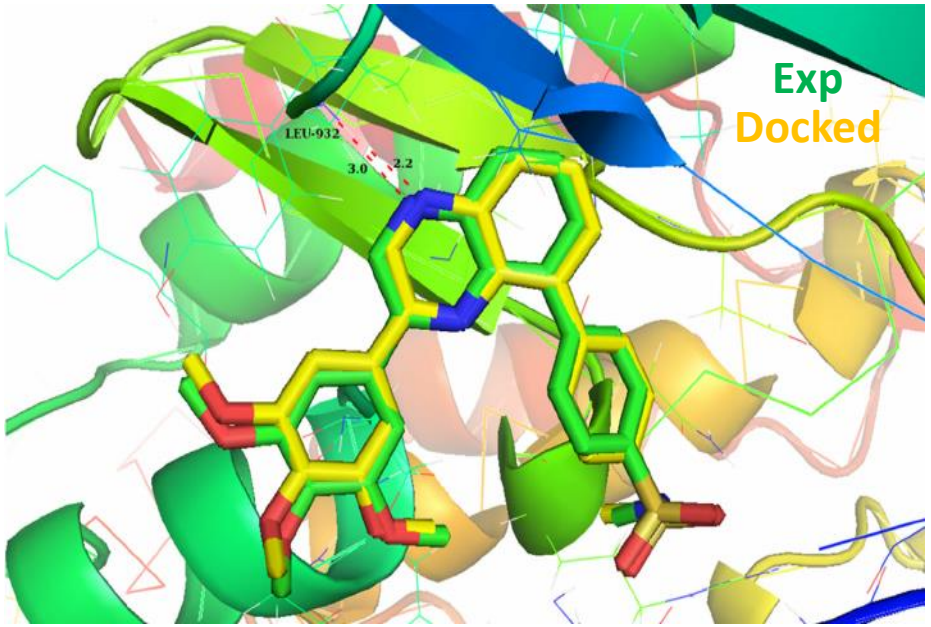
Two pillars of docking

Search algorithms:
Sampling/search algorithms help to identify the most energetically **favorable conformations of the ligand** within the protein's active site, taking into account their binding mode.



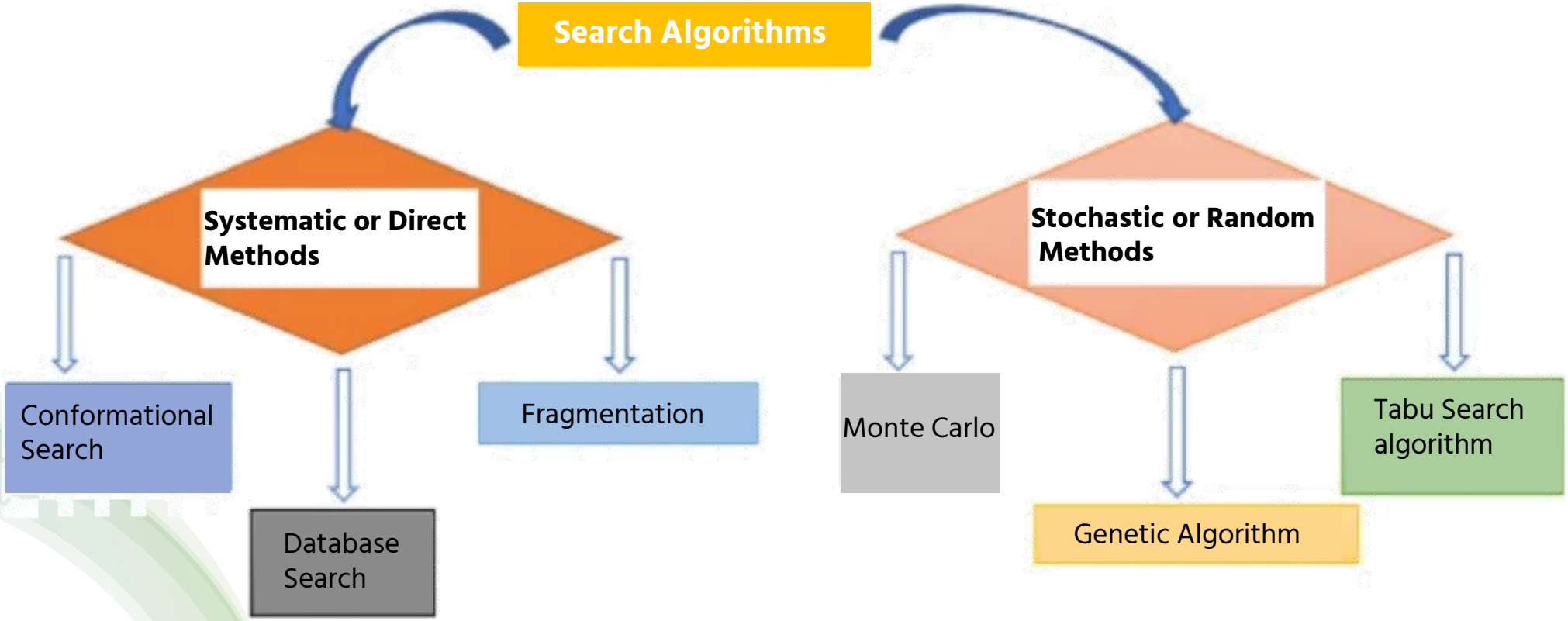
$$V_{torsional} = \sum_{dihedrals} \frac{V_n}{2}\left[1+\cos(n\omega-\gamma)\right]$$

Scoring functions:
to evaluate the ligand-receptor interactions in a way that **may discriminate the experimentally observed mode from others and estimate the binding affinity**.



Search Algorithms

Classification of the search algorithms

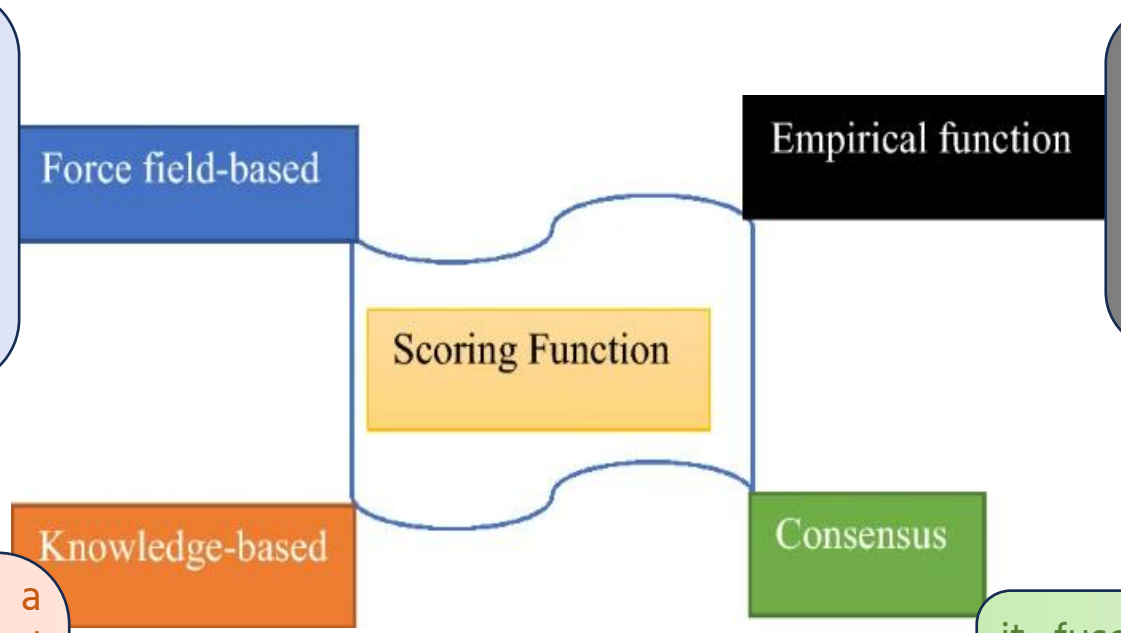


Agu, P.C. *et al. Sci. Rep.*, **13**, *2023*, 13398

Scoring Functions

Classes of scoring function mechanisms

Adds the contribution of non-bonded interactions including van der Waal forces, hydrogen bonding, and Columbic electrostatics, and bond-like angle bonding and torsional deviation. (i.e AutoDock, DOCK)



It relies on repeated linear relapse analysis of a prepared set of complex structures using protein–ligand complexes with known binding affinities (i.e AutoDock, DOCK)

Exploits structural information of a collection of known protein–ligand complexes (i.e stored in Protein Data Bank) to provide a score that is close (or proportional) to the experimentally derived value (good balance between accuracy and calculation efficiency) (i.e AutoDock Vina)

it fuses the evaluations or orders acquired through multiple evaluation methods in various arrangements.

Speed *vs* accuracy

Agu, P.C. *et al. Sci. Rep.*, **13**, *2023*, 13398

Scoring Functions

Classes of scoring function mechanisms

Interdisciplinary Sciences: Computational Life Sciences (2019) 11:320–328
https://doi.org/10.1007/s12539-019-00327-w

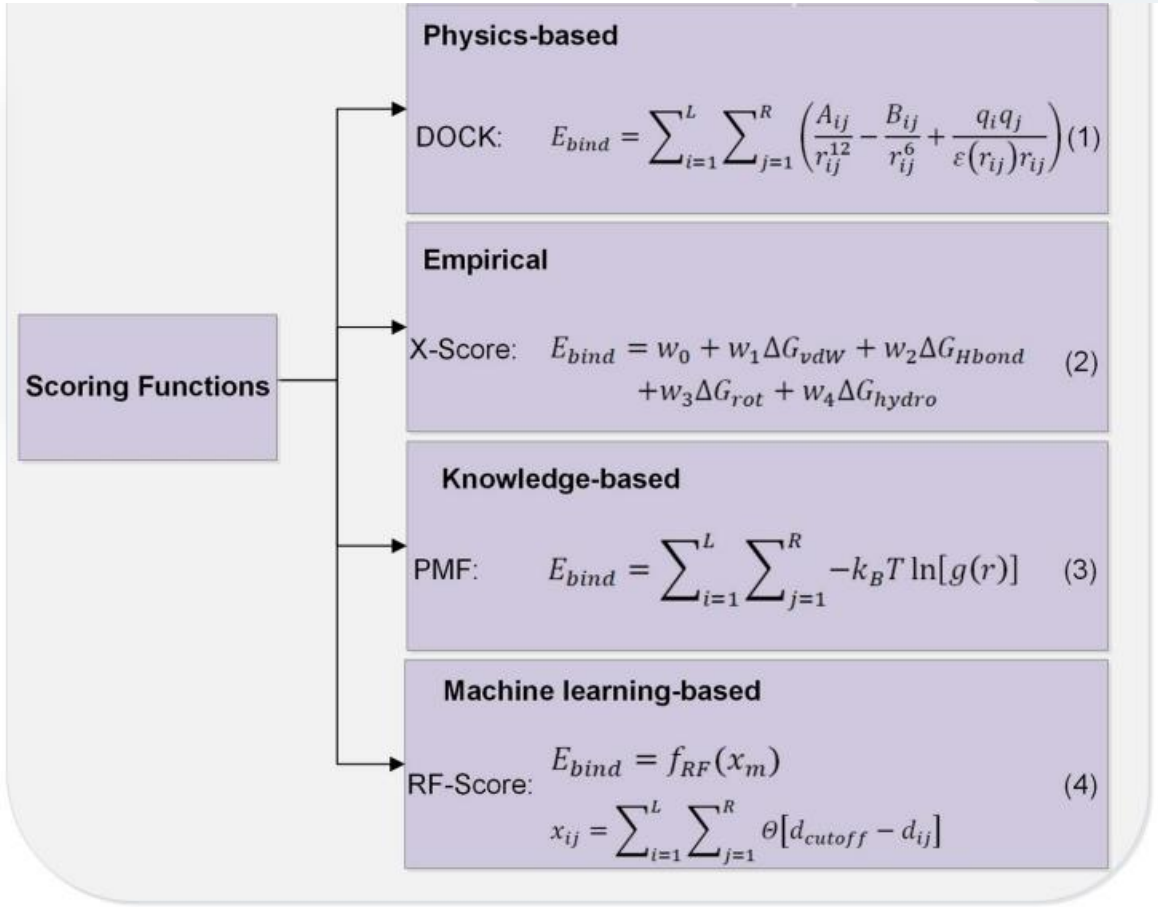
REVIEW

Check for updates

An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking

Jin Li[1,2] · Ailing Fu[3] · Le Zhang[1,4,5,6]

The reliability of molecular docking depends on the **accuracy of the adopted scoring function**, which can guide and determine the ligand poses when thousands of possible poses of ligand are generated.



Agu, P.C. *et al. Sci. Rep.*, **13**, *2023*, 13398

Force Field based Scoring Functions

Assumptions

Affinities are estimated by summing the strength of intermolecular van der Waals and electrostatic interactions between all atoms of the two molecules in the complex. In addition, the desolvation energies of the ligand and of the protein are also take into account

Advantages

- FF terms are well studied and have physical basis

Disadvantages

Electrostatics often are overestimated, leading to systematic problems in ranking complexes

General function form

AutoDock 4

$$\Delta G_{binding} = \Delta E_{vdw} + \Delta E_{electrostatic} + (\Delta E_{H\text{-}bond}) + \Delta G_{desolvation}$$

Coulombic terms

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij})\, r_{ij}} \right)$$

VDW parameters

Agu, P.C. *et al. Sci. Rep.*, **13**, *2023*, 13398

Empirical Scoring Functions

Assumptions

Are developed to reproduce experimental affinity data based on the idea that it is possible to correlate the free energy of binding to a set of non-related variables.

The ligand *internal-energy* is related to the loss of flexibility of the ligand upon binding, and consequently, to the reduction of the number of ligand accessible conformations upon binding that promotes the **"entropic loss" that is unfavorable to the binding affinity.**

Advantages

Fast & direct estimation of the binding affinity

Disadvantages

- Discrepancy in the binding affinity
- Heavy dependence on the placement of hydrogen atoms

General function form

Schrödinger

GlideScore:

$$\Delta G_{bind} = \Delta G_{lipophilic} + \Delta G_{coulomb} + \Delta G_{h\text{-}bond} + \Delta G_{vdW} + \Delta G_{rot} + \Delta G_{aromatic} + \Delta G_{int\text{-}energy} + \Delta G_{solvation}$$

Agu, P.C. *et al. Sci. Rep.*, **13**, *2023*, 13398

Knowledge-based Scoring Functions

Statistical information from the PDB complexes

Assumptions

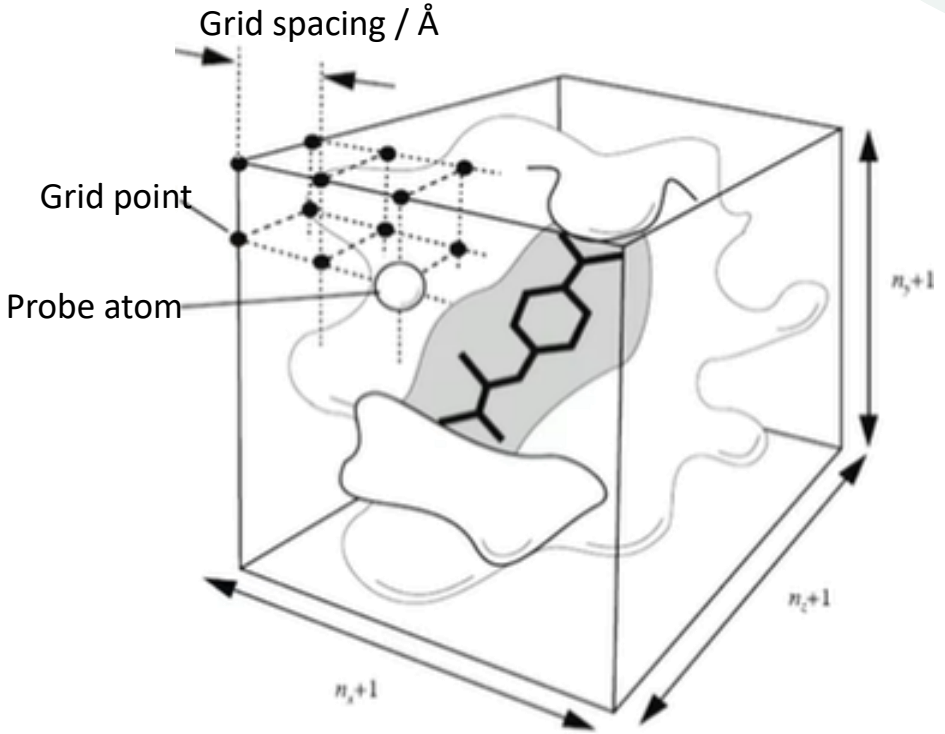
- An observed crystallographic complex represents the optimum placement of the ligand atoms to the receptor atoms
- The Boltzmann hypotheses convert such frequencies into an effective interaction energy.
- Designed to reproduce the experimental structures rather than binding energies

Advantages

Similar to empirical, but more general (much more distance data than binding energy data)

Disadvantages

PMF are typically pair-wise, while the probability to find atoms A and B at a distance *r* is non-pairwise and depends also on surrounding atoms



Boltzmann constant

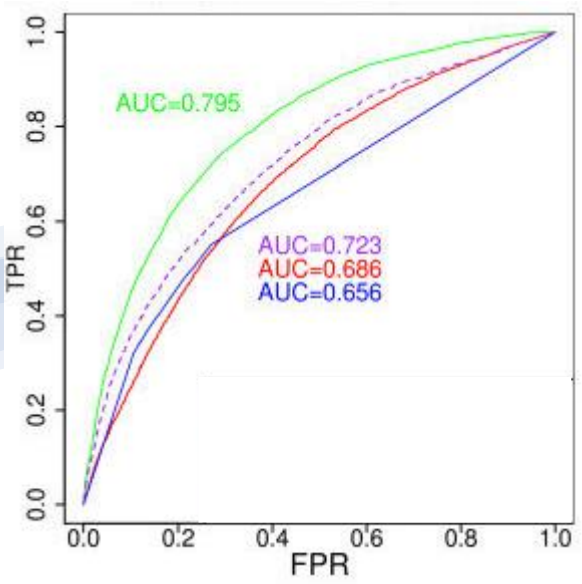
$$w(r) = -\kappa_B T \log[g(r)], g(r) = \frac{\rho(r)}{\rho * (r)}$$

Strengths and weaknesses of classical scoring functions

Can docking scoring functions guarantee success in virtual screening?

Assumptions

A perfect docking or VS protocol should be able to rank only true hits at the beginning of its ranked list



Real applications

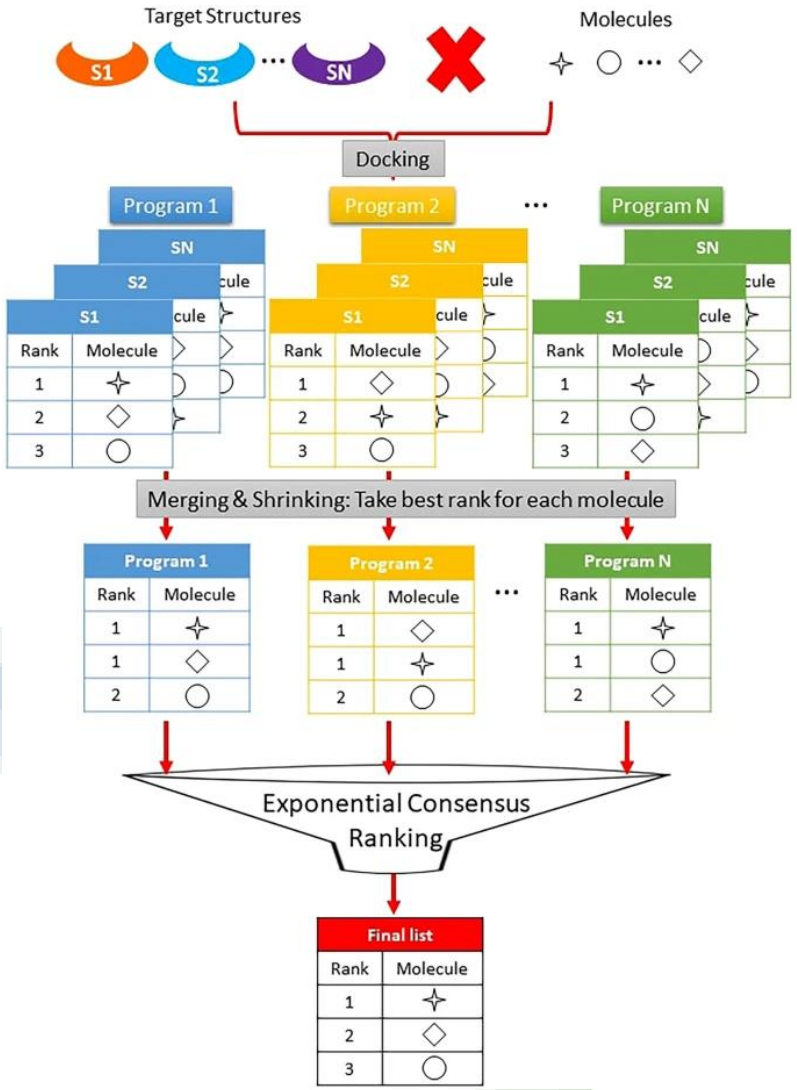
This is not the case because of the complexity of the molecular recognition process

scoring functions that make approximations to evaluate the binding affinity between receptors and a small molecule

Possible solution

Comparative studies of scoring functions

Consensus Docking to Improve the Reliability of Docking



consensus scoring is a valuable method for obtaining consistent hit-rates across diverse targets, while **reducing the number of false positive.**

Exponential Consensus Ranking (ECR)

$$p(r_i^j) = \frac{1}{\sigma} \exp\left(-\frac{r_i^j}{\sigma}\right),$$

$$P(i) = \sum_j p(r_i^j) = \frac{1}{\sigma} \sum_j \exp\left(-\frac{r_i^j}{\sigma}\right)$$

The final score of each molecule *i* is defined as the sum of the exponential score for all of the scoring functions *j*

Palacio-Rodríguez, K. et al. *Sci Rep* **9**, 5142 (2019)

Docking Software: what to know

- Sensitivity of the parameters (including the starting conformation)

- Adaptability to additional scoring functions

- Ability for iteratively refining docking parameter

- Speed, user interface, I/O structural file formats

- Code availability and ugrading possibily

Docking Software: what to know

International Journal of *Molecular Sciences*

MDPI

Review

Key Topics in Molecular Docking for Drug Design

Pedro H. M. Torres [1], Ana C. R. Sodero [2], Paula Jofily [3] and Floriano P. Silva-Jr [4,*]

Software	Posing	Scoring	Availability
Vina	Iterated Local Search + BFGS Local Optimiser	Empirical/Knowledge-Based	Free (Apache License)
AutoDock4	Lamarckian Genetic Algorithm, Genetic Algorithm or Simulated Annealing	Semiempirical	Free (GNU License)
GOLD	Genetic Algorithm	Physics-based (GoldScore), Empirical (ChemScore, ChemPLP) and Knowledge-based (ASP)	Commercial
Glide	Systematic search + Optimisation (XP mode also uses anchor-and-grow)	Empirical	Commercial
Surflex	Fragmentation and alignment to idealised molecule (Protomol) + BFGS optimisation	Empirical	Commercial

Brief Introduction to the AutoDock Suite

The **AutoDockSuite**, is free open source software for the computational docking of small molecules to macromolecular receptors.

Complementary docking engines

AutoDock4 — general-purpose docking of ligands to proteins

AutoDockVina — rapid docking of ligands

AutoDockFR — docking with flexible receptors

AutoDockCrankPep — docking of peptide ligands

Tools&Methods

Graphical User Interfaces - AutoDockTools (ADT)
- Raccoon2

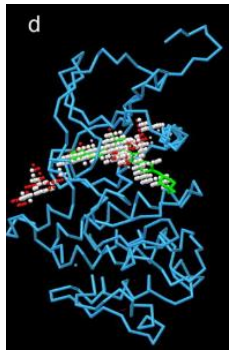
Specialized Docking Methods - Covalent Docking

Active Site Prediction - AutoLigand
- AutoGrid

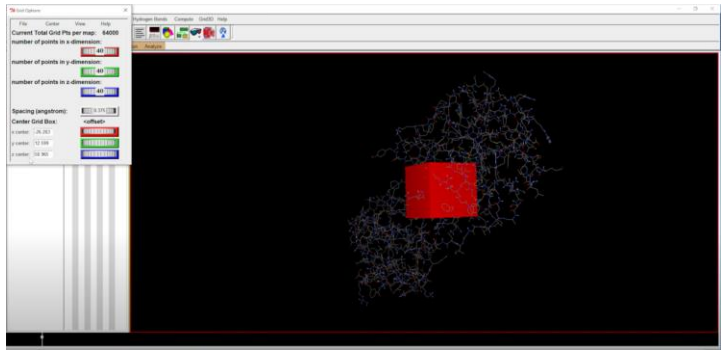
PROTOCOL

Computational protein–ligand docking and virtual drug screening with the AutoDock suite

Stefano Forli, Ruth Huey, Michael E Pique, Michel F Sanner, David S Goodsell & Arthur J Olson

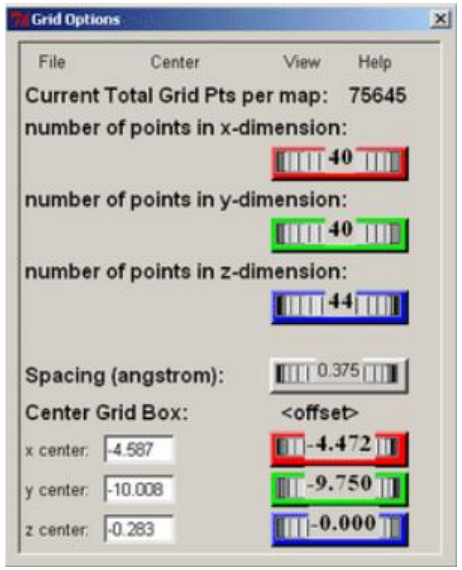


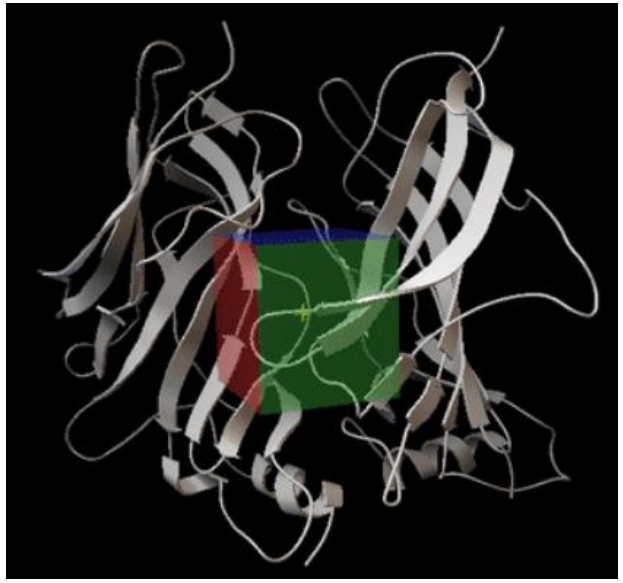
AutoLigand


Graphical User Interface of ADT

Brief Introduction to AutoDock4

AutoDock4 is a free open source software for the computational docking of small molecules to macromolecular receptors. Over the years, it has been modified and improved to add new functionalities, and multiple engines have been developed. The most recent version is **AutoDock-GPU**, an accelerated version of AutoDock4 that is hundreds of times faster than the original single-CPU docking code.





General rule: *The grid volume should be large enough to at least allow the ligand to rotate freely.*

How does it work on the receptor?

1. Precalculation of atomic affinities using **AutoGrid**

3D Grid maps of non-covalent interaction energies are pre-calculated over the protein **for each atom type in the ligand**

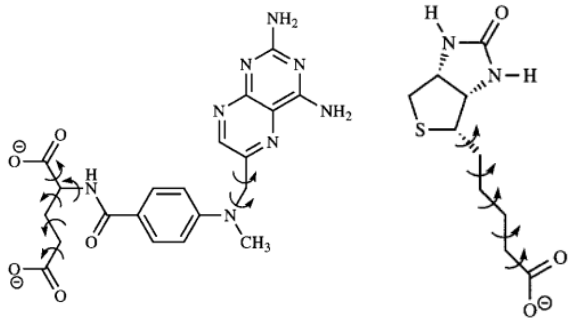
In addition, the **electrostatic potential** and **desolvation free energy grid maps** may also be calculated.

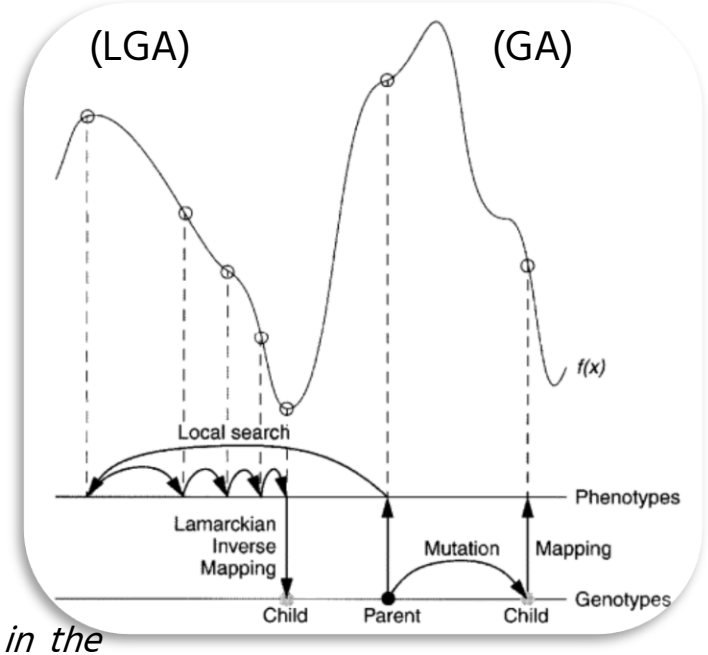
Grid maps are stored in plain text files with the extension '.map' and are required by AutoDock 4 to perform dockings.

```
autogrid4 -p input.gpf -l output.glg
```

Brief Introduction to AutoDock4

AutoDock4 is a free open source software for the computational docking of small molecules to macromolecular receptors. Over the years, it has been modified and improved to add new functionalities, and multiple engines have been developed. The most recent version is **AutoDock-GPU**, an accelerated version of AutoDock4 that is hundreds of times faster than the original single-CPU docking code.





General rule: *the more rotatable bonds in the ligand, the more difficult it will be to find a good binding modes.*

How does it work on the ligand?

2. Ligands are docked using the grid-maps information **AutoDock4**

Ligand conformational search methods

Local search method
Solis & Wets

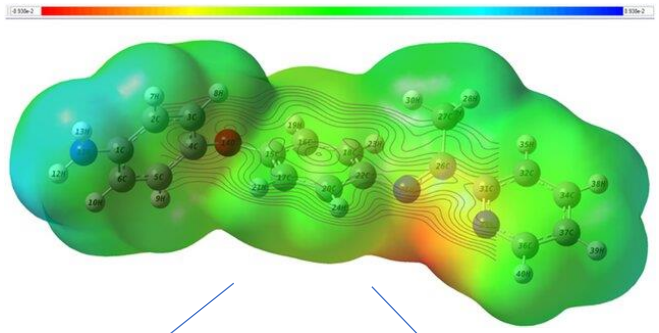
Global search methods
Genetic Algorithm (GA);

Hybrid search methods
Lamarckian Genetic Algorithm (LGA)

```
prepare_dpf4 -l ligand.pdbqt -r receptor.pdbqt
-o parameter_out.dpf
```

Introduction to AutoDock4

Ligand partial charges



Gasteiger-Marsili

molecule's electrostatic potential

summing over the atomic electron densities in the molecule

Kollman

AMBER force field and are based on the molecule's molecular orbital electron density

more physically accurate

How does it work on the ligand?

3. Atomic partial charge calculation in **AutoDock4**

Gasteiger and Kollman partial charges

The choice of charges will depend on the specific requirements of your study and the accuracy that you need for your predictions. You may consider trying both sets of charges and comparing the results to determine which set provides the best predictions for your aptamer-ligand system.

Forli S., et al. Nat. Protoc. 2016, 11:905-19

Brief Introduction to AutoDock4

Let's have a look to the Docking Parameter File (DPF)

```
ligand_types C N HD A OA fld
1r42_min_cap_allH.maps.fldmap
1r42_min_cap_allH.C.mapmap
1r42_min_cap_allH.N.mapmap
1r42_min_cap_allH.HD.mapmap
1r42_min_cap_allH.A.mapmap
1r42_min_cap_allH.OA.mapelecmap
1r42_min_cap_allH.e.mapdesolvmap
1r42_min_cap_allH.d.map
```

Parameters defining the grid maps to be used

```
move LAB.pdbqt
about 0.645 -0.957 [illegible]
```

Filename for the ligand to be docked

rotation center of the ligand

```
tran0 random
quat0 random
axisangle0 random
dihe0 random
```

Initial coordinates for the center of the ligand

```
rmstol 2.0
```

rms deviation tolerance for cluster analysis

```
ga_pop_size 150
ga_num_evals 25000000
ga_num_generations 27000
ga_elitism 1
ga_mutation_rate 0.02
ga_crossover_rate 0.8
ga_window_size 10
ga_cauchy_alpha 0.0
ga_cauchy_beta 1.0
set_ga
```

Parameters for Genetic Algorithm

```
sw_max_its 300
sw_max_succ 4
sw_max_fail 4
sw_rho 1.0
sw_lb_rho 0.01
ls_search_freq 0.06
set_psw1
```

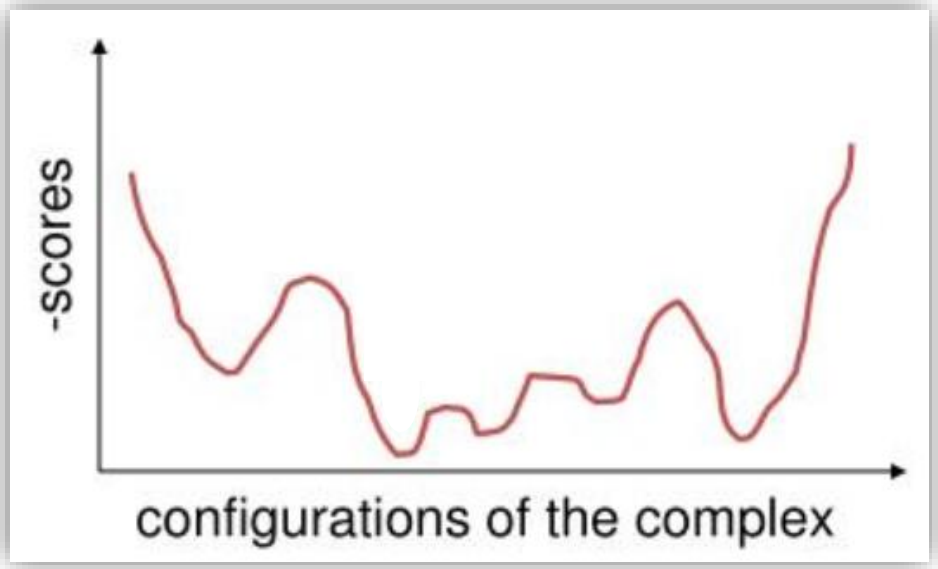
Parameters for local search (Solis & Wets)

```
ga_run 20
```

invokes Lamarckian Genetic Algorithm search engine

Brief Introduction to AutoDock4

AutoDock4 is a free open source software for the computational docking of small molecules to macromolecular receptors. Over the years, it has been modified and improved to add new functionalities, and multiple engines have been developed. The most recent version is **AutoDock-GPU**, an accelerated version of AutoDock4 that is hundreds of times faster than the original single-CPU docking code.



How does it work on the complex?

2. Ligands are docked using the grid-maps information **AutoDock4**

Empirical free energy force field

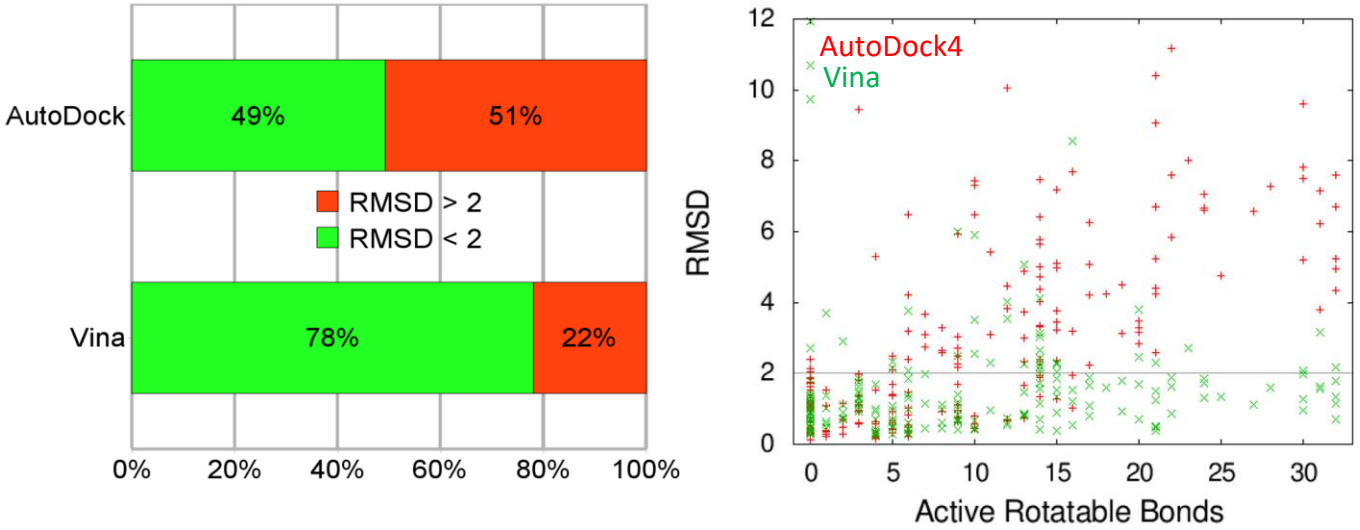
AutoDock4 adopts the physics-based force field scoring function with Van der Waals, electrostatic, and directional hydrogen-bond potentials derived from an early version of the **AMBER force field**

AutoDock4 *vs* AutoDock Vina

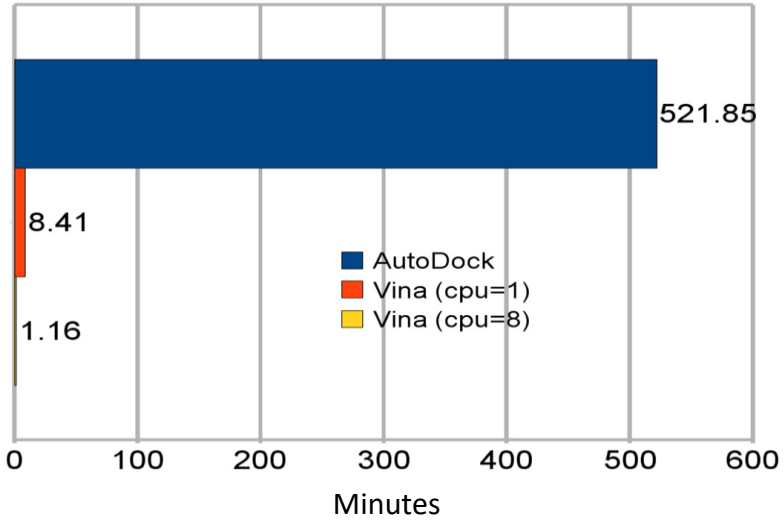
AutoDock Vina is considered to be the successor of AutoDock4.2 and comes with a new knowledge-based, statistical scoring function that replaces the empirical force field of AutoDock4. The advantages of Vina over AutoDock4.2 are its improved prediction accuracy and speed.

Furthermore, **AutoDock Vina** was designed to be compatible with the file format used for AutoDock 4

Accuracy



Speed



AutoDock4 *vs* AutoDock Vina

1. **Search algorithm (LGA)**

2. Semi-Empirical scoring function (based on the AMBER force field)
- electrostatic interactions, hydrogen bonds, desolvation energy, conformational entropy (too many torsions are problematic!)

3. Grid are calculated separately by running AutoGrid4

1. **Search algorithm (Monte Carlo+BFGS*)**

2. Hybrid scoring function (empirical+knowledge-based)
- Steric interaction (Gaussian, repulsion), hydrogen bond, hydrophobic, and torsion terms
- lacks in electrostatics and solvation terms

3. It calculates the grid charges internally

*Broyden–Fletcher–Goldfarb–Shanno

AutoDock Vina 1.2.0

AutoDock Vina

AutoDock 4

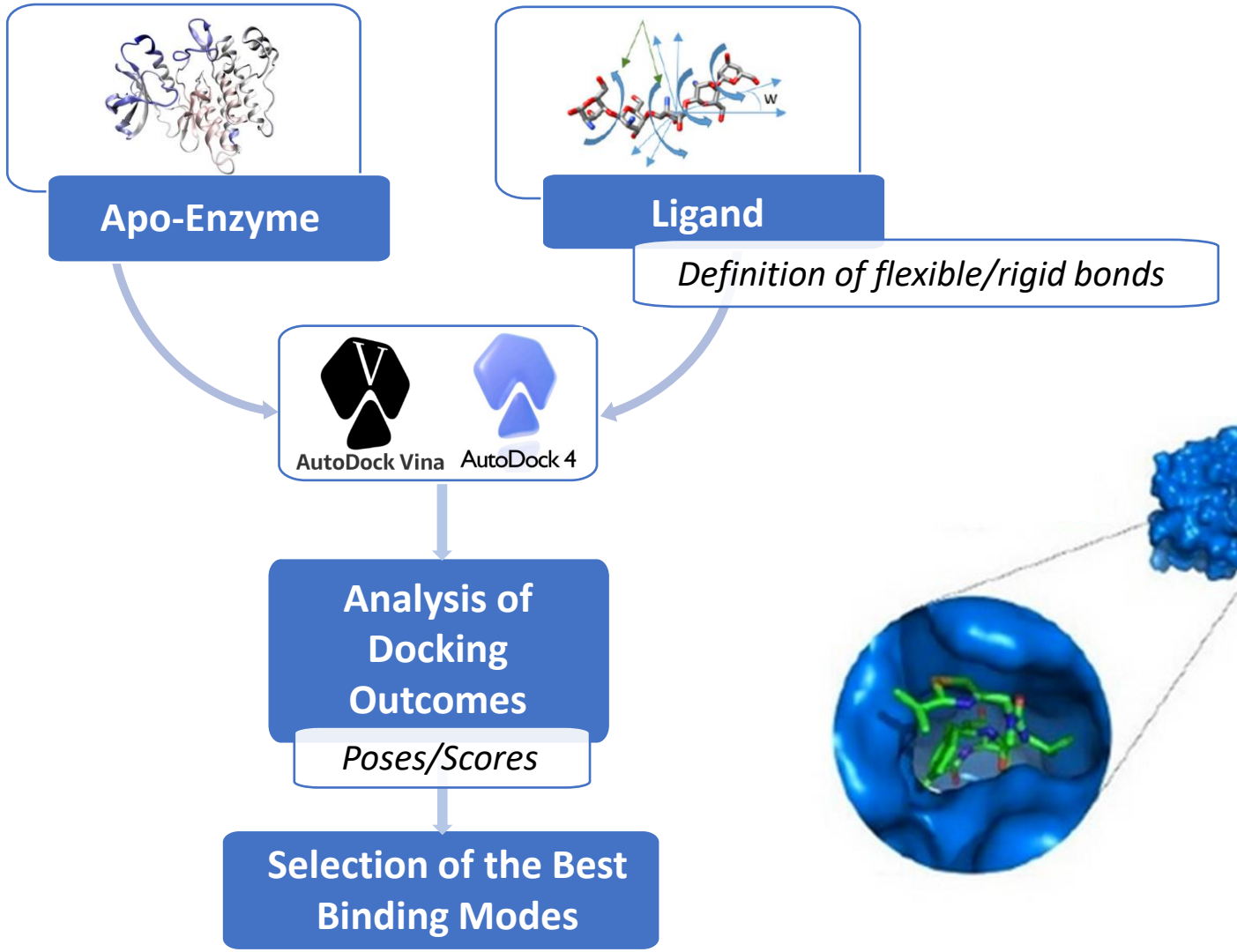
AutoDock Vina

unification of the features of the
AutoDock4 and AutoDock Vina programs

- new atom types
- optionally load external grid map files
- explicit water molecules
- Sampling of macrocycles

Eberhardt J, Santos-Martins D, Tillack AF, Forli S. *J Chem Inf Model.* 2021

Workflow of a docking study



Conclusions

- Determination of the binding mode and affinity between the constituent molecules in molecular recognition is crucial to understanding the interaction mechanisms and to designing therapeutic interventions
- The scoring function is a key element of a protein-ligand docking algorithm
- No perfect docking scoring function with optimal performance in every problem exists
- All current scoring functions have advantages and limitations
- One of the strategies commonly used to overcome these limitations is consensus scoring
- By using more than one docking program to predict the binding pose can lead to a significant improvement in the true positive ranking performance.

Thank you for your attention

Now it’s time to dock!