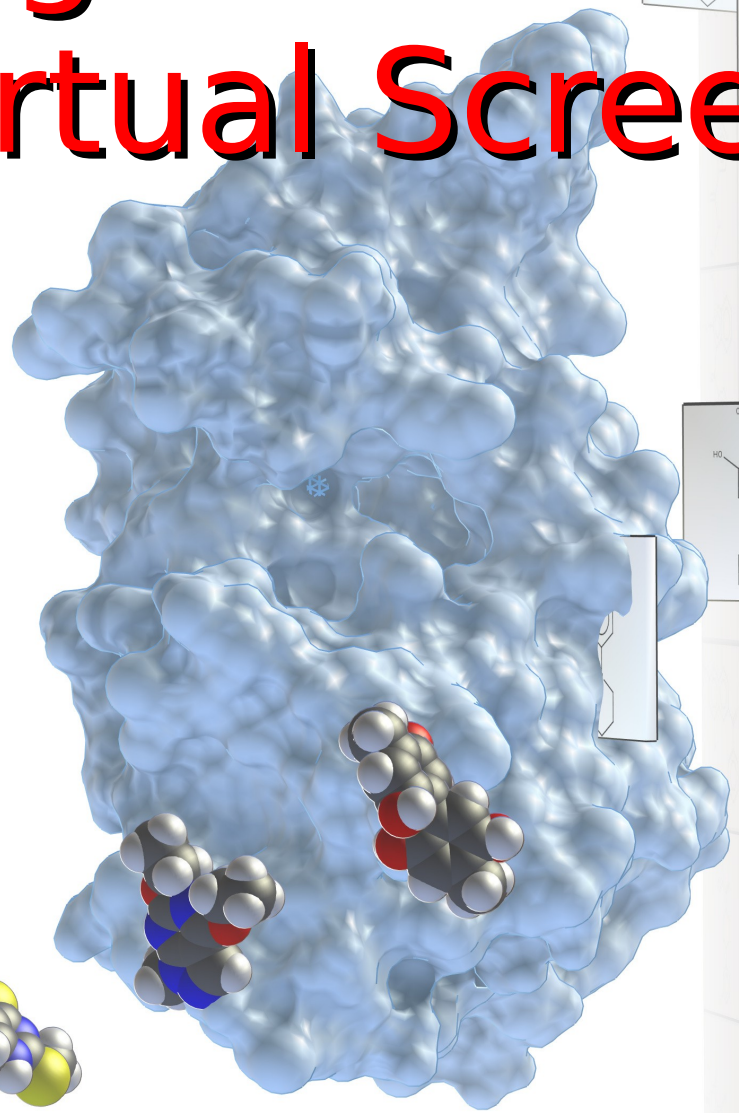
The Scripps Research Institute®

Using AutoDock for Virtual Screening



CUHK Croucher ASI Workshop 2011

Stefano Forli, PhD
Prof. Arthur J. Olson, Ph.D

Molecular Graphics Lab

AutoDock

Screening and Virtual Screening

The ultimate tool for identifying active compounds is the biological test:

High-Throughput Screening

Expensive (both money and time)

Can be automated but it still needs a lot of human intervention

Not all assays can be automated

Screening and Virtual Screening

Compounds can be pre-screened *in silico* enriching the ligand set

High-Throughput Screening

Cheap (saves both money and time)

Can be easily automated

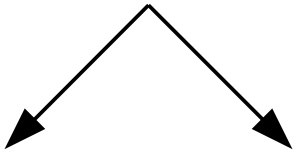
Dramatic reduction of the number of:
- compounds to be tested
- false negative

Virtual Screening

Definition

“Search for compounds with a defined biological activity using a computational model”

It's a *knowledge-based* method



Ligand based

(pharmacophores,QSAR...)

Structure based

(docking)

Horvat, D., “A virtual screening approach applied to the search for trypanothione reductase inhibitors”, **1997**, *Journal of Medicinal Chemistry* (40), 2412-2423

Virtual Screening

Advantages

Relatively cheap filter
(save both time and money)

Enrich ligand libraries

Exploit the increase of target structures
(structural genomics and crystallography)

Allow to test *in silico* the “druggability” of
new targets

Virtual Screening

Advantages

Relatively cheap filter
(save both time and money)

Enrich ligand libraries

Exploit the increase of target structures
(structural genomics and crystallography)

Allow to test *in silico* the “druggability” of new targets

Disadvantages

Often inaccurate

Scoring-function dependent

There is no method that's better than others

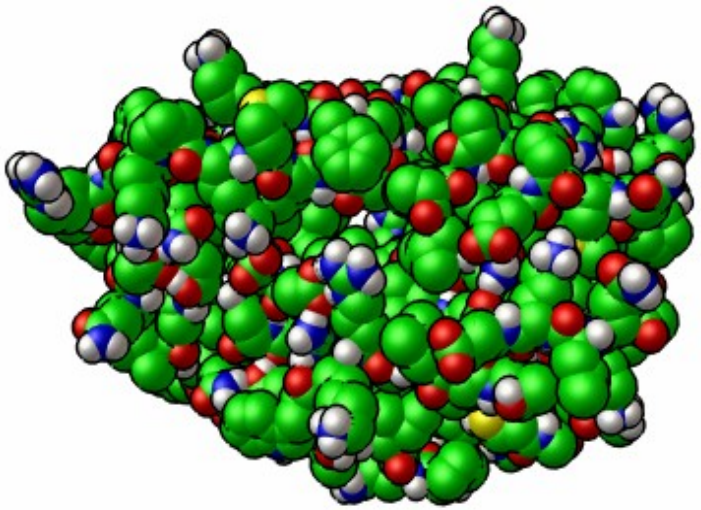
Strongly dependent on:
- target
- search method
- chemical space sampled

Always provides an answer
(McMaster competition 2005)

The Goal

Identify a molecule able to bind to a target providing a biological function

K_i / Energy



RECEPTOR

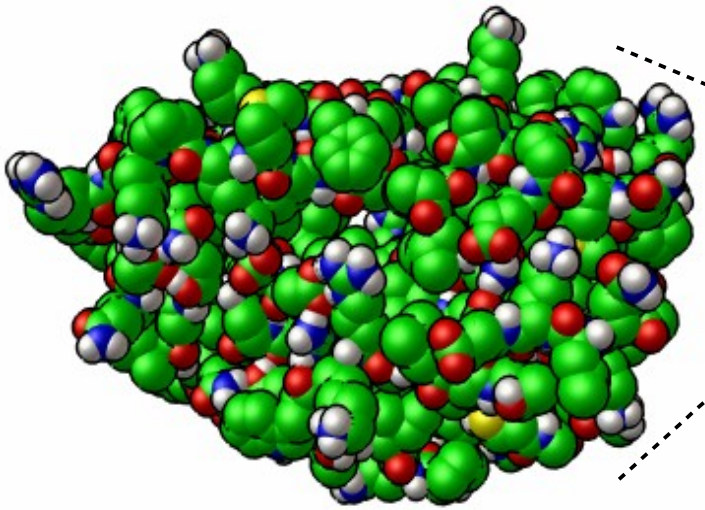
The Scripps Research Institute®

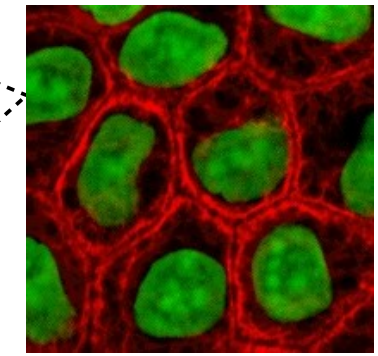
The Goal

Identify a molecule able to bind to a target providing a biological function

K_i / Energy

Unusual elements (Pt, Ru, U...)
Reactive chemical groups
Over/Under-functionalization
Partition coefficient (logP)





RECEPTOR

CELL

The Scripps Research Institute®

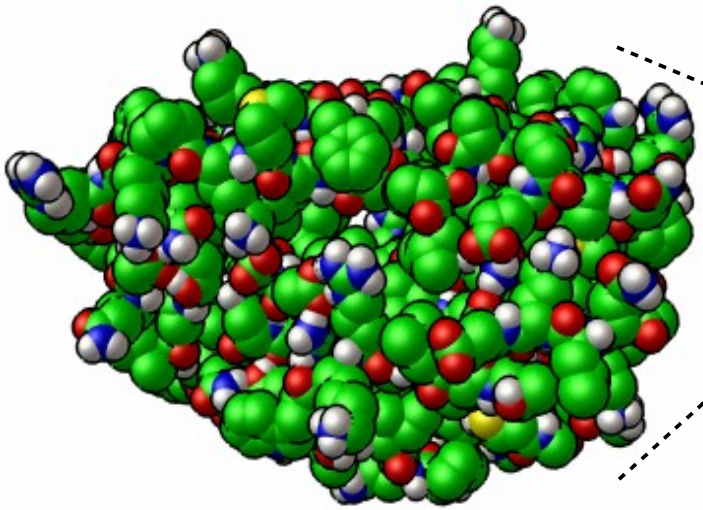
The Goal

Identify a molecule able to bind to a target providing a biological function

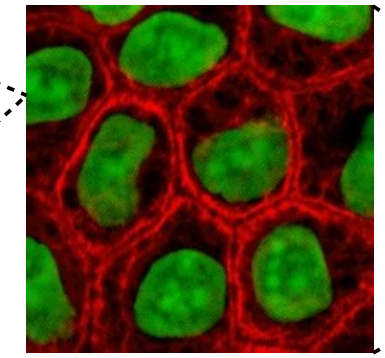
Absorption
Distribution + **Tox**
Metabolism
Excretion

K_i / Energy

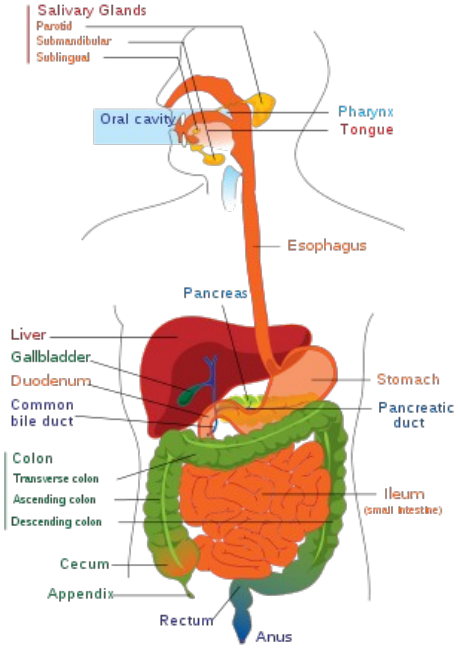
Unusual elements (Pt, Ru, U...)
Reactive chemical groups
Over/Under-functionalization
Partition coefficient (logP)



RECEPTOR



CELL

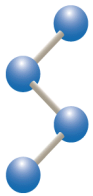


BODY

The Scripps Research Institute®

Virtual Screening
The Question

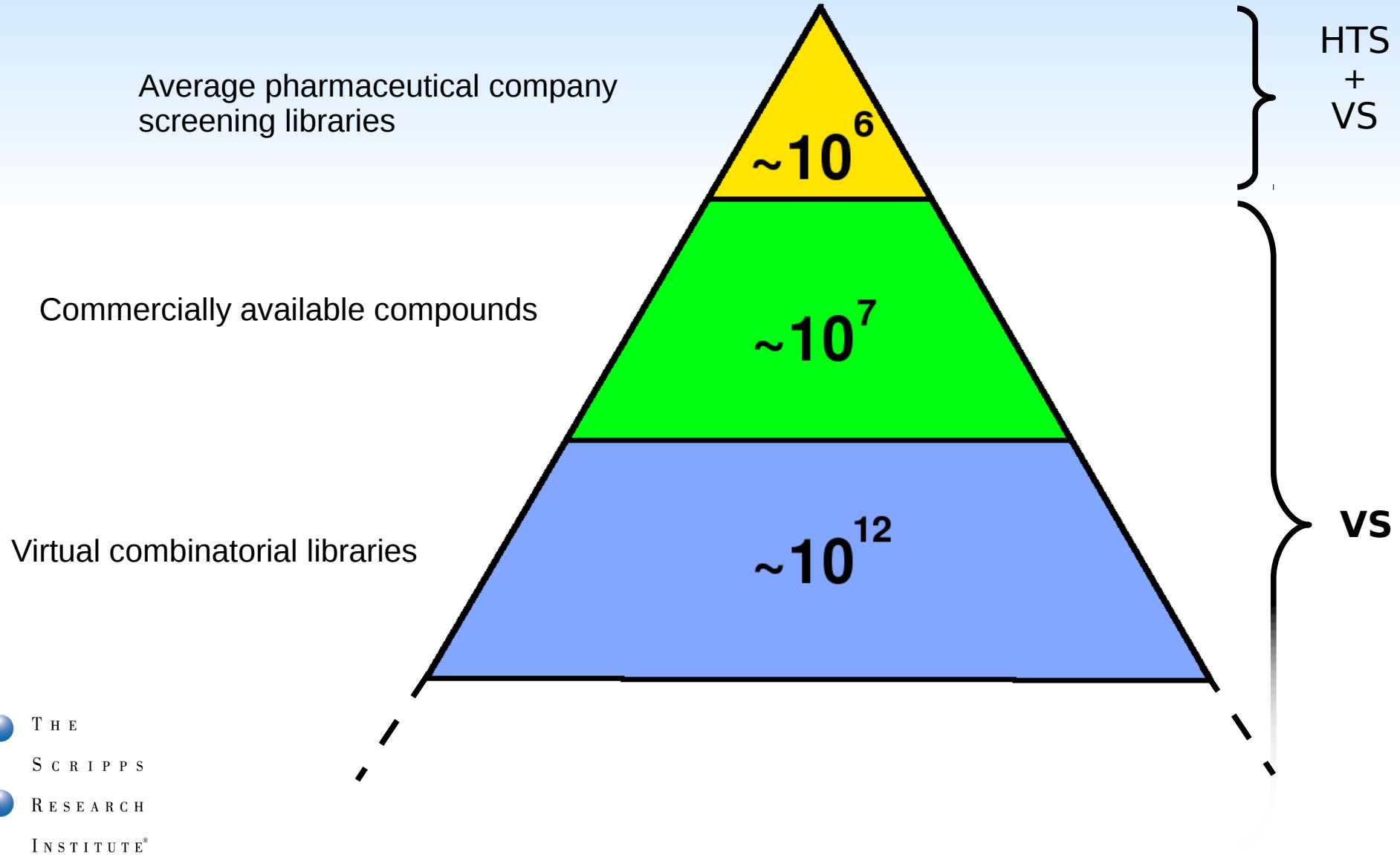
Which ligand binds in a target structure ?



The Scripps Research Institute®

Where to look for the answer

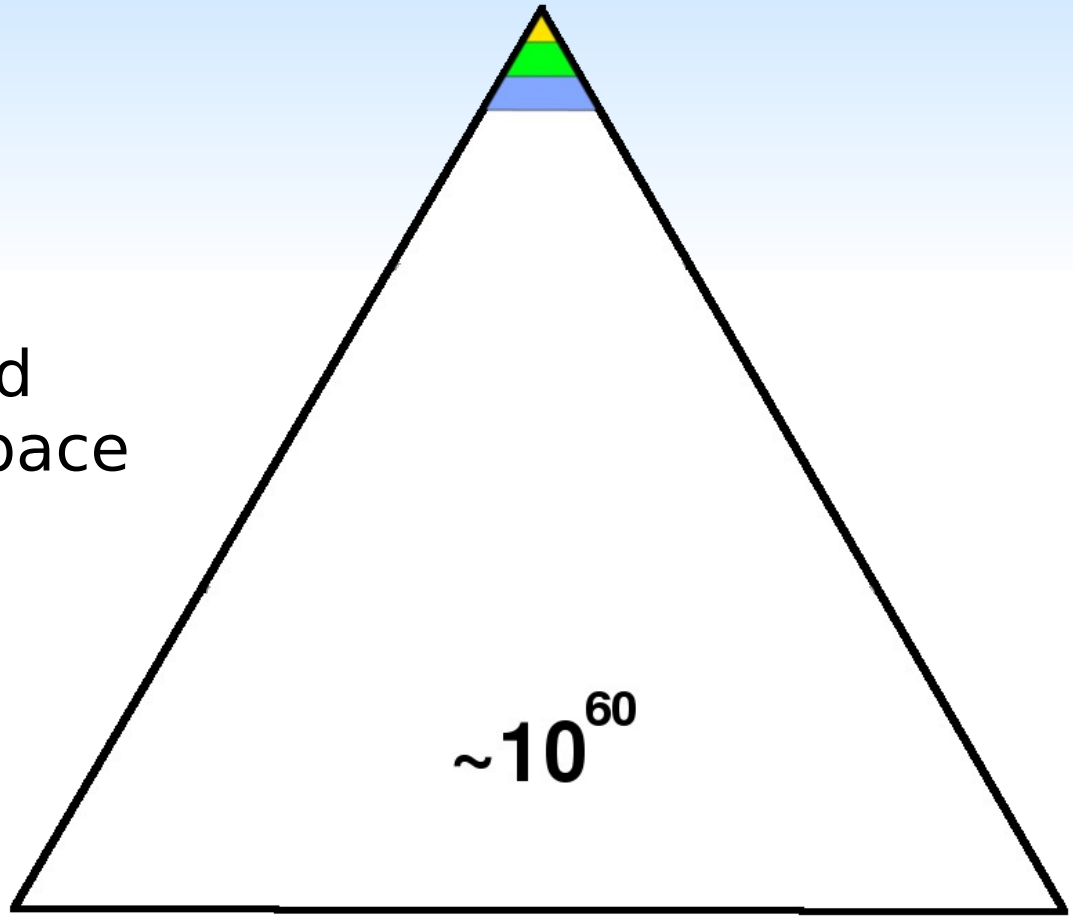
The Chemical Space



Where to look for the answer

The Chemical Space

Estimated chemical space

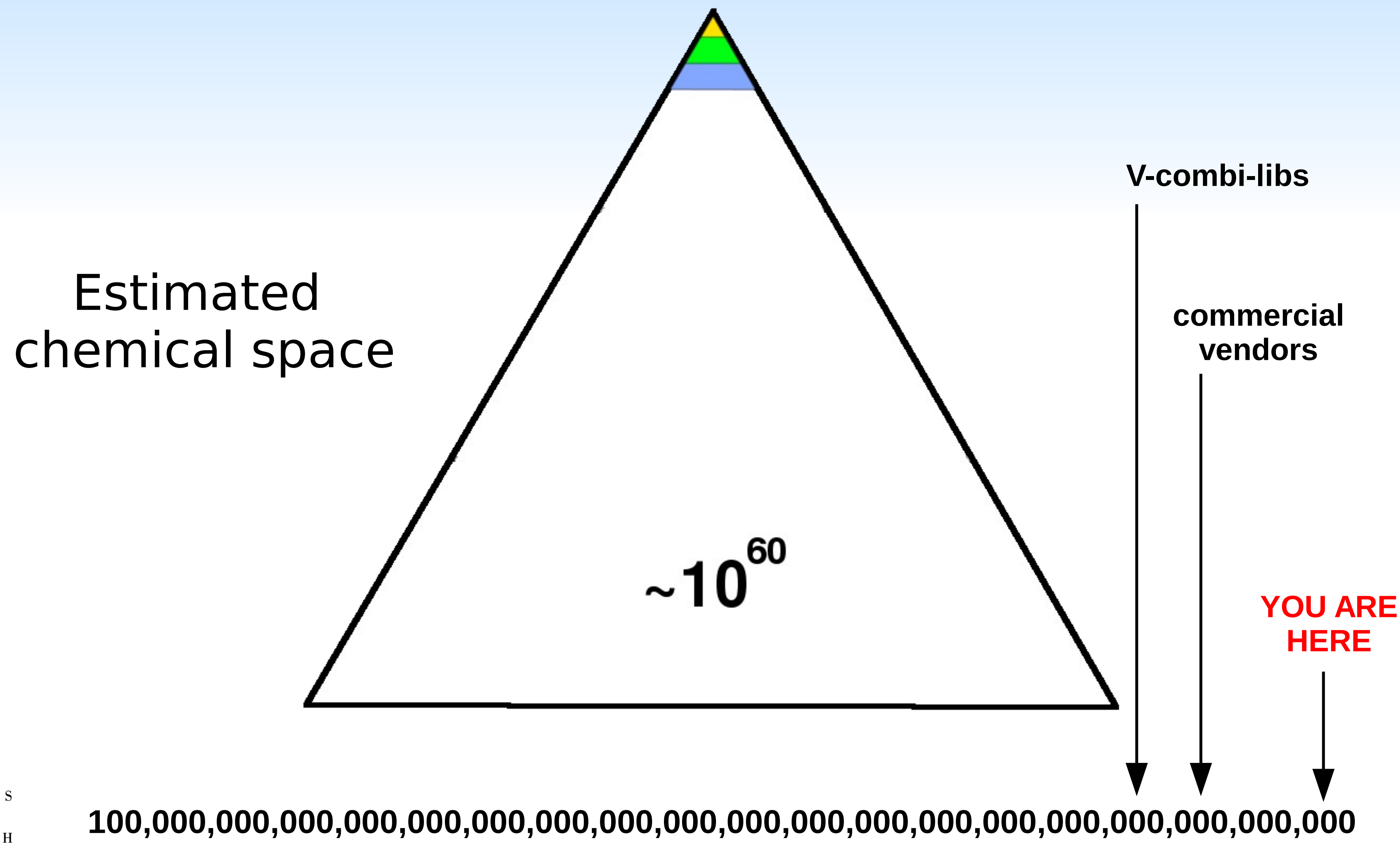


THE SCRIPPS RESEARCH INSTITUTE®

100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000

Where to look for the answer

The Chemical Space



The Scripps Research Institute

What the answer looks like

Hit low/medium target affinity

Lead sub-optimal target binding affinity

To be chosen for further development a **lead** compound should have the following properties:

- relatively simple chemical features (suitable for combinatorial/med-chem optimization, no/few chiral centers)

- well-established SAR series (similar compounds/chemical groups should present similar activity)

- good ADME properties

- [OPTIONAL] favorable patent situation

What the answer looks like

Filtering '*rules*'

Drugs 'Rule of Five' (Lipinski rule)

Hydrogen bond donors <= 5
Hydrogen bond acceptors <= 10
Molecular weight <= 500 dalton
Log$P_{w/o}$ < 5

← Approved drugs

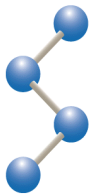
Hit Fragments 'Rule of Three'

Molecular weight <= 300 dalton
HB donor/Acceptors <= 3
C*log*P <= 300
Nrot <= 3

← ASTEX frag hits

HTS efforts by using Lipinski-filtered libraries led to few micro-molar hits

LEADS are not DRUGS



THE SCRIPPS RESEARCH INSTITUTE®

"Rules" are good in principle, but they require to sample a huge chemical space to give really effective molecules.

What the answer looks like

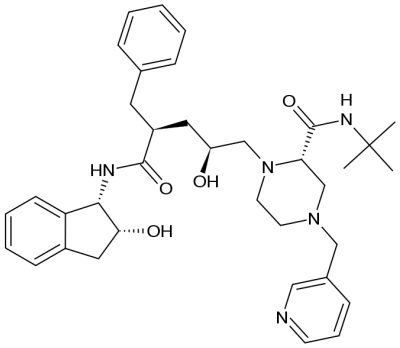
Rules 'Exceptions'

The nature and location of the target must be take into account for properties profile:

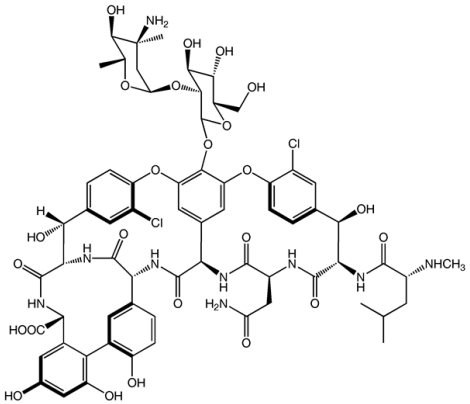
CNS molecule
(lipophilic blood-brain-barrier)

gastro-intestinal antibiotic
(highly soluble)

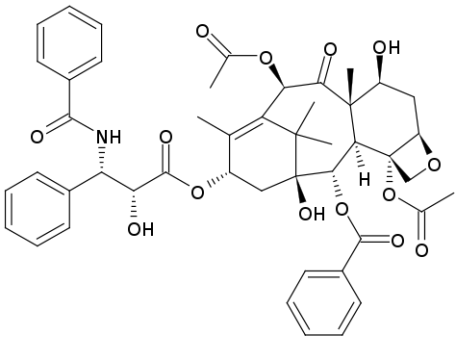
Natural compounds & Pro-drugs and “last resort” compounds



Indinavir



Vancomycin

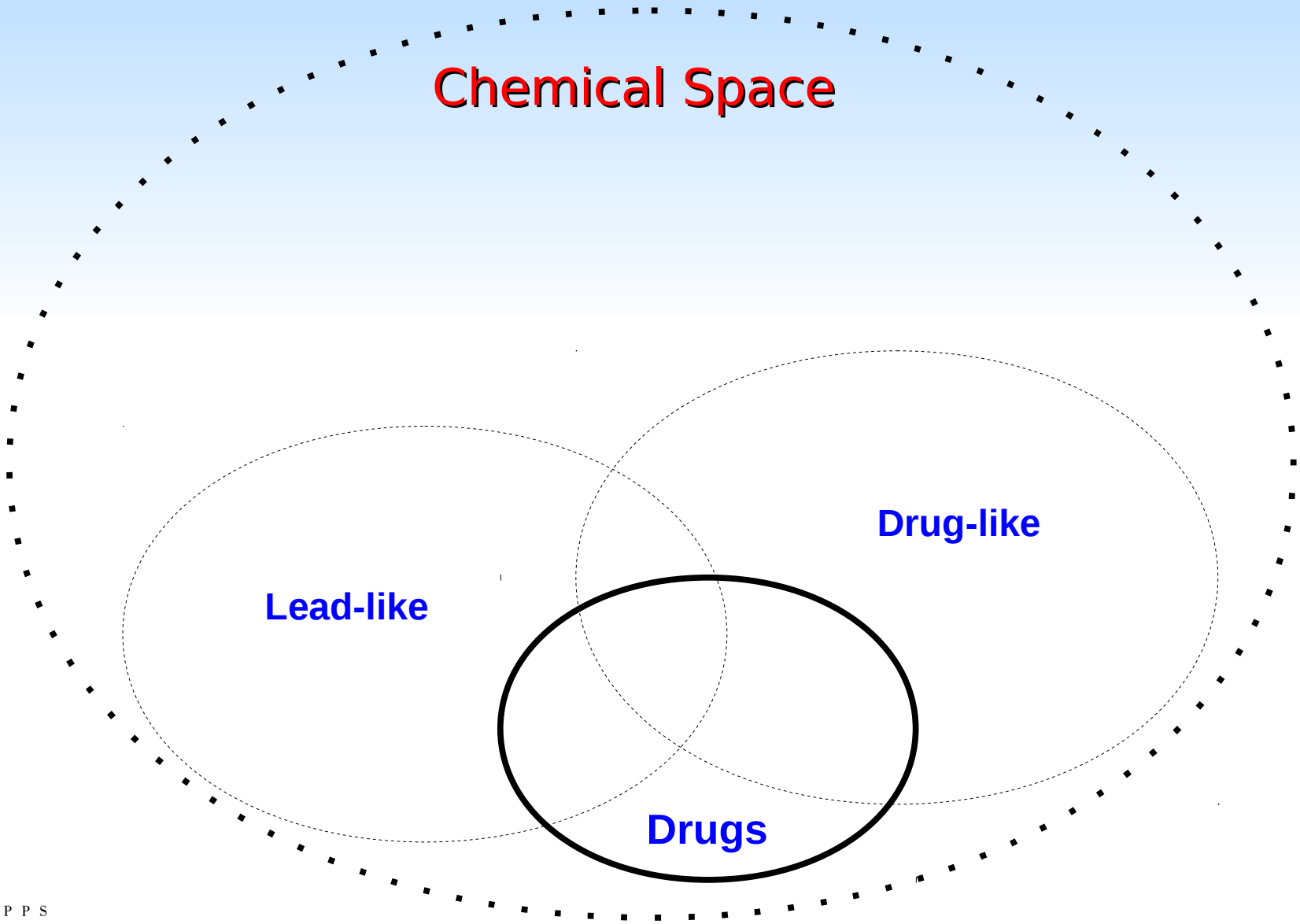


Paclitaxel

The Scripps Research Institute®

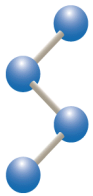
<u>The ultimate tool for identifying hits is the High-throughput screening</u>

What the answer looks like



Virtual Screening
The Question

Which LIGANDS are *likely* to bind in a target structure ?



The Scripps Research Institute®

Where the answer should be found

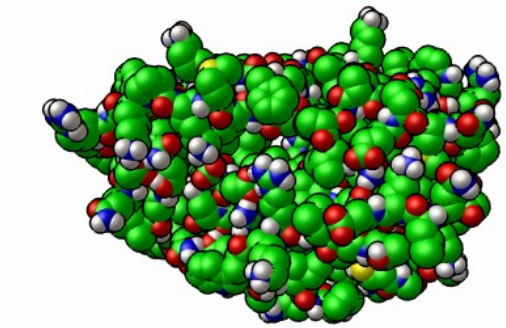
Target state(s)

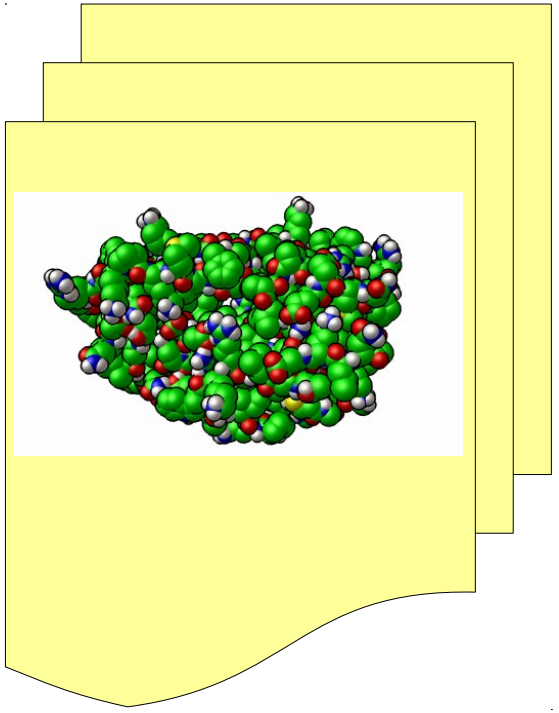
Hitting a moving target:

- *functional states (active-inactive)*

- *dynamic states (temperature)*

- *protonation/complexation states*





The Scripps Research Institute®

Virtual Screening

The Question

Which ligands are *likely* to bind the most probable state(s) of my target structure

Virtual Screening Hints

Prepare target and select ligand libraries with care

- Filter unusual elements
- Reliable 3D geometries
- Protonation states/tautomers

Reduce the space of your search

- diversity sets
- generic filtering
- target specific filtering (lipophilic VERSUS hydrophobic binding sites)

Use all available information to select results

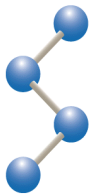
- mutagenesis, SAR...

Try to sample different conformations of the protein

- reduce false negative

Use reference compounds whenever available

- Useful for comparing results with ligands with known activity



Available ligand libraries

PubChem

http://pubchem.ncbi.nlm.nih.gov/

All biological data related to a compound

2D structures

ZINC

http://zinc.docking.org/

A free database of commercially available compounds for virtual screening

109 commercial compound suppliers, 30×10^6 compounds (non-unique)

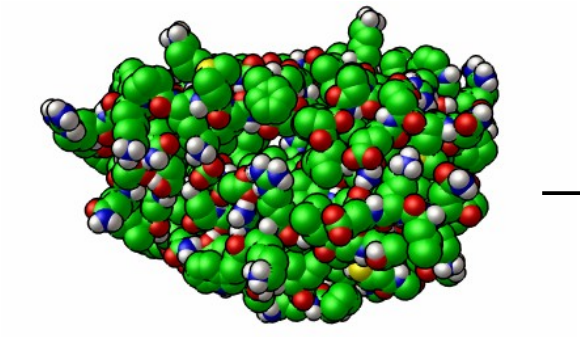
The Scripps Research Institute®

Irwin and Shoichet (2005) *J. Chem. Inf. Model.* **45**(1), 177-82

AutoDock Single Docking

Input
(PDB, Mol2)

PDBQT
(partial charges,torsions...)



PDBQT
(protonation states, waters, cofactors...)

Maps

GPF

Ligand
Ligand
Ligand

DPF

DOCKING

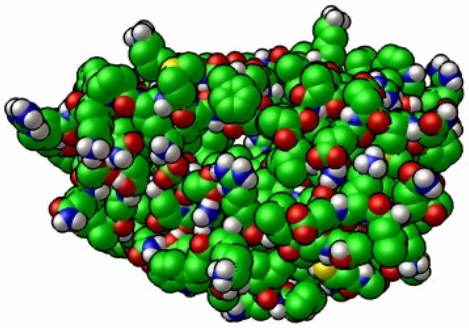
The Scripps Research Institute®

AutoDock VS

Input
(PDB, Mol2)

FILTER

PDBQT
(partial charges,torsions...)



PDBQT
(protonation states,
waters, cofactors...)

Maps

GPF

Ligand
Ligand
Ligand

DPF

MULTIPLE DOCKINGS

The Scripps Research Institute®

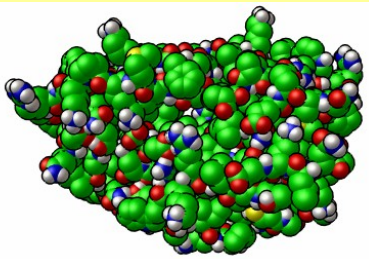
AutoDock VS

Input
(PDB, Mol2)

FILTER

PDBQT
(partial charges,torsions...)

R.C.



PDBQT
(protonation states,
waters, cofactors...)

Maps

GPF

Ligand
Ligand
Ligand

DPF

MULTIPLE DOCKINGS

The Scripps Research Institute®

How spot a good answer?

Ligand properties used for results analysis

Energy: AutoDock score

Ligand Efficiency : $$\Delta g = \frac{-RT \log K_d}{N_{\text{non-hydrogen atoms}}}$$

Cluster analysis:
- multiple poses clustering tolerance
- cluster size
- energy range

Knowledge-base analysys
- chemical similarities with known binders
- mutagenesis data
- structure/sequence homology

THE SCRIPPS RESEARCH INSTITUTE®

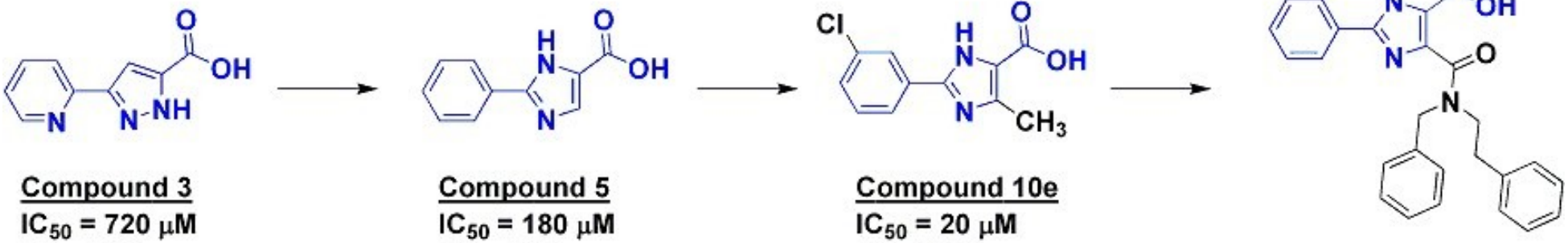
REDUCE THE NUMBER OF RESULTS TO ANALYZE AND (HOPEFULLY) ENRICH THE QUALITY

How spot a good answer?

Ligand properties used for results analysis

OPTIMIZATION

ligand efficiency



Compound 3
IC_{50} = 720 μM
LE = 0.31

Compound 5
IC_{50} = 180 μM
LE = 0.36

Compound 10e
IC_{50} = 20 μM
LE = 0.40

Compound 20
IC_{50} = 0.83 μM
LE = 0.25
GI_{50} = 13 μM

The Scripps Research Institute®

Potter AJ, et al. Bioorg Med Chem Lett. 2010 Jan 15;20(2):586-90

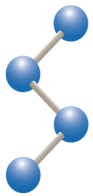
How spot a good answer?

Results clustering

```
Number of distinct conformational clusters found = 2,  out of 100 runs,
Using an rmsd-tolerance of 2.0 A

        CLUSTERING HISTOGRAM
        ____________________

_____________________________________________________________________________
     |             |     |             |     |
Clus | Lowest      | Run | Mean        | Num | Histogram
-ter | Binding     |     | Binding     | in  |
Rank | Energy      |     | Energy      | Clus|    5    10   15   20   25   30   35
_____|_____________|_____|_____________|_____|____:____|____:____|____:____|____:___
   1 |      -7.52  |  14 |      -7.52  |   7 |#######
   2 |      -7.39  |  60 |      -7.39  |  93 |##############################################...
_____|_____________|_____|_____________|_____|_____________________________________________

Number of multi-member conformational clusters found = 2, out of 100 runs.
```

BEST ENERGY?

MOST POPULATED CLUSTER?



The Scripps Research Institute®

AutoDock VS

Input
(PDB, Mol2)

FILTER

PDBQT
(partial charges,torsions...)

R.C.

PDBQT
(protonation states,
waters, cofactors...)

AutoDock VS
RACCOON

+

Fox

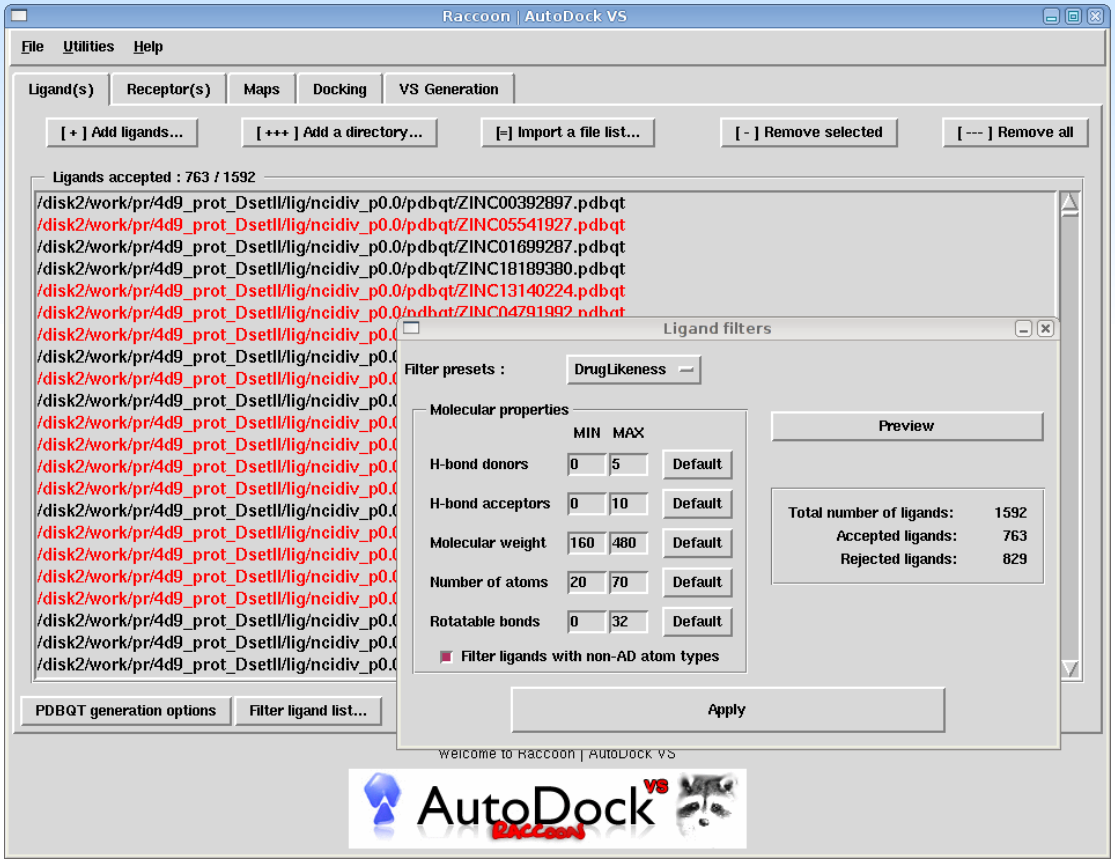
GPF

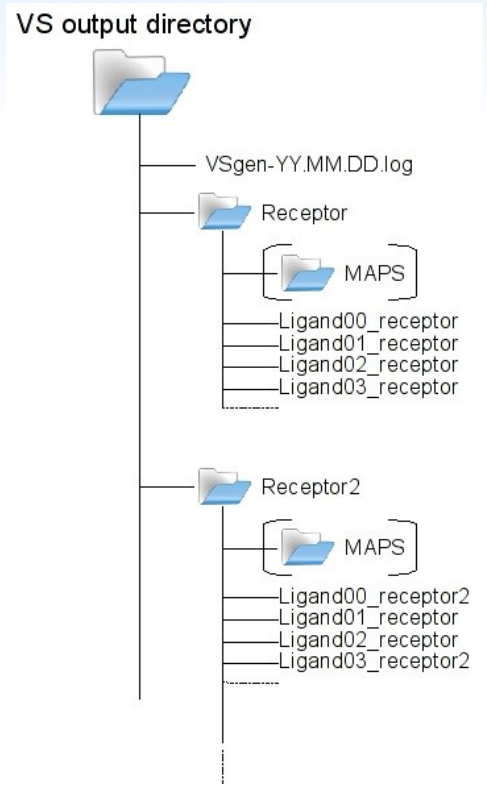
DPF

Ligand
Ligand
Ligand

AutoDockVS | Raccoon

- input preparation and filtering
- file-system organization
- parameter files generation
- automated calculation scripts
- generation data logging

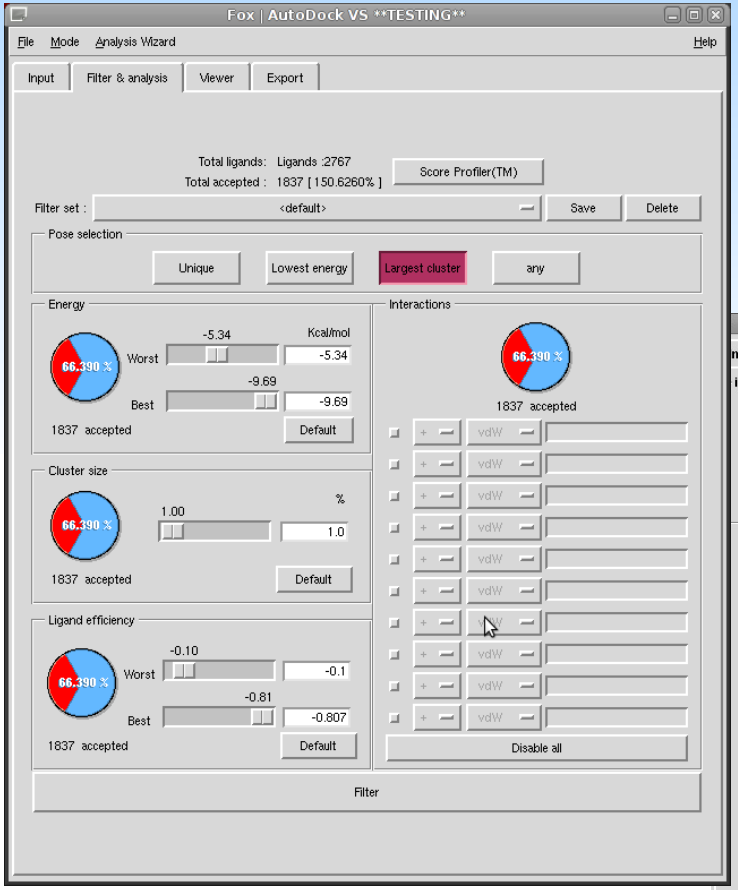


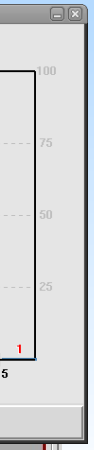


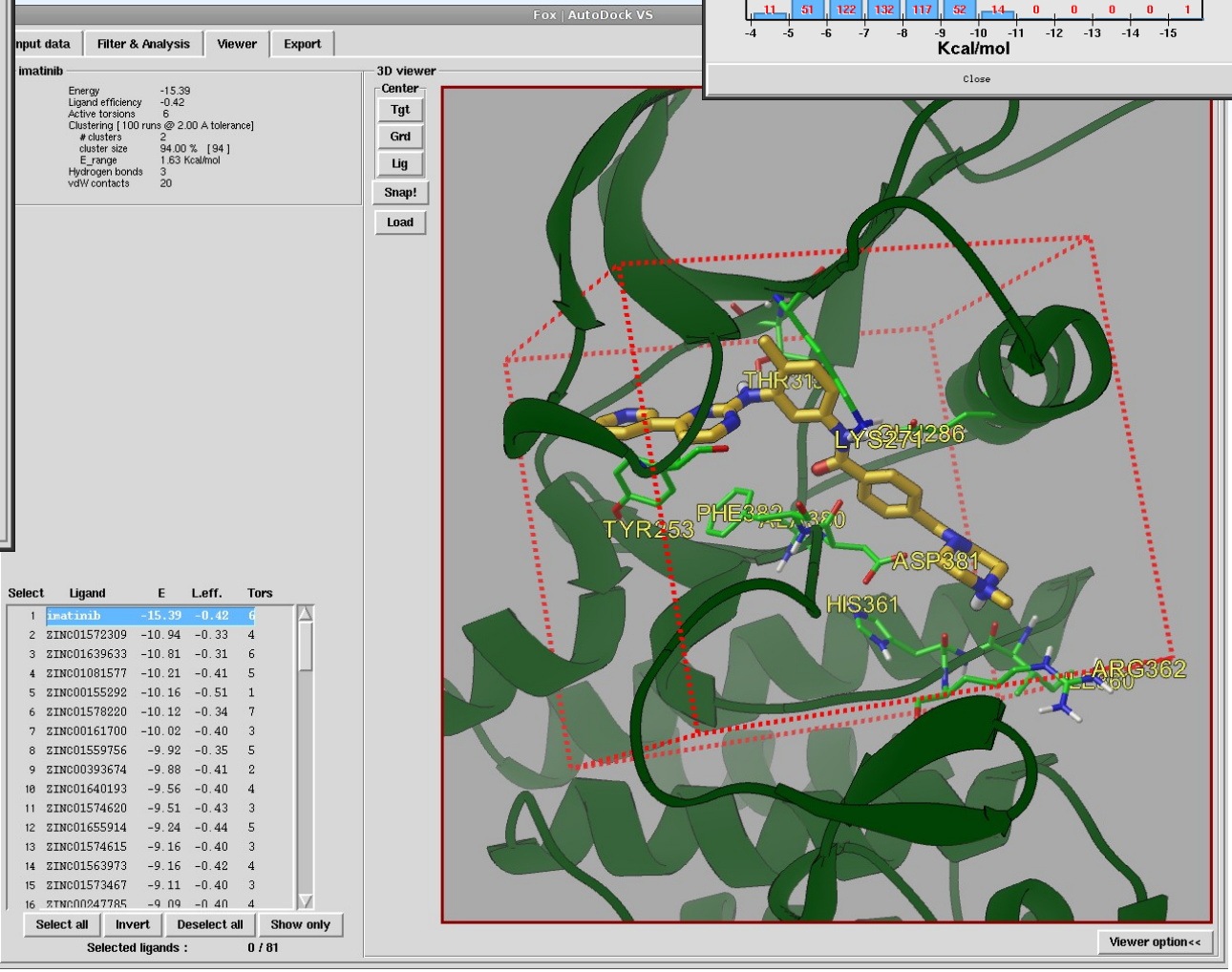
THE SCRIPPS RESEARCH INSTITUTE®

AutoDock VS RACCOON

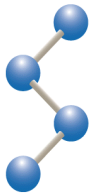
AutoDockVS | Fox







- clustering
- results analysis
- report tools



THE SCRIPPS RESEARCH INSTITUTE®

How to obtain good answers

Virtual Screening Hints

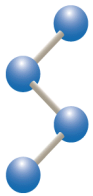
Pre-processing

- Choose with care which ligands to include in the screening
- Select representative target state(s)

Post-processing

- Efficiently filter results:
- *avoid chemical complexity*
- *search for specific interactions (polar residues)*
- *use both energy score and ligand efficiency*

- Use knowledge-driven criteria
- *use target information (function, mutagenesis)*
- *use known binders references (if available)*



Recommended readings

“Is there a difference between leads and drugs? A historical perspective”
Oprea, T., I., Davis, A., M., Teague, S .,J., Leeson, P., D.J.Chem. Inf. Comput.Sci. 2001, 41, 1308-1315

“A 'rule of three' for fragment based lead discovery?”
Congreve, M., Carr, R., Murray, C., Jhoti, H. 2001, Drug Discov. Today, 2003, v8, n19, p876

“Virtual screening - what does it give us?”
Köppen H. Curr Opin Drug Discov Devel. 2009 May;12(3):397-407

”Ligand efficiency: a useful metric for lead selection”
Hopkins AL, Groom CR, Alex A. Drug Discov Today. 2004 May 15;9(10):430-1.

"Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings"
C.A. Lipinski; F. Lombardo; B.W. Dominy and P.J. Feeney (1997). . Adv Drug Del Rev 23: 3–25