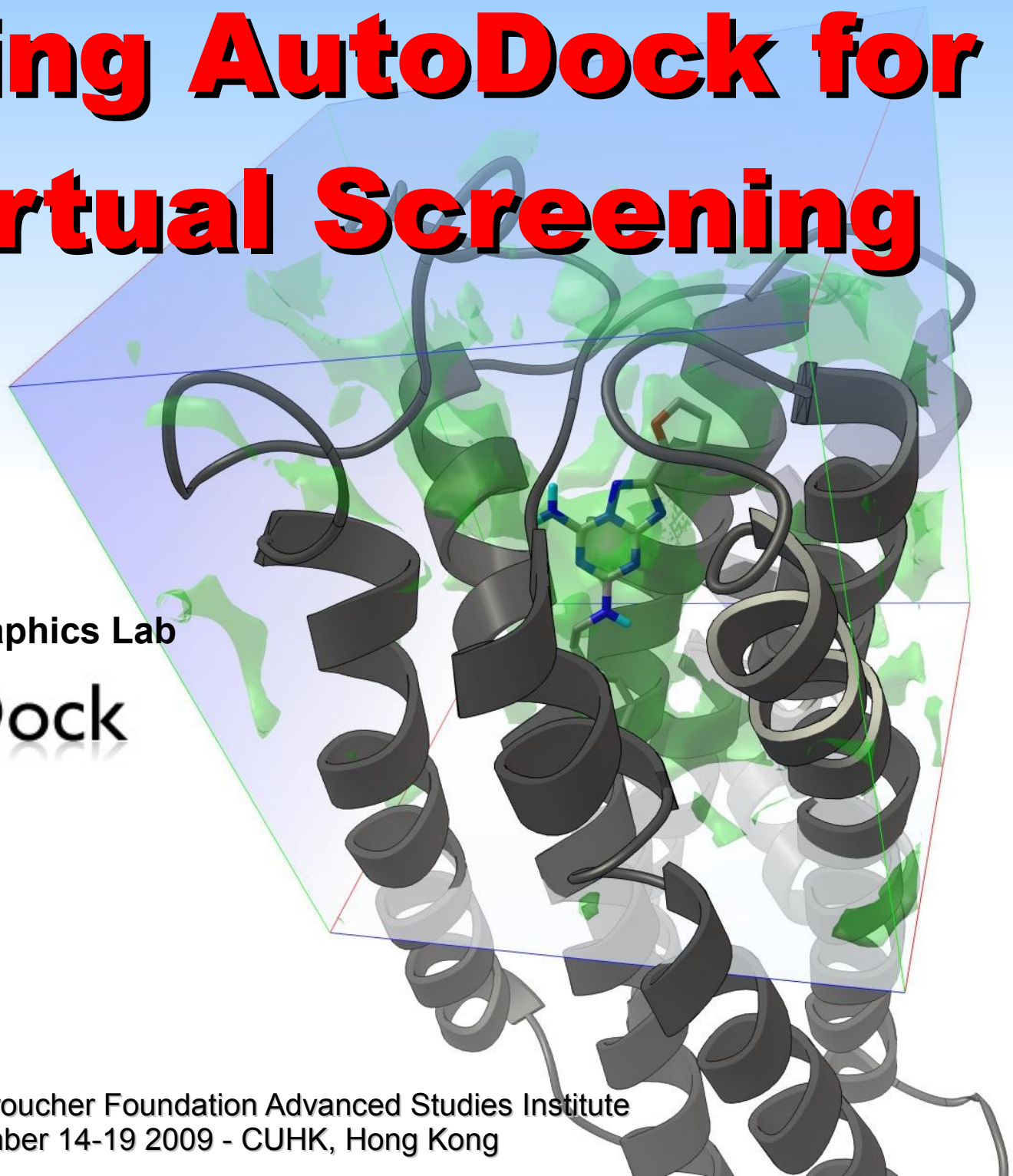
Using AutoDock for Virtual Screening



Stefano Forli
Ruth Huey
Garrett M. Morris
William Lindstrom

The Molecular Graphics Lab

AutoDock

The Scripps Research Institute®

The Croucher Foundation Advanced Studies Institute
December 14-19 2009 - CUHK, Hong Kong

Virtual Screening Session

Outline

Introduction to Virtual Screening

- Definition
- The goal of a vs
- Preparing the inputs

Hands-on Tutorials

- Raccoon

Introduction to Virtual Screening

The Scripps Research Institute®

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

Virtual

High-Throughput Screening

- Cheap (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)

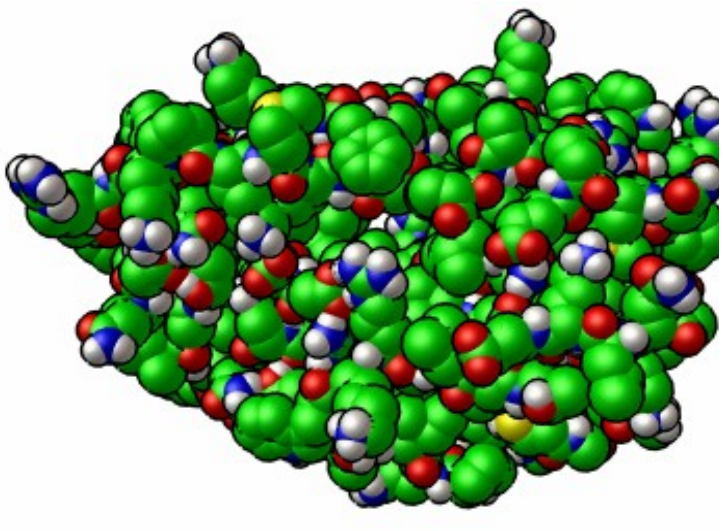
THE SCRIPPS RESEARCH INSTITUTE®

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

The Goal

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

K_i / Energy



RECEPTOR

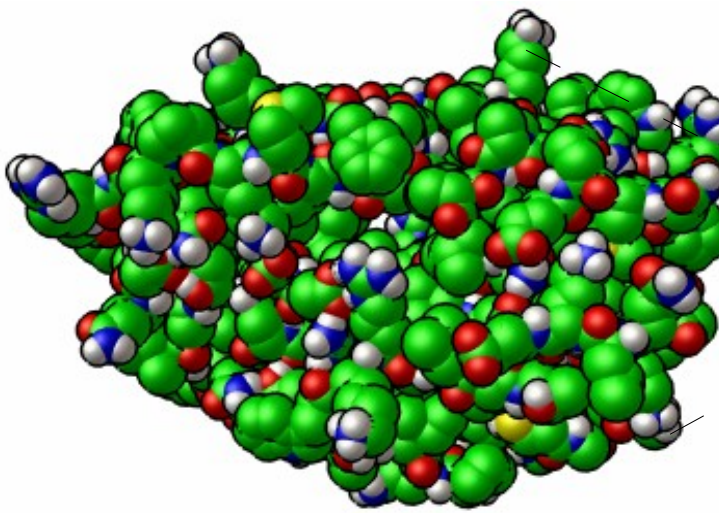
The Scripps Research Institute®

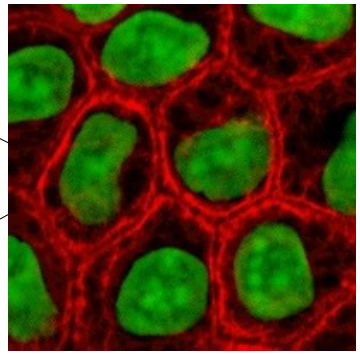
The Goal

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-functionalized cpds
Partition coefficient





RECEPTOR

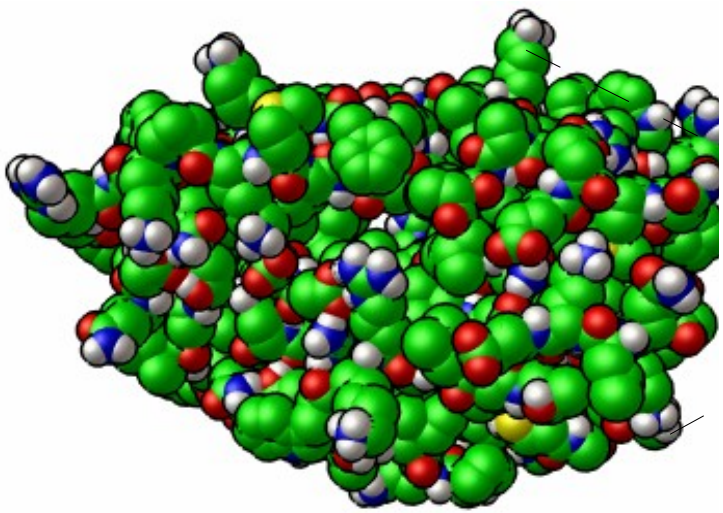
CELL

The Scripps Research Institute®

The Goal

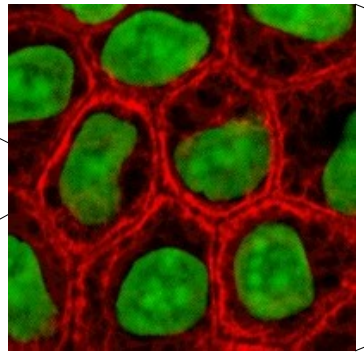
Identify a molecule able to bind to a target providing a biological function

K_i / Energy



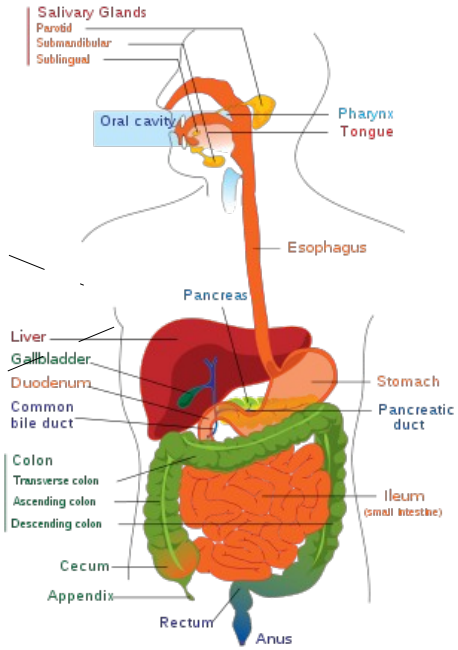
RECEPTOR

Unusual elements (Pt, Ru, U…)
Reactive chemical groups
Over/Under-functionalized cpds
Partition coefficient



CELL

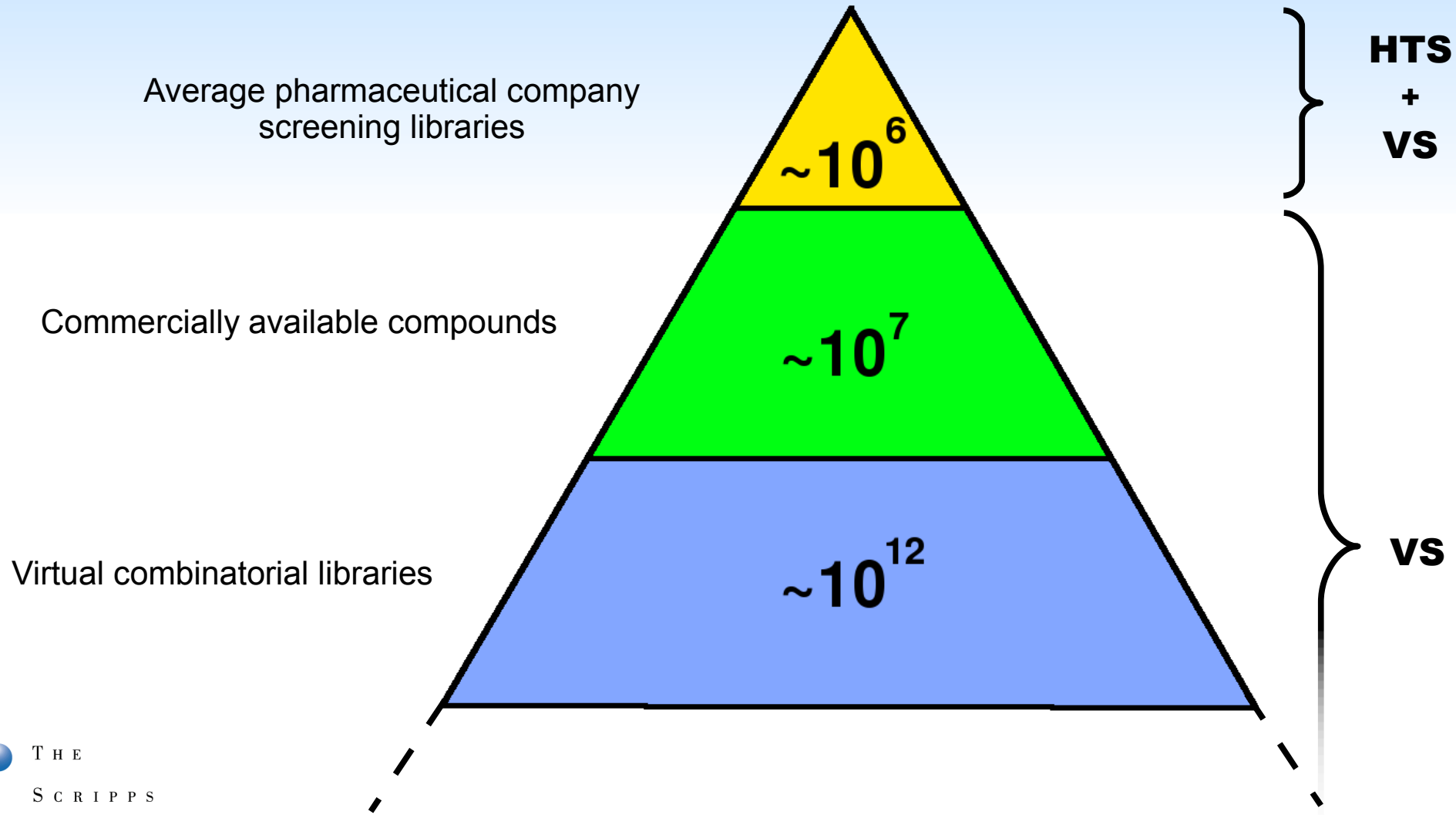
Absorption
Distribution + **Tox**
Metabolism
Excretion



BODY

The Scripps Research Institute®

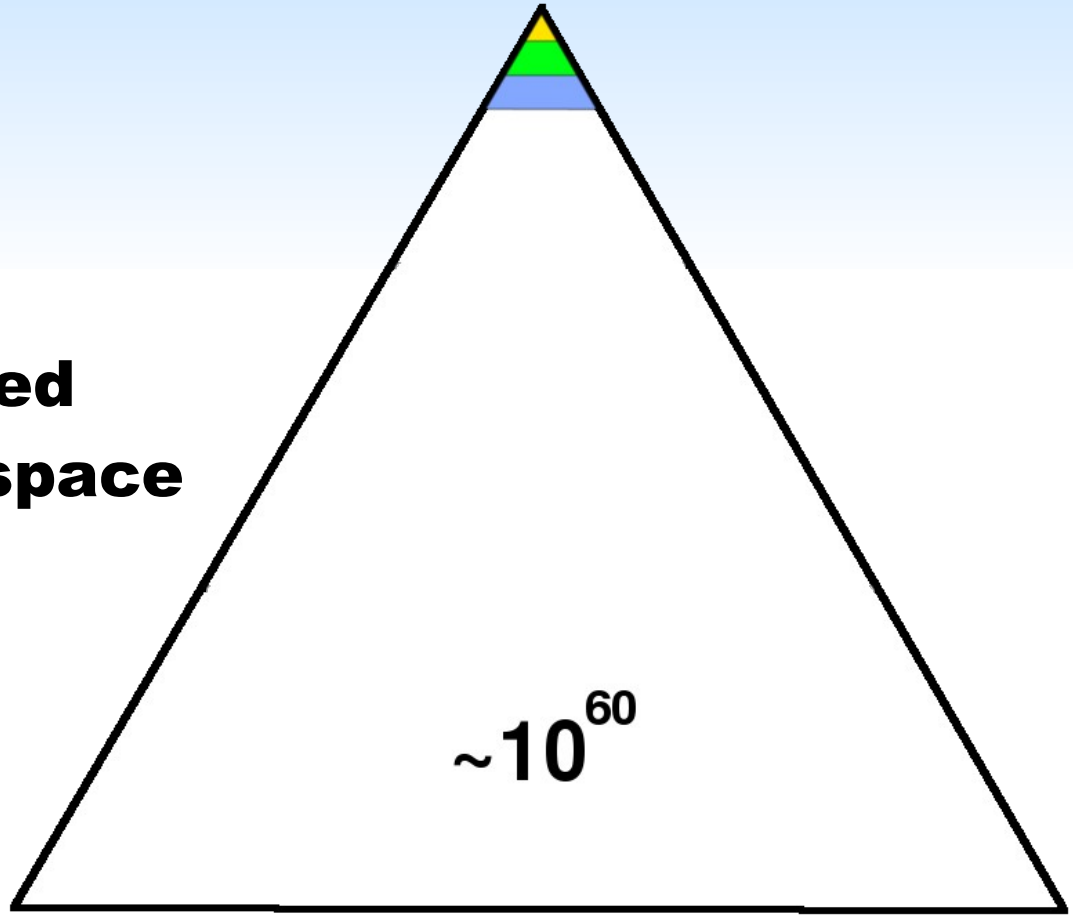
Where to look for: The Chemical Space



The Scripps Research Institute®

Where to look for: The Chemical Space

Estimated chemical space



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

The Scripps Research Institute®

What we are looking for

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)
- belong to a well-established SAR series (similar compounds and or similar chemical groups should present similar activity)
- good ADME properties
- favorable patent situation

What we are looking for

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 micromolar 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 we are looking for

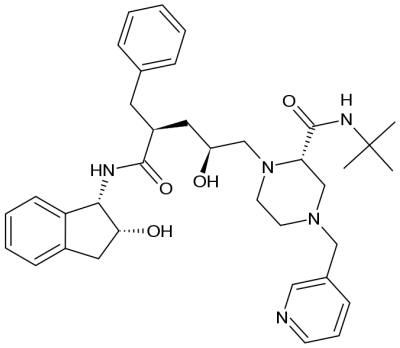
Rules 'Exceptions'

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

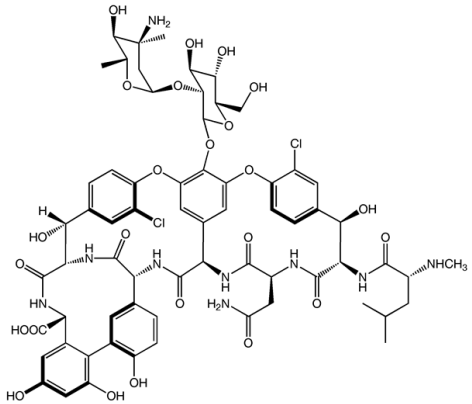
CNS molecule
(lipofilic blood-brain-barrier)

gastro-intestinal antibiotic
(highly soluble)

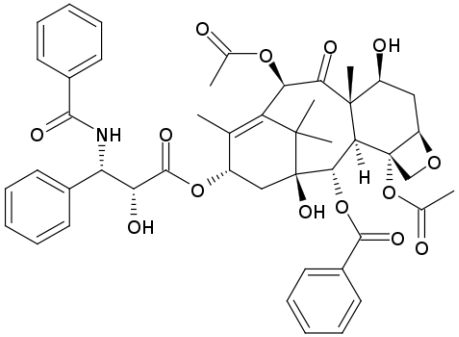
Natural compounds & Pre-drugs and “last resort” compounds (i.e. nitrogen mustards, cisplatin...)



Indinavir



Vancomycin

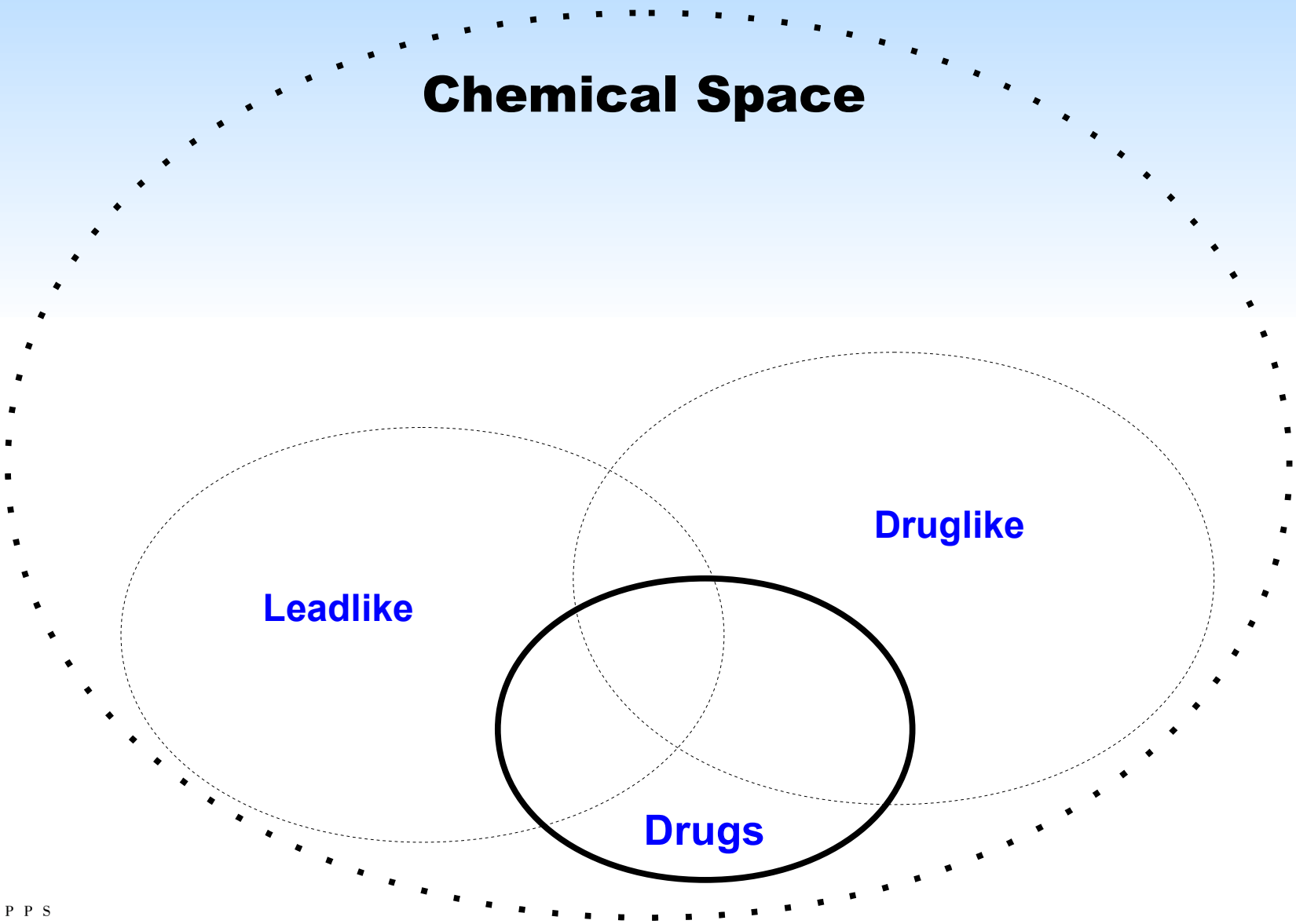


Paclitaxel

The Scripps Research Institute®

The ultimate tool for identifying hits is the High-throughput screening

What we are looking for



The Scripps Research Institute®

Virtual Screening

Advantages

- Relatively cheap filter
  (save both time and money)
- Enrichment factor of libraries
- Given the rise of structural genomics
  and crystallography
- It allows to test *in silico* the “druggability”
  of new targets

Virtual Screening

Advantages

- Relatively cheap filter
  (save both time and money)
- Enrichment factor of libraries
- Given the rise of structural genomics and crystallography
- It allows 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)

Virtual Screening Hints

- **Prepare both target and ligand libraries with care**
  Filter unusual atom types (AutoDock atom types)
  Reliable 3D geometries
  Protonation states and tautomers

- **Reduce the space of your search**
  diversity sets
  generic filtering
  target specific filtering (lipofilic binding site v.s. hydrophobic binding site)

- **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

Target preparation

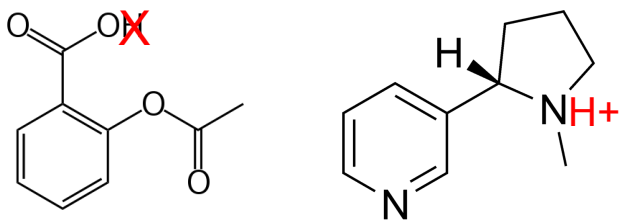
- **Inspect the protein structure**
  Check the quality of the target structure, missing residues...
- **Protonation states**
  Inspect the protonation states
- **Check side chains alternate conformations**
  Select the most suitable conformation when alternate choices are available
  Take into account the use of flexible residues, if necessary
- **Waters and cofactors**
  Some waters (or co-factors) can be essential for the binding... or not...

Compound libraries preparation

Pre-processing of ligand libraries can affect dramatically the quality of results

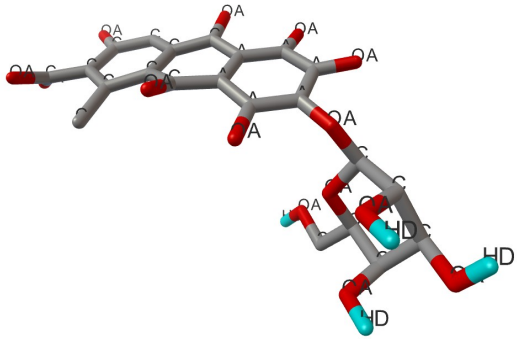
Aspirin (SMILES) : OC(=O)C1=C(C=CC=C1)OC(=O)C

- protonation states
- geometry optimization
- tautomer generation
- custom partial charges



Aspirin **Nicotine**

PROTONATION STATE



OPTIMIZED GEOMETRY

Common file formats:

1D: SMI (multistructure)

2/3D: MOL2 (multistructure)
SDF (multistructure)
PDB

The Scripps Research Institute®

Ligand libraries

Prepare your molecules 2D->3D

Manually

Avogadro [OPEN SOURCE]
http://avogadro.openmolecules.net/wiki/Main_Page

Automatically

Corina [COMMERCIAL, DEMO AVAILABLE]
http://www.molecular-networks.com/online_demos/corina_demo.html

ProDRG2 server [FREE BUT LIMITED]
http://davapc1.bioch.dundee.ac.uk/prodrg/index.html

ChemAxon Marvin [FREE FOR ACADEMICS]
http://www.chemaxon.com/product/marvin_land.html

OpenBabel [OPEN SOURCE]
http://openbabel.org/wiki/Main_Page

Ligand libraries

PubChem

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

Search unique chemical structures using names, synonyms or keywords, SMILES, schetched. Links to available biological property information are provided for each compound

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+FDA approved drugs,metabolites...

$30x10^6$ compounds (non-unique)

- different tautomers/protonation states
- reliable 3D structures
- pre-filtered for not allowed chemical groups and too complex structures

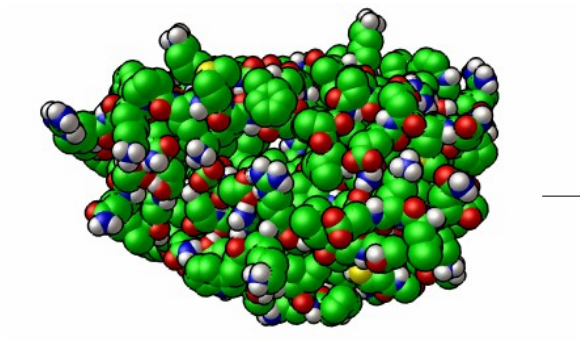
The Scripps Research Institute®

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

AutoDock Single Docking

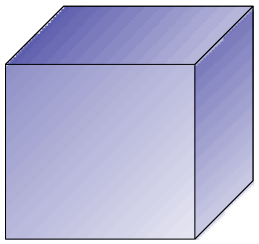
Input
(PDB, Mol2)

PDBQT
(partial
charges,torsions...)

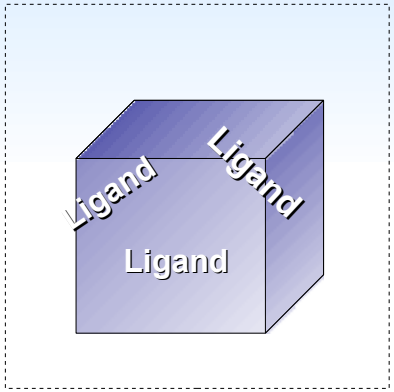


PDBQT
(protonation states,
waters, cofactors...)

Maps



GPF

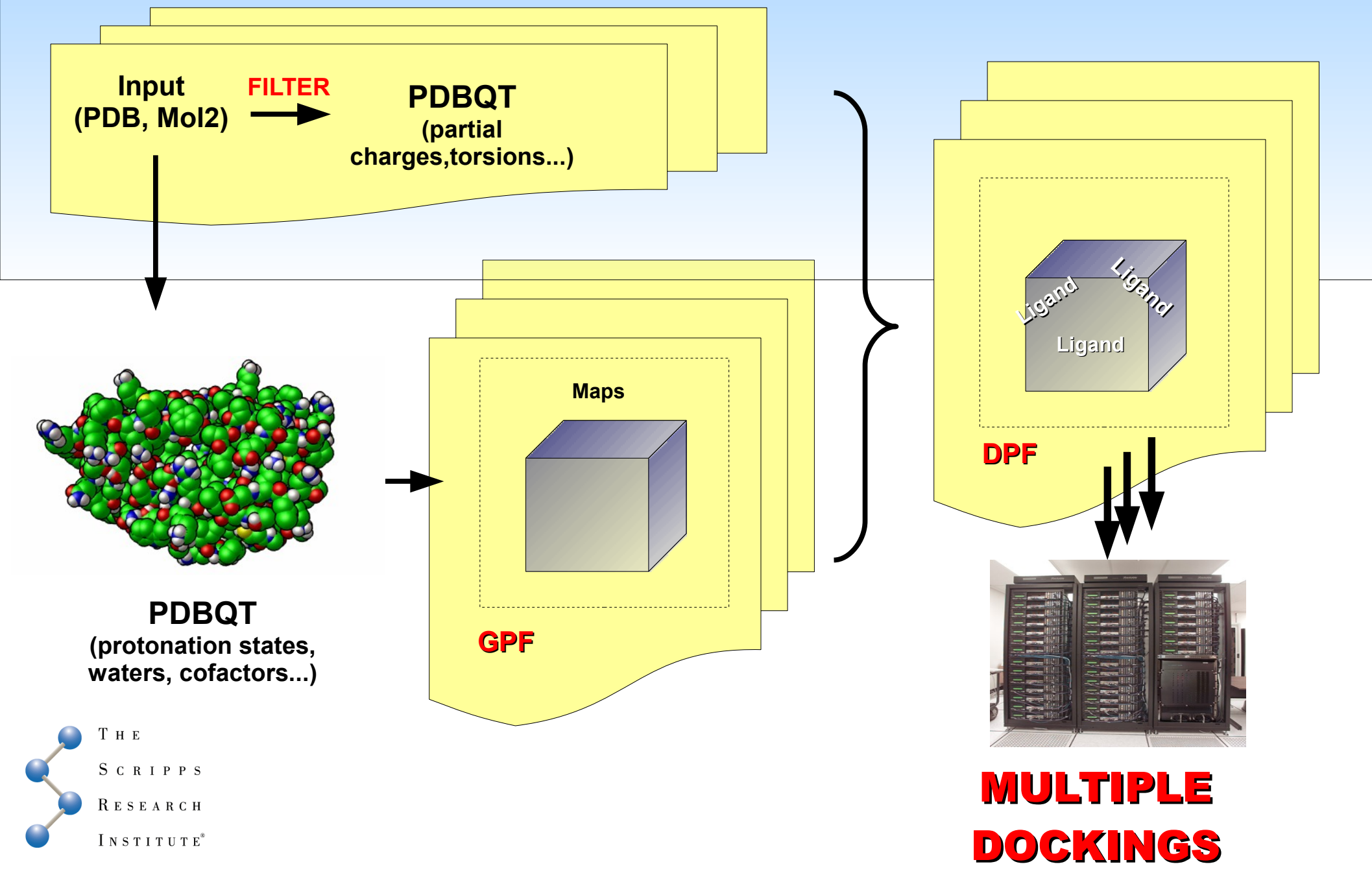


DPF

DOCKING

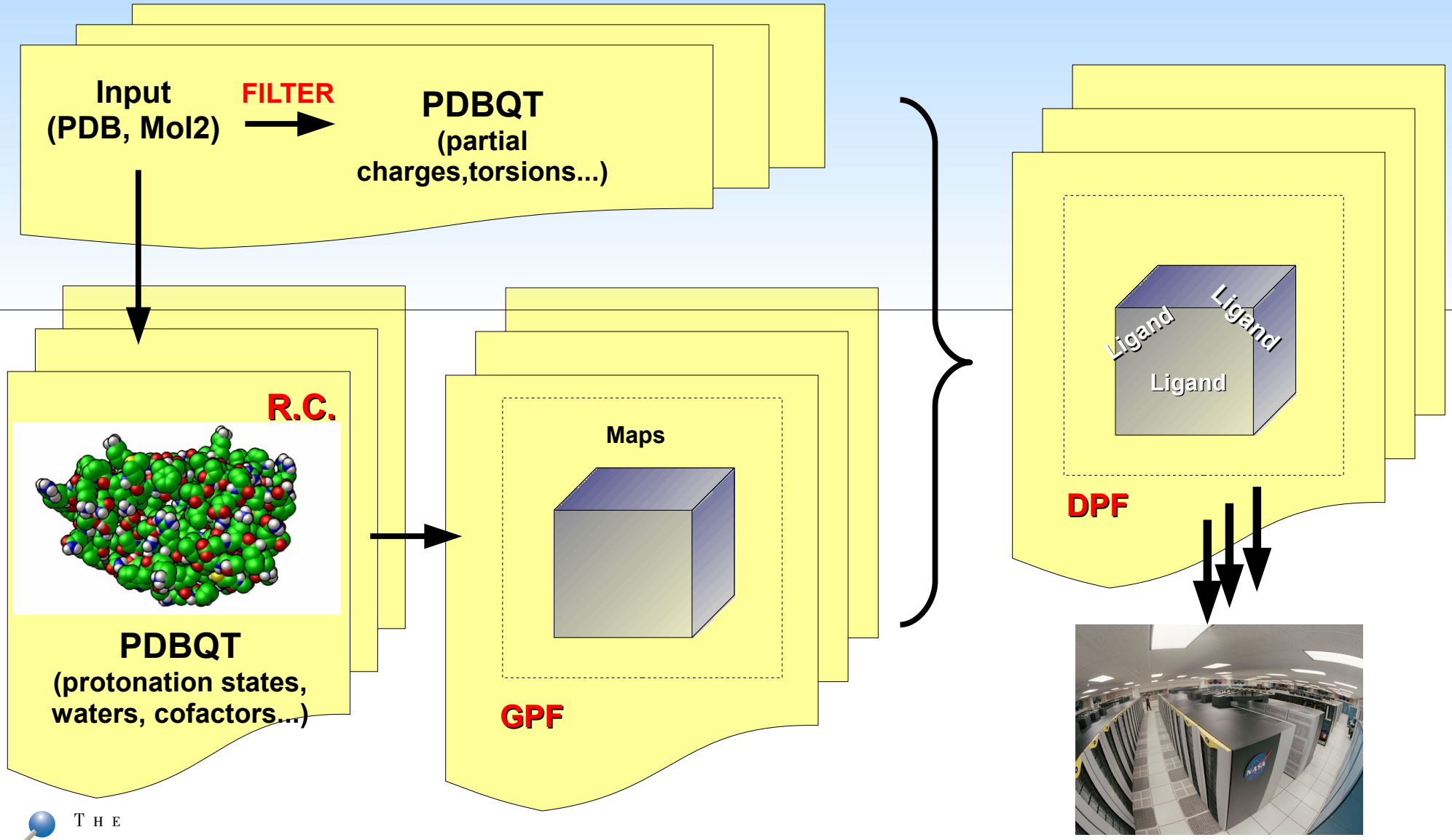
The Scripps Research Institute®

AutoDock VS



The Scripps Research Institute®

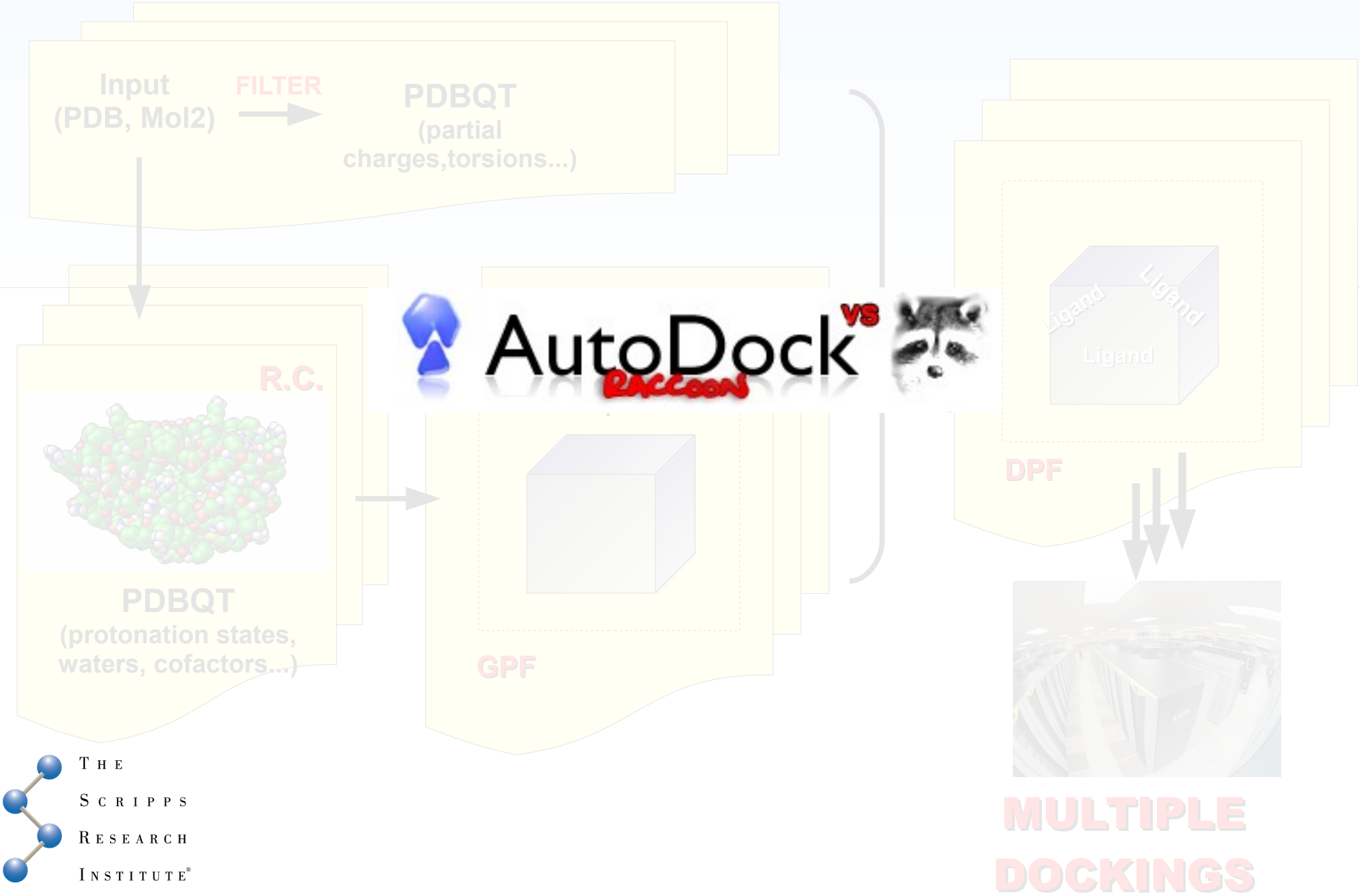
AutoDock VS



MULTIPLE DOCKINGS

The Scripps Research Institute®

AutoDock VS



AutoDock VS RACCOON

The Scripps Research Institute®

Run the software

Double-click on the Raccoon link on your desktop

Workflow

Select the inputs for the VS

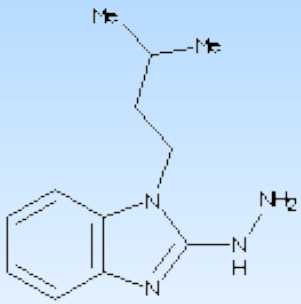
- Define the ligands
- Define the receptor(s)
- Define the grid box
- Define the docking parameters

Generate the VS jobs

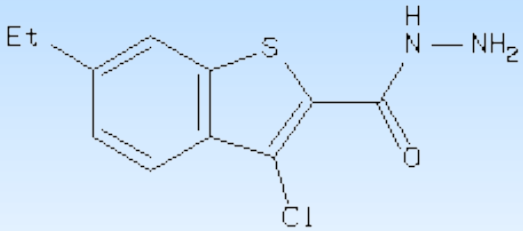
Run the VS jobs

Analyze results

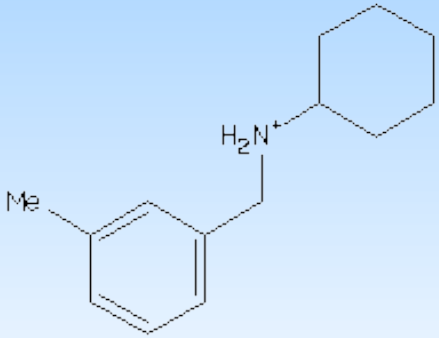
Ligand library



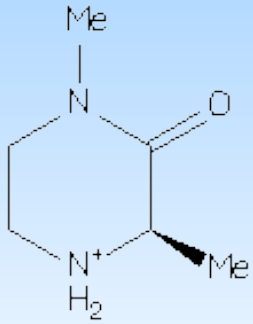
00572736



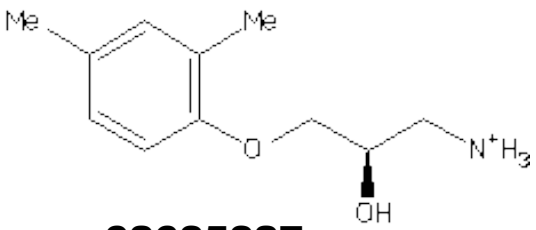
02387687



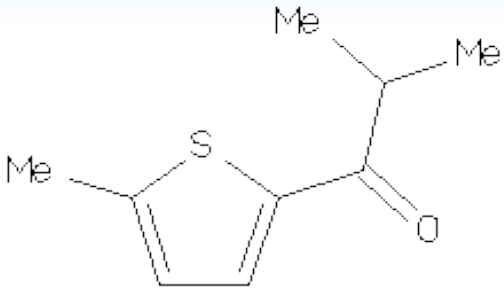
02390501



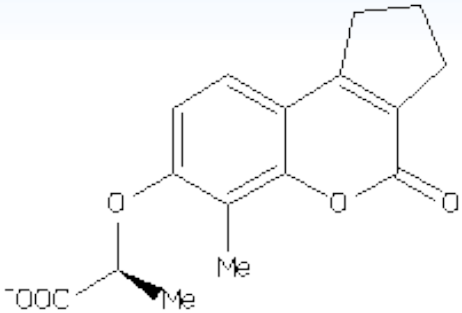
15015303



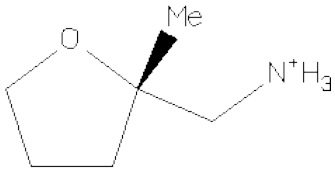
02685827



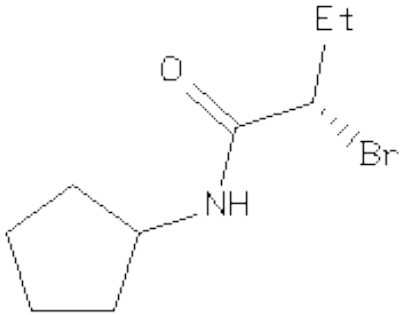
02753817



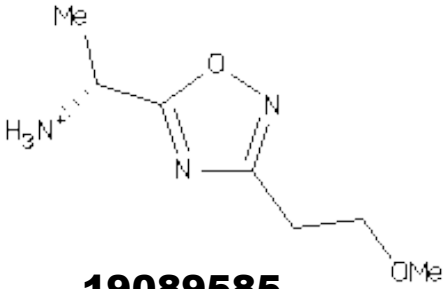
04387707



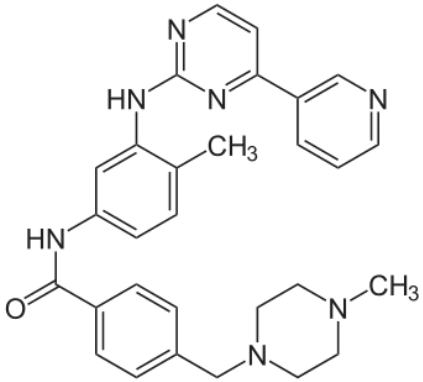
06067294



09355876



19089585



Imatinib

The Scripps Research Institute®

Analyze results

```
Number of distinct conformational clusters found = 2,  out of 87 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 |  80 |##########################################...
_____|____________|_____|____________|_____|_____________________________________

Number of multi-member conformational clusters found = 2, out of 87 runs.
```

BEST ENERGY?

MOST POPULATED CLUSTER?

Analyze results

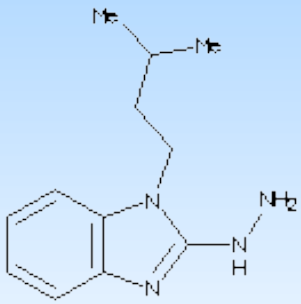
```
Virtual screening summarization from user entropia on atara [Thu Dec 10 18:01:37 PST 2009]
==============================================================================

        Summarize results in    :       /disk2/work/hongkong/raccoon/results/1iep_receptor
 Results sorted based on the lowest energy conformation in the
 largest cluster to find your best docking.
-----------------------------------------------------
 largestCl_dlgfn    #runs #cl #LC    LE_LC     rmsd_LC #ats #tors
-----------------------------------------------------
 imatinib_1iep_receptor/imatinib_1iep_receptor, 10,  2,  9,-15.4100, 38.6005,  40, 6
 ZINC04387707_1iep_receptor/ZINC04387707_1iep_receptor, 10,  5,  4, -9.1000, 59.2051,  21, 3
 ZINC02387687_1iep_receptor/ZINC02387687_1iep_receptor, 10,  4,  7, -8.3300, 61.9815,  19, 3
 ZINC02390501_1iep_receptor/ZINC02390501_1iep_receptor, 10,  5,  4, -7.8100, 52.9734,  17, 3
 ZINC00572736_1iep_receptor/ZINC00572736_1iep_receptor, 10,  7,  3, -6.9000, 56.8724,  19, 4
 ZINC02685827_1iep_receptor/ZINC02685827_1iep_receptor, 10,  6,  4, -6.4000, 57.1118,  18, 6
 ZINC19089585_1iep_receptor/ZINC19089585_1iep_receptor, 10,  2,  9, -6.1100, 60.6241,  15, 5
 ZINC02753817_1iep_receptor/ZINC02753817_1iep_receptor, 10,  4,  7, -5.9300, 63.2043,  11, 2
 ZINC09355876_1iep_receptor/ZINC09355876_1iep_receptor, 10,  3,  5, -5.8400, 58.0156,  13, 3
 ZINC15015303_1iep_receptor/ZINC15015303_1iep_receptor, 10,  2,  8, -5.3000, 56.4589,  11, 0
 ZINC06067294_1iep_receptor/ZINC06067294_1iep_receptor, 10,  5,  4, -5.0900, 55.4987,  11, 2
```

Ligand Efficiency

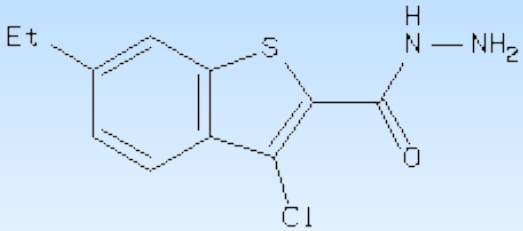
$$\Delta g = \frac{-RT \log K_d}{N_{\text{non-hydrogen atoms}}}$$

The Scripps Research Institute®

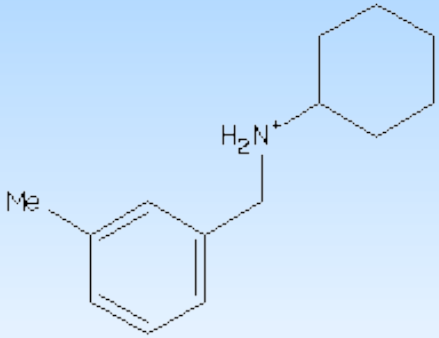
Ligand library



00572736

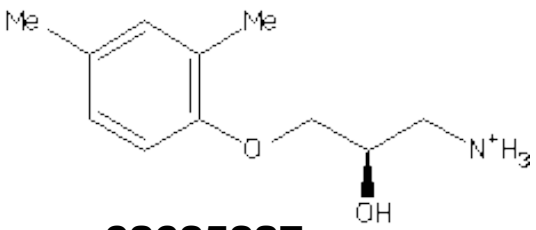


02387687

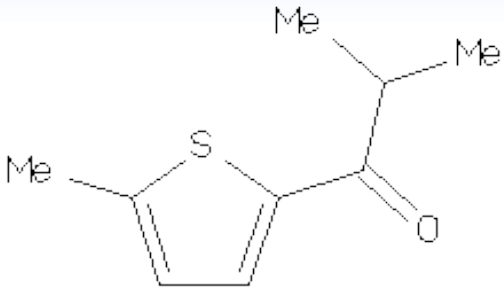


02390501

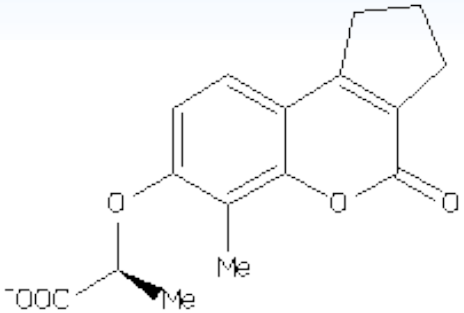
15015303



02685827

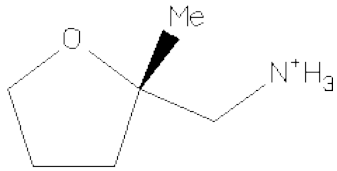


02753817

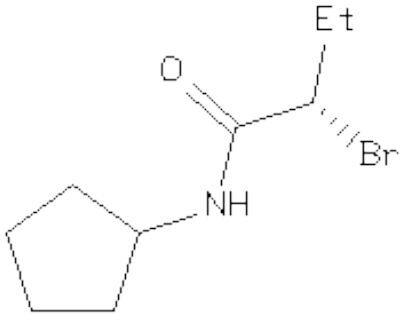


04387707

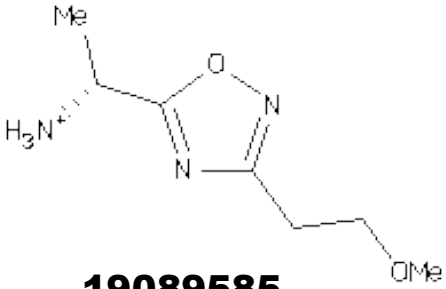
VIRTUAL HIT



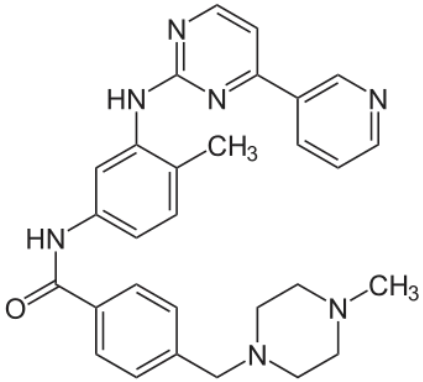
06067294



09355876



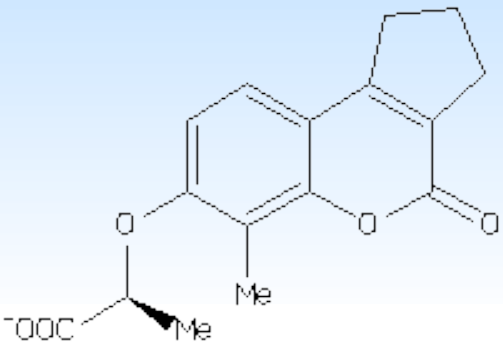
19089585



Imatinib

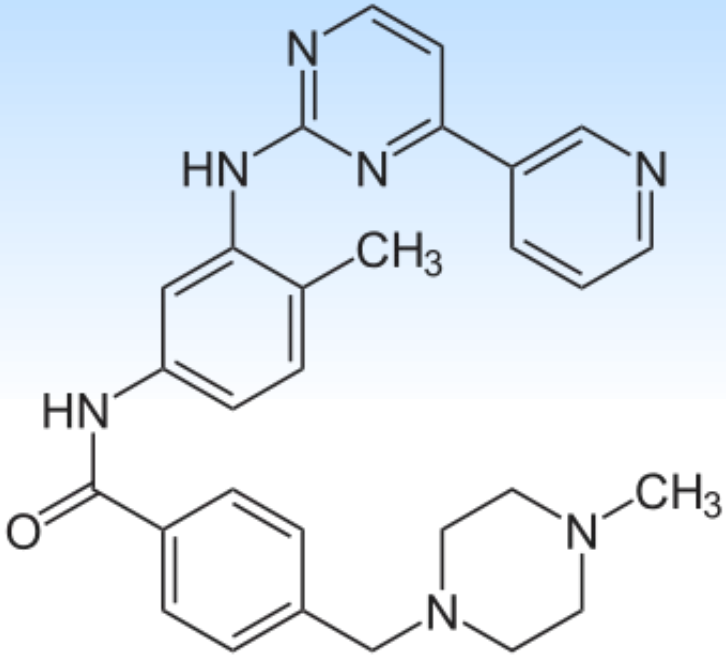
The Scripps Research Institute®

Ligand library



04387707

VIRTUAL HIT



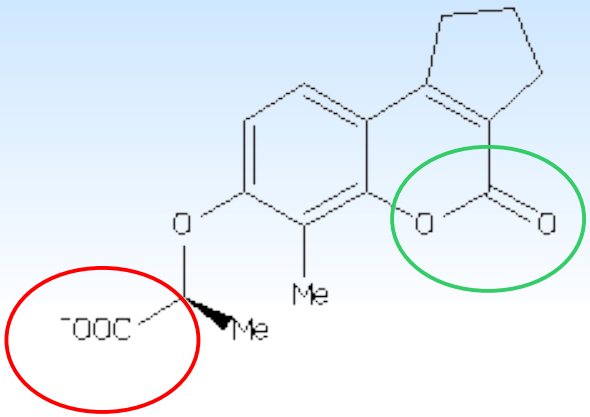
imatinib_xray.pdb

COMPARE THE CONTACTS

- where the interaction is similar
- where the interaction is <u>different</u>
- biological assay to evaluate the activity

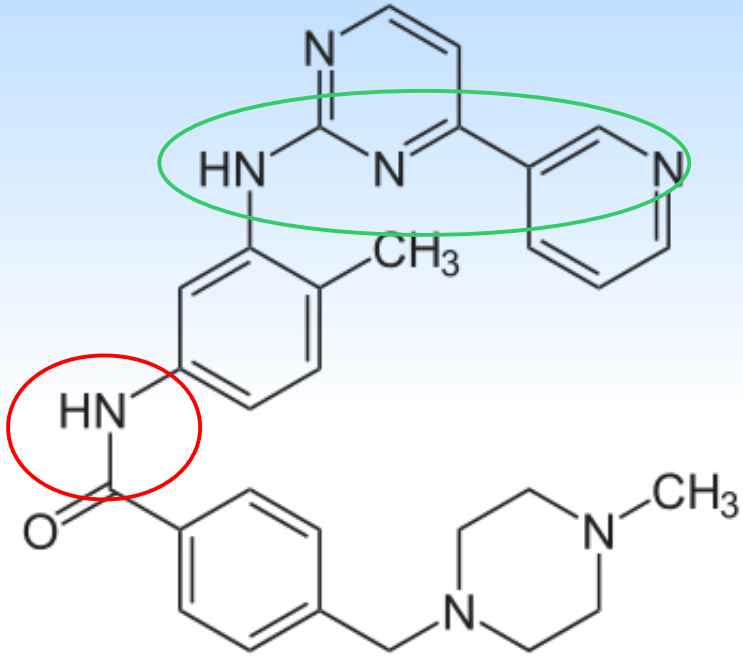
The Scripps Research Institute®

Ligand library



04387707

VIRTUAL HIT



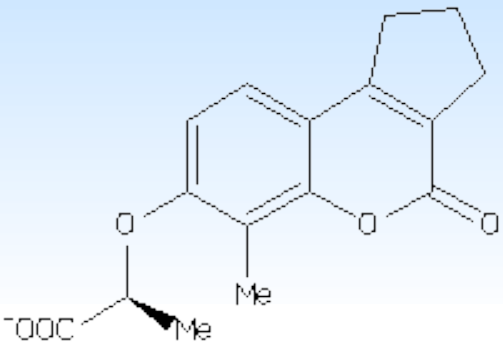
imatinib_xray.pdb

COMPARE THE CONTACTS

- where the interaction is similar **(Met318,Thr315)**
- where the interaction is <u>different</u> **(Lys271)**
- biological assay to evaluate the activity

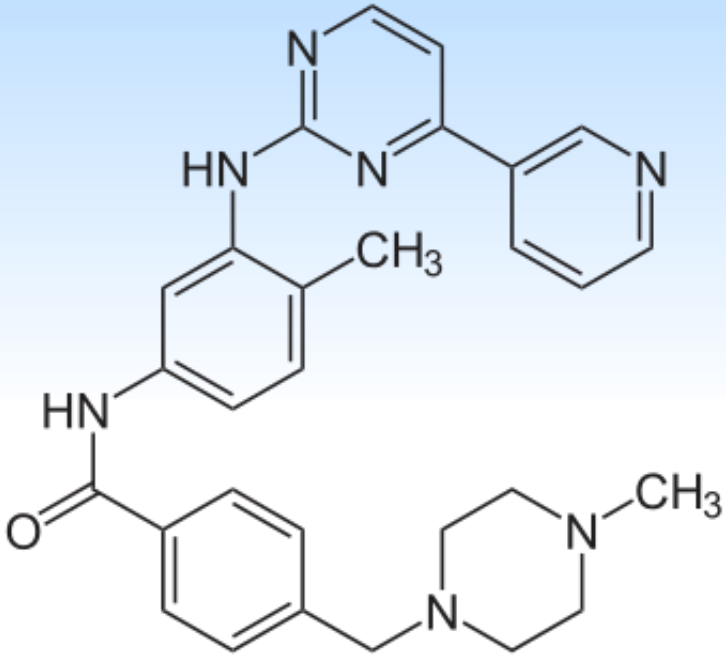
The Scripps Research Institute®

Ligand library



04387707

CONFIRMED HIT



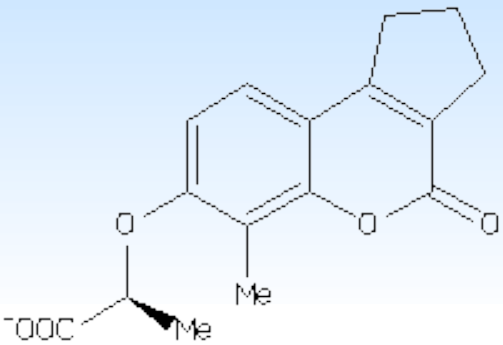
imatinib_xray.pdb

HIT-TO-LEAD

- visualize grid interaction maps (*ADT tutorial, exercise 9*)
- identify where to grow the hit-molecule
- similarity search, focused libraries

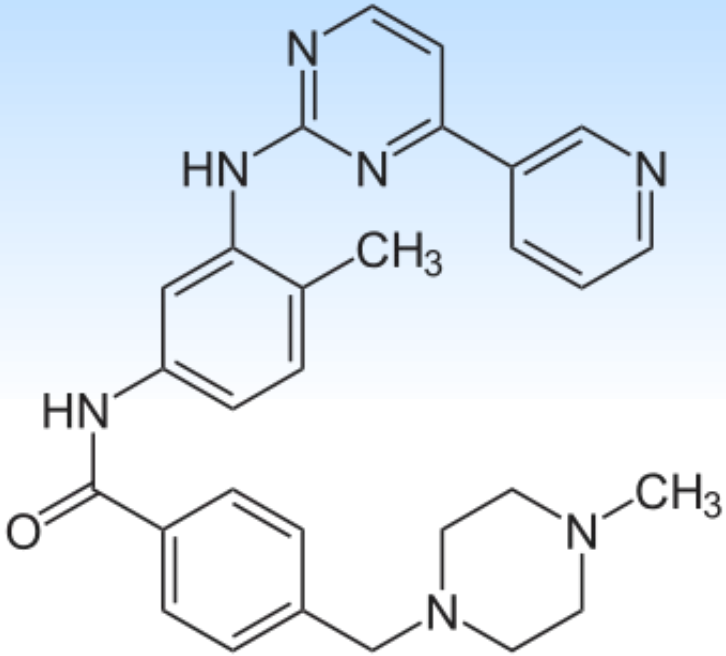
The Scripps Research Institute®

Ligand library



04387707

CONFIRMED HIT



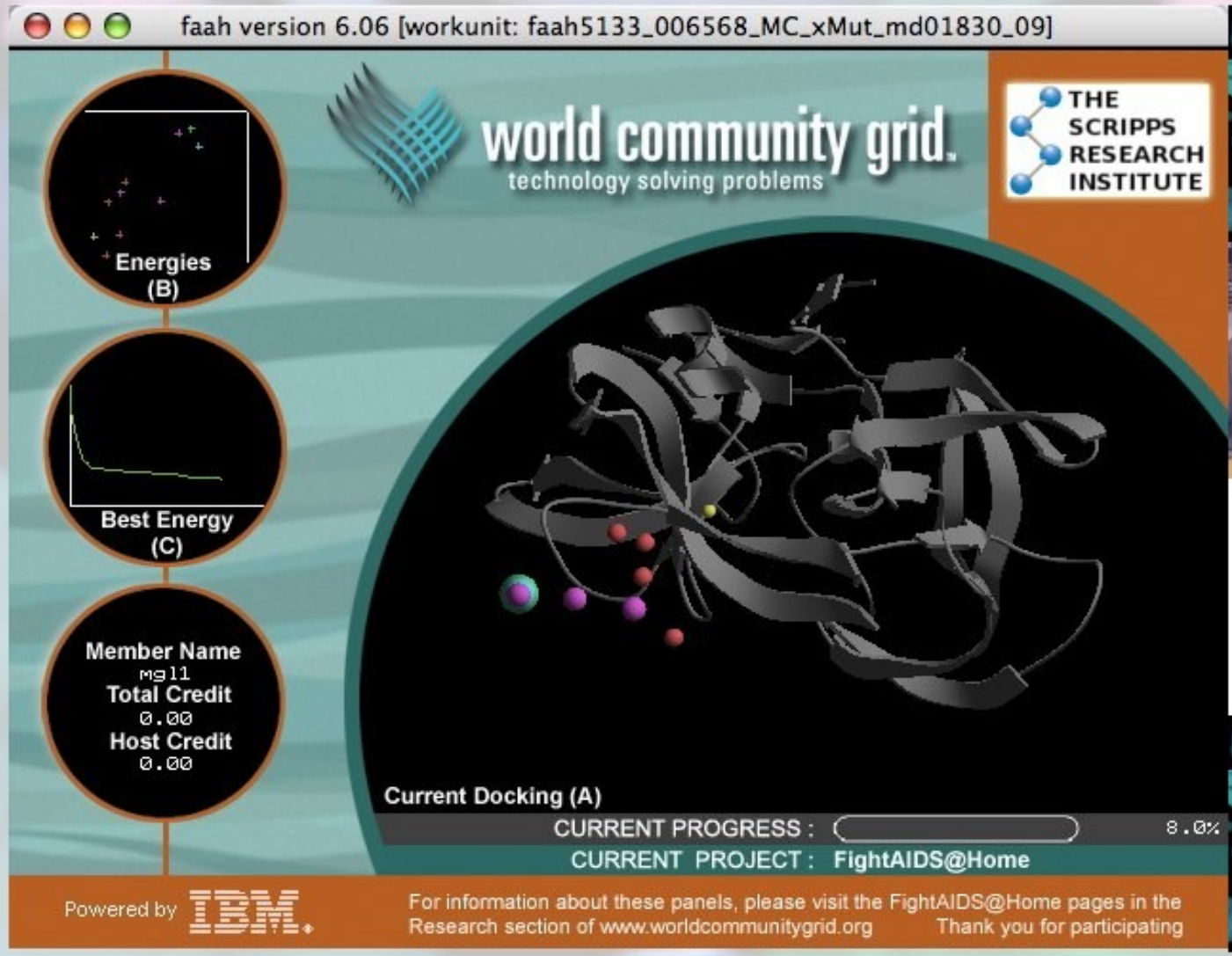
imatinib_xray.pdb

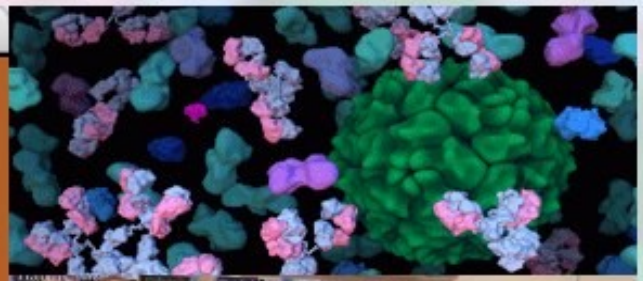
HIT-TO-LEAD

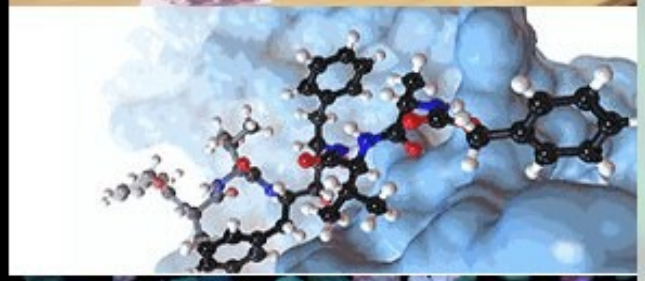
- visualize grid interaction maps
- identify where to grow the hit-molecule
- similarity search, focused libraries
- **PATENT IT**!

The Scripps Research Institute®

FightAIDS@Home







THE SCRIPPS RESEARCH INSTITUTE®

http://fightaidsathome.scripps.edu/

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

For some of the pictures used in this presentation: Wikipedia, CC and GNU-FDL