Using AutoDock for Virtual Screening

David S. Goodsell
Ruth Huey
Alex Perryman
(Garrett Morris, William Lindstrom)
Virtual Screening

**Definition of Virtual Screening:**

Use of high-performance computing to analyze large databases of chemical compounds in order to identify possible drug candidates.


**Virtual Screening is also known as:**
- High-Throughput Docking
- High-Throughput Virtual Screening
Why Use Virtual Screening?

- VS is a computational filter
  - reduces the size of a chemical library to be screened experimentally,
    —Saves time & money (thousand or million fold reduction)

- May improve likelihood of finding a good compound
  - as opposed to random screening
  - enhanced “hit rates”
  - evaluate virtual combinatorial libraries before synthesized

- In the “post-genomic” era, many new targets will be discovered…
HTS versus VS

- High Throughput Screening (HTS):
  - Tests activity *in vitro*.
  - Assays are not infallible (false negatives).
  - Chemical synthesis & testing are expensive.

- Virtual Screening (VS):
  - Computes binding activity *in silico*.
  - VS is also known as “vHTS”.

- HTS and VS are complementary:
  - Use VS to exclude compounds which are predicted not to bind, helping to “enrich” the library…
  - VS can also help to identify false-negatives in HTS
Different Types of Libraries

Which library you choose depends…

- Comprehensive (> ~500,000 compounds)
  - search in the dark

- Diversity-based to cover ‘chemical space’
  - efficient search in the dark

- “Focused” or “Targeted” for lead identification
  - e.g. filtered by 2D or 3D pharmacophores
  - search with a flashlight

- “Focused” or “Targeted” for lead optimization
  - focussing the spotlights

- Combinatorial Libraries
Comparison of Libraries


www.univ-orleans.fr/SCIENCES/ICOA/eposter/eccc9/ECCC9i.htm
NCI Diversity Set

~140,000 compounds

> 1.0 gram available

71,756 compounds

Diversity based on unique 3-point pharmacophores

1,990 compounds

How it was built...
Chem-X (Oxford Molecular Group) was used.

(1) Defined 3-point pharmacophores based on hydrogen bond acceptor, hydrogen bond donor, positive charge, negative charge, aromatic, hydrophobic, acid, base and defined distance intervals.

(2) Generated a set of ~1,000,000 pharmacophores for all acceptable conformations of each structure.

(3) A diverse subset was built up by comparing all pharmacophores for each acceptable conformation, adding the structure to the set if it had 5 or more new pharmacophores.
SMILES

- **Simplified Molecular Input Line Entry Specification**
- A string of letters, numbers and other characters that specify the atoms, their connectivity, bond orders, & chirality
Small Molecule Structures

- **Sources of Small Molecule Structures:**
  - **CCDC’s Cambridge Structural Database**
    - [http://www.ccdc.cam.ac.uk/products/csd/](http://www.ccdc.cam.ac.uk/products/csd/)
  - **NCI, National Cancer Institute**
  - **PubChem**
  - **ZINC, ZINC Is Not Commercial**
    - [http://blaster.docking.org/zinc/](http://blaster.docking.org/zinc/)

- **More information:**
  - **Molecular Docking Web**
    - [http://www.scripps.edu/mb/olson/people/gmm/index.html#SmallMolecules](http://www.scripps.edu/mb/olson/people/gmm/index.html#SmallMolecules)
Converting 1D & 2D to 3D

- **Corina**
  - 1000 structures for free
  - Specify input as SMILES or sketch using JME
  - [http://www2.ccc.uni-erlangen.de/software/corina/free_struct.html](http://www2.ccc.uni-erlangen.de/software/corina/free_struct.html)

- **PRODRG**
  - Specify input as PDB, MDL MOL or sketch using JME; returns PDBQ format
  - [http://davapc1.bioch.dundee.ac.uk/programs/prodrg/](http://davapc1.bioch.dundee.ac.uk/programs/prodrg/)

- **ZINC**
  - Specify input as SMILES
  - [http://blaster.docking.org/zinc/](http://blaster.docking.org/zinc/)
Strategy

- Find the 3D structure and inhibition constant $K_i$ of a complex of your desired target with an inhibitor (‘positive control’)
- Perform a “re-docking” on your positive control to verify your input files and parameters are reasonable.
- Note the predicted binding free energy (BFE) from AutoDock
- This energy, plus the standard deviation in the predicted BFE of the AutoDock 3 force field, 2.1 kcal/mol, forms the threshold above which we will be looking for “hits”, molecules with better BFE than the positive control’s BFE.
- Add the positive control inhibitor(s) to your library before
e.g. AICAR Transformylase

- AutoDock 3 was used to screen the NCI Diversity Set, 1990 compounds, against AICAR transformylase, an enzyme involved in the purine biosynthetic pathway.

- AutoDock Parameters used:
  - 5 million evals per run
  - 100 runs per compound

- Took about 2 weeks using 32 nodes of the Scripps Research Institute’s “redfish” Linux cluster (circa 2003)

Chenlong Li
VS & Kinetic Inhibition Results

- **In silico:**
  44 top compounds, $E_{binding} \leq -13.0$ Kcal/mol

- **In vitro:**
  10 are insoluble in water
  18 precipitate in buffer solution
  8 out of 16 soluble compounds bind
    (50% success)

Chenlong Li et al., *J. Med. Chem.* ASAP article. Web Release Date: December 2, 2004
Tyrosine Phophatase 1B (PTP1B)

- **HTS (in vitro) of 400,000 compounds**
  - 300 hits with IC$_{50} < 300$ μM
  - 85 validated hits with IC$_{50} < 100$ μM
  - 0.021% hit rate ( = 85 / 400,000)
  - many violate Lipinski rules

- **VS (in silico) of 235,000 compounds (DOCK)**
  - 365 high-scoring molecules
  - 127 validated hits with IC$_{50} < 100$ μM
  - 34.8% hit rate ( = 127 / 365)
  - hits are more drug-like

Single Docking v. Library Screen

- Use GUI
- Data in one directory
- Prepare input files:
  - Ligand PDBQ
  - Receptor PDBQS
  - GPF
  - DPF
- One AutoGrid calculation
- One AutoDock calculation
- Analyze Results

- Use scripts
- Data in tree structure
- Prepare input files:
  - Library of Ligand PDBQ files
  - Receptor PDBQS
  - 2 GPFs
  - Library of DPFs
- Two AutoGrid calculations
- Submit jobs to cluster
- Rank Results; Analyze best
Virtual Screening Tutorial Map

- **diversity.sdf** → **Exercise 1**
- ***.pdb** → **Exercise 2**
- ***.pdbq** → **Exercise 3**
- **ligand_dict.py**

- **x1hpv.pdb** → **Exercise 4**
- **x1hpv.pdbqs** → **Exercise 5**
- **x1hpv_*.gpf** → **Exercise 6**

**X1hpv*map***

- **ind_x1hpv.dpf**
- **Ind_x1hpv.dlg**

- ***_x1hpv.dpf** → **Exercise 8**
- ***_x1hpv.dlg** → **Exercise 9**
Virtual Screening Tutorial

Directory Structure

/home/ADT

$VSTROOT

VSTutorial

Results

VirtualScreening

scripts

classes

Ligands

Receptor

Dockings

diversity*_x1hpv
... diversity*_x1hpv

ind_x1hpv
Virtual Screening Tutorial
Directory Structure

/home/ADT

$VSTROOT

/home/ADT/VSTutorial

/home/ADT/VSTutorial/diversity.sdf
/home/ADT/VSTutorial/scripts/ex*.csh
/home/ADT/VSTutorial/*.py
/home/ADT/VSTutorial/Results
/home/ADT/VSTutorial/VirtualScreening

/home/ADT/VSTutorial/VirtualScreening/etc
/home/ADT/VSTutorial/VirtualScreening/Ligands
/home/ADT/VSTutorial/VirtualScreening/Receptor
/home/ADT/VSTutorial/VirtualScreening/Dockings

/home/ADT/VSTutorial/VirtualScreening/Dockings/diversity0001_x1hpv
/home/ADT/VSTutorial/VirtualScreening/Dockings/diversity*_x1hpv
/home/ADT/VSTutorial/VirtualScreening/Dockings/ind_x1hpv
General Comments

- use **pwd** and **ls** often

  If you are unfamiliar with the Unix command line and/or navigating around a hierarchical file system, use the **pwd** (print name of current/working directory) and **ls** (list directory contents) shell commands as much as you need to stay oriented in the file system. It's always helpful to draw a quick picture.

- use **man** often

  Use the **man** command copiously. Unix has a very useful on-line manual that you can read with the **man** command. For example, if you can't remember how to use the **ls** command to list a directory contents with file modification dates, type **man ls**. This will display the on-line manual page which describes the **ls** command and allow you quickly learn how do it. **man -k** is a useful option when you can't remember the name of the command you want to read about. (Look up **man -k** in the on-line manual by typing **man man**).
Unix Shell Commands Used

- ls
- pwd
- cd
- mkdir
- ..
- man
- setenv
- echo
- foreach
- >
- |
- cp
- ln -s
- cat
- more
- head
- tail
- wc
- grep
- sort
- awk
- sed
- vi or emacs