BCMB422

Intro to Kraken Project

Ensemble Docking

Sally R. Ellingson, PhD Candidate
Center for Molecular Biophysics, UT/ORNL (Advisor: Jerome Baudry)

SALLY R. ELLINGSON

BS in Computer Science and BS in Mathematical Sciences
Florida Institute of Technology, Melbourne, FL

Pursuing PhD in Genome Science and Technology (GST)
Interdisciplinary Graduate Minor in Computational Science (IGMCS)
University of Tennessee, Knoxville (UTK)

NSF-IGERT Fellow (Scalable Computing and Leading Edge Innovative Technologies)

Graduate Research Assistant at UT/ORNL Center for Molecular Biophysics

Research Interest: Intersection of high-performance computing and drug discovery

sellings@utk.edu

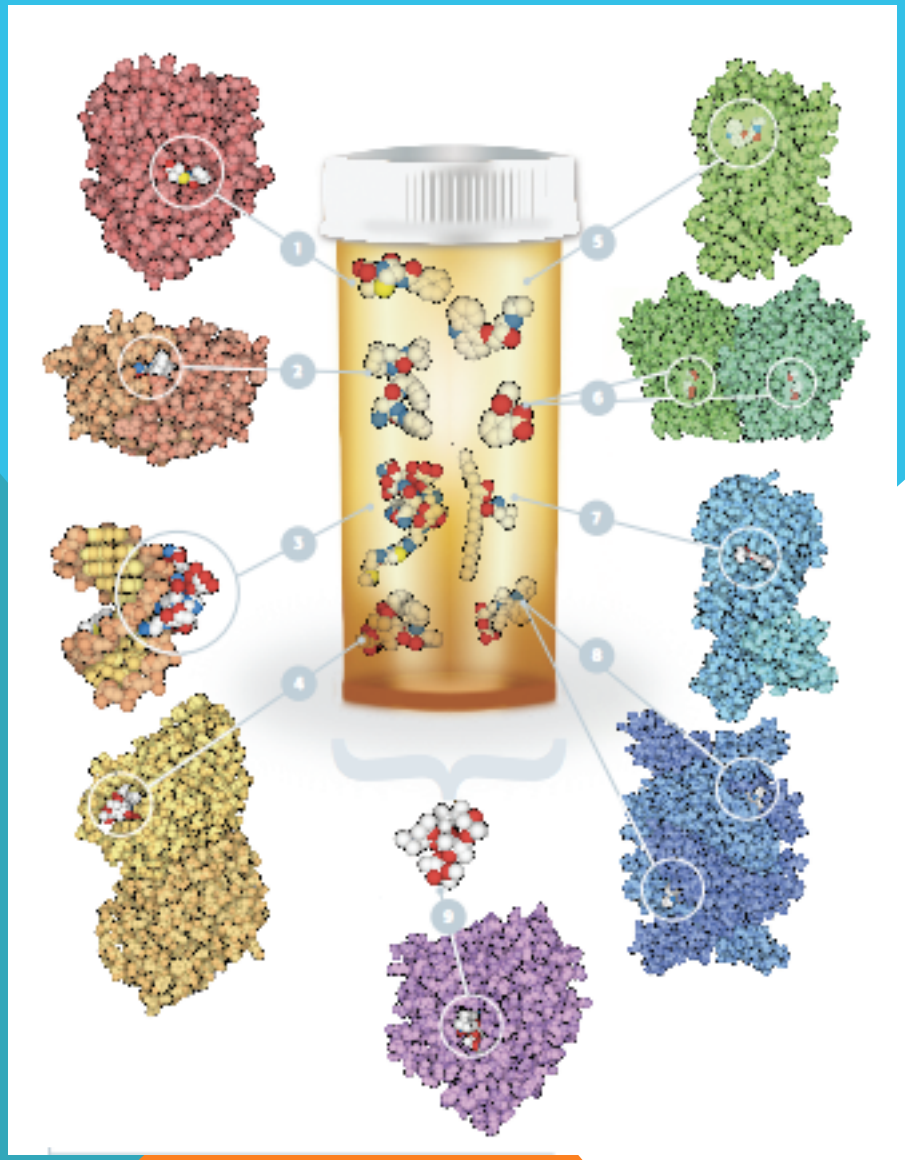
OUTLINE

- What is …
  - a drug?
  - virtual docking?
  - a high-throughput screen?
  - enrichment?
  - ensemble docking?

- How do we do it on supercomputers?

- What will we do in the rest of this workshop?

HOW DO DRUGS WORK?



Antibiotics & Antivirals

1 2

Antibiotics and antiviral drugs are specific poisons. They need to kill pathogenic organisms like bacteria and viruses without poisoning the patient at the same time. Often, these drugs attack proteins that are only found in the targeted bacterium or virus and which are crucial for their survival or multiplication. For instance, **penicillin** attacks the enzyme that builds bacterial cell walls, and HIV protease inhibitors like **saquinavir** attack an enzyme that is needed for HIV maturation.

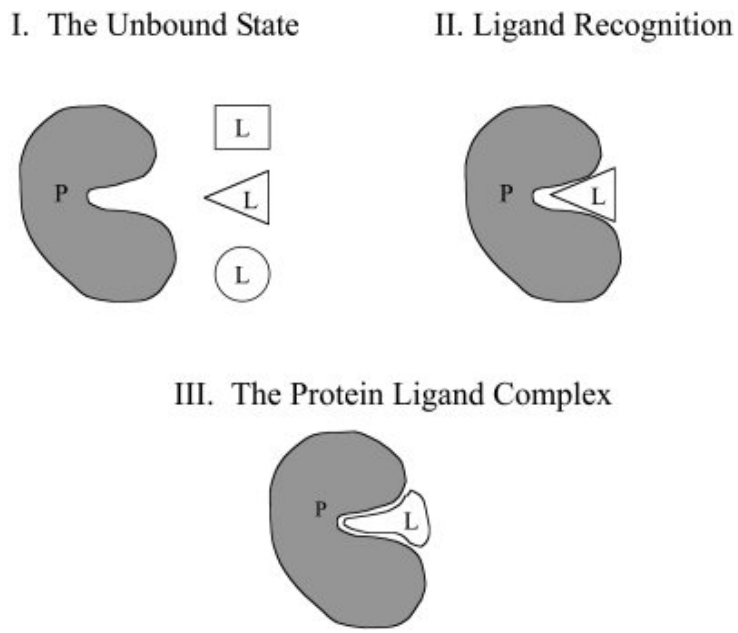
1. D-alanyl-D-alanine carboxypeptidase with penicillin (1pwc)
2. HIV protease with saquinavir (1hxb)

DRUGS are small molecules that bind to proteins and modify their actions

WHAT IS MOLECULAR DOCKING?

Efficiently…

- **Predicts conformation of a protein-ligand complex**
- **Predicts binding affinity of the ligand to the protein**



(+) Reproduce correct bound conformation

(+) Assign better scores to high-affinity ligands than to decoys (enrichment)

(-) Generate scores that correlate with measured binding affinities

Docking engines such as Autodock4 and Autodock Vina allow for **virtual binding experiments** with atomic level detail on a desktop computer

AUTODOCK VINA

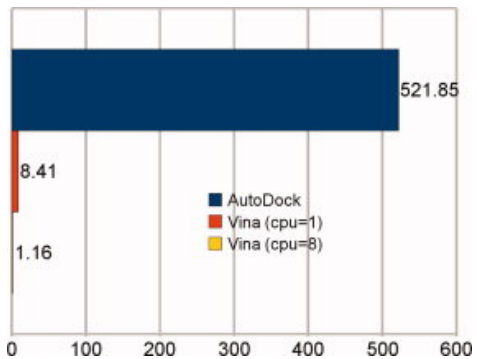
DOCKING ENGINE FOR MULTI-PROTEIN SCREENING

Scoring function: machine-learning approach

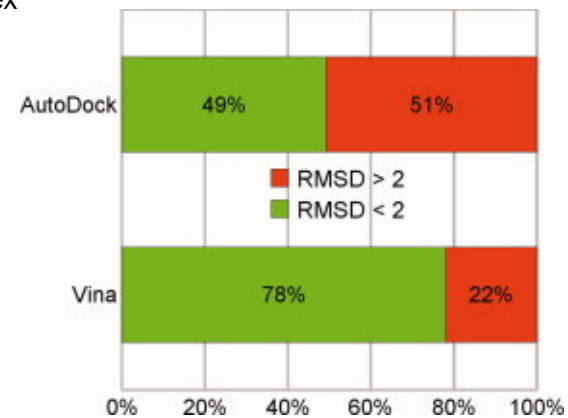
Conformational search: iterated local search global optimizer

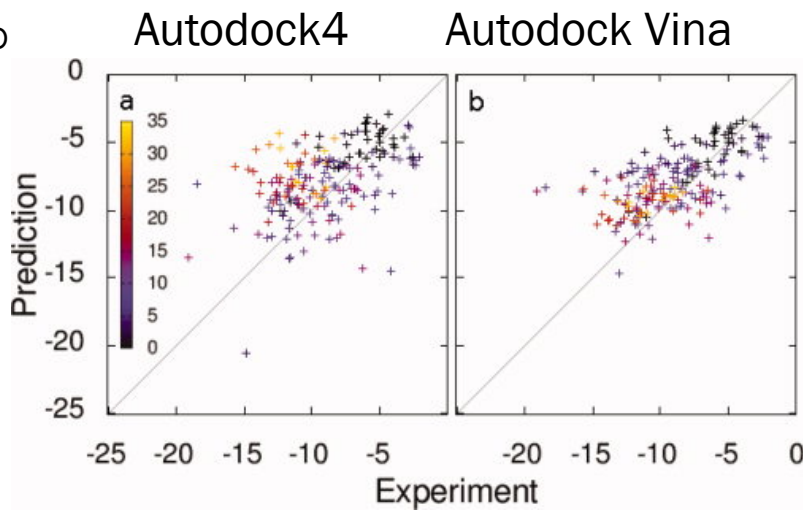
step mutation, local optimization, Metropolis acceptance criterion

Average time in minutes per complex



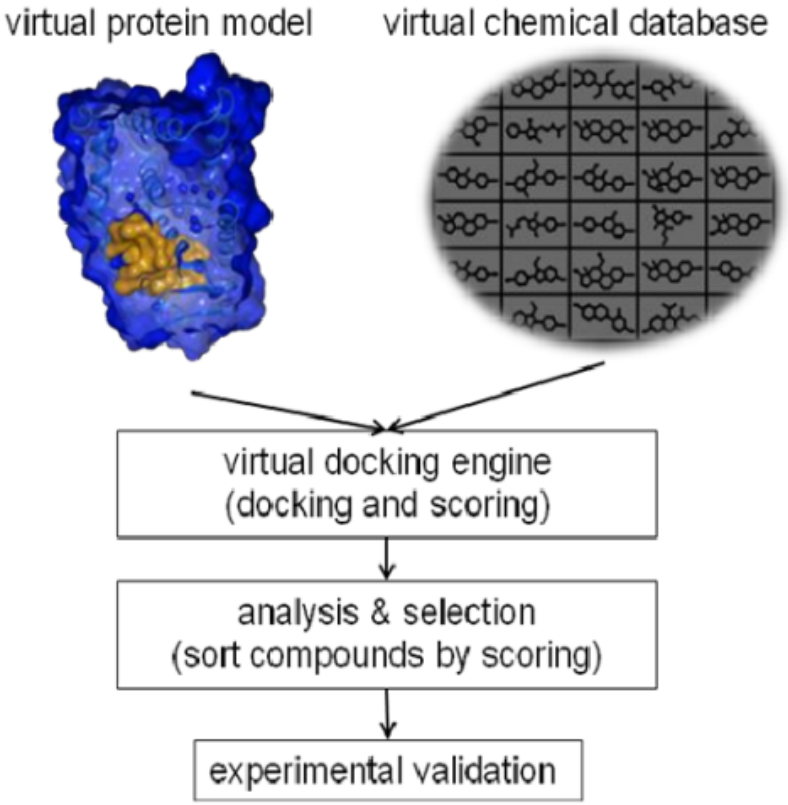
Predicted structures within 2 Å RMSD

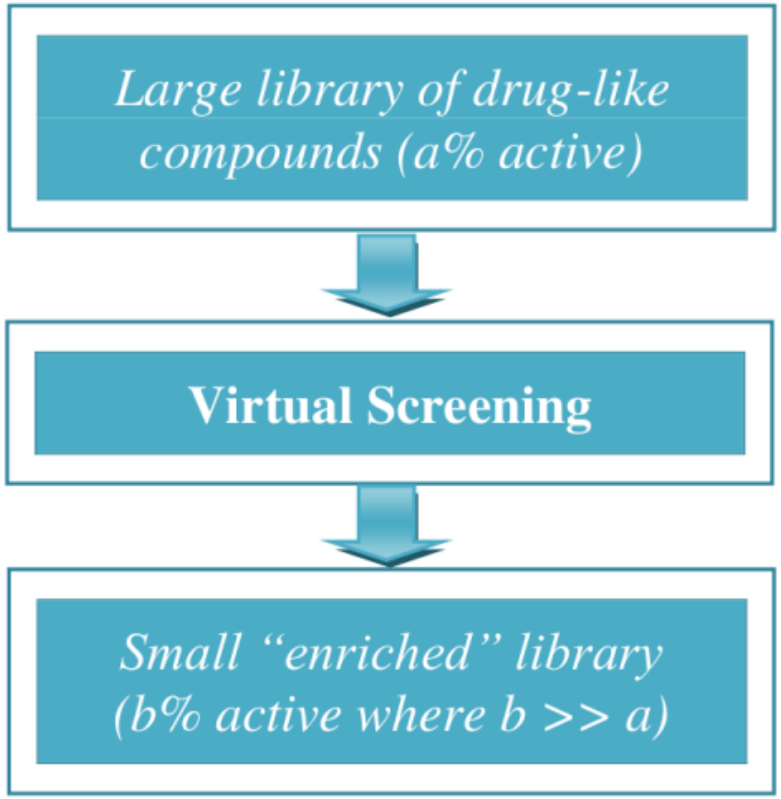




Trott, O. and Olson, A. J. (2010), AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem., 31: 455–461. doi: 10.1002/jcc.21334

WHAT IS A VIRTUAL HIGH-THROUGHPUT SCREEN?



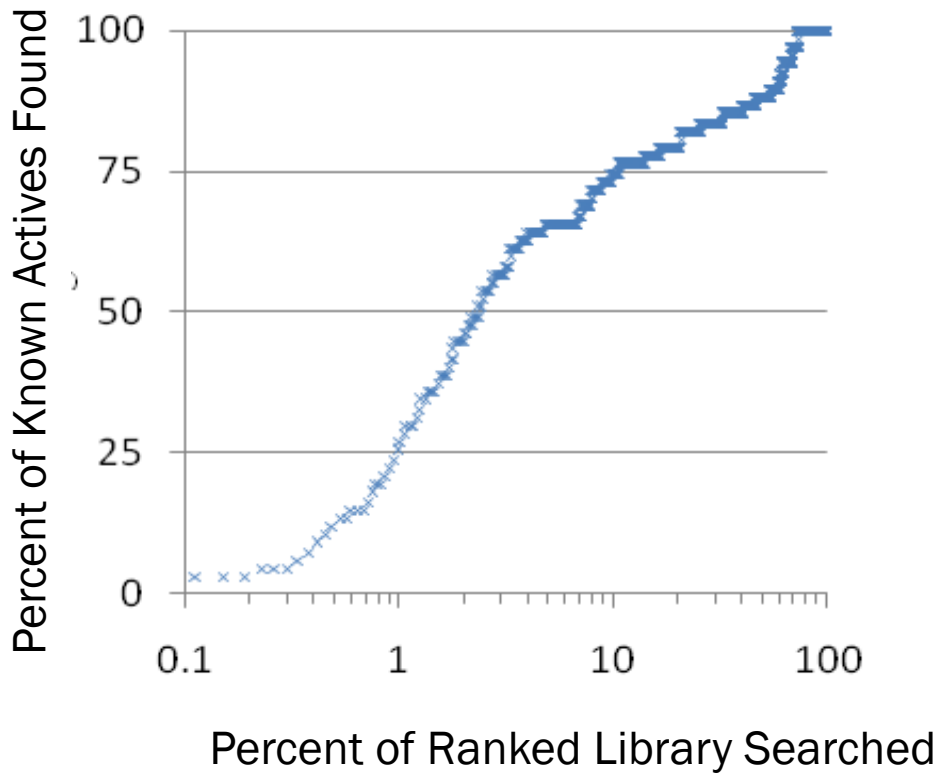


HOW DO WE ASSESS THE ENRICHMENT OF A DOCKING METHOD?

Benchmark of known active and decoy compounds

HTS

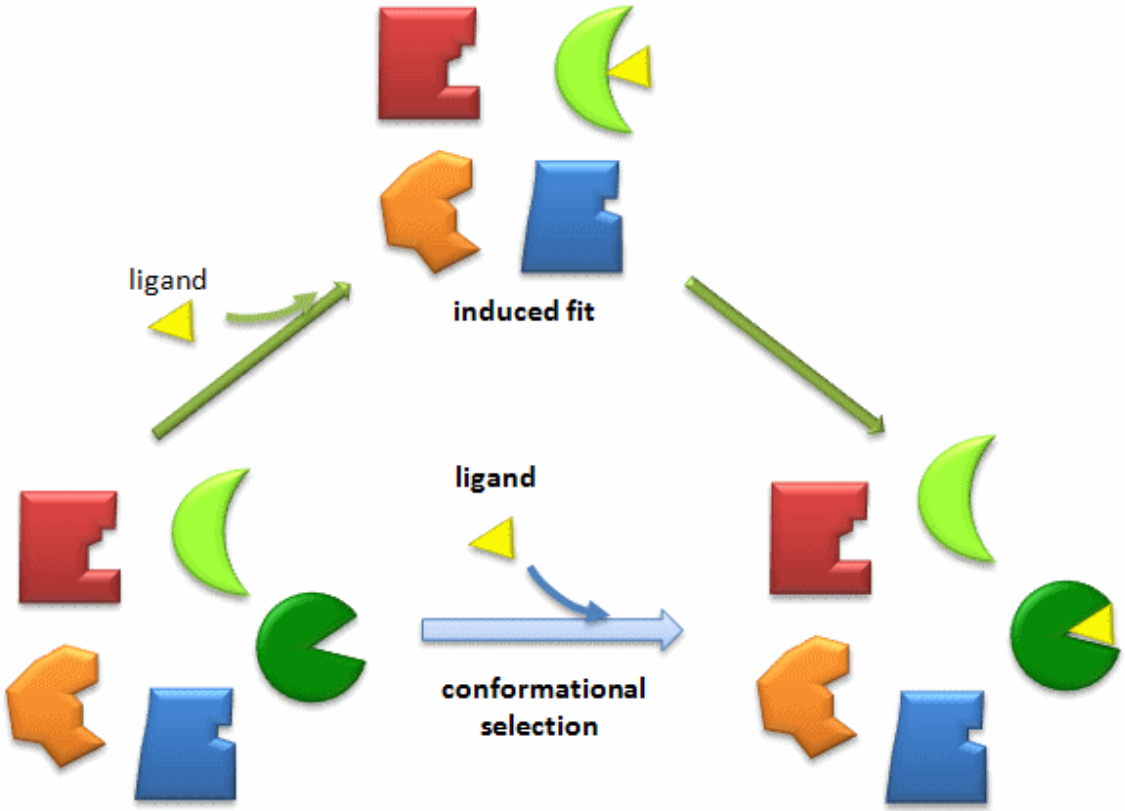
Ranked list of compounds
candidate1
candidate2
candidate3
⋮



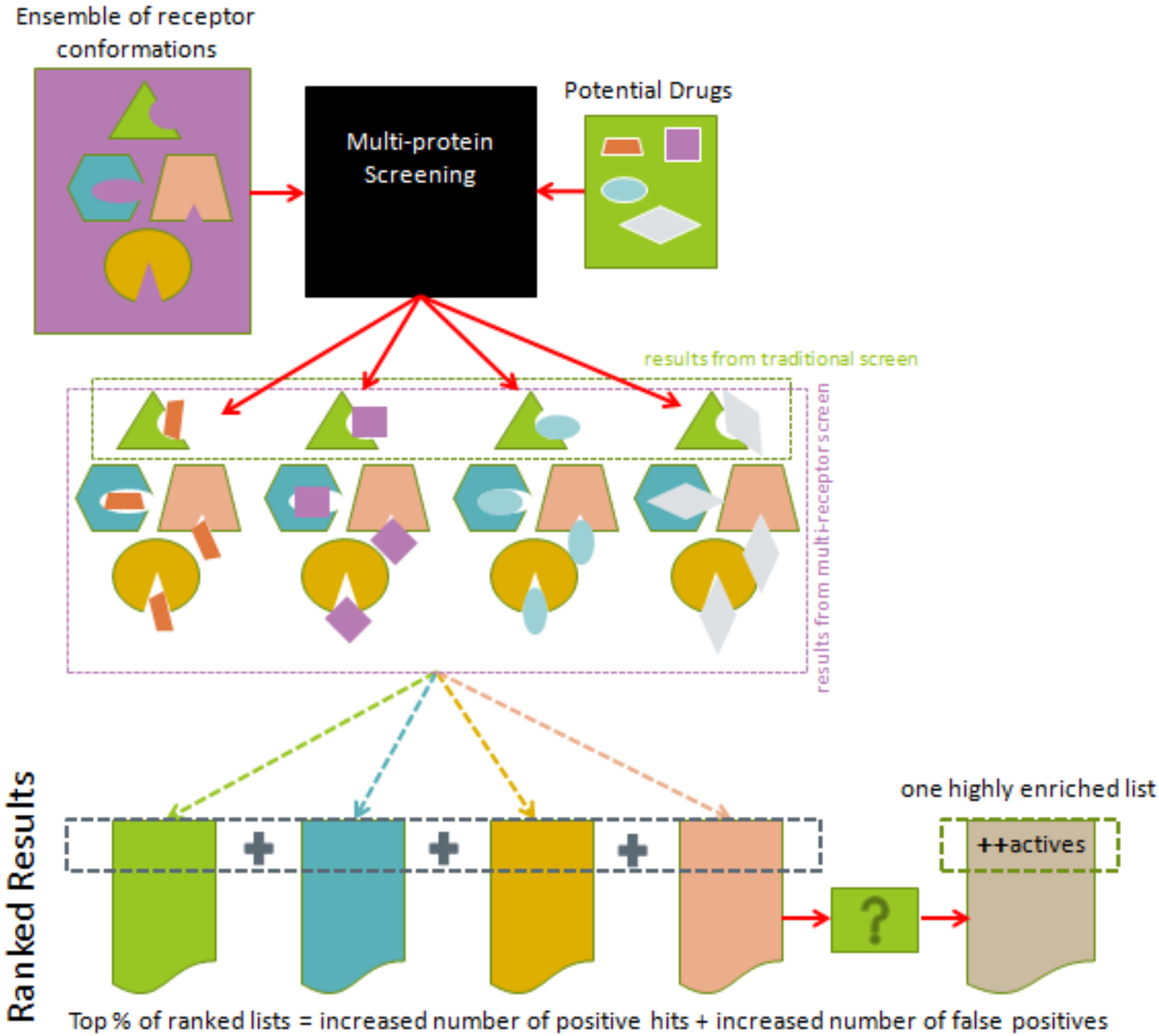
WHAT IS ENSEMBLE DOCKING?

Proteins are not static structures but are treated mostly so in virtual docking.

Proteins may exist is many druggable conformational states.



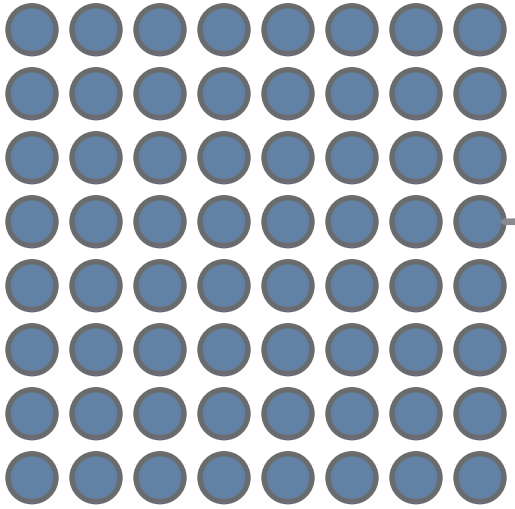
ensemble docking



HOW DO WE DO IT ON SUPERCOMPUTERS?

Even Distribution of Complexity (EDC)

Every worker determines its own set of non-overlapping tasks



worker
id = 31
total_complexity = n
average_complexity = n / #workers
get_work_set(id, average_complexity)
Dock(work_set)

$$\text{total_complexity} = \sum^{\text{library}} \text{tdof}$$

tdof = torsional degrees of freedom in ligand

WHAT WILL WE DO IN THE REST OF THIS WORKSHOP?

1) estrogen receptor agonist

Good enrichment with a crystal structure

In class, walk-through of creating an enrichment plot using only the crystal structure. All files are prepared, you must login and execute commands on Kraken. Extra credit will be offered.

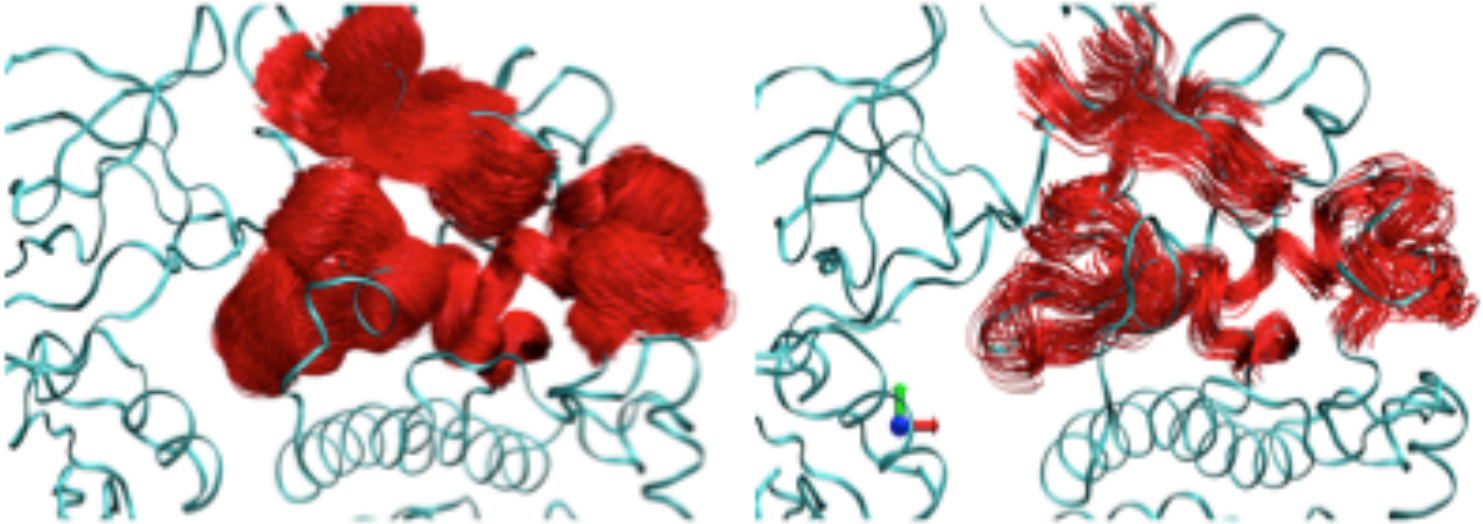
2) tyrosine-protein kinase SRC

Poor enrichment with a crystal structure

In class project, work at own pace and finish on own time. Prepare files to do an ensemble screen with SRC. Everyone will use slightly different parameters and create enrichment plots in order to decide which method worked the best.
This is novel research!!!

http://cmb.ornl.gov/~sek/bcmb422.html

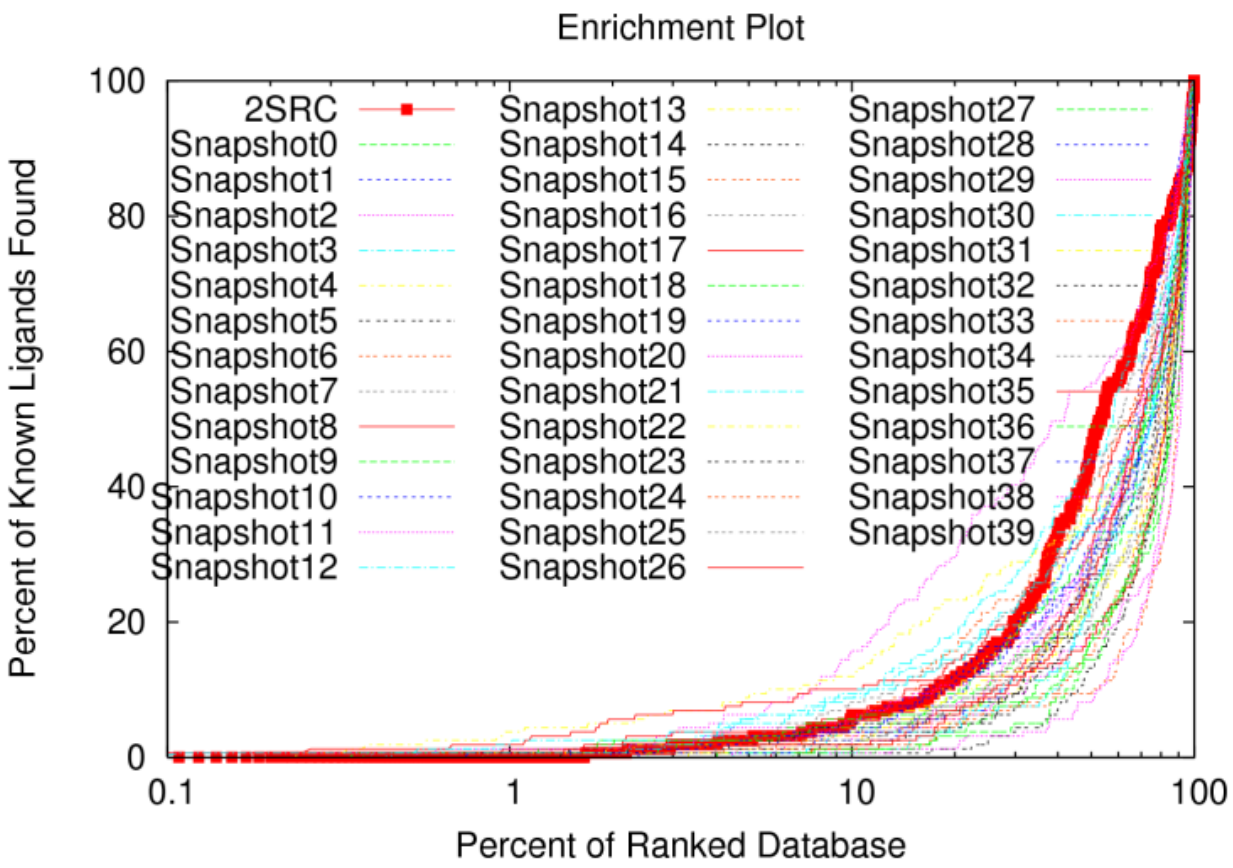
TYROSINE-PROTEIN KINASE SRC



2000 snapshots from a 200ns molecular dynamics simulation

40 distinct states identified with clustering

red indicates the binding site region


Enrichment Plot
Percent of Known Ligands Found
Percent of Ranked Database
100
80
60
40
20
0
0.1
1
10
100
2SRC
Snapshot0
Snapshot1
Snapshot2
Snapshot3
Snapshot4
Snapshot5
Snapshot6
Snapshot7
Snapshot8
Snapshot9
Snapshot10
Snapshot11
Snapshot12
Snapshot13
Snapshot14
Snapshot15
Snapshot16
Snapshot17
Snapshot18
Snapshot19
Snapshot20
Snapshot21
Snapshot22
Snapshot23
Snapshot24
Snapshot25
Snapshot26
Snapshot27
Snapshot28
Snapshot29
Snapshot30
Snapshot31
Snapshot32
Snapshot33
Snapshot34
Snapshot35
Snapshot36
Snapshot37
Snapshot38
Snapshot39

ACKNOWLEDGEMENTS

Genome Science and Technology, UT

Center for Molecular Biophysics, UT/ORNL

SCALE-IT, NSF/IGERT
Scalable Computing and Leading Edge Innovative Technologies

XSEDE and National Institute for Computational Sciences

Georgetown University

NIH-CTSA

Center for Molecular Biophysics UT / ORNL

GENOME SCIENCE AND TECHNOLOGY

SCALE-IT
Scalable Computing and Leading Edge Innovative Technologies

QUESTIONS??????