Open Science Grid

Post ASP 2012:
Protein Docking
at the IU School of Medicine

Protein Docking with IU School of Medicine

- SPLINTER - Structural Protein-Ligand Interactome
- Used autodock-vina – "…open-source program for drug discovery, molecular docking and virtual screening…"
- Frist run - docked ~3900 Proteins with 5000 Ligands for a total of ~19M docked pairs.
- Submitted via command line to Condor using Pegasus on the OSG-XSEDE submission node
- Infrastructure is set and new runs can be easily started

What does this look like? Executable Wrapper

```
#!/bin/bash
set -e

# Wrapper for Autodock Vina this shell script will import a receptor a ligand and the center
coordinates
# and convert the output to a .mol2 format. Afterward it will remove the unnecessary files so only
the
# .mol2 is returned
# Version 1.2 - Rob Quick - Added file clean up to streamline output

if [ $# -ne 5 ]; then
    echo "Usage: vina  "
    exit 1
fi

chmod a+x vina
#Run vina
./vina --cpu 1 --receptor $1.pdbqt --ligand $2.pdbqt --center_x $3 --center_y $4 --center_z $5 --
size_x 21 --size_y 21 --size_z 21 --out $1-$2.pdbqt
#Run pdbqt to mol2 conversion
./pdbqt2mol2.py $1-$2.pdbqt $2.mol2 $1-$2.mol2
#Get rid of unnecessary files
rm $1-$2.pdbqt
rm $1-$2.pdbqt-new

echo $HOSTNAME
```

Open Science Grid

What does this look like? Condor Submit File

```
Universe = vanilla
Executable = vina_wrapper.sh
Log = vina.log
Error = error/vina.err.$(Process)
should_transfer_files = YES
when_to_transfer_output = ON_EXIT

Output = output/3DLXA1-ZINC00260992.mol2
Arguments = 3DLXA1 ZINC00260992 11.923 26.684 0.672
transfer_input_files = /home/rquick/autodock_vina_1_1_2_linux_x86/bin/vina,/home/yx5/test/
receptors/3DLXA1/3DLXA1.pdbqt,/home/yx5/test/ligands/pdbqt/ZINC00260992.pdbqt,/home/yx5/test/
ligands/mol2/ZINC00260992.mol2,/home/yx5/pdbqt2mol2.py
Queue

Output = output/3DLXA1-ZINC02913596.mol2
Arguments = 3DLXA1 ZINC02913596 11.923 26.684 0.672
transfer_input_files = /home/rquick/autodock_vina_1_1_2_linux_x86/bin/vina,/home/yx5/test/
receptors/3DLXA1/3DLXA1.pdbqt,/home/yx5/test/ligands/pdbqt/ZINC02913596.pdbqt,/home/yx5/test/
ligands/mol2/ZINC02913596.mol2,/home/yx5/pdbqt2mol2.py
Queue

Output = output/3DLXA1-ZINC21637768.mol2
Arguments = 3DLXA1 ZINC21637768 11.923 26.684 0.672
transfer_input_files = /home/rquick/autodock_vina_1_1_2_linux_x86/bin/vina,/home/yx5/test/
receptors/3DLXA1/3DLXA1.pdbqt,/home/yx5/test/ligands/pdbqt/ZINC21637768.pdbqt,/home/yx5/test/
ligands/mol2/ZINC21637768.mol2,/home/yx5/pdbqt2mol2.py
Queue
```

www.biodrugscreen.org

Structural Protein-Ligand Interactome (SPLINTER)

Welcome! This is a service of the laboratory of Samy Meroueh at the Indiana University School of Medicine.

A Computational Resource to Enable Chemical Genomics and Drug Discovery

The portal is constructed around SPLINTER (Structural Protein-Ligand Interactome), an interactome that predicts the interaction of thousands of small organic molecules with thousands of proteins through structure-based molecular docking and scoring. A list of the top-ranking small-molecule ligands is provided for each protein target in SPLINTER along with purchasing information for experimental characterization. The site can also be used to rank proteins for individual active compounds to provide a list of potential targets for experimental validation in cells or *in vivo*.

Start

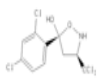
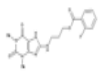
References:

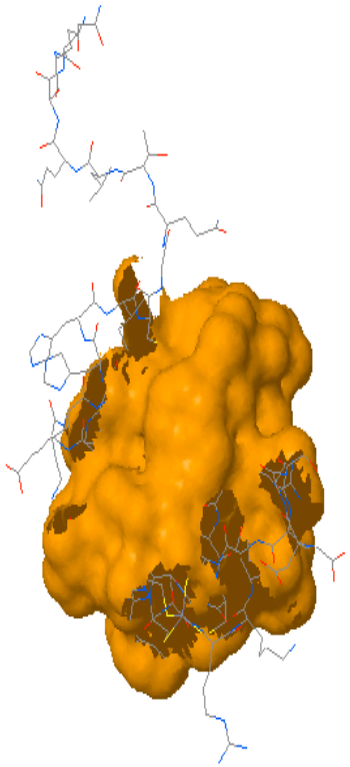
Li, L., J. Li, M. Khanna, I. Jo, J.P. Baird, and S.O. Meroueh**, *Docking Small Molecules to Predicted Off-Targets of the Cancer Drug Erlotinib Leads to Inhibitors of Lung Cancer Cell Proliferation with Suitable In vitro Pharmacokinetic Properties.* ACS Med. Chem. Lett. 2010; 1(5):229-233.

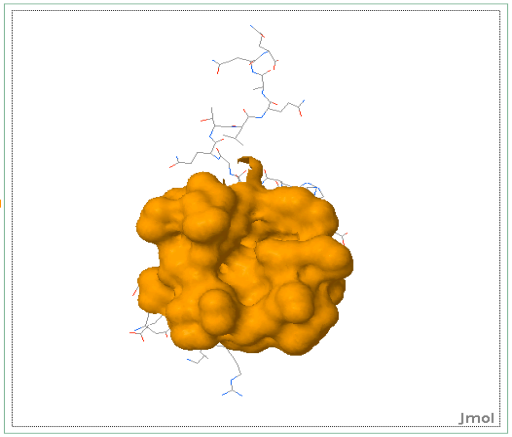
Li, L., K. Bum-Erdene, P.H. Baenziger, J.J. Rosen, J.R. Hemmert, J.A. Nellis, M.E. Pierce, and S.O. Meroueh**, *BioDrugScreen: a computational drug design resource for ranking molecules docked to the human proteome.* Nucleic Acids Res. 2010; 38(Database issue):D765-73.

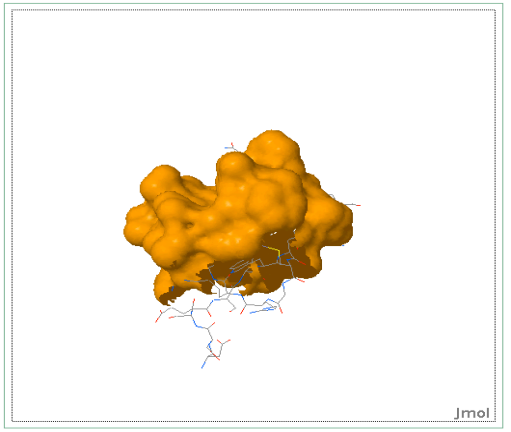
Open Science Grid

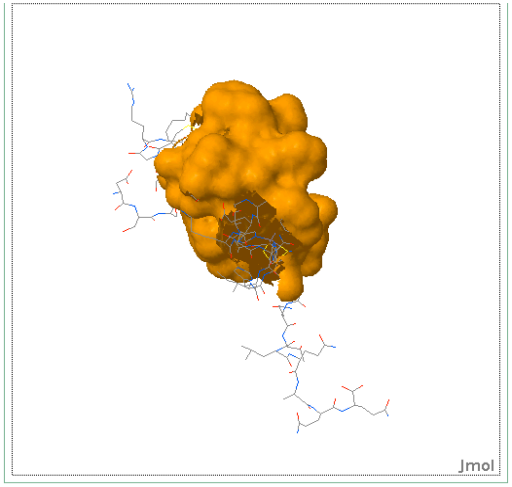
Various rotations of Protein CBFA2T1 (Cyclin-D-related protein) (Eight twenty one protein) (Protein ETO) (Protein MTG8) (Zinc finger MYND domain-containing protein 2)

Target	Ligand Image	Rank
ZINC27470710 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		1
ZINC01228697 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		2
ZINC04741379 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		3
ZINC09611213 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		4
ZINC02810311 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		5
ZINC02945941 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		6
ZINC00404256 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		7
ZINC20038318 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		8
ZINC04039256 VIEW POSE MOL2 ORDER INFO OTHER TARGETS		9



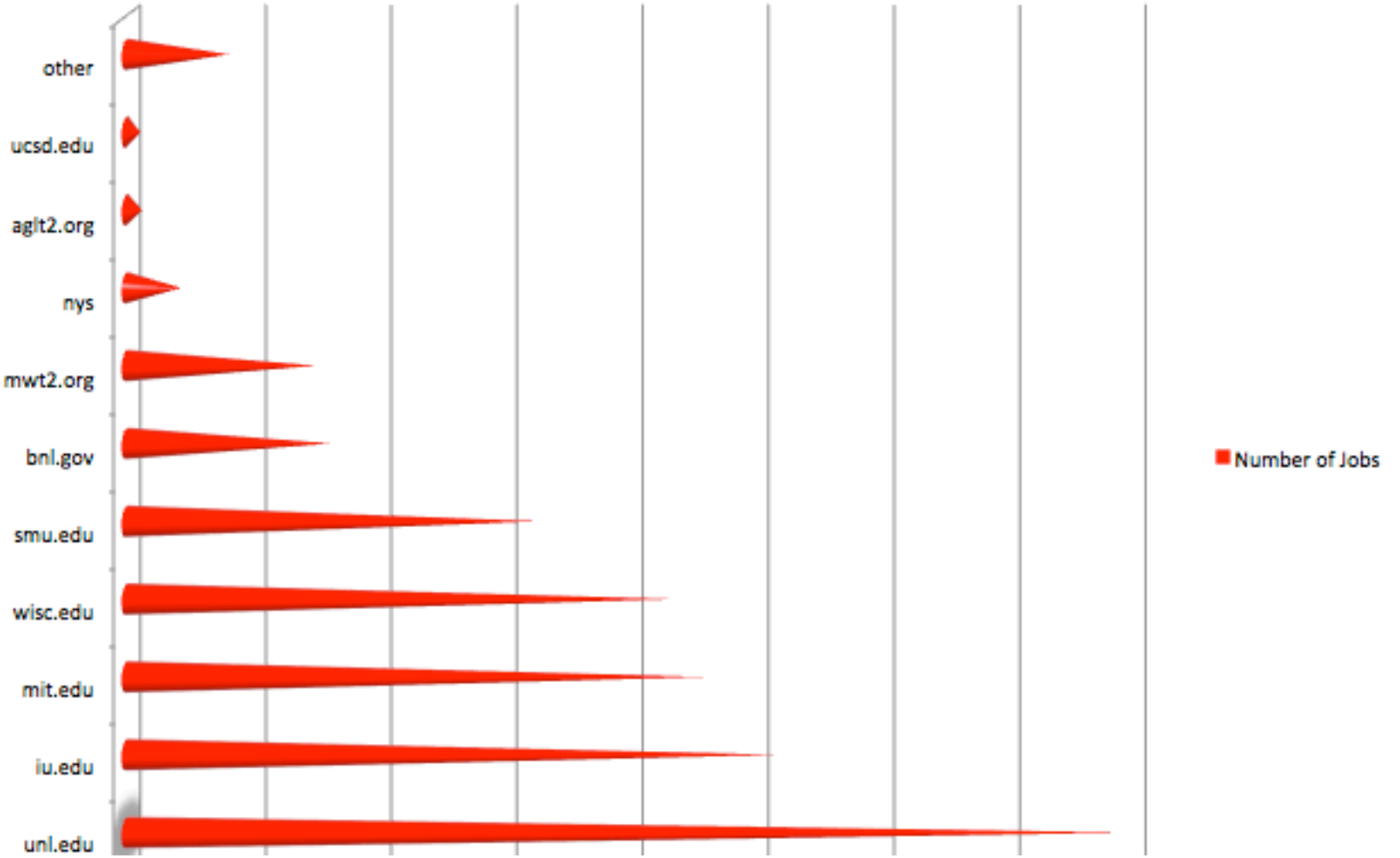




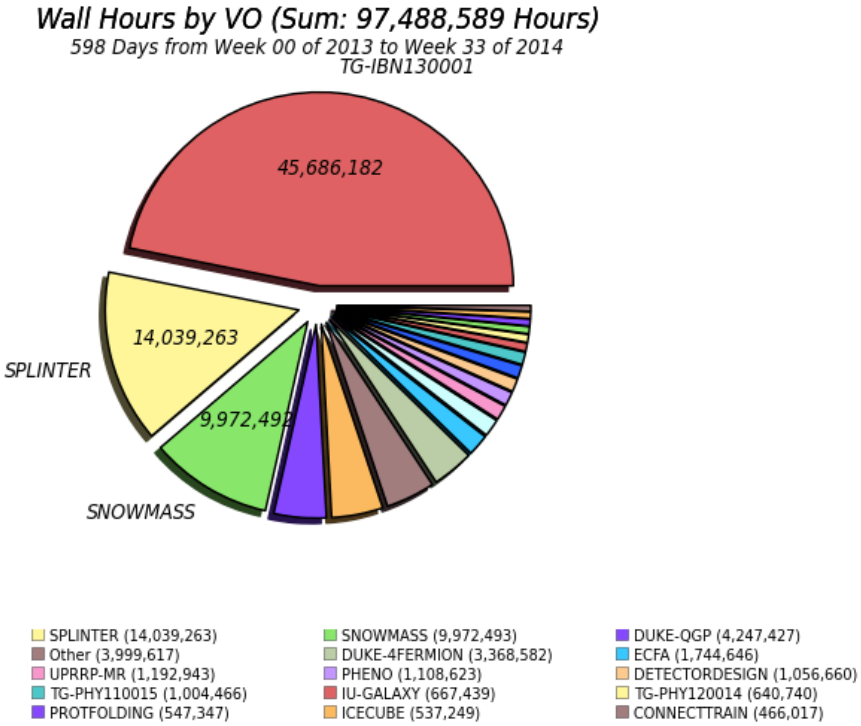


Open Science Grid

Where did these jobs run? First ~1M Jobs



Some Numbers



- Amazon S3 Computing $0.073/hour
- $1.025M Compute Only
- Data Transfer and Storage Not Included

Questions