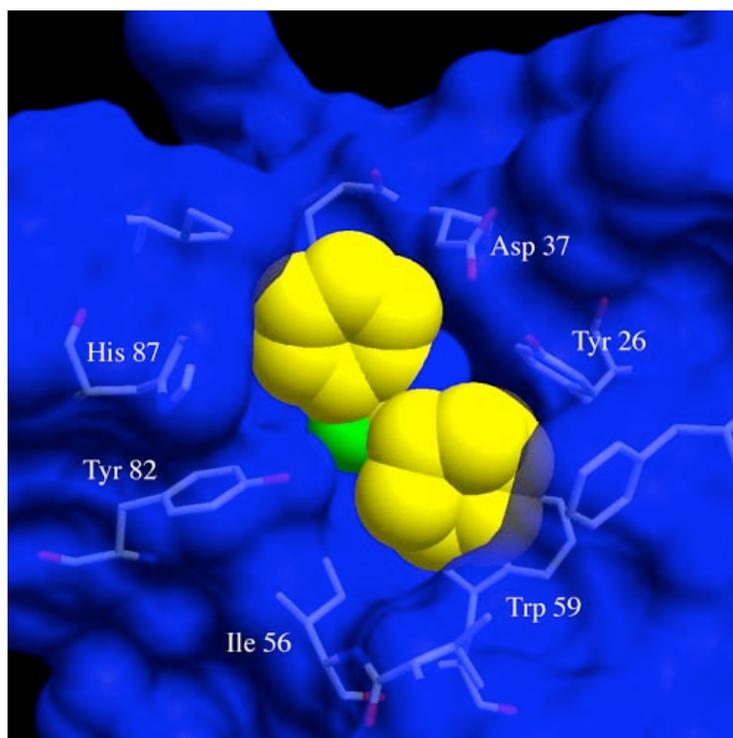


---

# An In Silico Drug Screening Pipeline on BlueGene/L



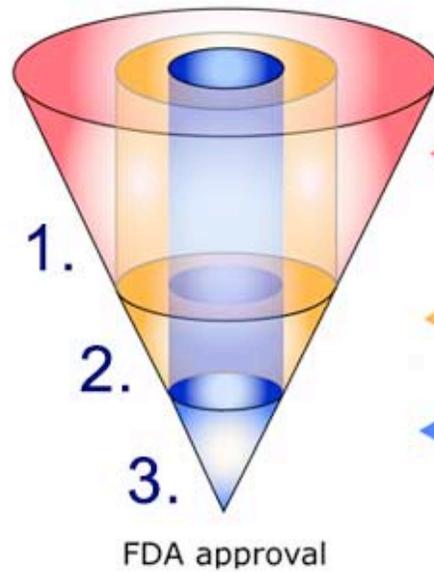
Fangfang Xia

The University of Chicago

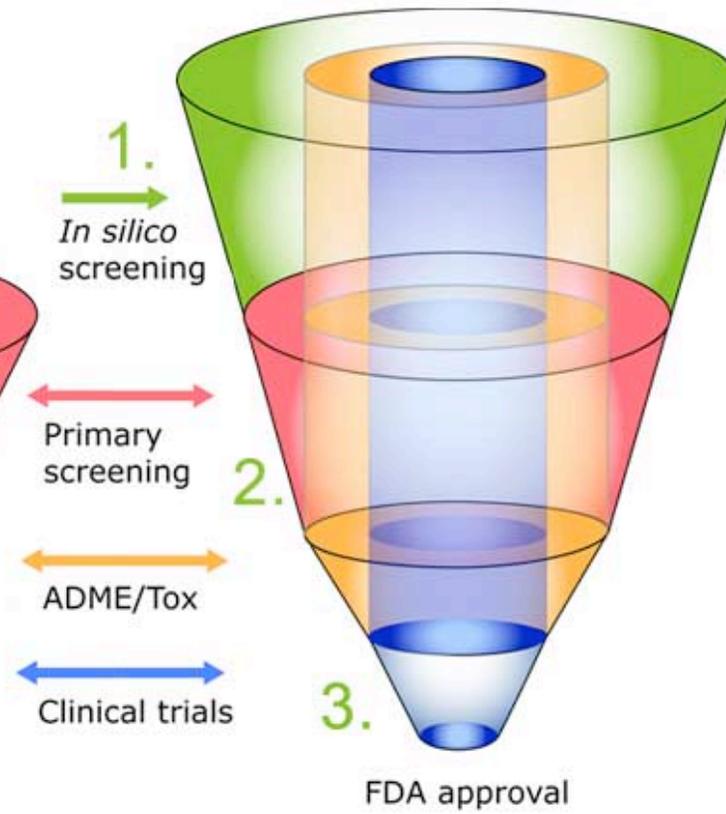
---

# In silico Screening

Classic Drug  
Discovery Funnel



Optimised Drug  
Discovery Funnel



---

# Participants



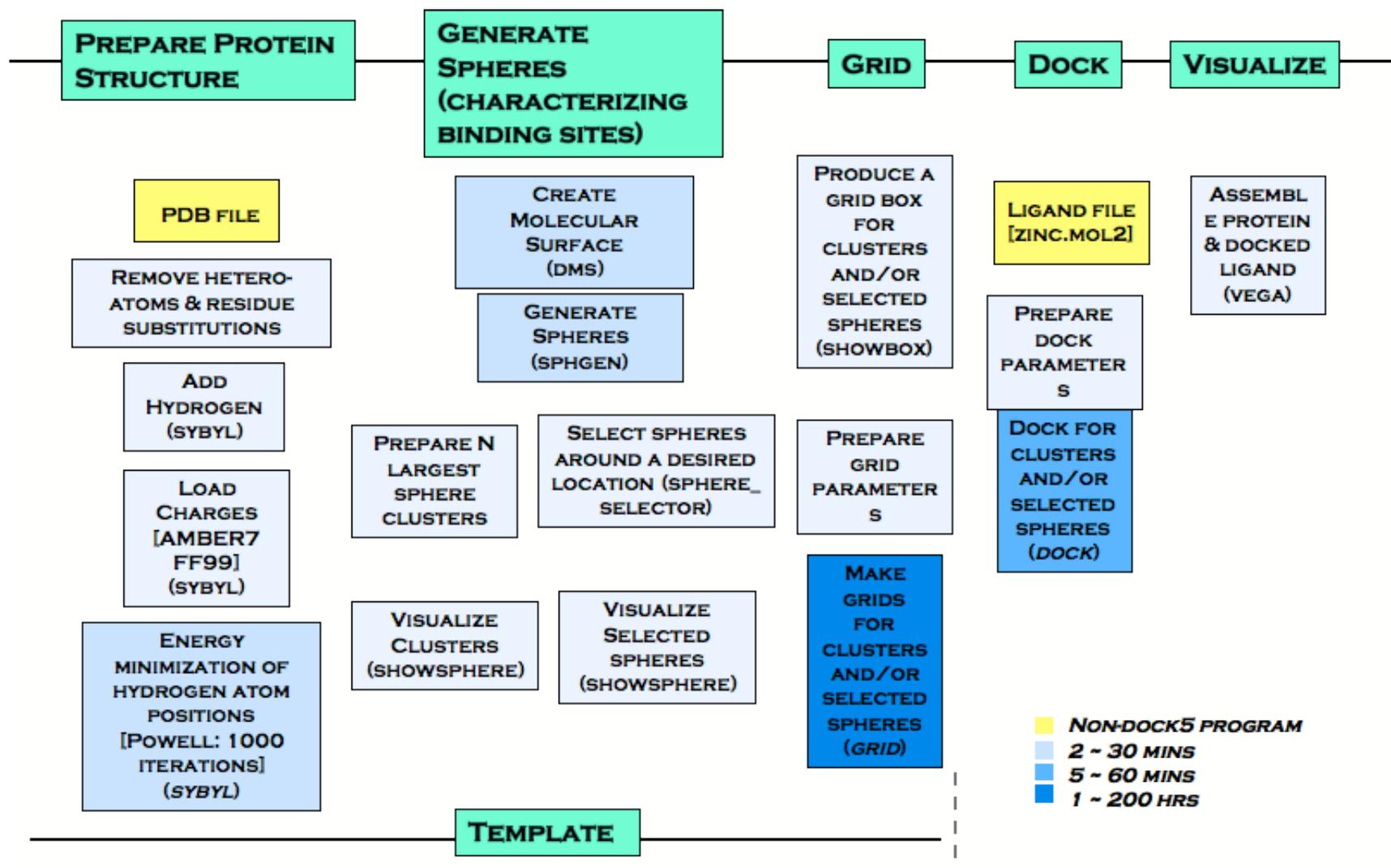
- Rick Stevens (co-PI), Ross Overbeek (co-PI)
  - Matt Cohoon, Mike Kubal, Kaitlyn Hwang, Jenifer Zinner, Gohar Margaryan, Mindy Shi, Fangfang Xia
-

---

# Gene-to-Screen Workflow

1. IDENTIFY GENES/PROTEIN TARGETS FROM THE LITERATURE
    - KNOWN ANTIBIOTIC TARGETS AND KNOWN ANTIBIOTIC RESISTANCE FACTORS
    - ESSENTIAL GENES AND VIRULENCE ASSOCIATED FACTORS
  2. DETERMINE FUNCTIONAL ROLE OF EACH TARGET
    - COMPARATIVE ANALYSIS (CLUSTER ANALYSIS, SUBSYSTEM RECONSTRUCTION)
  3. SEARCH FOR KNOWN INHIBITORS OF EACH TARGET
    - LITERATURE MINING AND COMPUTATIONAL SCREENING (SEE STEP 8)
  4. DETERMINE DEGREE OF CONSERVATION ACROSS SPECIES
    - PHYLOGENY AND SEQUENCE ALIGNMENT
    - CHARACTERIZATION OF THE ACTIVE SITE
    - EARLY SCREEN IN HUMAN AND MODEL SYSTEMS
  5. DETERMINE STRUCTURE OF EACH TARGET (PDB, COMPUTATION)
    - DATABASE SEARCH/SIMILARITY AND STRUCTURAL MODELING
  6. DETERMINE ACTIVE SITE OF EACH TARGET
    - COMPUTATIONAL ANALYSIS OF EACH STRUCTURE
  7. DETERMINE DRUGABILITY OF EACH TARGET
    - SIZE OF POCKET, NUMBER OF POCKETS
  8. SCREENING OF COMPOUNDS FOR BINDING AFFINITY ETC.
    - COMPUTATIONAL AND HIGH-THROUGHPUT EXPERIMENTS
  9. TOXICITY SCREENING IN HUMAN AND MODEL SYSTEMS
    - COMPUTATIONAL AND HIGH-THROUGHPUT EXPERIMENTS
-

# DOCK5 pipeline



---

# Parallel Programming

- Ported DOCK5 (UCSF) and AutoDock3 (Scripps) to BG/L
  - Trivially parallelizable: Master/Slave w/ MPI
  - I/O patterns:
    - Master: parses one ligand database file and passes individual ligand information to slaves
    - Slave: does actual “docking” and generates a small output file with a docking score and optimal molecular configurations
-

---

# Visualization and Analysis

- Top ligands
  - Compare docking score and configuration w/ those of natural ligands
  - Look into very interesting results w/ protein visualization tools such as Protein Explorer
-

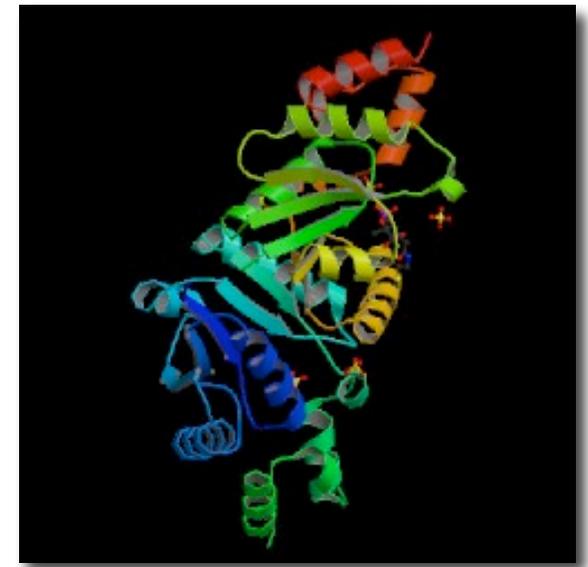
# Visualization and Analysis

Top 10000 comparison

http://ci.uchicago.edu/~fangfang/score/comp10kc.html

KEGG DAS Database distribution Automator Pr...ipt Actions Mayberrys ClieSource MCS Mac Software Google http://yersi...bDetails.cgi

ZINC ID	DOCK5 Energy	AutoDock Total Energy	AutoDock Inter-molecular Energy	Percentile (total)	Percentile (inter)
ZINC00167160	-56.060829	-9.410	-10.710	48.317%	48.055%
ZINC00236564	<b>-53.856201</b>	<b>-12.909</b>	<b>-10.967</b>	<b>96.374%</b>	<b>56.454%</b>
ZINC00073677	-53.075100	-3.404	-11.351	0.953%	68.388%
ZINC00286225	<b>-52.884354</b>	<b>-7.601</b>	<b>-13.008</b>	<b>21.548%</b>	<b>96.118%</b>
ZINC00226481	-52.646568	-7.673	-11.517	22.433%	73.067%
ZINC00082754	<b>-52.642052</b>	<b>-12.306</b>	<b>-11.462</b>	<b>92.362%</b>	<b>71.591%</b>
ZINC00284234	-52.624825	-7.355	-11.258	18.745%	65.755%
ZINC00167151	-52.552624	-8.136	-11.454	28.338%	71.364%
ZINC00034872	<b>-52.310295</b>	<b>-12.067</b>	<b>-11.537</b>	<b>90.237%</b>	<b>73.610%</b>
ZINC00289786	-51.856140	-6.168	-10.769	9.182%	50.033%
ZINC00267336	-51.214191	-8.739	-10.683	37.172%	47.145%
ZINC00308879	<b>-51.049255</b>	<b>-13.054</b>	<b>-12.040</b>	<b>96.993%</b>	<b>84.856%</b>
ZINC00243985	-50.957027	-3.907	-11.707	1.521%	77.883%
ZINC00087313	<b>-50.888405</b>	<b>-12.551</b>	<b>-12.204</b>	<b>94.295%</b>	<b>87.675%</b>
ZINC00289831	-50.819489	-5.612	-11.368	6.287%	68.889%
ZINC00308877	<b>-50.758175</b>	<b>-13.631</b>	<b>-13.000</b>	<b>98.716%</b>	<b>96.072%</b>
ZINC00291545	-50.735245	-9.276	-10.884	46.038%	53.776%
ZINC00291546	-50.590950	-8.326	-10.372	30.954%	37.054%
ZINC00180268	-50.178215	-2.994	-10.295	0.615%	34.712%
ZINC00142321	-50.074051	-10.384	-10.619	65.518%	44.965%
ZINC00074376	-49.973858	-10.633	-9.643	69.924%	17.920%
ZINC00309571	<b>-49.882210</b>	<b>-12.472</b>	<b>-12.246</b>	<b>93.722%</b>	<b>88.340%</b>
ZINC00142326	<b>-49.782463</b>	<b>-12.648</b>	<b>-12.403</b>	<b>94.923%</b>	<b>90.558%</b>
ZINC00115857	-49.771091	-11.346	-11.104	81.299%	60.952%
ZINC00267725	-49.737999	-7.621	-12.217	21.791%	87.903%



CoaD

---

# Status and Scalability

## ■ DOCK5

- ❑ Scales poorly beyond 256 processes
- ❑ Currently run with 32/64 processes
- ❑ Memory footprint often exceeds 512M
- ❑ Templates have to be prepared on big-endian machines

## ■ AutoDock3

- ❑ Scales well to 1024 processes
  - ❑ Runs in VN mode on BG/L
-

---

# Roadmap

- Current throughput
    - DOCK5: 10,000 successful screens per BlueGene/L rack (1024 nodes) per day
    - AutoDock3: 50,000 screens per rack per day
  - Searching for new antibiotics
    - 300 essential-gene-products x 3.3 million compounds  $\Rightarrow$  990 million drug docking computations (each one involves about 20 different computations)  $\Rightarrow$  over 10 billion jobs
  - Adding a neural network component to filter unpromising ligands based on high level features
-