

Drug Design Data Resource

*Leveraging Protein-Ligand Datasets for
Developing Better
Computer-Aided Drug Design Methods*

Vicki Feher, Rommie Amaro, Mike Gilson, UCSD

Stephen Burley, Rutgers & UCSD



Talk Overview

D3R's goals and synergies

2014 – 2016 Accomplishments

The 1st D3R Challenge & Workshop – What did we learn?

Plans



NIH-U01 Resource, Unique Purpose

5 year, \$3.7 million grant – R. Amaro, V. Feher, M. Gilson

PRIMARY GOAL:

Collect previously unpublished protein-ligand datasets and run community wide blinded challenges as benchmarks for better CADD algorithms

Synergy with RCSB PDB and other public databases

Co-award: S. Burley

Synergy with SAMPL and CASP/CAMEO

Synergy with Pharma and Academia

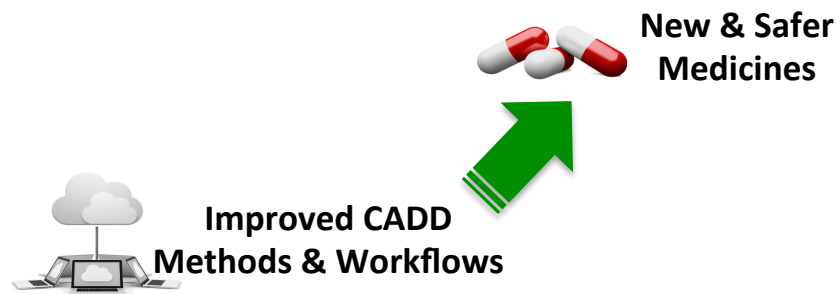
More predictive CADD methods benefit everyone!



D3R's Role



Unpublished
Datasets



Blinded
Challenges



ChEMBL

RCSB PDB
PROTEIN DATA BANK

PubChem



Protein-Ligand Datasets

overview of desirable features

20 to 100+ compounds with affinity data, including some known inactives

Five or more crystal structures with resolution $< 2.6\text{\AA}$, coordinates & mtz files

Experimental conditions, methods, uncertainties





Our first 2 years



Datasets Collected

Contributor	# of datasets	Curation
Abbvie - CSAR	1	Completed by CSAR
Genentech	1.5	Not needed
Roche	2	Completed by D3R
Baker/Stoddard	2	Completed by D3R
SGC - Oxford	1	Completed by D3R



Web Portal for Data, Workflows, Challenges & Community Activities

An Open Resource to Advance Computer-Aided Drug Design

Advancing the technology of computer-aided drug discovery through the interchange of high quality protein-ligand datasets, workflows and community-wide blind data challenges.

D3R Provides



read more >



CADD Datasets

D3R will make datasets available to the community.

Community Challenges

D3R will engage the community through challenges.

About

Welcome to the Drug Design Data Resource Community. D3R is funded in part by NIH grant 1U01GM111528 from the National Institute of General Medical Sciences.

Recent



Challenges

Home / CHALLENGES

D3R Grand Challenge 2015

Home /

Start Date: Sep 15, 2015

End Date: Feb 02, 2016

Welcome to D3R's inaugural Grand Challenge!

The Grand Challenges provide blinded unpublished datasets containing high quality crystal structures and binding affinity or potency data for testing and improving ligand-protein docking algorithms and their scoring protocols.

Each dataset is curated and embellished to challenge the methods, and their users, in a particular aspect of known docking protocol shortcomings.

Dataset 1

The HSP90 dataset, kindly donated by Abbvie, was expanded and curated by the CSAR group at University of Michigan. The ATP site of HSP90 has been the subject of many oncology drug discovery programs over the past 15 years and consequently has a large representation in the PDB and literature. However, the prevalence of water-mediated ligand-protein interactions and a ligand binding site that accesses multiple open and closed pocket conformations can make estimation of docking pose and ranked affinity a challenge.

This dataset has 8 crystal structures with resolution < 2.0Å, binding data for 180 compounds across five orders of magnitude and three chemical series, and over 50 inactive. The challenge has two stages, as follows:

Stage 1

Challenge: Predict the crystallographic poses of 8 ligands spanning all three chemical series, and predict affinities, or affinity rankings, for these ligands, and also for the other 174 ligands.

Provided Inputs: A) Protein-ligand co-crystal structures, 4 drawn from the PDB and 2 from the blinded dataset, that were solved with compounds of each chemical class in the series and prepared with hydrogens added. B) SMILES strings of the 8 ligands to be docked, of the 2 ligands in the newly revealed co-crystal structures, and of the additional 172 compounds for affinity prediction or ranking. C) benchmark IC50 values for relevant input structures. Note: There are Abbott publications for two of the chemical series of this Challenge containing crystal structures, SAR and related IC50s.

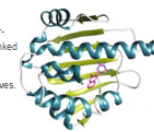
Outputs: A) Your predicted poses for the 8 ligands, in a coordinate system aligned with those provided in the Inputs. B) Your predicted affinities, or affinity rankings, for all 180 compounds. When Stage 1 closes, we will release the crystallographic poses of the 8 ligands.

Stage 2

Challenge: Predict the affinities, or affinity rankings, of all 180 ligands.

Inputs: Same as for Stage 1, supplemented by the co-crystal structures. **Correction: IC50s will be released once Stage 2 is over.**

Outputs: Your predictions of the affinities, or affinity rankings of all 180 compounds. NB. You are free to use additional public-domain protein structures and scientific literature to help you make your predictions in both stages. For example, if you prefer to dock the ligands into a different structure from the PDB, this is fine, so long as the structures you submit as your predictions are rotated and translated so they superimpose on the structures we provided.



Join the Challenge

The challenge has closed, but you can still...

Download the Data

HSP90

Stage 1 (09/15/2015 to 11/20/2015)

Please login first

Stage 2 (11/20/2015 to 02/01/2016)

Please login first

MAP4k4

Stage 1 (10/16/2015 to 12/16/2015)

Please login first

Stage 2 (12/17/2015 to 02/01/2016)

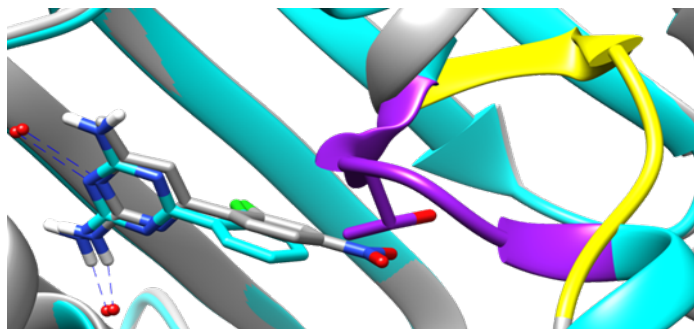
Please login first

Rules

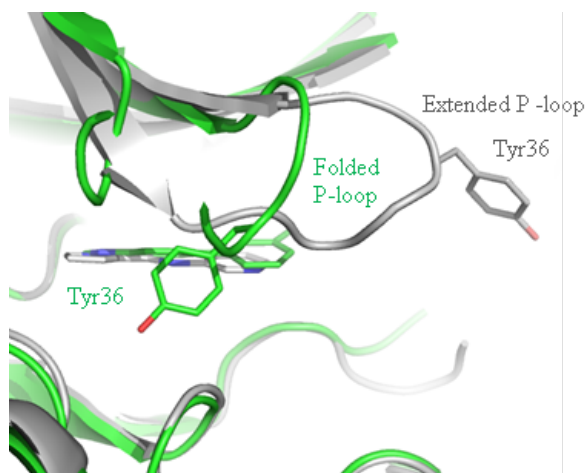
Read the Rules and Procedures

www.drugdesigndata.org

Grand Challenge 2015



HSP90 Dataset from Abbvie/CSAR



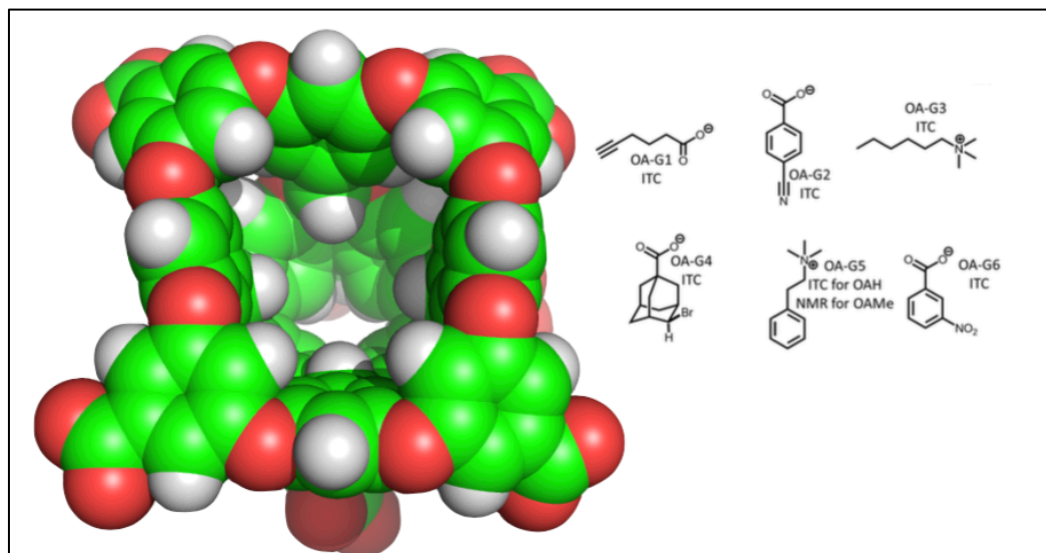
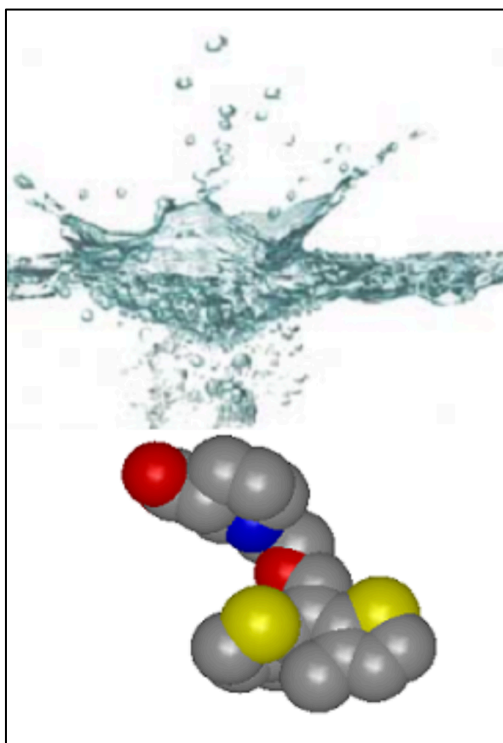
MAP4K4 Dataset from Genentech



SAMPL Challenge 2015

Model System Challenges

partition/distribution coefficients: ~50 sm. molecule cyclohexane/water
host-guest affinities: CBClip, OAH, OAMe



David Mobley, John Chodera, Mike Gilson, Genentech



1st Workshop March 9 – 11, 2016



Joint workshop for D3R and SAMPL participants
84 participants from academia & pharma

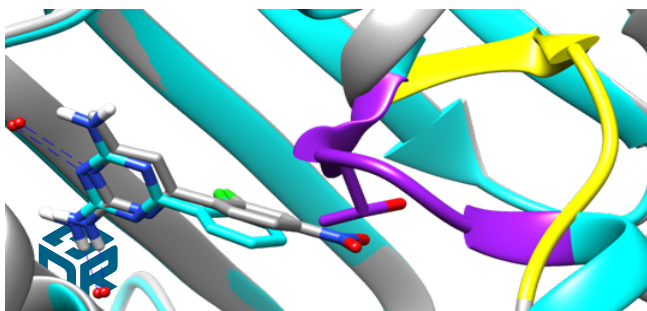
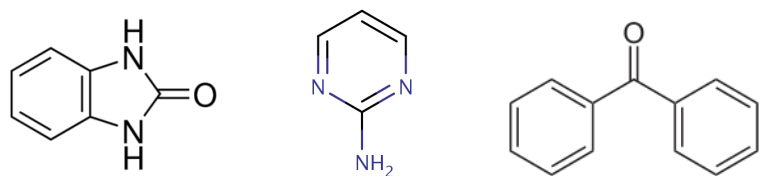


Grand Challenge 2015: Challenge Design

HSP90 Dataset from Abbvie/CSAR

180 IC50s; 5.2nM – 50uM,
3 chemical series, 33 inactives
8 blinded crystal structures, R = 1.6 -1.96Å

>200 crystal structures in the PDB
Water-mediated interactions
3 pocket conformations

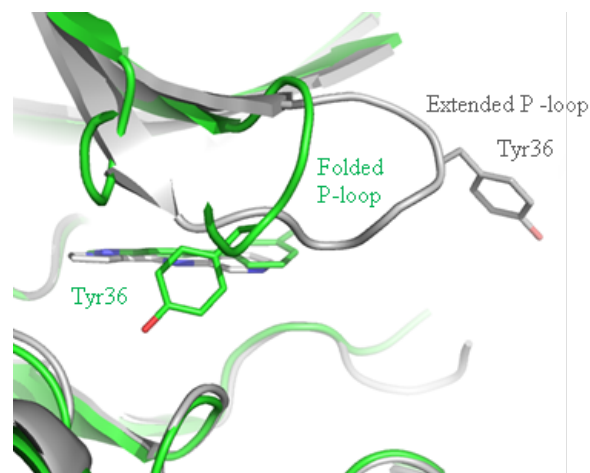


MAP4K4 Dataset from Genentech

18 IC50s; 3.1nM – 16.7uM,
Diverse chemical series
32 blinded crystal structures, R = 1.59 -3.04Å

5 crystal structures in the PDB

Two P-loop conformations



Grand Challenge 2015: Challenge Design & Participation

2 stages: Stage 1 – predict pose; submit rank
 Stage 2 – given poses, predict ranks

Use any public data available

Released 2 HSP90 structures with example chemotypes and hinted about water

3 small HSP90 subsets (n= 3, 4, 10) for alchemical free energy predictions

Challenge Component	N _{submit}
HSP90 Pose Predictions	47
MAP4K4 Pose Predictions	33
HSP90 Stage 1 Affinity Ranking	75 (41)
MAP4K4 Stage 1 Affinity Ranking	77 (40)
HSP90 Stage 2 Affinity Ranking	59 (30)
MAP4K4 Stage 2 Affinity Ranking	46 (26)
HSP90 Small Set Free Energies	18



Lessons

Successful prediction of docking poses depends on

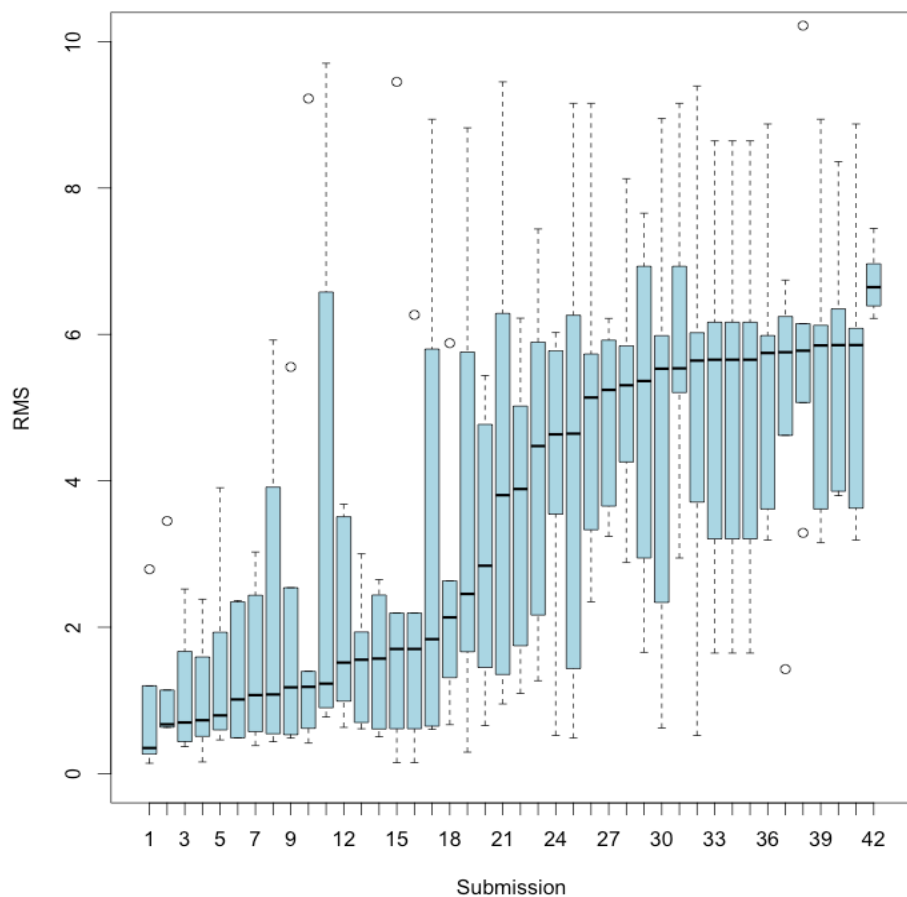
- the entire workflow**
- the amount of prior knowledge**

Ranking (affinity scoring) remains challenging even with known poses



HSP90 Pose Prediction Results

HSP90 Rank 1



P+L predictions = 5

Docking Programs for <math><2.0\text{\AA}</math> RMSD:

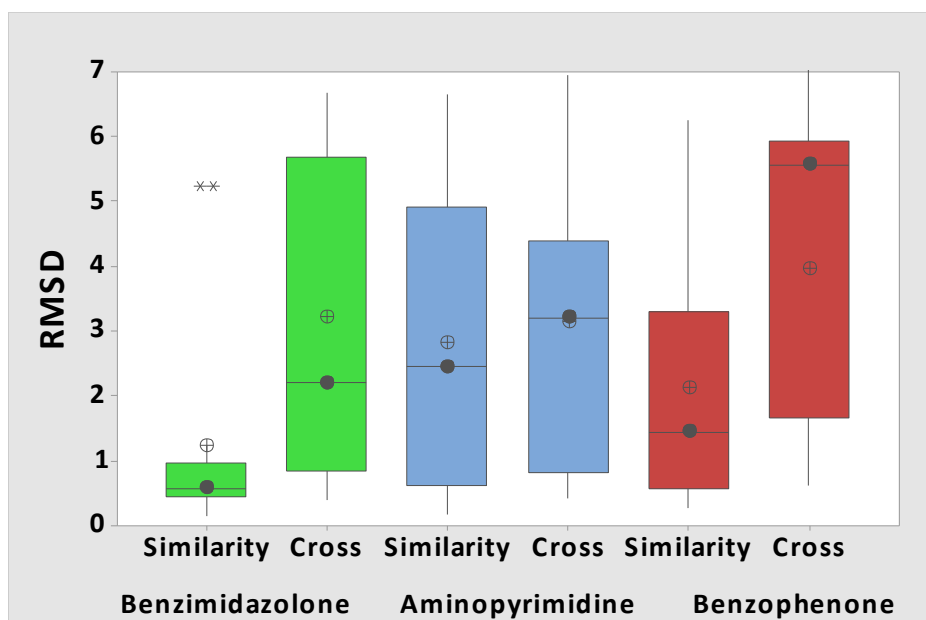
AutoDock/Vina	rDock
Glide	RosettaLigand
Gold	Smina
Surflex	

Docking Programs for >math>2.0\text{\AA}</math> RMSD:

AutoDock/Vina	rDock
Glide	RosettaLigand
Gold	Smina
MedusaDock	



Receptor Selection Impacted Success



N = 5

Knowledge-based receptor selection
shape/chemotype matches in PDB
GRIM rescoring

Included key waters

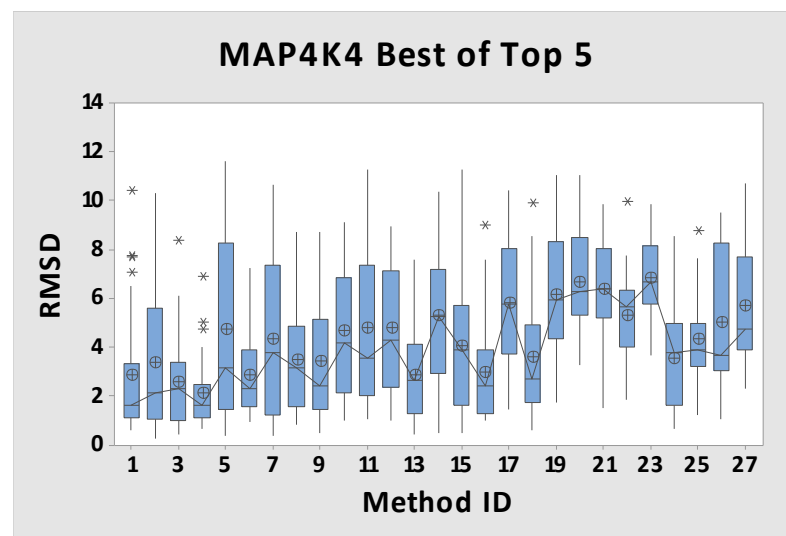
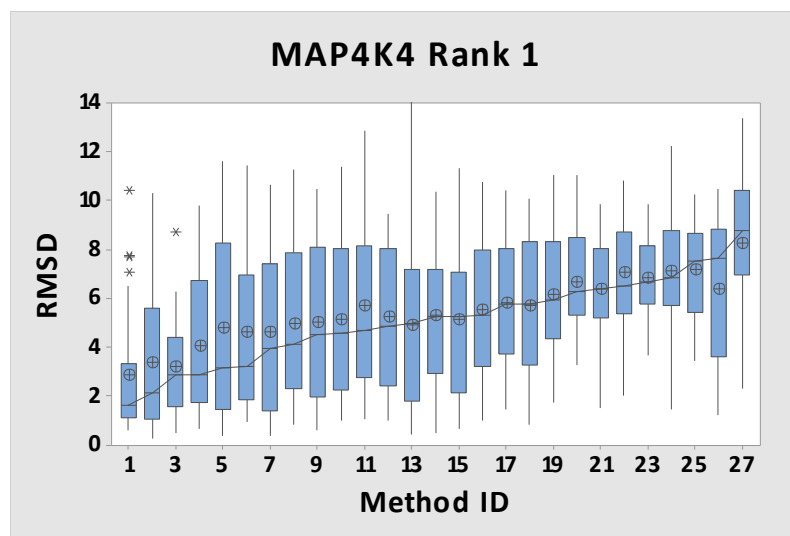
Visual inspection



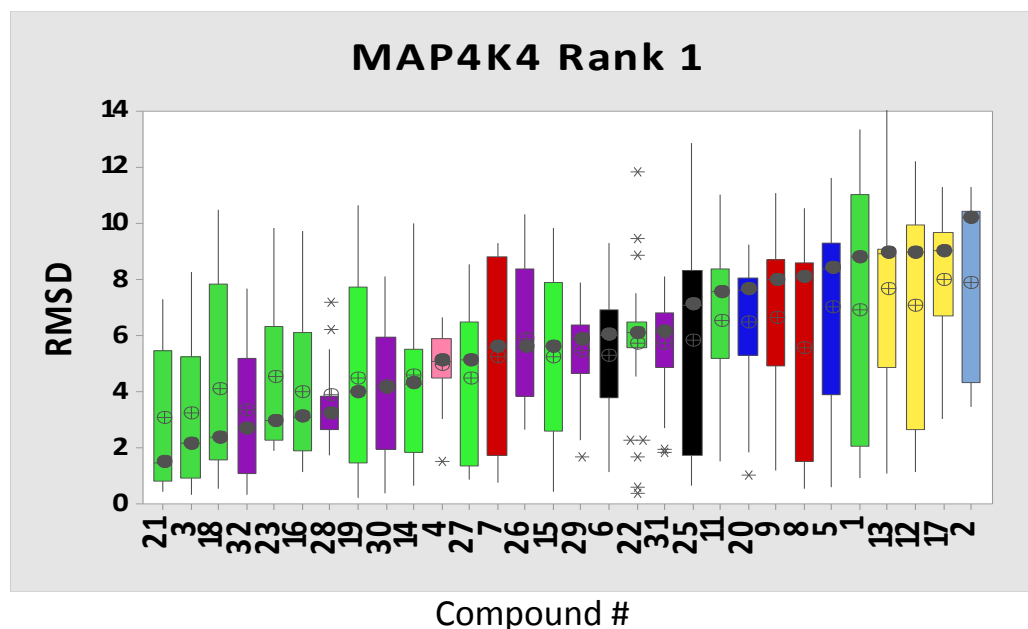
MAP4K4 Pose Prediction Results

Participants were allowed to submit up to 5 top poses per ligand-receptor pair

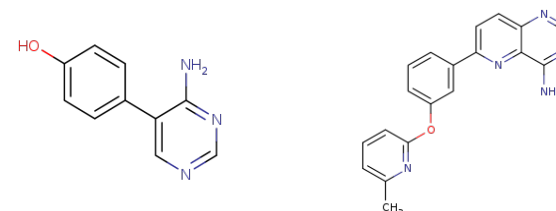
P+L predictions = 32



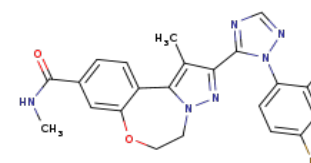
MAP4K4 Pose Results per Compound



Green = aminopyrimidine/aminopyrimidine



Yellow = benoxepin



Success does not correlate with MW (fragment vs. compound), logP, #rotamers, tautomers, ligand affinity, open/closed P-loop, or xtal resolution.

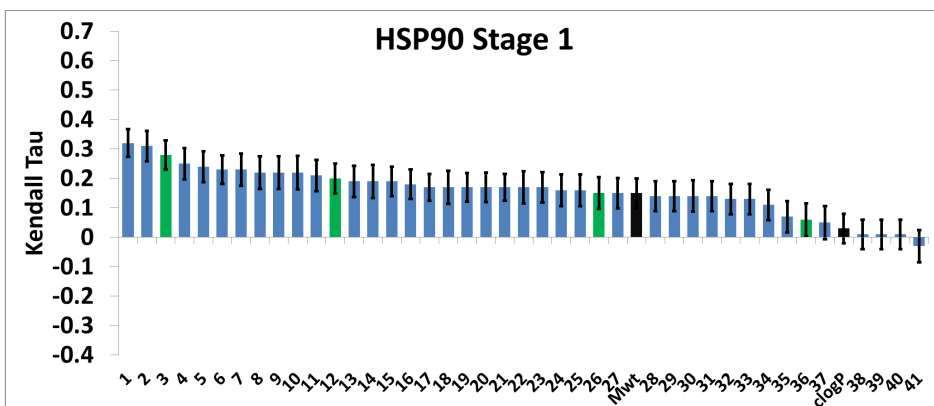
~ correlation to prior chemotype examples



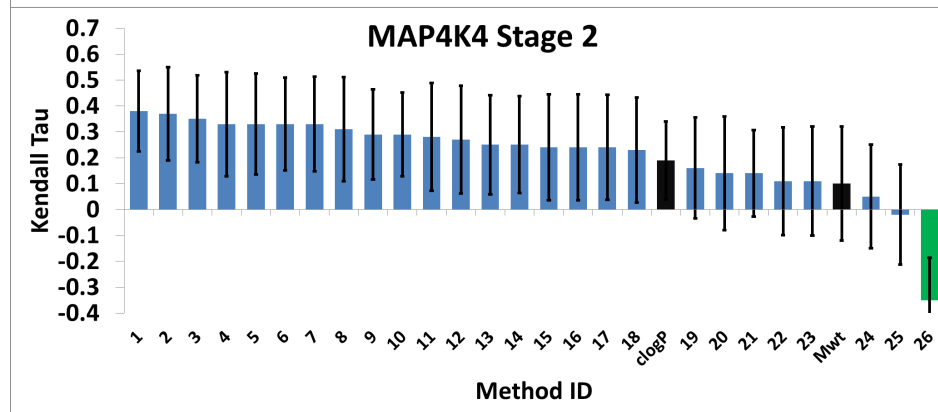
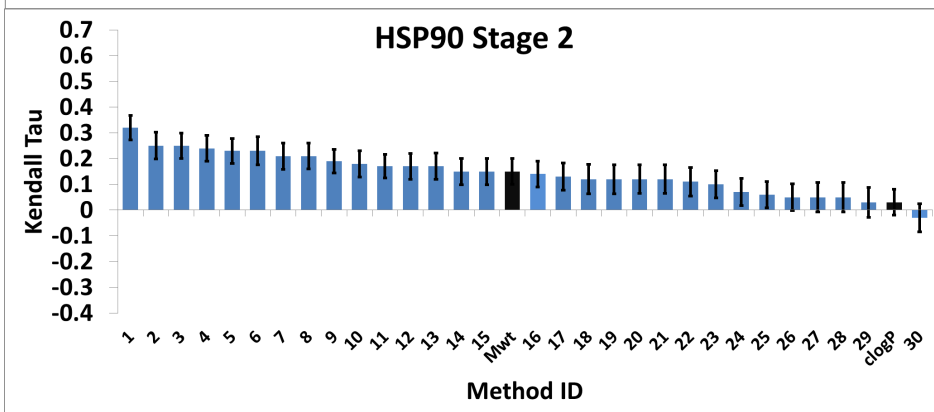
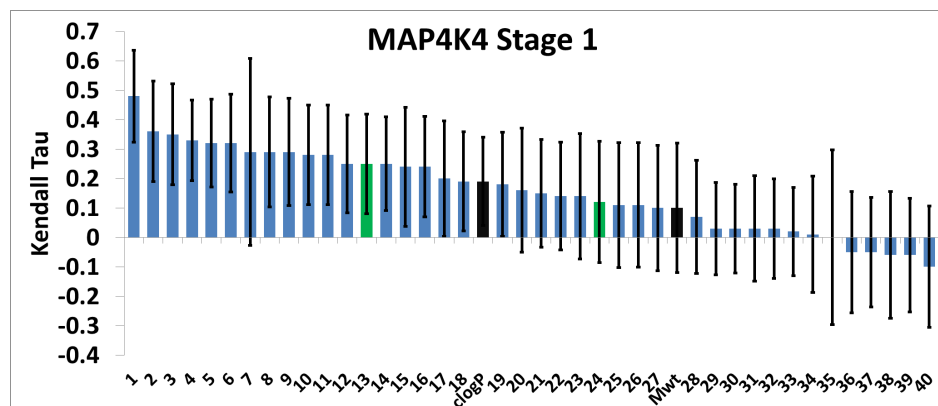
Potency Ranking



N = 179



N = 18



Methods

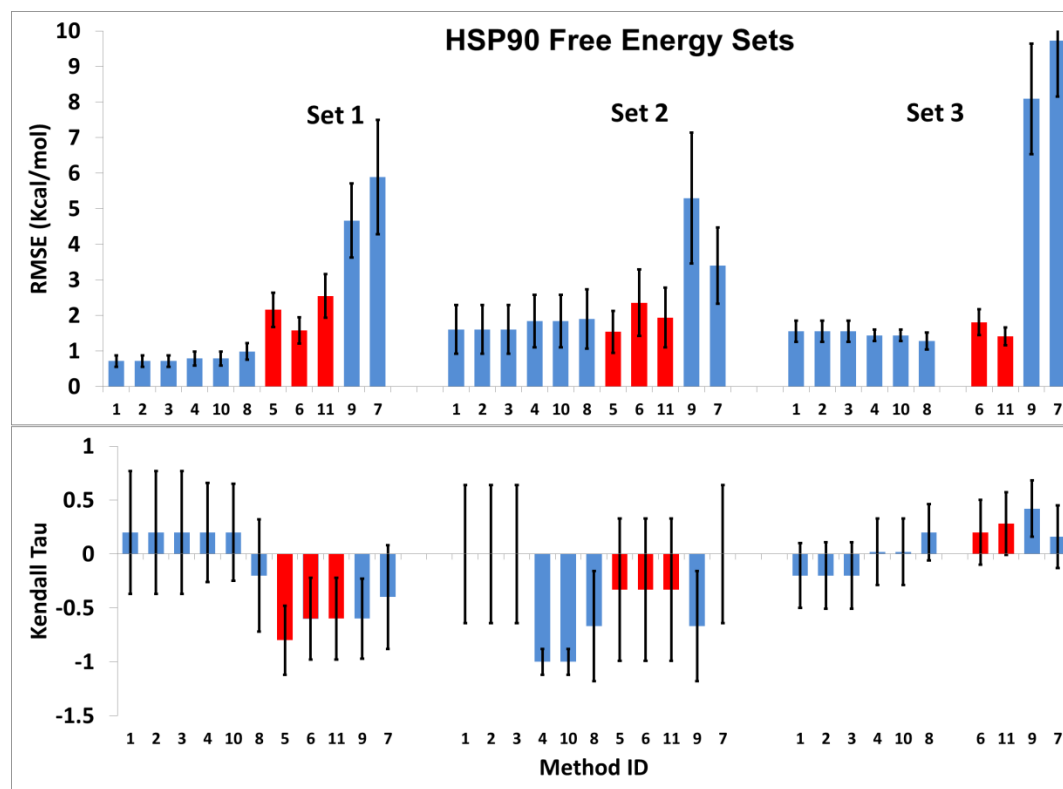
PLANTS, IT-Score, YADA, QSAR, Ensemble-QSAR, BEDAM, LASSO, GoldScore, CDOCKER

MOE_GBVI_WSA, MM-GBSA, RF-Score, Human Brain, CNNScore, QSite-Prime, Rosetta-Ligand,

MMPBSA, KECSA, ChemScore-PLP, GRIM, Pliers-MMGBSA, SMINA



Alchemical Free Energy Methods



N=5

N=4

N=10

Combinations of AM1BCC/GAFF, Merz-Kollman ESP/GAFF, AMBERff14SB, AMBER14SB-ildn, TIP3P



Lessons

Successful prediction of docking poses depends on

- the entire workflow**
- the amount of prior knowledge**

Ranking (affinity scoring) remains challenging even with known poses



Future Challenge Needs

Larger datasets are needed for statistical significance

Collect more detailed methods information (donors & participants)

Kd vs. IC50

Datasets & challenges designed for specific questions

More metrics



What's next:

JCAMD Special Issue: Challenge 2015 results

Challenge 2016-2: September 19th

Farnesoid X Receptor (FXR)

36 co-crystal structures, 102 affinity values, multiple chemotypes

FEP/Alchemical Methods Dataset: 2017?

More SAMPL

More P-L Datasets

Contributor	# of datasets	Completed agreement
Novartis	1	yes
Janssen	1	yes
GSK	1	yes
SGC-UNC	1	yes
Takeda	2-3	pending
Genentech	1-2	pending
Vertex	1	pending
Merck	1	pending



More Result Details



Tonight: Poster #260

D3R 2015 and 2016 challenges: Evaluation of predictions for the grand and mini challenges

Symon Gathiaka

D3R Team



Rommie Amaro



Mike Gilson



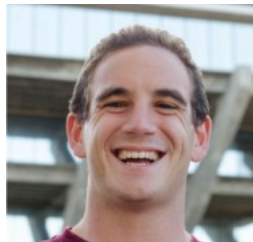
Stephen Burley



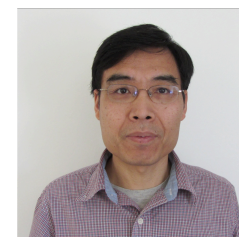
Symon Gathiaka



Shuai Liu



Jeff Wagner



Huanwang Yang



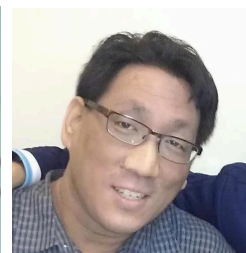
Jasmine Young



Chris Churas



Jeff Grethe



Mike Chiu



Pat Walters
Challenge Evaluator
at Relay



UC San Diego

Thanks to:

Challenge Participants!

Cathy Peishoff

CSAR & D3R dataset contributors: GSK, Vertex, Abbvie, Baker group, Genentech (Seth Harris), Roche, Novartis, Janssen

CSAR: Heather Carlson, Jim Dunbar, Jeanne Stuckey at Univ. of Michigan

Open Eye, CCDC, Schrödinger, CCG

NIH U01GM111528

JCAMD, Terry Stouch



Mike Gilson is founder and has equity interest in VeraChem, LLC.
Rommie Amaro is co-founder and has equity interest in Actavalon, Inc.
Vicki Feher has equity interest in Actavalon, Inc.

Ways to get involved

Participate in a Challenge

Next one starts September 19, 2016

Participate in CELPP

Donate a dataset

Attend a Workshop (Webinar in early 2017)

drugdesigndata@gmail.com

www.drugdesigndata.org

