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Collect previously unpublished protein-ligand datasets and run community wide blinded challenges as benchmarks for better CADD algorithms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Unpublished Datasets

Improved CADD Methods & Workflows

New & Safer Medicines

Workforce Training, Education

Blinded Challenges

CADD Community

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 < 2.6Å, coordinates & mtz files

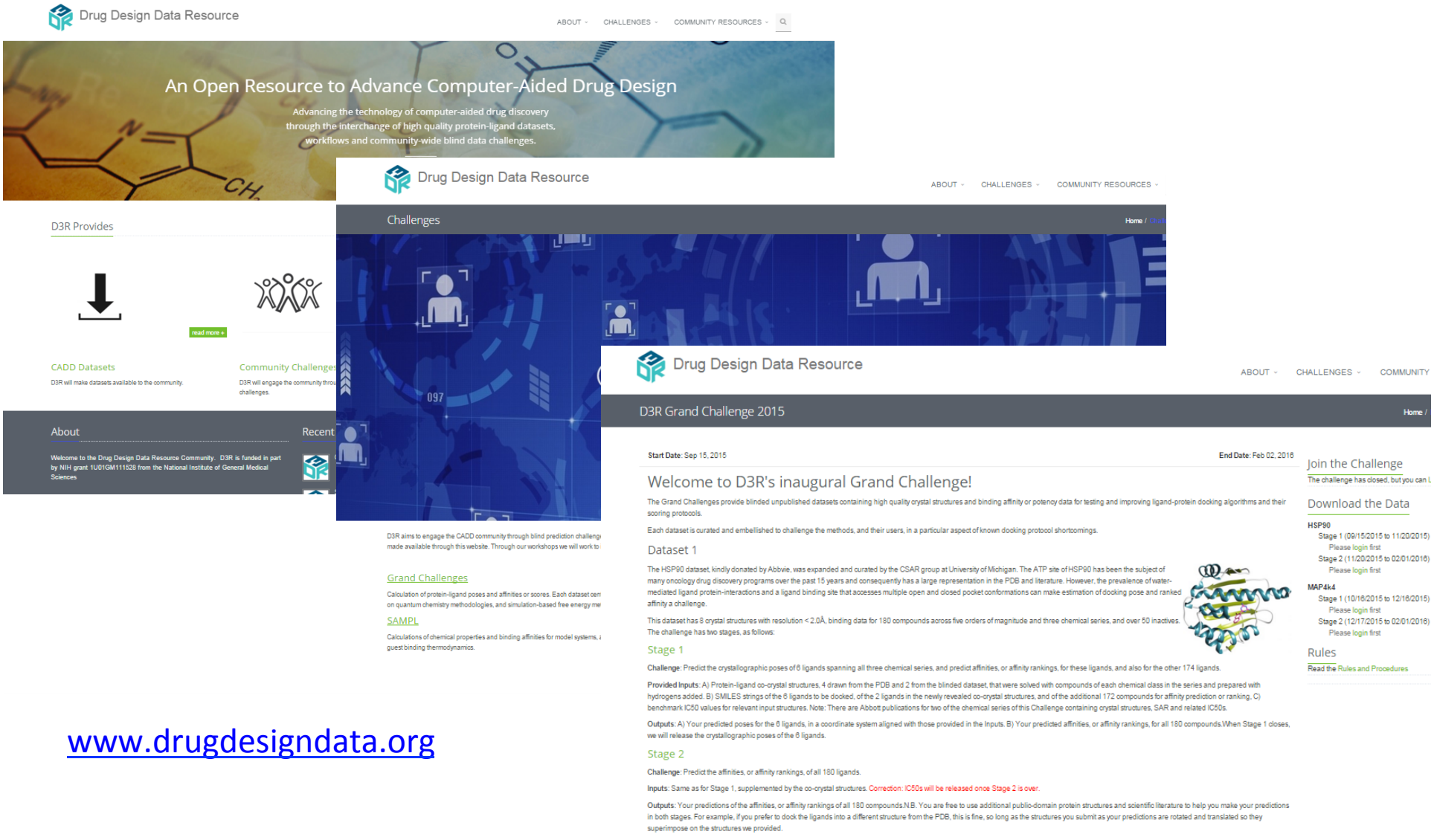
Experimental conditions, methods, uncertainties

Our first 2 years

Datasets Collected

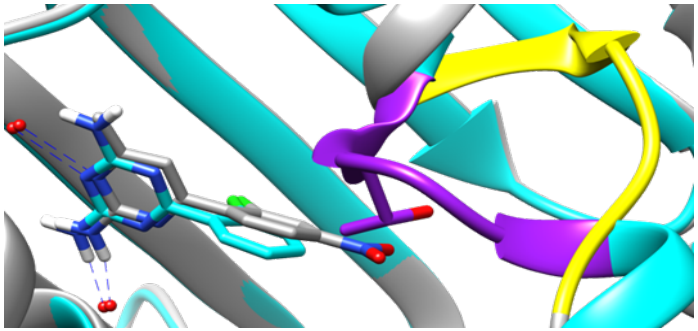
Contributor	# of datasets	Curation
Abbvie - CSAR	1	Completed by CSAR
Genentech	1.5	Not needed
Roche	2	Completed by D3R
Baker/Stoddard	2	Completed by D3R
SGC - Oxford	1	Completed by D3R

Web Portal for Data, Workflows, Challenges & Community Activities

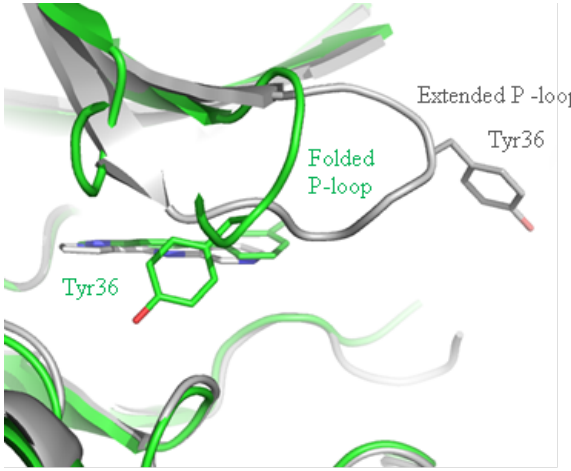


www.drugdesigndata.org

Grand Challenge 2015



HSP90 Dataset from Abbvie/CSAR



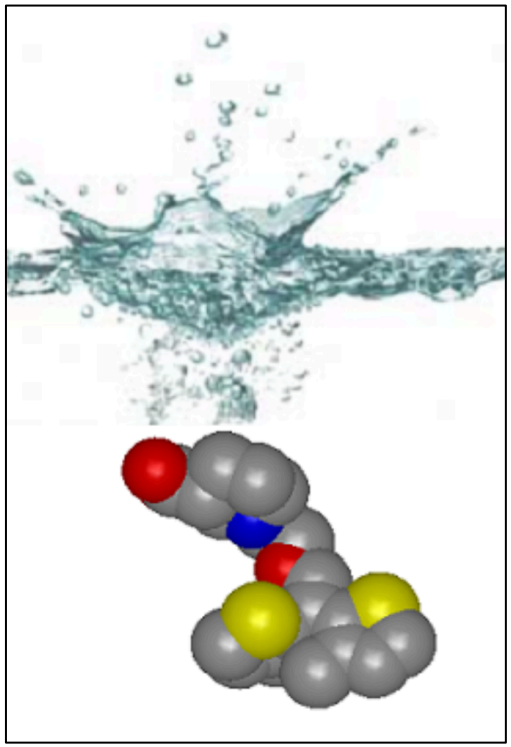
MAP4K4 Dataset from Genentech

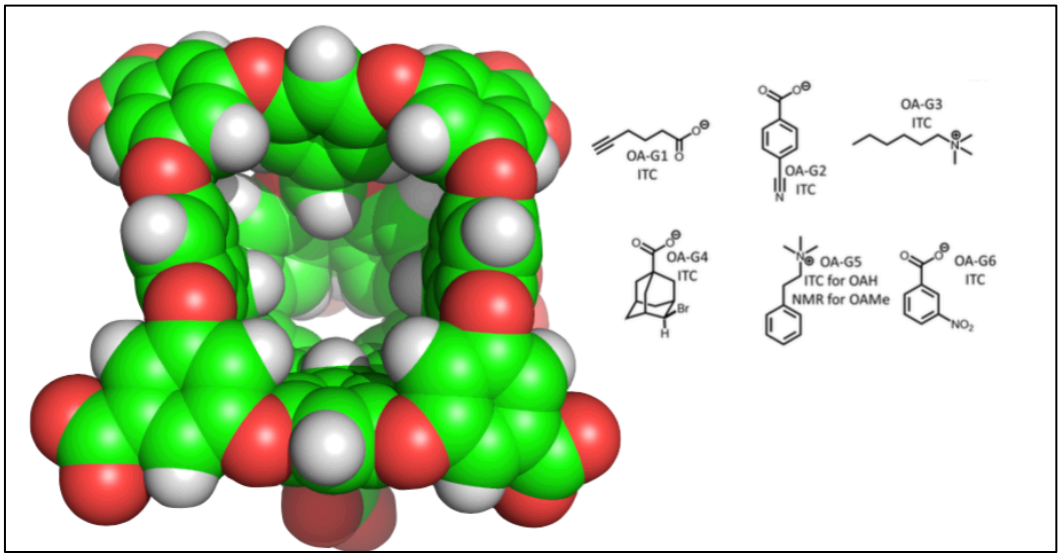
SAMPL Challenge 2015

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

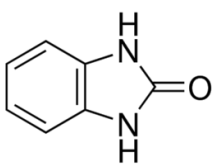
Joint workshop for D3R and SAMPL participants
84 participants from academia & pharma

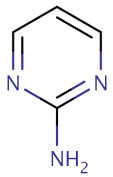
Grand Challenge 2015: Challenge Design

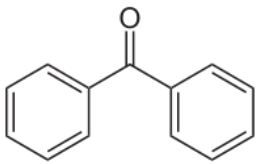
HSP90 Dataset from Abbvie/CSAR

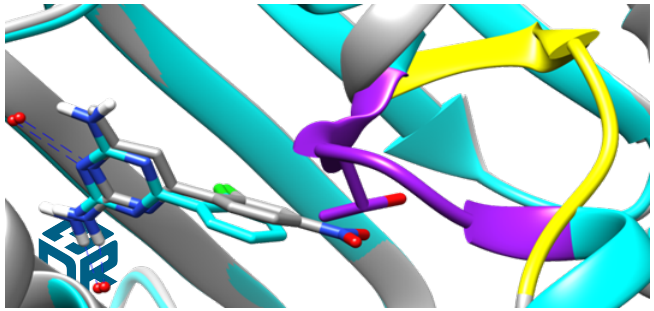
180 IC50s; 5.2nM – 50uM,
3 chemical series, 33 inactives
8 blinded crystal structures, R = 1.6 -1.96Å

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







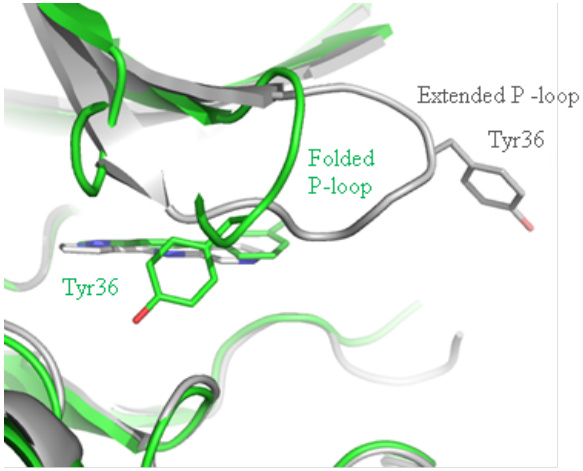


MAP4K4 Dataset from Genentech

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

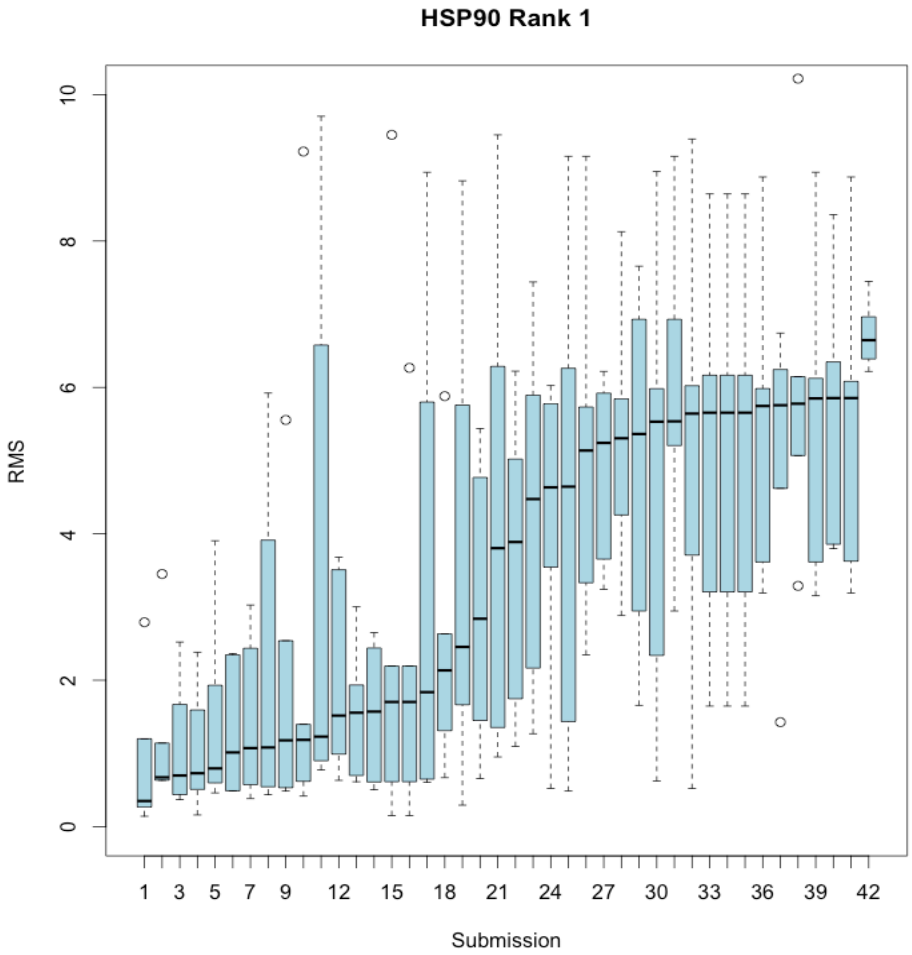
Successful prediction of docking poses depends on

- the entire workflow

- the amount of prior knowledge

Ranking (affinity scoring) remains challenging even with known poses

HSP90 Pose Prediction Results



P+L predictions = 5

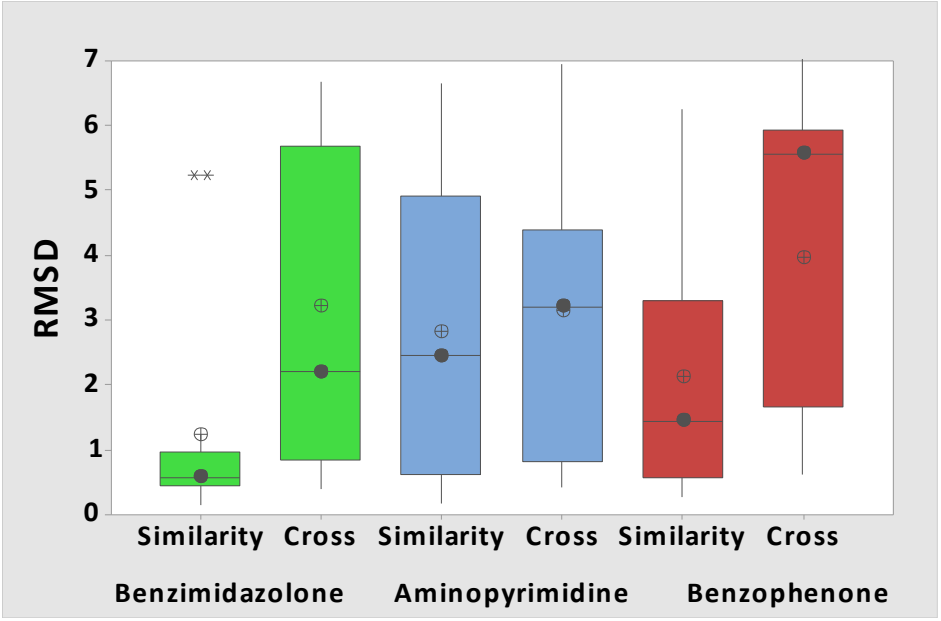
<u>Docking Programs for <2.0Å RMSD:</u>

AutoDock/Vina	rDock
Glide	RosettaLigand
Gold	Smina
Surflex	

<u>Docking Programs for >2.0Å RMSD:</u>

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

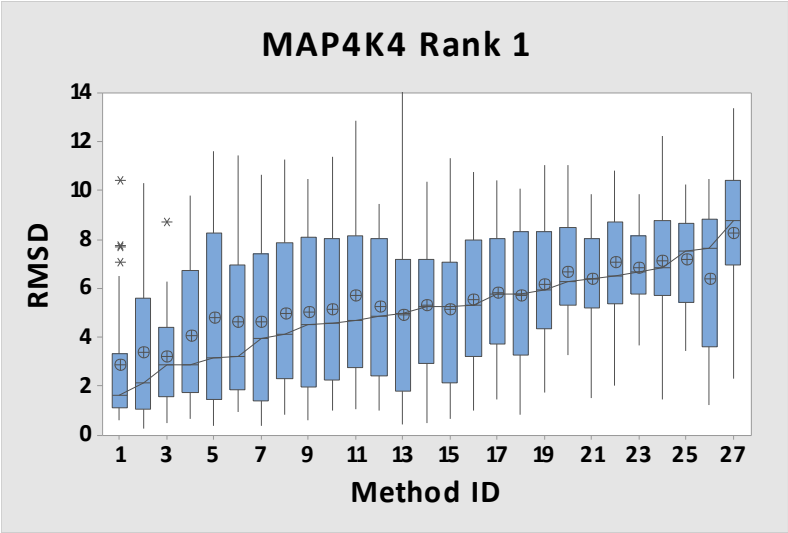
Included key waters

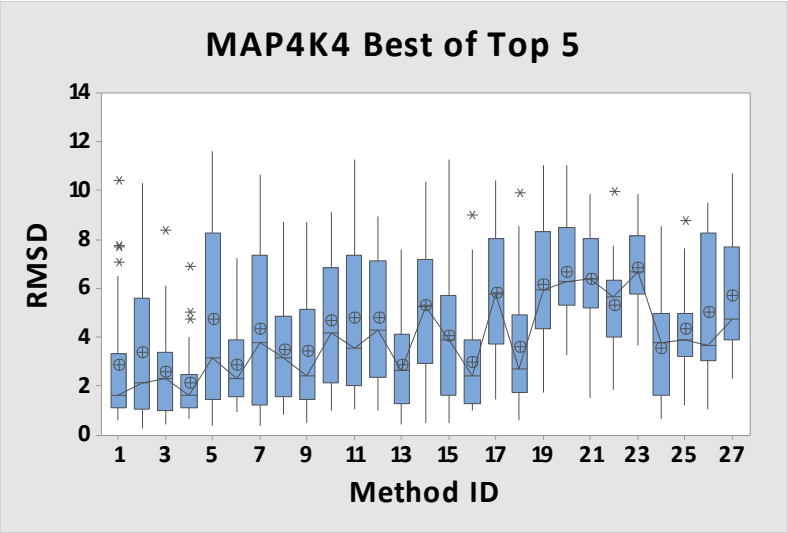
Visual inspection

MAP4K4 Pose Prediction Results

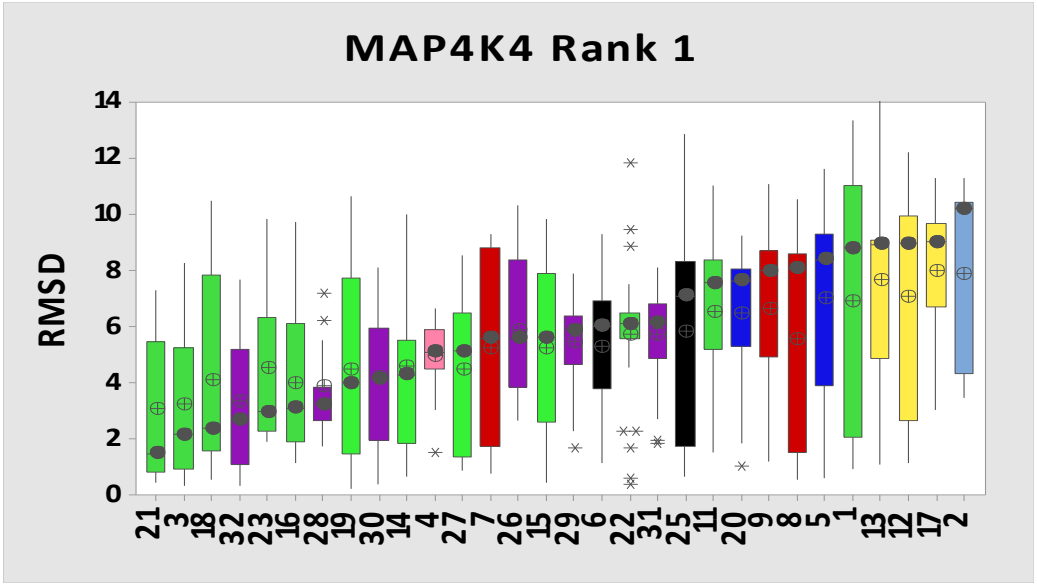
Participants were allowed to submit up to 5 top poses per ligand-receptor pair

P+L predictions = 32

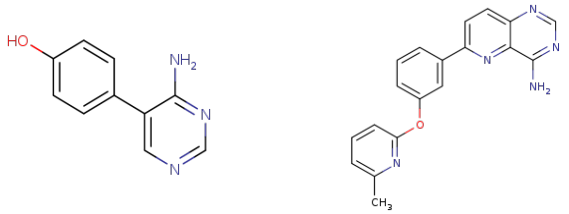




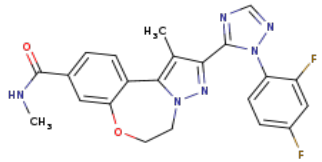
MAP4K4 Pose Results per Compound



Green = aminopyrimidine/aminoquinazoline



Yellow = benoxepin

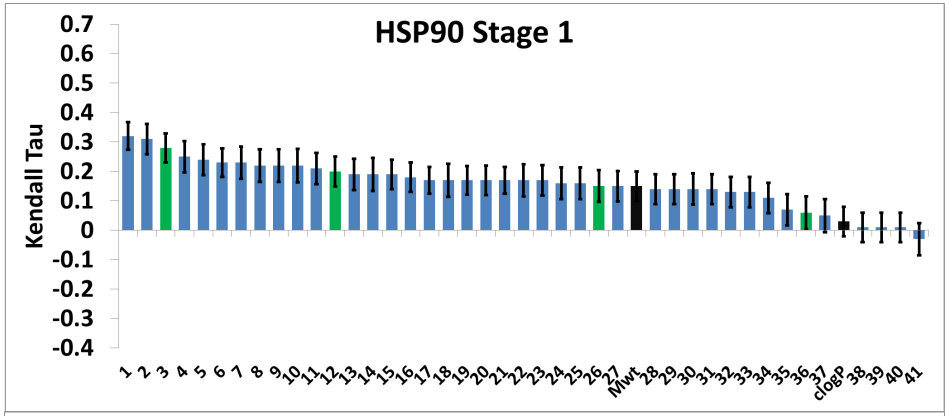


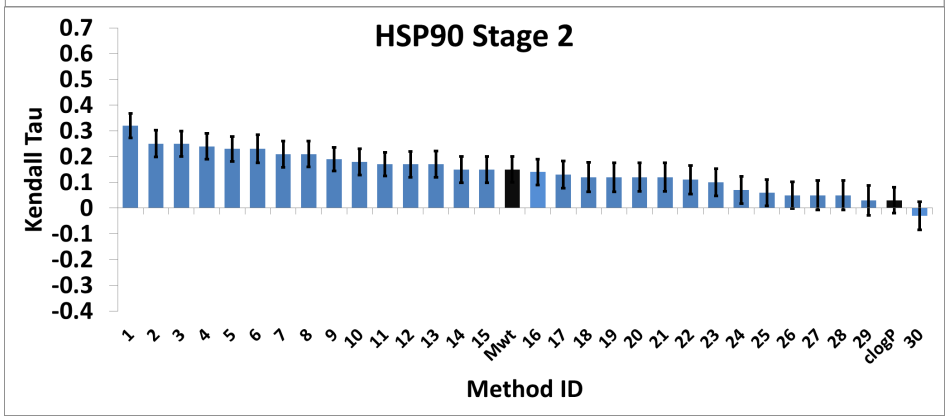
Success does not correlate with MW (fragment vs. compound), logP, #rotamers, tautomers, ligand affinity, open/closed P-loop,or xtal resolution.

~ correlation to prior chemotype examples

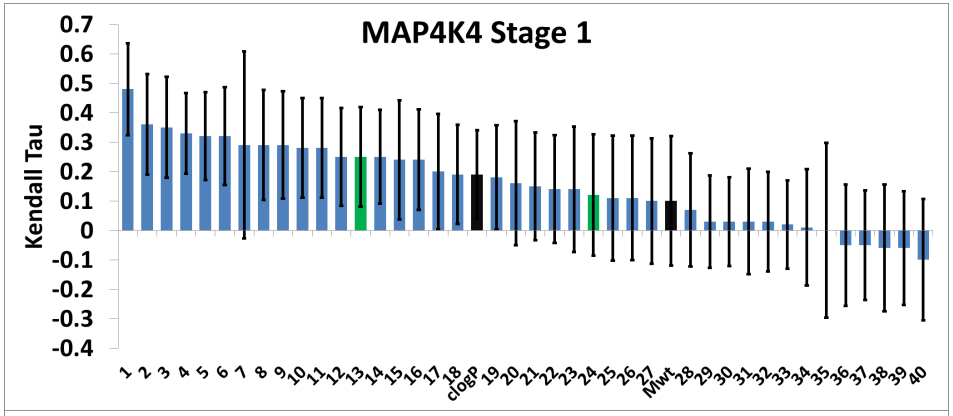
Potency Ranking

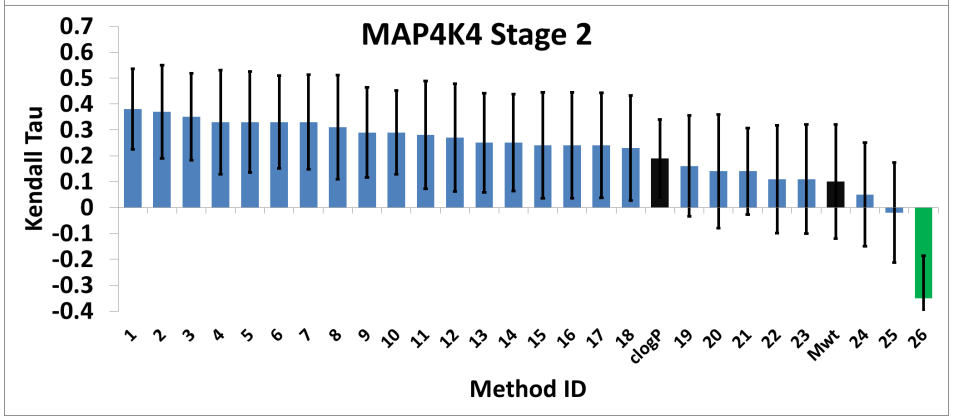
N = 179





N = 18

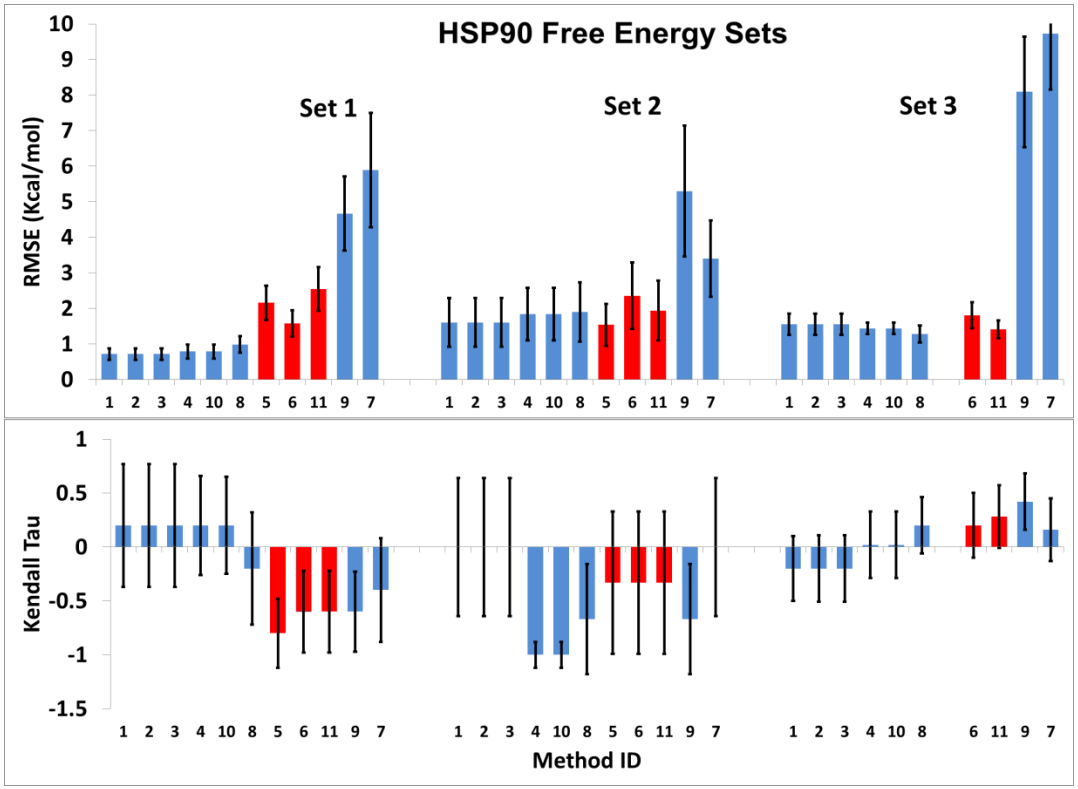




Methods

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

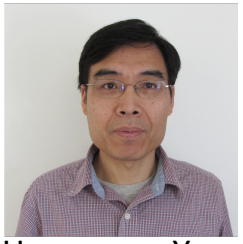

Mike Gilson


Stephen Burley

Symon Gathiaka


Shuai Liu


Jeff Wagner


Huanwang Yang


Jasmine Young


Chris Churas


Jeff Grethe


Mike Chiu

RUTGERS

RCSB PDB
PROTEIN DATA BANK

UC San Diego


Pat Walters
Challenge Evaluator
at Relay

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