

Drug Design Data Resource

Leveraging Data to Drive Progress in Protein-Ligand Modeling for CADD

Rommie Amaro, Stephen Burley^a, Vicki Feher and Mike Gilson
UC San Diego and Rutgers ^a



An aerial photograph of the University of California, San Diego campus. The image shows a dense cluster of modern academic buildings, green spaces, and parking lots. In the background, the coastline of San Diego is visible, with the blue ocean meeting the shore. The sky is clear and blue.

UC San Diego



Welcome to the first D3R workshop!

NIH-U01 Resource, Unique Purpose

Central Goal: Utilize previously unpublished datasets as benchmarks for developers of protein-ligand modeling technologies

Synergy with Public Databases: Public release of more industrial crystal structures and affinity data

Broader Goals: Utilize blinded datasets to drive improvement of all CADD technologies and to foster training and dissemination of methods

More predictive CADD methods benefit everyone!



D3R Team



Rommie Amaro



Vicki Feher



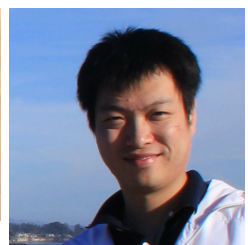
Mike Gilson



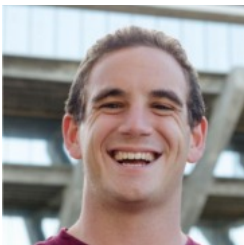
Stephen Burley



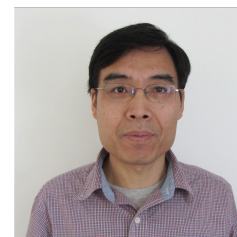
Symon Gathiaka



Shuai Liu



Jeff Wagner



Huanwang Yang



Jasmine Young



Chris Churas



Jeff Grethe



Mike Chiu

RUTGERS
RCSB **PDB**
PROTEIN DATA BANK

UC San Diego



Scientific Advisory Board



Aled Edwards
SGC



Charles Grimshaw
Takeda



Marti Head
GSK



David Mobley
UC Irvine



John Moulton
U Maryland



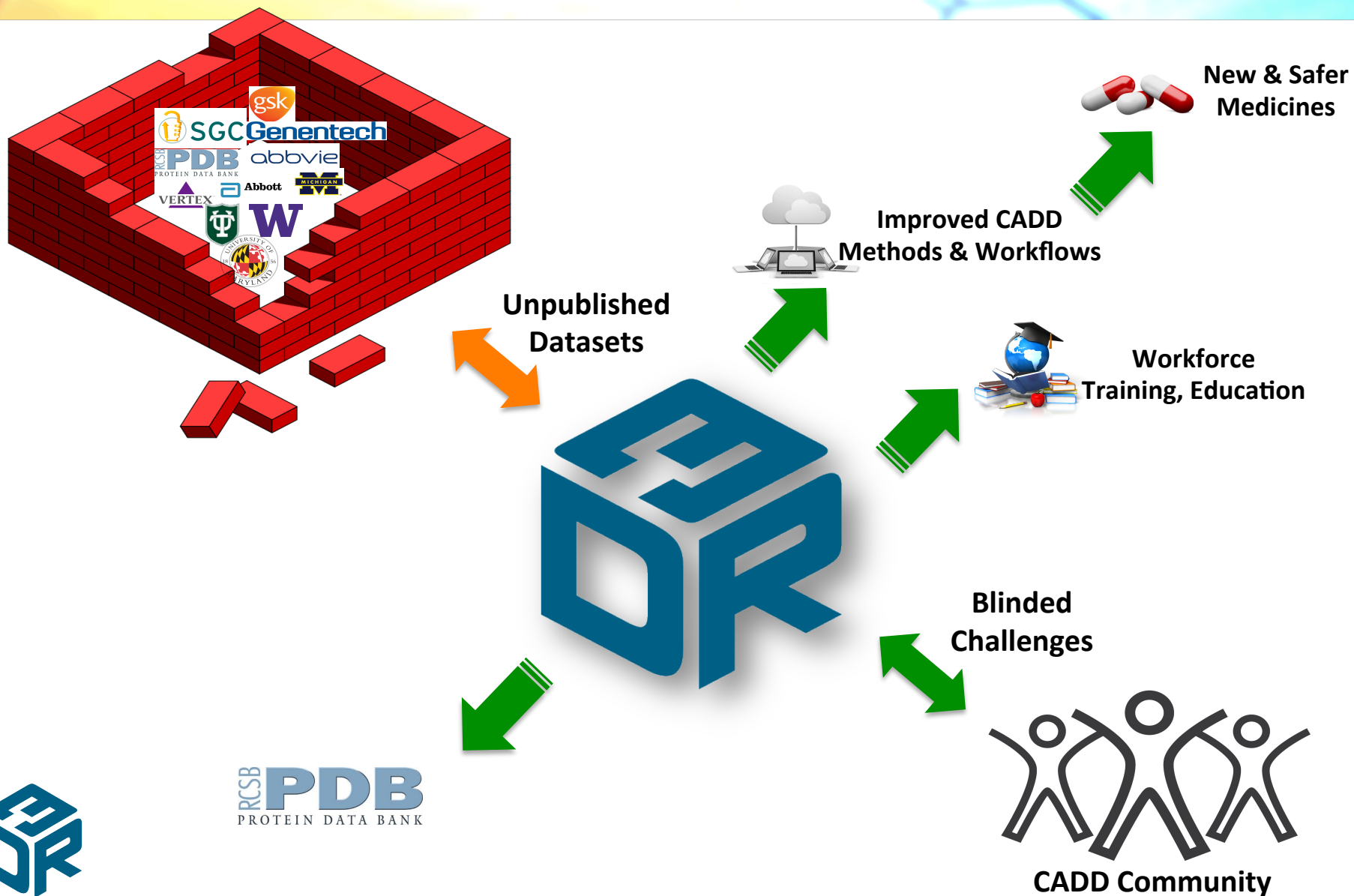
Adrian Roitberg
U Florida



Torsten Schwede
Biozentrum, Basel



D3R's Role & Vision



Unique Hub for CADD Community

Coherent CADD Datasets

Blinded challenges: Protein-ligand, model systems

Evaluation metrics

Capturing and disseminating workflows

Workshops

Networking, training, outreach, sabbaticals



D3R Challenge Types

Protein-Ligand Grand Challenges

Poses

Affinities

Model System Challenges (SAMPL); eg.,

partition/distribution coefficients

host-guest affinities

Continuous Evaluation of Ligand Pose Predictions



Continuous Evaluation of Ligand Pose Predictions

CELPP



100 – 200⁺ new co-crystal structures appear weekly in PDB
Automated, blinded docking enables statistical analyses
protocols and parameters
receptor and ligand types

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	<p>8:01 pm 1.pdbdownload</p> <p>11:01 pm 5.*.evaluation 6.resultwebdata</p>			<p>8:01 pm 1.makeblastdb 1.dataimport 2.blastnfilter 3.proteinligprep 4.webdata 4.fred 4.glide</p>		
<p>12:01 am Saturday to 2:59 pm Tuesday Tuesday User submissions accepted</p>		<p>3:00 pm Tuesday to 12:00 am Saturday No user submissions allowed during this time</p>			<p>12:01 am Saturday to 2:59 pm Tuesday User submissions accepted</p>	



Recent and Current Challenges

1st Grand Challenge

September 15, 2015 – February 1, 2016

2 protein – ligand datasets: HSP90 & MAP4K4

36 new crystal structures, 210 compounds

SAMPL5

3 host-guest series: CBClip, OAH, OAMe

~50 cyclohexane/water distribution coefficients

Challenge PL-2016-1 (active)

February 17, 2016 – March 22, 2016

2 protein-ligand datasets: 17-OHP, cholecalciferol

5 new crystal structures, 3 compounds



Web Portal for Data, 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 1U01GM111523 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 6 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 6 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.

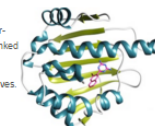
Outputs: A) Your predicted poses for the 6 ligands, in a coordinate system aligned with those provided in the Inputs. B) Your predicted affinities, or affinity rankings, for all 180 compounds. When Stage 1 closes, we will release the crystallographic poses of the 6 ligands.

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](#)

Dataset Donations & Activities

Individual Pharma Company contributors

Abbvie, Genentech, Vertex, GSK, ...

Academic Contributors

Univ. Michigan, Stoddard/Baker, Schiffer, Isaacs, Gibb, Chodera, ...

Structural Genomics Consortium

Cross-Pharma Consortium

GSK, Roche, Astex, Novartis, Merck, Pfizer, BMS, ...



Workshop Goals

Grand Challenge 2015 & SAMPL5 Scientific Reports

Feedback on challenges and other D3R activities

Synergies with RCSB PDB, TDT, CAMEO

D3R directions and announcements

Forthcoming protein-ligand challenges

CELPP



Thursday- Plenary

Overviews and perspectives (Moderator: Amaro)

D3R Introduction (Amaro, Gilson, Feher, UCSD)

NIGMS perspective (Peter Preusch, NIH)

CSAR and lessons learned (Heather Carlson, U. Michigan)

Grand Challenge 2015: poses and affinities (Moderator: Feher)

Overview of results (Pat Walters, Vertex)

Results 1 (Antonia Mey, U. Edinburgh)

Results 2 (Xinjun Hou, Pfizer)

Results 3 (Xiaoqin Zou, U. Missouri)

SAMPL5: distribution coefficients and host-guest affinities (Moderator: Gilson)

Overview of results (David Mobley, UC Irvine)

Experimental systems and data (Bas Rustenburg, MSKCC; Bruce Gibb, Tulane)

Distribution coefficients (Andreas Klamt, COSMOLogic)

Host-guest (Emillio Gallicchio, City U. of New York)

Poster session

Dinner

Friday- Plenary & Breakouts

General Discussion of D3R: dataset types, challenge schedules, workshop frequency...

CAMEO and CELPP (Torsten Schwede)

Group Photo

Morning breakouts: Protein-Ligand, SAMPL model systems

Lunch and SWOT analysis (Cathy Peishoff)

Afternoon breakouts: Protein-Ligand, SAMPL model systems

CADD status and directions (Cathy Peishoff)

Special issues of JCAMD (Terry Stouch)

Concluding remarks (Amaro, Feher, Gilson)



Friday- Breakouts

Morning

Protein-Ligand Track

Zhaofeng Ye
Nanjie Deng
David Koes

Model Systems Track

Lyle Isaacs (video)
Frank Pickard
Stefano Bosisio
Stefan Kast

Afternoon

Protein-Ligand Track

S. Burley: PDB–D3R Synergies
H. Jansen: TDT–D3R Synergies
V. Feher: CELPP and Future Grand Challenges

Model Systems Track

Florentina Tofoleanu
Kennie Merz
Ulf Ryde
Lyle Isaacs Q&A



Practicalities

Meals

Light breakfasts: today and tomorrow

Lunch today and tomorrow

Dinner today, here; on your own tomorrow

Shuttles

Both mornings 8:15 am. (Hotel shuttle for emergency backup)

Thursday evening: 7:30 pm after dinner

Friday evening: 5:00 pm

Posters

Staging in Seuss Room

Contact People

Ken Tomory and Jenny Chong



Questions?





Peter Preusch, Ph.D.

Chief, Biophysics Branch

Division of Cell Biology and Biophysics

National Institute of General Medical Sciences

National Institutes of Health

