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]

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Central Goal: Utilize previously unpublished datasets as benchmarks for developers of protein-ligand modeling technologies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

Adrian Roitberg
U Florida

Torsten Schwede
Biozentrum, Basel

D3R's Role & Vision

Unpublished Datasets

Improved CADD Methods & Workflows

New & Safer Medicines

Workforce Training, Education

Blinded Challenges

CADD Community

RCSB PDB PROTEIN DATA BANK

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

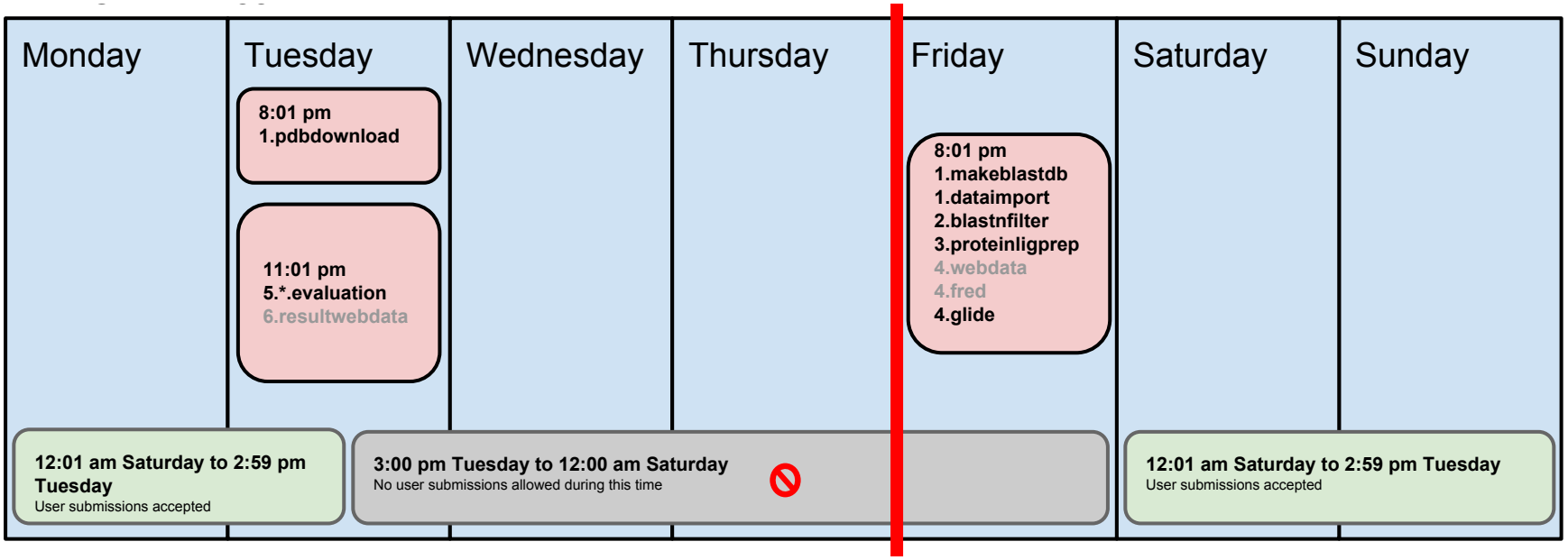
Continuous Evaluation of Ligand Pose Predictions

CELPP

100 – 200^{+} new co-crystal structures appear weekly in PDB

Automated, blinded docking enables statistical analyses

- protocols and parameters
- receptor and ligand types



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

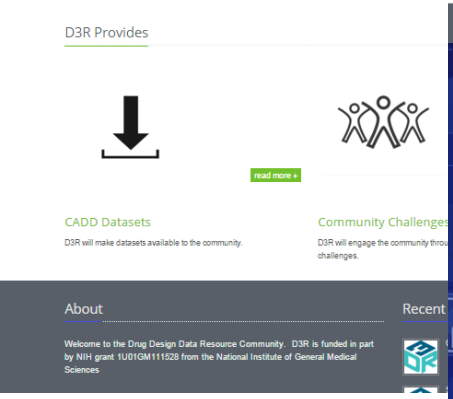
SAMPL5

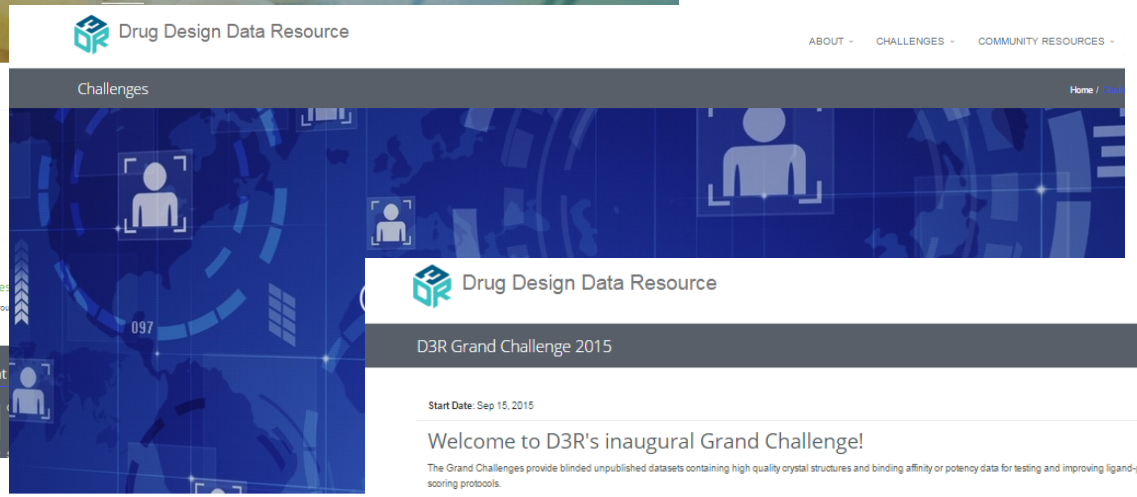
- 3 host-guest series: CBClip, OAH, OAMe
- ~50 cyclohexane/water distribution coefficients

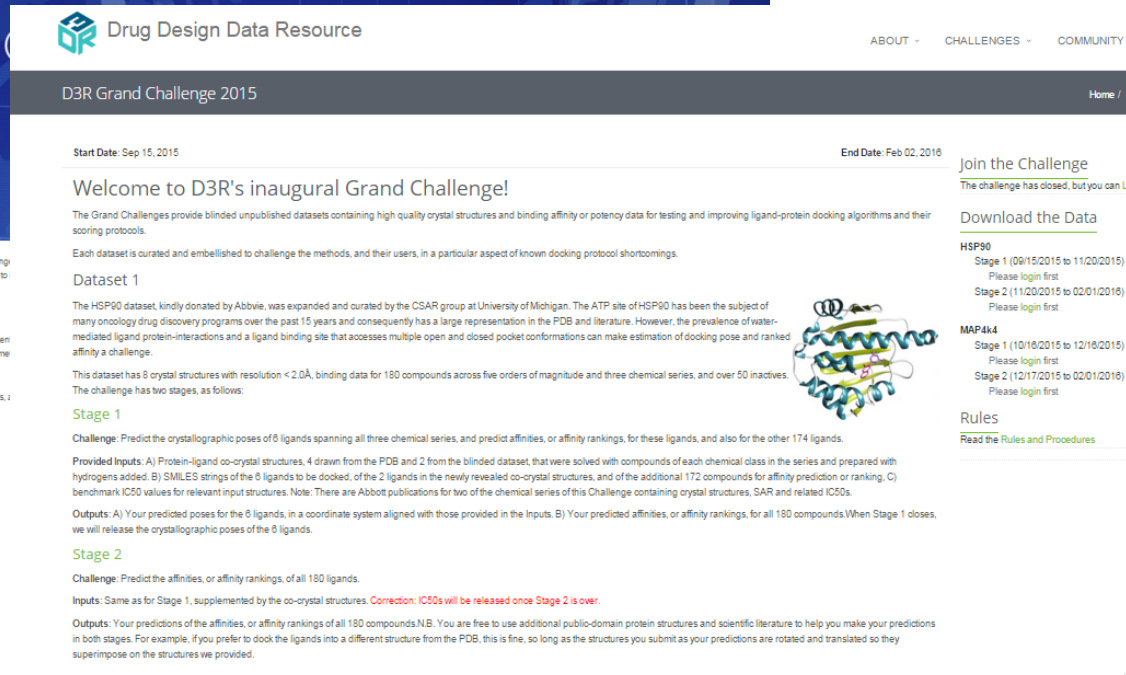
Challenge PL-2016-1 (active)

- February 17, 2016 – March 22, 2016
- 2 protein-ligand datasets: 17-OHP, cholecalciforerol
- 5 new crystal structures, 3 compounds

Web Portal for Data, Challenges, Community Activities







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