Welcome to the 2018 D3R workshop!

Workshop on Challenges in Docking and Screening

US National Institutes of Health, 2005

Participants from pharma, academia, US government

- Academic: Mike Gilson, Art Olson, Brian Shoichet
- Government: Chris Austin, Anne Chaka, Jayne Kapur, Janna Wehrle
- Pharma: Jeff Blaney, Wendy Cornell, Debbie Loughney, Cathy Peishoff, Emanuele Perola

Conclusions

- Computational predictions of poses and affinities need to improve
- Datasets from pharma could help

NIH National Institute of General Medical Sciences

Workshop report: http://bit.ly/2pLpy8C

Evaluation of Protein-Ligand Modeling Methods

Tests have been run with knowledge of the experimental results

Different methods have been tested for different systems

Value of blinded, community-wide, prediction challenges

NIH-U01 Resource, Unique Purpose

blinded prediction challenges to drive advances in CADD

NIH-funded initiative

CSAR 2010-2014 (Carlson, U. Michigan)

D3R 2014-present (Amaro & Gilson, UC San Diego)

Pharma as potential source of data

highly relevant

hitherto unpublished

Drug Design Data Resource (D3R)

Central Goal: Utilize previously unpublished datasets as benchmarks for developers of protein-ligand modeling technologies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

More predictive CADD methods benefit everyone!

D3R Project Team

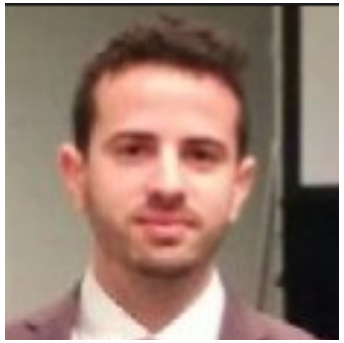
UC San Diego

RUTGERS RCSB PDB PROTEIN DATA BANK

Crystallography Group

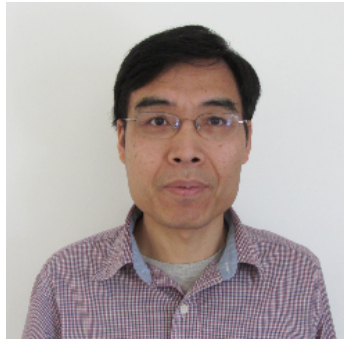

Rommie Amaro


Mike Gilson


Zied Gaieb


Conor Parks


Stephen Burley


Huanwang Yang

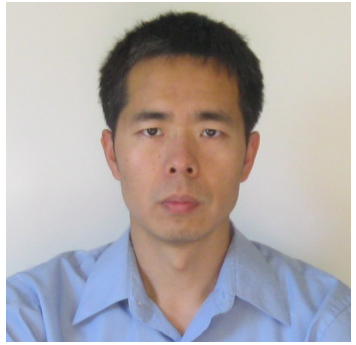

Jeff Wagner


Mike Chiu


Chris Churas


Jeff Grethe


Jasmine Young


Chenghua Shao

D3R Scientific Advisory Board

Aled Edwards
SGC

Charles Grimshaw
Takeda

David Mobley
UC Irvine

John Moult
U Maryland

Adrian Roitberg
U Florida

Torsten Schwede
Biozentrum

Martin Stahl
Roche

Unique Hub for CADD Community

Coherent CADD datasets

Blinded challenges: Protein-ligand, model systems

Evaluation metrics

Capturing and disseminating workflows

Workshops and networking

Challenge Types

Grand Challenges: ligand-protein poses and affinities

SAMPL: affinities, physical properties of simpler systems
with David Mobley, John Chodera, & Michael Shirts

CELPP: automated, weekly pose prediction challenge

Grand Challenges

Stage 1: Predict poses and affinities of multiple ligands for a protein

Stage 1b: Release co-crystal structures without ligands to enable self-docking (Isolates evaluation of docking algorithm)

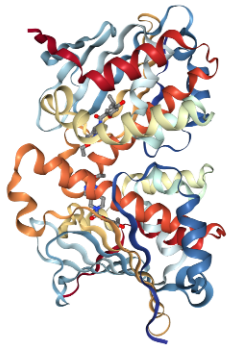
Co-crystal structures with ligands released

Stage 2: Predict affinities again

All data released, deposited to PDB, BindingDB

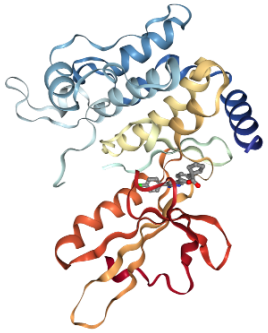
Grand Challenge 2015

35 participants, 355 submissions



HSP 90: focus on potency predictions

Data from Abbvie and Carlson's CSAR project
8 cocrystal structures (.6-2.0 Å resolution)
180 IC50s (5 nM-20 µM)
Three series: benzimidazolones, aminopyrimidines, benzophenone-like
Varied water-mediated interactions; open/closed conformations

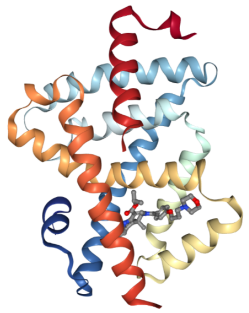


MAP4K4: focus on pose predictions

Data from Genentech
30 cocrystal structures (1.6 – 2.5 Å resolution)
18 IC50 data (3.1 nM - 10 µM)
Diverse chemotypes binding in ATP site
Open/closed P-loop structures

Grand Challenge 2

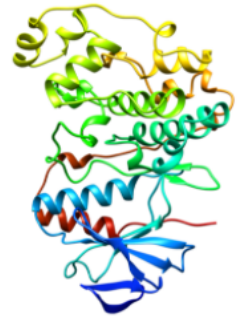
49 participants, 262 submissions



Farnesoid X Receptor (FXR): poses and potencies

Data from Roche
36 cocrystal structures (resolutions <2.6Å)
102 IC50s (0.3 nM-260 µM)
Three series + misc: sulfonamides, benzimidazoles, spiros
Helix shifts and varied water-bridges

Grand Challenge 3

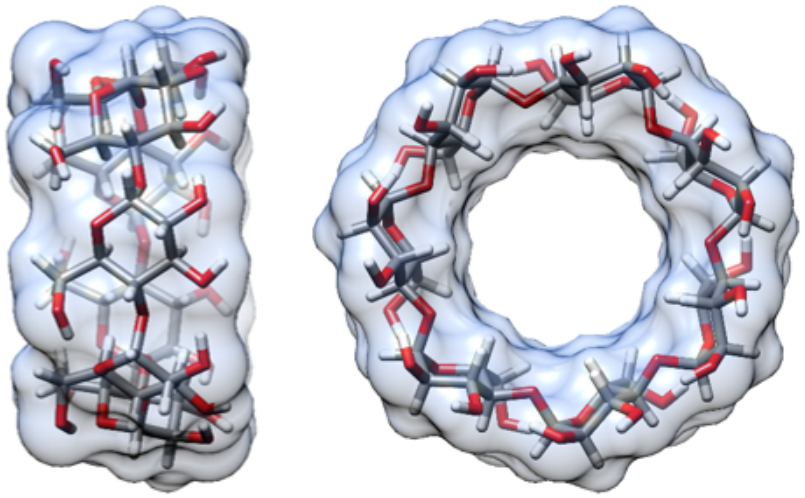


Kinase K_ds

SGC-UNC/DiscoverX

Cathepsin S poses & IC50s	Selectivity	Activity Cliffs	Activity Cliffs	Mutations
Janssen Pharmaceuticals	**VEGFR2**	**JAK2**	**TIE2**	**ABL1**
24 cocrystal structures, 3.0 Å	85 (0.62 to >10^4 nM)	17 (53 to >10^4 nM)	18 (3.4 to >10^4 nM)	12 (49 to >10^4 nM)
136 IC50s, 3 – 8500 nM	**JAK2**			
	89 (0.66 to >10^4 nM)			
	p38-α			
	72 (0.28 to >10^4 nM)			
27 participants	11 participants	6 participants	6 participants	6 participants
303 submissions	94 submissions	25 submissions	32 submissions	11 submissions

SAMPL Blinded Prediction Challenges



β-cyclodextrin

Started 2008

Small molecule hydration free energies
Nicholls, Mobley, Guthrie, Chodera, Bayly, Cooper, Pande

Simple model systems, e.g.,

Host-guest binding affinities
Water-organic phase partition coefficients
Small molecule pKa values
Small molecule hydration free energies

Advantages vs. protein-ligand challenges

Calculations far easier to converge
Troubleshooting by isolation of specific issues
Reduced ambiguity (protonation states, missing residues…)

https://en.wikipedia.org/wiki/SAMPL_Challenge

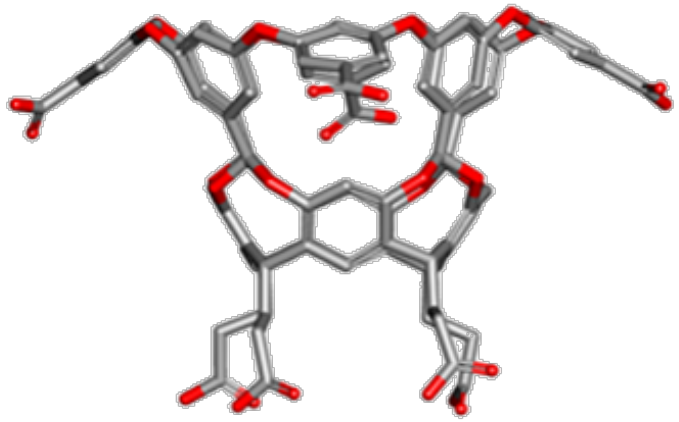
SAMPL6 Host-Guest and pKa Challenges

Host-Guest: 124 submissions from 6 groups

pKas: 95 submissions from 10 groups

Deep Cavity Cavitand Hosts

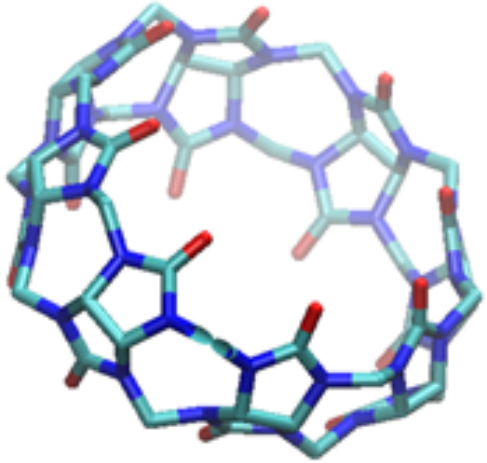
Bruce Gibb, Tulane U.



8 guest molecules with both
OA and TEMOA host variants

Cucurbit[8]uril Host

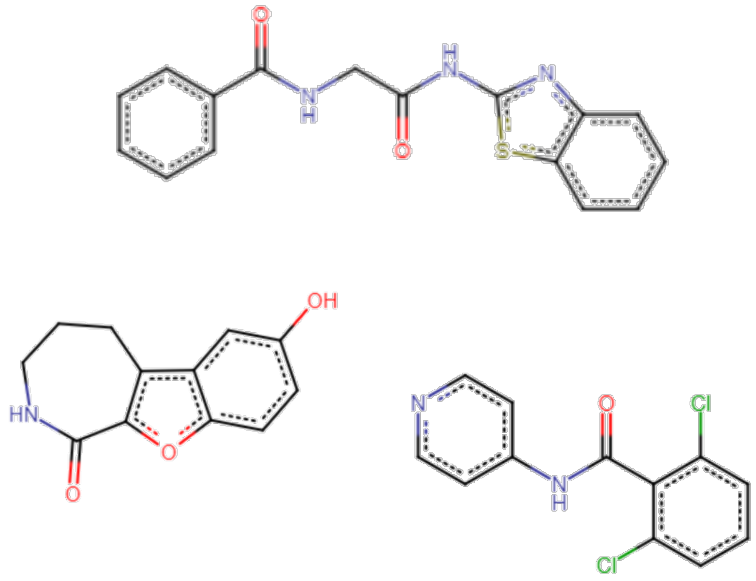
Lyle Isaacs, U. Maryland



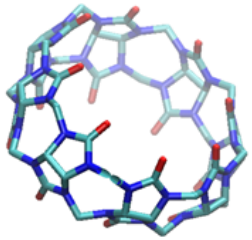
10 guest molecules

Small Molecule pKas

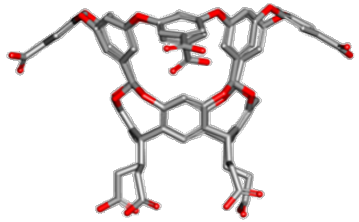
John Chodera, SKMCC

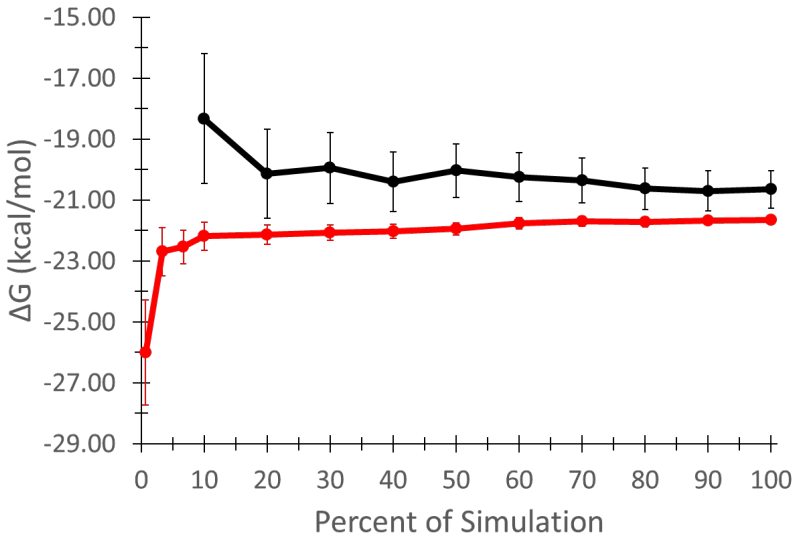


25 compounds



New SAMPLing Challenge





Tests efficiency of conformational sampling methods

Binding free energy convergence with

- Number of energy evaluations
- Wall clock time
- Total CPU time

Host-guest systems with specific setups

Wide Range of Protein-Ligand Methods

Pose prediction variations

- Software packages; .e.g. AutoDock Vina, Glide, rDOCK, Gold, RosettaLigand, Surflex
- Ligand overlay often used; e.g. ROCS, PoPSS
- Relaxation and rescoring; e.g. Molecular dynamics, MMPB/SA
- Combinations; e.g. Gold-PlantsPLP-rDock, RosettaLigand-Omega-ROCS, Surflex-Grim

Affinity prediction and ranking

- Ligand-based
- Structure-based
  - “Low resolution” docking and scoring
  - “High resolution” free energy methods

Machine-learning

What we have learned…

https://drugdesigndata.org/about/what-we-have-learned

Accuracy of docking and scoring correlates poorly with software choice, and successful pose prediction depends on other methodological factors; e.g.,

- ligand and protein preparation
- choice of protein conformation
- treatment of xtal waters.

Pose prediction benefits from use of known ligand-protein cocrystal structures; e.g., by ligand overlay

Human inspection and intervention do not consistently improve results

Accuracy of poses used correlates poorly with scoring accuracy

Application of free energy methods to host-guest systems points to need for better force fields

Explicit solvent free energy methods have not yet outperformed faster scoring methods

Rigorous evaluation of predictions is non-trivial and can be controversial

From Our GC Participants…

It has made me more aware of the challenges of sampling. I've been working on better ways to include this into our protocols and methods.

I would pay more attention to the receptor conformations and flexibility.

The D3R challenges allowed us to validate our docking protocol

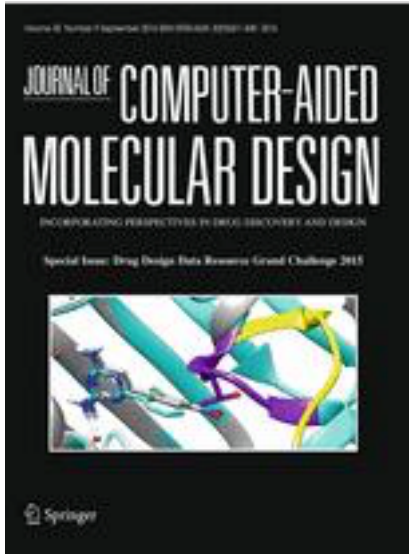
Docking seems to be improved by machine learning and I plan to incorporate such approaches.

... it will definitely change the I do docking to avoid or minimize false positives.

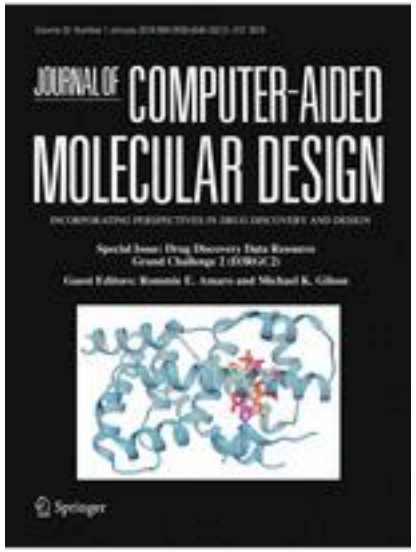
It has made me pleasantly surprised when a scoring function actually delivers a useful result and makes me very skeptical of people who blindly trust the score that they get.

Special Issues in JCAMD

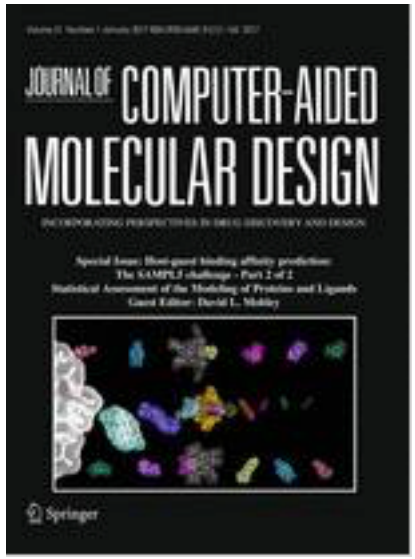
thanks to Terry Stouch, Senior Editor-in-Chief



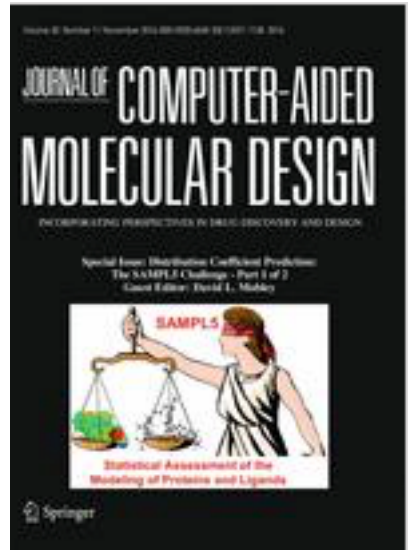
GC 2015
14 articles, 2016



GC2
23 articles, 2017



SAMPL5 1/2
12 articles, 2016



SAMPL5 2/2
17 articles, 2017

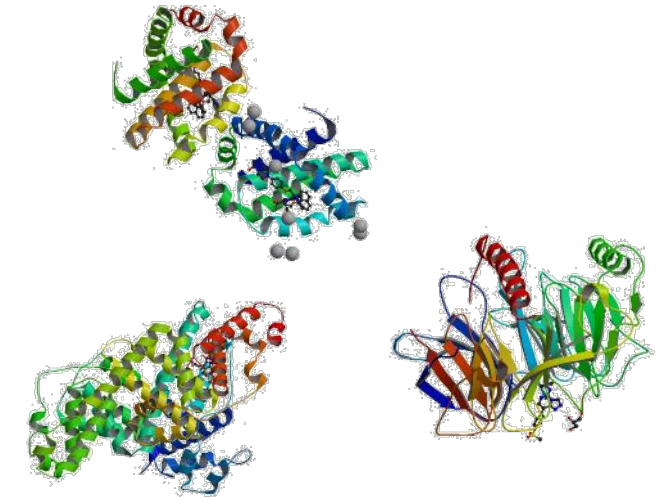
Toward Greater Statistical Power

Continuous Evaluation of Ligand Pose Predictions (CELPP)

Saturday

PDB pre-release
InChIs
Protein sequences
Forthcoming IDs

D3R scripts
Eliminate trivial ligands
Pick protein structures



Sunday

D3R releases InChIs and protein structures for docking

D3R opens for submissions

Tuesday

D3R submission window closes
PDB releases structures

Wednesday

D3R evaluates predictions against released structures

Capturing Complex Workflows

Method 1
OMEGA, SHAFTS, Amber11

Method 2
GLIDE-CCDC-GOLD, Amber14, MMGBSa

Method 3
WaterMap, SHAPE Screening, Structural Interaction Fingerprint, DFT/B3LYP/6-31G*, GLIDE-SP-XP, Induced-fit-docking, Emodel/GlideScore-SP, Binding Pose Metadynamics

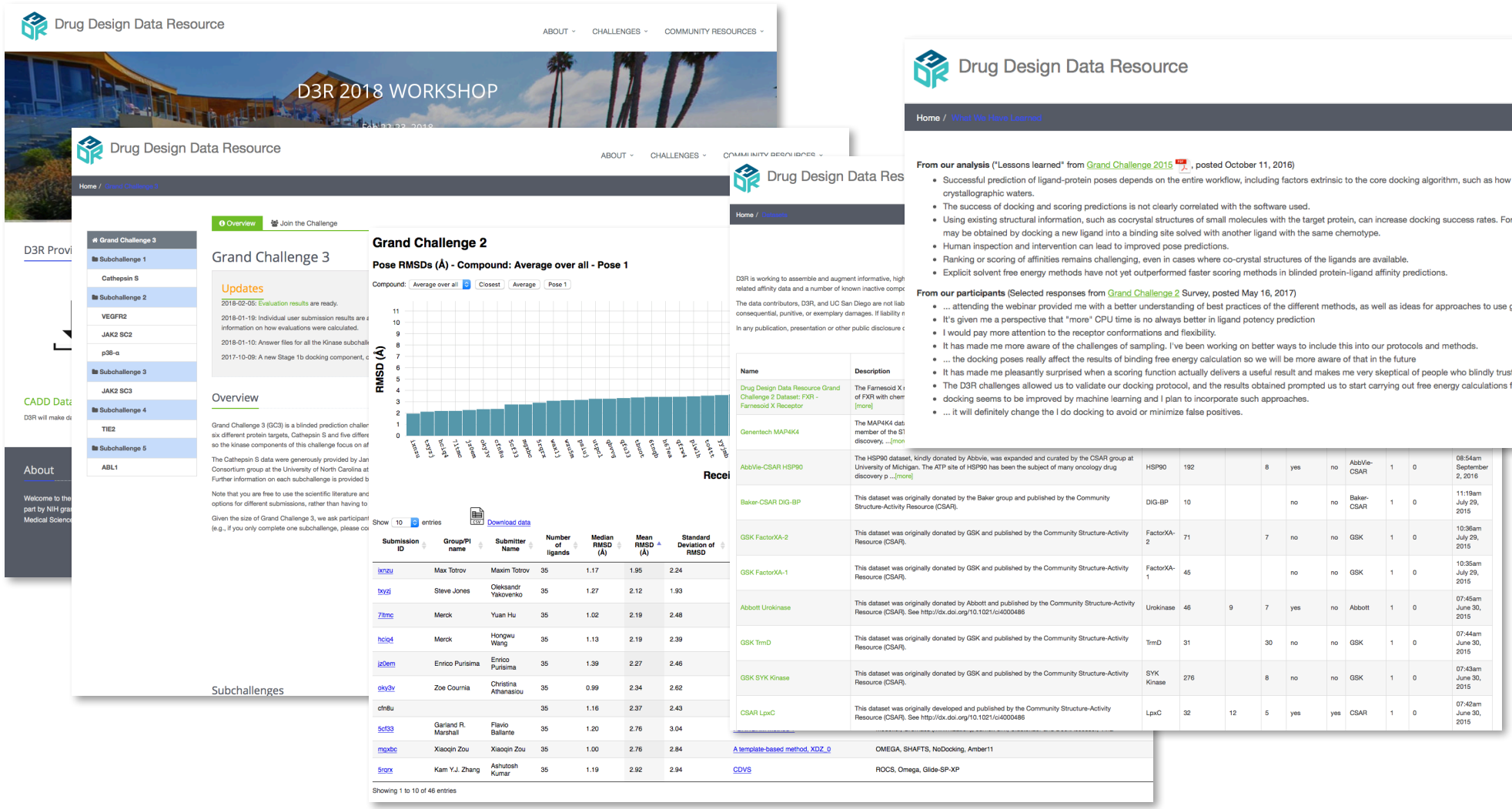
Full description of methods

Reproducibility

Evaluation on new datasets

Application to drug design projects

Web Portal for Data, Challenges, Community Activities



https://drugdesigndata.org

Thursday, February 22 at Scripps Institute of Oceanography Forum (SIO)

8:15 AM	Walk or Ride Share from La Jolla Shores Hotel to SIO Breakfast on-site
9:00-9:30 AM	*Welcome and D3R Update* — Mike Gilson and Rommie Amaro, D3R, UC San Diego
9:30-9:40 AM	*NIH Perspective* — Peter Lyster, NIGMS, NIH
9:40-10:15 AM	*Evaluation Overview of GC3* — Mill Lambert & Neysa Nevins, GSK
10:15-10:30 AM	**BREAK**
10:30-10:50 AM	*GC3 Participant Talk 1* — Maxim Totrov, MolSoft
10:50-11:10 AM	*GC3 Participant Talk 2* — Guo-wei Wei, Michigan State
11:10-11:30 AM	*GC3 Participant Talk 3* — David Koes, University of Pittsburgh
11:30-11:50 AM	*GC3 Participant Talk 4* — Ashutosh Kumar, RIKEN
11:50-12:10 PM	*Open Discussion on Evaluation Metrics* — D3R Moderator
12:10-1:30 PM	**LUNCH AT SIO**
1:30-1:40 PM	*SAMPL6 Intro* — John Chodera, MSKCC
1:40-1:55 PM	*SAMPL6 Host-Guest Intro and Overview* — Andrea Rizzi, MSKCC
2:00-2:15 PM	*SAMPL6 Host-Guest Participant 1* — Michail Papadourakis, Edinburgh (video presentatic
2:15-2:30 PM	*SAMPL6 Host-Guest Participant 2* — Marie Laury, Washington University
2:30-2:50 PM	*SAMPLing Challenge Overview and Results* — Andrea Rizzi, MSKCC
2:50-3:05 PM	**BREAK**
3:05-3:25 PM	*SAMPL6 pKa Intro and Overview* — Mehtap Isik, MSKCC
3:25-3:40 PM	*SAMPL6 Participant 1* — Samarjeet Prasad
3:40-3:55 PM	SAMPL6 Participant 2 — Qiao Zeng, NIH
3:55-4:10 PM	*SAMPL6 Participant 3*— Marvin Waldman, Simulations Plus
4:15-6:00 PM	Poster Session, Sunset (5:30PM) with Snacks and Liquid Refreshments
6:00-7:30 PM	**DINNER ON THE PATIO AT SIO**
7:30 PM	Walk or Ride Share Back to La Jolla Shores Hotel

Friday, February 23 at Scripps Institute of Oceanography Forum (SIO)

8:15 AM	Walk or Ride Share from La Jolla Shores Hotel to SIO Breakfast on-site
9:00-9:10 AM	*Introduction to Day 2* — Mike Gilson and Rommie Amaro
9:10-9:40 AM	*A Longitudinal View of the Grand Challenges* — Pat Walters, Relay Therapeutics
9:40-10:00 AM	*A Longitudinal View of the SAMPL Challenges*— David Mobley, UC Irvine
10:00-10:20 AM	*D3R Lessons Learned* — Alexandre Bonvin, Utrecht U (Video Conference)
10:20-10:40 AM	*SAMPL Lessons Learned* — Bogdan Iorga, ICSN, CNRS
10:40-11:00 AM	**BREAK**
11:00-11:20 AM	*Continuous Evaluation of Ligand-Protein Predictions (CELPP)* - Jeff Wagner, D3R, UC San Diego
11:20-11:40 AM	*GC and CELPP: Workflows and Insights* —Xiaoqin Zou, U. Missouri
11:40-12:00 AM	*MoISSI and Workflows* — John Chodera, Memorial Sloan Kettering
12:00-12:30 PM	*Open Discussion on Enabling Adoption of Workflows*
12:30-1:30 PM	**LUNCH AT SIO**
1:30-1:45 PM	**Group Photo**
1:45-2:00 PM	*Upcoming D3R and SAMPL Challenges*—Mike Gilson, Rommie Amaro, David Mobley
2:00-2:10 PM	*GC3 and SAMPL6 Special Issues* — Terry Stouch, Journal of Computer Aided Molecular Design
2:10-3:15 PM	D3R Community Feedback Discussion *Datasets—Attributes, Types, Size, Number/Challenge* *Challenges and Workshops—Timing, Type, Frequency* *Challenge Evaluations—Website Posting* *Future funding / support of blinded prediction challenges*
3:15- 3:30 PM	*Wrap-up and Conclusions*—Mike Gilson and Rommie Amaro
3:30 PM	**Workshop Concludes**
4:00-5:00 PM	SAB Meeting (Closed Session)
5:00-5:30 PM	SAB Session with D3R PIs
6:00 PM	SAB Dinner at La Jolla Shores Hotel

Practicalities

Meals

Light breakfasts: today and tomorrow
Lunch today and tomorrow
Dinner today, here; on your own tomorrow

Shuttles

Both mornings 7:30am and 7:45am
Thursday evening: 8:05pm and 8:15pm
Friday: 3:30pm, 3:45pm and 5:30pm

Posters: On walls, please use blue tape provided

Contact People

Megan Murphy
Iris Villanueva
Anyone from the D3R team

Acknowledgements

D3R Team at UCSD and Rutgers
D3R Scientific Advisory Board
SAMPL co-organizers: Profs. D. Mobley, J. Chodera, M. Shirts
Dr. Terry Stouch and the JCAMD team
Drs. Peter Lyster, Peter Preusch, and Janna Wehrle, National Institutes of Health
Data Contributors: Janssen Pharma, SGC-UNC, Chodera Lab, Gibb Lab, Isaacs Lab, others
External evaluators: Drs. Neysa Nevins, Mill Lambert, Pat Walters
All challenge participants