

TDT & D3R: PROGRESS TOWARDS FULLY ENABLED STATE-OF-THE-ART WORKFLOWS FOR DRUG DISCOVERY

Teach – Discover – Treat

An initiative to provide high quality computational chemistry tutorials to impact education and drug discovery for neglected diseases,
(founded under the umbrella of the COMP Division of the ACS)



Hanneke Jansen, TDT

March 2016

www.tdtproject.org

Outline

- TDT: Vision, results and impact
- Opportunities to improve and innovate the TDT concept
- Partnership between TDT and D3R

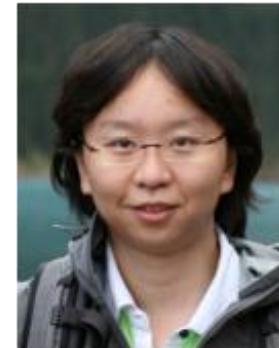
TDT Steering Committee



Hanneke Jansen
Novartis



Rommie Amaro
UCSD



Jane Tseng
National Taiwan University



Wendy Cornell
IBM Watson Lab



Emilio Esposito
exeResearch



Pat Walters
Vertex

What is needed to increase impact of computational drug discovery efforts

Where we are: room for improvement

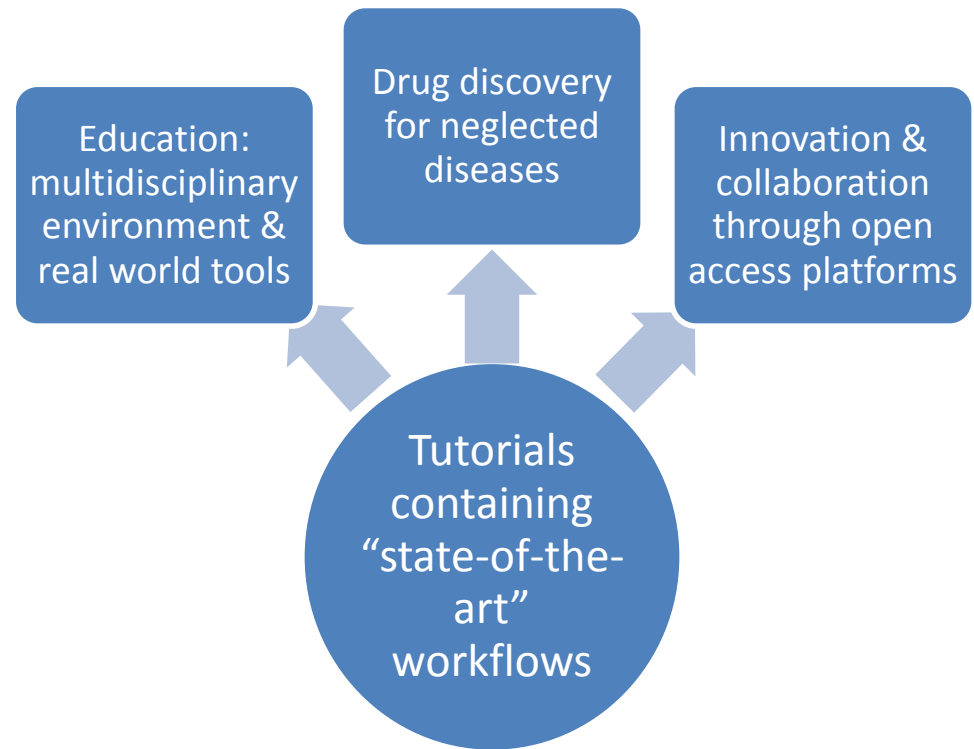
- Recognized contributions to drug discovery
- Predictions aren't perfect and our problems are challenging
- Many publications of new methods not enabled (data, source code)
- Anybody entering the field has a hard time finding state-of-the-art tools

Where we need to go: shore up the foundation for innovation

- “Recommit to reproducibility” so we learn from what works, as well as what fails
- Strengthen the education of drug discovery scientists across disciplines
- Enable state-of-the-art computational tools and methods

Enable “state-of-the-art” by encouraging the creation of high quality tutorials for computational workflows

- Solicit tutorials through competitions
 - Emphasis on fully enabled workflows (all data, models and parameters) that inspire the next experiment
 - Neglected disease angle: Partnerships for held-out test data **and** for executing on drug discovery
- “Open access” to impact drug discovery in all environments



Tutorials disseminated through
www.tdtproject.org

Judging Criteria & Judging Panel

- Scientific content
 - Quality of the underlying science
 - Relevance to drug discovery
- Presentation and clarity
 - Objectives clearly defined
 - Logical workflow
- Education benefit
 - Sufficient background
 - Stimulates interest in molecular modeling
 - Gateway to further research
- Reproducibility
 - Easily reproduced
 - Software available
 - Ease of installation



Steve Johnson
Bristol-Myers Squibb



Rajarshi Ghuha
NIH (NCATS)



Yvonne Martin
Abbott (Retired)



Steve Dixon
Schrödinger



Pat Walters
Vertex

TDT launched in 2012: two rounds of competition completed, 6 challenges

1. Structure-based design: New chemotypes for Dihydroorotate dehydrogenase (DHODH; Malaria)
2. Hit finding starting with HTS data-set from a Malaria phenotypic screen
3. Structure-based design: MedChem strategy & optimization for phosphodiesterase target in sleeping sickness
4. Virtual screening for parasitology drug target, the calcium-dependent protein kinase 1 (CDPK1) of *Eimeria tenella*
5. Structure-based design: High quality chemical starting points for *Pf* Lysyl tRNA synthetase (Krs1; Malaria)
6. Open innovation: A call for innovative workflows, models & tools to support drug discovery for neglected diseases

TDT operational notes

- Interest & global participation increased from round-1 to round-2
 - Round-1: 6 submissions, 1 overall winner, 3 runner-up awards
 - Round-2: 16 submissions, 1 overall winner, 2 runner-up awards
 - Awards: travel awards to present work at ACS meeting
- Additions in round-2
 - Held-out test-data available from TDT website: participants could assess their own performance before announcing externally
 - Prediction awards, separate from overall awards => use proposed compound lists for discovery efforts

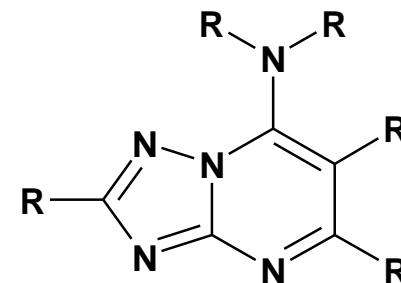
First fully completed challenge: New chemotypes for Dihydroorotate dehydrogenase, DHODH

- Case study with validated target, SAR set, and crystal structures
 - Partnership Meg Phillips (University of Texas Southwestern Medical Center)
 - Held-out test data: crystal structure of a new chemotype
- Workflow to start from SAR and structural information
 - Build models that can rationalize the SAR and be used to predict the binding mode of a new chemotype
- Drug discovery follow-up to identify new chemotypes
 - Apply predictive model to catalogue of commercial compounds and prioritize compounds based on affinity and novelty
 - Commitment at launch: At least 100 compounds will be acquired and tested

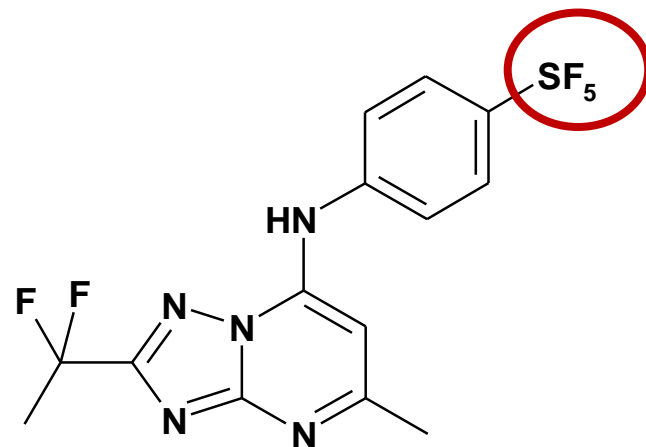
Triazolopyrimidine inhibitors of plasmodium DHODH show *in vivo* efficacy

- Dihydroorotate dehydrogenase, DHODH, catalyzes 4th step in *de novo* pyrimidine biosynthetic pathway
 - Malaria parasite uniquely vulnerable due to lack of salvage enzymes
- Extensive SAR set (193 compounds, 4 orders of magnitude)*
 - Includes a substituent that is challenging for some forcefields, pentafluorosulfanyl (-SF₅)
- Crystal structure of several analogues available

* J.Med.Chem, 2008, 3649-3653; J.Med.Chem, 2009, 1864-1872
J.Med.Chem, 2011, 3935-3949; J.Med.Chem, 2011, 5540-5561

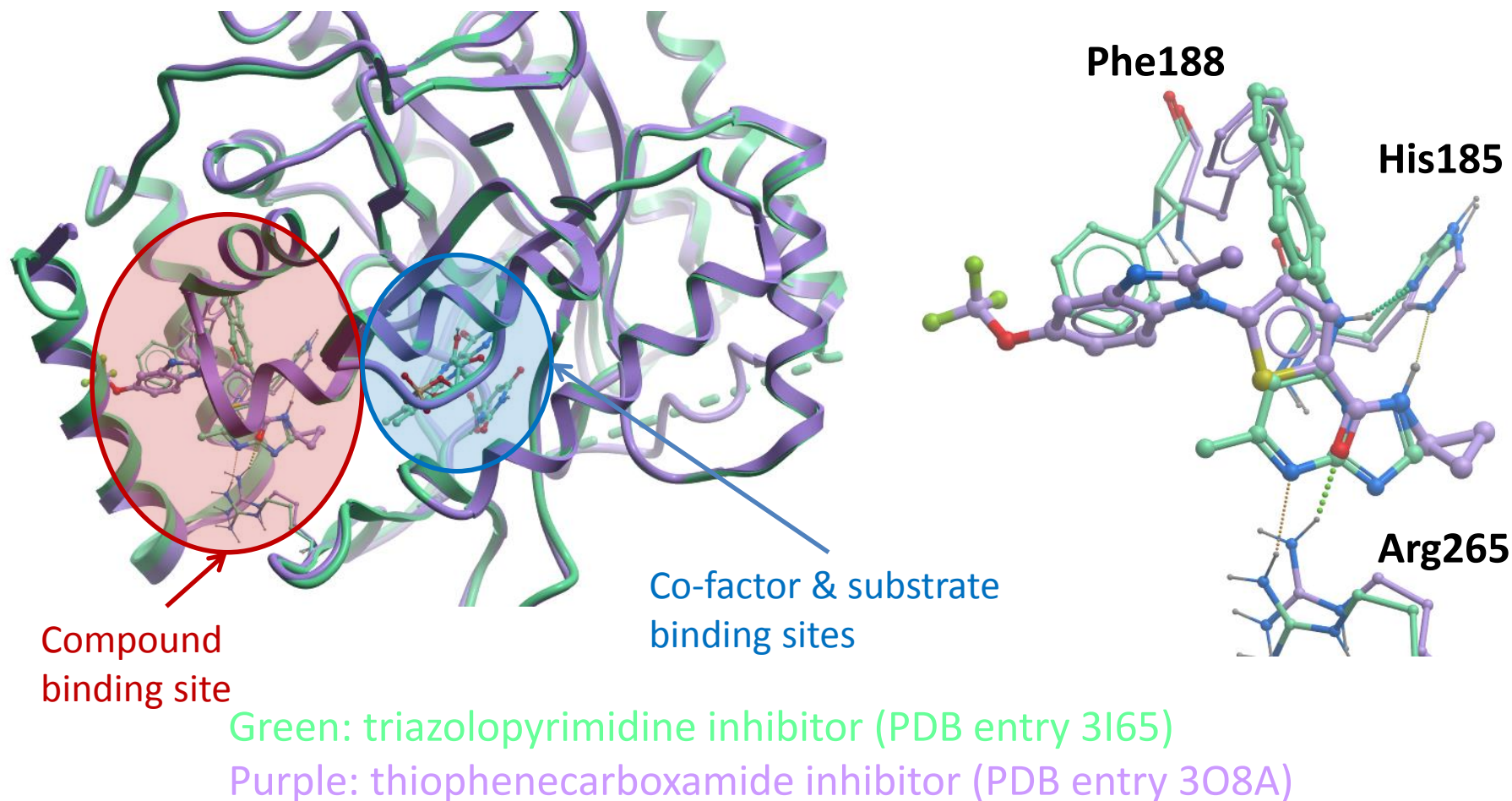


Generic structure for SAR set

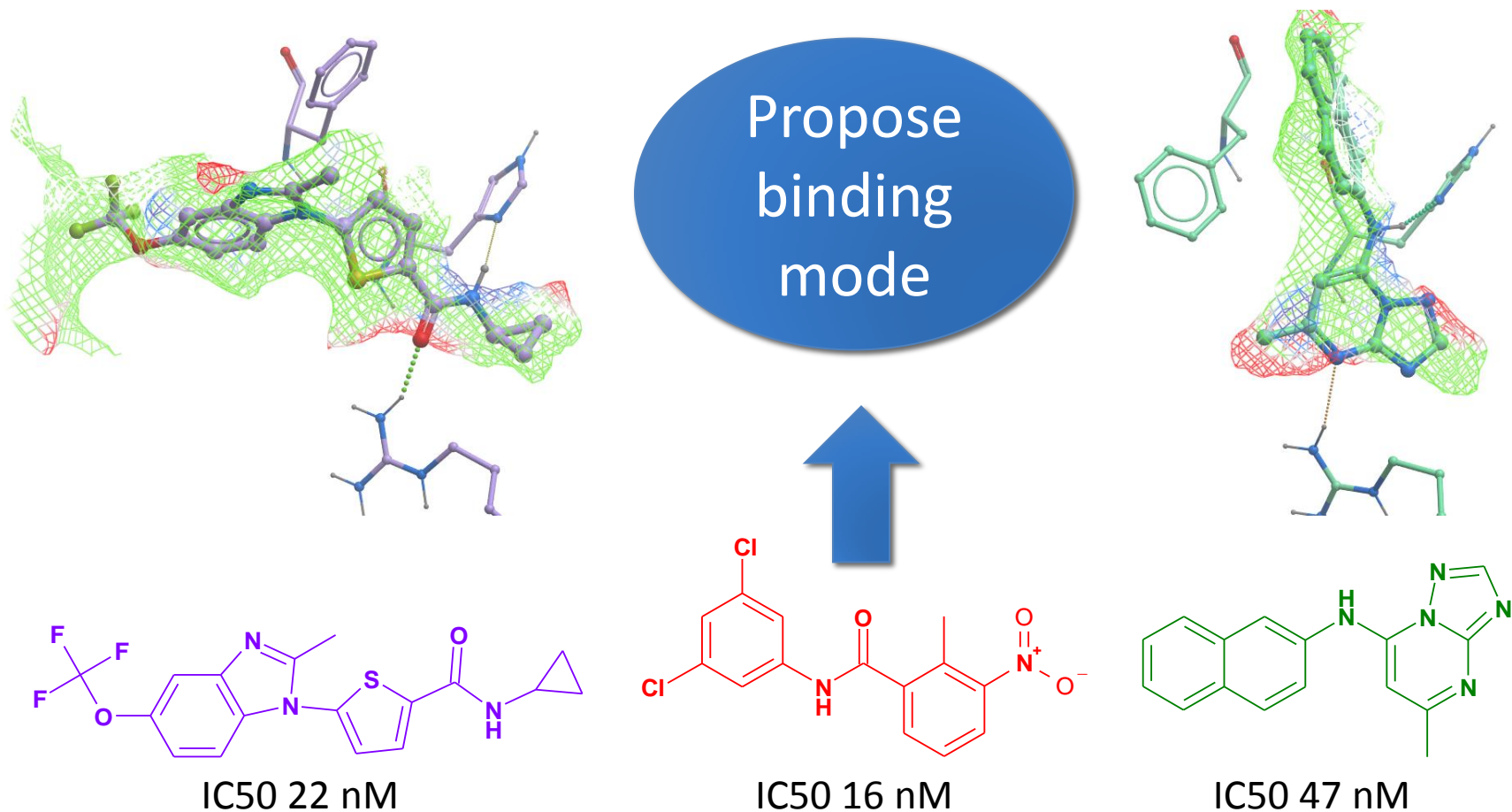


Selected candidate
Efficacious in *P. falciparum*
murine model; QD dosing

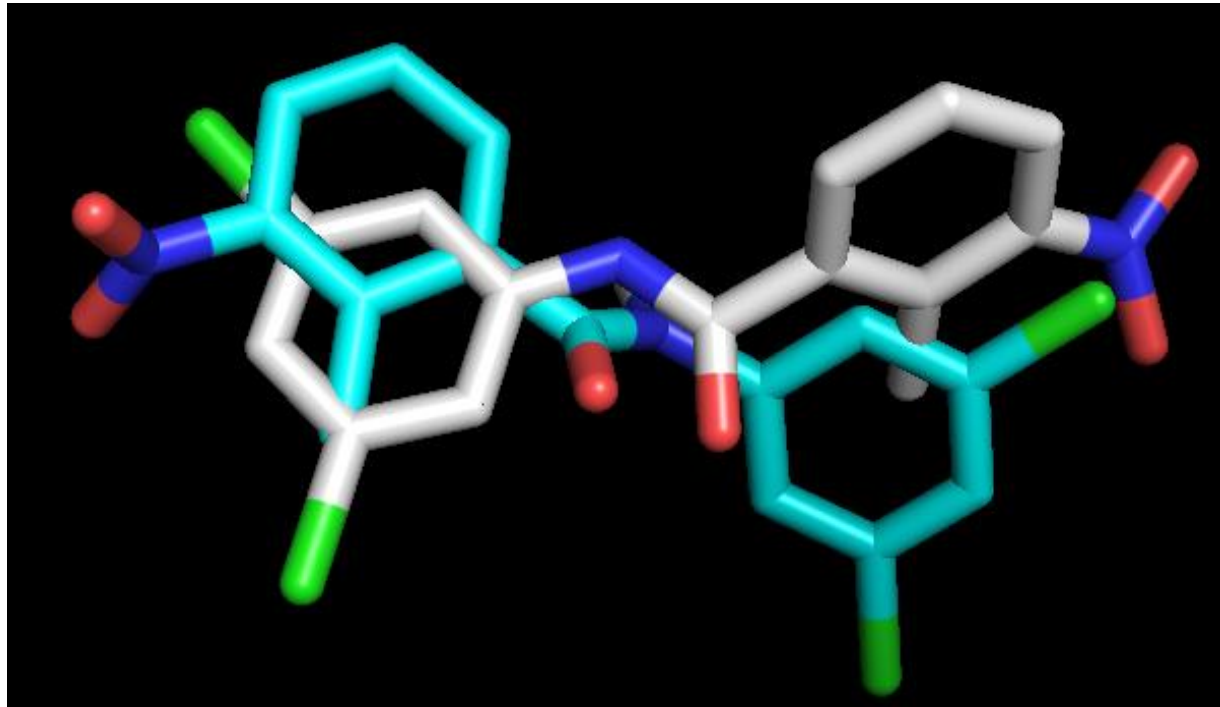
DHODH structures with different chemotypes highlight importance of protein plasticity



DHODH: Two chemotypes with crystal structures, one with unknown binding mode



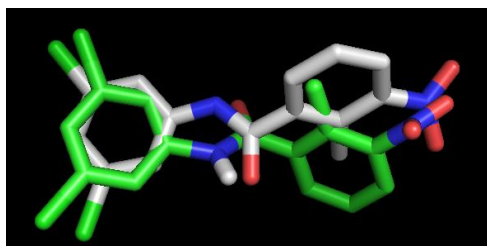
One Entry was “Flipped” (but used the right protein conformation)



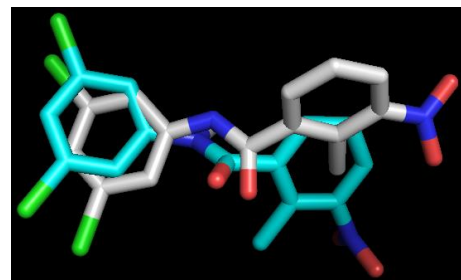
7.4 Å RMS

X-ray structure is in white

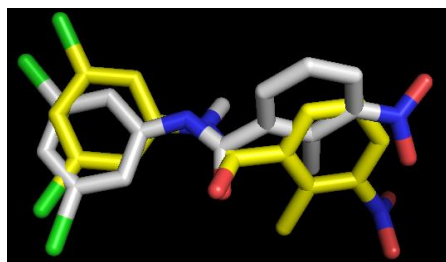
Other Entries Were Somewhat Closer



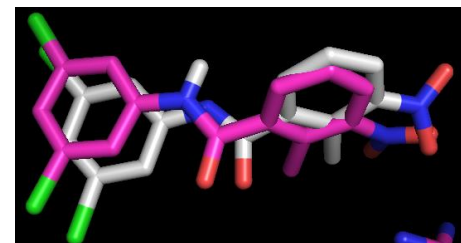
3.0 Å RMS



3.3 Å RMS



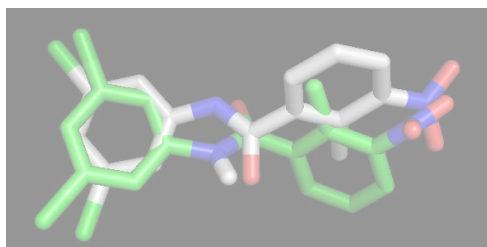
2.7 Å RMS



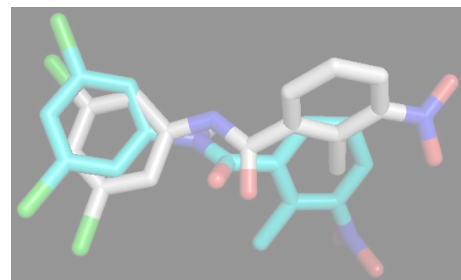
2.2 Å RMS

X-ray structure is in white

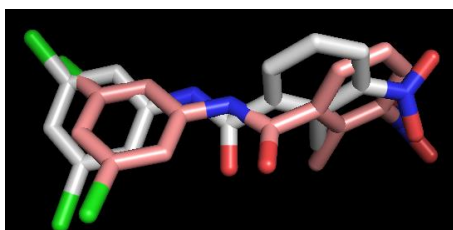
“Optimization” Doesn’t Always Optimize



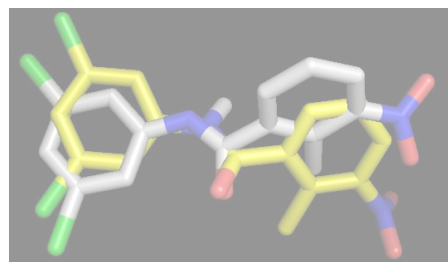
Minimized 3.0 Å RMS



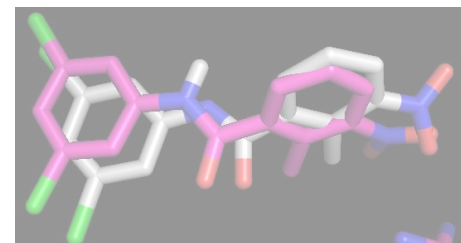
Minimized 3.3 Å RMS



Not Minimized 1.8 Å RMS



Minimized 2.7 Å RMS



Minimized 2.2 Å RMS

X-ray structure is in white

Drug discovery follow-up: Screening for new chemical matter for Malaria target DHODH

- 977 compounds from 2 submissions screened at 10 μM and 1 μM at the Phillips Research Group, University of Texas-Southwestern Medical Center
 - Overall winner: Paolo Tosco, Department of Drug Science and Technology, Torino, Italy
 - First runner-up: David Koes & Carlos Camacho, Department of Computational and Systems Biology, University of Pittsburgh
- Overall hit-rate across both submissions 6.2%
- Dose-response follow-up
 - $\text{IC}_{50} < 10 \mu\text{M}$ for 61 hits
 - 24 hits that do not belong to previously identified scaffolds but most are still close analogues to known anti-malaria compounds
 - Joint publication from one of the TDT teams and the TDT partner!

First TDT-enabled publication and disclosure of new chemical matter

A Teach-Discover-Treat Application of ZincPharmer: An Online Interactive Pharmacophore Modeling and Virtual Screening Tool

David Ryan Koes^{1*}, Nicolas A. Pabon¹, Xiaoyi Deng², Margaret A. Phillips², Carlos J. Camacho^{1*}

¹ Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, United States of America, ² Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, 6001 Forest Park Blvd, Dallas, TX, United States of America

* dkses@pitt.edu (DRK); ccamacho@pitt.edu (CJC)

Abstract

The 2012 Teach-Discover-Treat (TDT) community-wide experiment provided a unique opportunity to test prospective virtual screening protocols targeting the anti-malarial target dihydroorotate dehydrogenase (DHODH). Facilitated by ZincPharmer, an open access *online interactive pharmacophore search of the ZINC* database, the experience resulted in the development of a novel classification scheme that successfully predicted the bound structure of a non-triazolopyrimidine inhibitor, as well as an overall hit rate of 27% of tested active compounds from multiple novel chemical scaffolds. The general approach entailed exhaustively building and screening sparse pharmacophore models comprising of a minimum of three features for each bound ligand in all available DHODH co-crystals and iteratively adding features that increased the number of known binders returned by the query. Collectively, the TDT experiment provided a unique opportunity to teach computational methods of drug discovery, develop innovative methodologies and prospectively discover new compounds active against DHODH.

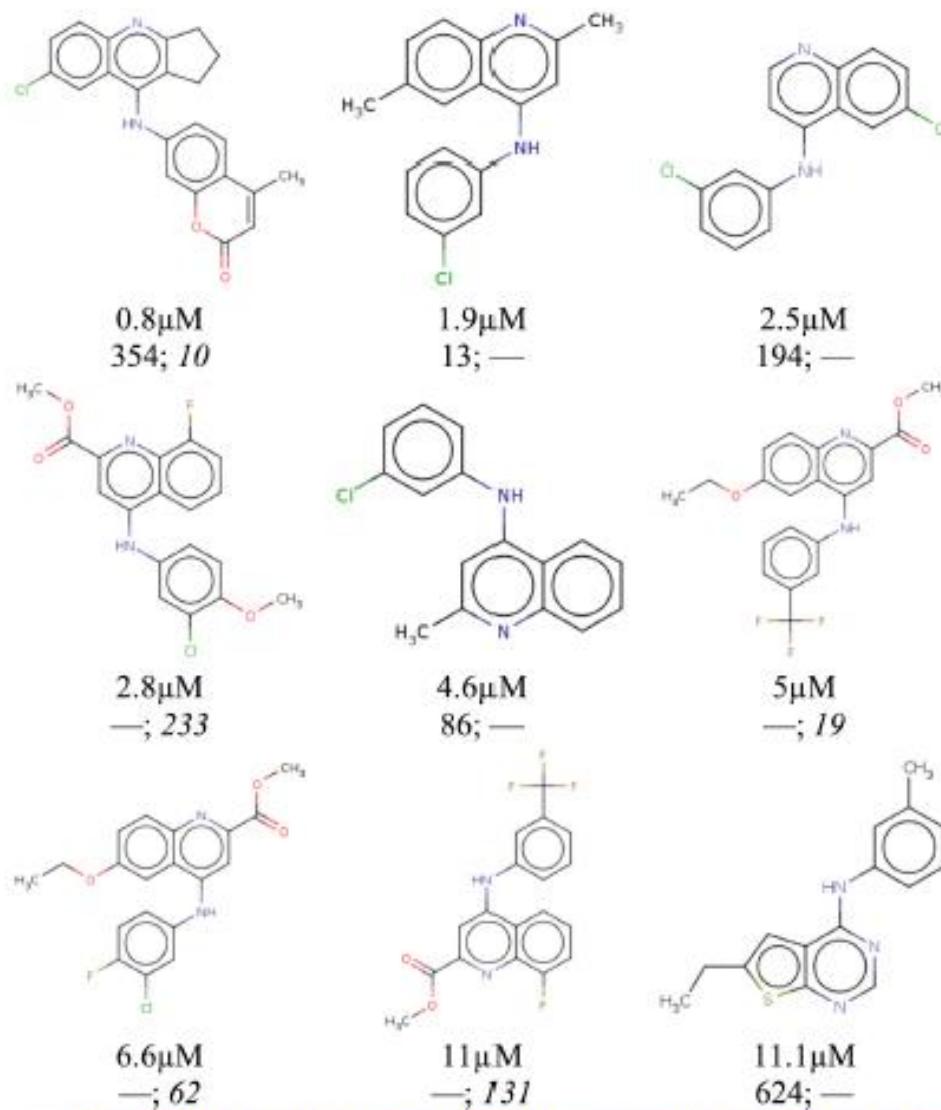


Fig 7. The nine active compounds with unpublished chemical scaffolds identified by the exercise.

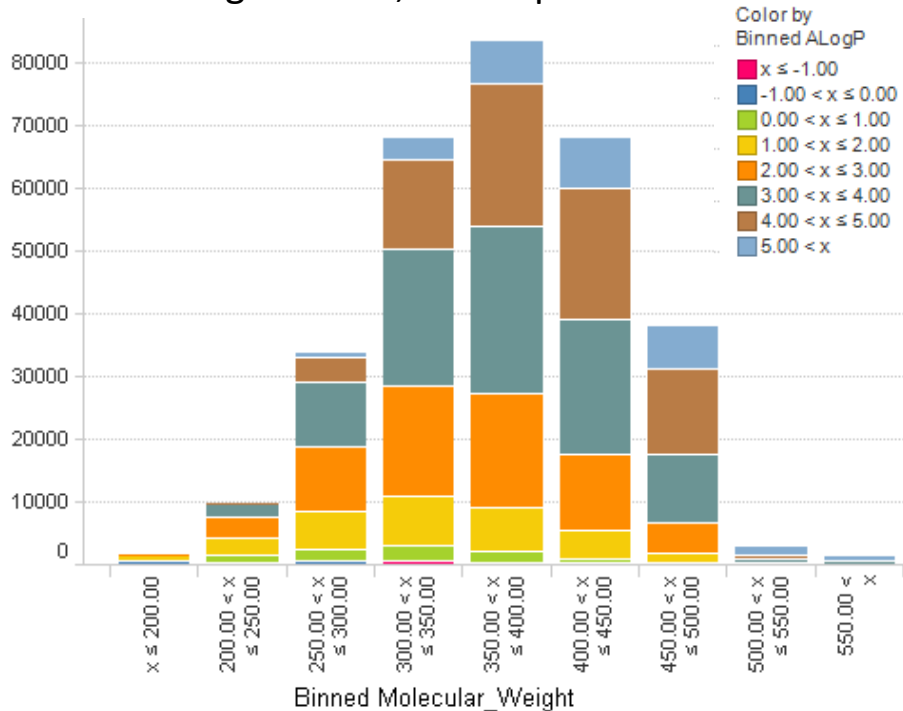
Compounds are shown with their measured IC_{50} and their rank within the Vina ranked list and the custom scored list (italic). The one compound present in both lists was the only novel compound to demonstrate sub-micromolar affinity. All compounds have a Tanimoto similarity coefficient of less than 0.16 with respect to the ligand used to define the pharmacophore as computed by OpenBabel with FP2 fingerprints.

Almost completed challenge: Workflow to analyze HTS data & build predictive models for further hit finding

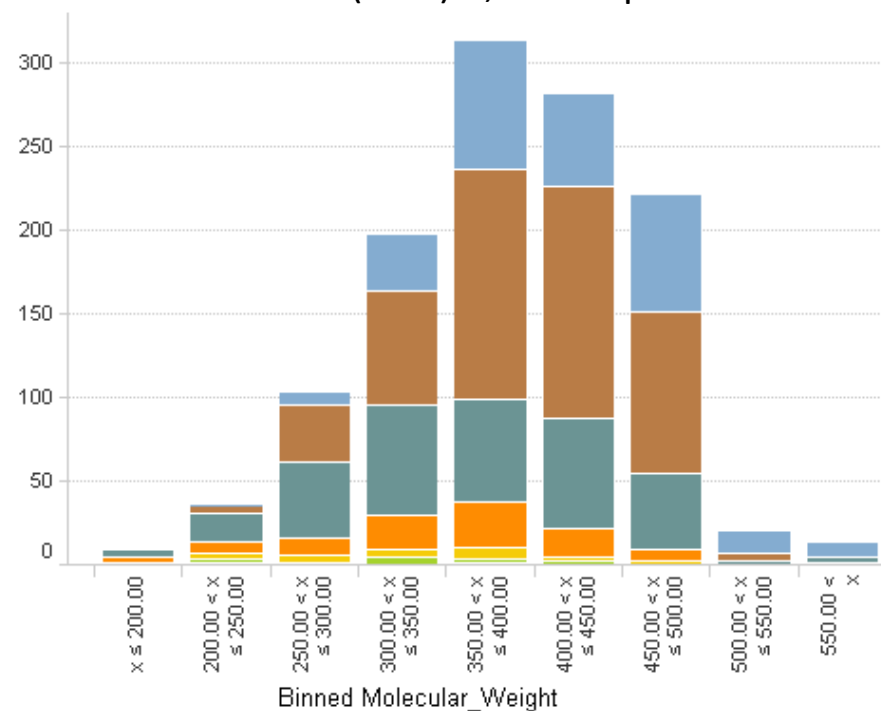
- Malaria case study with HTS hit list from phenotypic screen and training set of compounds with confirmed IC50 data
 - Partnership with Anang Shelat and Kip Guy (St. Jude)
 - Held-out test data: IC50 data for 1,056 compounds
- Workflow to start from HTS hit list and prepare models for further hit finding
 - Analysis of single concentration screening data: hit list triaging, selection of compounds for IC50
 - Building and validating a predictive activity model from training set with confirmed IC50 data, including predicting activity in a held-out test set
- Drug discovery follow-up to identify new hits
 - Apply predictive model to catalogue of commercial compounds and prioritize based on predicted activity
 - At least 100 compounds will be acquired and tested in Malaria whole-cell assay

St. Jude HTS data-set for Malaria phenotypic screen

Screening Set: 305,568 cmpds



Validated Hits (IC50): 1,189 cmpds



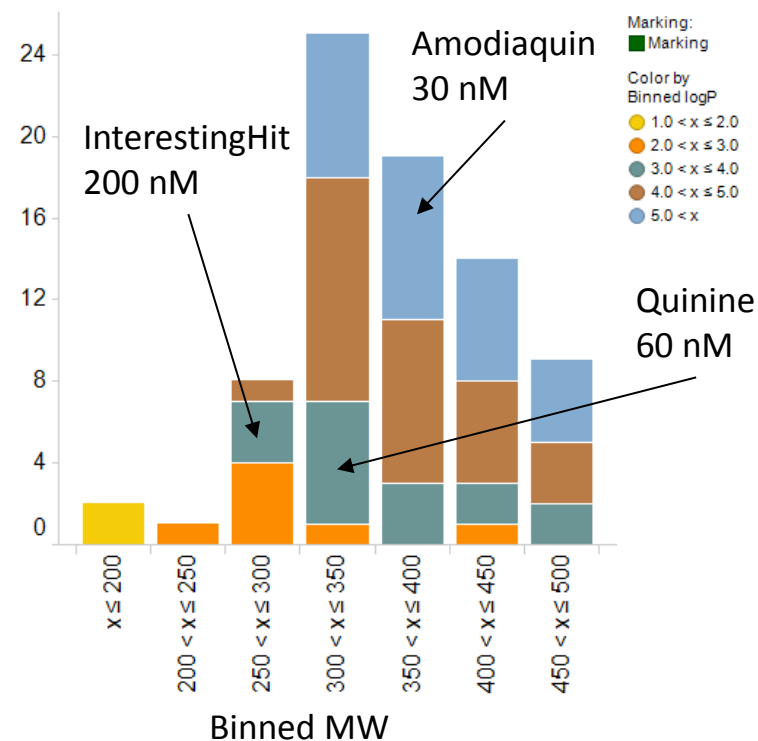
Validated hits show shift to higher MW and increased hydrophobicity compared to the originating screening set

Nature, 2010, 311-315; Chemistry & Biology, 2012, 116-129

Note: also data available from GSK and Novartis screening efforts

Screening for new chemical matter: 68% hit-rate in commercial compound set

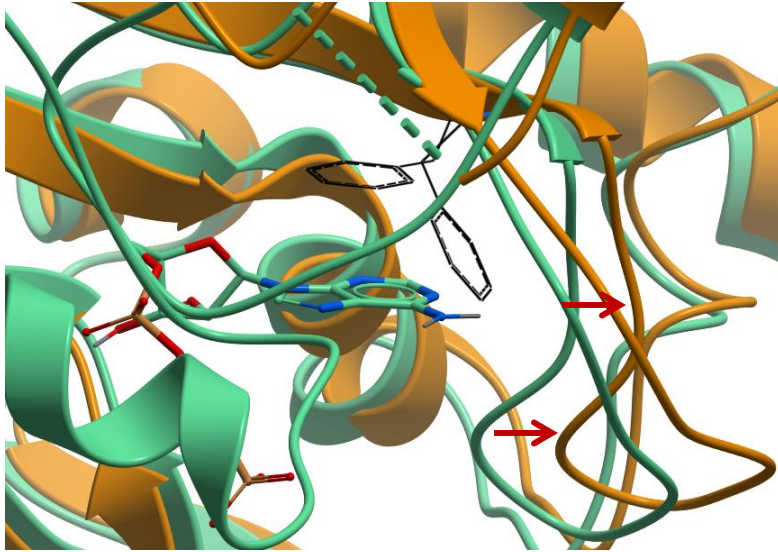
- 114 molecules from two submissions assayed in dose-response: 78 compounds that give a fit with efficacy >40% and at least 2 points above the noise
 - Overall winner: Sereina Riniker & Greg Landrum, Novartis (using iPython notebook!)
 - Prediction-award winner: Santiago D Villalba and Floriane Montanari, Institute of Molecular Pathology & Department of Pharmaceutical Chemistry, University of Vienna, Austria
 - 4 compounds present in both lists; each submission had a known anti-malarial (quinine and amodiaquin)
- “The hit rate in this experiment is extremely high => however, many of these hits are close analogs from the SJ data set”
- Interesting SAR findings (to be published) and some good low MW starting points



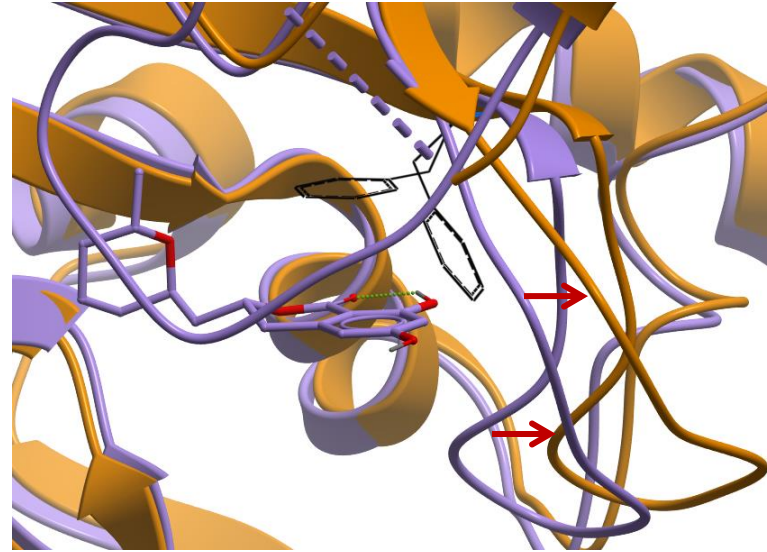
A challenge with some valuable lessons: structure-based drug discovery for *Pf* Lysyl tRNA synthetase

- Malaria case study on *Plasmodium falciparum* Lysyl tRNA synthetase, *Pf* Krs1, with one known selective inhibitor and an apo-structure in public domain
 - Partnership with Chris Walpole, Structural Genomics Consortium (Toronto, Canada)
 - Held-out test data: Screening data for compounds in TCAMS (Tres Cantos Anti Malarial Set)
 - Note: Data did not exist at the time the challenge launched; delays in getting assay validated; only one hit from TCAMS => do not launch challenge if you do not have the held-out test data
- Workflow to start from *Pf* Krs1 apo-structure and prepare models to select new compounds for screening
 - Preparation of crystal structure for virtual screening, including generation of binding pose for Cladosporin
 - “Computationally-derived binding poses for cladosporin could be compared with a co-crystal structure if that becomes available in the right time frame” => Structure available in PDB July 16th, 2014 (2 months after announcement of winners)
 - We did not point out a protein flexibility challenge and 1 of the 3 submissions missed the fact that the apo-structure could not be used “as-is” => need to annotate tutorials to flag cases with a significant oversight

The *Pf* apo-structure does not accommodate ATP



ATP-binding site of Ksr1 from human with ATP (green) overlaid with apo *Pf* (orange) shows slight differences in purine-binding loops. Major difference in rotamer for Phe342 in *Pf* apo structure would prevent proper binding of ATP.



ATP-binding site of *Pf* Ksr1 cocrystallized with cladosporin (purple) compared to apo *Pf* Ksr1 (orange) shows that ligand induces same binding site conformation as ATP does for human Ksr1: purine binding loops move in and Phe rotamer allows for ligand binding.

TDT Summary

- Tutorials with workflows available for education and drug discovery
- New chemical matter discovered and disclosed
- Datasets created from prospective predictions
 - What are the predictions for these compounds from the workflows from all the other participants who did not select these compounds?
 - Note: 51 compounds screened for one of the other challenges gave 0% hits. Would the other workflows for that challenge have predicted that?
- Challenge examples highlighting valuable lessons for SBDD: binding site flexibility, use of apo-structures, challenging atom-types, hit-rate vs novelty
- What is next?

Opportunities to improve and innovate the TDT concept

- What we can do better
 - Annotate tutorials: abstracts, keywords & quality, especially with respect to any serious oversight
 - Continuity for challenges
- Bringing TDT to the next level
 - Pro-actively gathering and creating workflows and tutorials
 - Interns-in-industry to move published methods to open access platforms
 - Published methods that address general drug discovery challenges and assessment that open access tools will work
 - Sabbaticals-from-industry to create tutorials and teach
 - Professionally maintained server & online community
 - Accounts available with environment set-up to run the highest rated workflows in tutorial or discovery setting
- D3R partnership can help with the improvements and innovation!

Partnership between TDT and D3R

- TDT will host competition data (training set, held-out test sets, screening results), tutorials, and workflows with D3R
 - TDT will gain a robust web platform to host the data and results for its past and future competitions
 - TDT will retain its current website at www.TDTproject.org
- D3R will gain drug discovery related datasets and a diverse set of tutorials and workflows
- TDT will have an option to partner with D3R on challenges, adding a tutorial/workflow component to D3R challenges
- D3R will also provide an outlet for TDT to announce future competitions and to engage with the scientific community
- Interns-in-industry? Sabbaticals-from-industry?

Summary

- TDT impact
 - Tutorials with workflows available for education and drug discovery
 - New chemical matter discovered and disclosed
 - Datasets created from prospective predictions
 - Challenge examples highlighting valuable lessons for SBDD: binding site flexibility, use of apo-structures, challenging atom-types, hit-rate vs novelty
- Opportunities to improve and innovate
 - Leverage partnership with D3R: access to data and (annotated) tutorials & workflows; continuity in challenges
 - Access workflows from industry and general industry experience through “interns-in-industry” and/or “sabbaticals-from-industry” effort

Acknowledgements

- TDT colleagues: Rommie Amaro (UCSD), Jane Tseng (Natl. Taiwan Univ.), Wendy Cornell (Merck), Pat Walters (Vertex), Emilio Esposito (exeResearch)
- TDT partners: Kip Guy & Anang Shelat (St. Jude); Meg Phillips (University of Texas Southwestern Medical Center); Mike Pollastri (Northeastern University); Chris Walpole, (Structural Genomics Consortium, Toronto, Canada); Kayode K Ojo & Wesley C Van Voorhis (CERID: Center for emerging and re-emerging infectious diseases) and Dustin Maly & Erkang Fan (University of Washington); Klaus Gubernator (eMolecules)
- TDT sponsors: KNIME, OpenEye, Cresset, CCG, Simulations Plus, Schrodinger, CDD, Gilead, Vertex, Novartis, Merck, ChemDiv, ChemBridge, Enamine, Vitas-M, exeResearch, ACS COMP Division, ACS Innovative Projects Grant
- TDT participants
- D3R