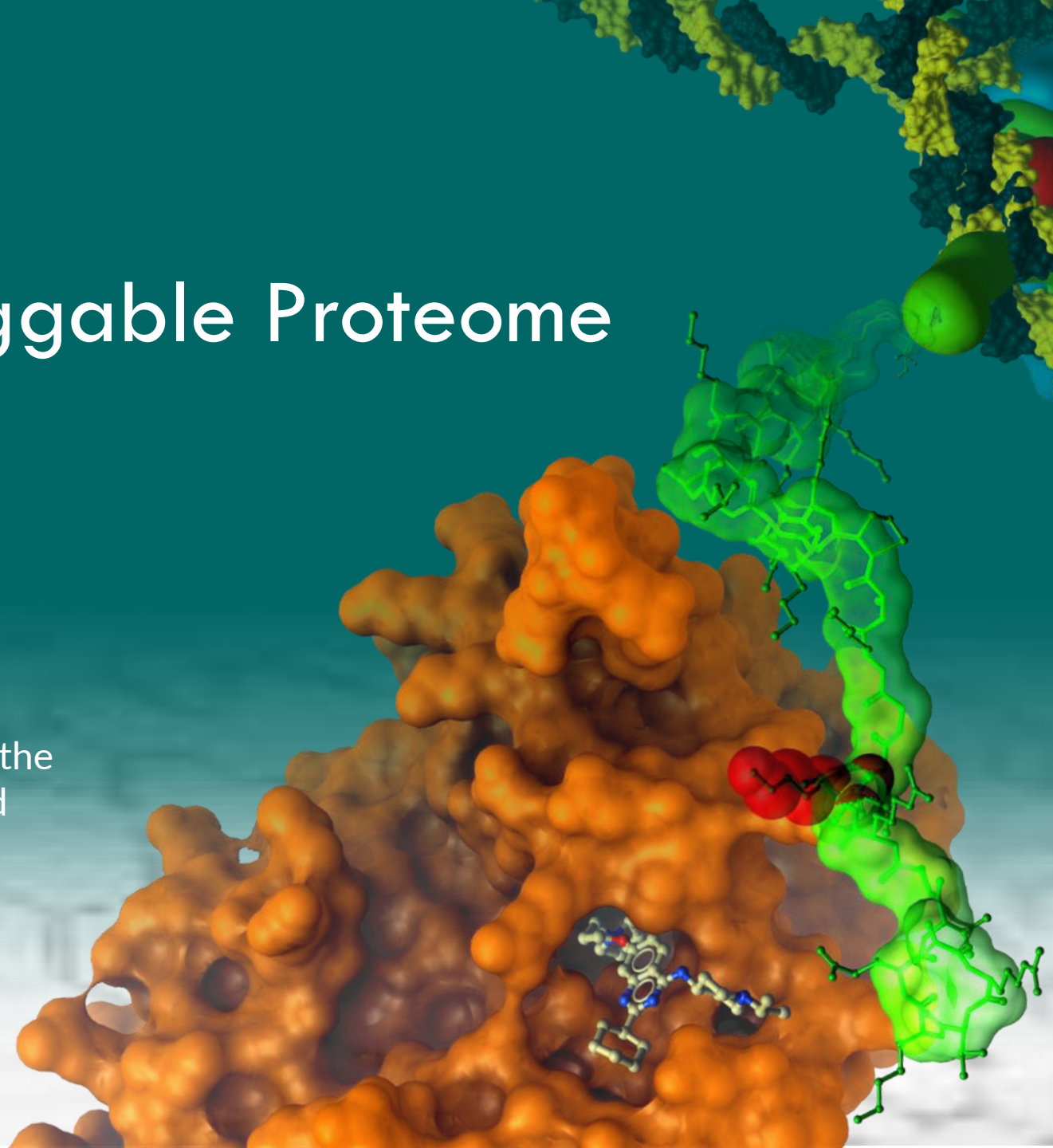
Expanding the Druggable Proteome

Stefan Knapp
Goethe-University Frankfurt am Main
Institute of Pharmaceutical Chemistry

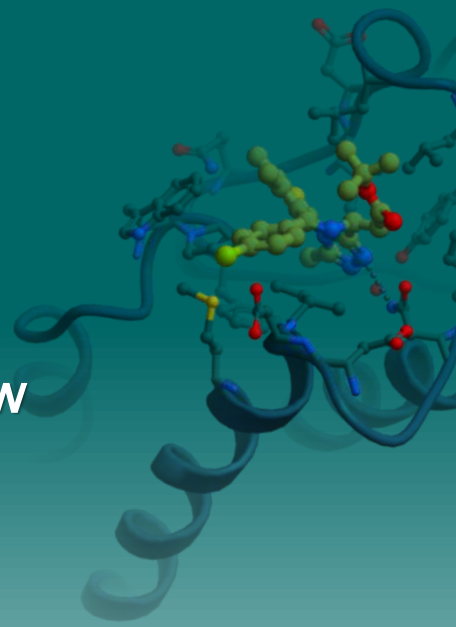
Advancing Drug Discovery: A Webinar Series of the National Academies of Sciences, Engineering and Medicine
November 14, 2024



Introduction to our lab

SGC

- Operations started in June 2004
- Aggregate funding >450 MEuro
- PPP: Government agencies, EU, charities & leading pharma companies
- +200-strong team in Toronto, Frankfurt, Chapel Hill, McGill, KI and UCL
- Open Access Policy:
  - **Promptly placing results, reagents and know-how in the public domain**
  - **SGC scientists never file patents**



Main outputs:

- **High Throughput Structural Biology (>4000 structures deposited)**
- **Renewable Antibodies/Binders**
- **Patient-Cell Derived Assays**
- **Chemical Probes (~190) and Chemogenomics Libraries (1000 targets)**

GenomeCanada · Innovative Medicines Initiative · Ontario · Bill & Melinda Gates foundation · The Eshelman Foundation Wilmington • North Carolina · Boehringer Ingelheim · janssen · BAYER · Pfizer · Takeda · Genentech · Bristol Myers Squibb™ · MERCK

Evolving science and partners to address pressing global needs

SGC

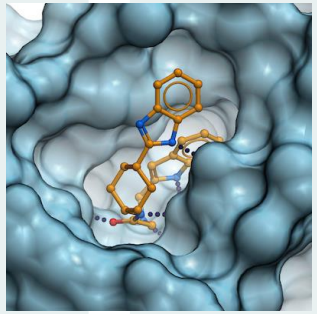
2004 | 2007 | 2011 | 2015 | 2020 | 2025

Phase I | Phase II | Phase III | Phase IV | Phase V

Scientific Focus

Structural Genomics

Human Protein Families
Human Parasite Proteins



Chemical Genomics

TARGET 2035

Bio-active Chemical Tools

Data driven DD

Machine Learning

Patient-derived disease assays

Industry Partners

GlaxoSmithKline

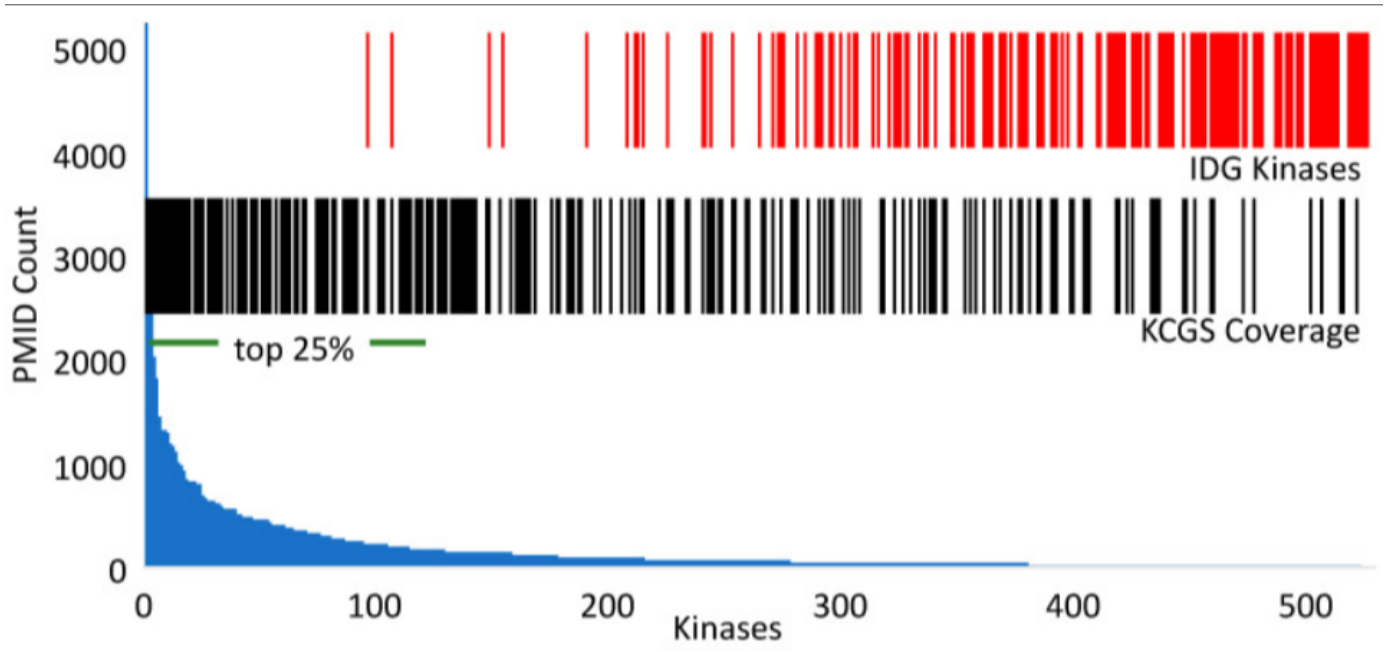
GlaxoSmithKline, Pfizer, MERCK, NOVARTIS

GlaxoSmithKline, Takeda, Pfizer, janssen, Lilly, abbvie, Boehringer Ingelheim, BAYER, NOVARTIS

Takeda, Boehringer Ingelheim, janssen, MERCK, Pfizer, abbvie, BAYER, MSD, NOVARTIS

Genentech, janssen, MERCK, Pfizer, Boehringer Ingelheim, Takeda, BAYER, Bristol Myers Squibb

Technology defines the druggable proteome



TARGET 2035

- Established target families are larger than initially predicted
- Assay technology has not been developed for a larger fraction of established families

Size of the druggable families – for example kinases ?

- Manning et. al. : 514 kinases
  - ePK: 478
  - Eukaryotic like kinases (eLK), differ in substrate binding lobe
  - Atypical fold (aPK)
- Kinhub Web resource: 522 proteins
- IDG dark kinases: 163
- Uniprot set: 684

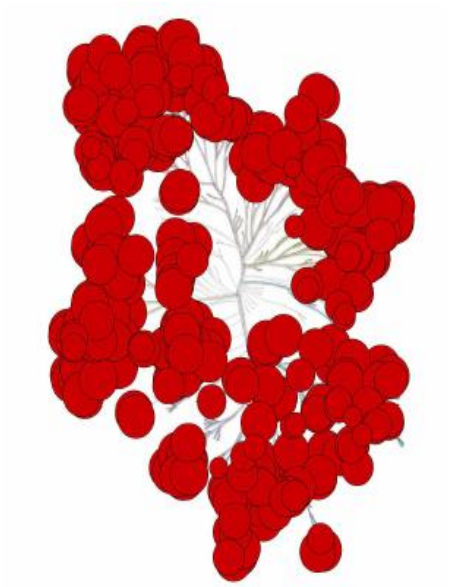
Druggable proteome is rapidly expanding

- Druggability of many7if not most protein families unknown
- Binding sites in protein-protein interfaces
- Potential of new technologies and modalities such as molecular glues and/or PROTACs ?

Chemical Tools for poorly explored target families

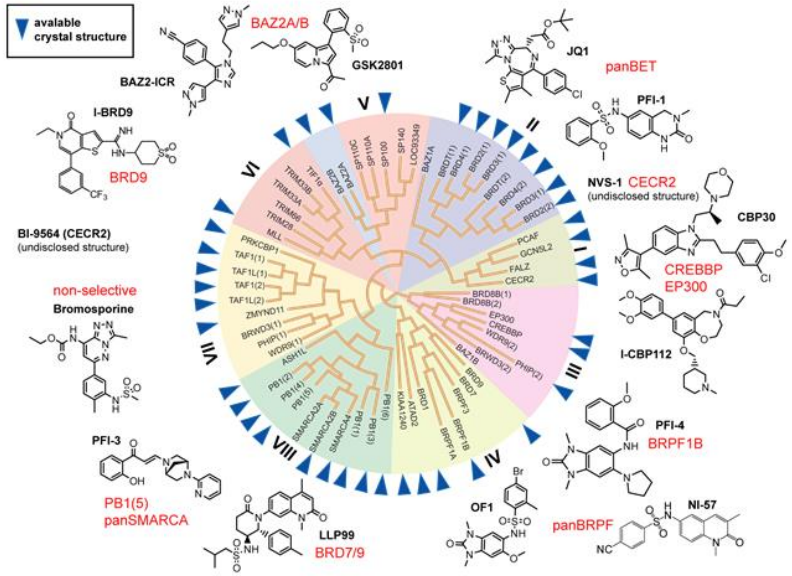
Kinases

~300 high potency CG cpds
~ 100 chemical probes



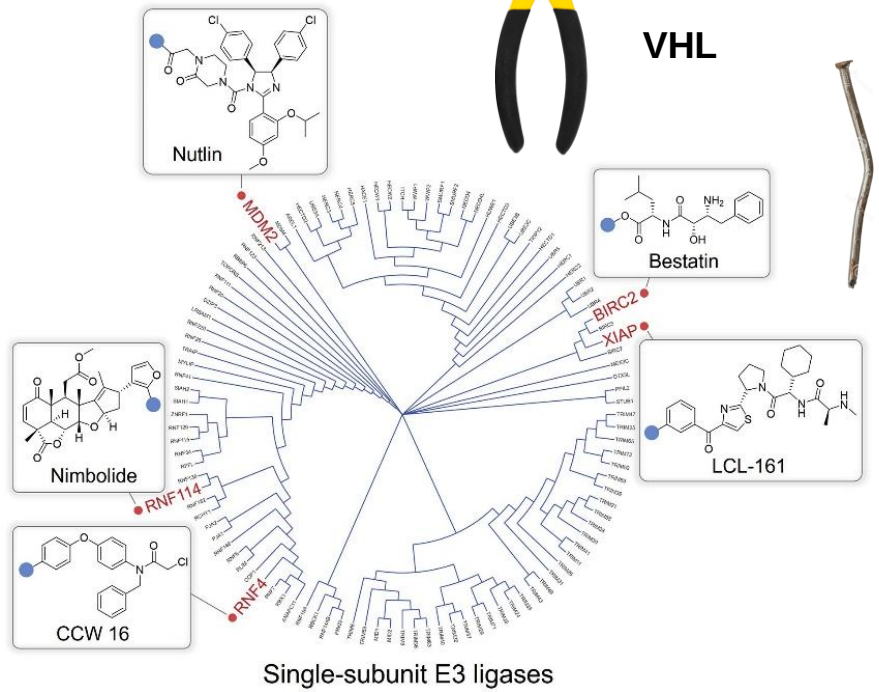
Bromodomains

chemical probes for most subfamilies



>90% of PROTACs use CRBN or VHL ligands

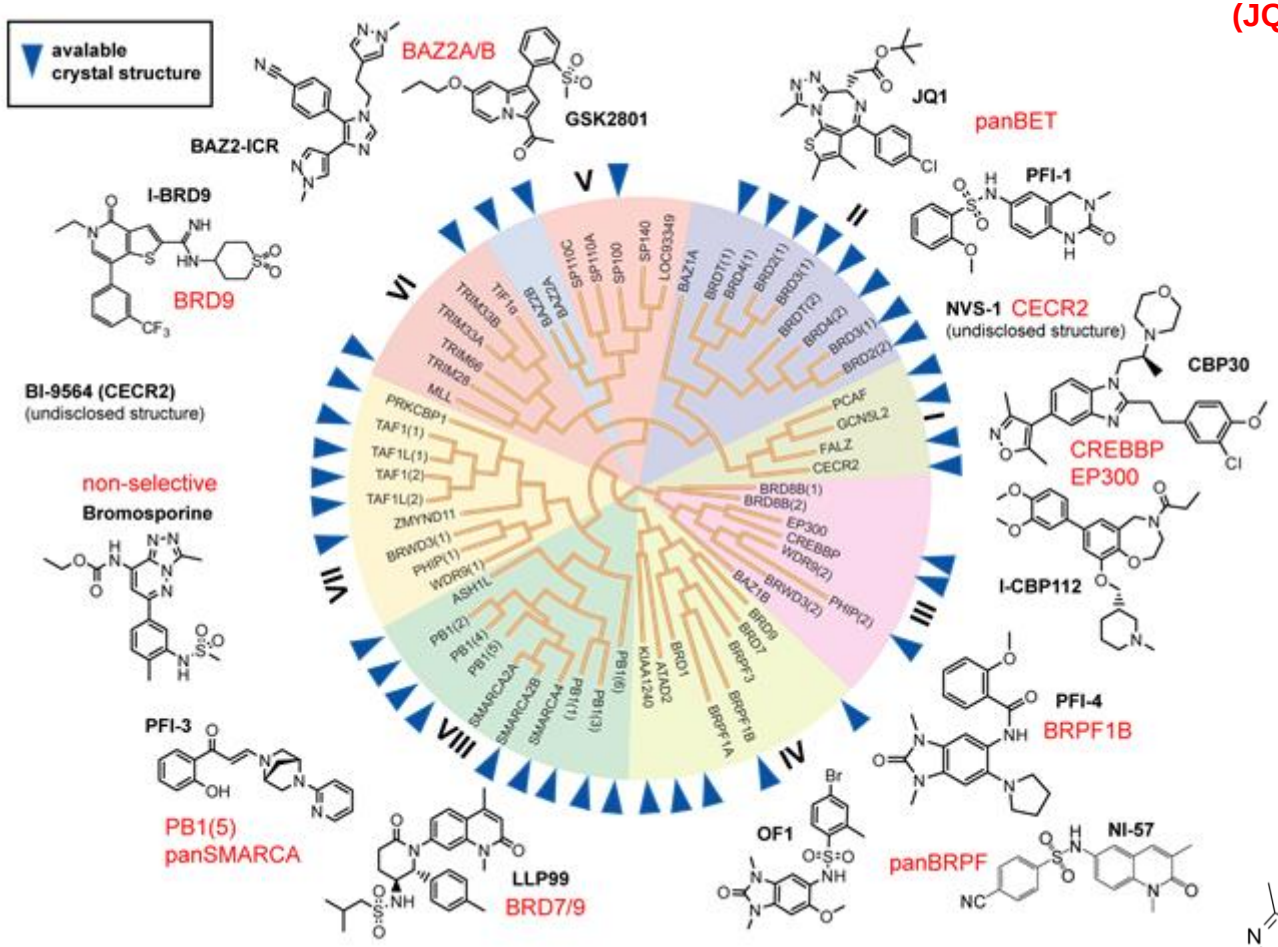
E3 Ligases



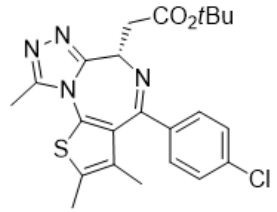
Single-subunit E3 ligases

Will the pre-competitive release of chemical probes limit commerciallization?

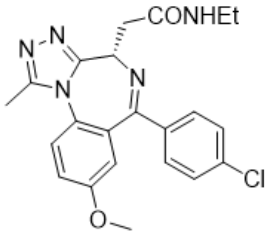
Comprehensive bromodomain inhibitor sets



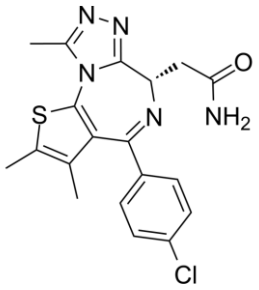
Pan BET BRD chemical probe (JQ1)

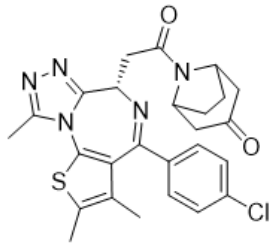


(+)-JQ1

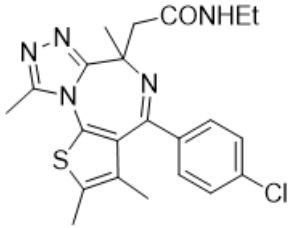


I-BET GSK



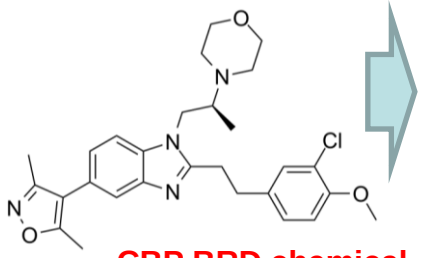


BAYER

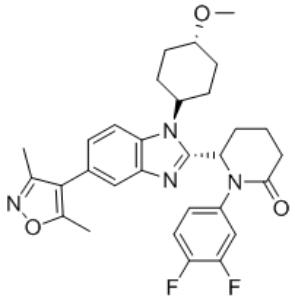


Constellation

44 clinical studies



CBP BRD chemical probe (CBP30)



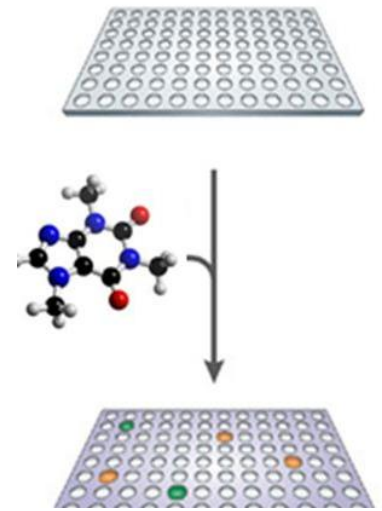
2 clinical studies on CBP inhibitors in oncology (Inobrodib)

CBP BRD clinical candidate drug

How to explore new target areas ?

Assay technology:

- **Fragment based**
- **Experimental screening (direct binding assays) (ASMS)**
- **Experimental + ML/AI: DEL-ML**
- **In silico**



How do we best validate most successful screening technology?

Data for AI/ML predictions?

nature

Explore content | About the journal | Publish with us | Subscribe

nature > editorials > article

EDITORIAL | 17 May 2023

For chemists, the AI revolution has yet to happen

Machine-learning systems in chemistry need accurate and accessible training data. Until they get it, they won't achieve their potential.

In silico benchmarking: CACHE

- Google's DeepMind demonstrated in 2021 that AI could predict the structure of most soluble proteins with accuracy comparable to experimental determination.
- CASP set the stage for this earthquake in the field
  - CASP is an international benchmarking competition for protein structure prediction
  - CASP has been running for 30 years

CACHE Is a Prospective Hit Finding Competition

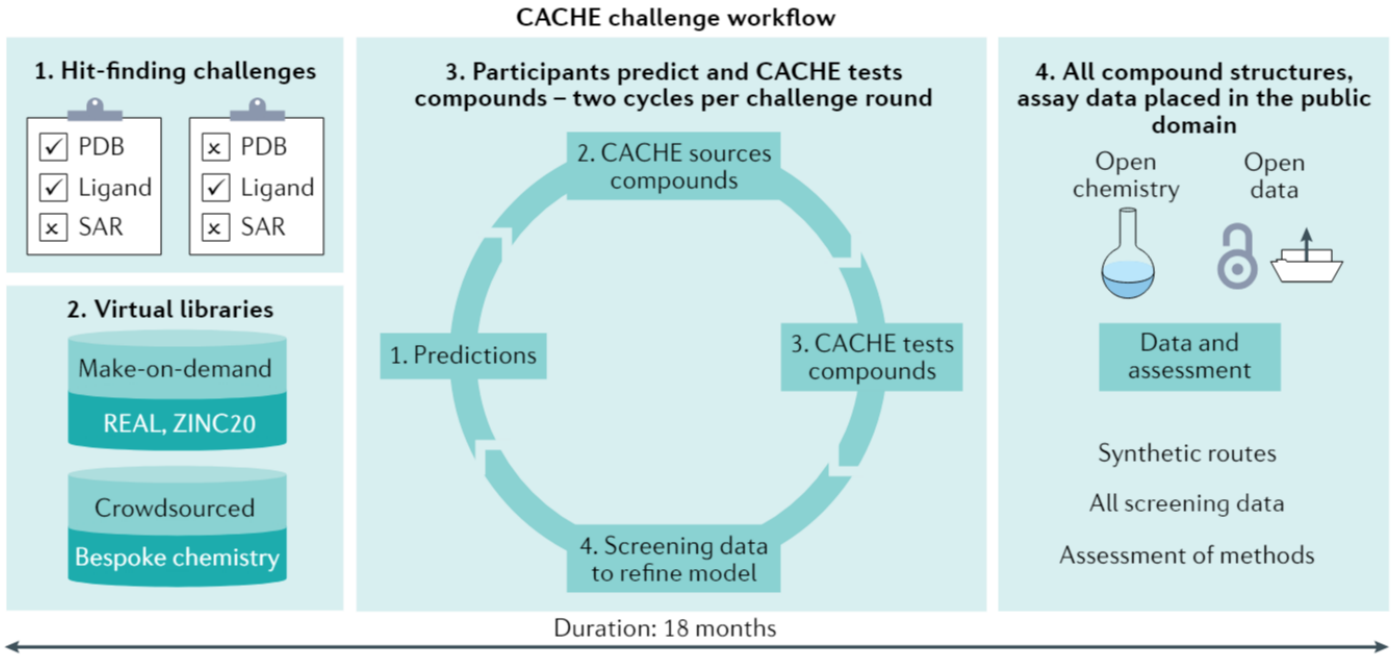
The next frontier is the computational design of drug-like ligands for any given protein target.

CACHE (Critical Assessment of Computational Hit-finding Experiments) is modeled after CASP

How are we helping drive progress? Organizing benchmarking

NATURE REVIEWS | CHEMISTRY

CACHE (Critical Assessment of Computational Hit-finding Experiments): A public-private partnership benchmarking initiative to enable the development of computational methods for hit-finding



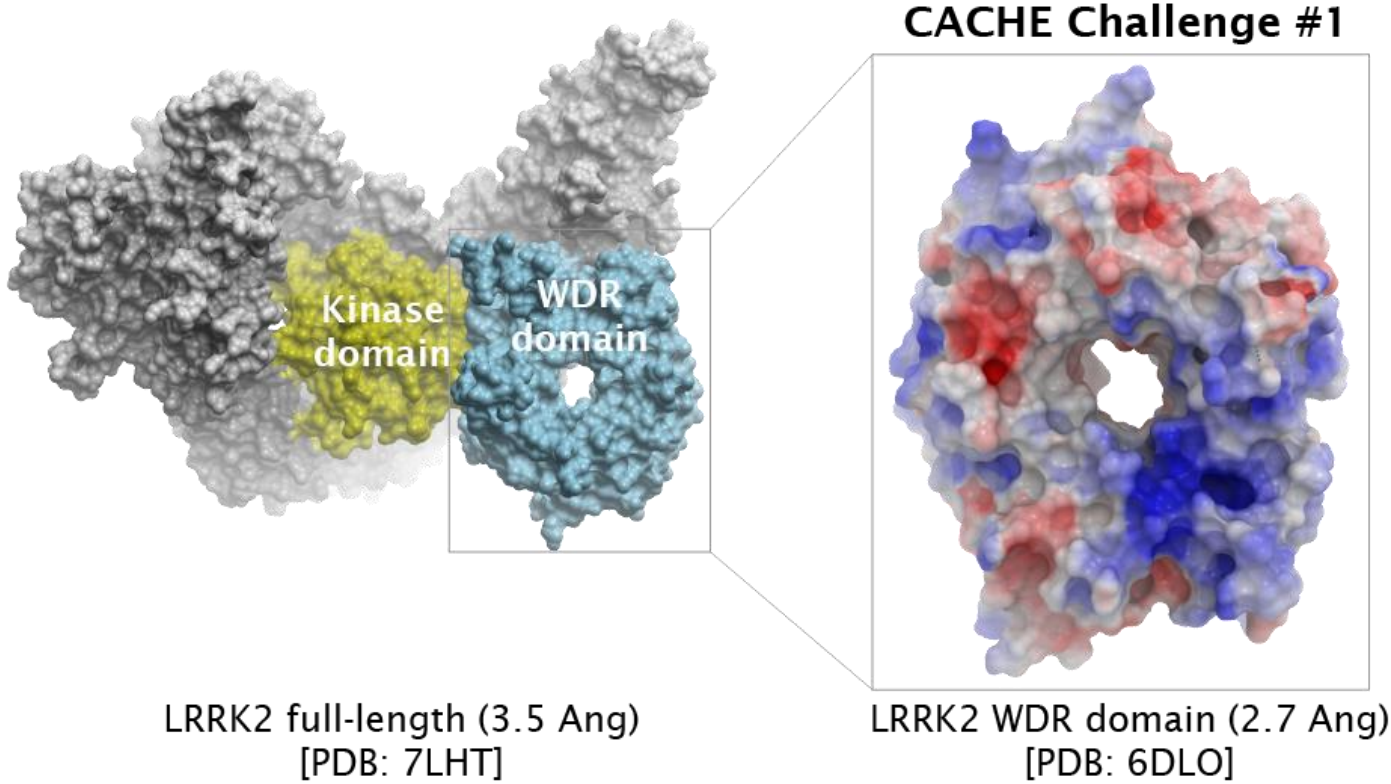
CACHE is a unique initiative that aims at engaging participants in 'hit finding' challenges which consist of the following steps:

1. Nomination of targets by CACHE
2. Curation of a virtual compound library by CACHE
3. Prediction of hits by participants and experimental testing by CACHE
4. All compounds, structures and assay data publicly available

SGC

Challenge #1: WDR domain of LRRK2

- LRRK2 is the most commonly mutated gene in familial Parkinson's Disease
- Drug candidates targeting the kinase domain are in preclinical or early-stage clinical trial
- Most kinase inhibitors stabilize the closed form of LRRK2, associated with the pathogenic formation of LRRK2 filaments
- The WDR domain in LRRK2 may be important for recruiting LRRK2 signalling partners or for binding to tubulin
- Targeting the WDR domain of LRRK2 is an attractive and underexplored alternative to kinase inhibitors



Deniston et al. Nature 2020

Challenge #1: WDR domain of LRRK2

Call for applications closed on February 1st 2022

35 Applications

United States - 11
Canada - 5
China - 4
Germany - 3
France - 2
Japan - 2
Brazil - 1
Czech Rep - 1
Denmark - 1
Estonia - 1
India - 1
Italy - 1
Sweden - 1
UK - 1

Highlights

- Top players in the field
- Top universities
- Mix of commercial and in-house software
- Mix of physics-based and AI methods, often combined
- Mix of approaches (fragments, pharmacophore, docking, generative) often combined

Areas

- Academia - 25
- Biotechs – 7
- Government – 2
- Pharma -1

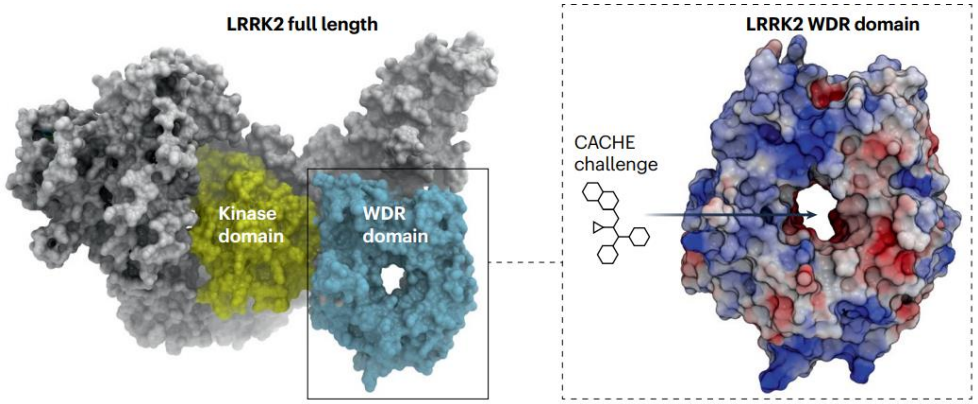
CACHE Challenge #1

In silico hit ID for less precedented targets

When can AI deliver the drug discovery hits?

The CACHE hit-finding competition highlights the potential of AI to identify small molecules that bind to hard-to-drug targets – and the long road ahead for these computational screening approaches.

By Asher Mullard

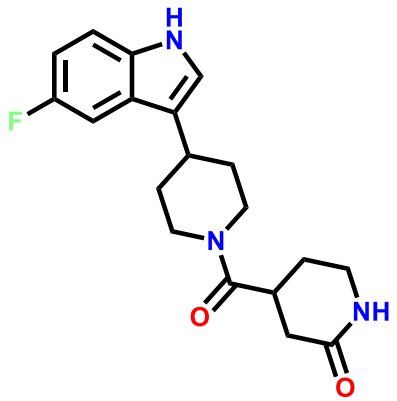


nature reviews drug discovery

https://www.nature.com/articles/d41573-024-00036-0

Challenge #1

- 23 teams predicted >2000 cmpds
- Overall <1% hit rate
- 20-70 µM potency range



- High-throughput docking as input for generative design, selection from Enamine RealSpace & filtering for drug-likeness
- Best hit: LRRK2 SPR K_D 48 µM
- 3 analogs providing first SAR
- Hit expansion via Enamine RealSpace for additional SAR

https://cache-challenge.org/results-cache-challenge-1

SGC

Developing a Chiral ASMS Platform for Proteome-Wide Hit Discovery and Characterization

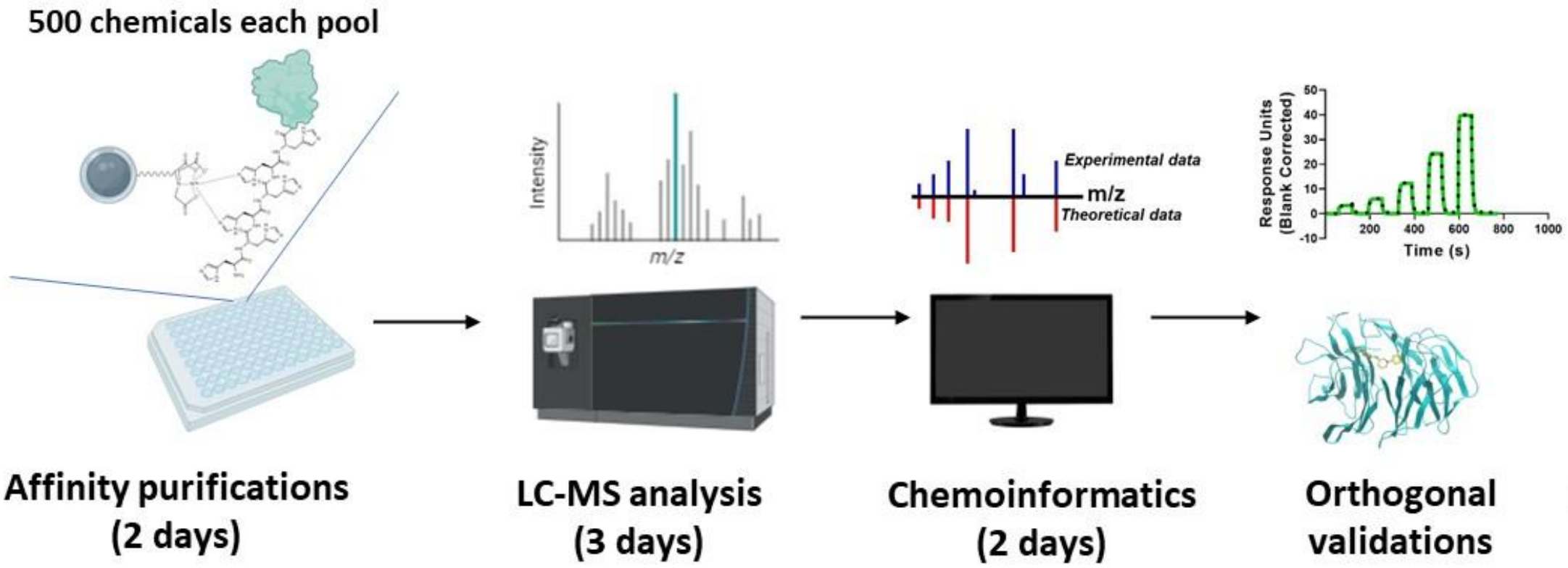
Levon Halabelian (co-lead)

Jianxian Sun

Xiaoyun Wang
(PDF)

Diwen Yang
(PhD)

Experimental Screening: ASMS @ SGC

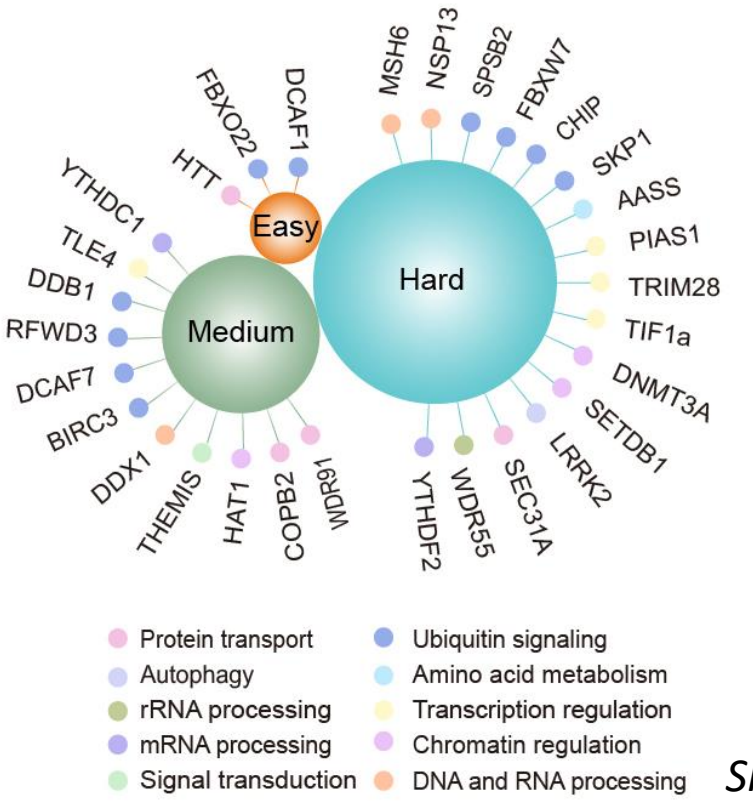


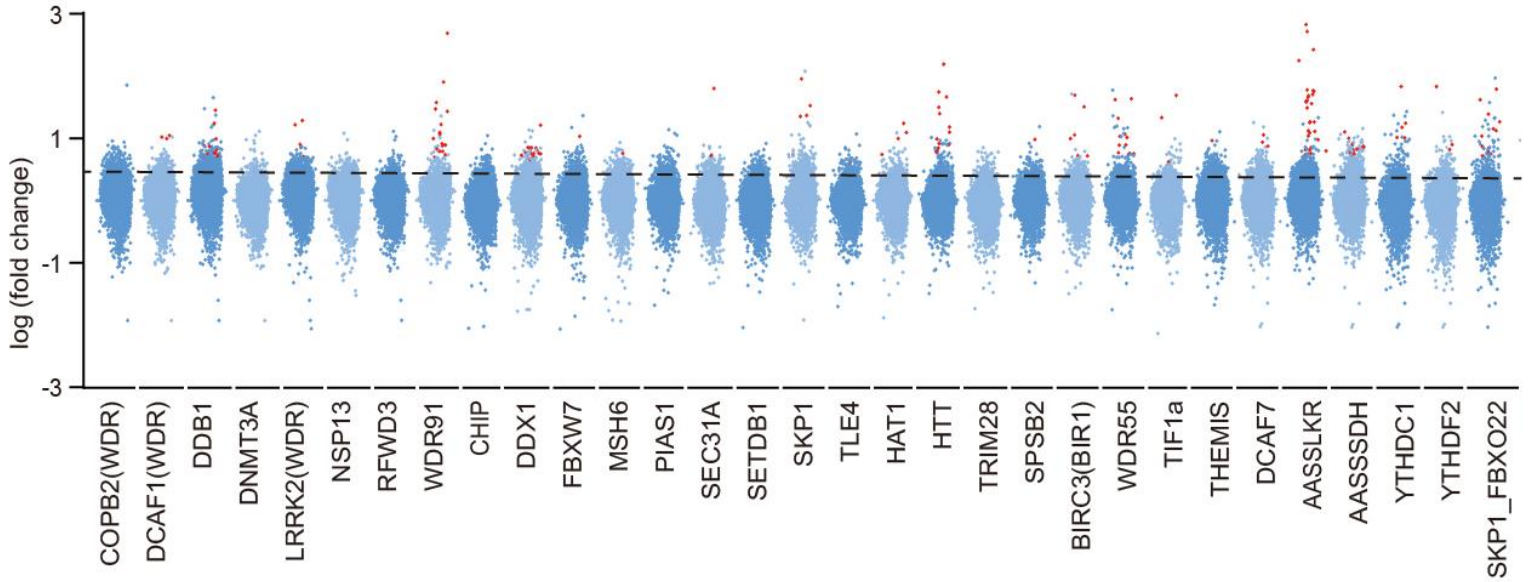
460K library: Enamine Real HLL-460 Library

Library v2.0:
12K picked by SGC partners

Throughput: ~400 proteins against 15K chemicals per year

Targets were structurally diverse and represented a range of predicted ligandability



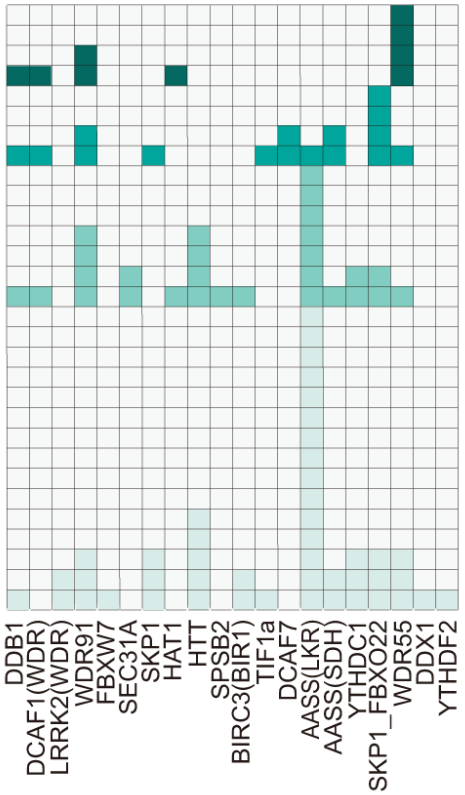


Sheridan et al., J. Chem. Inf. Model. 2010

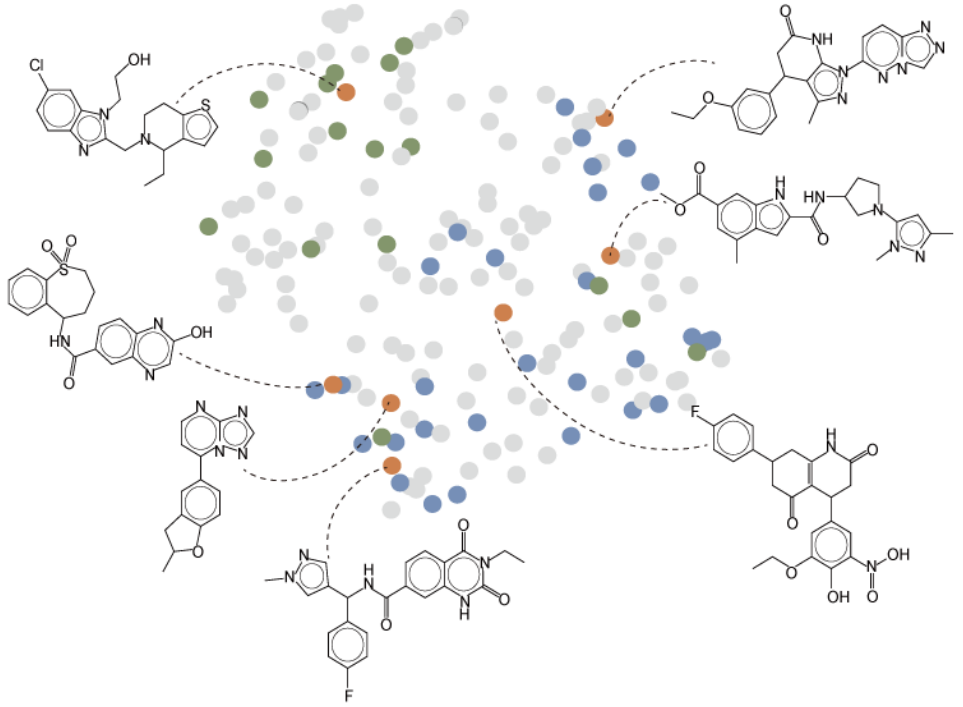
-31 targets including 3 ‘easy’ (DLID>1), 11 ‘medium’ (0.5<DLID<1), and 16 ‘hard’ (DLID<0.5) targets

-81 ASMS hits (>10) were detected for 19 targets: 3 ‘easy’ (100%), 7 ‘medium’ (64%), 9 ‘hard’ targets (56%)

Many (but not all) predicted binders were confirmed in orthogonal biophysical assays



Level 1: Co-crystal structure
Level 2: Orthogonal validation
Level 3: Chiral selectivity
Level 4: High fold change

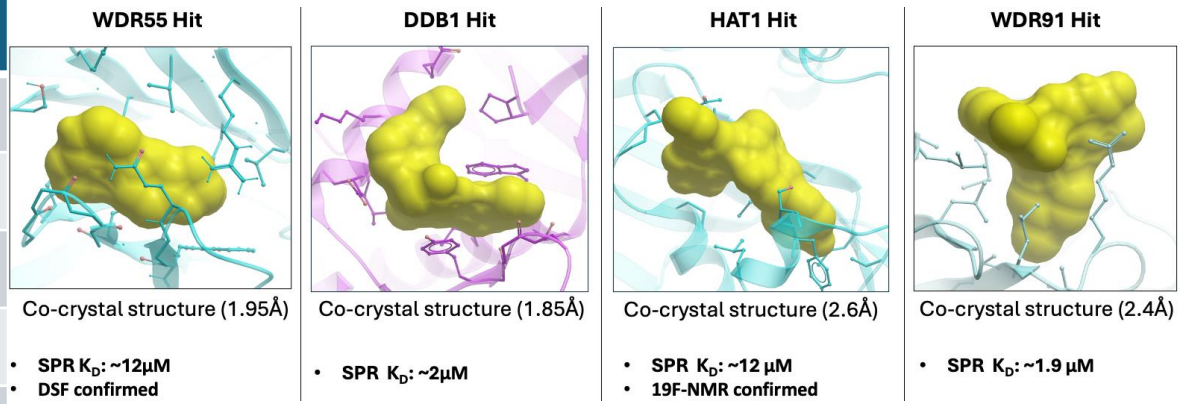


SPR and chiral selectivity
Chiral selectivity
SPR validated
Fold change > 5

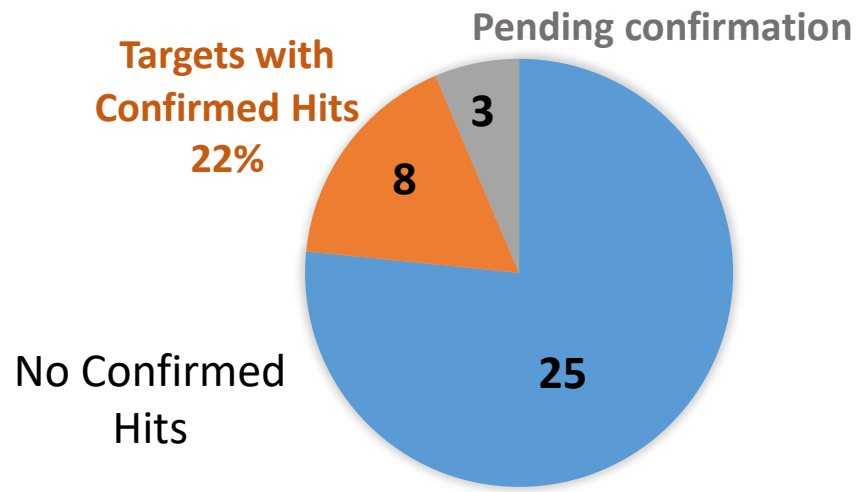
- 25 hits were validated for 10 targets (10/31, 32%) by SPR; and 9 hits were validated by co-crystal structures
- 10 enantioselective hits were validated
- First-in-class ligands were discovered for 2 targets including WDR55 and HAT1

ASMS screening summary

New Targets with confirmed hits	SPR K_D (µM)	Xray/CryoEM structure
DDB1 (2 hits)	~ 2 µM	Yes
WDR91 (2 hit)	~ 1.9 µM	Yes
HAT1 (1 hit)	~ 12 µM	Yes
WDR55 (4 hits)	~ 10-56 µM	Yes
SEC31A (1 hit)	~ 30 µM	In progress
SKP1 (2 hits)	~ 20-50 µM	No
FBXO22-SKP1 (3 hits)	~ 4-10 µM	In progress
Huntingtin-HAP40 (1 hit)	~ 15 µM	In progress



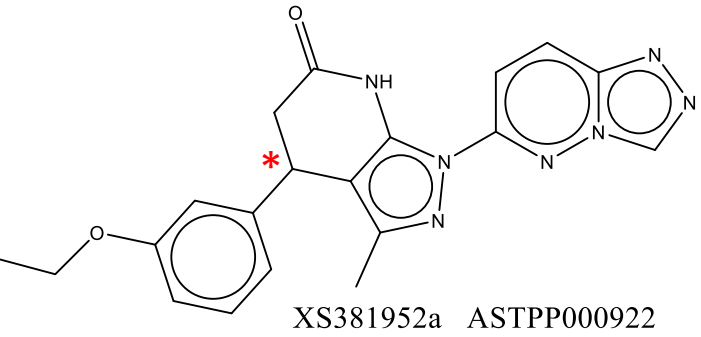
To Date: 36 Targets were screened with the 9K library



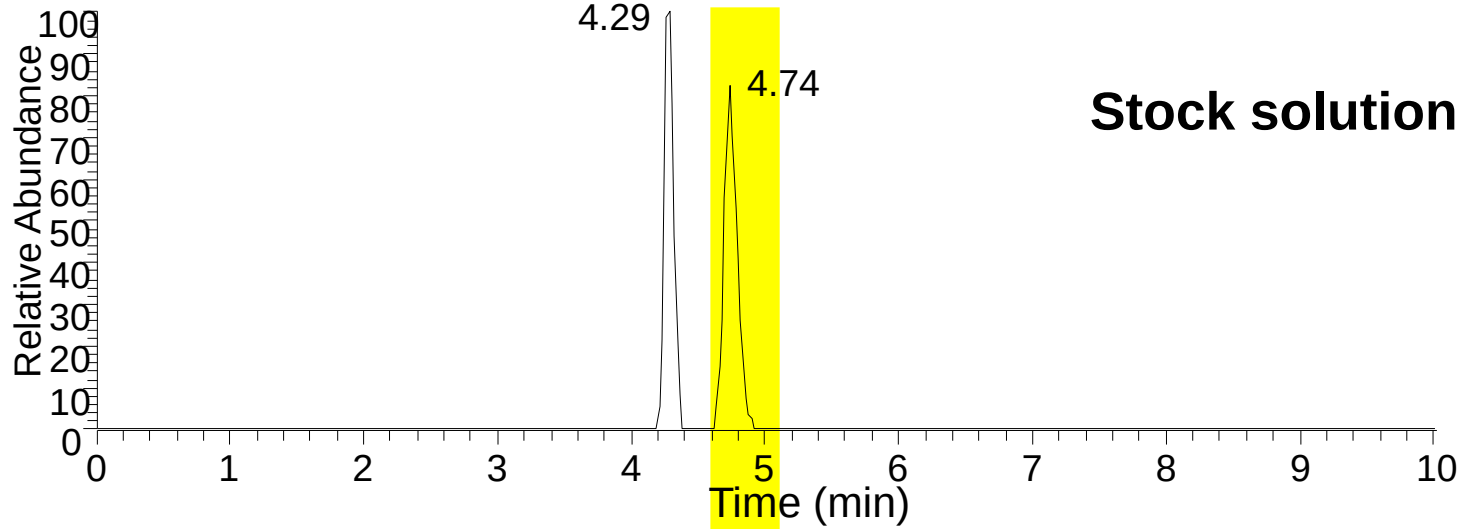
Community Targets with hit confirmation in progress

- THEMIS CABIT domain (T. Beyett)
- AASS LKR-GFP (Wyatt Yue target)
- AASS SDH-GFP(Wyatt Yue target)

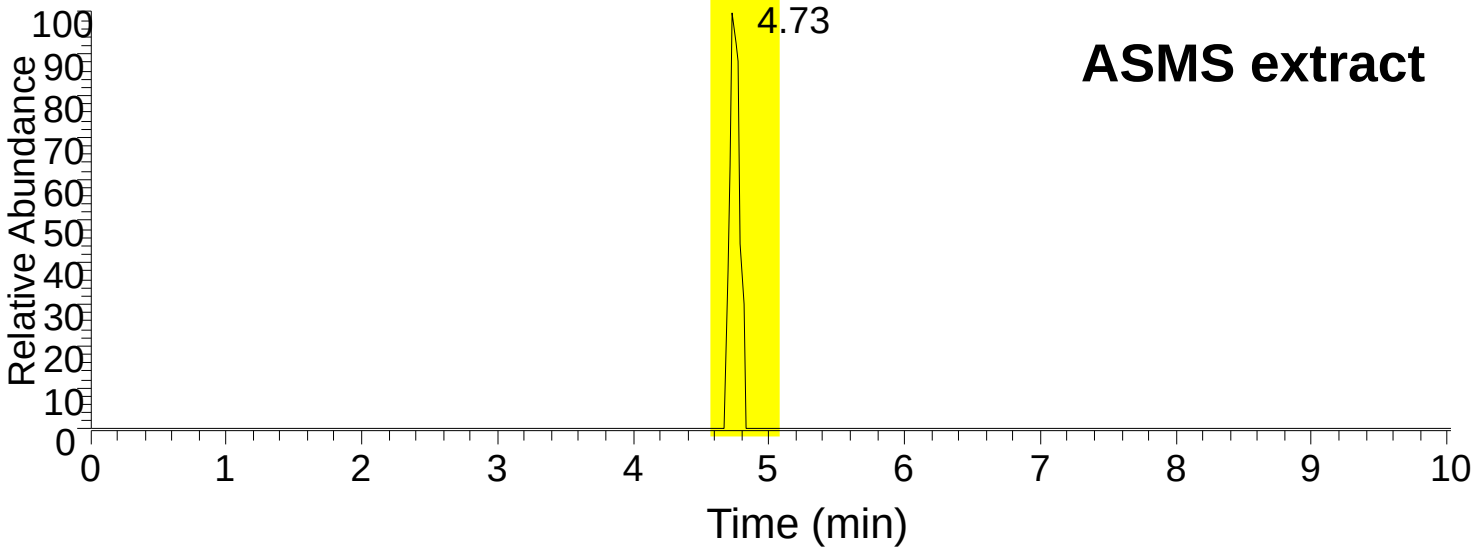
The DDB1 hit displays enantiomeric selectivity



XS381952a ASTPP000922



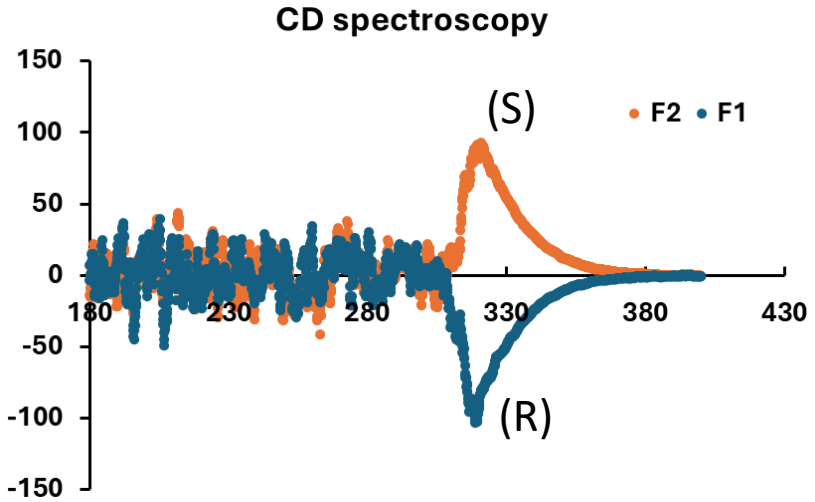
Stock solution

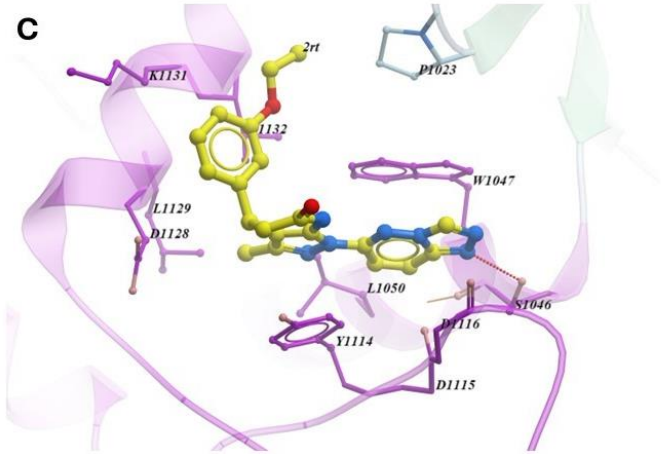


ASMS extract

The 2nd enantiomer was selectively enriched

Verification of enantiomer-specific binding for DDB1



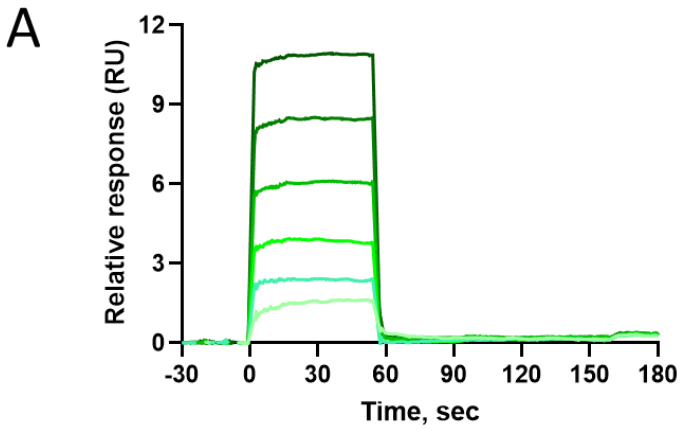


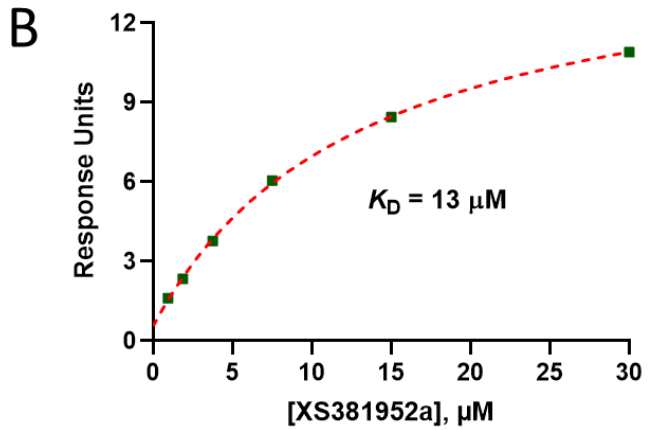
SPR:

K_D (S)=13 µM

K_D (R)=85 µM

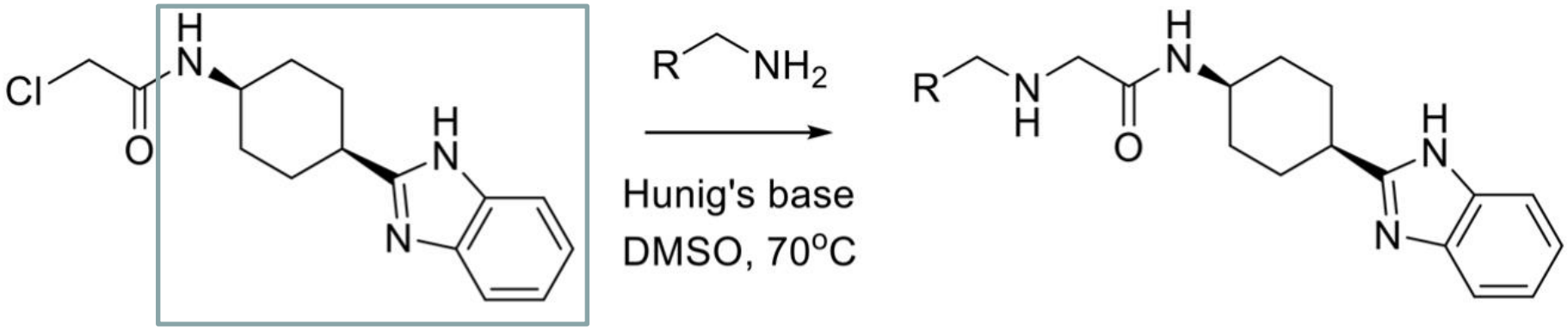
The (S) enantiomer selectively binds to DDB1 protein

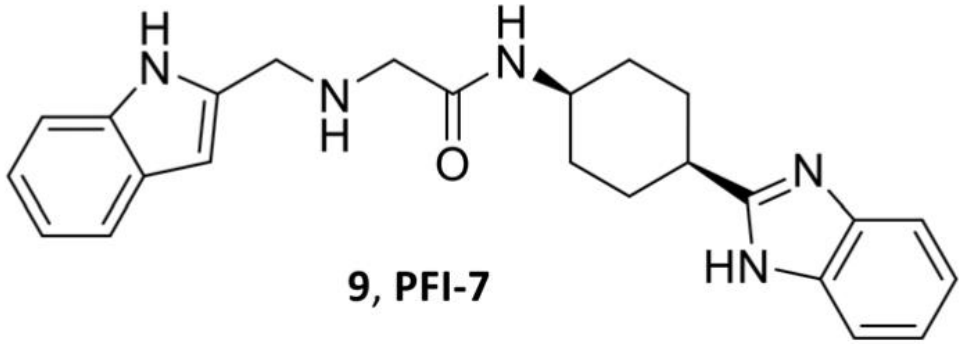




AS-MS Screening

GID4: mass spectrometry (ASMS) primary screen (500 000 compounds)

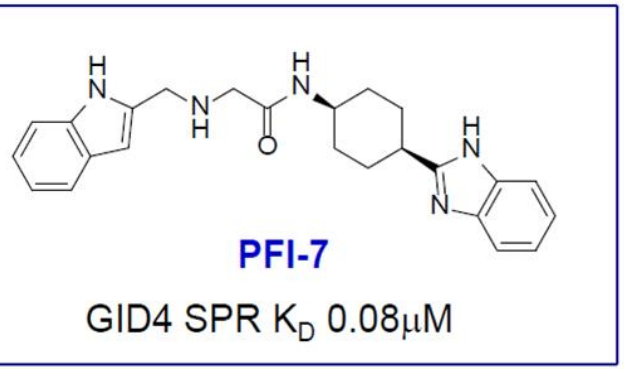


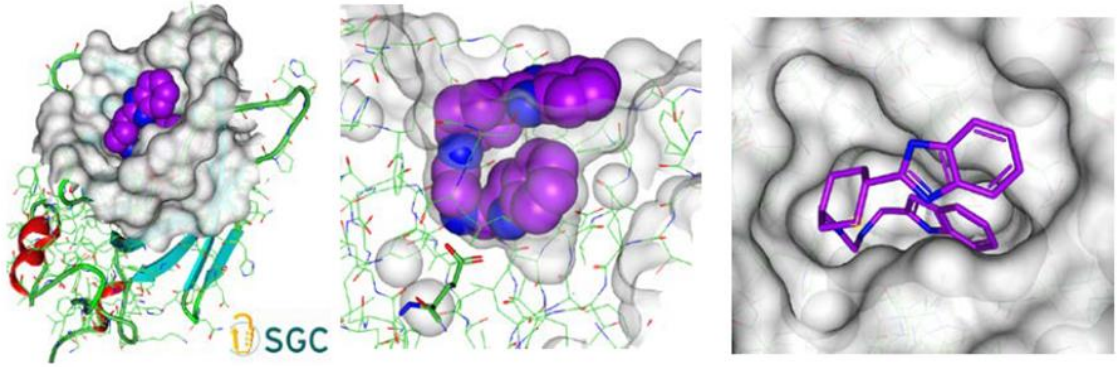


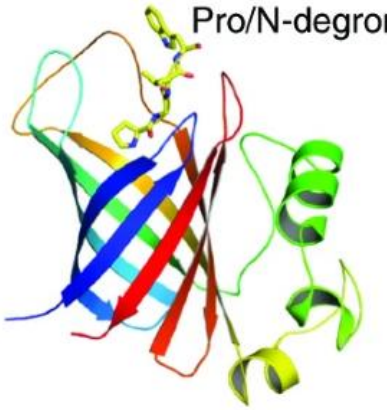
9, PFI-7

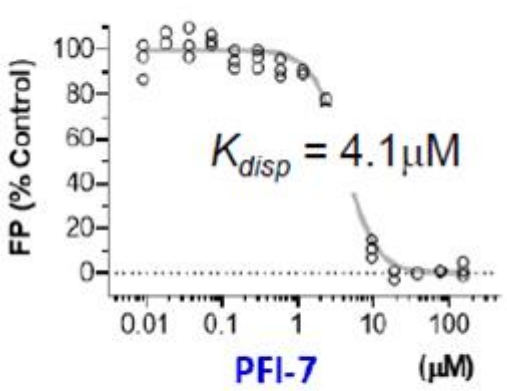
New E3 ligands targeting GID-4

- Gid4 was shown to target the gluconeogenic enzymes Fbp1, Icl1, and Mdh2 in yeast
- N-terminal Pro residues (Pro/N-degron; N-terminal Pro followed by a small residue)

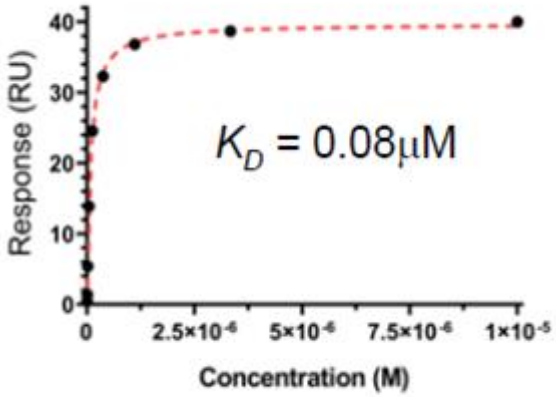






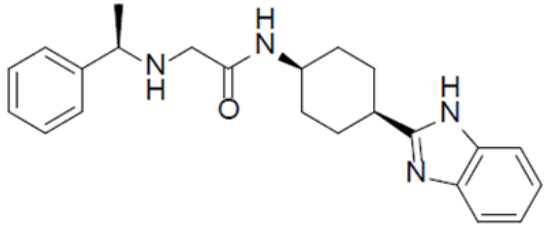


Biochemical Assay

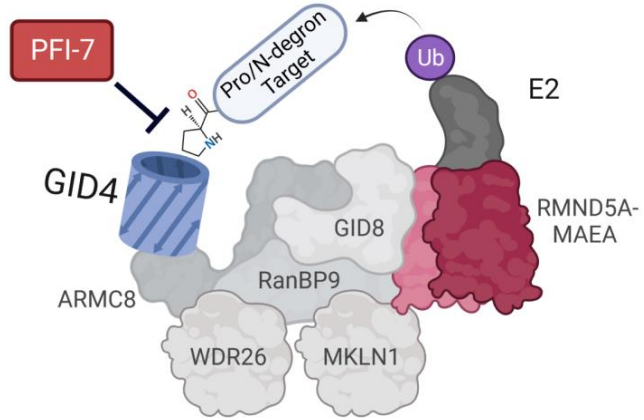


SPR

Neg. control

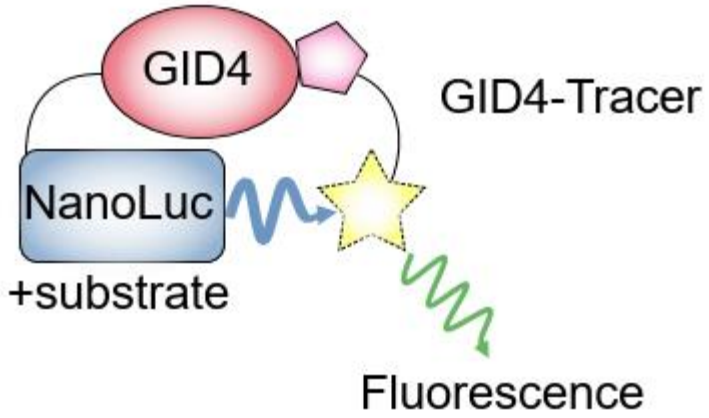


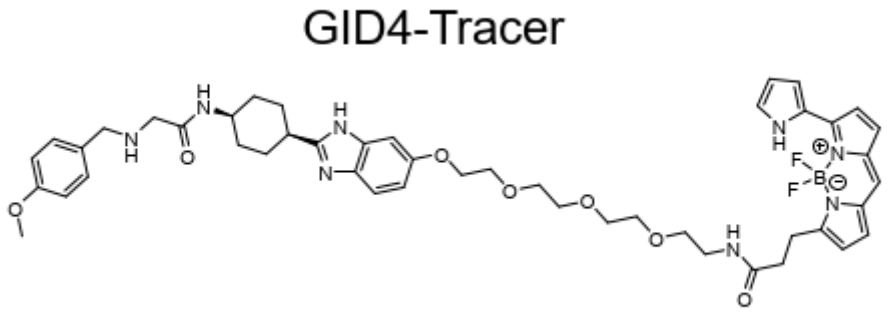
PFI-7N

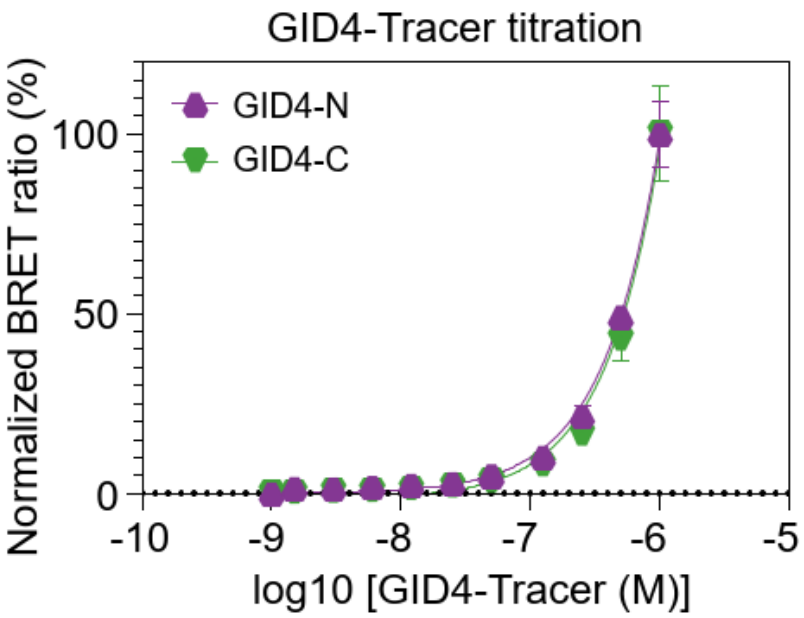


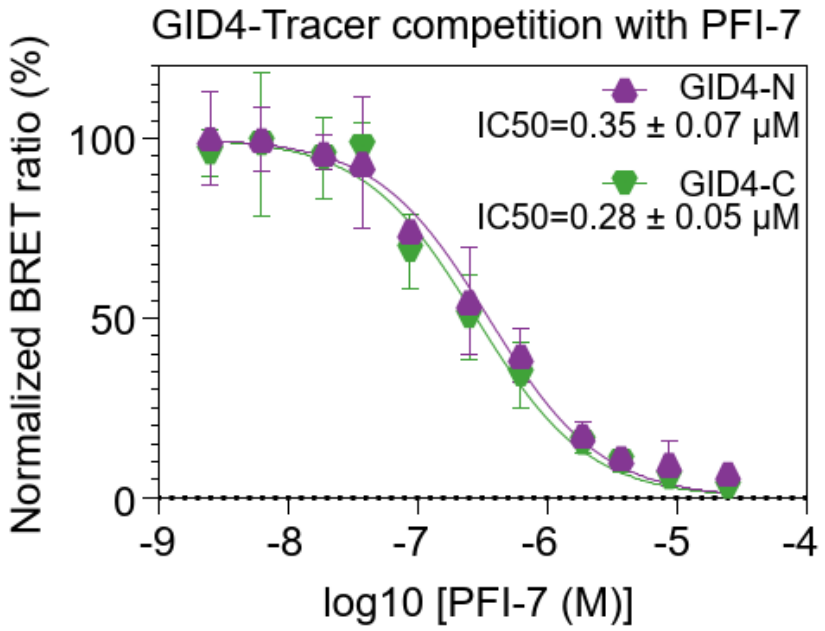
Nature Chemical Biology, in press

Tracing ligand binding to GID-4

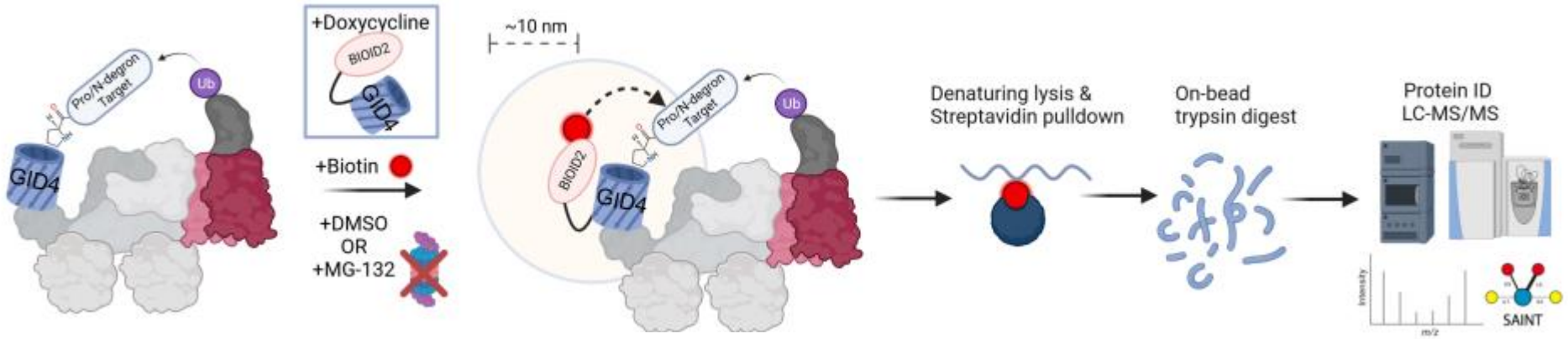


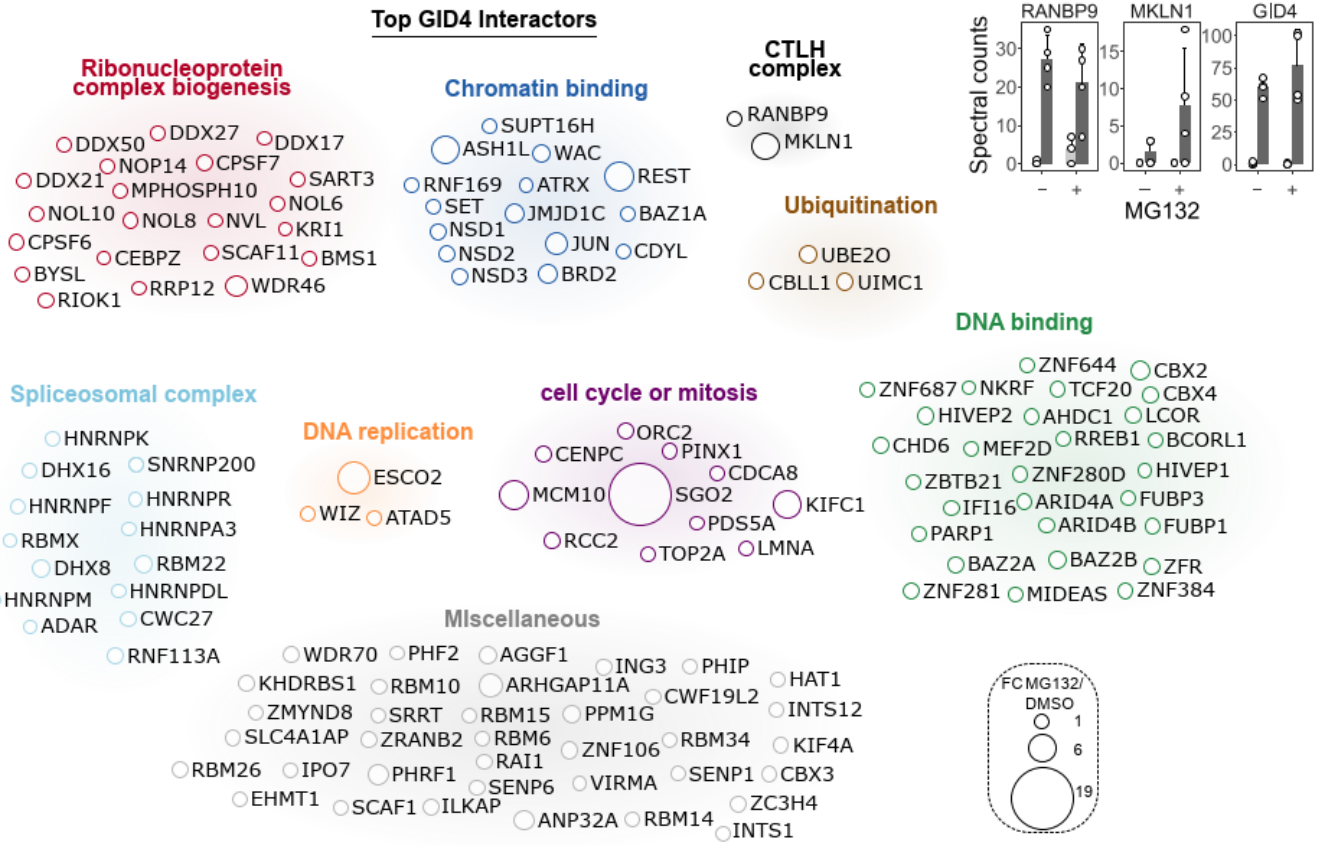




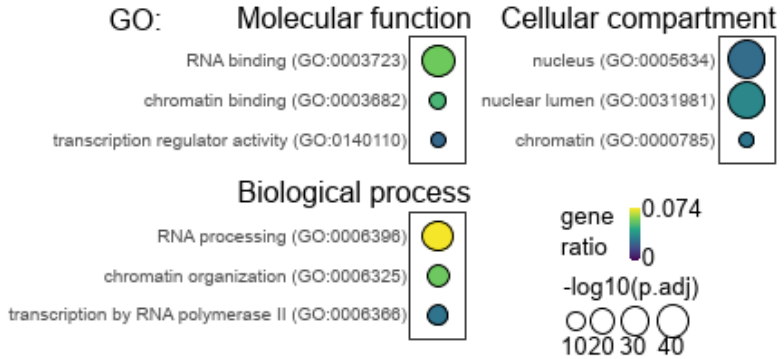


GID-4 Interactors in human Cells



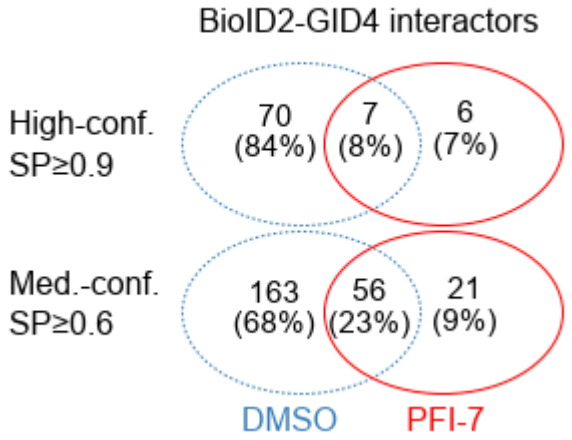


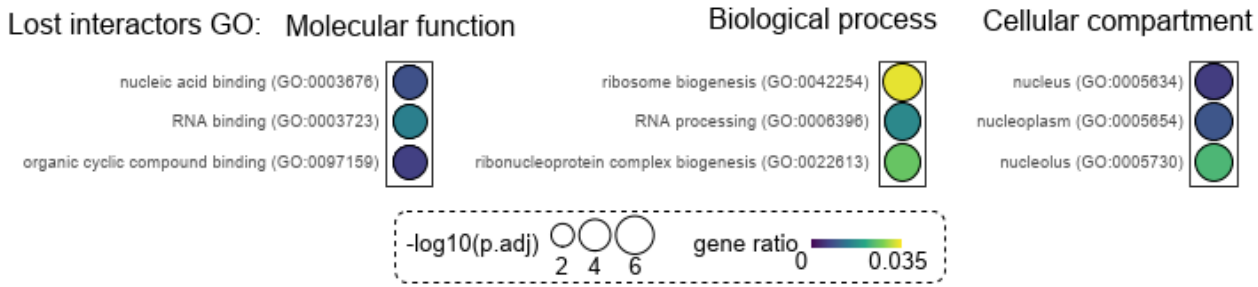
GID-4 interactors

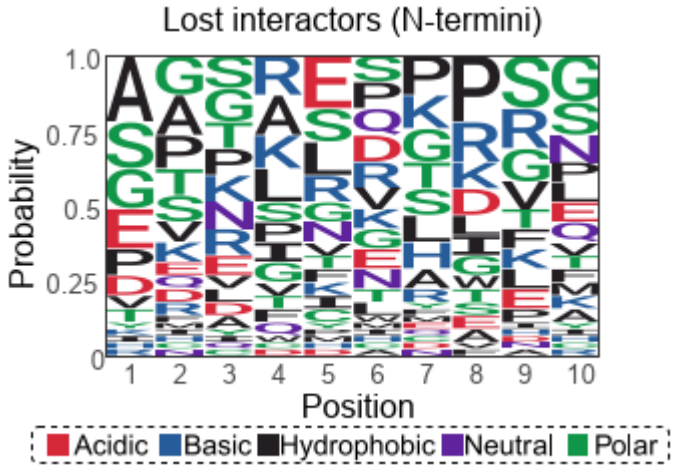


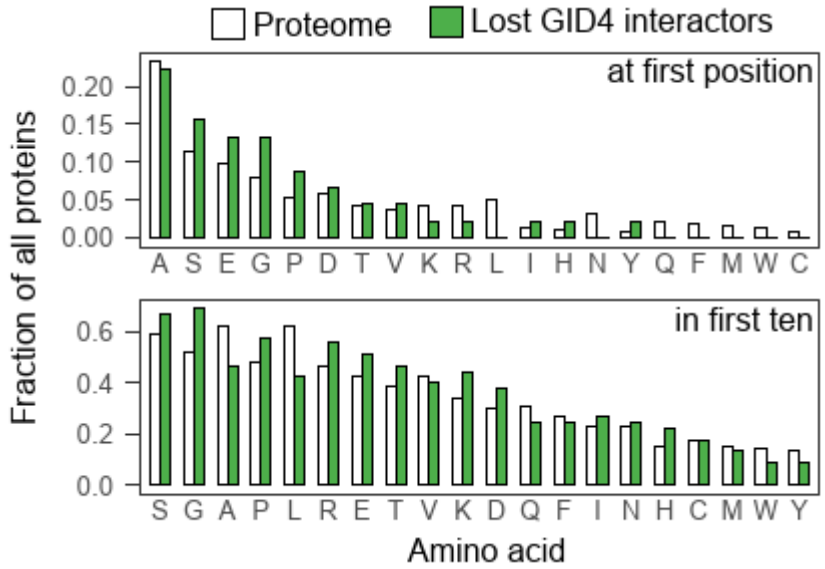
Nature Chemical Biology, in press

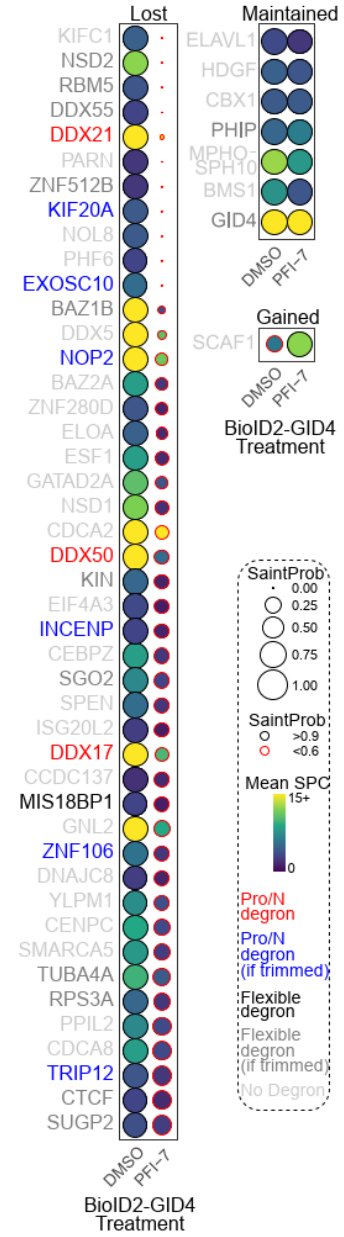
Consequences of GID-4 E3 inhibition









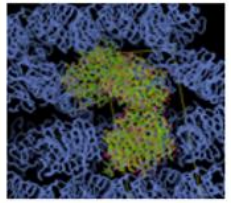


Nature Chemical Biology, in press

HTP Crystallographic Fragment Screening

Crystallisation screening
3 crystals forms: $C222_1$, $P2_12_12_1$, P422

Suitable: $P2_12_12_1$
(pocket open to solvent channel)



1 **Search for 'right' crystal form**

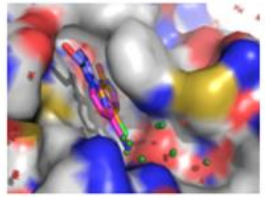
2 **Reproducible crystallisation**

Defining condition for crystallising frozen protein with glycerol.

>80 crystals/plate using 3 repeating conditions with defined range of precipitant.

Crystals tolerate soaking and DMSO.

In-house manual soaking:
>50 fragments =>
more than 40 datasets; 3 ligands found



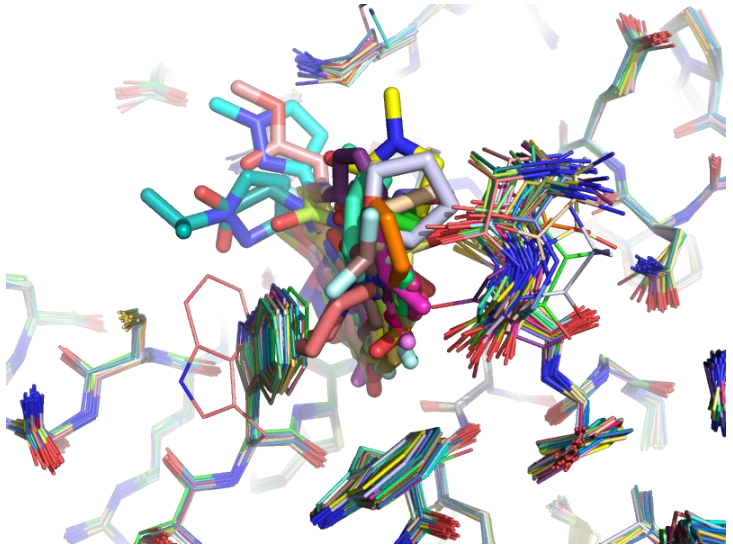
3 **In-house, manual test soaking**

XCHEM Diamond Fragment Screening

4 **XCHEM (and MAX IV) fragment screening**

XCHEM – full campaign of >760 cpds

MAXIV - >200 cpds tested



950 diffraction data sets
52 with fragments

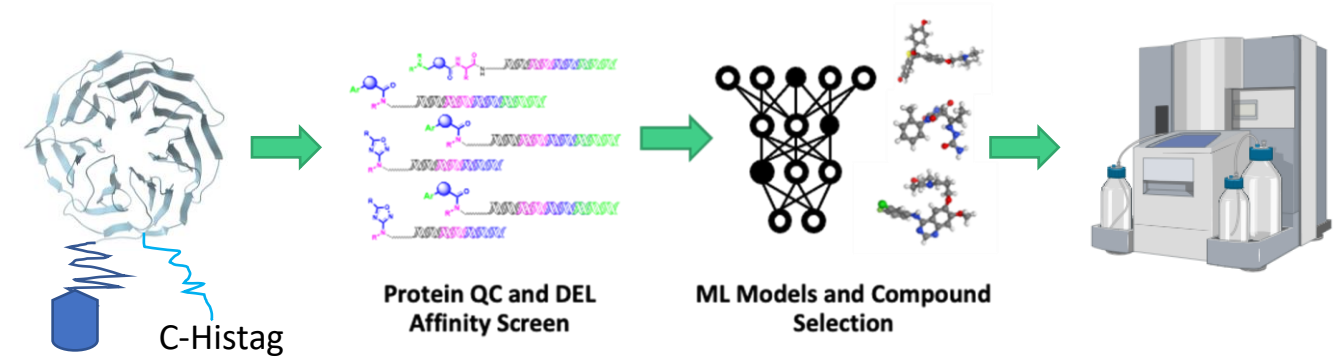
Development of pharmacophore models

DEL + ML hit finding pilot with XChem/ZebiAI/XChem and Google

pubs.acs.org/jmc Article

Machine Learning on DNA-Encoded Libraries: A New Paradigm for Hit Finding

Kevin McCloskey,‡ Eric A. Sigel,‡ Steven Kearnes, Ling Xue, Xia Tian, Dennis Moccia, Diana Gikunju, Sana Bazzaz, Betty Chan, Matthew A. Clark, John W. Cuozzo, Marie-Aude Guié, John P. Guilinger, Christelle Huguet, Christopher D. Hupp, Anthony D. Keefe, Christopher J. Mulhern, Ying Zhang, and Patrick Riley*

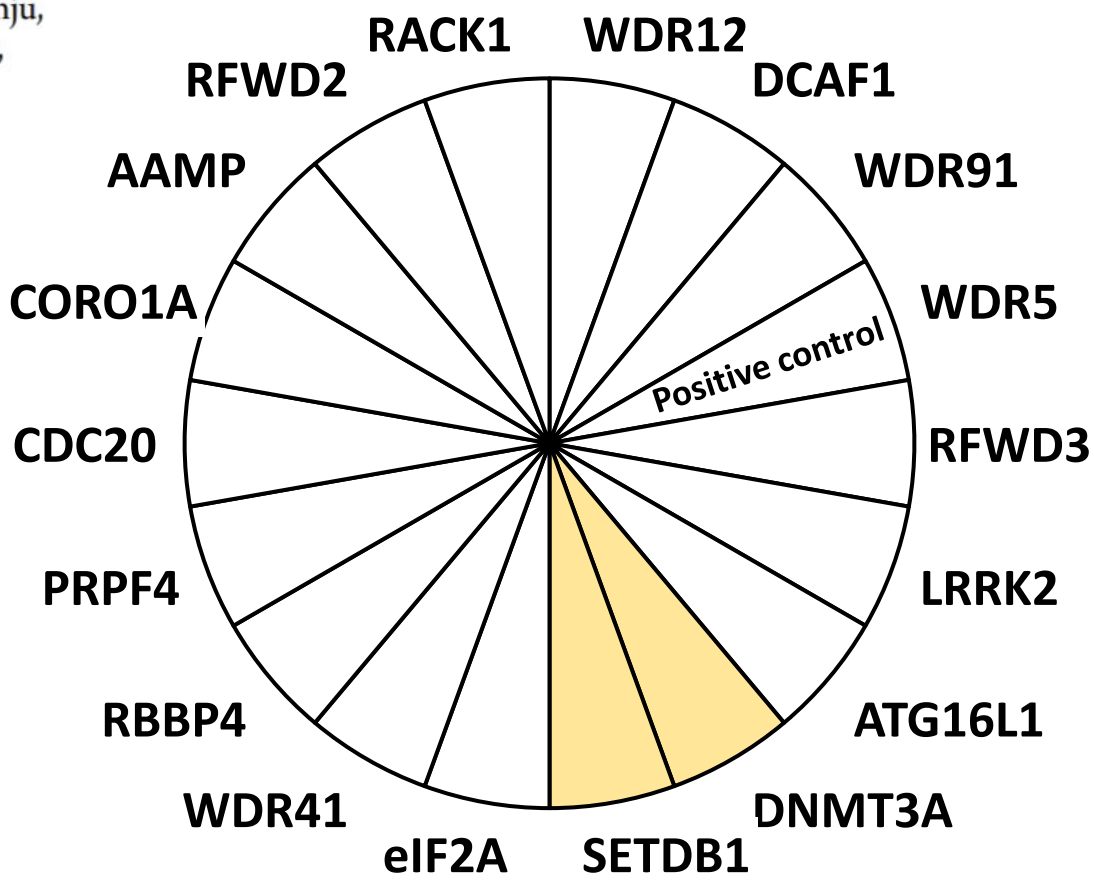


SGC | X-Chem | ZebiAI | SGC

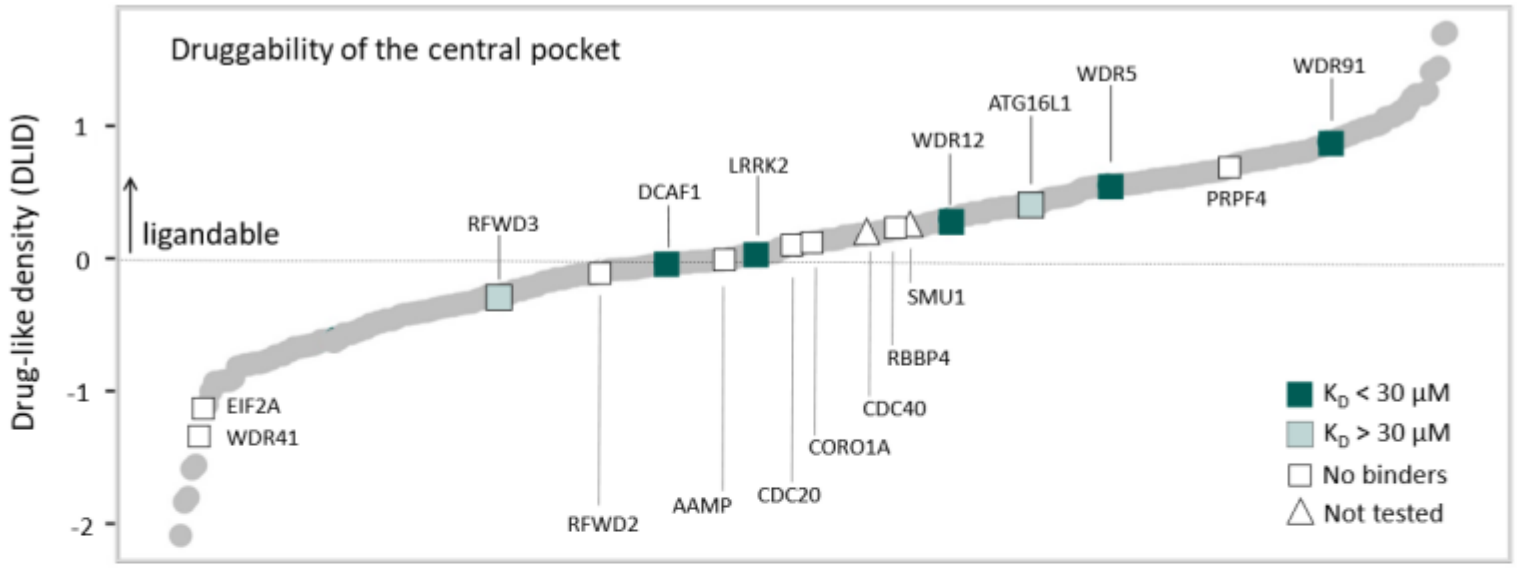
Google

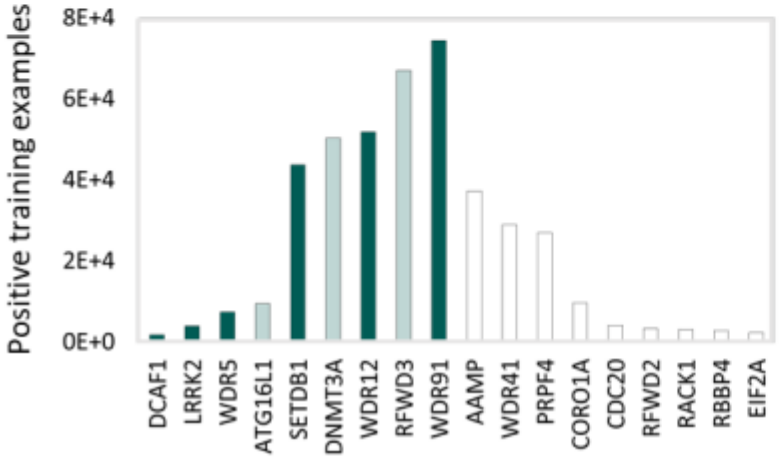
to predict commercially available compounds

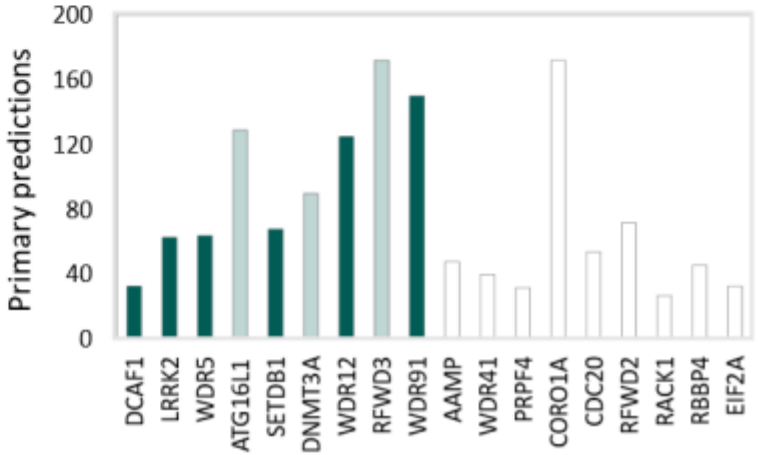
Unprecedented targets for DEL+ML screening



DEL-ML Screening



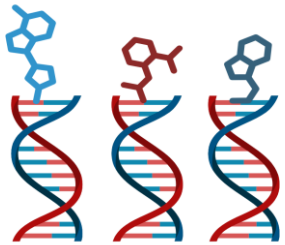




Process applied to DCAF1

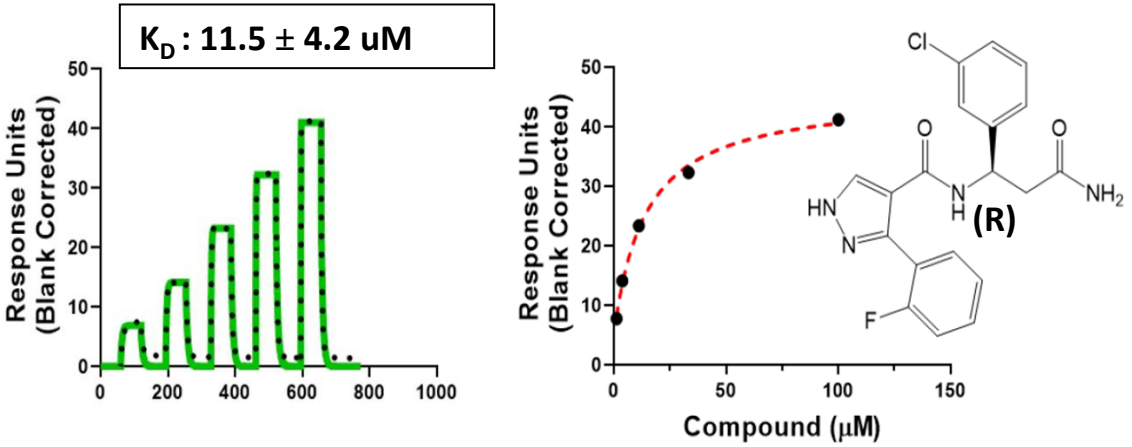
X-CHEM TAKE THE LEAD. DRIVE DISCOVERY.

zebi AI



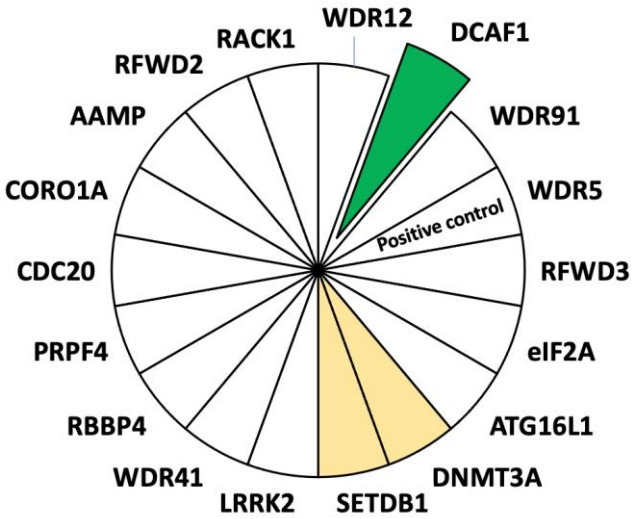
>2K DEL selections

Exp. DATA ⟷ ML DATA

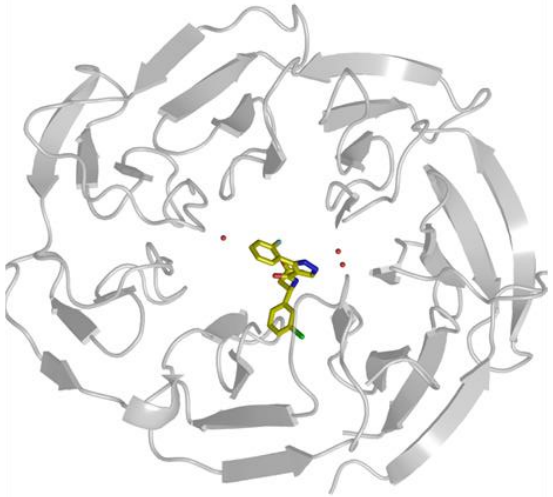


114 billion compounds screened by DEL

- **33** compounds selected through ML from Enamine REAL library



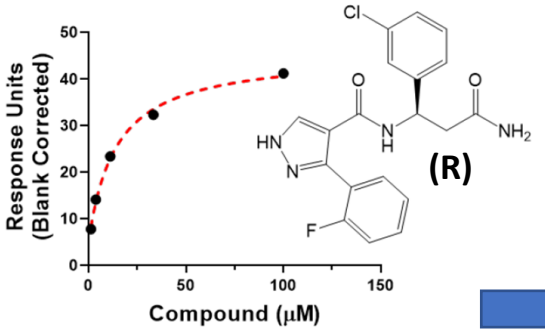
2.3Å X-ray structure



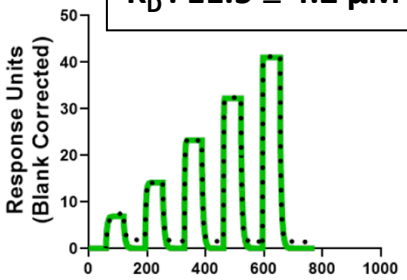
<u>Experimental screening:</u>
Alice Li, Fengling Li, Masoud Vedadi

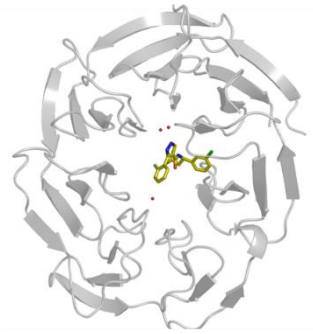
Serah Kimani, PDF

DCAF1: DEL+ML hit optimization

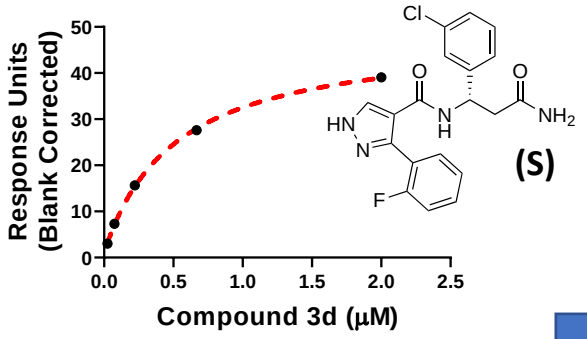


K_D : 11.5 ± 4.2 μM

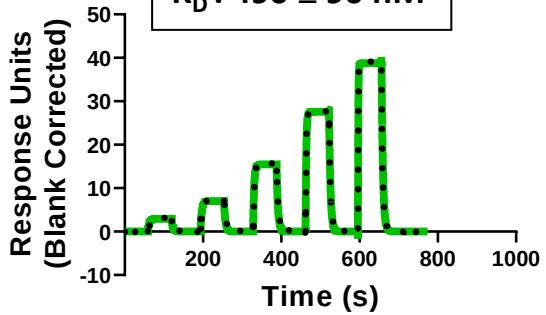


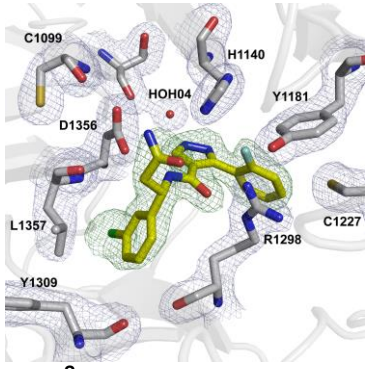


2.3Å X-ray structure

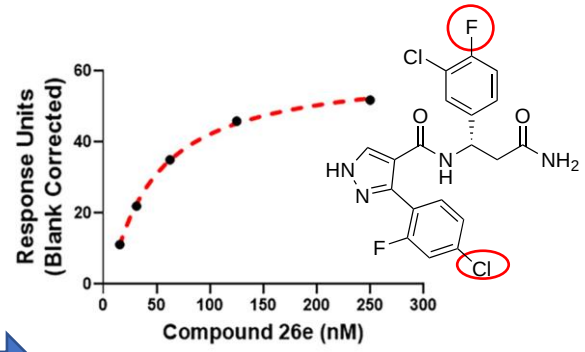


K_D : 490 ± 90 nM

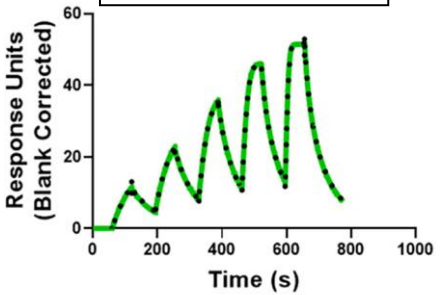


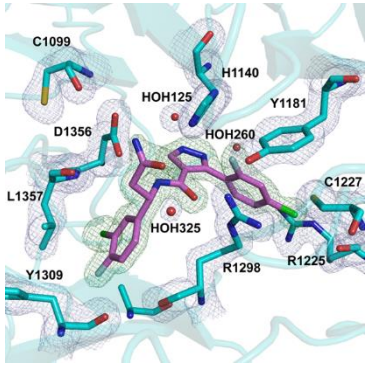


1.9Å X-ray structure
(PDB ID:7UFV)



K_D : 38 ± 1.5 nM





1.55Å X-ray structure
(PDB ID:8F8E)

RETURN TO ISSUE | < PREV ARTICLE NEXT >

Discovery of Nanomolar DCAF1 Small Molecule Ligands

Alice Shi Ming Li, Serah Kimani, Brian Wilson, Mahmoud Noureldin, Héctor González-Álvarez, Ahmed Mamai, Laurent Hoffer, John P. Guilinger, Ying Zhang, Moritz von Rechenberg, Jeremy S. Disch, Christopher J. Mulhern, Belinda L. Slakman, John W. Cuozzo, Aiping Dong, Gennady Poda, Mohammed Mohammed, Punit Saraon, Manish Mittal, Pratik Modh, Vaibhavi Rathod, Bhashant Patel, Suzanne Ackloo, Vijayaratnam Santhakumar, Magdalena M Szewczyk, Dalia Barsyte-Lovejoy, Cheryl H. Arrowsmith, Richard Marcellus, Marie-Aude Guié, Anthony D. Keefe, Peter J. Brown*, Levon Halabelian*, Rima Al-awar*, and Masoud Vedadi*

Cite this: *J. Med. Chem.* 2023, 66, 7, 5041−5060
Publication Date: March 22, 2023
https://doi.org/10.1021/acs.jmedchem.2c02132
Article Views 12646 | Altmetric 28 | Citations 6
LEARN ABOUT THESE METRICS
Share | Add to | Export

Data Management and Access Strategy

- 2TB (compressed) to 50 Tb (non-compressed) of protein-ligand interaction data
- Rigorous data management integrated with experimental methods.
- Collaboration between experimentalists and data scientists.

- Cloud-based database designed to accommodate over 300 Tb of data
- Central repository for the Target 2035 screening datasets
- Experimentally validated protein-small molecule binding data (positive and negative)
- FAIR principles (Findable, Accessible, Interoperable, Reusable)
- Comprehensive documentation of experimental protocols and findings
- **Machine learning-ready data datasets**

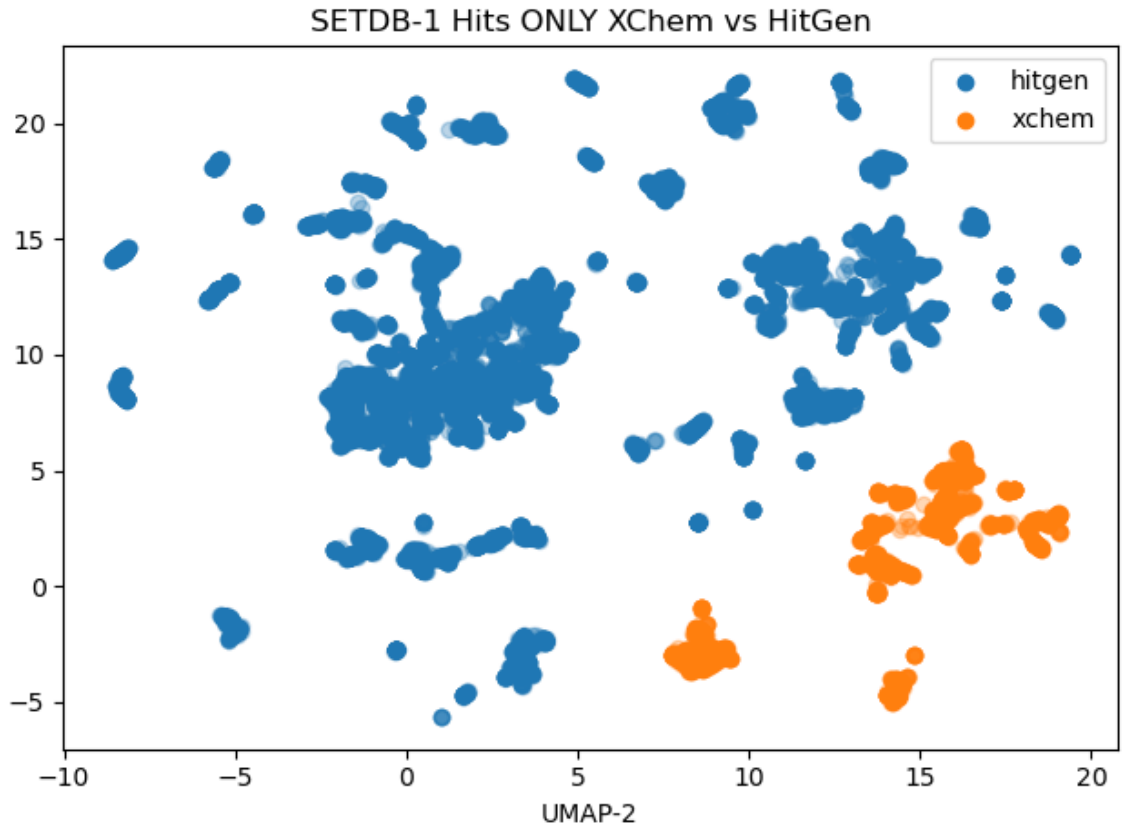
Datasets

Target Name ↑↓	Description	Selection Date ↑↓	Partner ↑↓	Data Dictionary	Download Dataset
WDR12	To identify binders to the target	2023-08-28	HitGen		Public access coming soon
SETDB1	To identify binders to the target	2023-10-20	X-Chem		Public access coming soon
RFWD3	To identify binders to the target	2023-10-20	X-Chem		Public access coming soon
DNMT3A	To identify binders to the target	2023-10-20	X-Chem		Public access coming soon

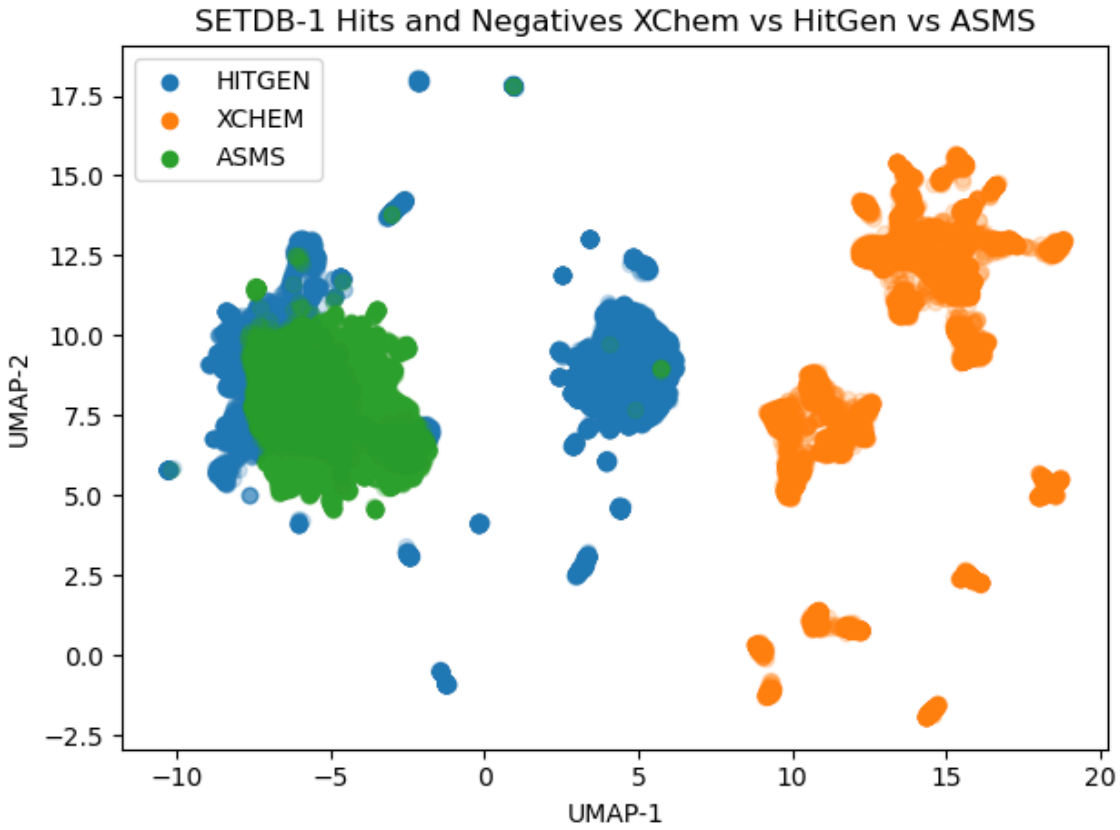
SGC https://aircheck.ai/

What we have learned so far?
Comparing HitGen and XChem Chemical Space

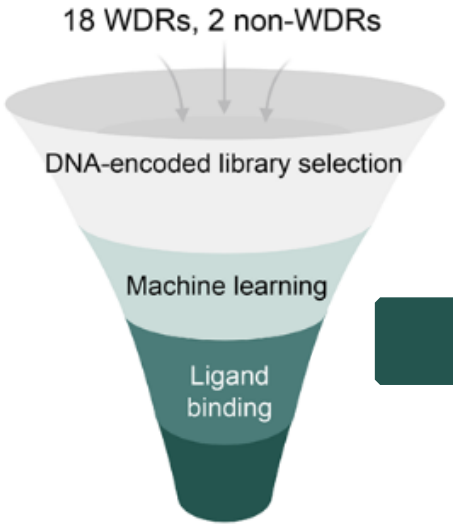
SETDB FP Dataset – Positive Training Sets ONLY

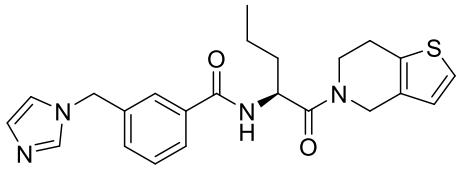


SETDB FP Dataset – Positive & Negative Training Sets



Example 1: From DEL-hit to a chemical probe for WDR5

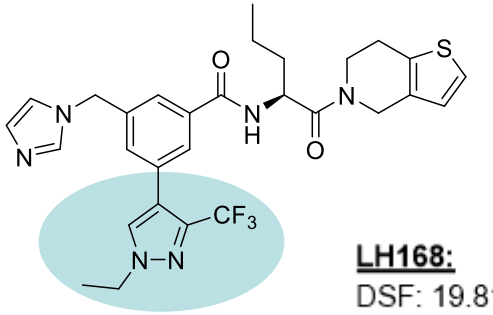




DEL-Hit Toronto (MR43378) [11]

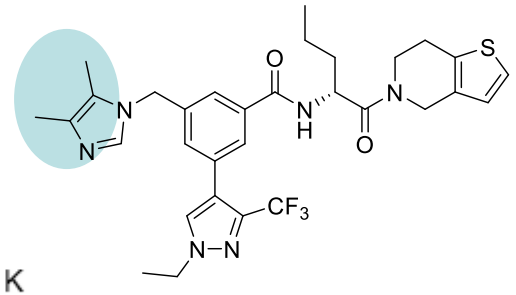
K_D (SPR): **16 µM**

SAR



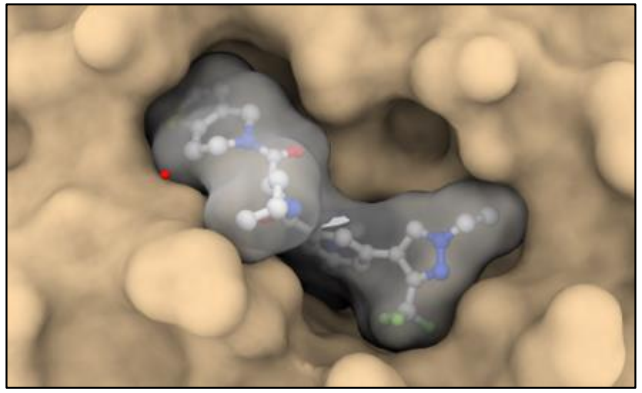
WDR5 probe (LH168)

LH168:
DSF: 19.81 K
K_D (ITC): 13 nM
IC_{50}(NanoBRET): intact: < 10 nM



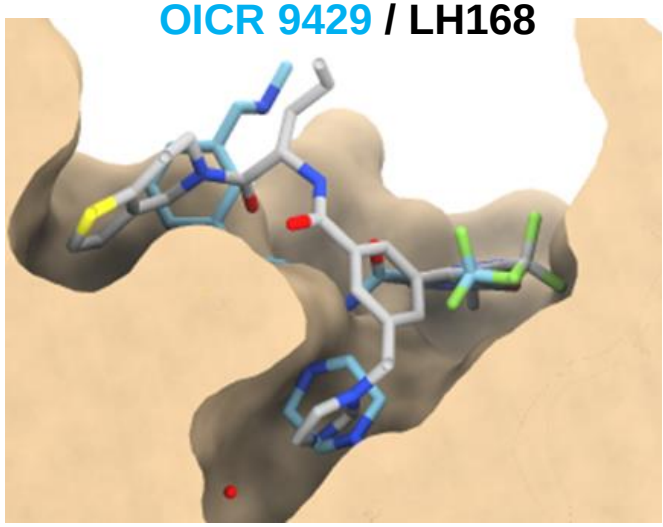
negative control (LH222)

LH168 bound to WDR5



Resolution 1.7 Å

OICR 9429 / LH168



AIRCHECK Team

Nabin Bagale

James Wellnitz

Shaghayegh Reza

Matthew Boccalon

Farnoosh Khodakarami

Rafael Counago

Maxwell Morgan

Matthieu Schapira

Benjamin Haibe-Kains

BHK LAB

MAINFRAME

MAchine learning **I**nnovation **N**etwork **F**or **R**esearch to **A**dvance **ME**dicinal chemistry

Join Us

Contact:
Dr. Albert A. Antolin
aantolin@idibell.cat

As of October 28, 2024

BHK LAB

Target 2035

A PROBE FOR EVERY PROTEIN

- Who: International federation of biomedical scientists (public & private sector)
- Aim: Develop and apply new technologies to create by the year 2035 chemogenomic libraries, chemical probes and/or biological probes for the entire human proteome.

Interested to share your experience, knowledge and/or results?

→ Contact: Target2035@thesgc.org

TARGET 2035

www.target2035.net

Acknowledgements

Susanne Muller-Knapp

Dinh Nguyen Ngu To
Claudia Tredup
Theresa Ehret
Amelie Menge
Johannes Dopfer
Lewis Elson
Hanna Holzmann
Sandra Häberle
Natalie Schneider
Saran Aswathaman
Sivashanmugam

Chemistry

Thomas Hanke
Sandra Röhm
Francesco Greco
Joshua Gerninghaus
Sebastian Mandel
Nicolai Raig
Athina Zerva
Ralf Braden
Aleksandar Lucic
Marko Mitrovic

PROTACs

Václav Němec
Janik Weckesser
Nebojsa Miletic
Adrian Haag
Anton Ruben Hamann
Lasse Hoffmann

Assays

Krishna Saxena
Martin Schwalm
Florian Born
Frederic Farges
Christopher Lenz

Biochemistry/PX

Andreas Krämer
Yeojin Kim
Adarsh Kumar
Guiqun Wang
Rezart Zhubi

TACTIC
Deutsche Krebshilfe
HELFEN. FORSCHEN. INFORMIEREN.

Collaborators

Merck
Paul Gehrtz
Beate Opelt
Corinna Rettig
Johannes Krieger

Goethe Universität Frankfurt am Main

SGC

efpia imi

EUbOPEN
Enabling & Unlocking Biology in the OPEN

Frankfurt Cancer Institute

SGC/companies

- Pfizer
- Genentech
- Boehringer
- BAYER
- Takeda
- Jansen
- BMS

Promega
Matt Roberts
Kristin Huwiler

SGC
Aled Edwards
Cheryl Arrowsmith
Tim Wilson

Xiaoyun Wang
Jianxian Sun
Diwen Yang
Santha Santhakumar
Hui Peng
Levon Halabelian

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

DKTK German Cancer Consortium

Funding:

DKTK
BMBF (Proxidrugs)
Krebshilfe
MJFF
DFG
SGC
LOEWE (TRABITA)
LOEWE (FCI)

PROXIDRUGS