SGC

www.thesgc.org

Aggregating, mining and sharing structural chemical biology data

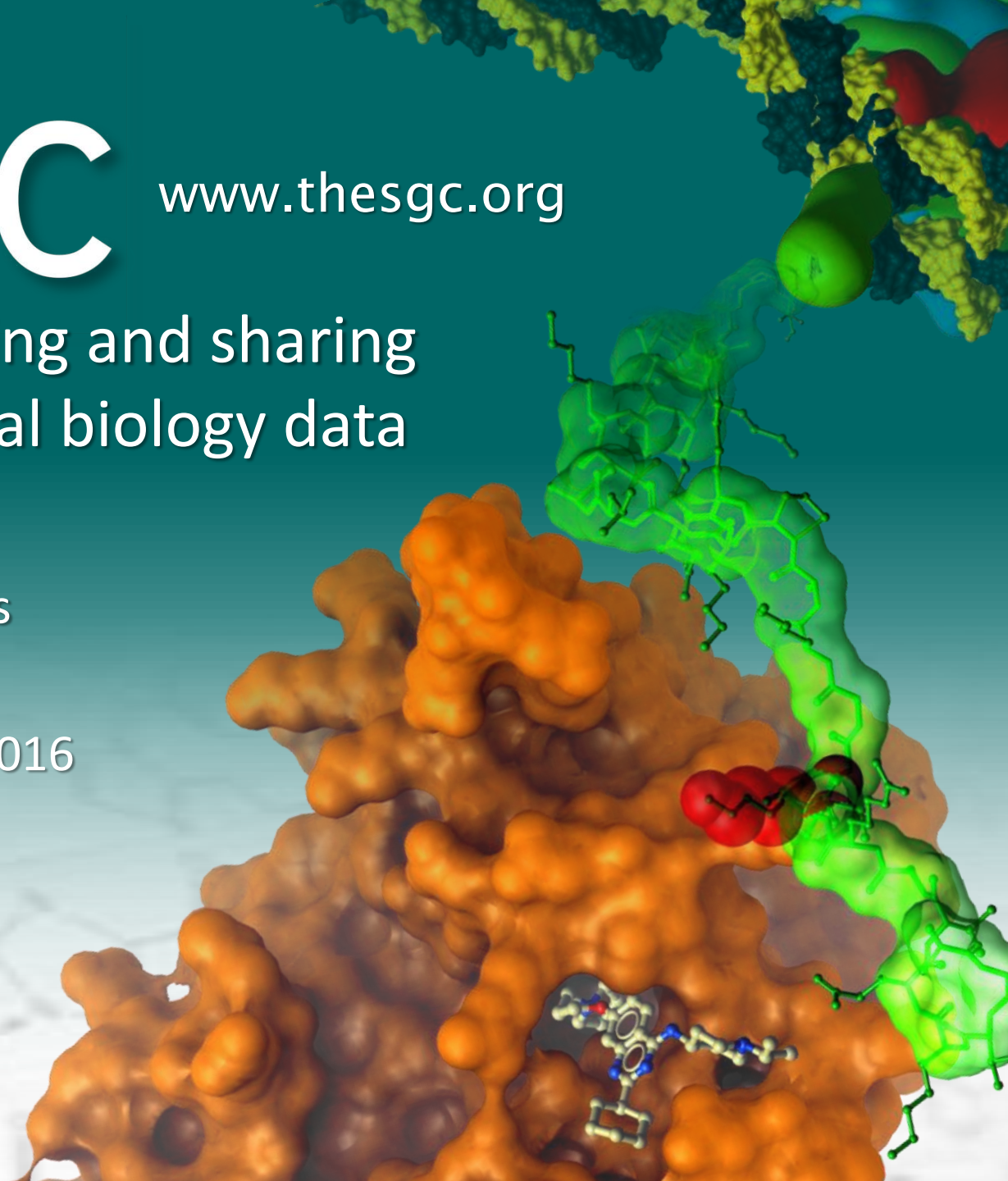
Brian Marsden

PI, Research Informatics

SGC & Kennedy Institute

University of Oxford

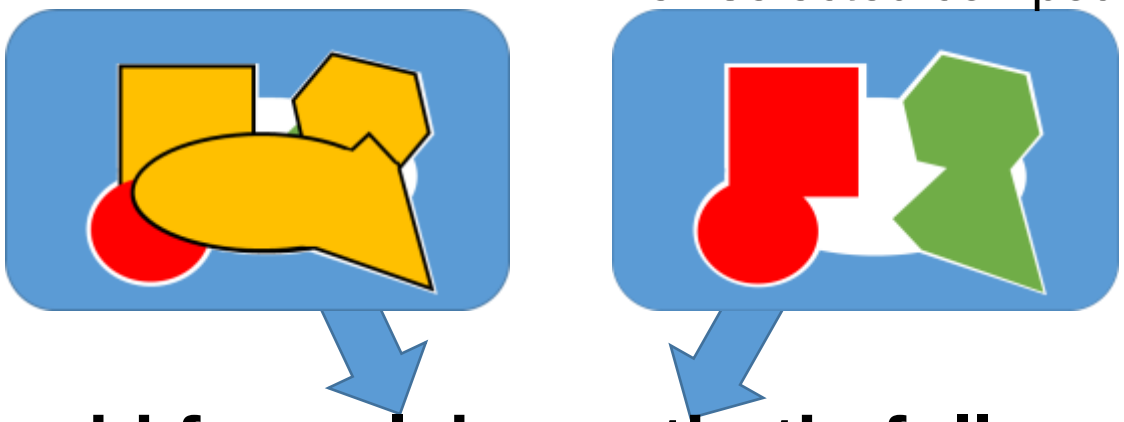
MolSoft UGM, March 2016



XCHEM – X-RAY GUIDED CHEMICAL SYNTHESIS

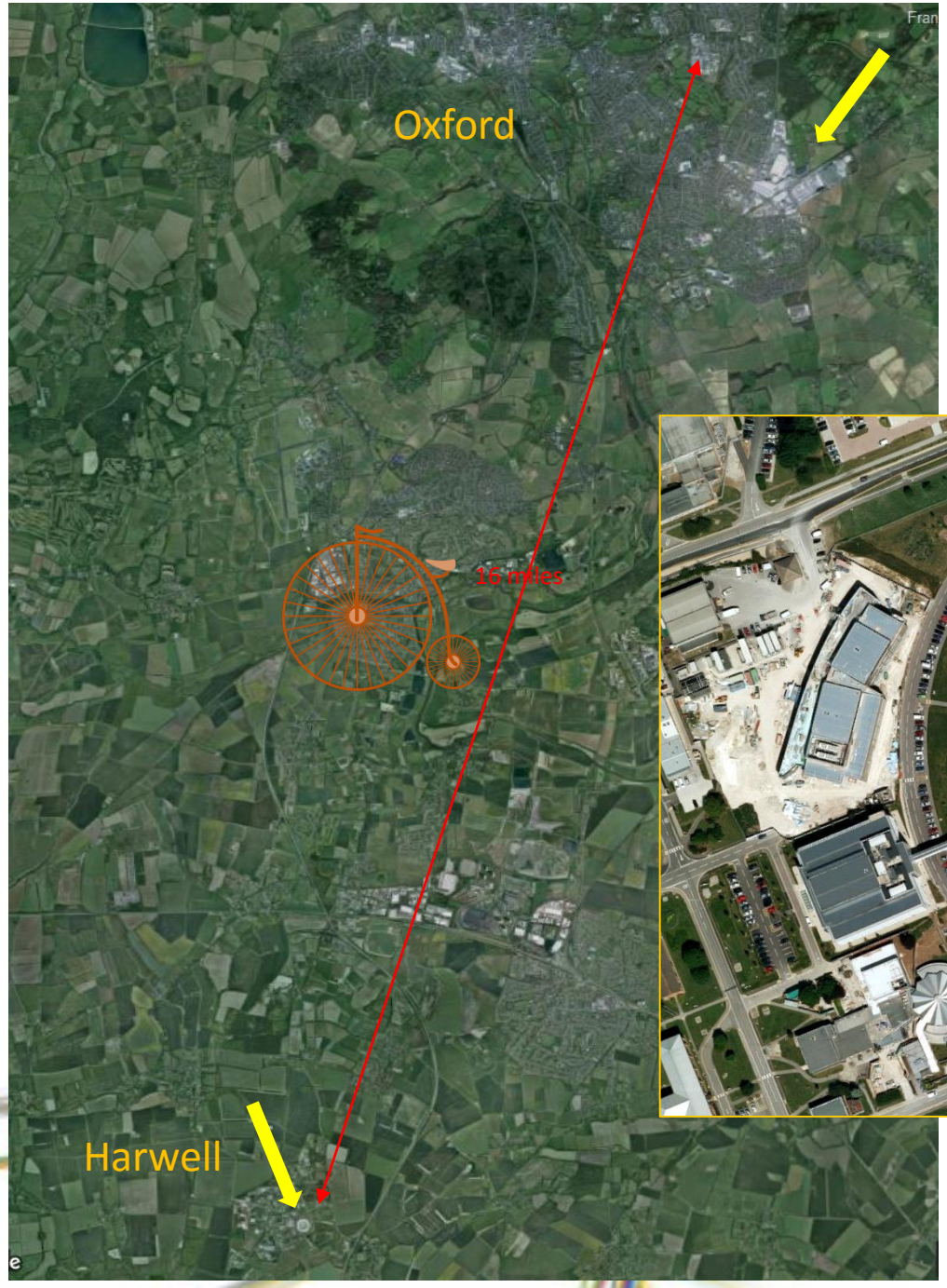
Most sensitive detection technique:
Soaking and X-ray crystallography

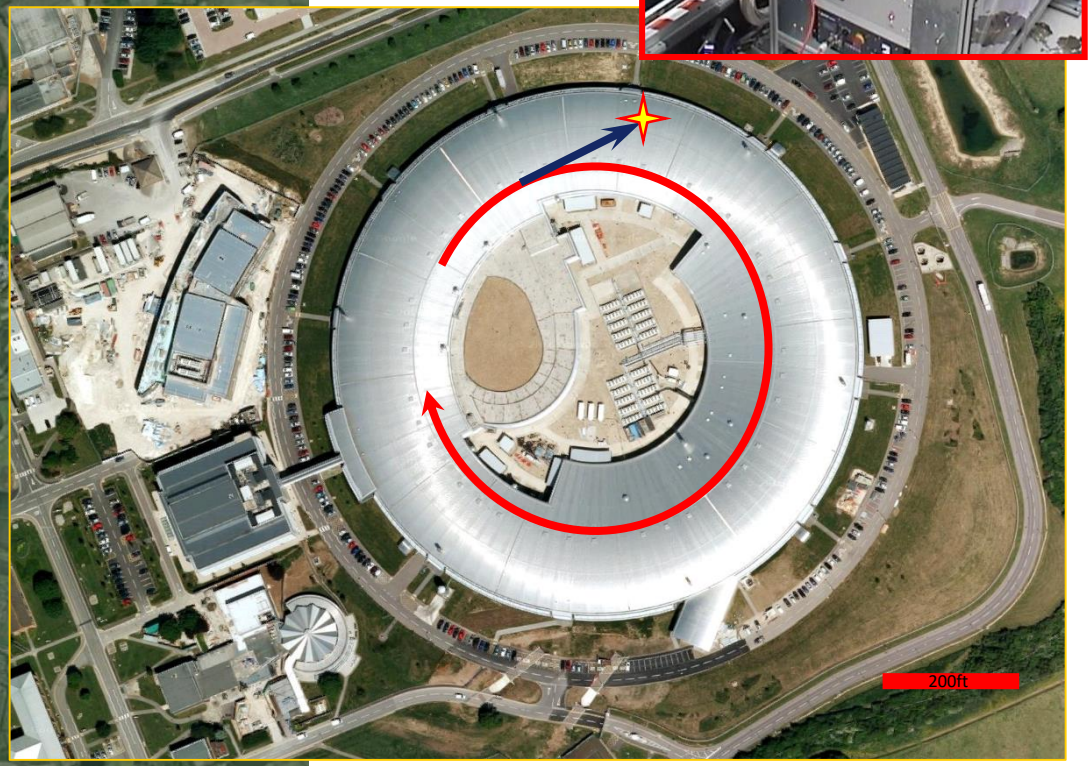
Computational strategies + Well-selected compounds



Rapid formulaic synthetic follow-up

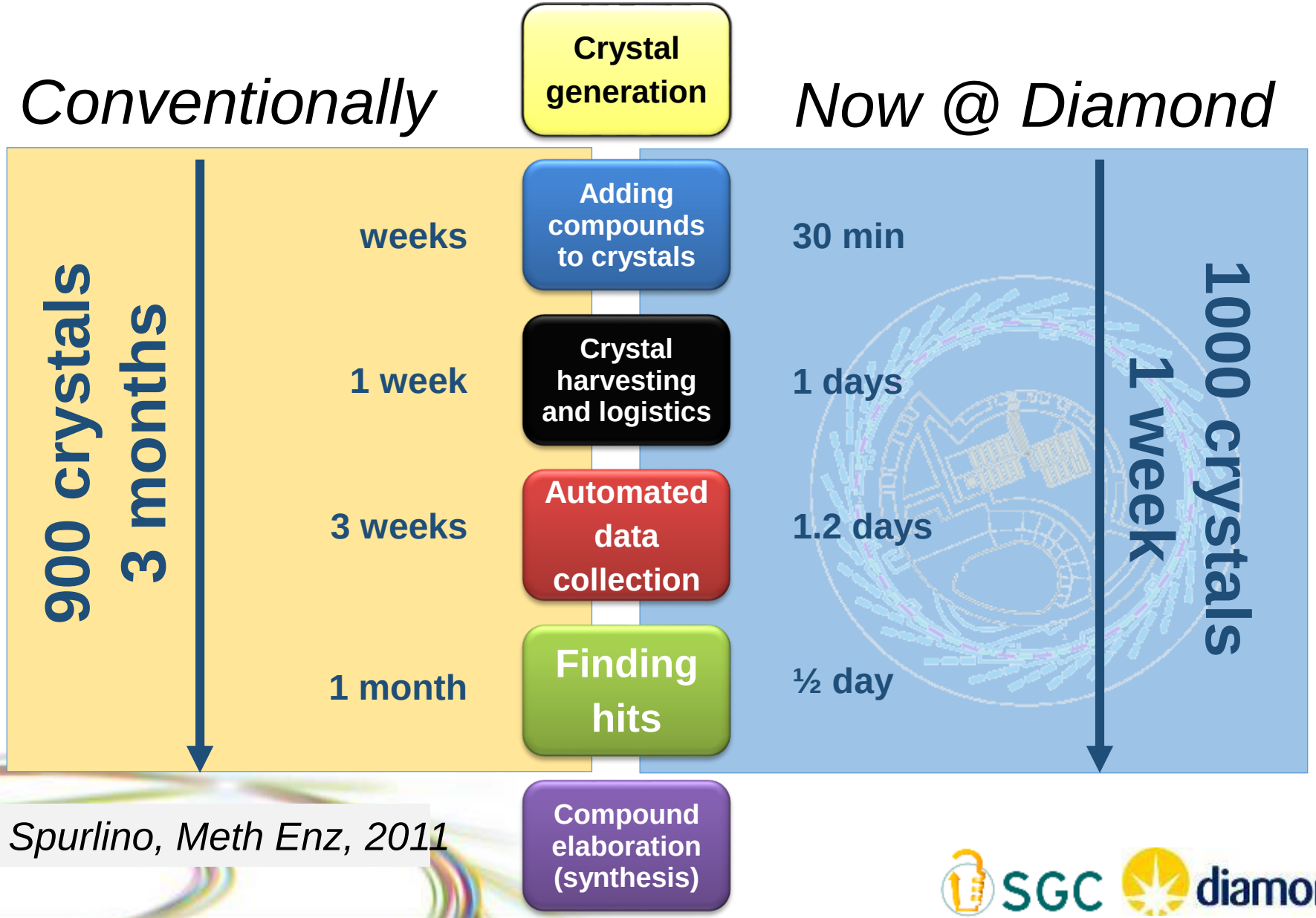
- **Confirm usefulness of weak binders** *(no assay!)*
- **Confirm relevance of *allosteric* binders** *(no assay!)*
- **Move to potency rapidly and cheaply**





SGC diamond

XChem: New Benchmark



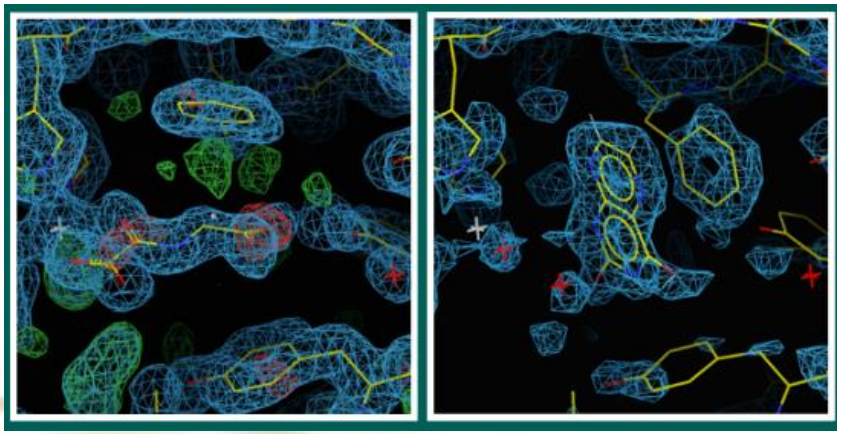
Spurlino, Meth Enz, 2011

SGC diamond

XChem highlights

- Users since October 2015
- Officially opened: November 2015 (media coverage)
- Users to date: 10 academic, 5 industry
- Targets: >20. Crystals: >13,000. Hits: >>250
- *cf.* Astex (industry leader since 2000) – ~4/year

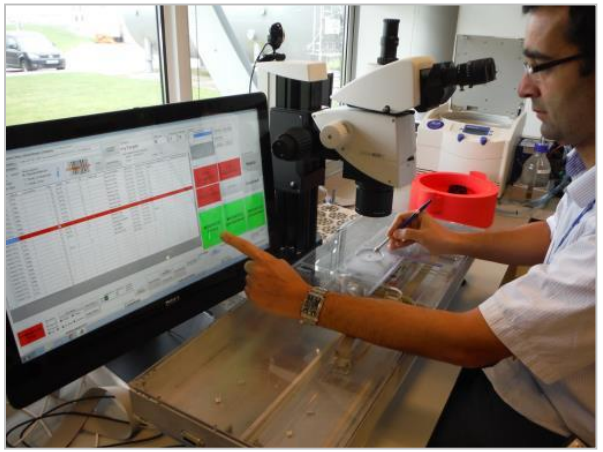
PANDDA: 3D background correction



Standard maps *Corrected maps*

New paradigm in crystallography

Shifter: robot-assisted crystal harvesting



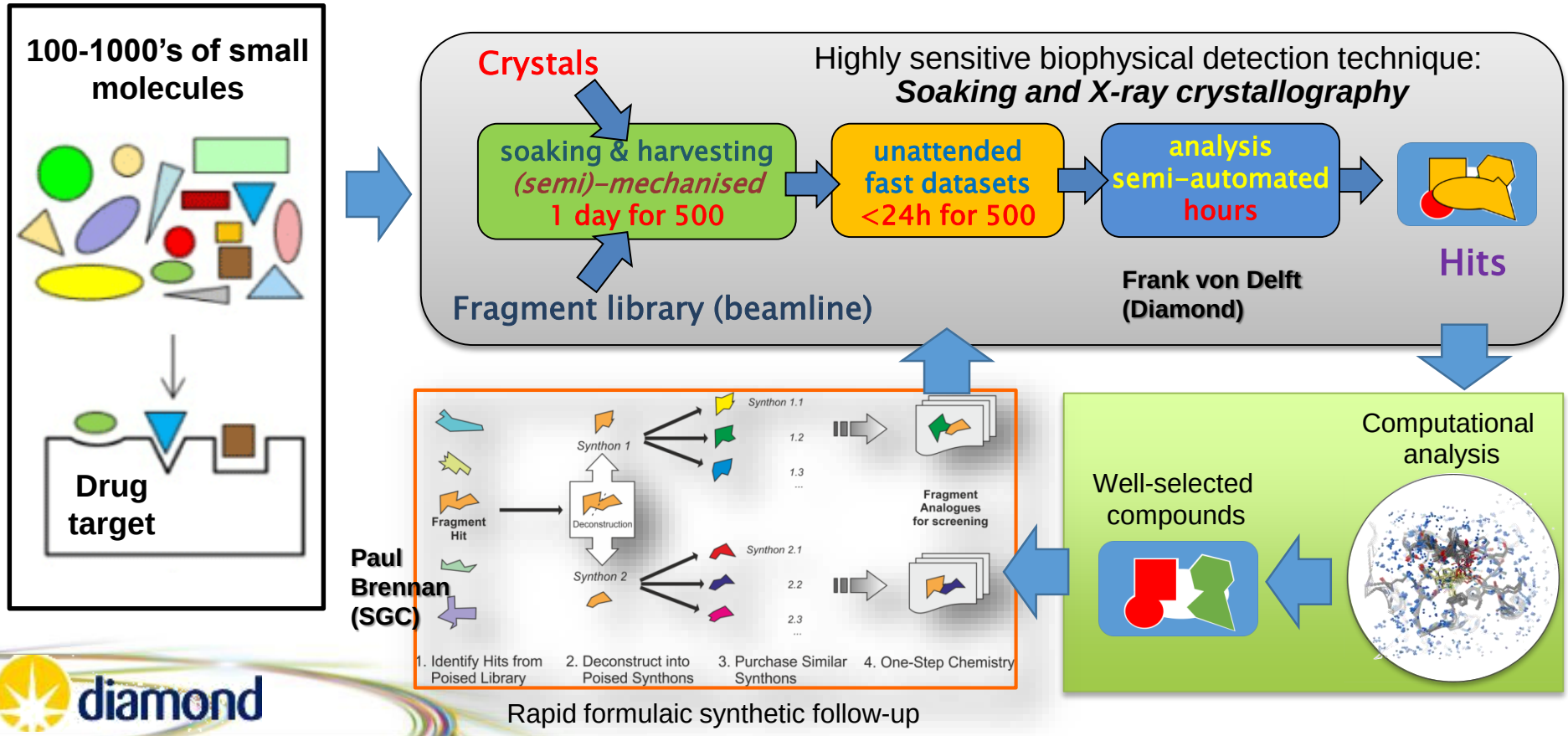
Up to 200 crystals / per hour
Fully recorded experiments

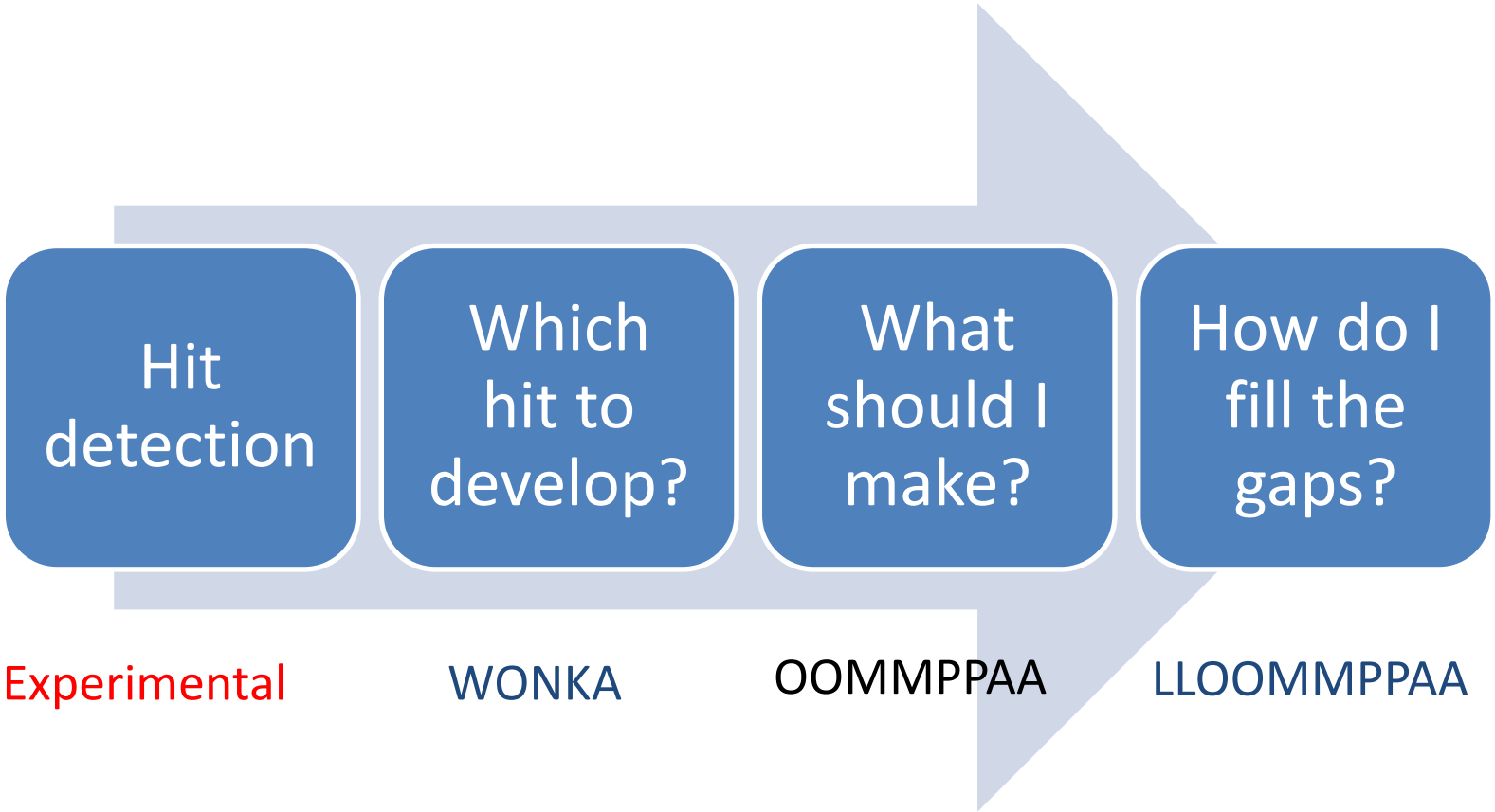
XCHEM – X-RAY GUIDED CHEMICAL SYNTHESIS

Problem: How can structural biology prime and iteratively accelerate early stage drug discovery in a timely fashion

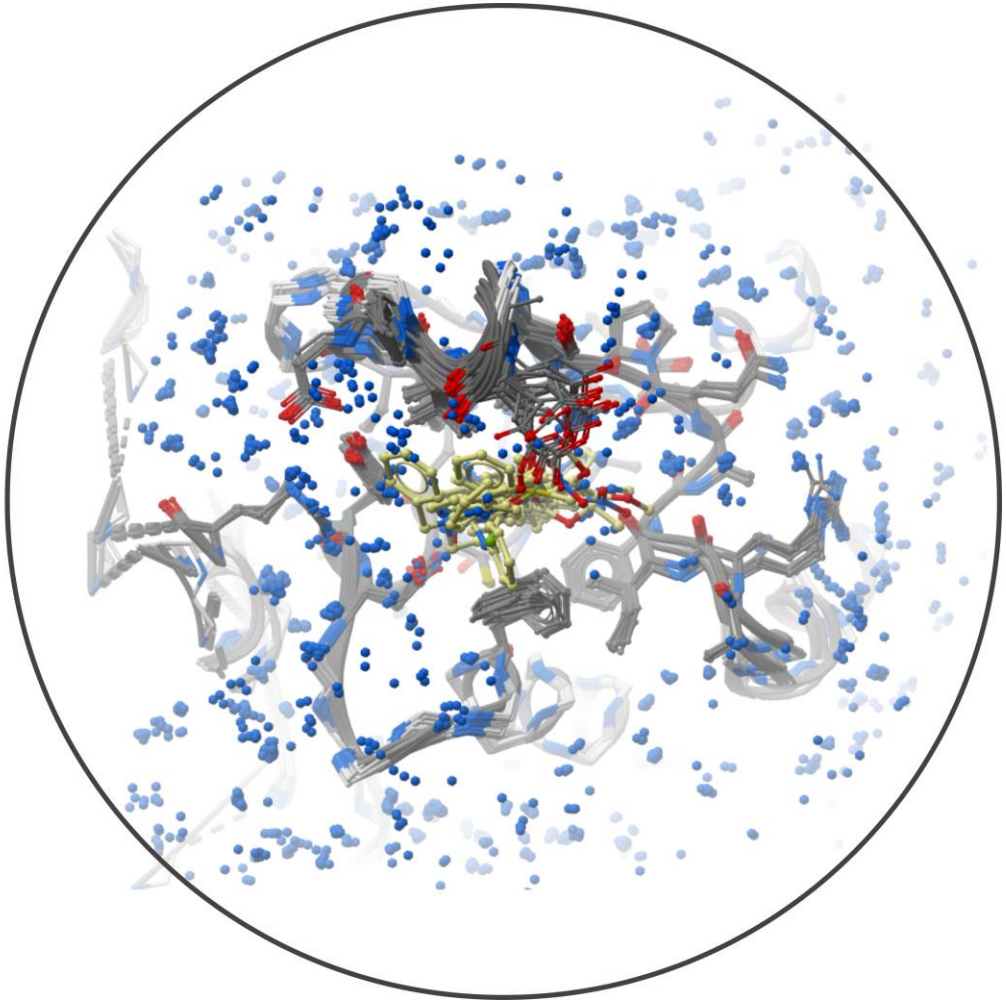
History: Protein crystallography and data analysis cycle times too long

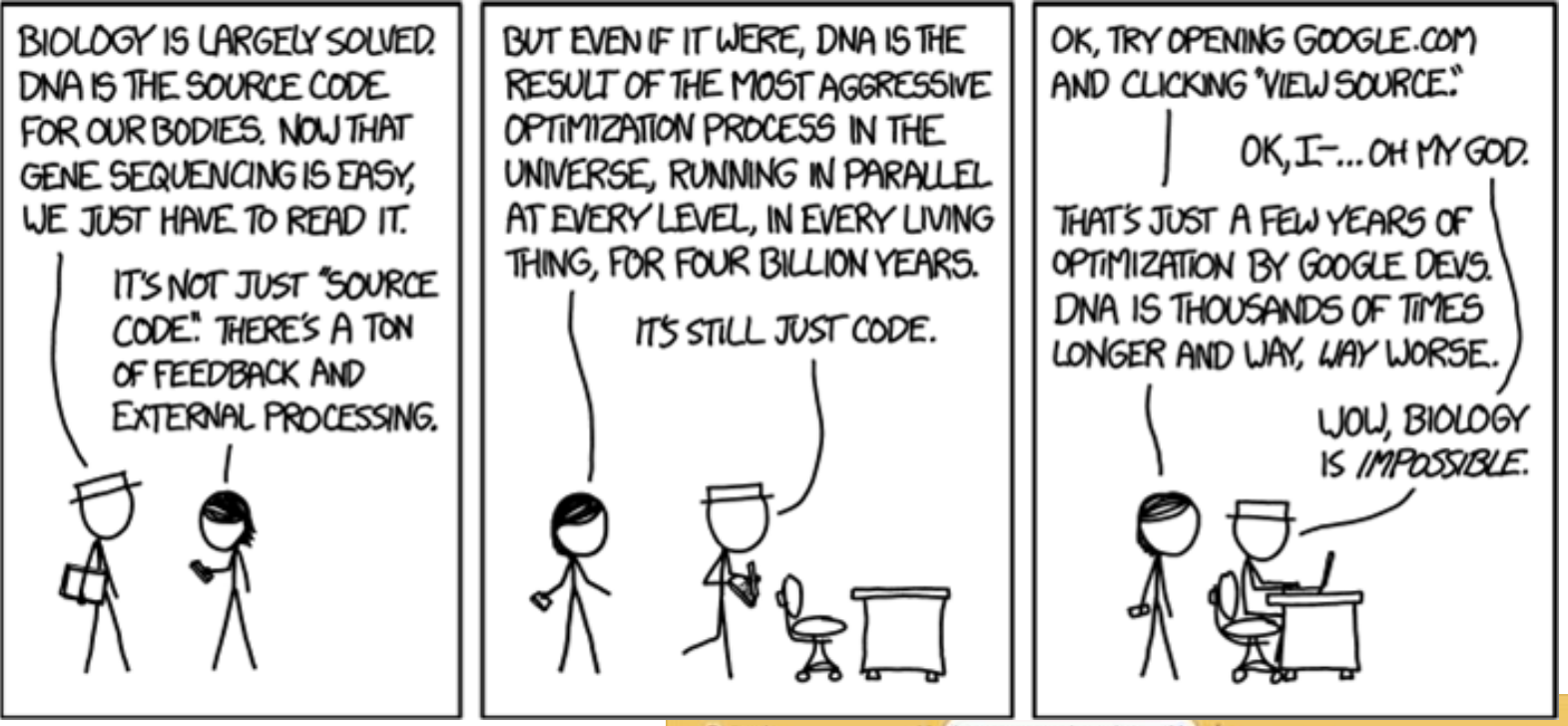
Solution: Diamond I04-1 beamline & SGC's XChem fragment-based soaking approach

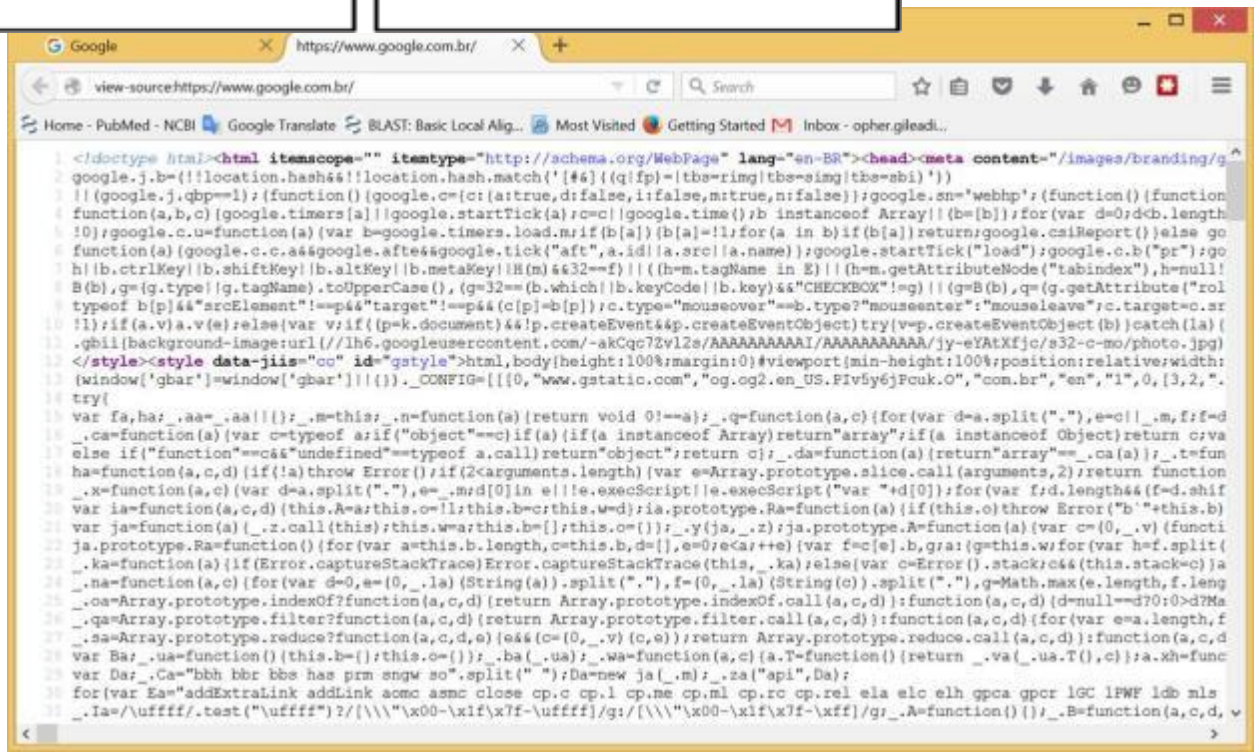



Hit detection
Which hit to develop?
What should I make?
How do I fill the gaps?
Experimental
WONKA
OOMMPPAA
LLOOMMPPAA

Experimental Data

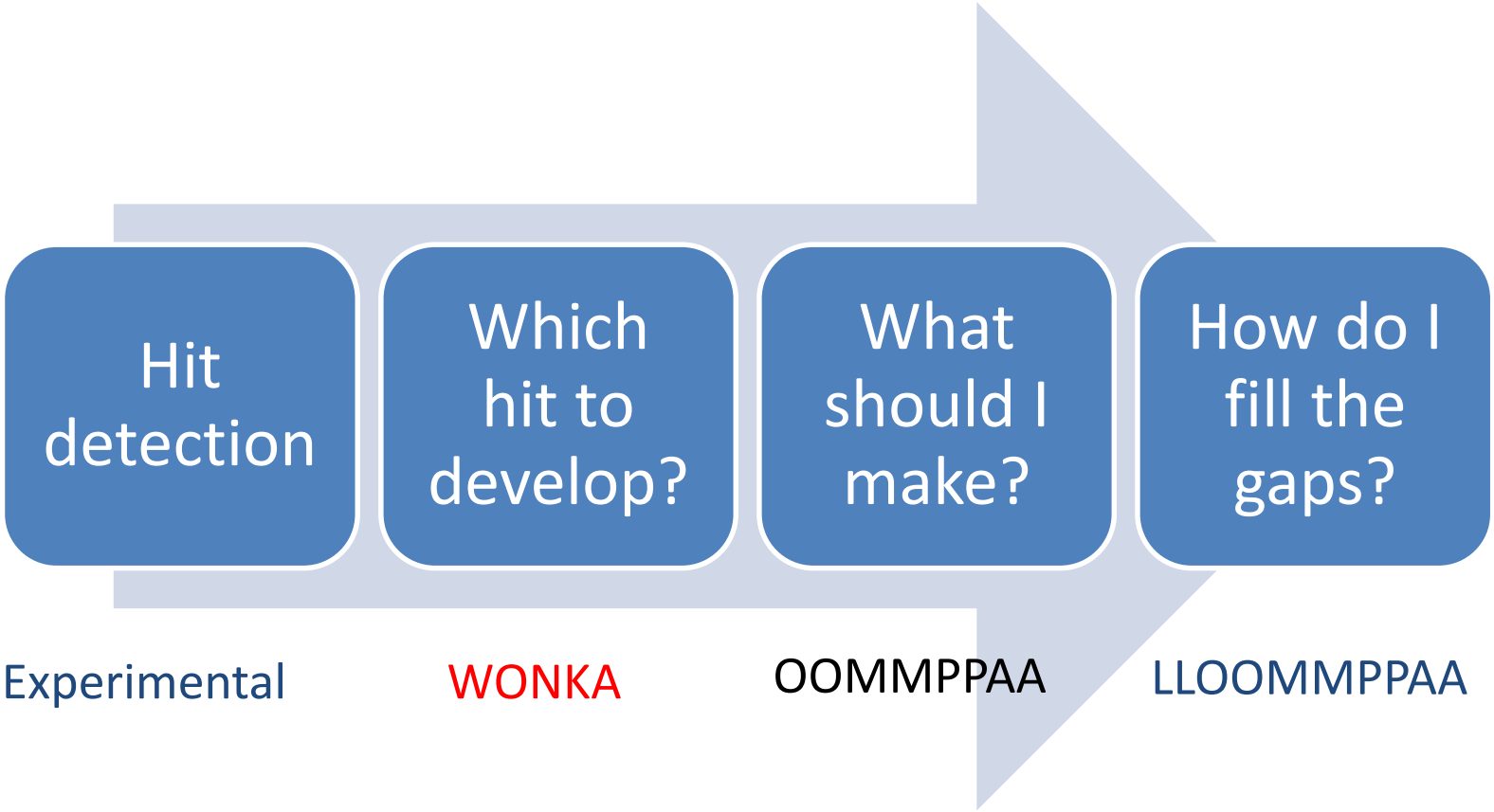






It's Biology.

Deal with it!


Hit detection
Which hit to develop?
What should I make?
How do I fill the gaps?
Experimental
WONKA
OOMMPPAA
LLOOMMPPAA

WONKA

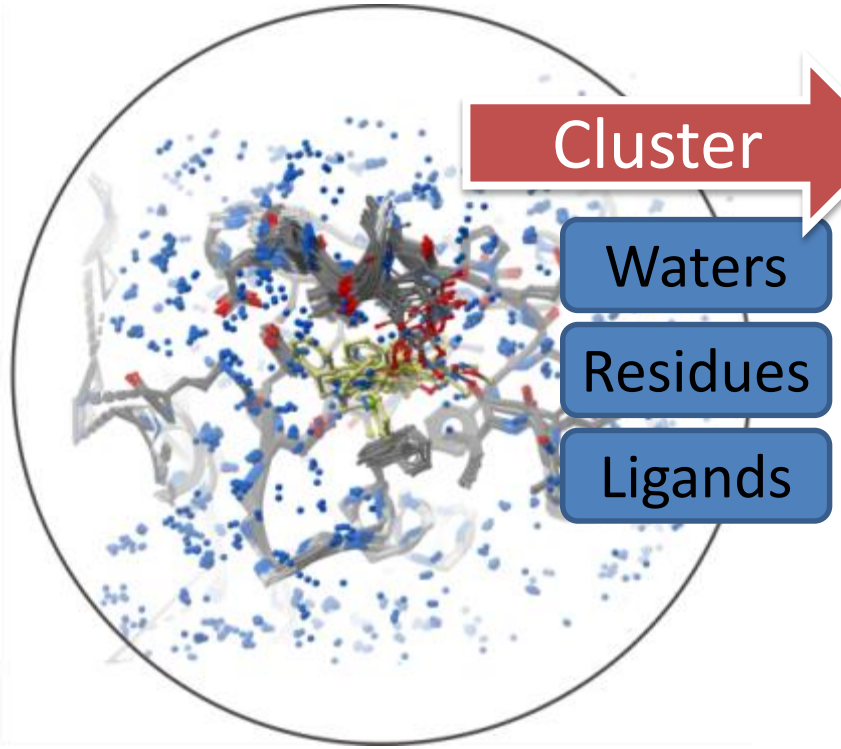
Make sense of this mess

- Automatically, objectively identify features
- Display, navigate
- Annotate, share, discussion thread

➔ *Natural end-point for screening experiment*

Anthony Bradley

http://wonka.sgc.ox.ac.uk



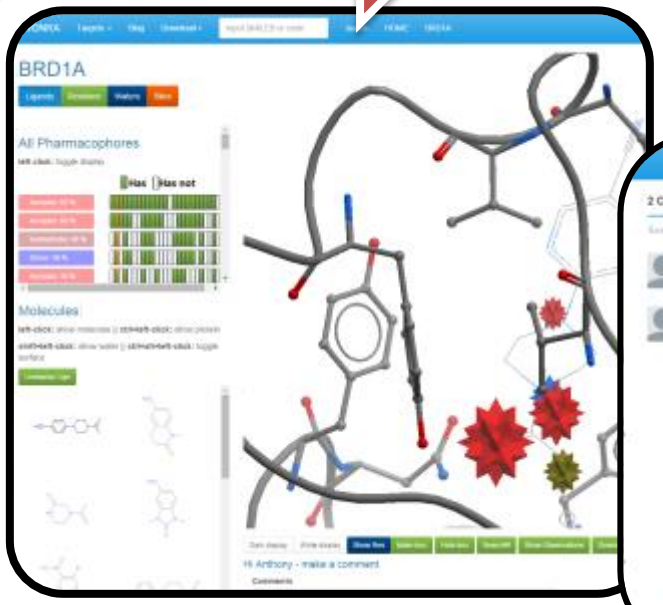
Cluster

Waters

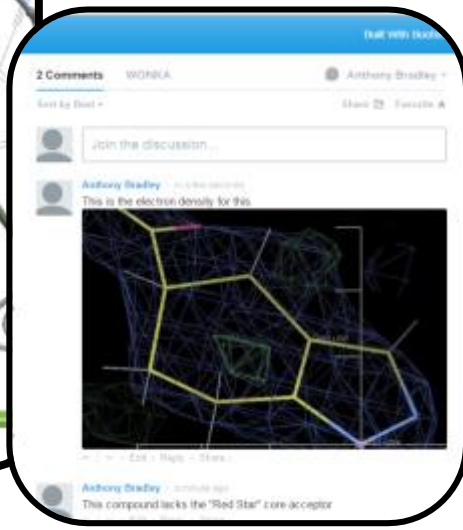
Residues

Ligands

View

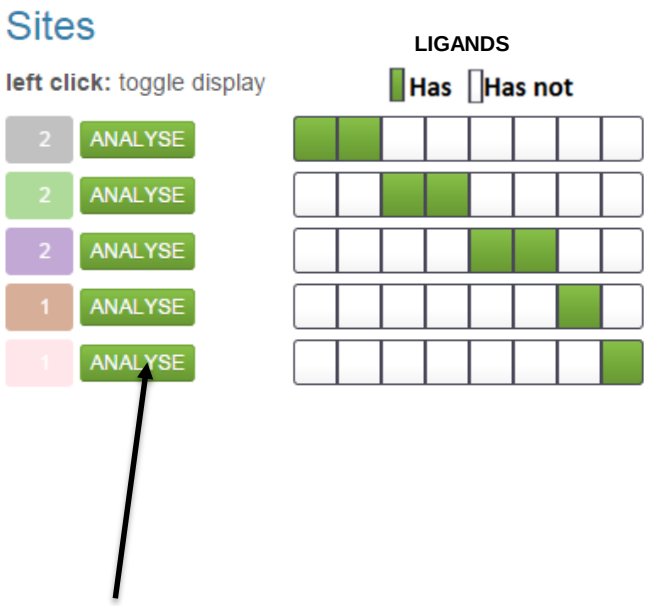


Annotate

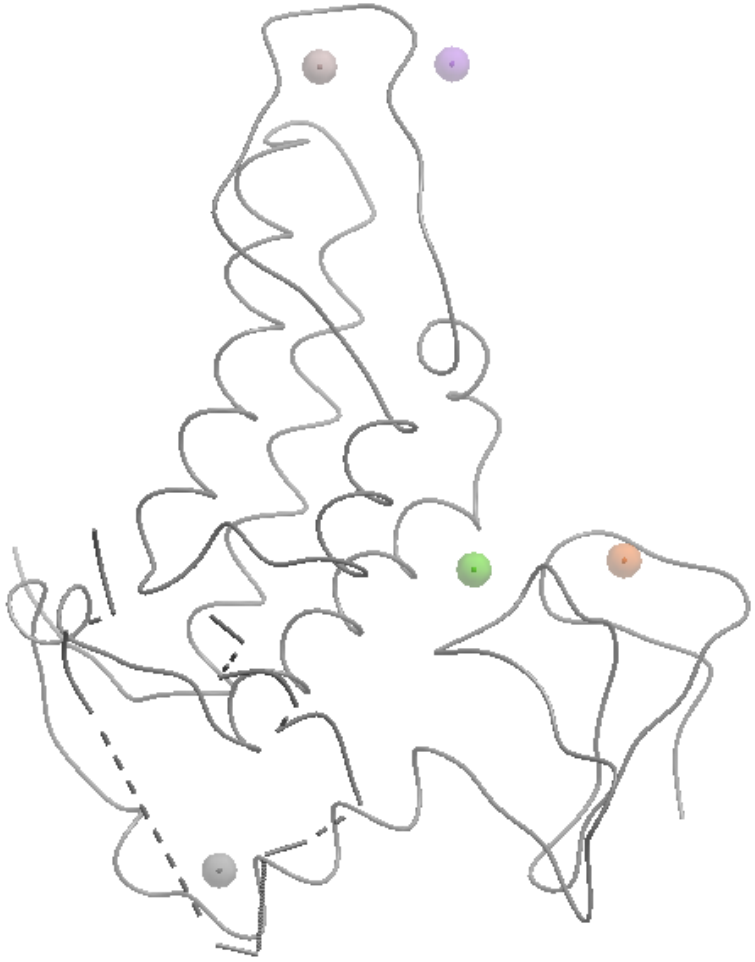


Journal of Computer-Aided Molecular Design **29**(10):963-973

WONKA – KEY BINDING SITES

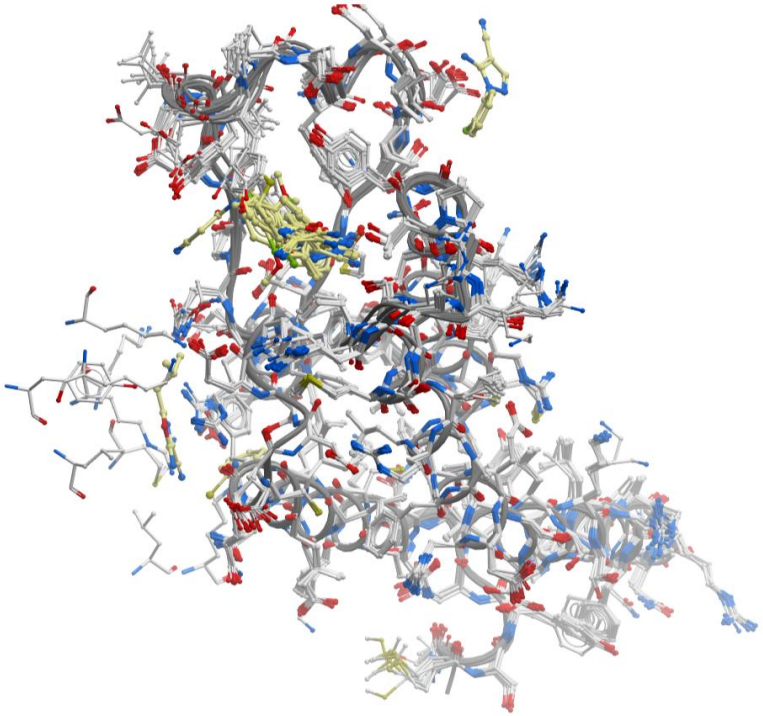


Site by site analysis

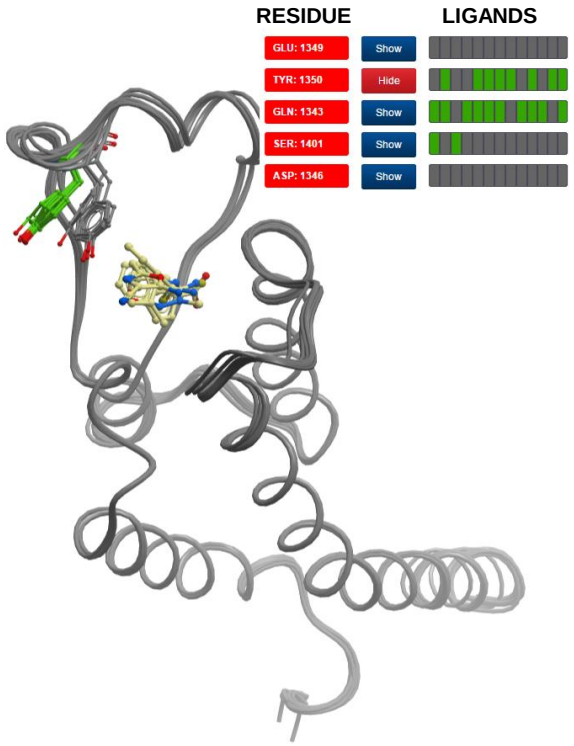


WONKA – KEY RESIDUES

Raw data

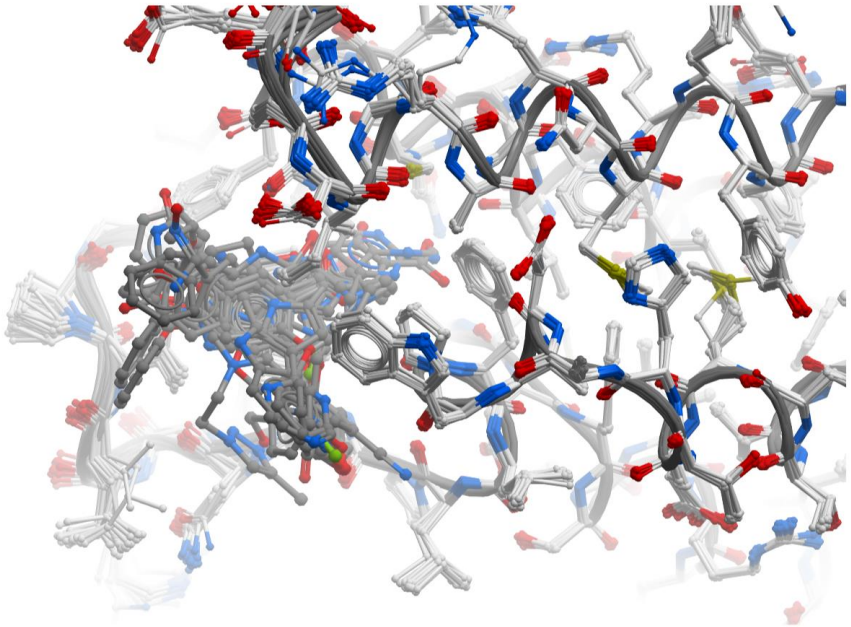


WONKA

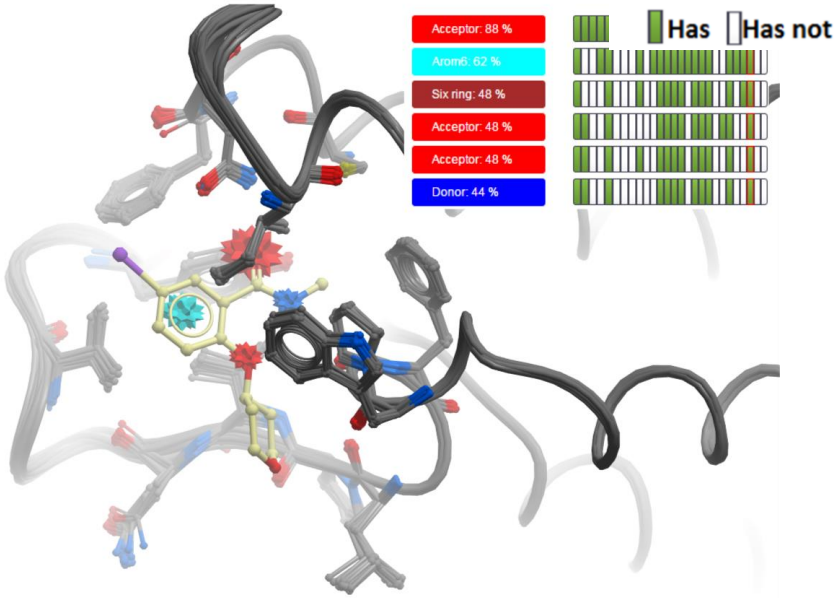


WONKA – LIGAND FEATURES

Raw data

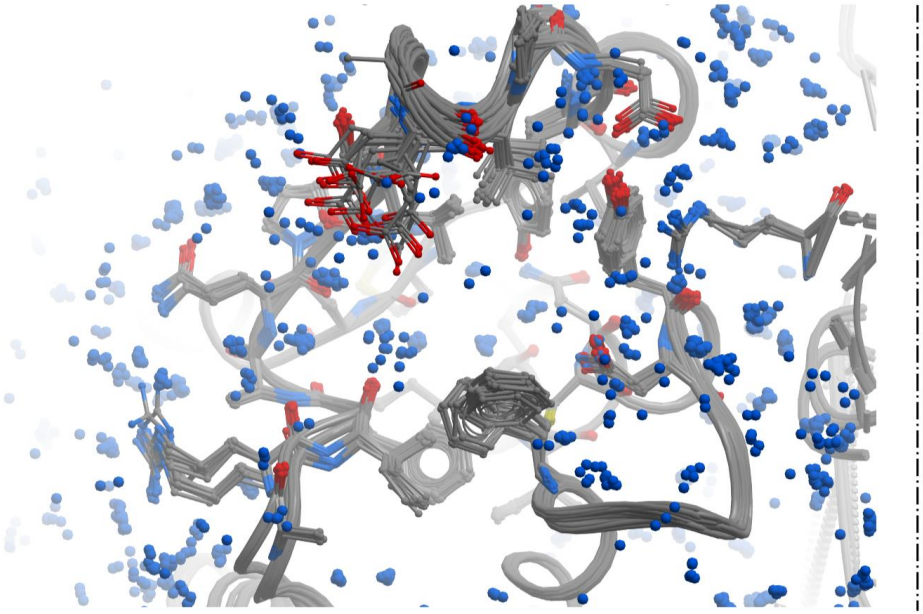


WONKA

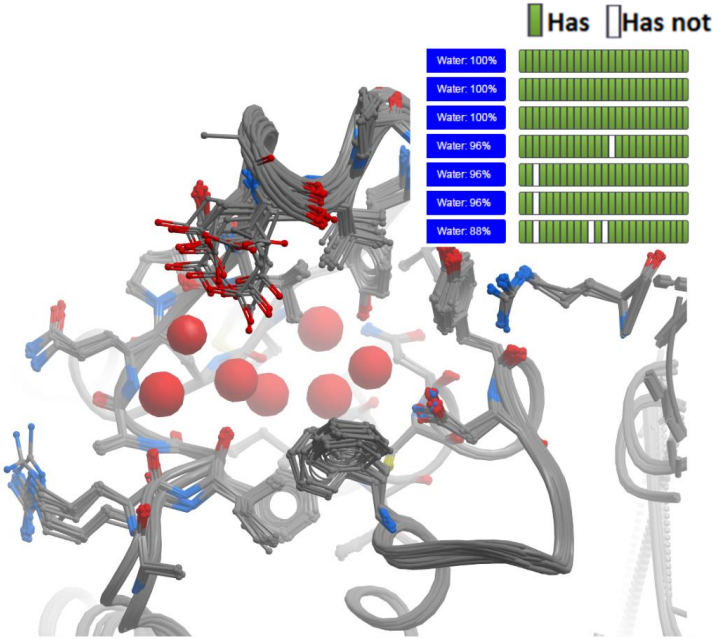


WONKA – Conserved Water Molecules

Raw data



WONKA

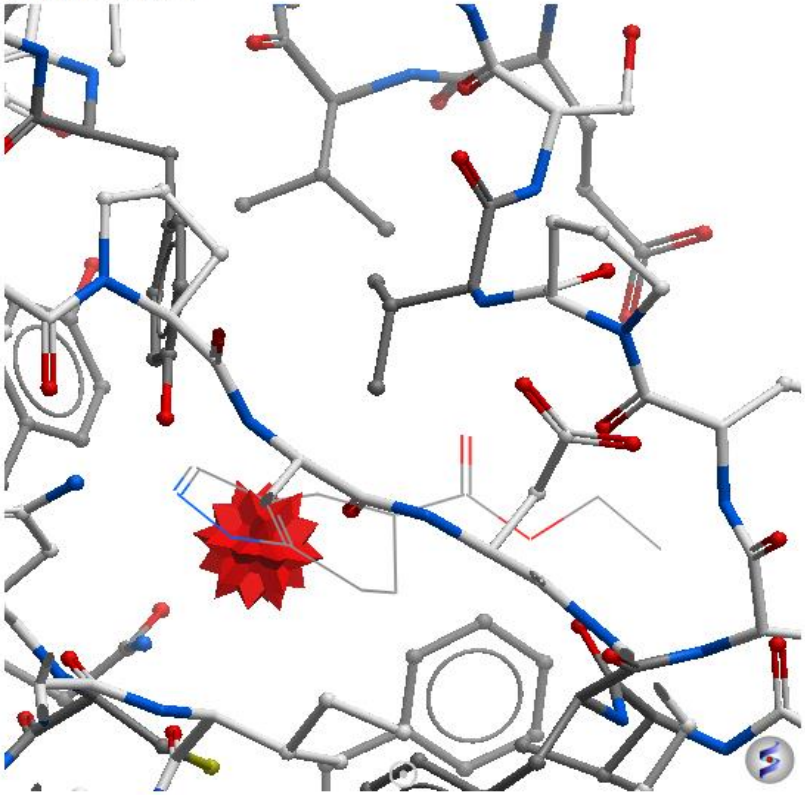


WONKA – Sharing Observations

WONKA Targets Blog Download HOME BRD1A Built With Bootstrap

Anthony:

Cool BRD1 structure



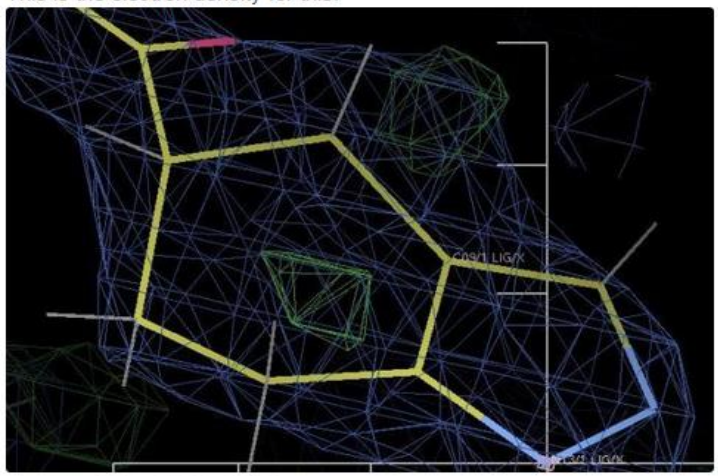
2 Comments WONKA Anthony Bradley

Sort by Best Share Favorite

Join the discussion...

Anthony Bradley - in a few seconds

This is the electron density for this.



Edit • Reply • Share ›

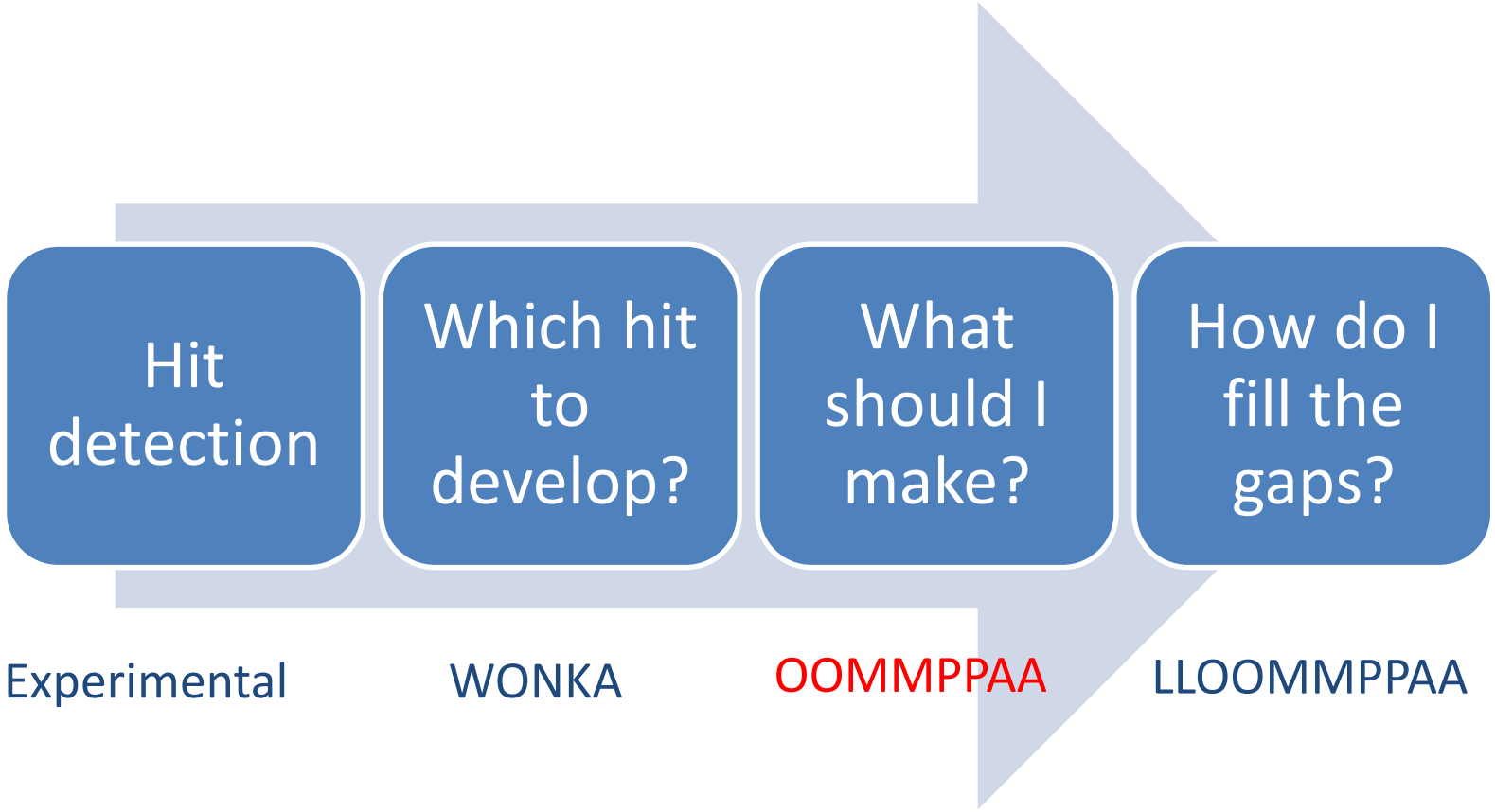
Anthony Bradley - a minute ago

This compound lacks the "Red Star" core acceptor

Edit • Reply • Share ›

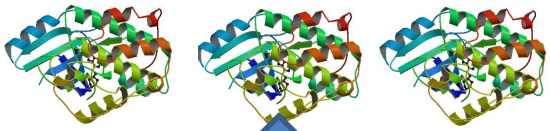
DISQUS

OOMMPPAA



http://oommppaa.sgc.ox.ac.uk

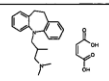
OOMMPPAA – 3D MMP Analysis



CONTEXT

3D MMP analysis

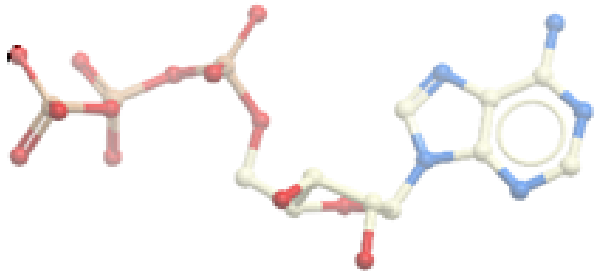
ACTIVITY

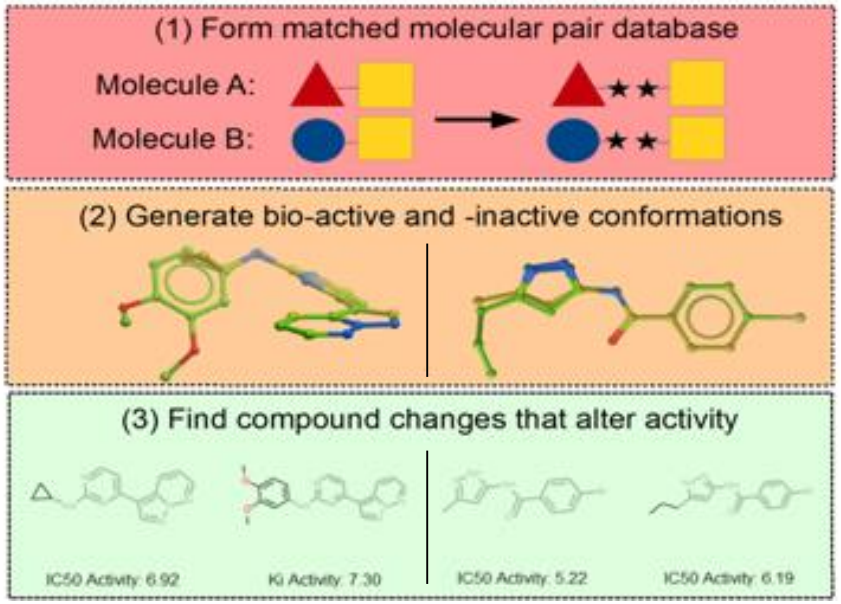
Structure	MW	% Inh. @ 1 µM	% Inh. @ 5 µM	IC_{50} range (µM)
	263.39	39	58	1-2.5
	313.87	51	65	<1
	410.52	28	48	1-2.5
	351.32	50	64	<1
	316.88	34	60	<1
	302.85			2.5-5

OOMMPPAA – Exploring data in 3D space

a) **Activity data**

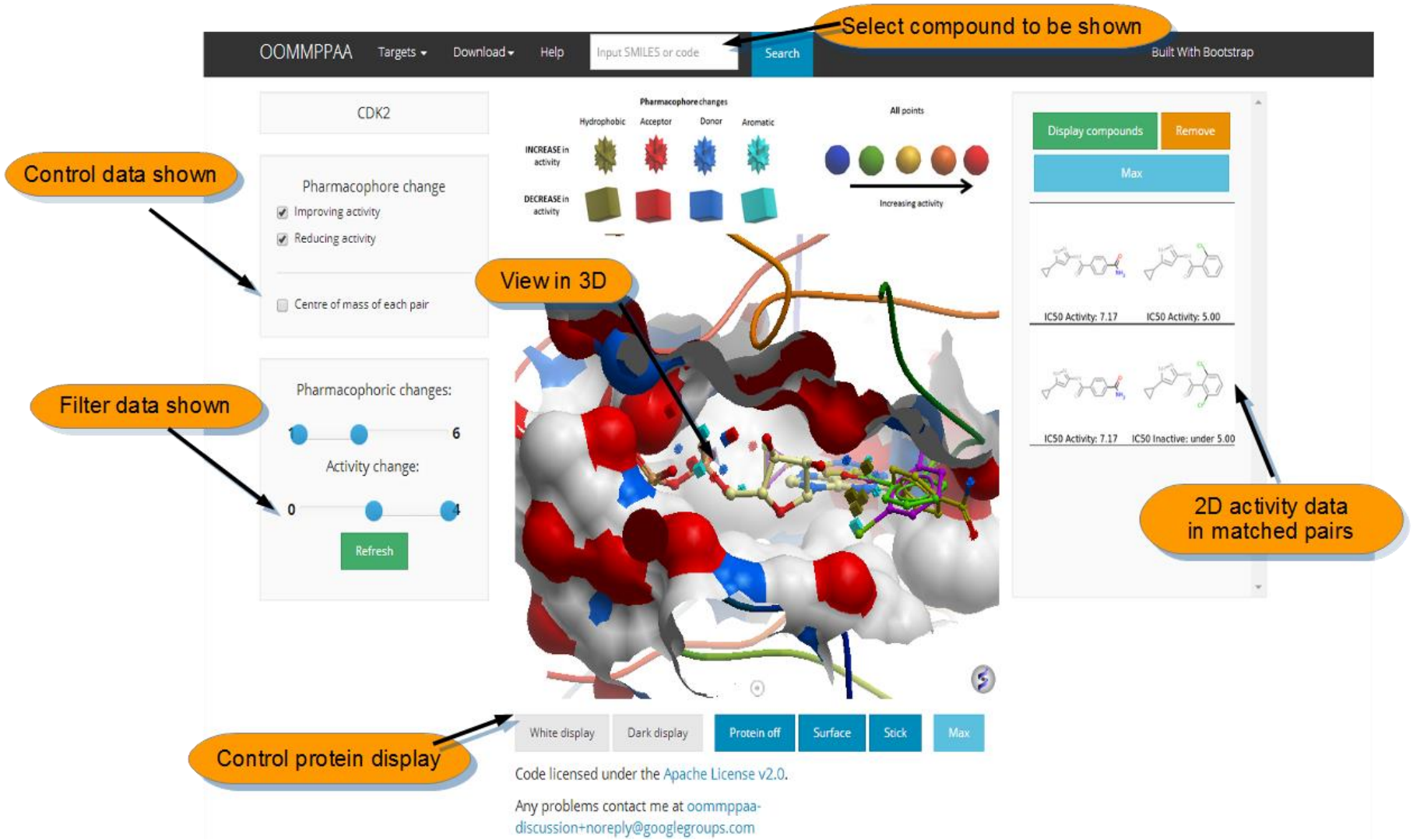
b) **Complexed ligand structures**





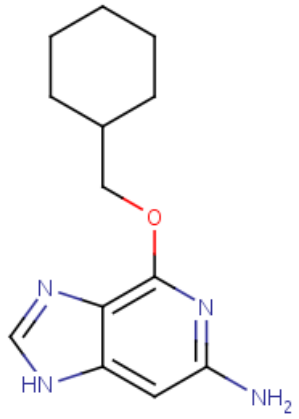
Journal of Chemical Information and Modeling 54(10):2636-2646

OOMMPPAA – Interactive Visualisation

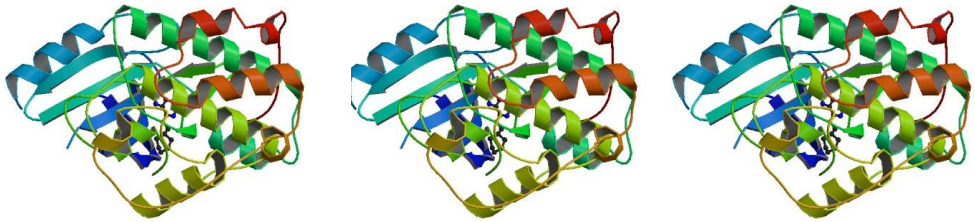


http://oommppaa.sgc.ox.ac.uk

Lead – IC50 13µM

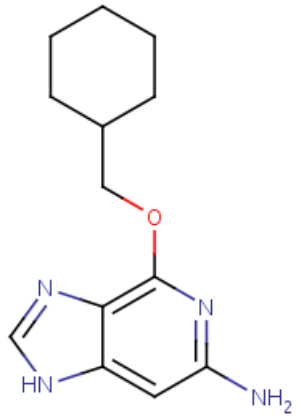


Structural data

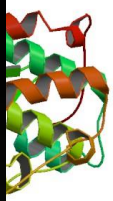


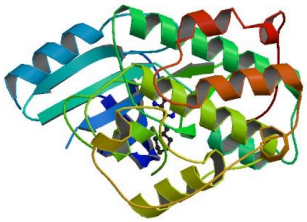
Activity data

Lead – IC50 13μM



Structural data

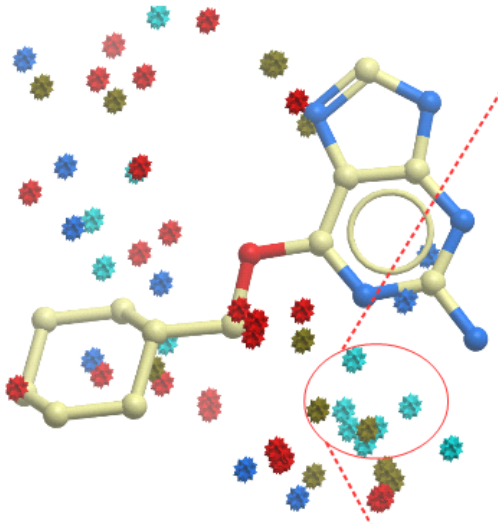




data

What compound do we make next?????????

OOMMPPAA – CONDENSING SAR



Pharmacophore changes
Increase in activity

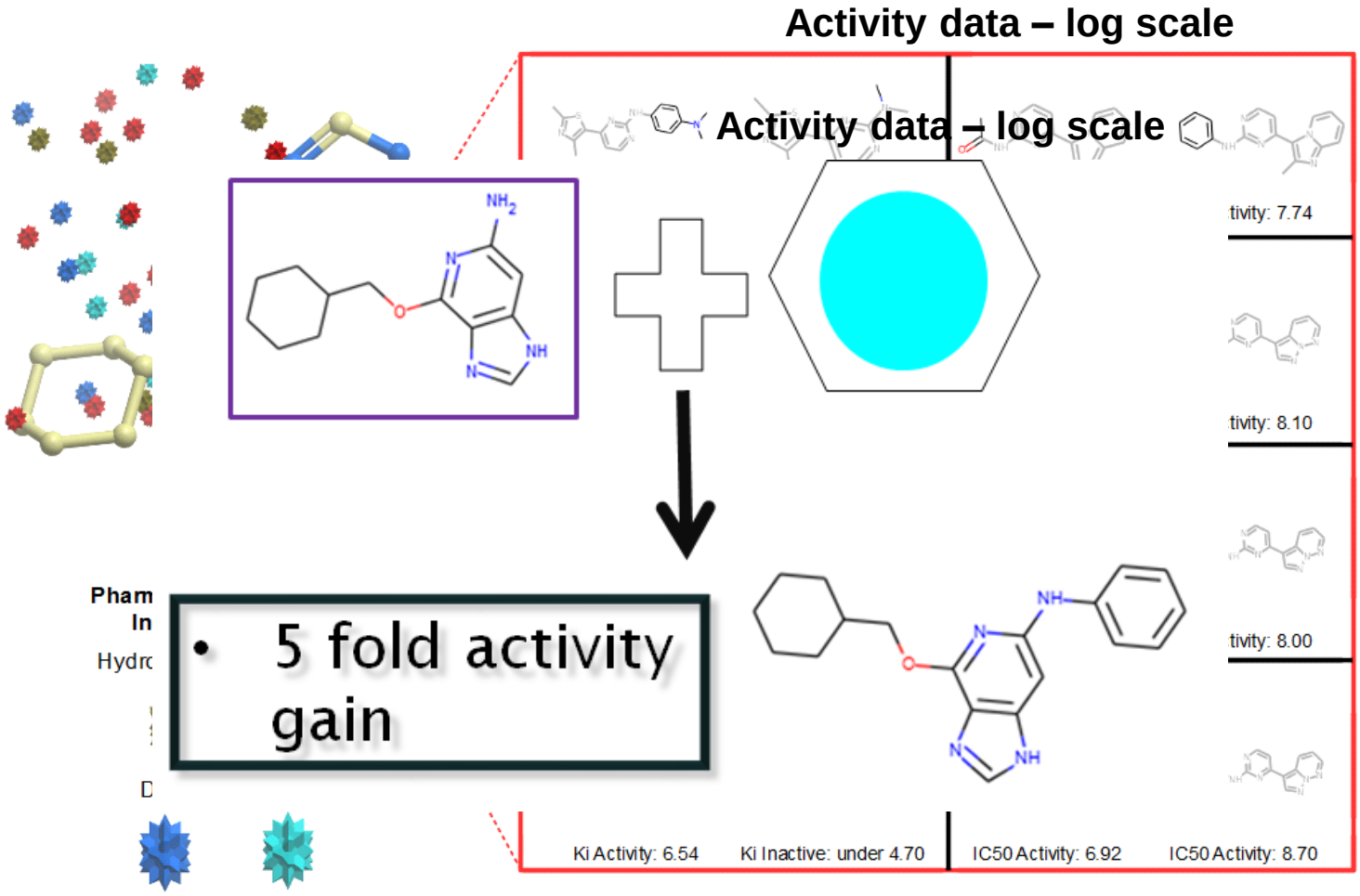
Hydrophobic Acceptor

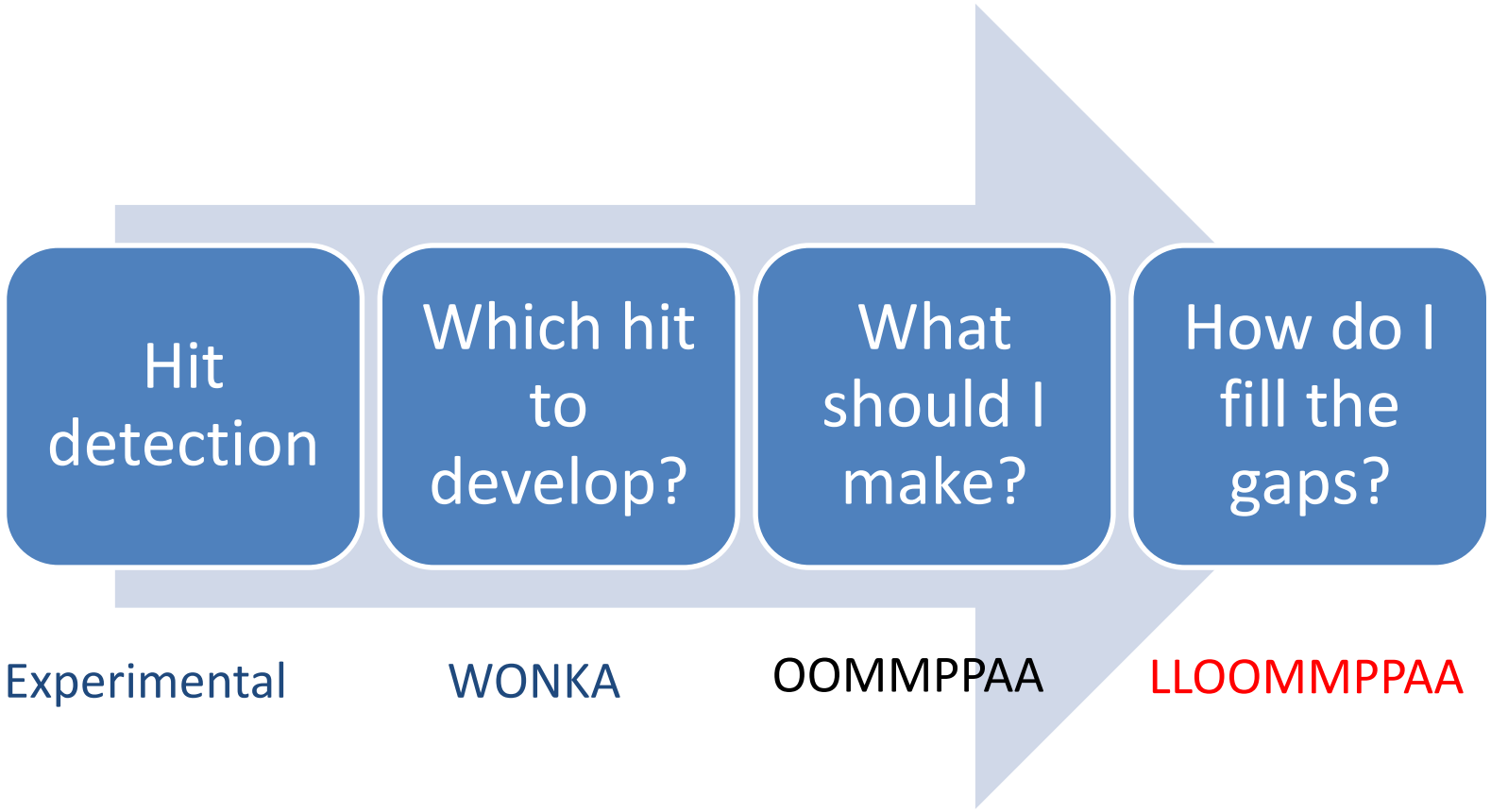
Donor Aromatic

Activity data – negative log scale

Ki Activity: 6.60	Ki Inactive: under 4.70	IC50 Activity: 5.54	IC50 Activity: 7.74
IC50 Activity: 6.92	IC50 Activity: 8.52	IC50 Activity: 6.92	C50 Activity: 8.10
Ki Activity: 6.60	Ki Inactive: under 4.70	IC50 Activity: 6.92	IC50 Activity: 8.00
Ki Activity: 6.54	Ki Inactive: under 4.70	IC50 Activity: 6.92	IC50 Activity: 8.70

OOMMPPAA – CONDENSING SAR




Hit detection
Which hit to develop?
What should I make?
How do I fill the gaps?
Experimental
WONKA
OOMMMPPAA
LLOOMMMPPAA

OAKLEY SAYS – HOW SHOULD I FILL THE GAPS?

Oakley Cox – Medicinal Chemist

We tried cloning Oakley...

Idea 2: Use Computers

Chemical universe

- 10^{60} Compounds
- Huge inaccessible space

Accessible universe

- What could I make?
  - Hit discovery
  - *In silico* library design
  - Medicinal chemists

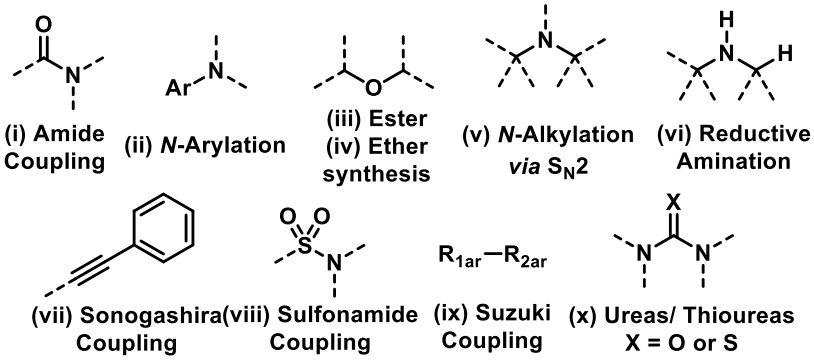
Efficient universe

- Which thirty compounds most efficiently explore all the potential options?

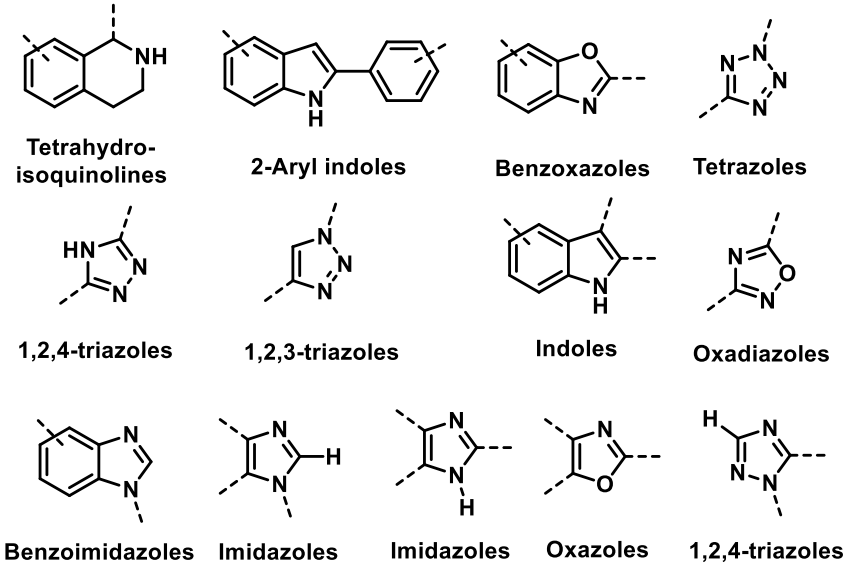
Poised Fragments

62% of med chem reactions performed in 2008 could be classified by one of ten reactions

Core Scaffolds



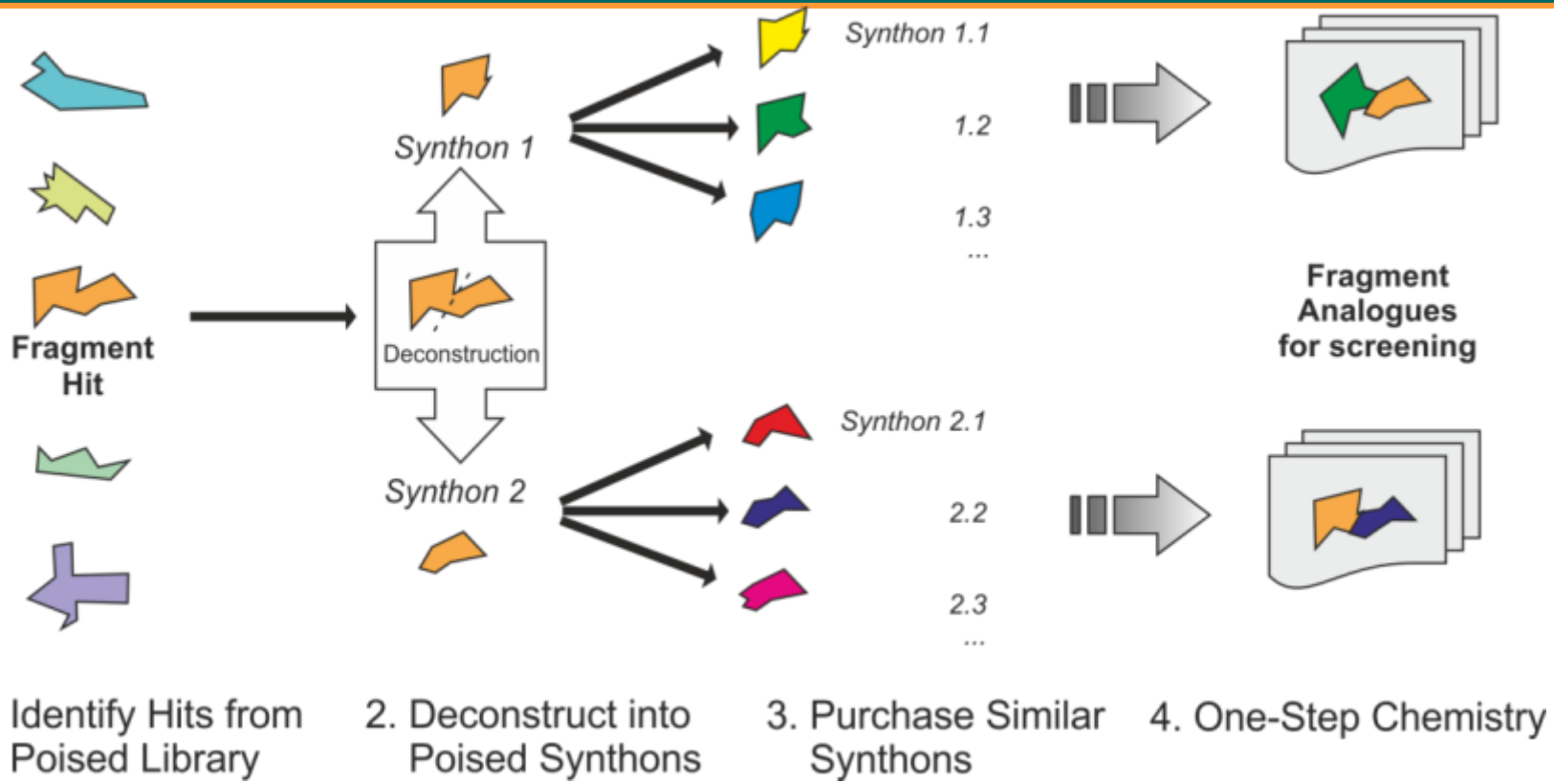
(xi) *Heterocycle Formation*



S. D. Roughley & A. M. Jordan, *J. Med. Chem.*, 2011, **54**, 3451-3479

M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, E. Jacoby and S. Renner, *J. Chem. Inf. Model.*, 2012, **52**, 1167-1178

Oakley – Using Poised Fragments



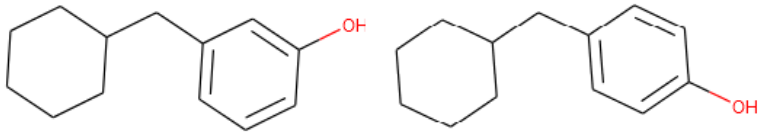
Poised reactions are:

- robust and reliable;
- make drug-like products;
- be possible using commercially available starting points;
- compatible with a range of substrates.

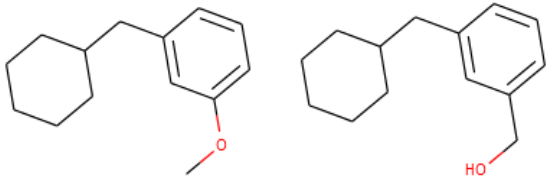
Oakley Cox
(SGC: Paul Brennan)

Defining relevant diversity

2D Diversity



3D Shape diversity



3D Pharmacophore diversity

- But are the changes relevant?

Protein-ligand interaction diversity

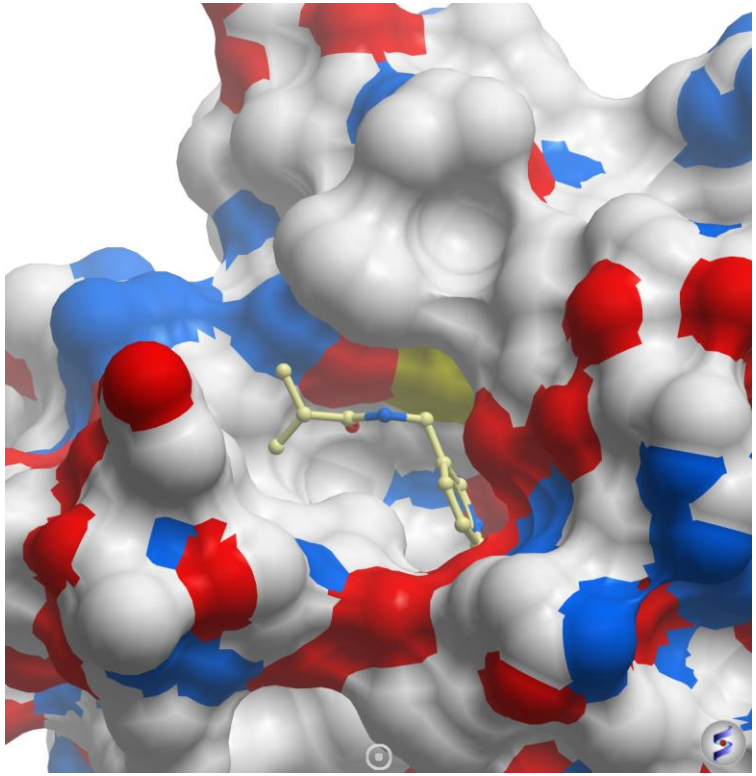
Relevance

Example

- Morgan fingerprint
- MACCS
- Shape Tanimoto
- USR CAT
- LLOOMMPPAA

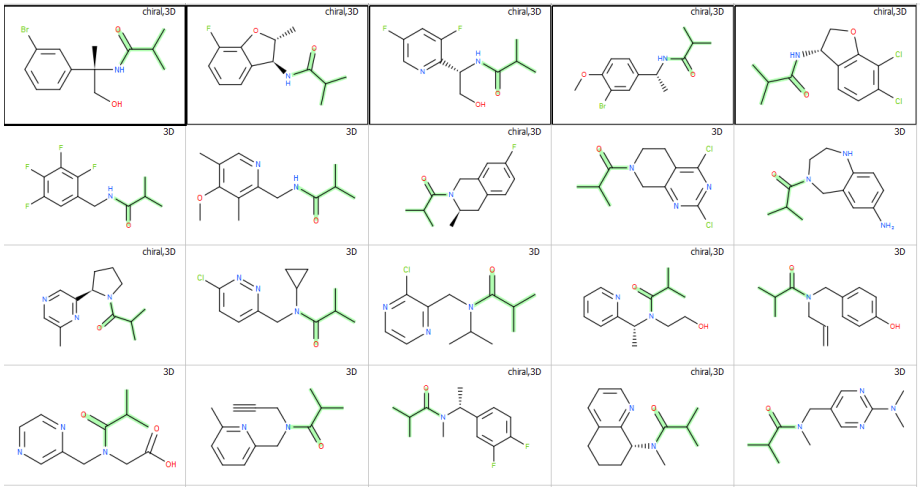
LLOOMMPPAA - METHOD

Complexed ligand structure



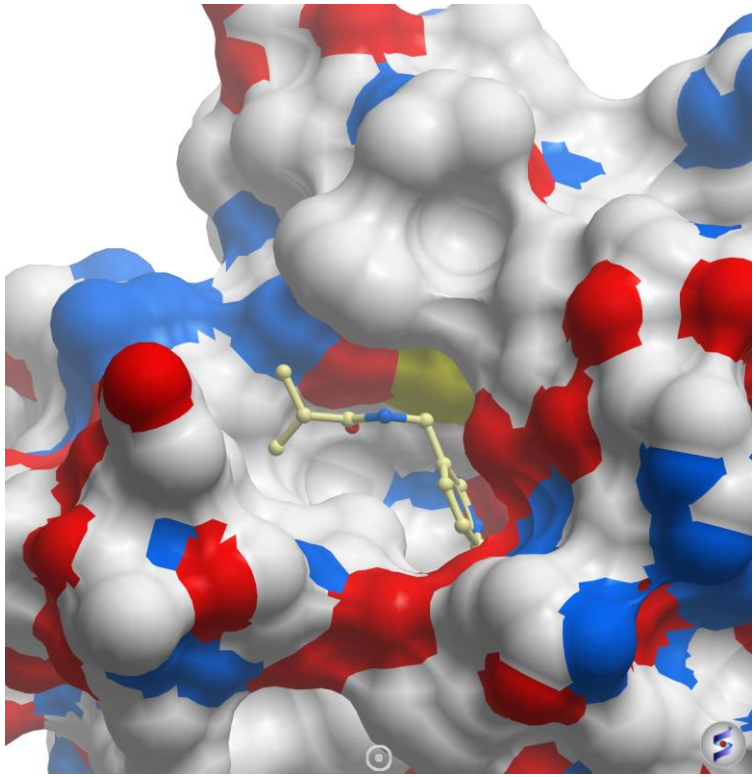
1)

Define follow ups – simple reactions (amide coupling)



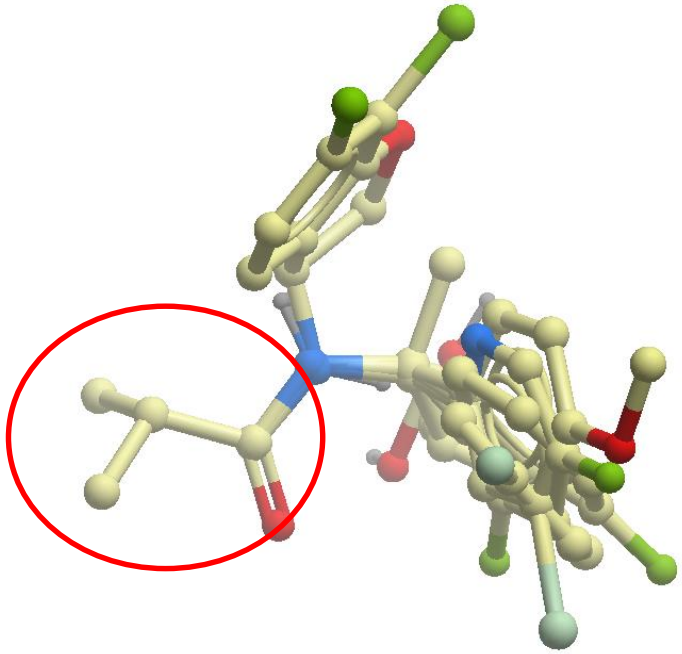
LLOOMMPPAA - METHOD

Complexed ligand structure



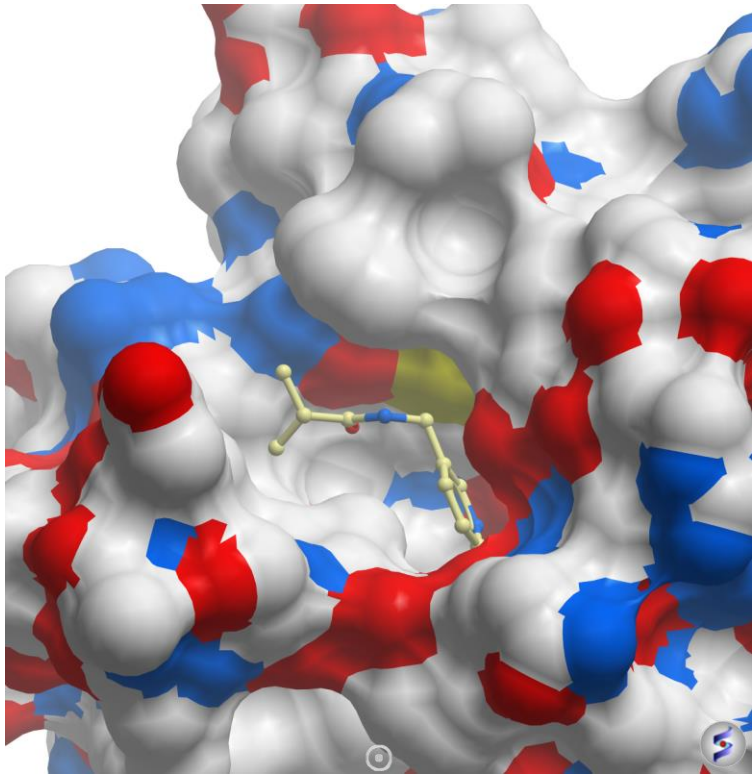
1) **Define follow ups**

2) **Generate diverse conformations – keep shared core static**



LLOOMMPPAA – Method

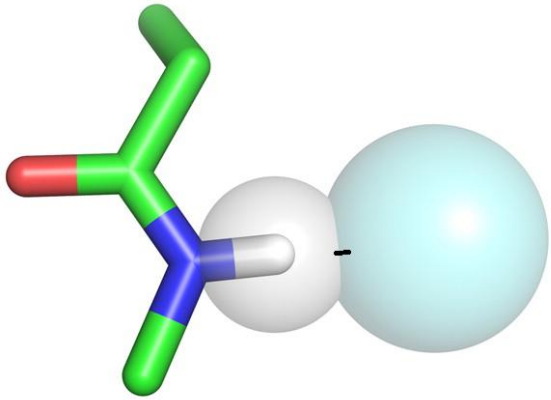
Complexed ligand structure



1) **Define follow ups**

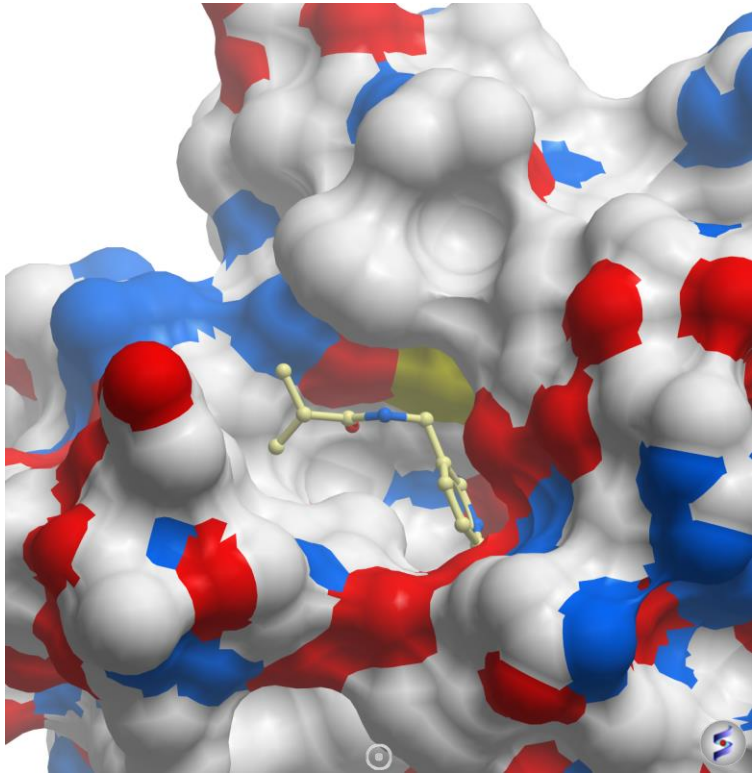
2) **Generate diverse conformations**

3) **Remove VdW clashes with protein**



LLOOMMPPAA - METHOD

Complexed ligand structure



1) **Define follow ups**

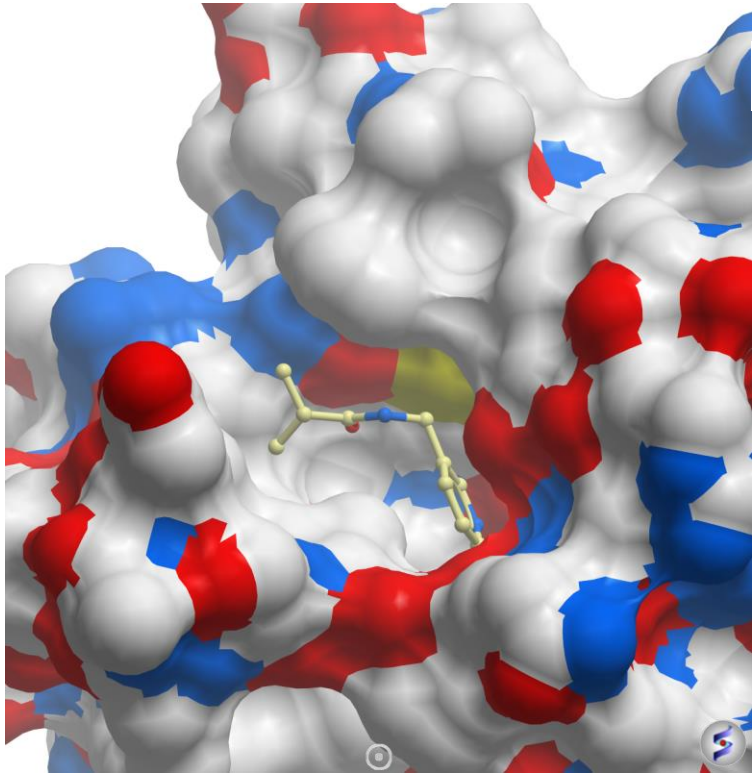
2) **Generate diverse conformations**

3) **Remove VdW clashes with protein**

4) **Find protein-ligand interaction fingerprints (PLIFS)**

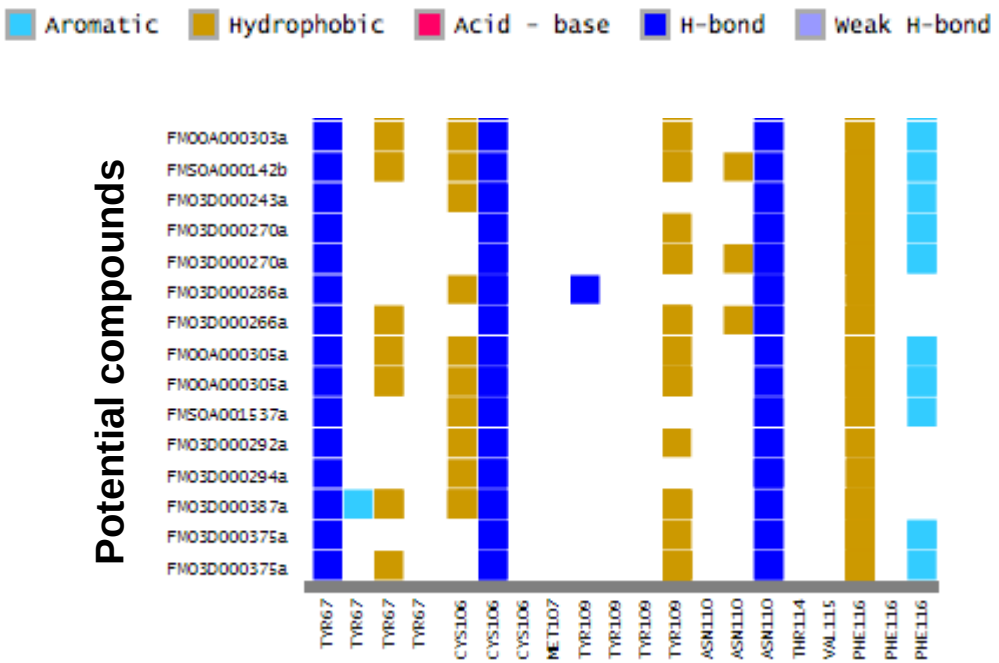
LLOOMMPPAA - METHOD

Complexed ligand structure



4)

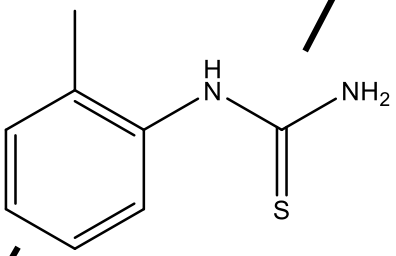
Find protein-ligand interaction fingerprints (PLIFS)



PHIP(2) - Thiourea Poised follow-ups

ID	IC_{50} (µM)	LE
FMOOA463	68	0.45
FMOOA365	128	0.45
FMOOA473	128	0.29
FMOOA315	140	0.36
FMOOA314	143	0.45
FMOOA470	194	0.35
FMOOA312	256	0.42
FMOOA476	276	0.45
FMOOA475	294	0.41
FMOOA471	295	0.35
FMOOA474	347	0.24
FMOOA316	414	0.39
FMOOA472	415	0.47
FMOOA465	423	0.34
FMOOA310	482	0.39
FMOOA466	536	0.31
FMOOA313	565	0.32
FMOOA468	566	0.38
FMOOA309	747	0.36
XST00942	768	0.4

Ten-fold increase in potency



Original fragment hit

ID	IC_{50} (µM)	LE
XST00942	768	0.4
FMOOA462	791	0.31
FMOOA477	1021	0.28
FMOOA308	1200	0.34
FMOOA467	1291	0.34
FMOOA363	1500	0.32
FMOOA357	1700	0.35
FMOOA321	1900	0.35
FMOOA369	2100	0.29
FMOOA319	2100	0.27
FMOOA358	2200	0.29
FMOOA362	2600	0.31
FMOOA367	2600	0.26
FMOOA311	3200	0.29
FMOOA359	3200	0.29
FMOOA361	4800	0.27
FMOOA360	>5000	<0.29
FMOOA461	>5000	<0.29
FMOOA460	>5000	<0.27
FMCAC421	>5000	<0.25
FMOOA368	>5000	<0.25
FMOOA364	>5000	<0.25
FMOOA318	>5000	<0.25
FMOOA317	>5000	<0.25
FMOOA366	>5000	<0.23
FMOOA320	>5000	<0.23
FMOOA469	>5000	<0.23
FMOOA464	>5000	<0.21

LLOOMMPPAA

CCF Thoughts

- WONKA/OOMMPPAA/LLOOMMPPAA
  - In use at a number of big pharma (Novartis, Roche, Merck, GSK)
- Build interactive interface to focus on the questions that medchemists are interested in
- Use technologies that provide beautiful and functional interfaces that work across architectures – activeICMJS!
  - Nothing else out there so feature-rich
  - Means we do not need to reinvent the wheel. Again.
- BUT: release early, release often

Supporting The Tools

- How to maintain and integrate these tools in existing workflows?
- Impact Software Engineer now in place
  - Available for consultancy
- Supports SGC tools + Prof. Charlotte Deane's Antibody modelling platform
- Please email brian.marsden@sgc.ox.ac.uk for more info

ACKNOWLEDGEMENTS

SGC RI Group
Brian Marsden
David Damerell
Anthony Bradley

SGC PX Group & Diamond
Frank von Delft
Tobias Krojer
Jia Ng

SGC Med Chem
Paul Brennan
Oakley Cox

OPIG Group
Charlotte Deane
Jean-Paul Ebejer
Saulo d'Oliveira
Nicholas Pearce

GSK
Ian Wall
Darren Green
Jammeed Hussain
Stephen Pickett

MolSoft
Ruben Abagyan
Eugene Raush
Max Totrov

FUNDING PARTNERS
The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA), Janssen, Merck & Co., Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome Trust