



[www.thesgc.org](http://www.thesgc.org)

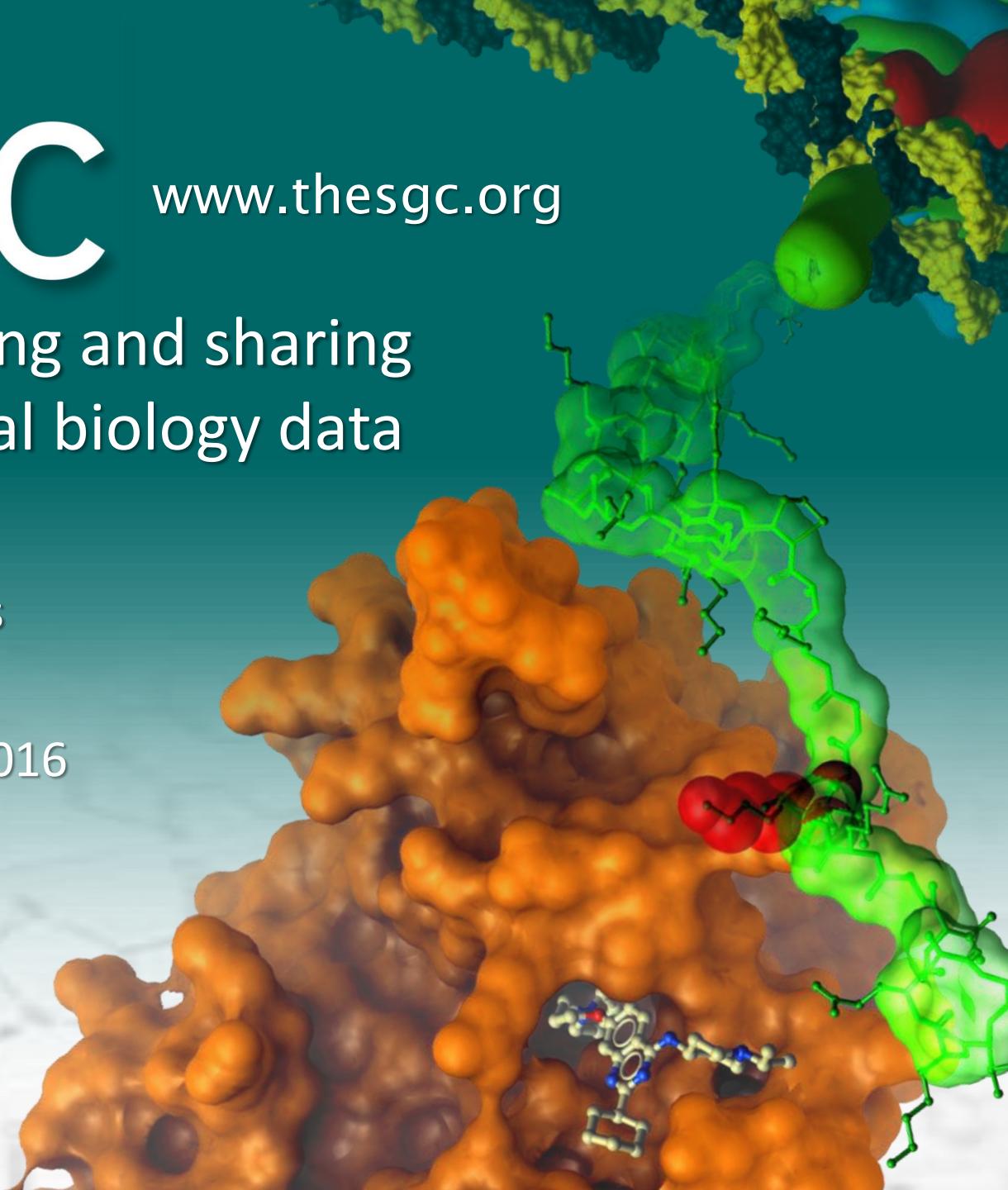
Aggregating, mining and sharing  
structural chemical biology data

Brian Marsden

PI, Research Informatics

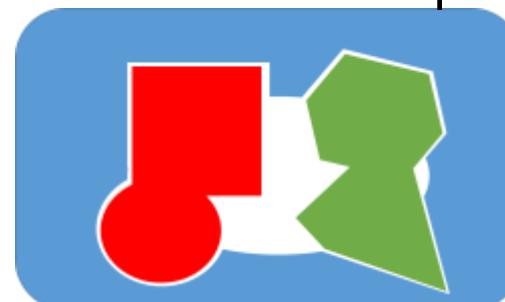
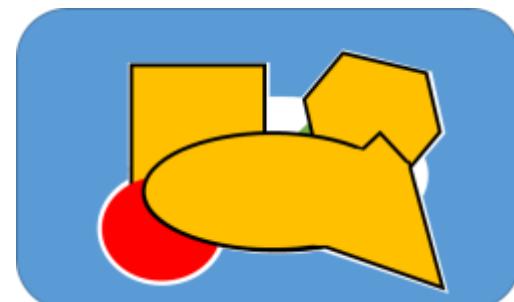
SGC & Kennedy Institute  
University of Oxford

MolSoft UGM, March 2016



## Most sensitive detection technique: *Soaking and X-ray crystallography*

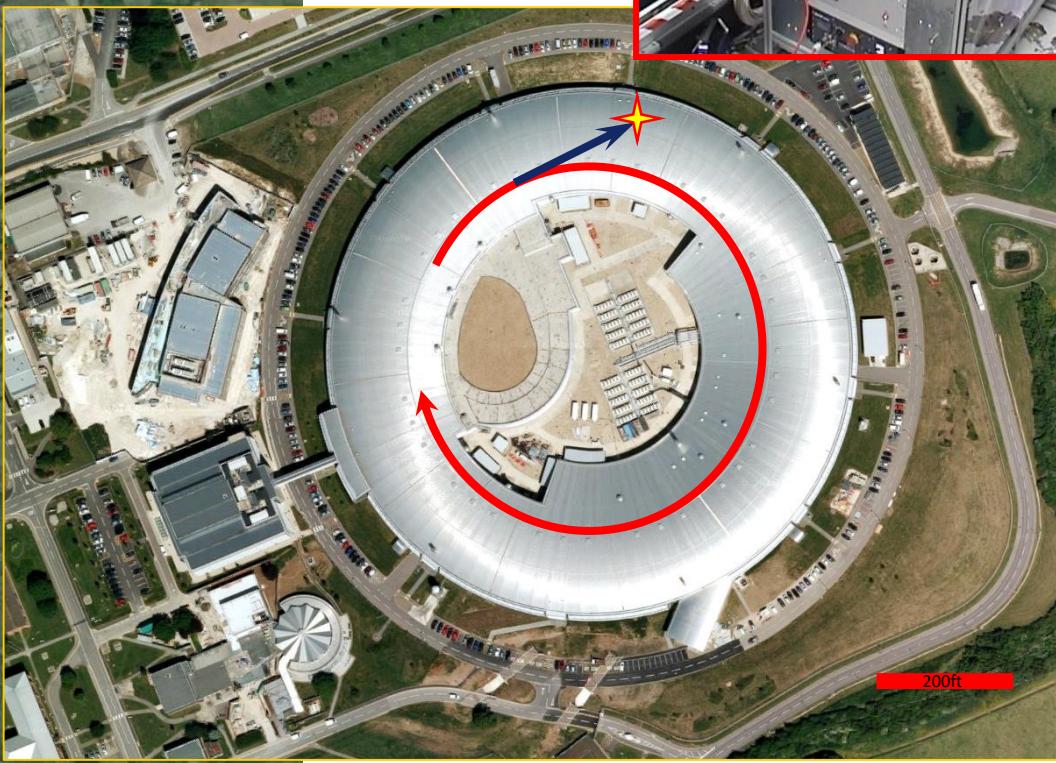
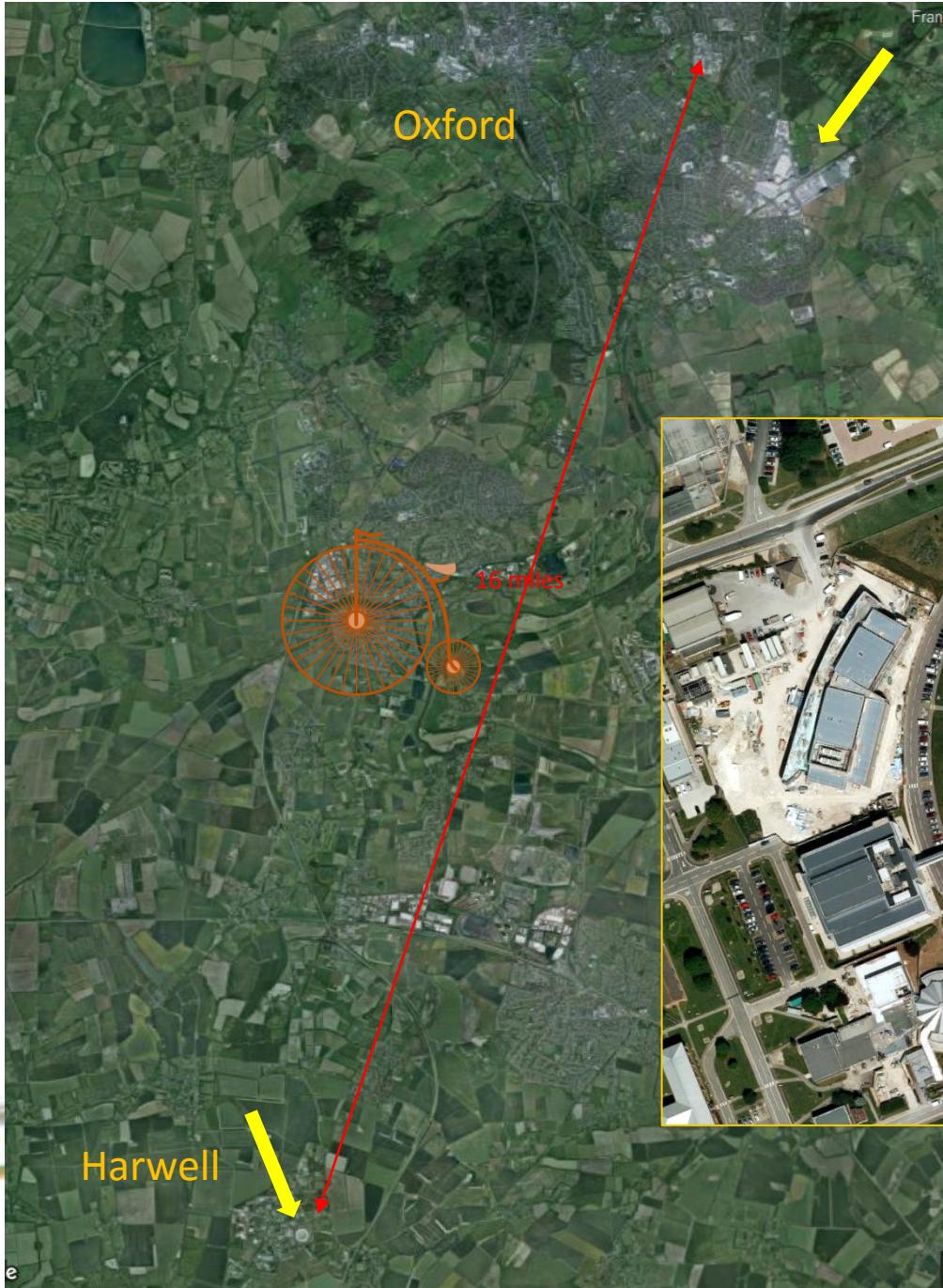
Computational strategies + Well-selected compounds



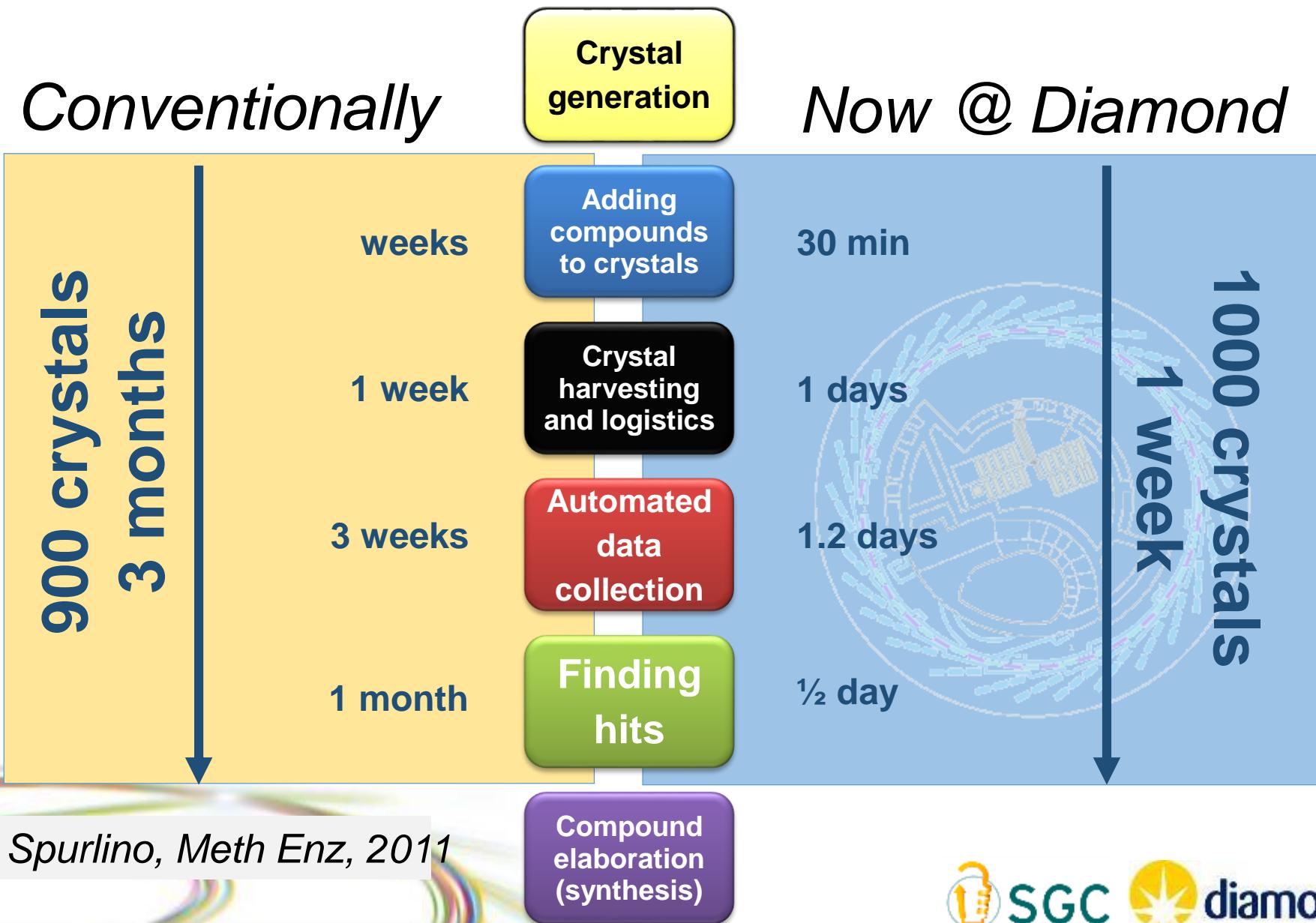
## Rapid formulaic synthetic follow-up



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



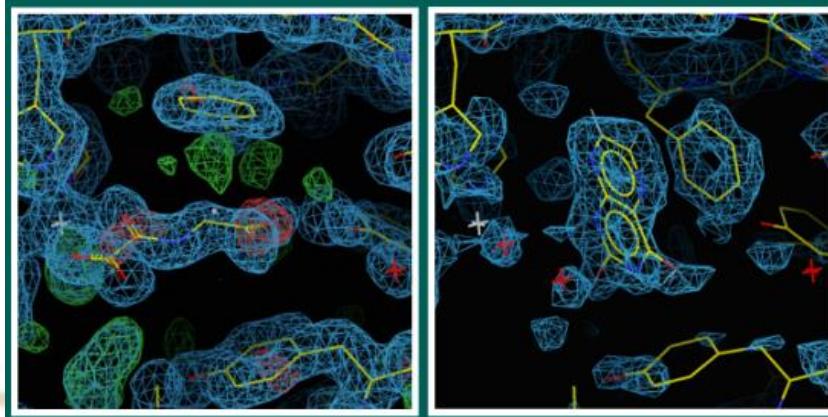
# XChem: New Benchmark



# 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

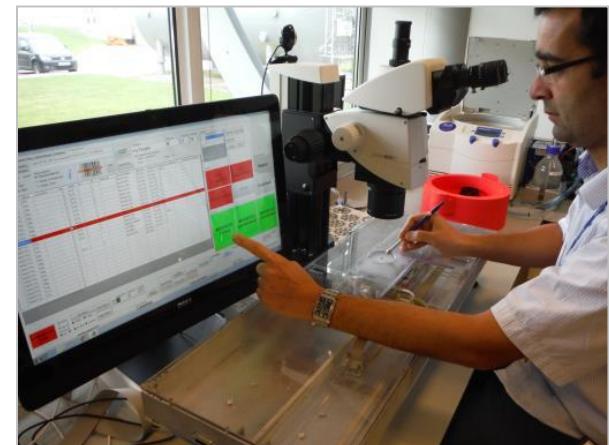


Standard maps

Corrected maps

New paradigm in crystallography

Shifter: robot-assisted  
crystal harvesting



Up to 200 crystals / per hour  
Fully recorded experiments

IND

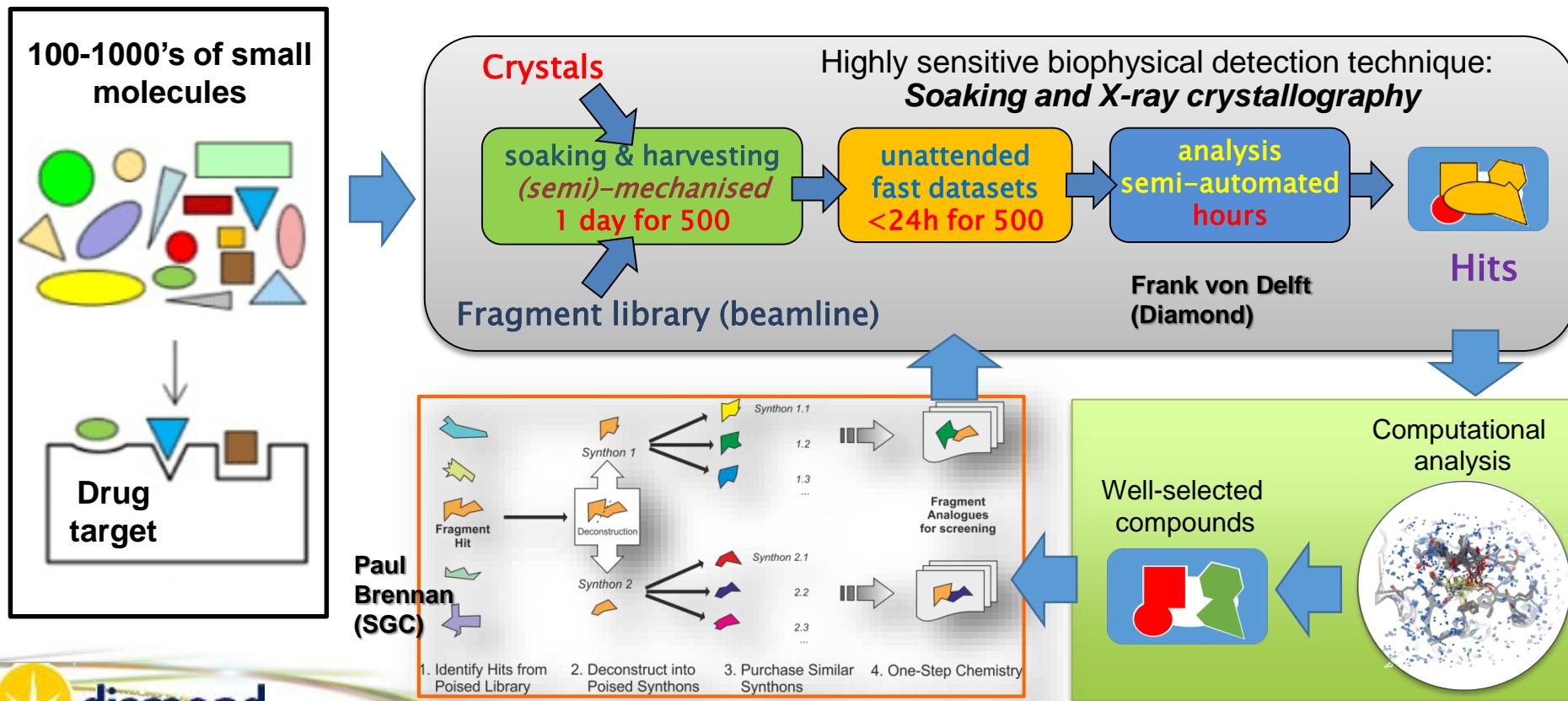
# XCHEM – X-RAY GUIDED CHEMICAL SYNTHESIS



**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?

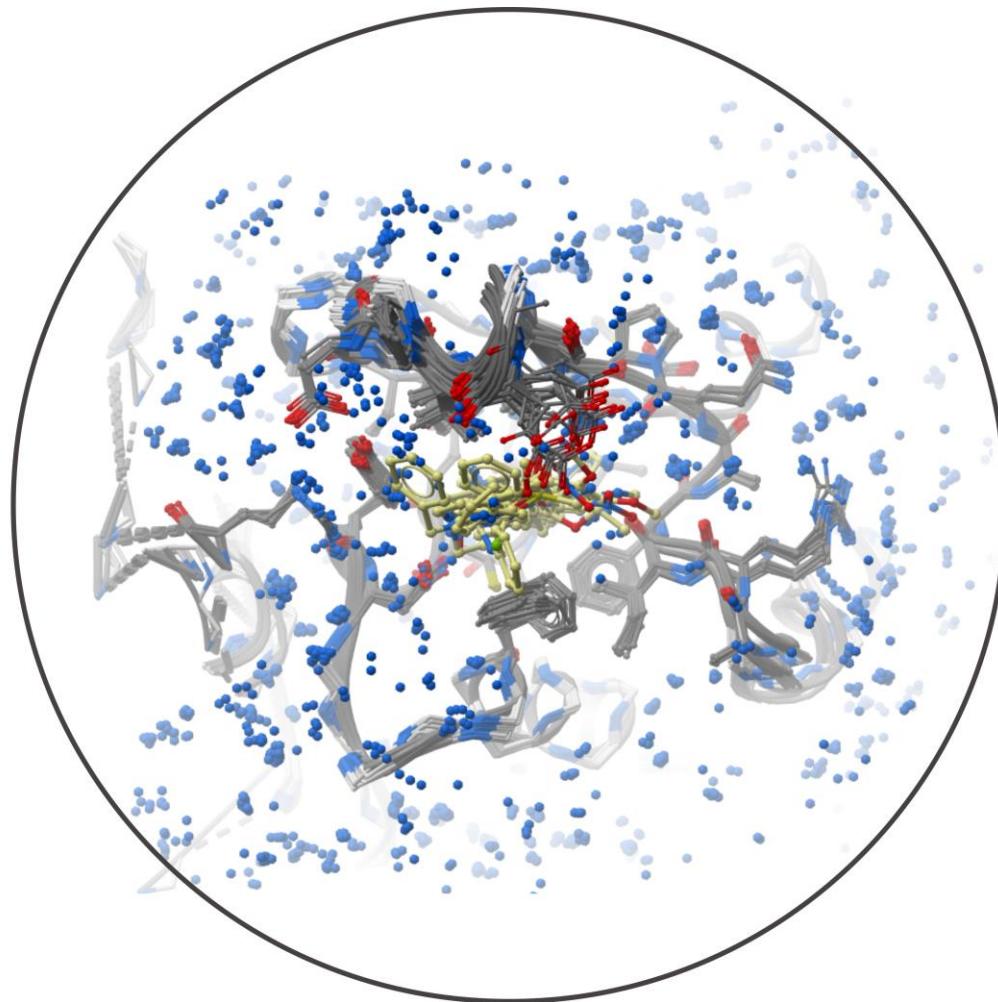
Experimental

WONKA

OOMMPPAA

LLOOMMPPAA

# EXPERIMENTAL DATA



BIOLOGY IS LARGELY SOLVED.  
DNA IS THE SOURCE CODE  
FOR OUR BODIES. NOW THAT  
GENE SEQUENCING IS EASY,  
WE JUST HAVE TO READ IT.

IT'S NOT JUST "SOURCE  
CODE". THERE'S A TON  
OF FEEDBACK AND  
EXTERNAL PROCESSING.



BUT EVEN IF IT WERE, DNA IS THE  
RESULT OF THE MOST AGGRESSIVE  
OPTIMIZATION PROCESS IN THE  
UNIVERSE, RUNNING IN PARALLEL  
AT EVERY LEVEL, IN EVERY LIVING  
THING, FOR FOUR BILLION YEARS.

IT'S STILL JUST CODE.



OK, TRY OPENING GOOGLE.COM  
AND CLICKING "VIEW SOURCE."

OK, I... OH MY GOD.

THAT'S JUST A FEW YEARS OF  
OPTIMIZATION BY GOOGLE DEV'S.  
DNA IS THOUSANDS OF TIMES  
LONGER AND WAY, WAY WORSE.

WOW, BIOLOGY  
IS IMPOSSIBLE.



It's Biology.

Deal with it!

A screenshot of a web browser window. The address bar shows "view-source:https://www.google.com.br/". The main content area displays the raw HTML source code of the Google homepage. The code is extremely long and complex, consisting of numerous lines of JavaScript, CSS, and HTML. It includes various functions like `getElementsByClassName`, `addEventListener`, and `createEvent`, as well as large blocks of CSS declarations and HTML structure.

```
<!doctype html><html itemscope="" itemtype="http://schema.org/WebPage" lang="en-BR"><head><meta content="/images/branding/g...>
  google.j.b={!location.hash&!!location.hash.match('!#')}({ifp=(!tbs=rimg){tbs=simg}{tbs=shb}})
  !(google.j.qbp=1);(function(){google.c||(c:={aitrue:d:false,i:false,m:true,n:false});google.sm="webhp";(function(){(function(a,b,c){google.timers[a]||google.startTick(a);c||google.time()};b instanceof Array){for(var d=0;d<b.length;d++)google.c.u=function(a){var b=google.timers.load.m;if(b[a])b[a]=1;for(a in b){if(b[a])return;google.csReport()}else go...>
  function(a){(google.c.c.a$=google.aft$).google.aft$=a;google.tick("aft",a.id);a.src||a.name);google.startTick("load");google.c.b("pr");go...>
  h||b.ctrlKey||b.shiftKey||b.altKey||b.metaKey||b.keyCode!=32&f||((b.m.tagName||b.m.getAttributeNode("tabindex"),h=null||B(b),g=g.type||g.tagName).toUpperCase(),(g32=(b.which||b.keyCode||b.key)&&"CHECKBOX"!=g)||{g=B(b),q=q.getAttribute("rol...>
  typeof b[p]!="rcElement"||p44(c[p]=b[p])&c.type=="mouseover"||b.type?"mousenter":"mouseleave";c.target=c.target=o.set...>
  l);if(a.vla.v){e:else{var vif=p.k.document&&p.createEvent&&p.createEventObject}try{v=p.createEventObject(b);catch(la)...>
  .gbii[background-image:url(/lh6.googleusercontent.com/-akCqc72v12s/AAAAAAAAl/AAARAAAAAAA/jy-eYAtXfjc/s32-c-mo/photo.jpg)
</style><style data-jiis="cc" id="gstyle">html,body{height:100%;margin:0}#viewport{min-height:100%;position:relative;width:100%}</style>
<div data-bbox="0 0 1000 1000" data-via="www.gstatic.com" data-vp="og.og2.en_US.Pfv5y6jPcuK.O" data-w="com.br" data-en="1" data-d="3,2," data-t="...>
  _CONFIG=[[0,"www.gstatic.com","og.og2.en_US.Pfv5y6jPcuK.O"],[1,"com.br","en","1",0,[3,2,"...>
  try{
    var fa,ha,_aa=_.aa||{};_m=this;_n=function(a){return void 0==a};_q=function(a,c){for(var d=a.split("."),e=c||_.m,f=d.length,...>
    _ca=function(a){var c=typeof a;if("object"==c){if(a){if(a instanceof Array) return "array";if(a instanceof Object) return "obj...>
    else if("function"=="c")undefined==typeOf a.call{return "object"};return c};_da=function(a){return"array"==_.ca(a)};_t=fun...>
    ha=function(a,c,d){if(!a)throw Error();if(!arguments.length){var e=Array.prototype.slice.call(arguments,2);return function(...>
    _x=function(a,c){var d=a.split("."),e=_m[d[0]]|e||!e.execScript||e.execScript("var "+d[0]);for(var f;d.length&#46;(f=d.length-1);f--){var ia=function(a,c,d){this.A=a>this.o||!this.b||this.c||this.w||d;ia.prototype.Ra=function(a){if(this.o)throw Error("b"+this.b);var ja=function(a){_.x.call(this);this.wa=this.b=[].push(this.o||{});ja.prototype.Ra=function(a){var c=_.v_(functi...>
    ja.prototype.Ra=function(){for(var a=this.b.length,c=this.b,d=[],e=0;e<a;e++){var f=c[e],b,gra:{g=this.wf||var he=f.split("...>
    _ka=function(a){if(Error.captureStackTrace)Error.captureStackTrace(this,_ka);else var c=Error().stack;c&#46;(this.stack=c)a...>
    _na=function(a,c){for(var d=0,e=0,_la=(String(a)).split("."),f=0,_la=(String(c)).split("."),g=Math.max(e.length,f.length),_oa=Array.prototype.indexOf.call(a,c,d){return Array.prototype.indexOf.call(a,c,d);_da=function(a,c,d){d=null==d?0:d>f.length?f:f;_sa=Array.prototype.filter.call(a,c,d);_da=function(a,c,d){function(a,c,d){d=null==d?0:d>f.length?f:f;_sa=Array.prototype.reduce.call(a,c,d,e&gt;=c[_v_.v()]);return Array.prototype.reduce.call(a,c,d);_fa=function(a,c,d){var Bar=_ua=function(){this.b=[].push(this.o||{});_ba(_ua);_wa=function(a,c){a.T=function(){return _va(_ua,T(),c)};_xa=func...>
    var Da_=Ca="bbb bbr bbs has prm sngw so_.split(" ");Da=new ja(_m);_za("api",Da);
    for(var Ea="addExtraLink addLink addmc asmc close cp.c cp.l cp.me cp.ml cp.rc cp.rel ela elc elh gpcg qpcg lgc lPFW ldb mls...>
    _Ia=\uffff/.test("\uffff")?/\\""\x00-\x1f\x7f-\uffff/g:/\\""\x00-\x1f\x7f-\x7f/g;_A=function(){},_B=function(a,c,d,...>
```

Hit  
detection

Which  
hit to  
develop?

What  
should I  
make?

How do I  
fill the  
gaps?

Experimental

WONKA

OOMMPPAA

LLOOMMPPAA

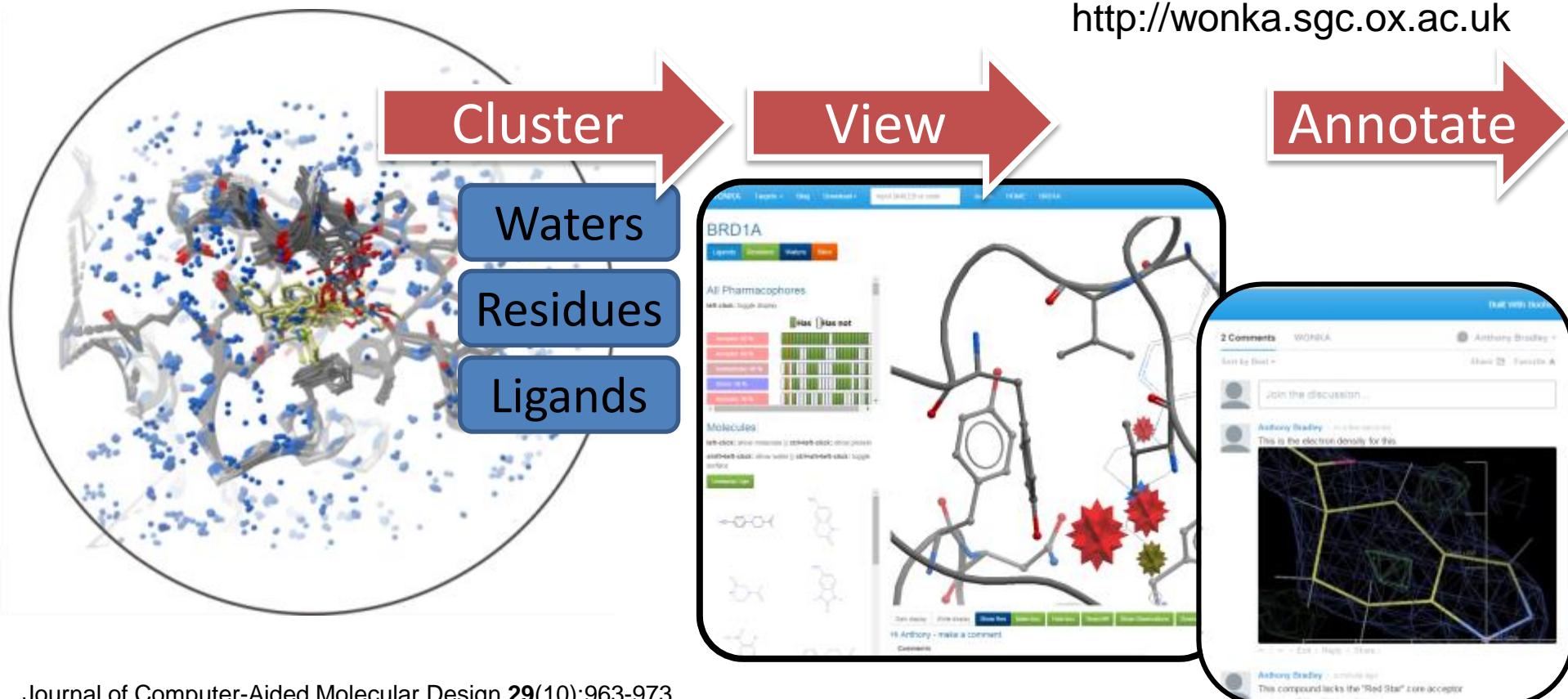
Make sense of this mess

- Automatically, objectively identify features
  - Display, navigate
  - Annotate, share, discussion thread
- *Natural end-point for screening experiment*



Anthony Bradley

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



# WONKA – KEY BINDING SITES

## Sites

left click: toggle display

2	ANALYSE
2	ANALYSE
2	ANALYSE
1	ANALYSE
1	ANALYSE

## LIGANDS

Has  Has not

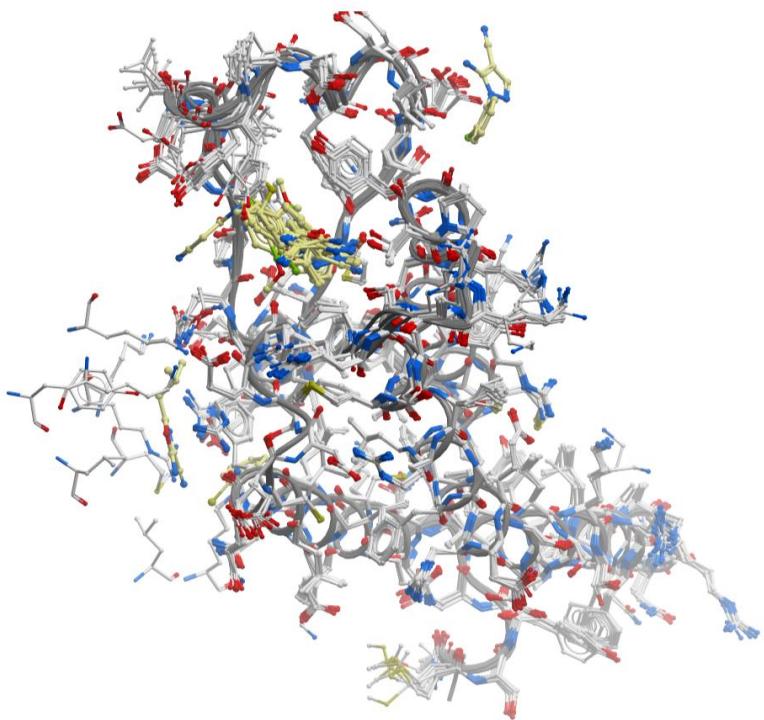
Has	Has				
		Has			
			Has		
				Has	
					Has

Site by site  
analysis

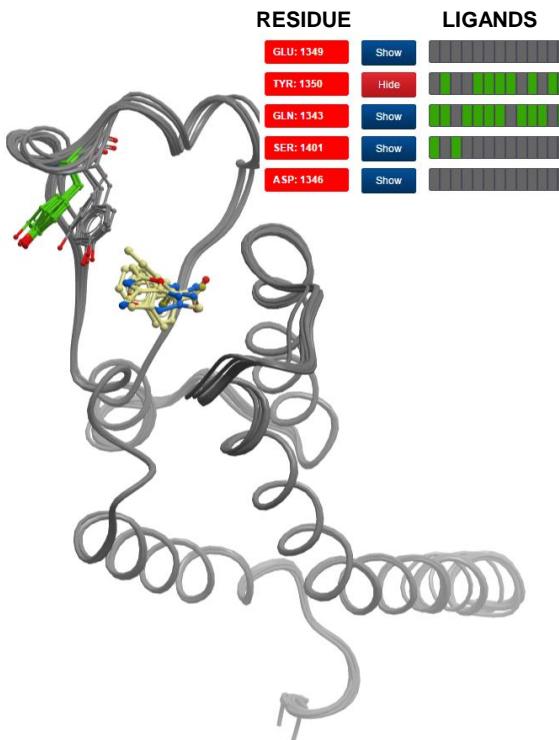


# WONKA – KEY RESIDUES

Raw data



WONKA



RESIDUE

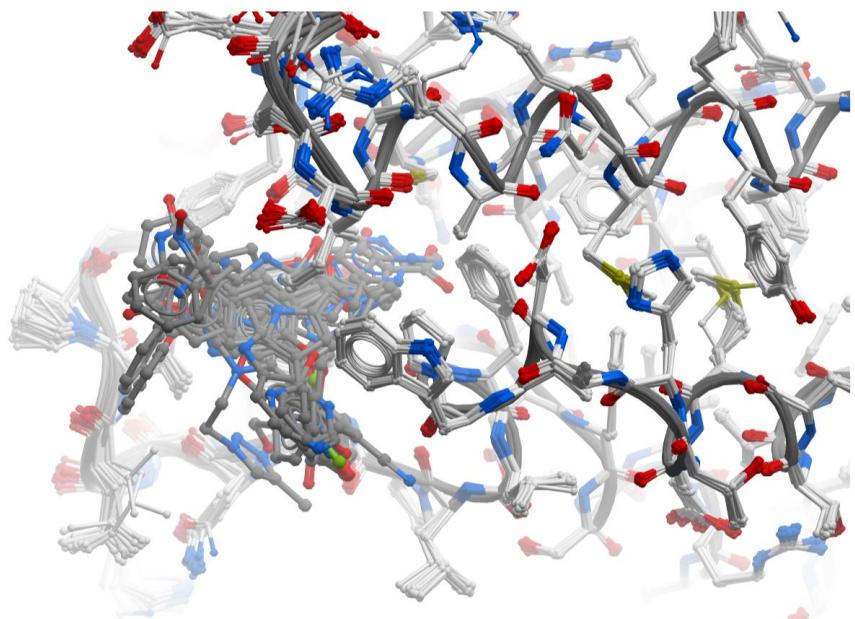
GLU: 1349  
TYR: 1350  
GLN: 1343  
SER: 1401  
ASP: 1346

LIGANDS

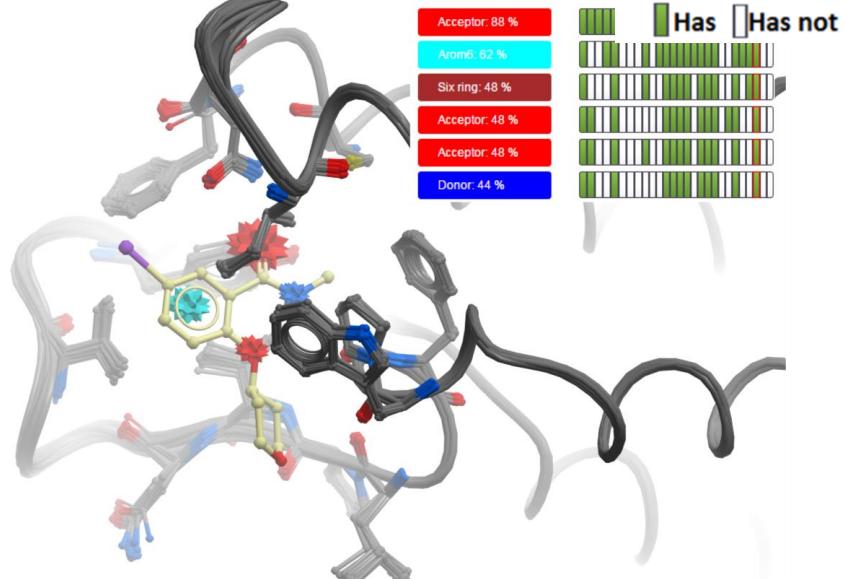
Show  
Hide  
Show  
Show  
Show

# WONKA – LIGAND FEATURES

Raw data



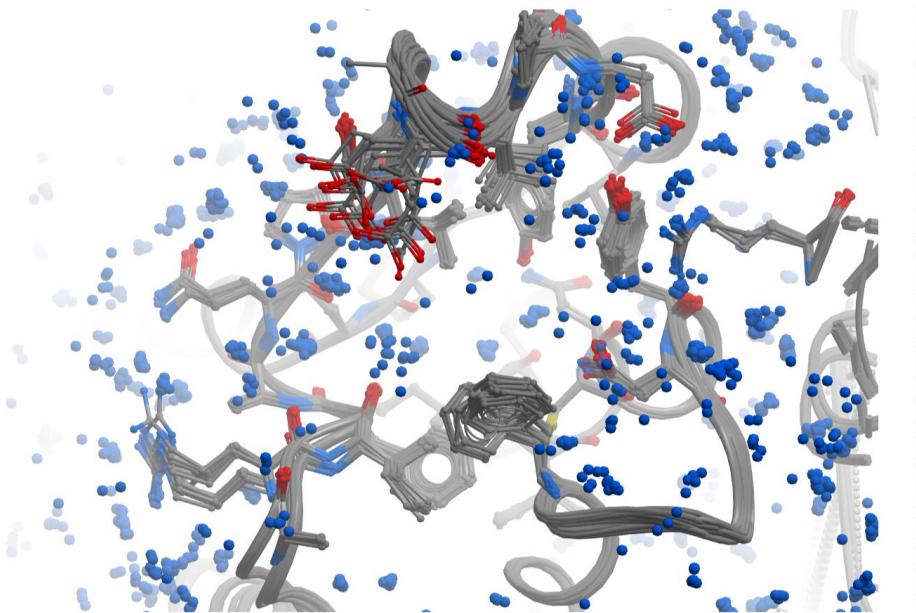
WONKA



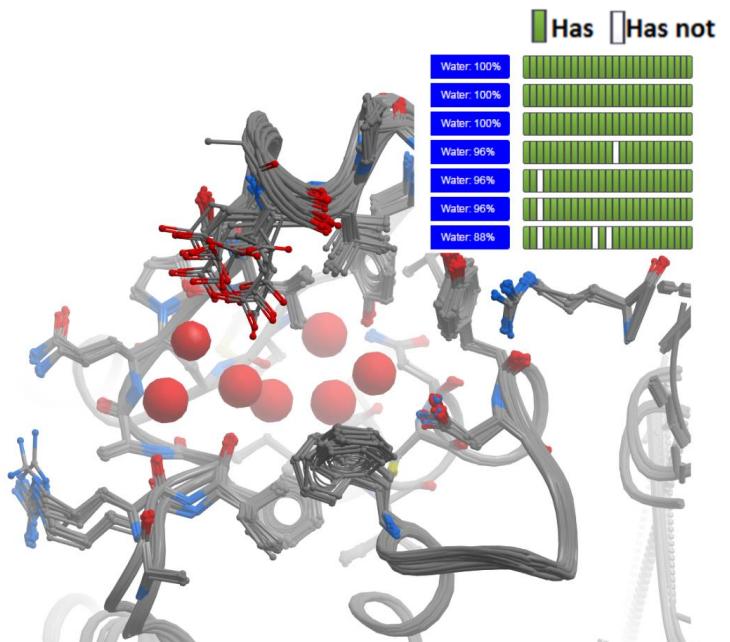
# WONKA – CONSERVED WATER MOLECULES



Raw data



WONKA



Has Has not

Water: 100%	Has
Water: 100%	Has not
Water: 100%	Has
Water: 96%	Has not
Water: 96%	Has
Water: 96%	Has not
Water: 88%	Has

# WONKA – SHARING OBSERVATIONS

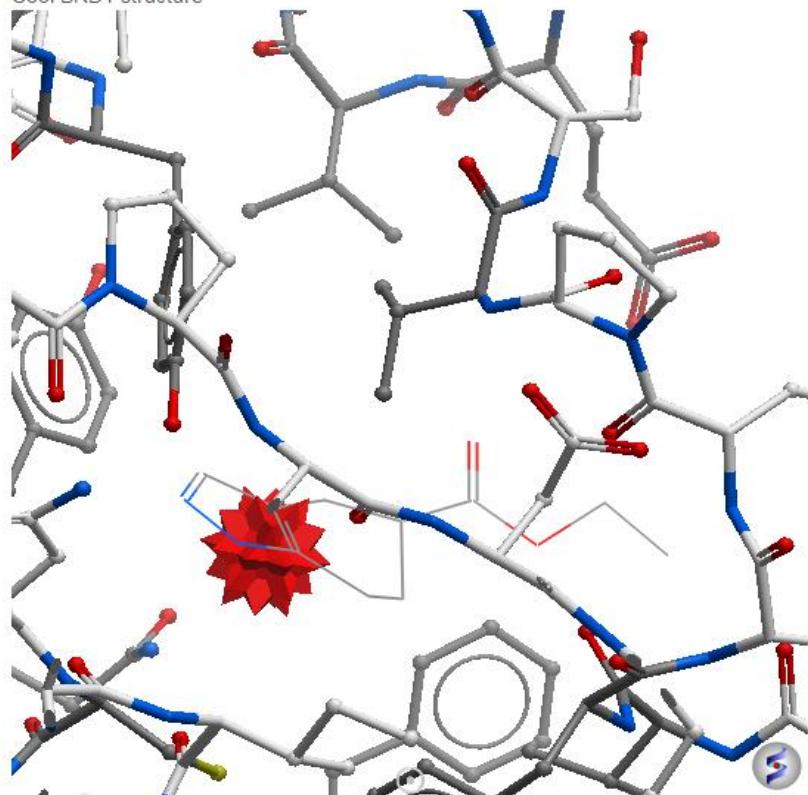


WONKA Targets ▾ Blog Download ▾ HOME BRD1A

Built With Bootstrap

Anthony:

Cool BRD1 structure



2 Comments

WONKA

Anthony Bradley ▾

Sort by Best ▾

Share Favorite

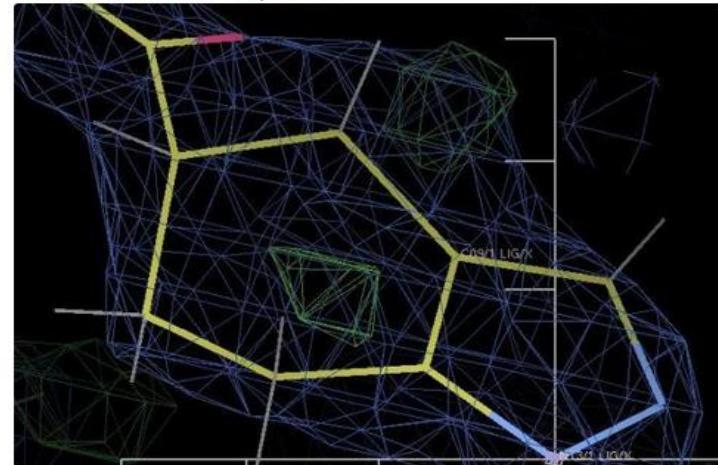


Join the discussion...



Anthony Bradley · in a few seconds

This is the electron density for this.



^ | v • Edit • Reply • Share >



Anthony Bradley · a minute ago

This compound lacks the "Red Star" core acceptor

^ | v • Edit • Reply • Share >

DISQUS

Hit  
detection

Which hit  
to  
develop?

What  
should I  
make?

How do I  
fill the  
gaps?

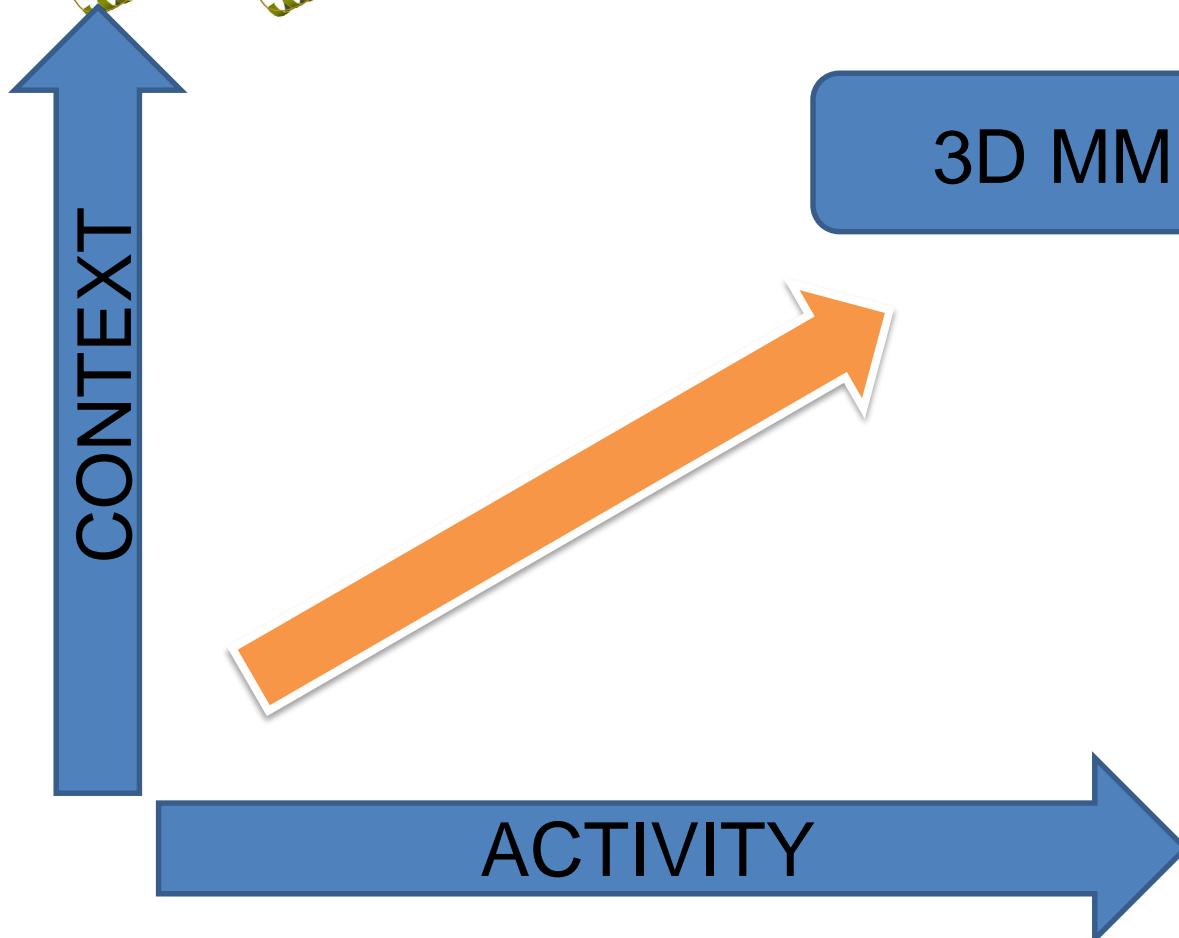
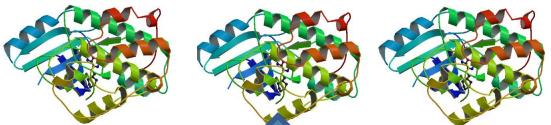
Experimental

WONKA

OOMMPPAA

LLOOMMPPAA

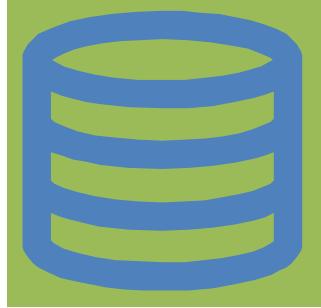
# OOMMPPAA – 3D MMP ANALYSIS



Structure	MW	% Inh. @ 1 μM	% Inh. @ 5 μM	IC <sub>50</sub> 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

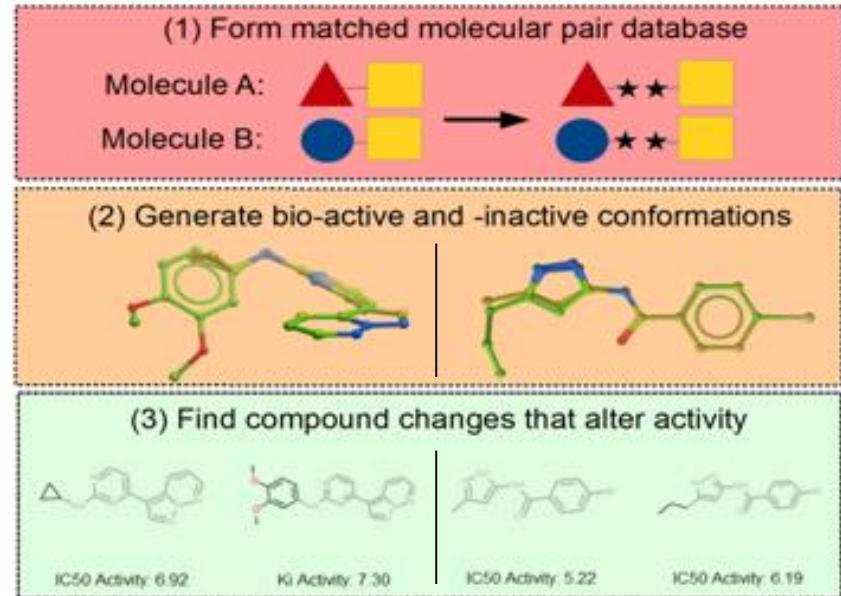
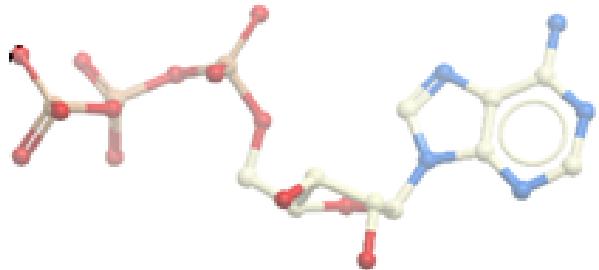
a)

## Activity data



b)

## Complexed ligand structures



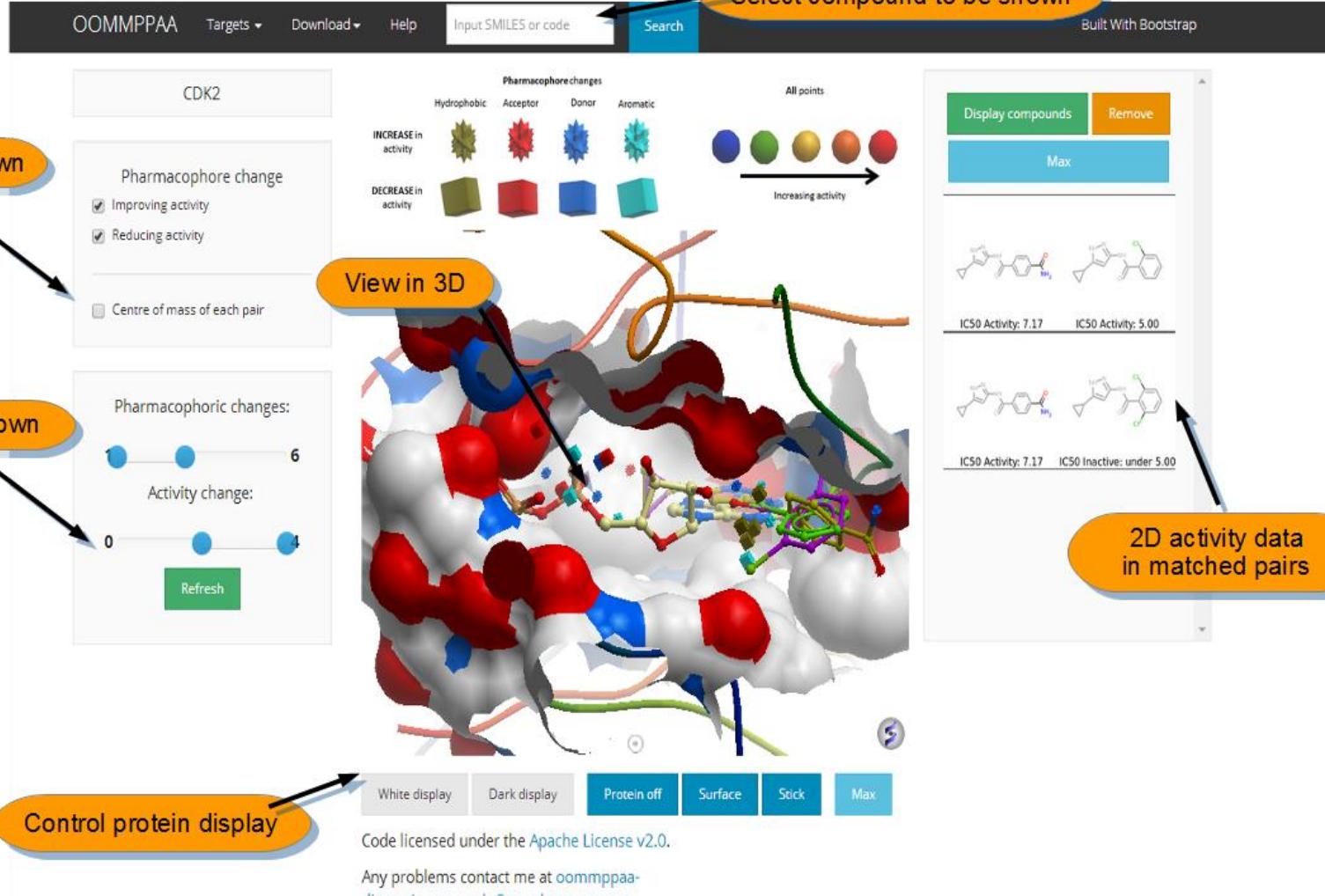
# OOMMPPAA – INTERACTIVE VISUALISATION

Control data shown → **Select compound to be shown**

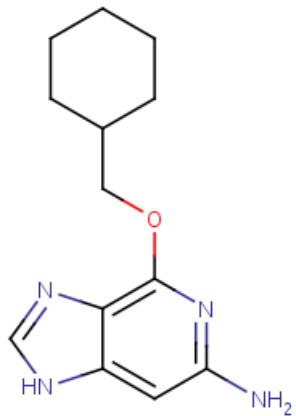
Filter data shown → **View in 3D**

Control protein display → **2D activity data in matched pairs**

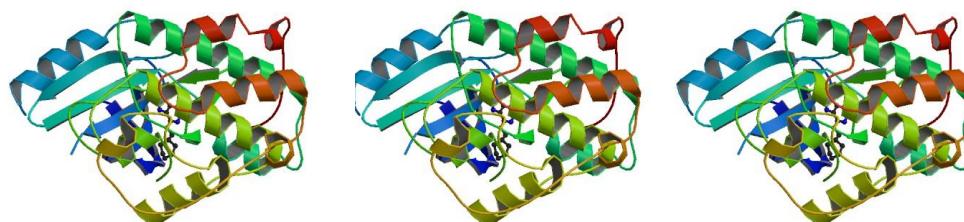
Code licensed under the Apache License v2.0.  
Any problems contact me at [oommppaa-discussion+noreply@googlegroups.com](mailto:oommppaa-discussion+noreply@googlegroups.com)



Lead – IC<sub>50</sub> 13μM



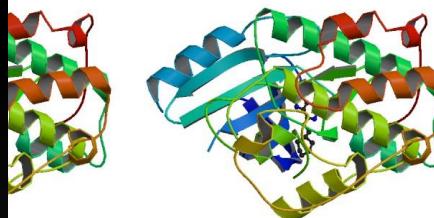
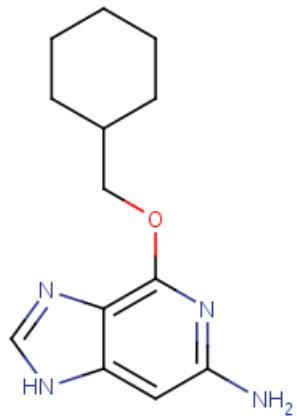
### Structural data



### Activity data



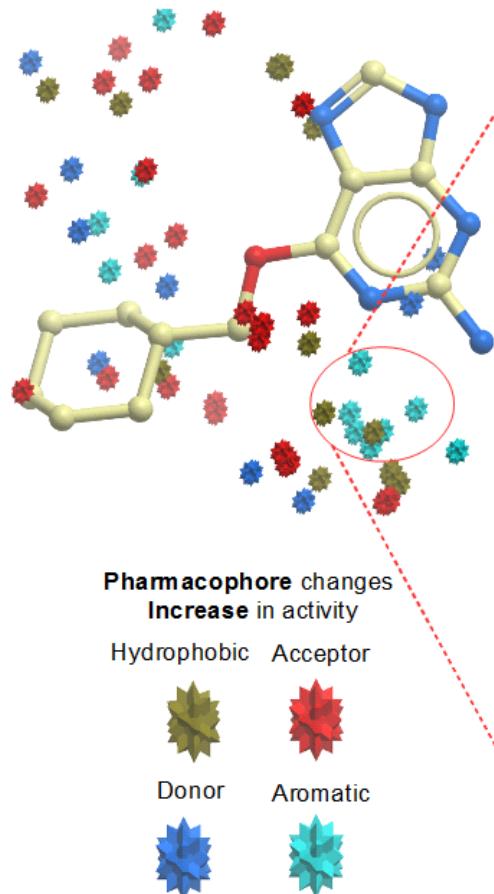
Lead – IC<sub>50</sub> 13μM



data



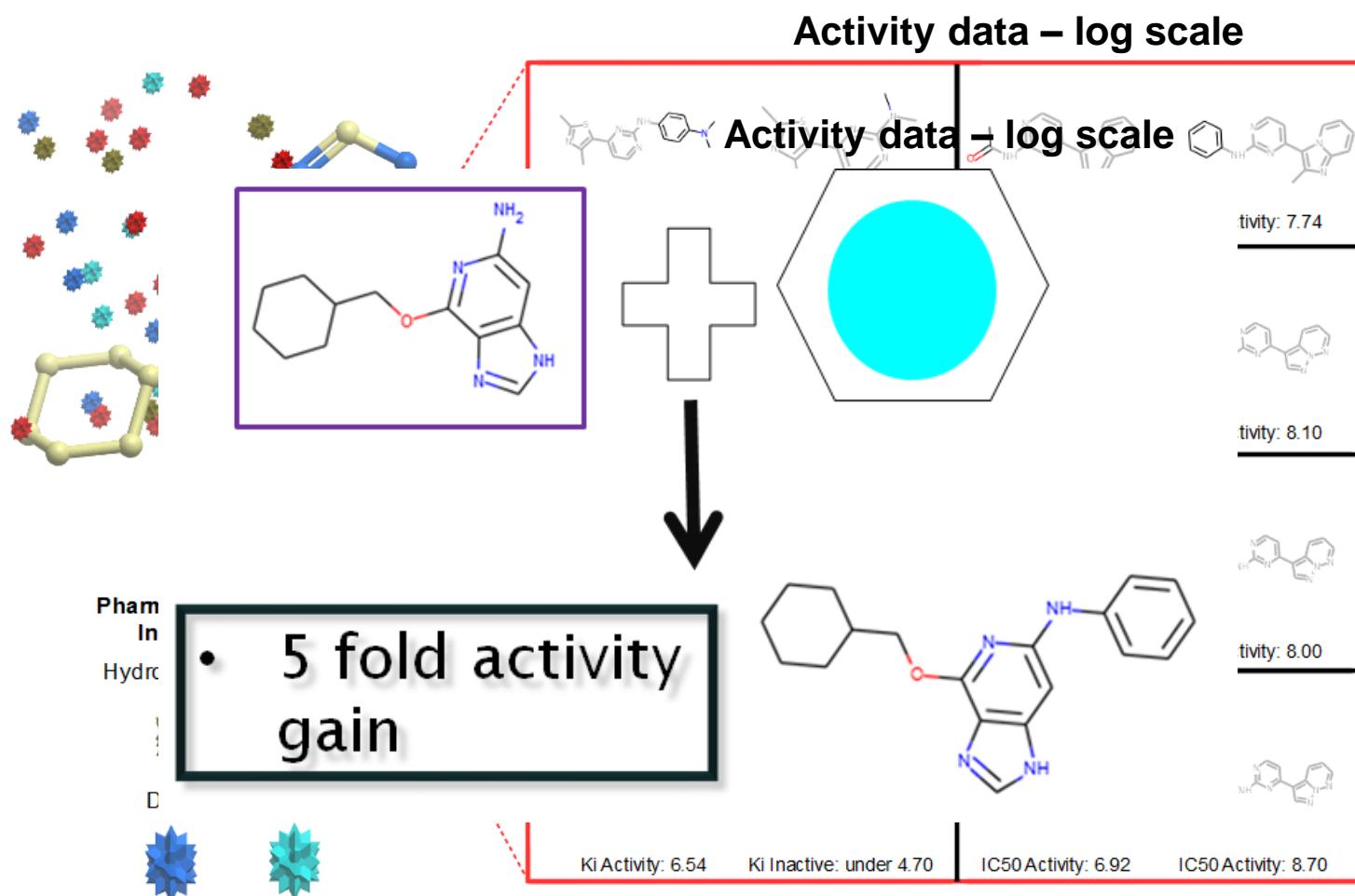
What compound do we make next?????????

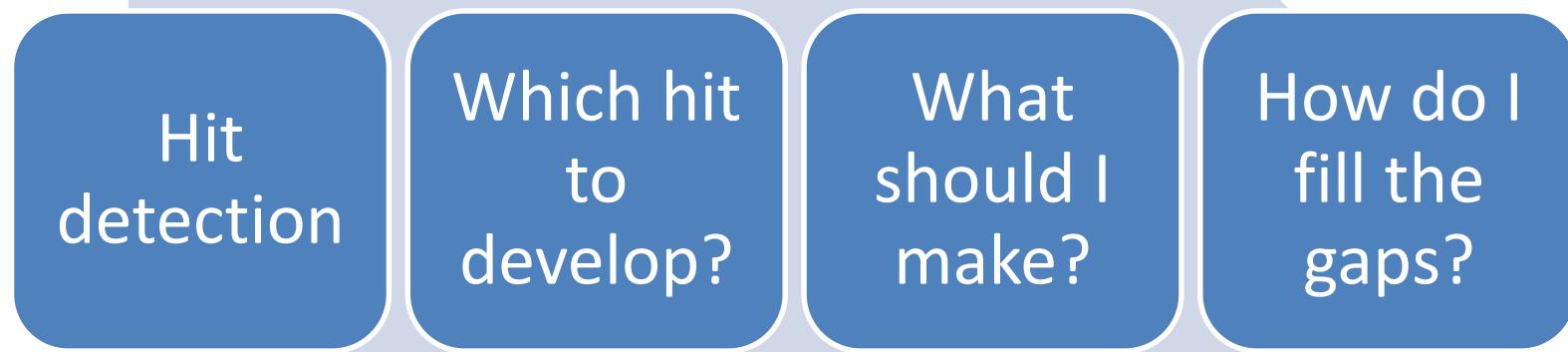


## Activity data – negative log scale

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





Experimental

WONKA

OOMMPPAA

LLOOMMPPAA

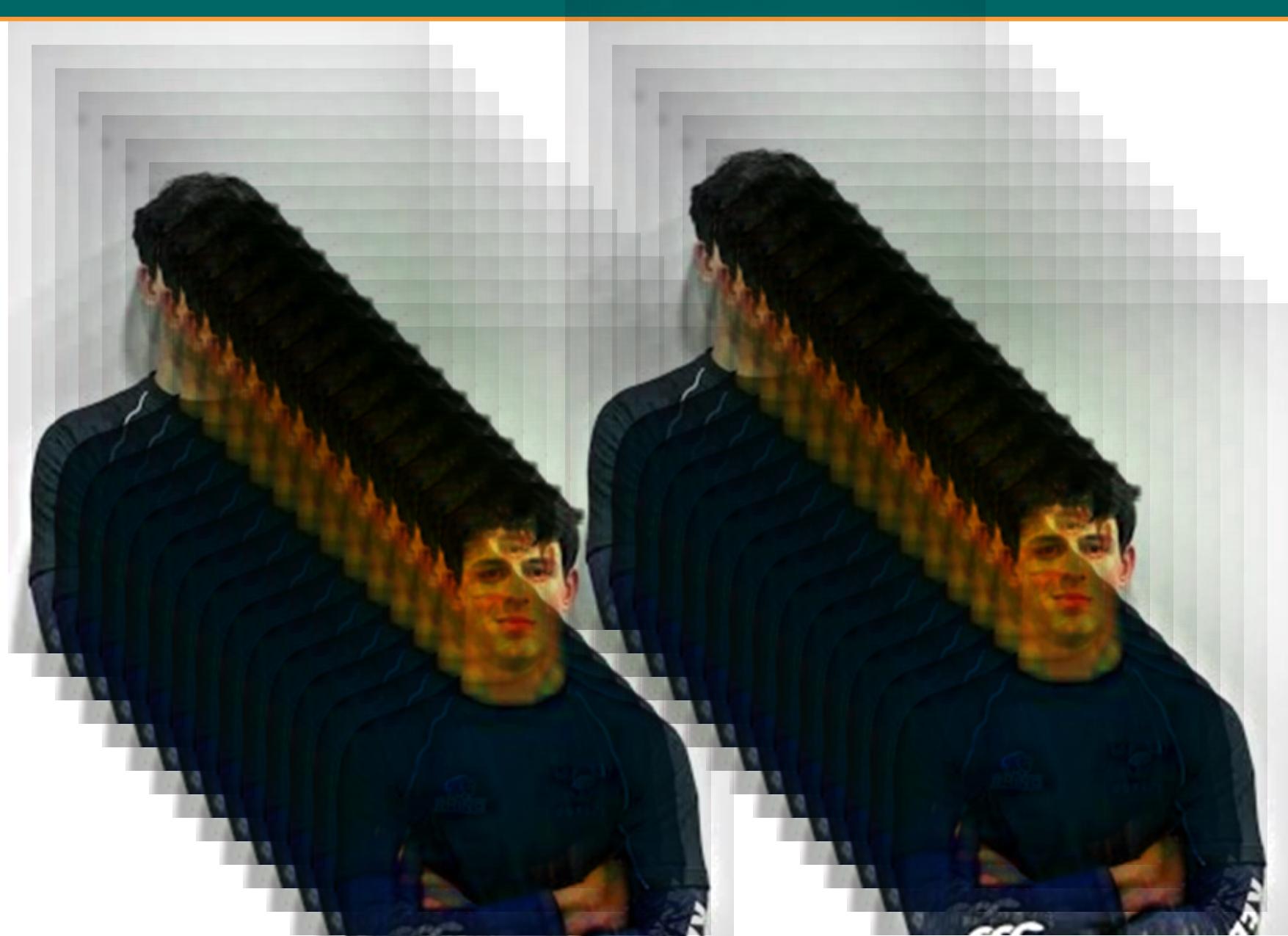
# OAKLEY SAYS – HOW SHOULD I FILL THE GAPS?



I could make 3,000  
compounds – but I only  
have time to make 30?

**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?

Virtual library  
design



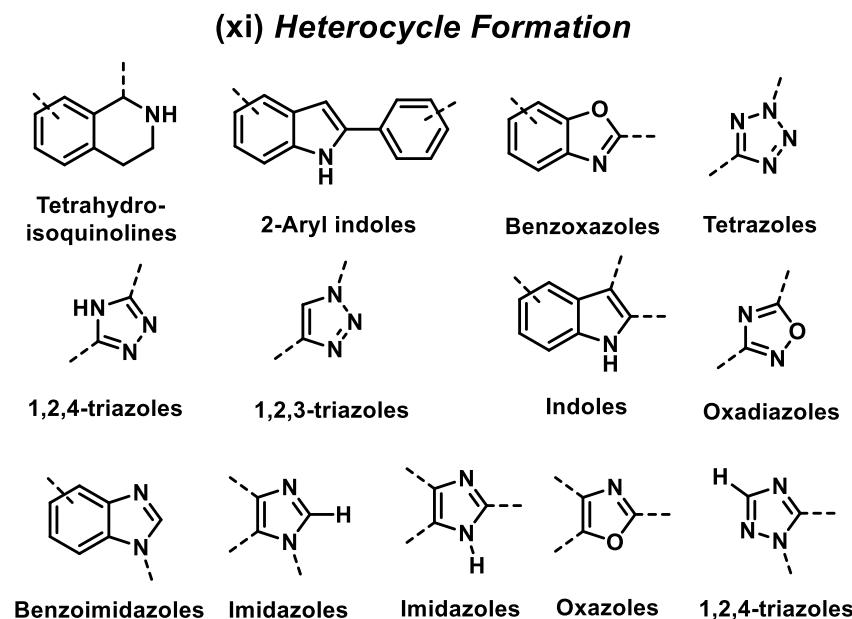
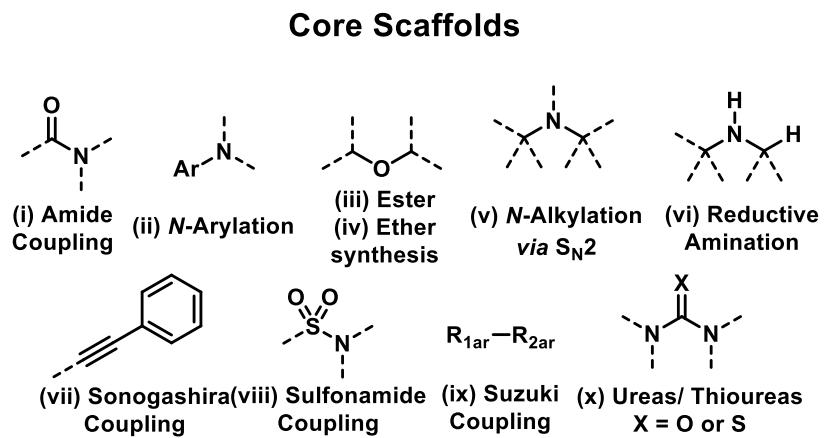
COMPUTER



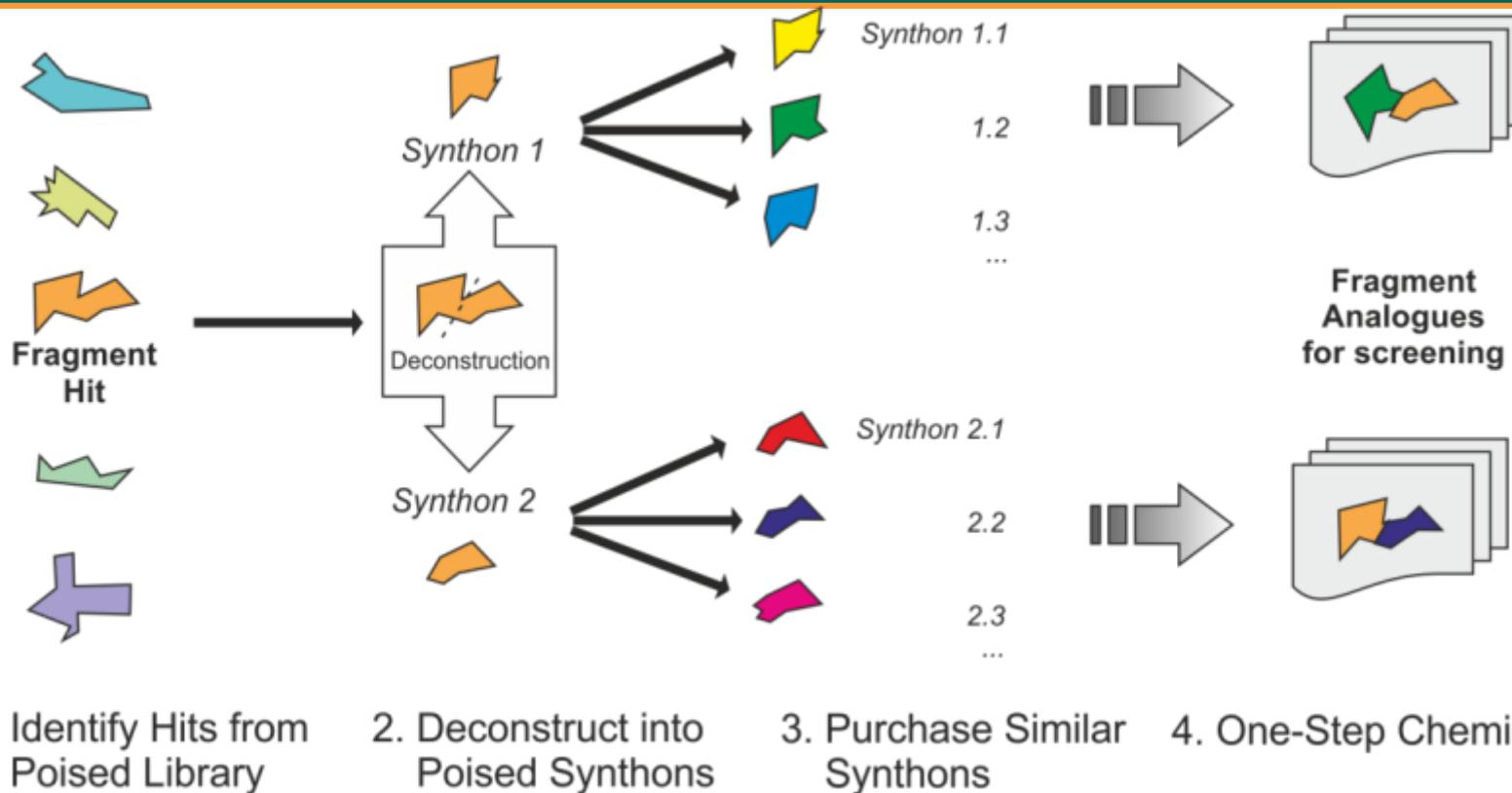
# POISED FRAGMENTS



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



# OAKLEY – USING POISED FRAGMENTS



## Poised reactions are:

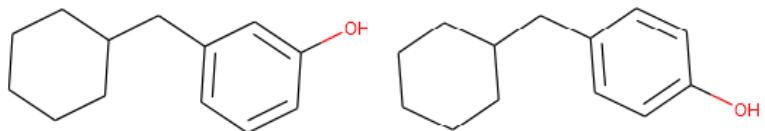
- robust and reliable;
- make drug-like products;
- be possible using commercially available starting points;
- compatible with a range of substrates.



Oakley Cox  
(SGC: Paul Brennan)

# DEFINING RELEVANT DIVERSITY

## 2D Diversity

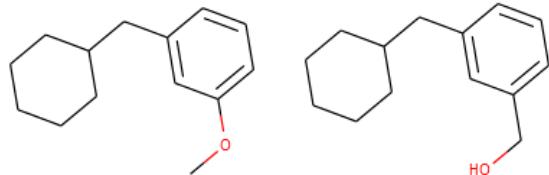


## Relevance

## Example

- Morgan fingerprint
- MACCS

## 3D Shape diversity



- Shape Tanimoto

## 3D Pharmacophore diversity

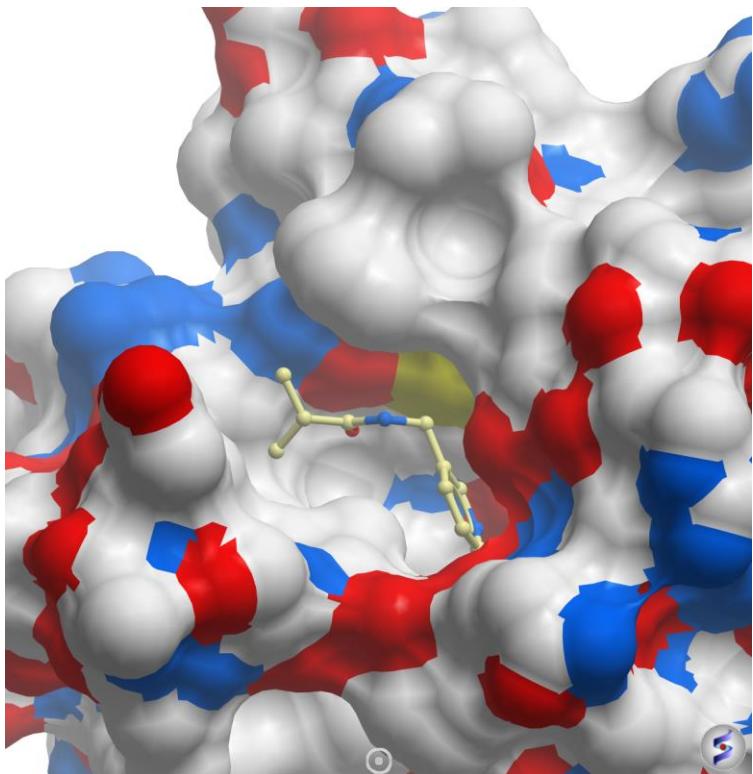
- But are the changes relevant?

- USR CAT

## Protein-ligand interaction diversity

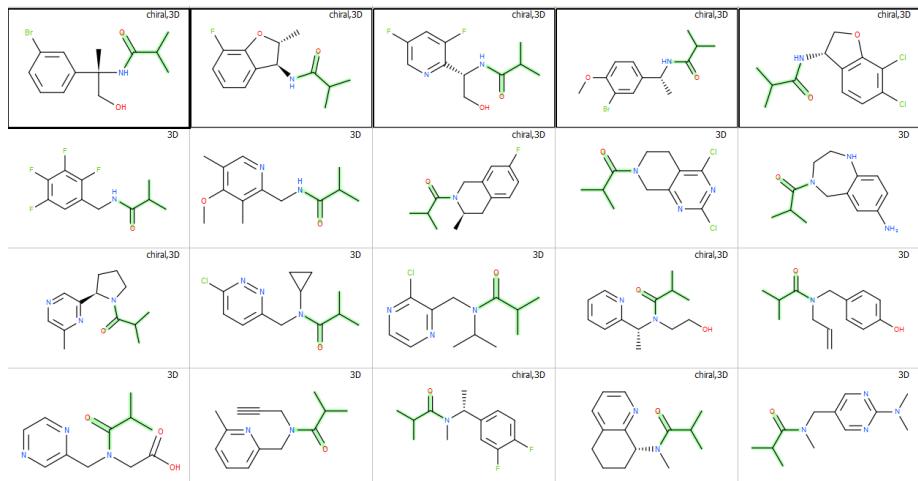
- LLOOMMPPAA

## Complexed ligand structure

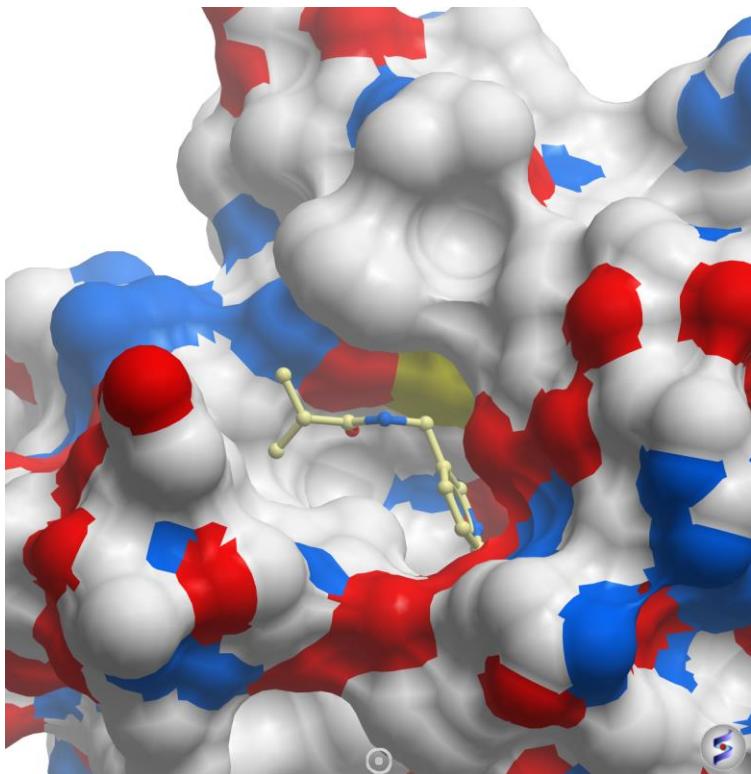


1)

Define follow ups – simple reactions (amide coupling)



## Complexed ligand structure

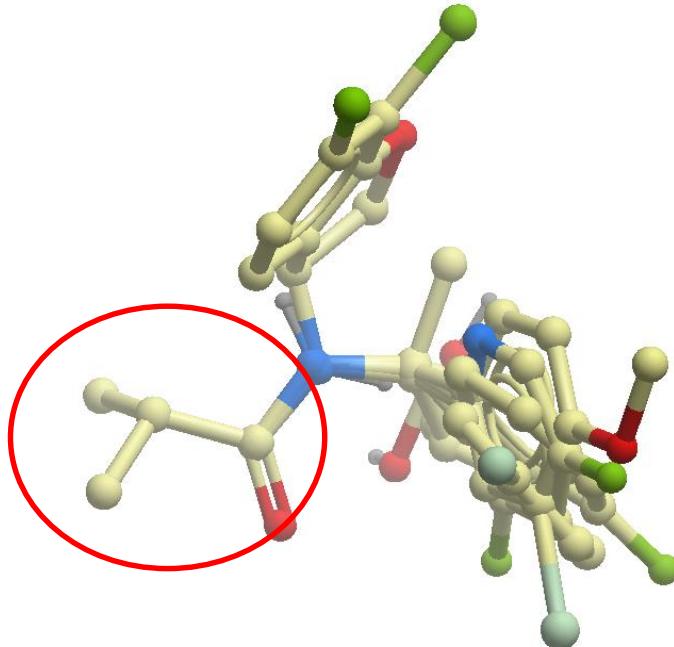


1)

**Define follow ups**

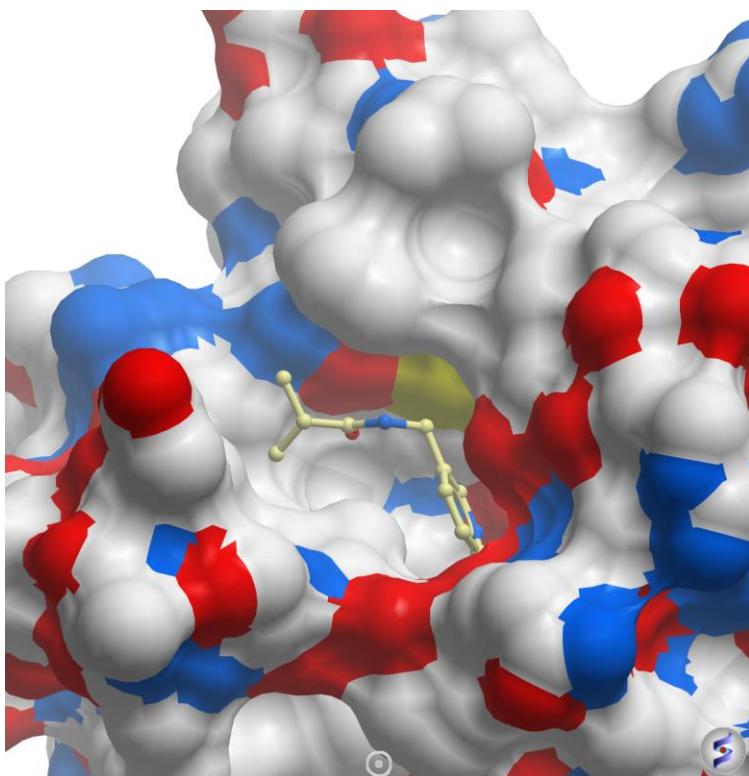
2)

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



## LLOOMMPPAA – Method

Complexed ligand structure



1)

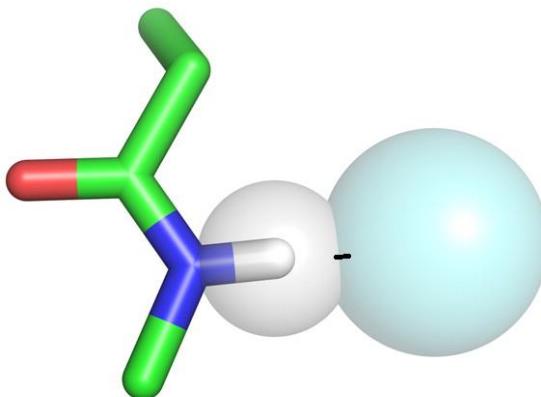
**Define follow ups**

2)

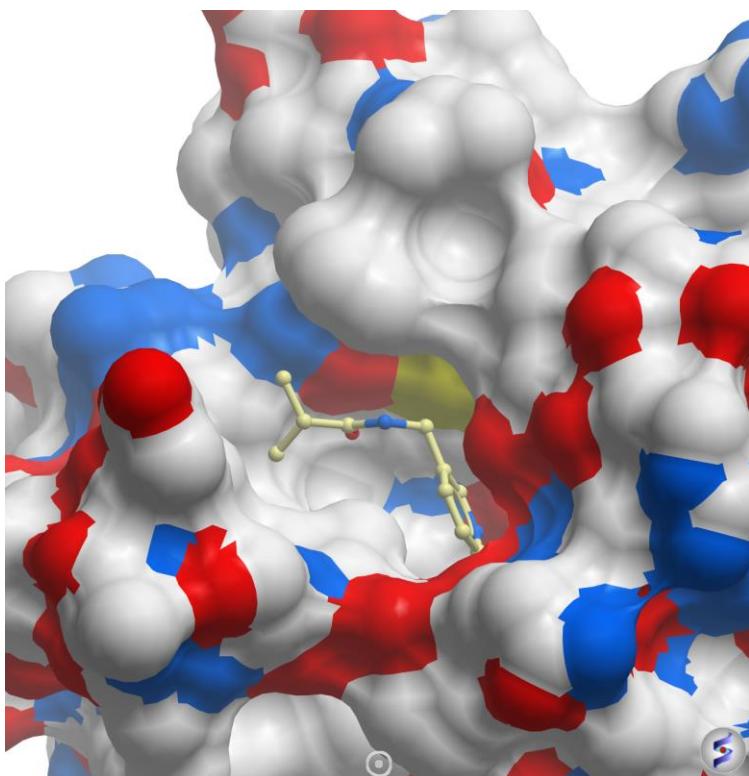
**Generate diverse conformations**

3)

**Remove VdW clashes with protein**



## Complexed ligand structure



1)

**Define follow ups**

2)

**Generate diverse conformations**

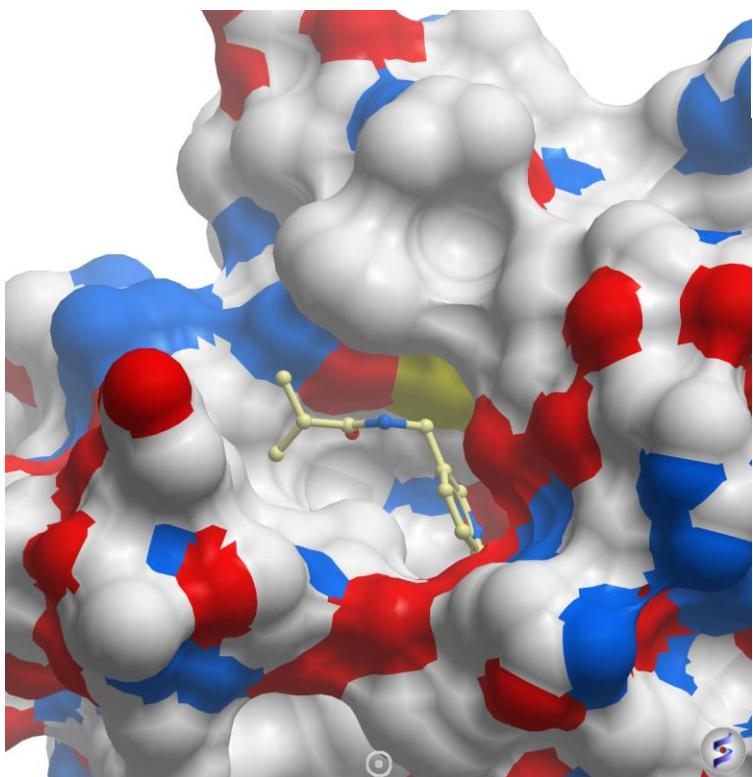
3)

**Remove VdW clashes with protein**

4)

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

## Complexed ligand structure

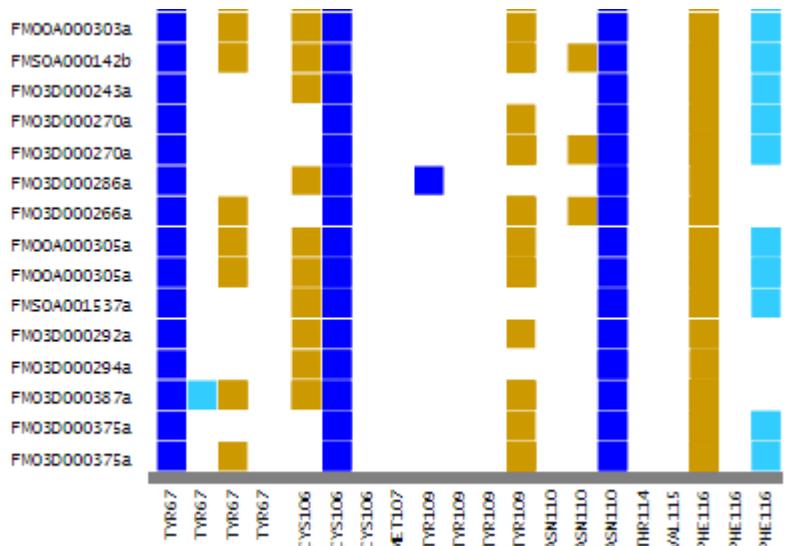


4)

## Find protein-ligand interaction fingerprints (PLIFS)

■ Aromatic ■ Hydrophobic ■ Acid - base ■ H-bond ■ weak H-bond

Potential compounds

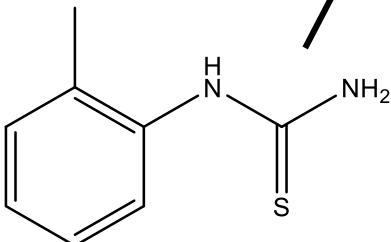


Protein receptors

# PHIP(2) - THIOUREA POISED FOLLOW-UPS

ID	IC <sub>50</sub> (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
<b>XST00942</b>	<b>768</b>	<b>0.4</b>

Ten-fold  
increase in  
potency



Original fragment  
hit

LLOOMMPPAA

ID	IC <sub>50</sub> (μM)	LE
<b>XST00942</b>	<b>768</b>	<b>0.4</b>
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

- 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

- 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](mailto: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