



UNIVERSITY OF
OXFORD



ChemBio Hub

ChemBio Hub - tools to help collaboration within the University of Oxford and beyond

Paul Barrett

Web Developer, ChemBio Hub, University of Oxford

github: [@paulbarrett79](#)

twitter: [@paulbarrett79](#)

email: paul.barrett@chembiohub.ox.ac.uk

RDKit UGM Zurich, 4th Sept 2015



UNIVERSITY OF
OXFORD



ChemBio Hub

Agenda for this talk:

- What is ChemBio Hub?
- What are we doing for scientists?
- How you can help.



About me

Web developer.

Been involved in a number of projects.

I learned python only recently.

RDKit user for approx. 1 year



UNIVERSITY OF
OXFORD



ChemBio Hub

How ChemBio Hub came about

ChemBio Hub came about from the SGC



Open source ethos

Cutting edge science

Links with pharma

www.thesgc.org

@thesgconline



How ChemBio Hub came about

- SGC model has been very successful
- Identified issues at other academic departments
 - Duplication of effort, searching linked work, enabling collaborations, reproducibility
- Even geographical locations have hindered collaborations!

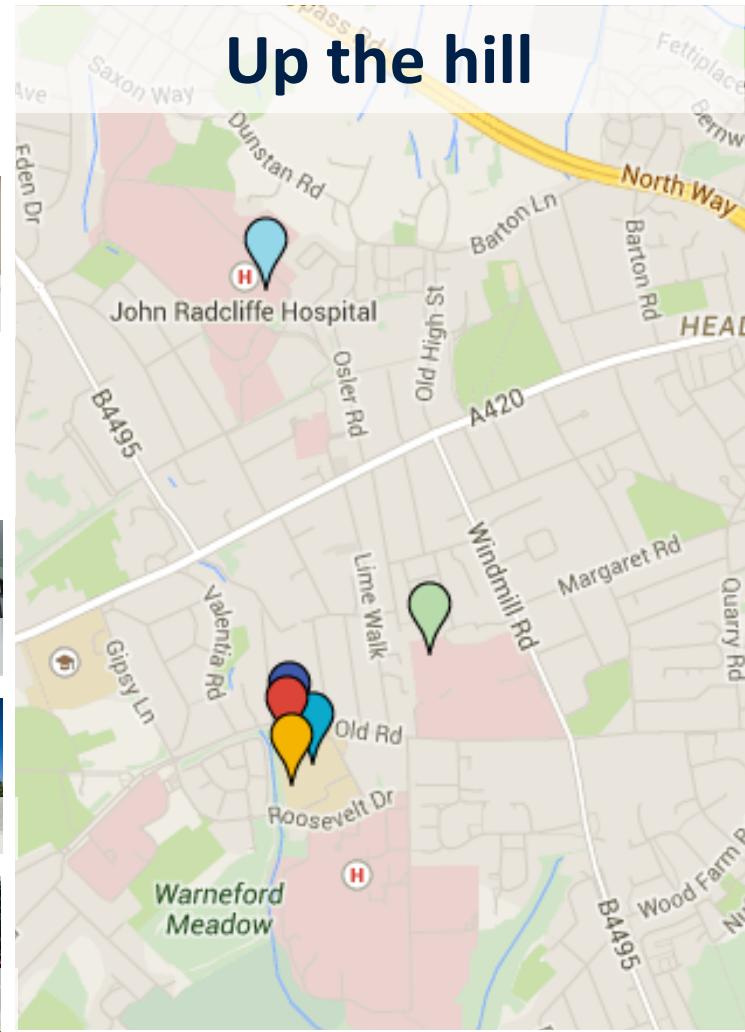


UNIVERSITY OF
OXFORD



ChemBio Hub

Chemical biology across Oxford





UNIVERSITY OF
OXFORD



ChemBio Hub

We initiated ChemBio Hub...

We secured funding from:



The ChemBio Hub vision:

- to provide the tools that will make it easier for Oxford University scientists to connect with colleagues to improve their research
- to satisfy funders that the data they have paid for is being managed according to their requirements,
- to make new alliances with pharma and biotech partners.



ChemBio Hub ethos

- Creating web based tools to help variety of research
- Create events and outreach to help these aims
- Taking existing tools where possible
- Open source code to help build a community

www.chembiohub.ox.ac.uk

@oxchembiohub



UNIVERSITY OF
OXFORD



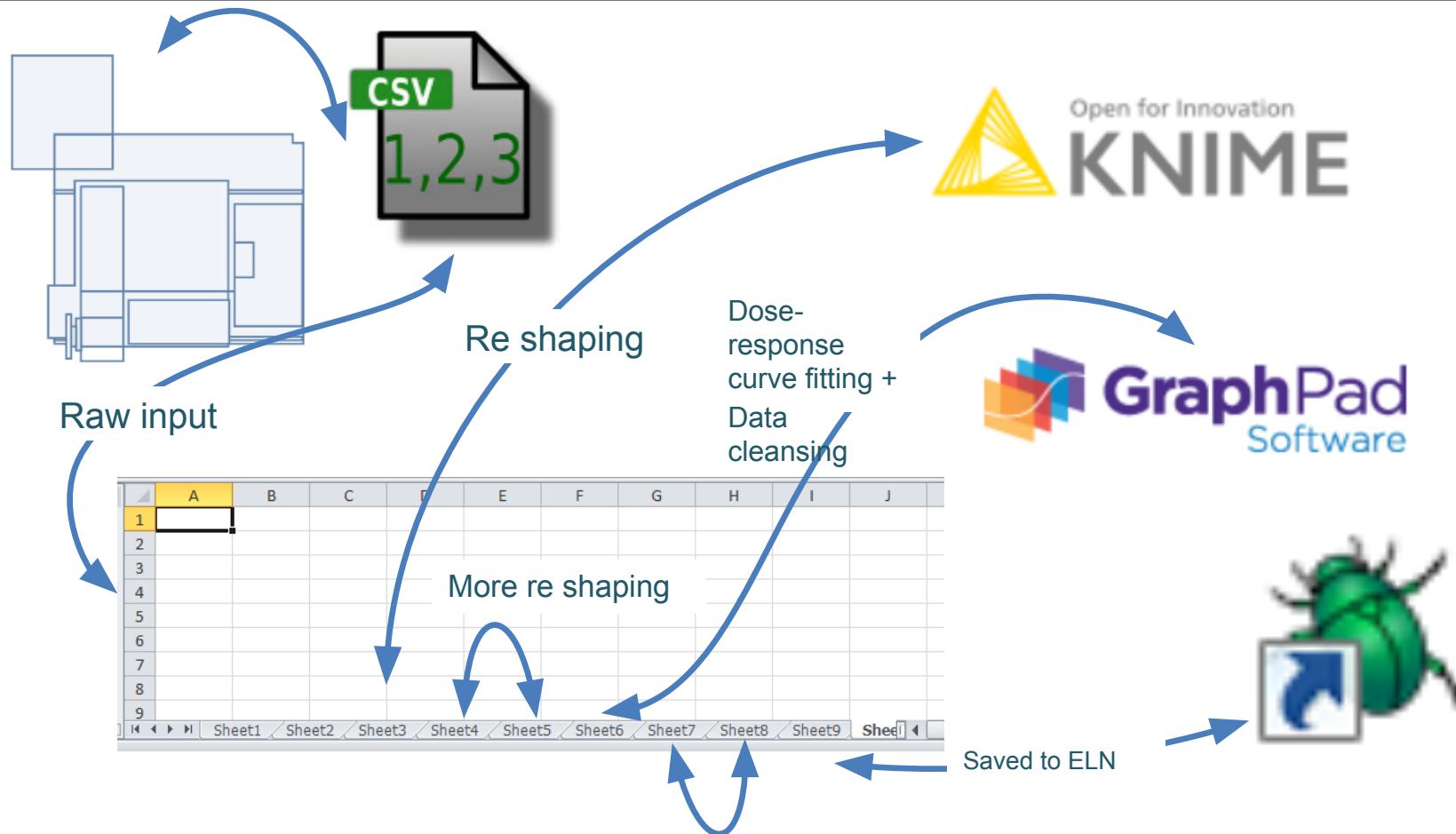
ChemBio Hub

ChemBio Hub Tools

The first utility we built was ChemBio Crunch.



Example Alpha Screen data flow



- The process is flexible and familiar but flawed and time consuming



We came up with - ChemBio Crunch

1) Upload raw data

Title*

BMG output file*
 No file selected.

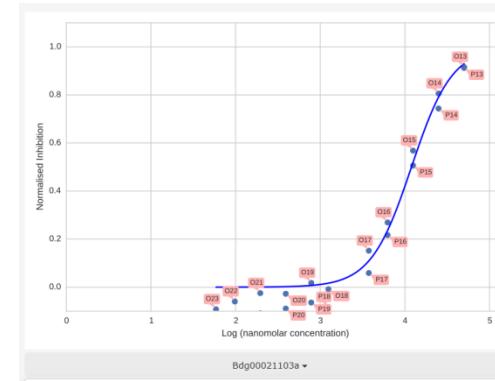
ESXX transfer file*
 No file selected.

Uploaded meta file
 No file selected.

2) Validate plates for systematic errors

	1	2	3	4	5	6	7	8	9	10	11
A	600761	596087	600267	592192	590748	592762	604252	599792	596182	592980	594548
B	500310	501095	500746	502544	517773	594281	570242	583205	567114	598714	559492
C	56582	65588	86735	119054	151981	225056	265777	326173	381292	439128	457483
D	52896	65075	85063	109326	145483	220879	235868	317623	367916	453334	459292
E	294196	383952	451478	487122	519441	534603	552425	544711	539778	548777	551988
F	301777	390602	430185	500137	528751	548435	546706	545642	557137	559623	565649
G	51813	59964	70471	101289	120599	214987	271700	354805	425486	470915	506235
H	50029	62377	69369	10276	121582	219610	273011	357599	422559	475798	497401
I	67381	110390	155553	226255	285950	403446	480243	496090	539581	563027	568898
J	65139	105630	164787	217284	274664	405842	459660	492936	544863	559834	564718
K	62859	110295	146889	196061	264472	368353	415948	484319	503952	523083	543077
L	67533	110162	149131	204839	254068	361646	422256	484389	507148	523226	531551
M	42104	56563	85177	126199	164426	282492	346541	421070	474696	490139	504412
N	50951	51870	83505	124639	175236	287584	358644	421577	486694	490713	520296
O	378169	559360	496033	577182	528316	564623	560120	574313	569487	572880	553394

3) Calculate IC50 for multiple plates



Mark as poor fit

4) Mark poor fits where observed

plate	coumpound	logIC50	ic50 (nM)	system_comments	user_comments	graph
2	BDG00021 060a	4.25927396	18166.6127	Low total inhibition, values could be inaccurate	no	
2	BDG00021 344a	4.8345796	68324.9943	Low total inhibition, values could be inaccurate	no	
2	BDG00019 351a	4.41049146	25733.0618	Low total inhibition, values could be inaccurate	no	

Export all workflow graphs as XLSX

5) Export and deposit in ELN
(Comments automatically generated where IC₅₀ may be inaccurate)



ChemBio Hub Crunch

[!\[\]\(d8f0f18404af9968c3d2c85c7c26102a_img.jpg\) ChemBio Crunch - The assay data analyser](#)

 Paul  Log Out

[◀ Prev Plate](#)

Plate [3]-01 (3 of 3 total plates imported)

[Next Plate ▶](#)[Export ▾](#)

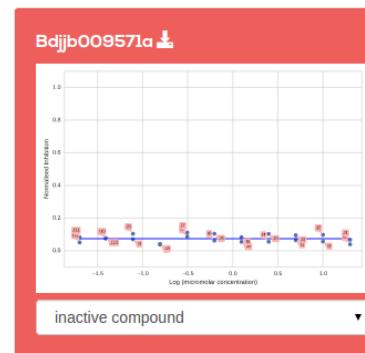
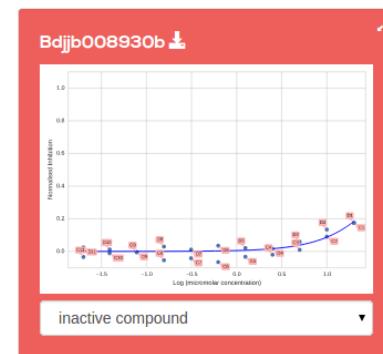
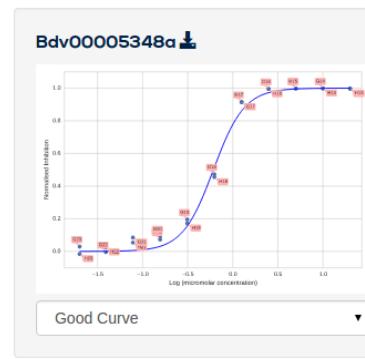
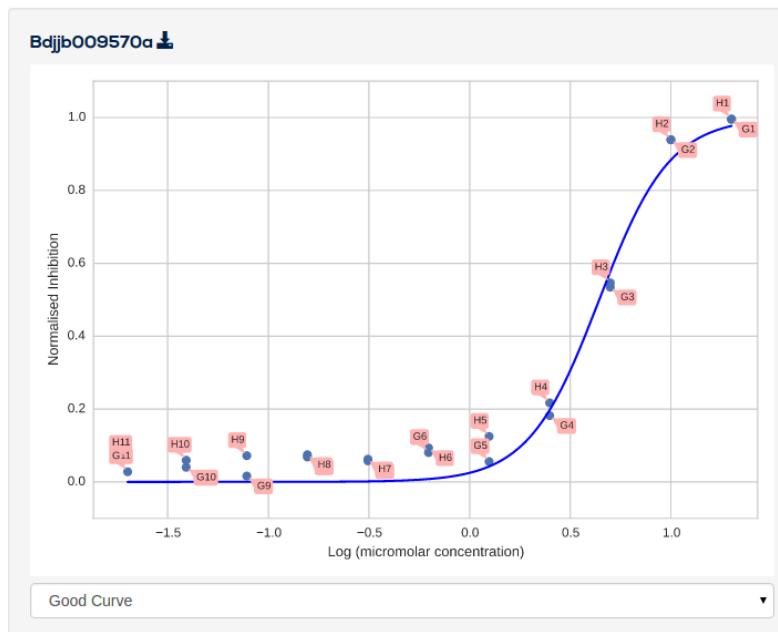
- Your data is displayed as a conditionally formatted heatmap
- Validate your data by eye.
- You can deselect wells as outliers by clicking the value in each table cell. Deselected cells will turn grey.
- You can deselect columns and rows by clicking on the label.
- You can re-select cells by clicking the value again.
- When you are finished, click Update and you will see the graphs below

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	940k	927k	911k	921k	896k	907k	894k	907k	876k	893k	909k	905k	898k	901k	885k	873k	882k	882k	891k	901k	915k	920k	919k	924k
B	863k	827k	829k	814k	829k	815k	801k	805k	801k	820k	803k	801k	795k	819k	823k	815k	809k	801k	850k	846k	847k	846k	851k	908k
C	635k	699k	761k	784k	793k	818k	800k	810k	772k	777k	794k	770k	422k	401k	511k	601k	674k	748k	752k	745k	779k	794k	794k	825k
D	633k	665k	723k	756k	752k	741k	759k	745k	772k	759k	749k	767k	378k	413k	500k	583k	665k	726k	743k	777k	773k	801k	797k	857k
E	319k	694k	776k	791k	773k	779k	755k	753k	745k	746k	754k	748k	2565	2356	27k	260k	560k	685k	738k	717k	741k	766k	772k	785k
F	310k	670k	758k	743k	738k	741k	727k	730k	729k	742k	745k	724k	1425	1938	27k	257k	582k	697k	744k	745k	739k	753k	788k	842k
G	3705	46k	357k	628k	725k	706k	724k	716k	756k	737k	747k	744k	3116	1045	1710	3990	65k	405k	618k	697k	703k	771k	745k	794k
H	3287	46k	348k	601k	672k	696k	720k	711k	713k	722k	718k	724k	1520	1083	1387	4009	64k	418k	636k	712k	727k	749k	780k	824k
I	738k	725k	719k	726k	704k	720k	701k	739k	714k	710k	729k	721k	2869	1634	2964	48k	311k	605k	666k	695k	702k	698k	722k	757k
J	716k	693k	695k	689k	726k	688k	683k	737k	688k	708k	705k	704k	1938	1634	2622	47k	317k	606k	661k	693k	712k	724k	735k	789k
K	801k	771k	755k	744k	733k	733k	727k	744k	724k	737k	739k	720k	760k	749k	726k	732k	748k	742k	764k	739k	766k	751k	771k	797k
L	758k	726k	731k	726k	721k	746k	717k	723k	722k	721k	709k	736k	736k	737k	734k	738k	746k	743k	755k	743k	750k	759k	801k	846k
M	806k	772k	747k	751k	743k	748k	752k	736k	742k	740k	742k	726k	747k	744k	751k	745k	748k	752k	761k	751k	741k	758k	780k	821k
N	798k	749k	750k	728k	729k	724k	730k	737k	735k	730k	726k	735k	740k	734k	738k	745k	763k	742k	749k	771k	786k	779k	805k	855k
O	100k	432k	637k	689k	720k	729k	724k	726k	728k	732k	713k	722k	728k	736k	724k	736k	740k	744k	753k	744k	733k	759k	783k	815k
P	84k	425k	675k	727k	754k	758k	767k	785k	763k	783k	769k	787k	785k	771k	782k	776k	782k	773k	791k	791k	790k	841k	852k	920k



ChemBio Hub Crunch

N	798k	749k	750k	728k	729k	724k	730k	737k	735k	730k	726k	735k	740k	734k	738k	745k	763k	742k	749k	771k	786k	779k	805k	855k
O	100k	432k	637k	689k	720k	729k	724k	726k	728k	732k	713k	722k	728k	736k	724k	736k	740k	744k	753k	744k	733k	759k	783k	815k
P	84k	425k	675k	727k	754k	758k	767k	785k	763k	783k	769k	787k	785k	771k	782k	776k	782k	773k	791k	790k	841k	852k	920k	





ChemBio Hub Crunch tech stack

ChemBio Crunch

Python Django / PostgreSQL



Python LMFit, Pandas framework,
Hstore filesystem

Allows chemists to process machine
output data quickly

Exports IC50 curves and reports





Results

- Production-ready web application written in 3 weeks
- ChemBio Crunch matches the GraphPad curve fit to within <1%
- Time to process plates cut by > 10x



UNIVERSITY OF
OXFORD



ChemBio Hub

ChemBio Hub Tools

The next utility we built was ChemReg.



Hasn't this already been fixed?

- Sure, there are lots of tools available.





UNIVERSITY OF
OXFORD



ChemBio Hub

ChemBio Hub ChemReg – registration complexity





ChemBio Hub ChemReg – registration complexity

- Stereochemistry - same bond organisation but 3D differences
- Tautomeric forms - interchangeable structures e.g. ketones and enols
- “Batches” - structural information exists already but “real world” data is different, for example commercial supplier, purity.
- Pan Assay Interference (PAINS) compounds may be less desirable to register for some but not others
- Private vs Public compounds / Security
- Blinded / Virtual compounds - do not initially have a structure

All of these issues mean you can't just “put” your compounds in a database with a generated ID.



UNIVERSITY OF
OXFORD



ChemBio Hub

ChemBio Hub ChemReg tech stack





elastic

+ Tastypie Web Services



ChEMBL



Open-Source Cheminformatics
and Machine Learning



PostgreSQL



ChemBio Hub ChemReg technology choices

Angular JS

- 2 way binding - minimises server calls
- Can still function with a different backend
- Part of modern web development flow



Django

- Compatibility with scientific libraries - Python
- Good access to web service architecture



ChEMBL & RDKit

- Well developed chemistry tools
- Strong user base
- Use of standards





Registration process

Projects ▾

ChemReg beta

Projects | Search | Help Paul ▾

[Back to Moghaddam Compounds project home](#)

Add compounds to Moghaddam Compounds

Data summary

Uploaded Records

19	with structures		
0	without structures		
0	not processed		

Unique Compounds

19	new to this project		
0	overlaps		

No duplicates found in uploaded data

19 new batches will be saved to the database and 0 batches will be ignored.

Start again Save these records

Single-page “wizard” walks chemists through the process of registration.
Can add compounds individually or as a batch



Registration process

 Edit and save your data below

19 results

[Remove Sorts](#)

[Remove Filters](#)

1 2 >

Structure	Row	Info	Action	Inchi Key	SGC_ID	Unmapped 	Map 
	1	• New	New Batch ▾	HZHLPMMAFRS...	K04056a		
	2	• New	New Batch ▾	JRCNZTHUXQ...	K04057a		
	3	• New	New Batch ▾	XUAXBQIDCW...	K04058a		
	4	• New	New Batch ▾	PHXBWCZPBJ...	K04059a		
	5	• New	New Batch ▾	UDJKXRBGHS...	K04060a		
							

Users can add standard format files (SDF, Chemdraw or Excel)
Validation at every step - Information about what is currently in the DB
Problems and issues can be identified early in the process



Registration process

Search

ChemDoodle®

68 results

List Edit



Gallery

10/page

UOX ID

Search multiple IDs

Add your IDs and click on them as they appear in the drop-down list.

Project

Brennan group

Added after

Added before

Functional Group

None

SMILES or SMARTS

Search SMILES or SMARTS string

Structural search type

Substructure Exact Match

Upload ID

Filter by project data values:

chem

ChemReg batch Identifier: Summary information

Chem_reg Identifier: Summary information

Supplier: BDH CHEMICALS LTD

Chem_reg batch Identifier: hFTO clones

Chem_reg Identifier: Summary information

Chem_reg Identifier: hFTO clones

Chem_reg Identifier: mFTO clones



Registration process

Batch 29778 - UOXLCO6BXY Copy ID Copy CDXML Brennan group

Compound Information

Added By	Registered on
Adam Hendry	17-08-2015
SMILES	AlogP
CCOC1=C[C@@](CC)	3.51
Standard INCHI	Mol Weight
InChI=1S/C11H20O/c:	168.15
Standard INCHI key	Mol Formula
VEMFOXUQCJLHCH-	C11H20O

Project Data Toggle Edit Mode

Custom Identifier	Amount
Name and Synonyms	Physical Form
Date of Synthesis	Supplier
Storage Conditions	Library
Keywords	Library well ID



ChemBio Hub ChemReg enables...

- Addition of extra custom data relevant to their project
- Extra data fields can be stipulated or required per-project - differentiation between project and user-specific (supplementary) fields
- Autocomplete searching for these fields.
- Results can be exported as SD or Excel file in two clicks for auditing etc.



RDKit plays a large role in this application

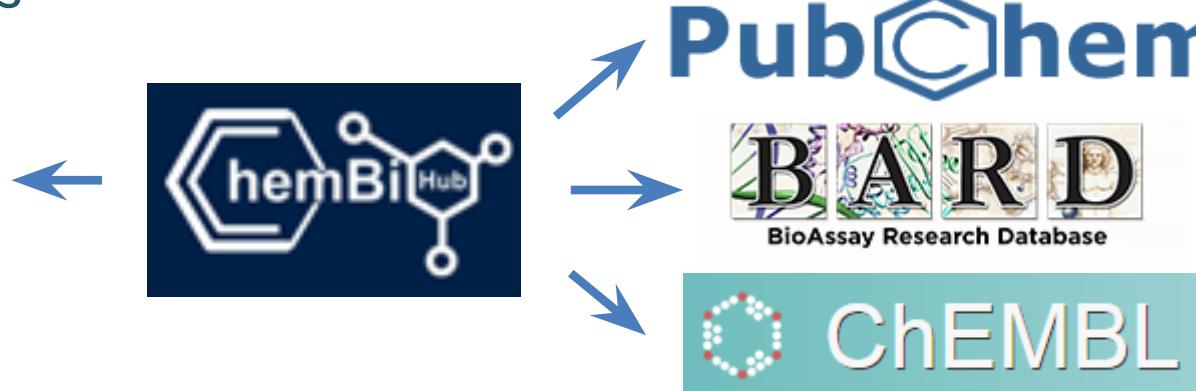
- Allows registration to occur!
 - Comparison of existing compounds
 - Warnings of PAINS or other undesirable compounds
 - Substructure searching
- For future ChemReg development will allow:
 - more complex structural searches
 - profiling and auto-searches



Data Transfer and Deposition

- ChemBio Hub tools will allow users to deposit their data in external services

DATABANK



- Where deposition APIs do not exist e.g. ChEMBL– we can help build them
- Long term archiving in <https://databank.ora.ox.ac.uk/>



UNIVERSITY OF
OXFORD



ChemBio Hub

ChemBio Hub Tools

Under development - ChemBio Hub AssayReg

- Data repository for assays and results
- Allow easier sharing to public repos and other scientists



How you can help

We need to build a community around these tools

- For continuing development and support
- For feedback and improvement
- To spread the project ethos

Try the tools!

<http://github.com/thesgc/>
info@chembiohub.ox.ac.uk



Thanks to RDKit UGM for listening!

Acknowledgements

ChemBio Hub team:

Brian Marsden

Andrew Stretton

Karen Porter

Adam Hendry

(Michael O'Hagan)

SGC team:

David Damerell

Anthony Bradley

James Bennett

Oleg Fedorov

Octovia Monteiro

Funders: NDM, Wellcome Trust, Higher Education Funding Council for England, John Fell Fund