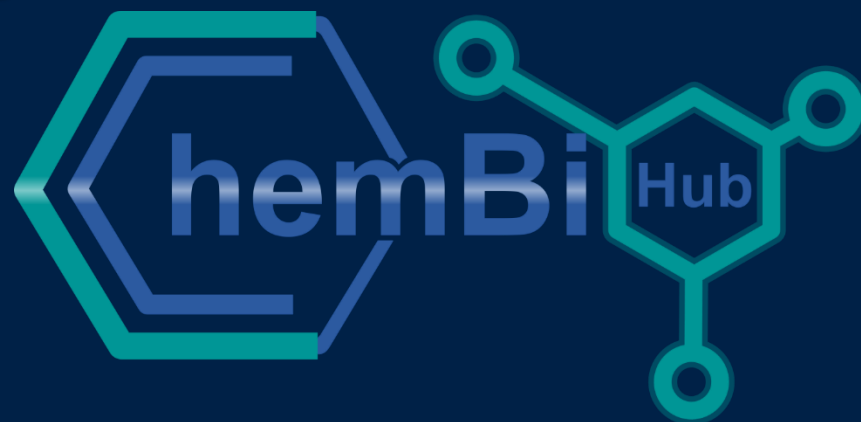


Things to do with Data: Taming University Chemical Biology Knowledge



<http://chembiohub.ox.ac.uk>

Brian Marsden
SGC, NDM
Kennedy Institute of Rheumatology, NDORMS

Covering:

- Who am I and why am I here?
- Chemical Biology – what's that?
- The basic science (in 10 minutes)
- So, what's the problem?
- How are we tackling it?
- What have we done so far?
- What's next?

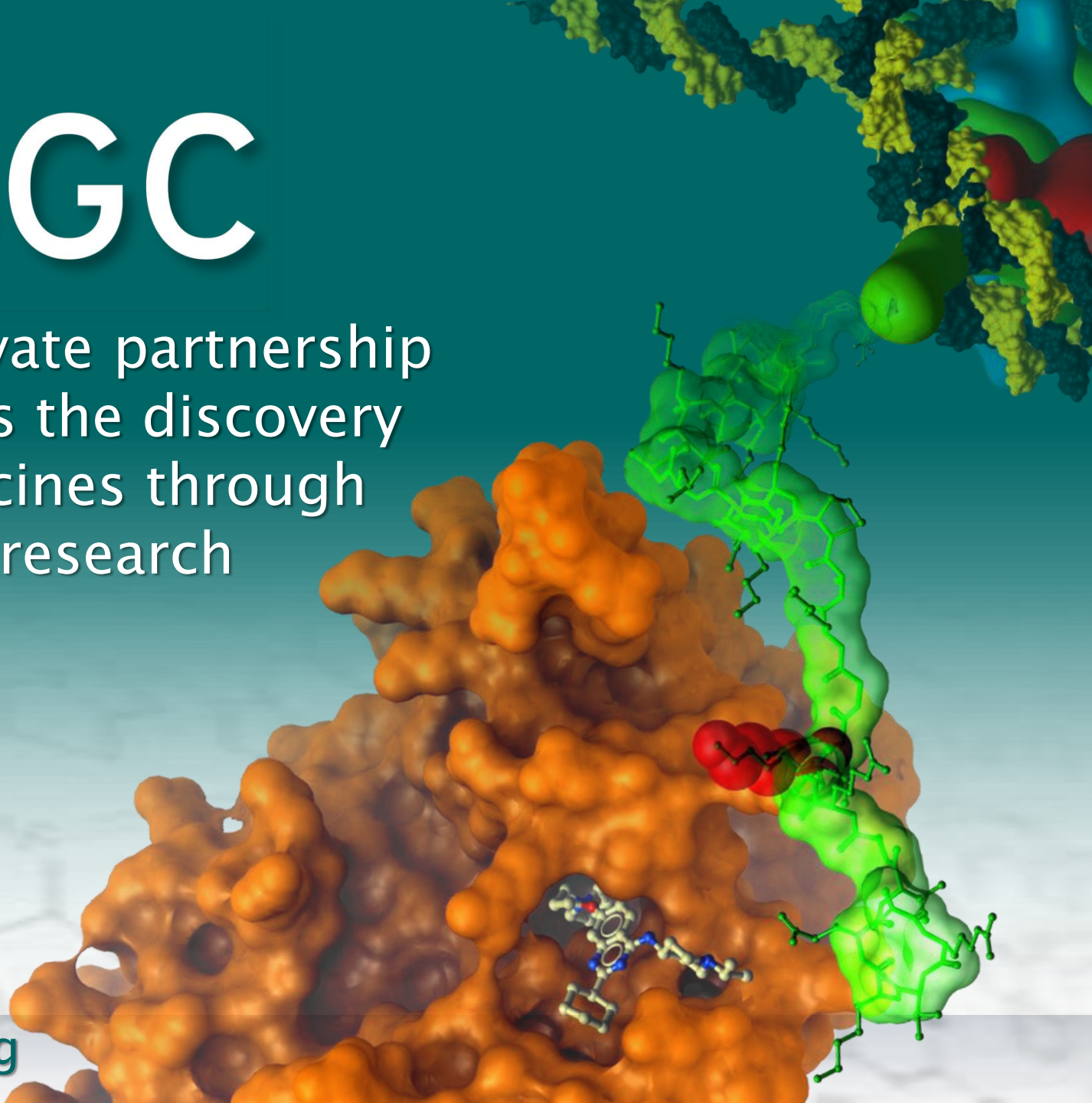


SGC

A public-private partnership
that supports the discovery
of new medicines through
open access research



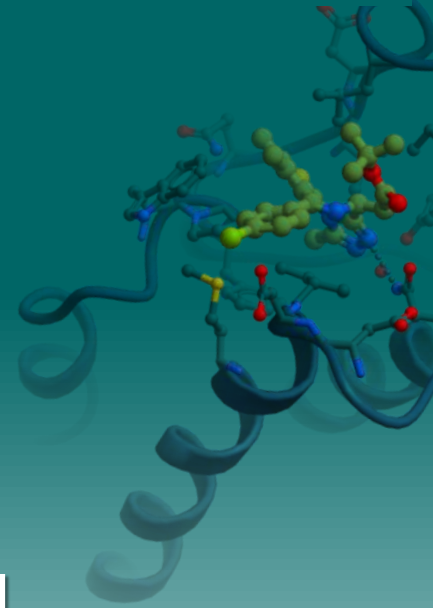
www.thesgc.org





SGC at a glance

- Operations started in June 2004
- 3 government agencies, Wellcome Trust and 10 leading pharma companies
- Aggregate funding in excess of \$300M
- +200*-strong team in Oxford, Toronto & Campinas
- Open Access Policy:
 - Promptly placing results, reagents and know-how in the public domain
 - SGC scientists **never** file patents



What is chemical biology?

- Chemical biology means different things to different people
- Are the projects below examples of chemical biology...?
- Designing molecules to act as high-energy foodstuffs



What is chemical biology?

- Chemical biology means different things to different people
- Are the projects below examples of chemical biology...?
- Using nature's chemistry to produce clean energy



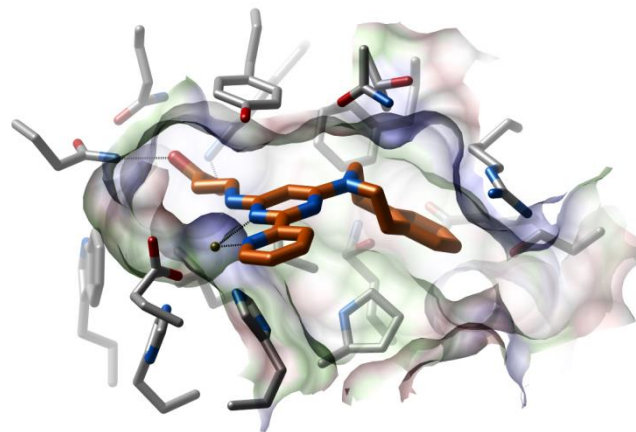
What is chemical biology?

- Chemical biology means different things to different people
- Are the projects below examples of chemical biology...?
- Tweaking biological structures at the atomic level

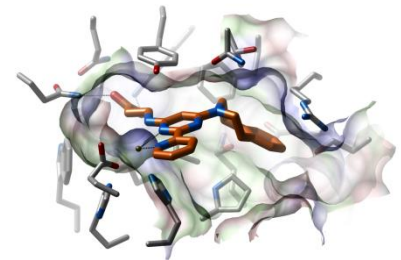


What is chemical biology?

- Chemical biology means different things to different people
- Are the projects below examples of chemical biology...?
- Testing the effect of chemicals on biological molecules and cells



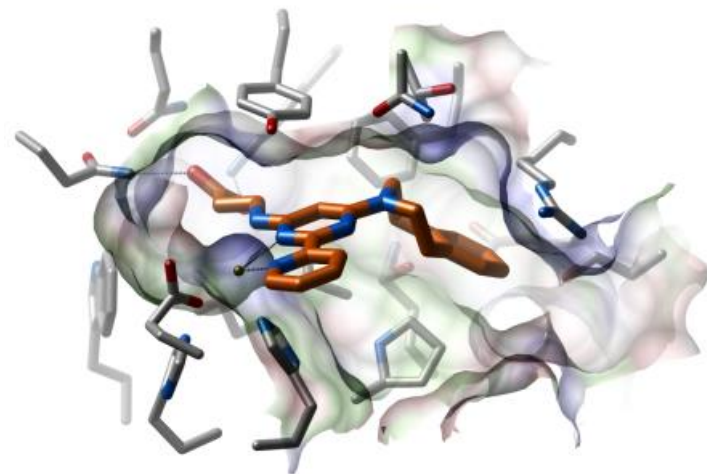
What is chemical biology?



- All these (and more!) can be considered as chemical biology...
 - ...but also as medicine, physics, engineering, nanotechnology, drug discovery, synthetic biology, biochemistry, etc...
- Too diverse to deal with all at once!

What *sort* of chemical biology?

- ChemBio Hub initial focus:
 - Effect of chemicals on biological molecules and cells
- Why?
 - Expertise of our group
 - Significant amounts of data generated across the University
 - Demand from pharmaceutical industry for this data
 - Complements public data curation initiatives
e.g. ChEMBL, PubChem



And now for the Science 'Bit'...



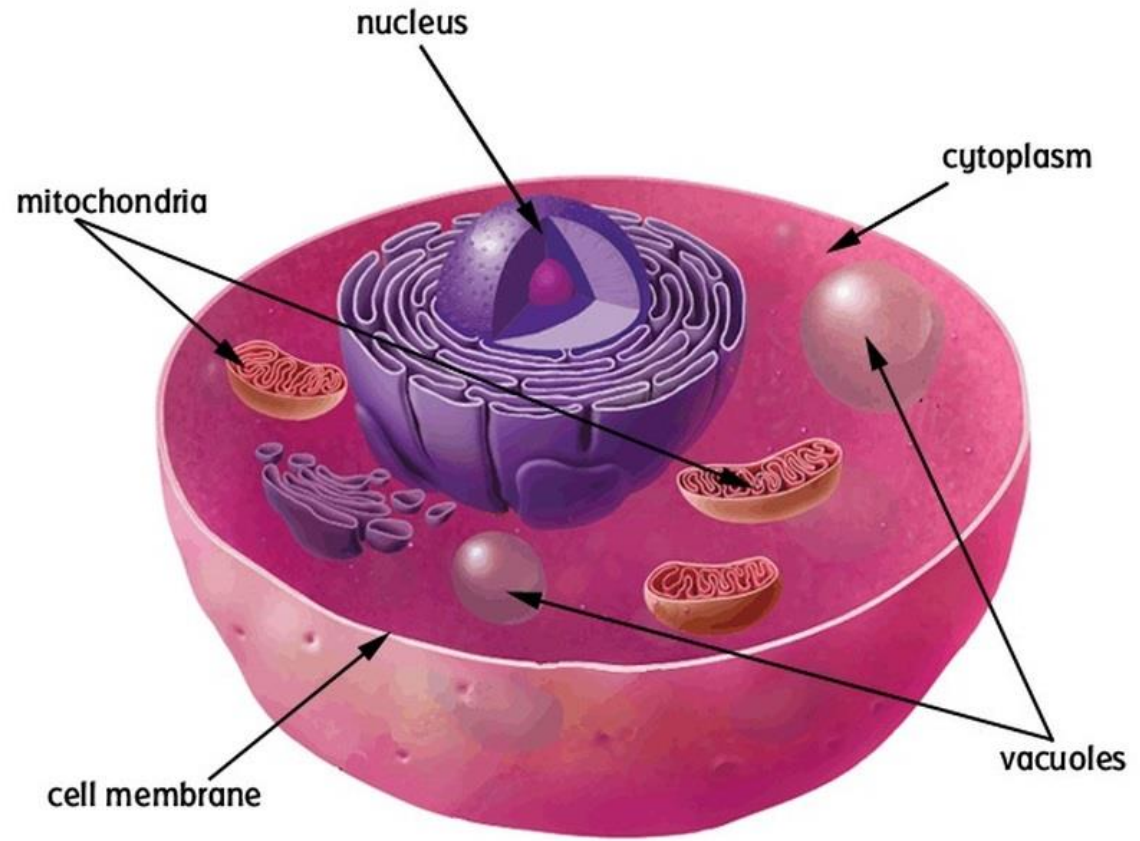
Cells

The basic unit of living things

Over 200 distinct cell types

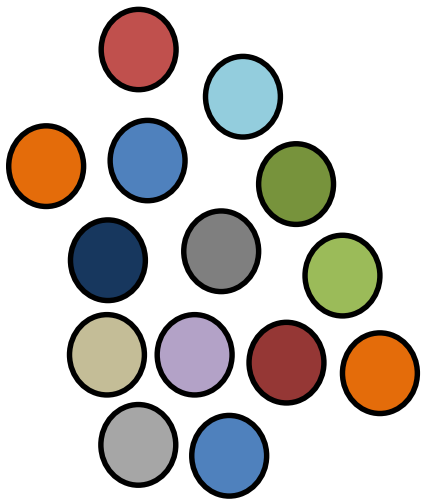
Core components include

- Nucleus
- Mitochondria
- Cytoplasm
- Cell membrane
- Millions of proteins



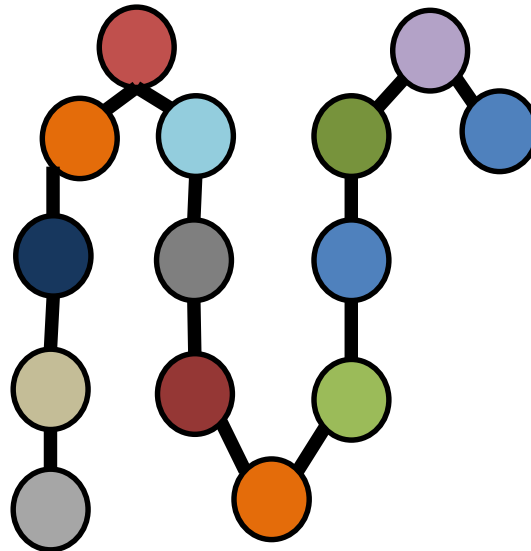
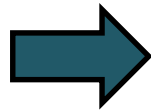
What is a protein?

Proteins – Large biological molecules consisting of one or more long chains of amino acid residues



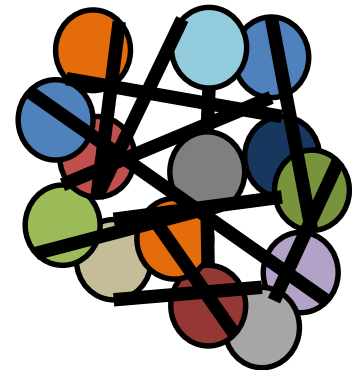
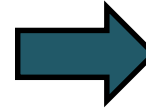
Amino Acids:

Proteins are made up of individual building blocks



Polypeptide:

When assembled they form a peptide chain

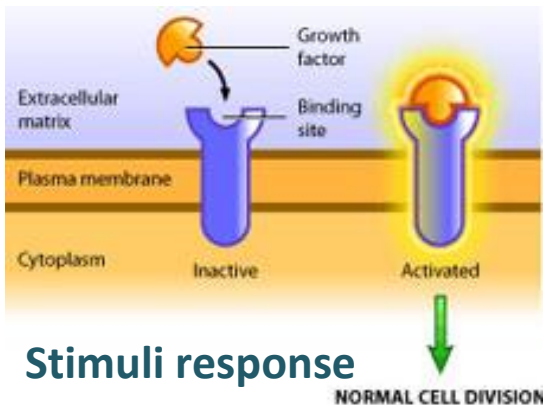
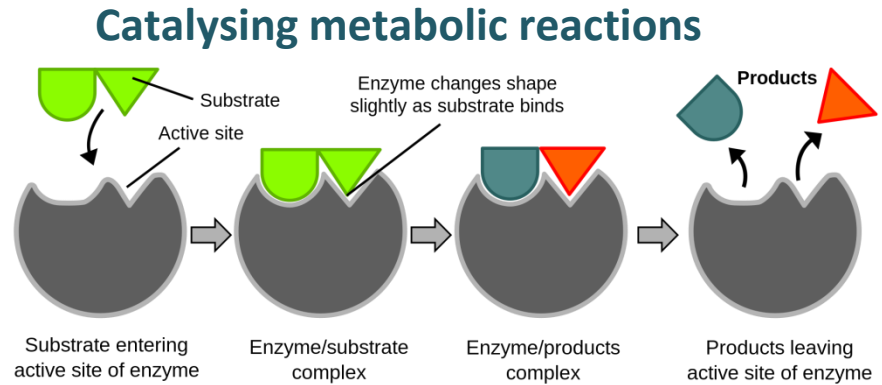
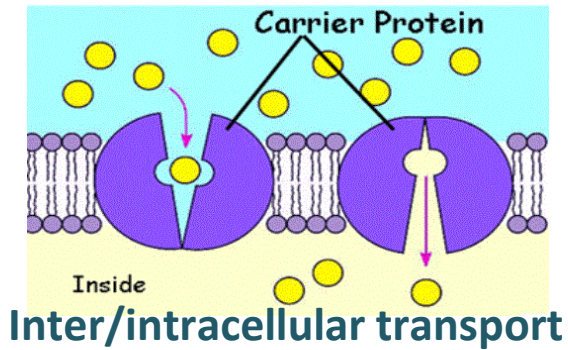


Protein:

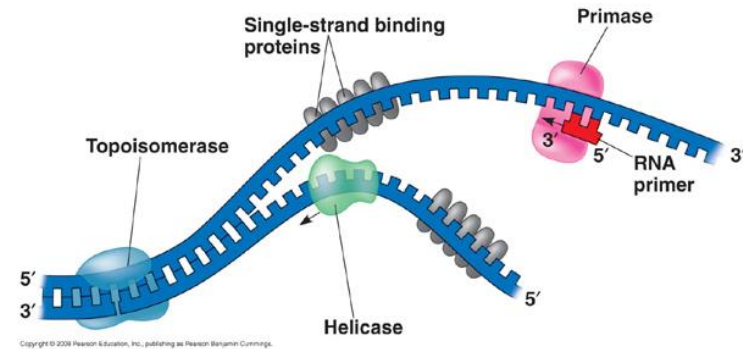
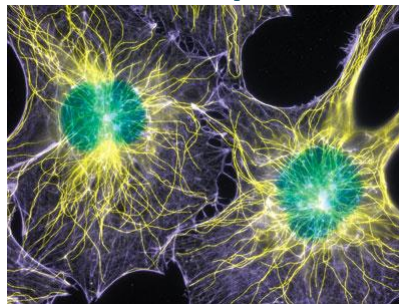
Specific folding of peptide results in a protein

What do proteins do?

Proteins are essential to life and play a significant role in almost every biological function, including:



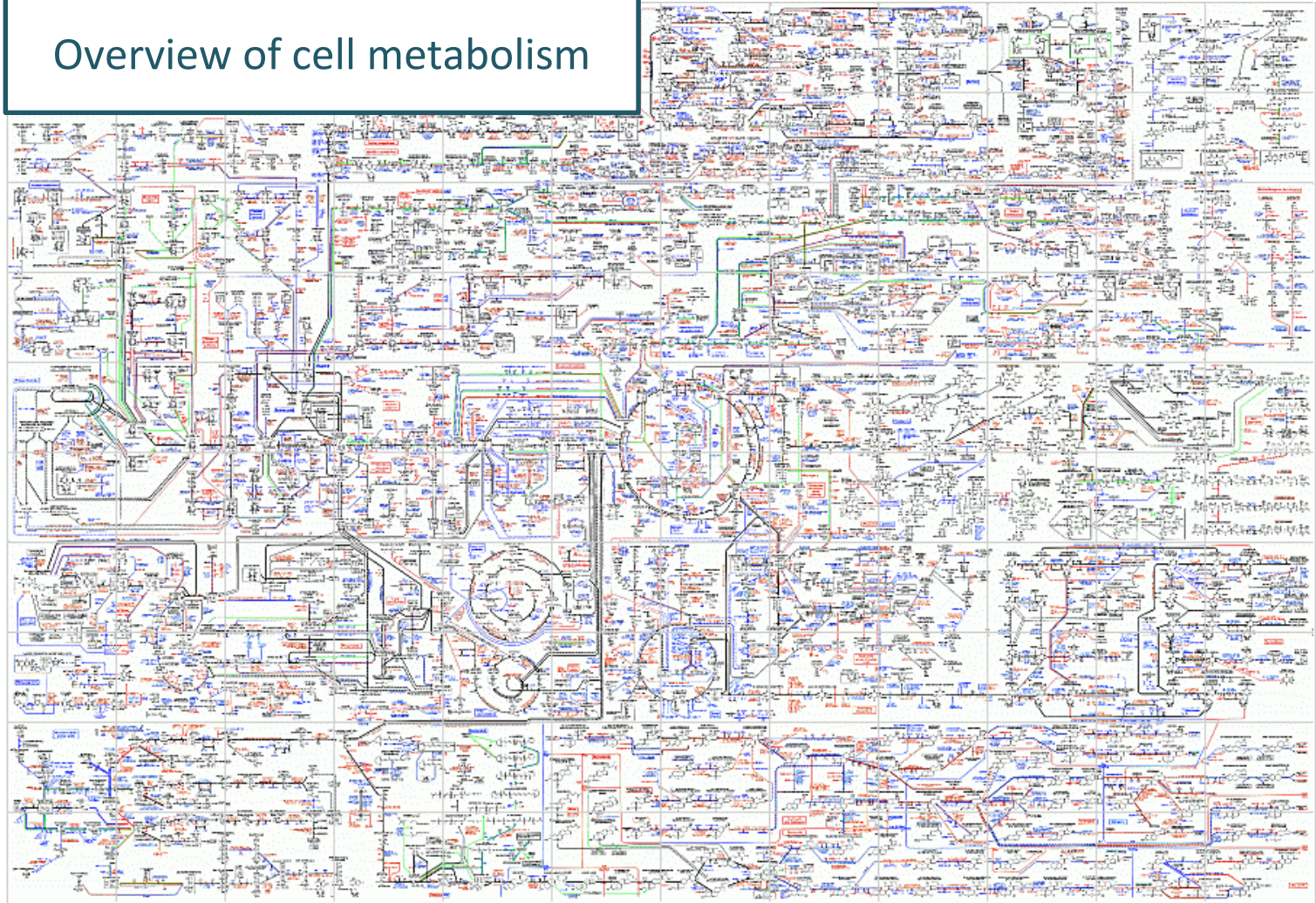
Structural components



DNA replication

Controlled protein expression is essential

Overview of cell metabolism



But it can go wrong

**WE ARE
MACMILLAN.
CANCER SUPPORT**



CANCER
RESEARCH
UK

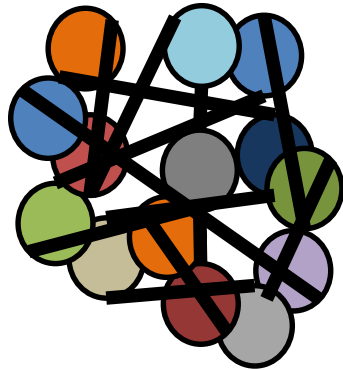
The
Breast
Cancer



Research
Foundatio



Abnormal protein
regulation can lead to
a wide variety of
diseases



TOGETHER WE CAN



**STOP
DIABETES**

American Diabetes Association.

DIABETES UK
CARE. CONNECT. CAMPAIGN.



British Heart
Foundation



HEART &
STROKE
FOUNDATION OF
NEWFOUNDLAND
& LABRADOR

Finding answers. For life.

A real-life example

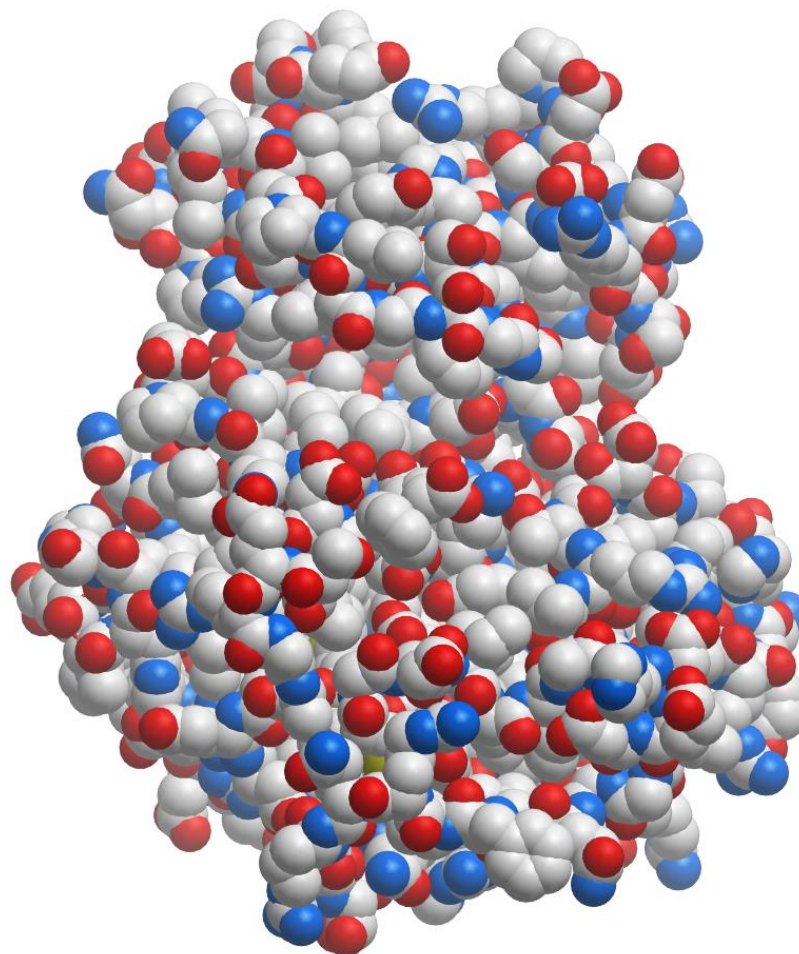
Human PIM1

A protein kinase.

It puts a phosphate on other proteins as a means to cascade important signals within the cell.

When this goes wrong, cancer results.

A small-molecule can help to prevent this by inhibiting the protein's default action.



A real-life example

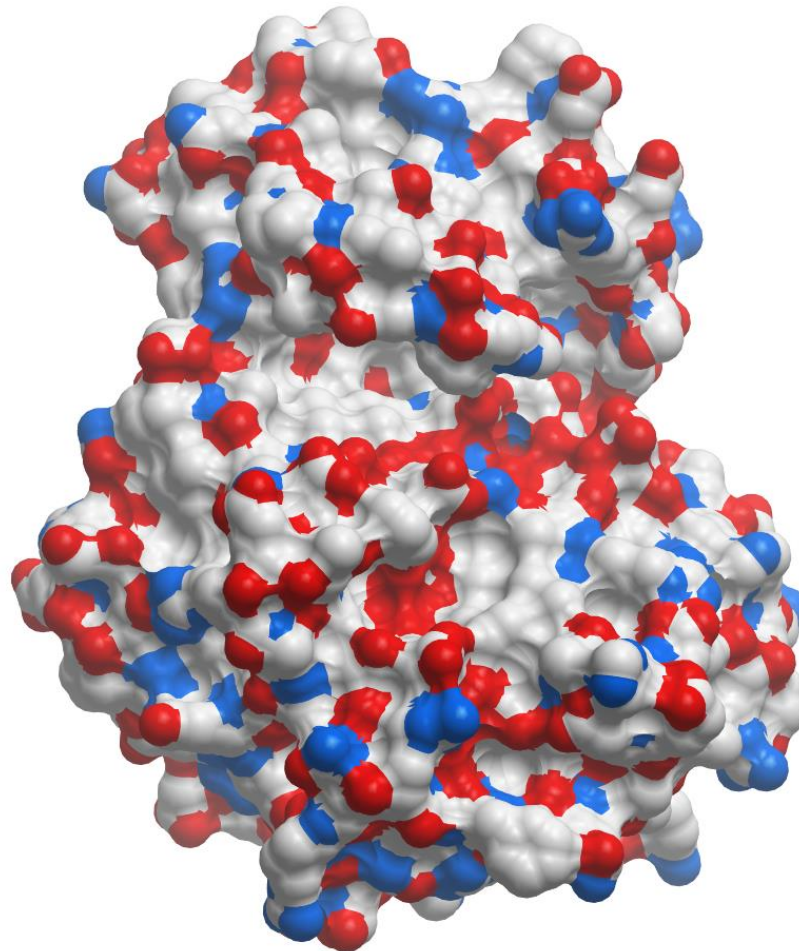
Human PIM1

A protein kinase.

It puts a phosphate on other proteins as a means to cascade important signals within the cell.

When this goes wrong, cancer results.

A small-molecule can help to prevent this by inhibiting the protein's default action.



A real-life example

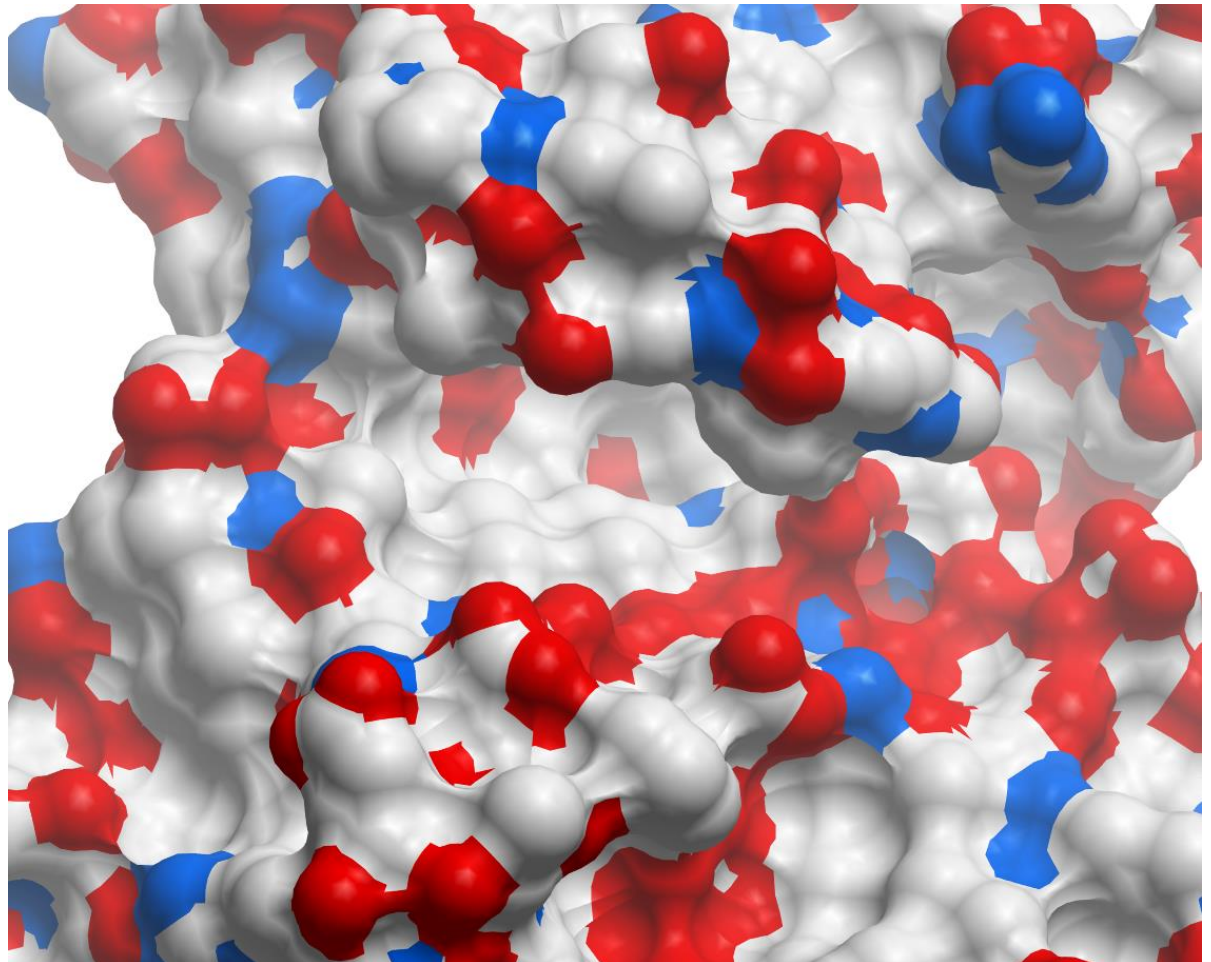
Human PIM1

A protein kinase.

It puts a phosphate on other proteins as a means to cascade important signals within the cell.

When this goes wrong, cancer results.

A small-molecule can help to prevent this by inhibiting the protein's default action.



A real-life example

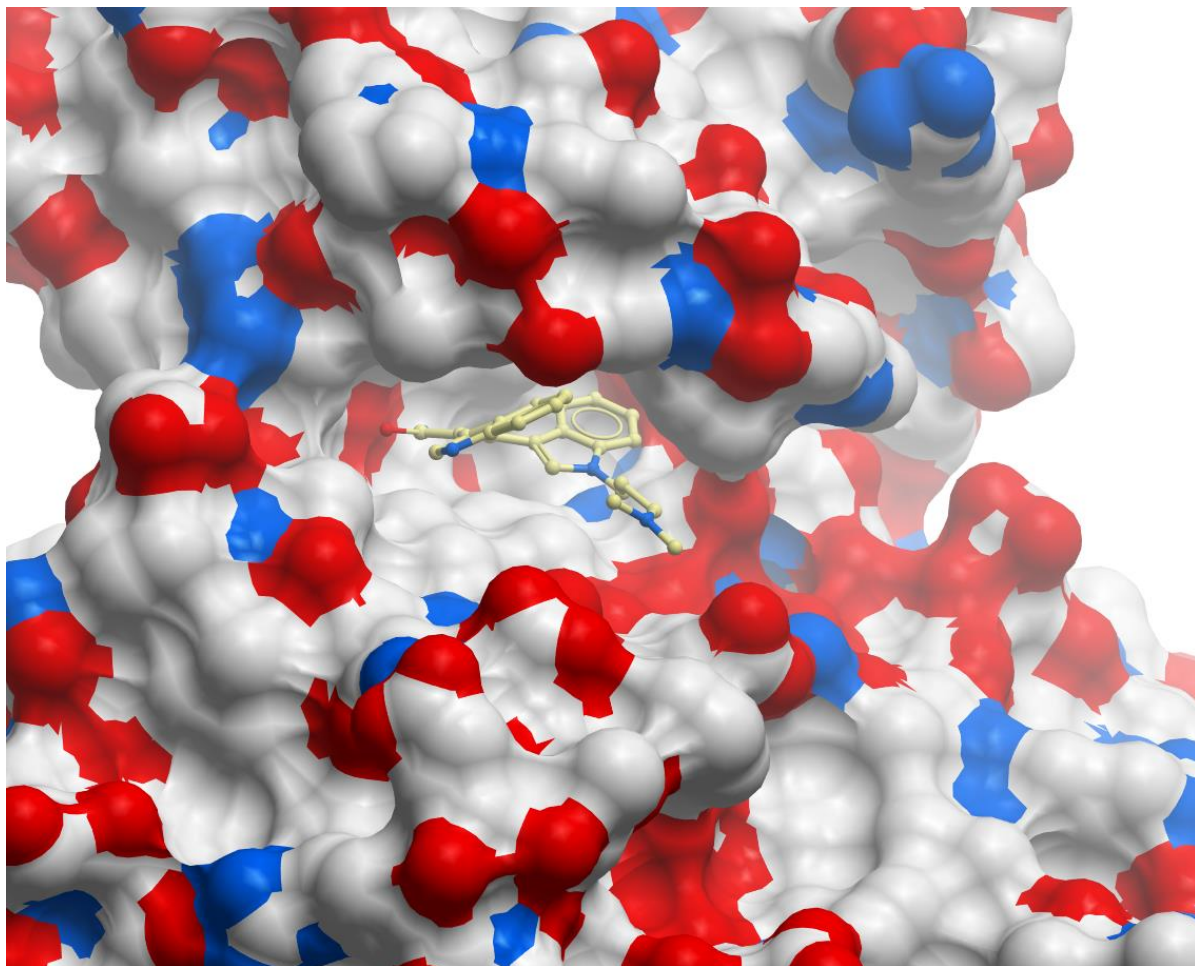
Human PIM1

A protein kinase.

It puts a phosphate on other proteins as a means to cascade important signals within the cell.

When this goes wrong, cancer results.

A small-molecule can help to prevent this by inhibiting the protein's default action.



A real-life example

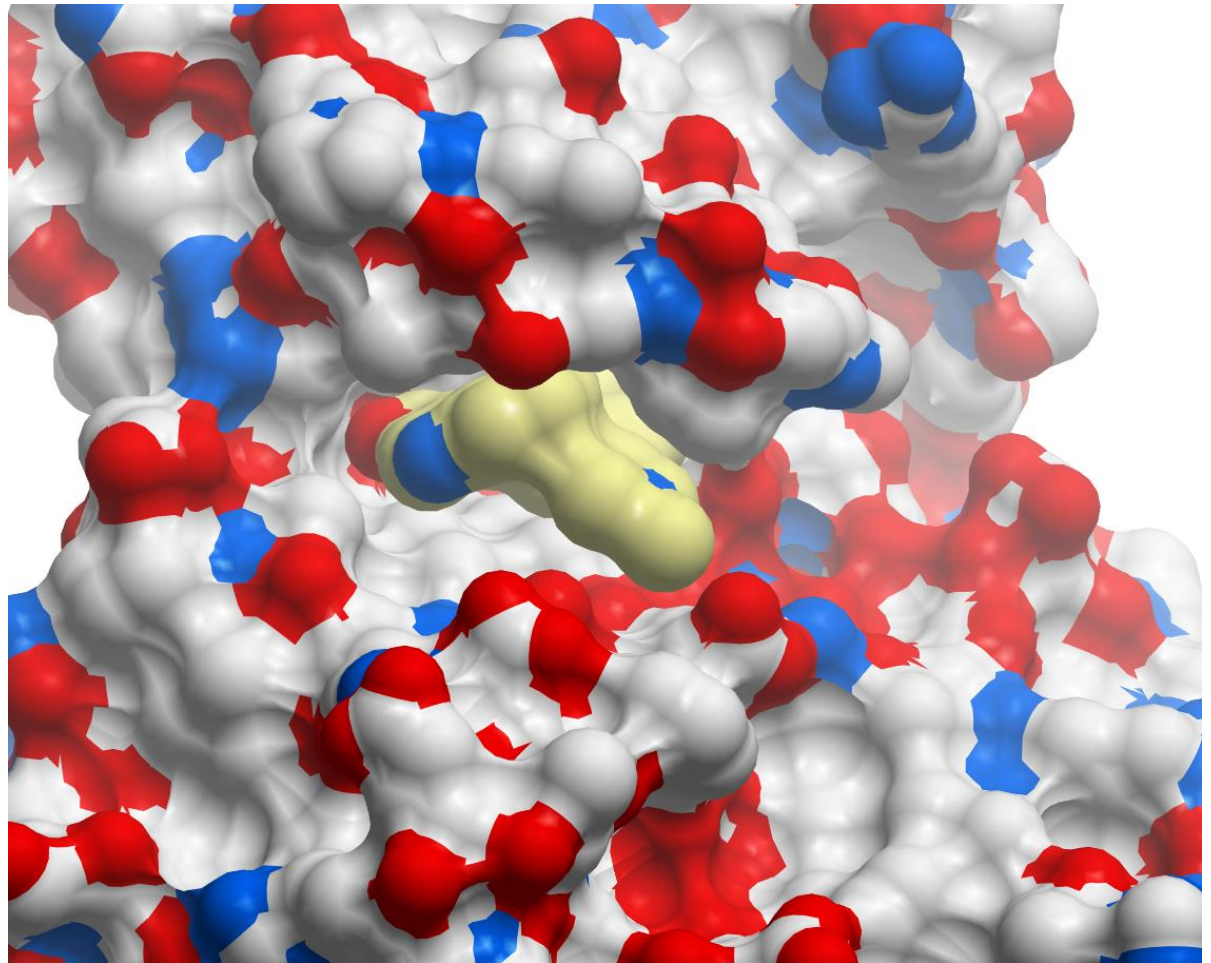
Human PIM1

A protein kinase.

It puts a phosphate on other proteins as a means to cascade important signals within the cell.

When this goes wrong, cancer results.

A small-molecule can help to prevent this by inhibiting the protein's default action.



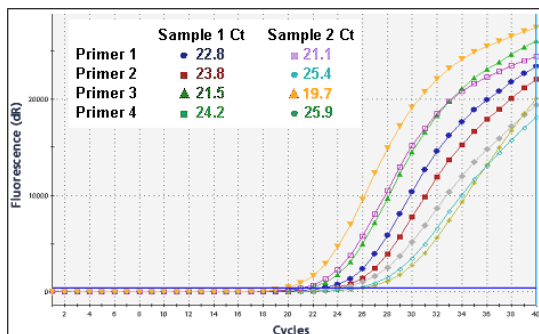
What is an assay?

“an analytic procedure for qualitatively assessing or quantitatively measuring the presence, amount or functional activity of a target entity (the analyte)”.

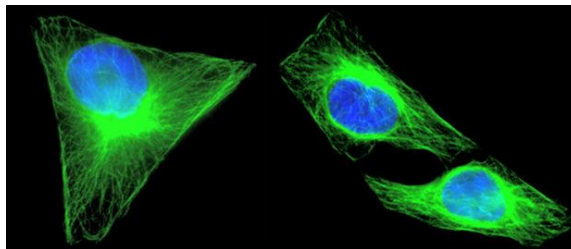
- Depending on the question, there are any number of potential assay types available
- Picking the most appropriate assay is crucial
- But knowing what is available and how to implement it is not easy

Kinetic assays
Bioassays
Ligand binding assays
Fluorescence assays
Colony count assays
Photometric assays
Absorbance assays
Immunoassays

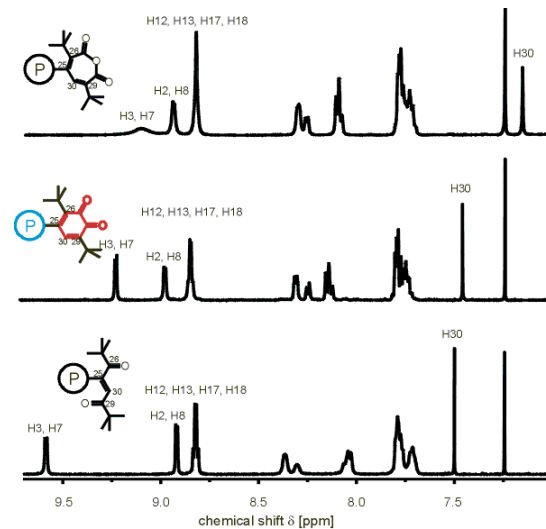
Generic assay (meta) data capture is very challenging



qPCR data

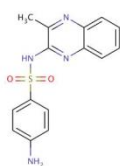
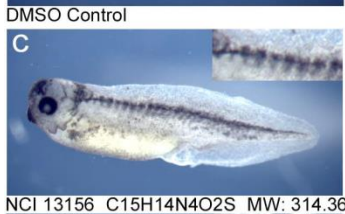


Immunostaining



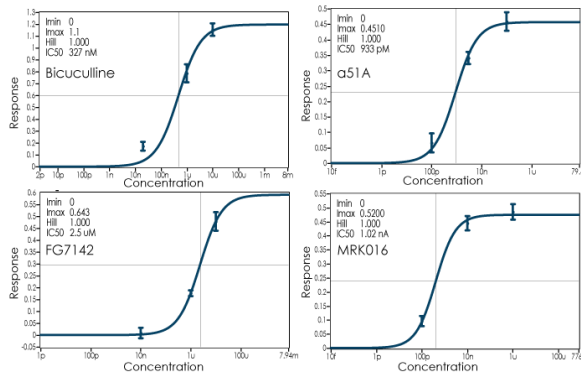
NMR spectra

There is an incredibly diverse variety of assay data types

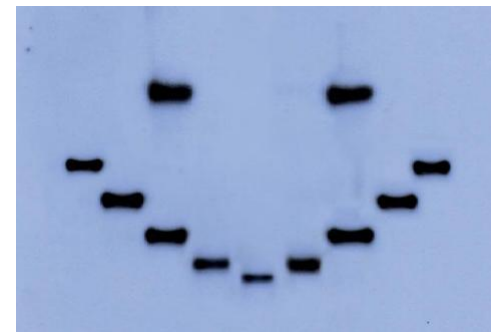


NCI 13156 C₁₅H₁₄N₄O₂S MW: 314.36

Phenotypic data



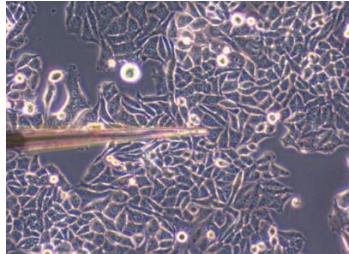
IC₅₀ data



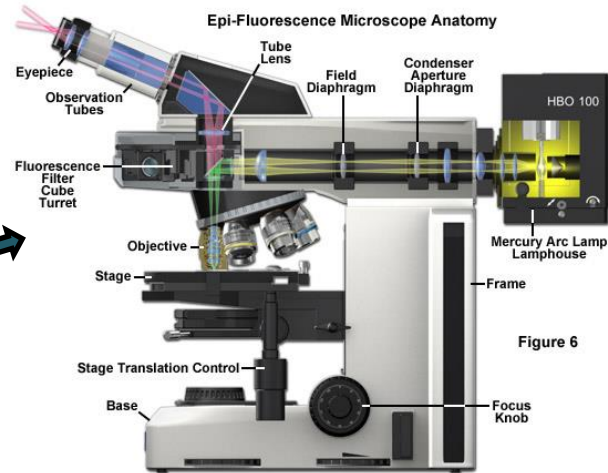
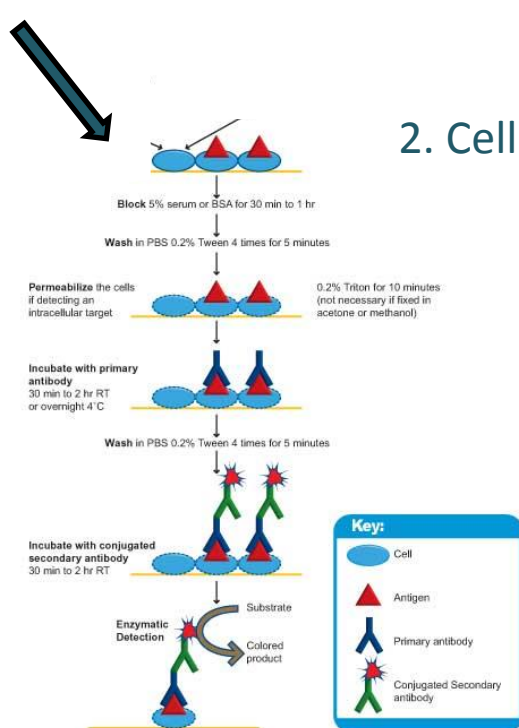
Western Blot

Generic immunofluorescence protocol

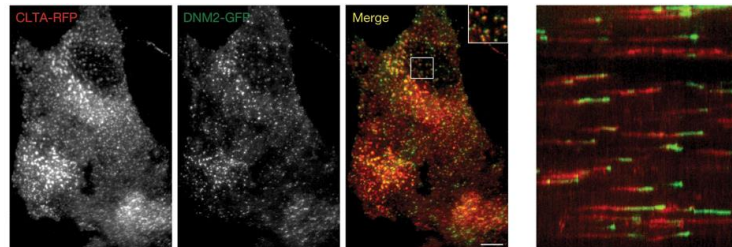
1. Cell culture



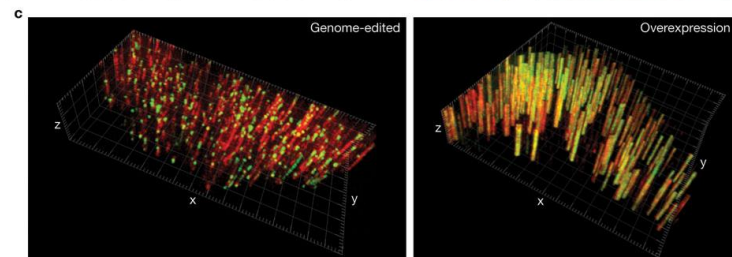
2. Cell staining



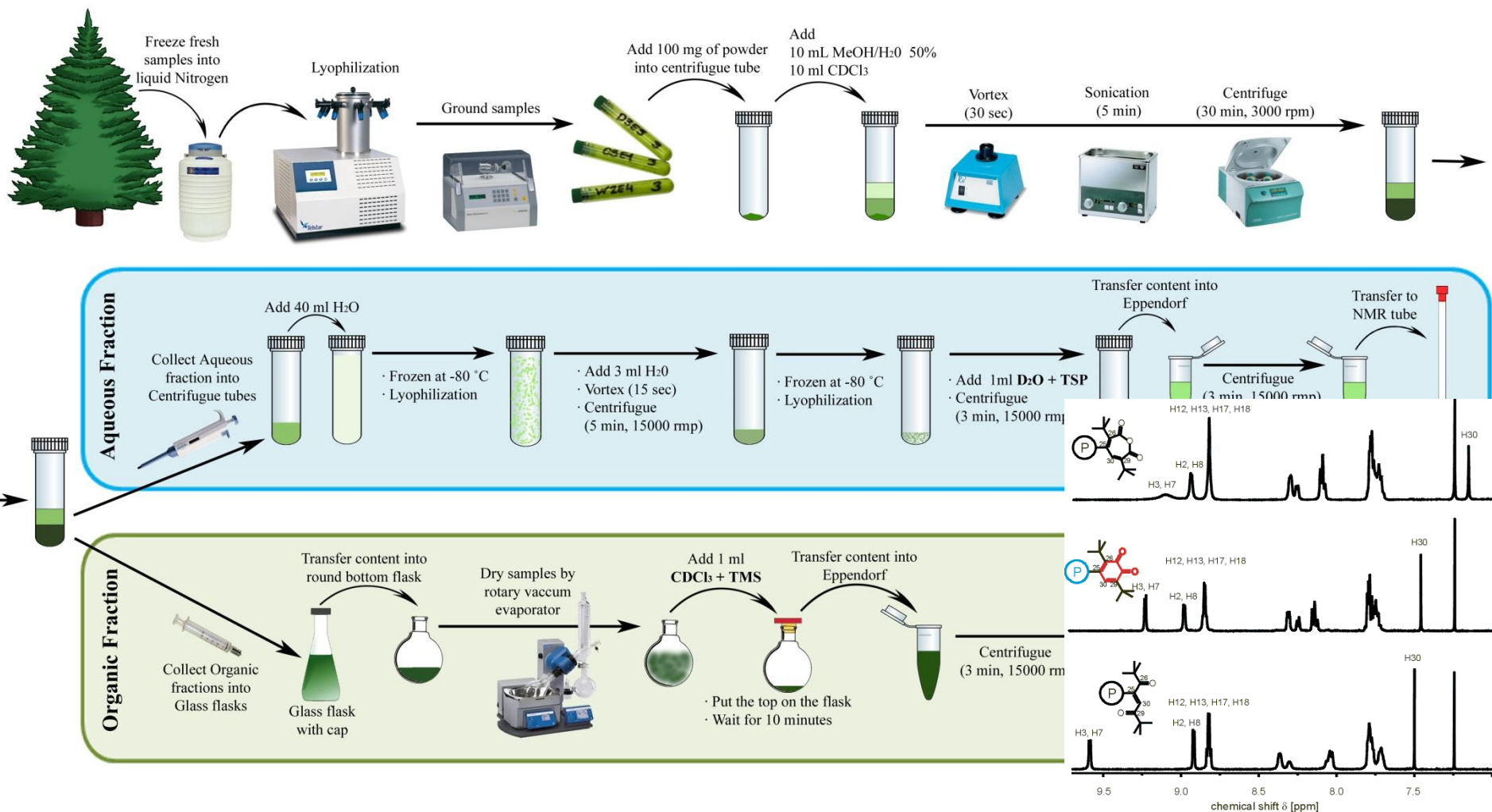
3. Fluorescence microscopy



4. Image analysis



Generic NMR protocol





Target focused small molecule drug discovery

Drug Target

Drug Candidate

Target
Discovery

Lead
Discovery

Lead
Optimization

Pre-
Clinical

Ph I

Ph II

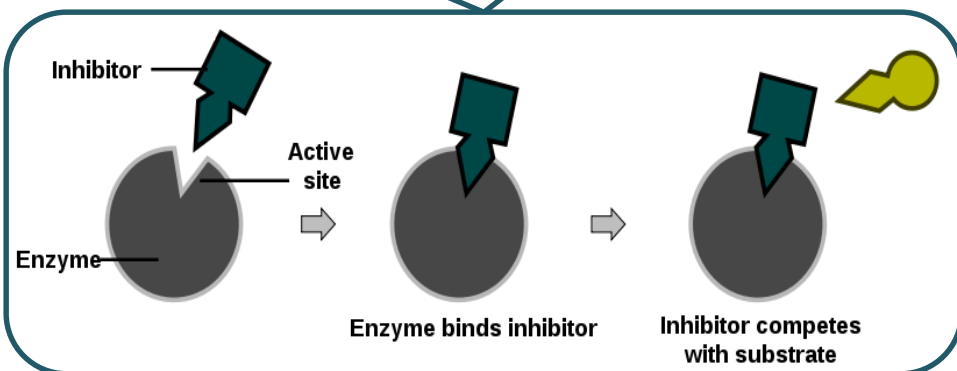
Ph III

Approval
Launch

Discovery

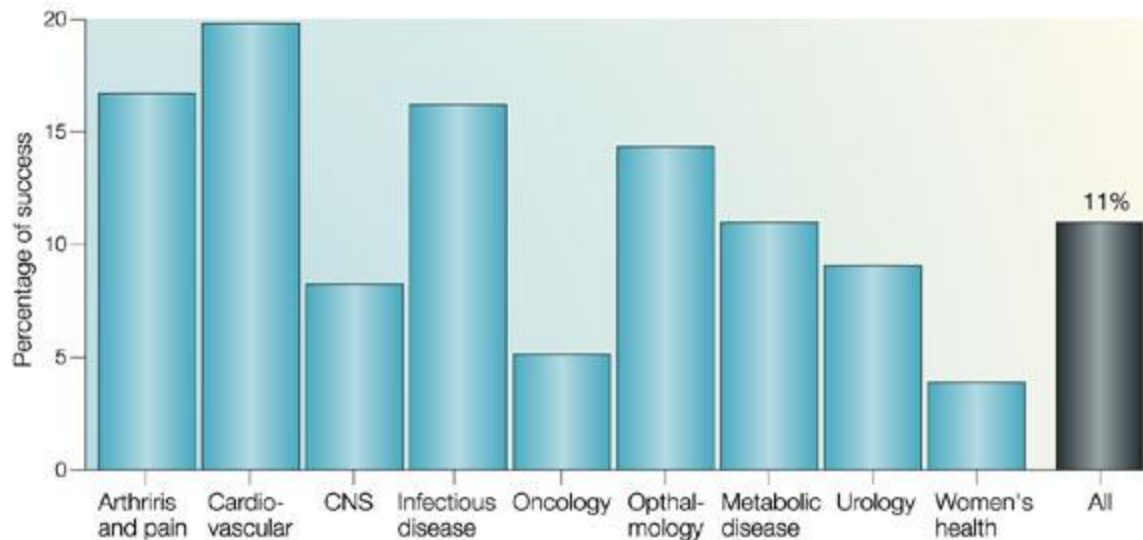
Development

Clinical



Carefully selected small molecule inhibitors modulate a specific protein to induce a therapeutic effect.

Drug discovery is hard



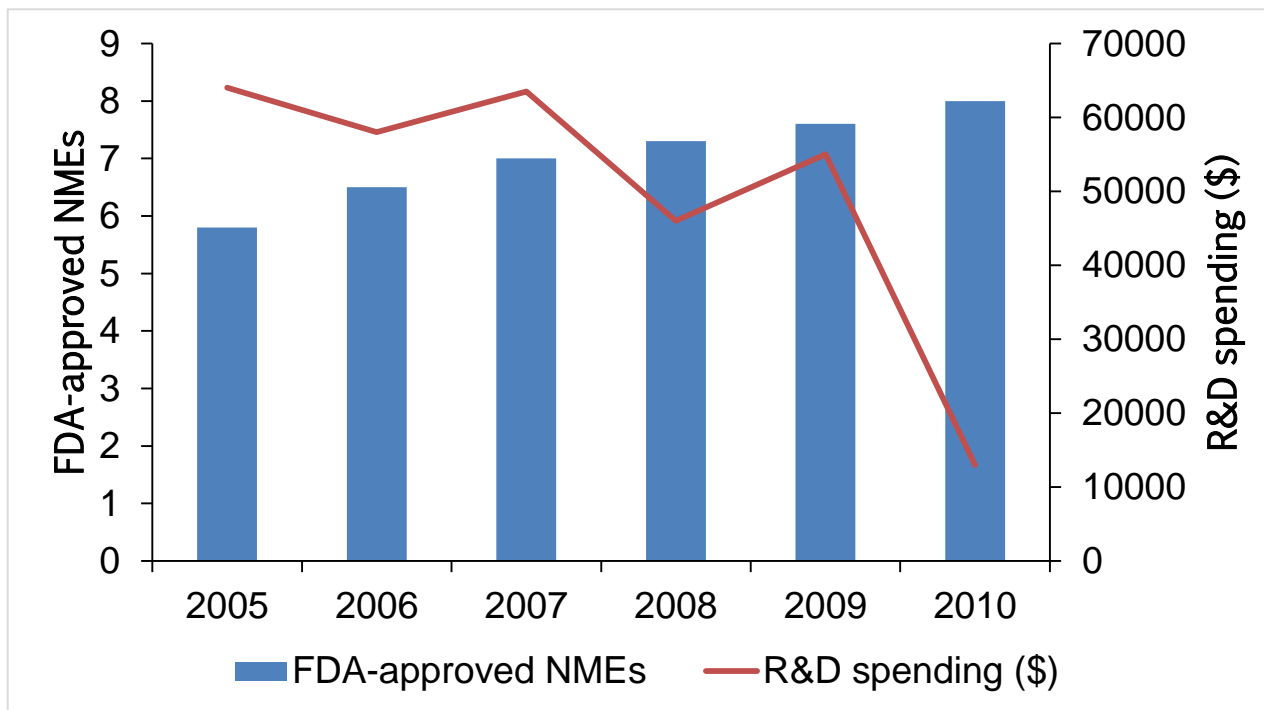
Drug approval rates by therapeutic area

Many factors to consider when considering drug design –

- Chemical properties
- Adsorption, distribution, metabolism and excretion
- Pharmacokinetic profile
- Toxicity profile
- Off targets affects
- Target binding efficacy

Attrition is high

Summary of large pharma productivity between 2005-2010



Current attrition rates are unacceptably high
Time for a change in tactic

Industry is moving to Open Innovation strategy

- Looking for new basic science in academia
- Why?
 - Access ‘blue-sky’ research
 - New assays, targets, starting points...
- Carrot?
 - Academic gets access to industry facilities and know-how



OK, fine.

But what's the problem in Oxford?

Challenges for University Researchers

- **Good science**

i.e. data management

- **Efficiency**

duplication of effort

- **Fulfilling Grant Conditions**

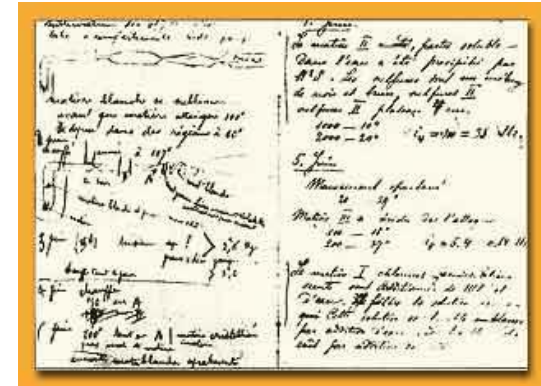
open data

- **Personal progression**

finding effective collaborators

- **Outcomes**

partnerships with industry to bridge the gap



wellcome trust

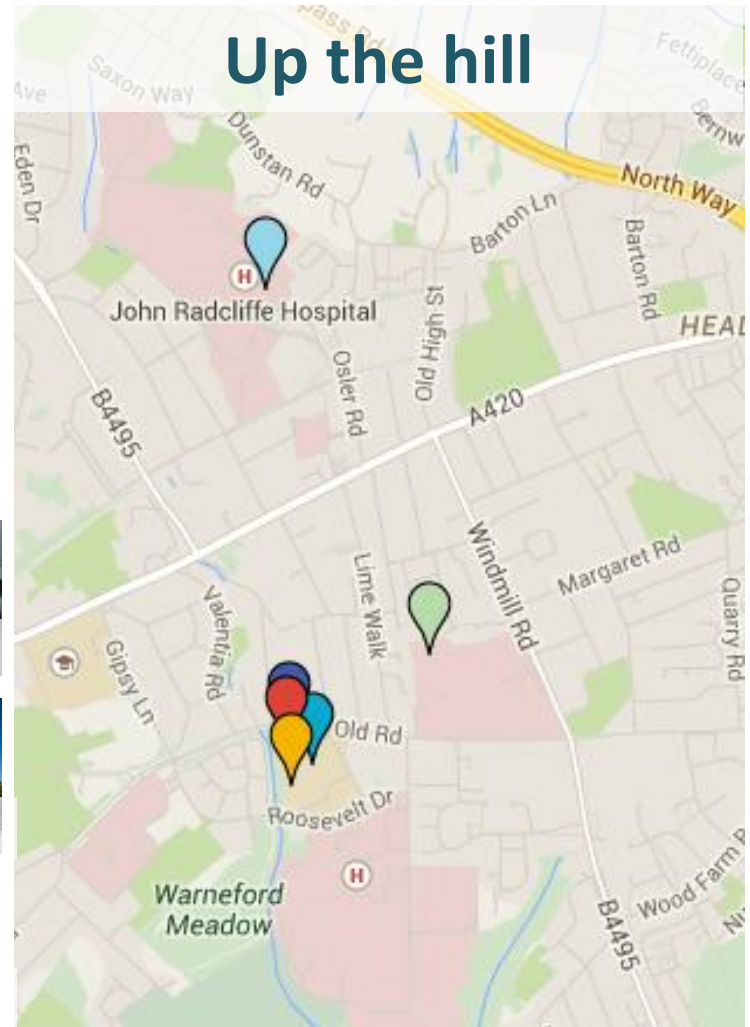


What sparked this all off?

“It is currently very difficult for one research group within the University to know if another group has expertise in a certain assay or screening technique. There is also no mechanism for one group to know if another group has compounds of interest for either target discovery or lead discovery.”

NDM/ISSF University-wide chemical biology audit - 2013

Chemical biology across Oxford



ChemBio

We approached them about:

- What they were doing
- What their goals were
- Whether they were open to collaboration

Lots don't use a computer, paper and pen and pencil

Sharing was not valued
Searching and being



"OF COURSE YOU CAN'T REPLICATE MY EXPERIMENT. THERE'S A SECRET INCANTATION THAT YOU HAVE TO CHANT, AND I'M NOT TELLING IT TO ANYONE."

say

asked them

(ing)

currently did



the benefit
ies

Hasn't this already been fixed?

Sure, there are lots of tools available.



We initiated what became known as ChemBio Hub

In response to the audit findings 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.

The solution – ChemBio Hub

Capture data, reagents/compounds, expertise –

- With assistance and curation
- In a central repository

Controlled levels of access –

- Within group, department, Oxford or externally

Example outcomes:

- The ‘go-to’ location for all aspects of University Chemical Biology
- Ability to discover tool compounds against target/protein of interest
- Identification of possible translational routes
- Pushing data externally, attracting pharma funding towards novel targets

To provide:

Better **reproducibility**

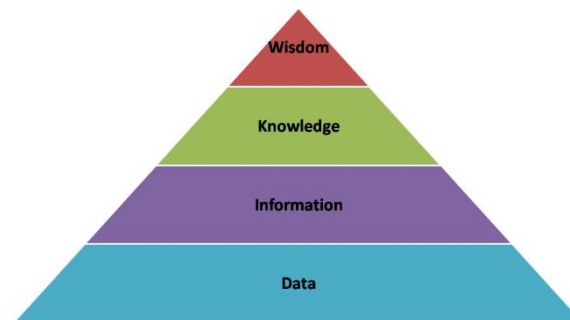


“..though the testimony of a single witness shall not suffice to prove the accused party guilty of murder; yet the testimony of two witnesses ...shall ordinarily suffice to prove a man guilty”

More **discoverable**: easily find out *who* else has worked on similar compounds, what effort is going into similar *targets*, what similar *techniques* or *equipment* are in use.

Better **tools**: to visualise data and aid interpretation

- Saves time in interpretation
- Makes troubleshooting much easier
- Clarifies communication



So, we brought together:

A project manager:

Web developers:

A Knowledge Exchange coordinator:

And a data scientist:



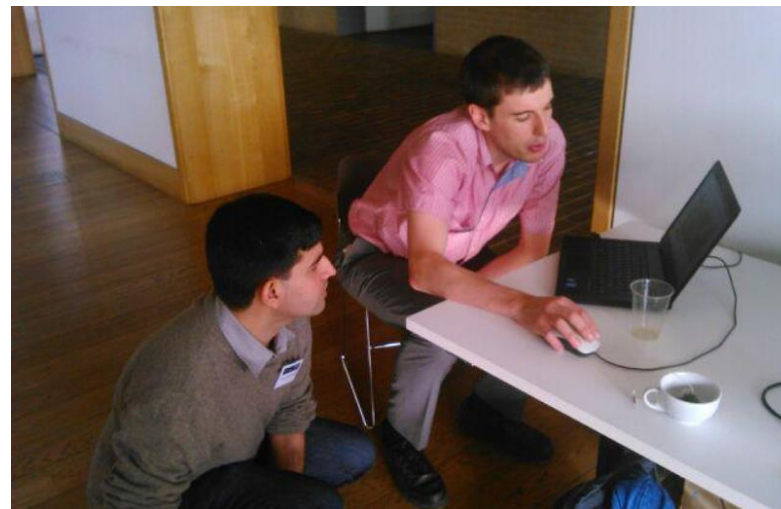
to tackle the problem.

The current funding will support the project until June 2016.

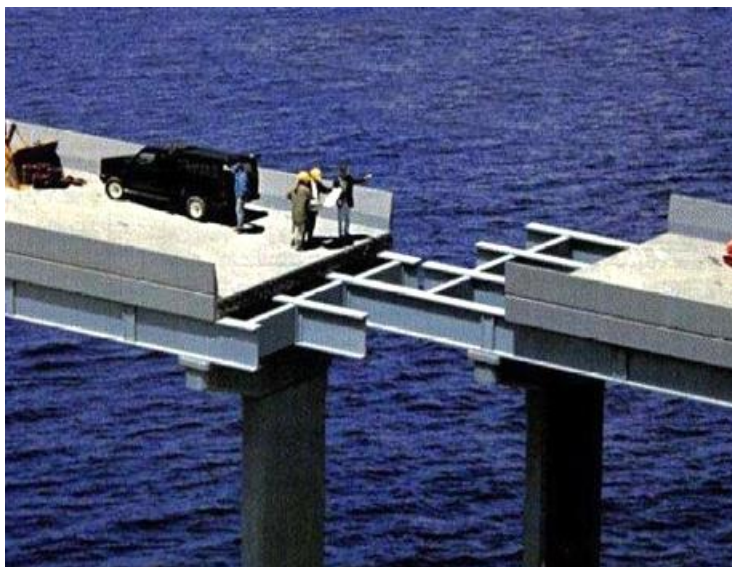
How have we tackled it?



One bite at a time



Getting value to users



Rather than risk this:

- Months of investigation and analysis
- Designing ‘the’ solution
- Unveiling to users – only to discover we’ve missed the mark

We deliver functionality every 2 weeks:

- With a named individual in mind – fixing a specific problem
- Getting feedback straight away
- Without fear of negative findings – all results are of interest and influence further actions

Enough background.

Show us some substance!

Just a little bit more context

The first utility we built was ChemBio Crunch.

We took an existing workflow:

And we improved on it:

- Simplified the data flow
- Improved speed of processing
- Fixed quality issues along the way
- And made visualisation and sharing easier

We came up with - ChemBio Crunch

1) Upload raw data

2) Validate plates for systematic errors

3) Calculate IC_{50} for multiple plates

Title*

BMG output file*

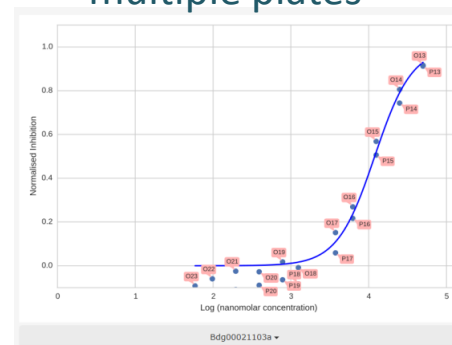
No file selected.

ESXX transfer file*

No file selected.




Uploaded meta file

No file selected.



Mark as poor fit

4) Mark poor fits where observed

	A	B	C	D	E	F
1	plat:	compound:	logIC50	ic50 (nM)	system_comments	use: graph
2	2	BDG00021 060a	4.25927396	18166.6127	Low total inhibition, values could be inaccurate	no 
3	2	BDG00021 344a	4.8345796	68324.9943	Low total inhibition, values could be inaccurate	no 
4	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 IC_{50} may be inaccurate)

Results

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

“I was just looking at the concentration discrepancies you mentioned and it looks to me like there has been copy and paste error or the template is out of date... If this is the case then it is the kind of thing that would be eliminated by using the new utility!”

ChemBio Hub Crunch

🏠 ChemBio Crunch - The assay data analyser

👤 Paul [Log Out](#)

◀ Prev Plate

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

Next Plate ▶

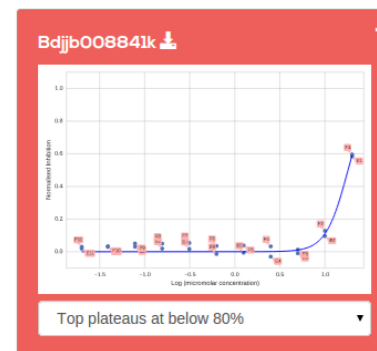
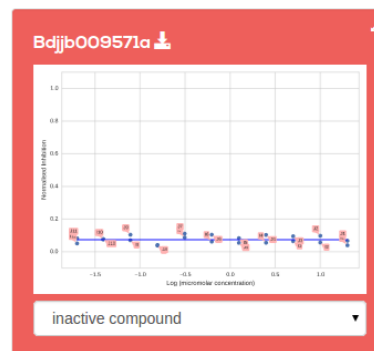
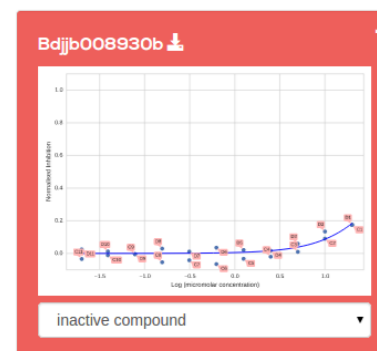
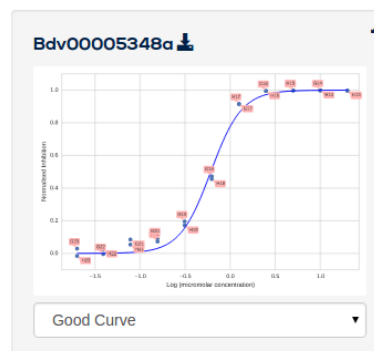
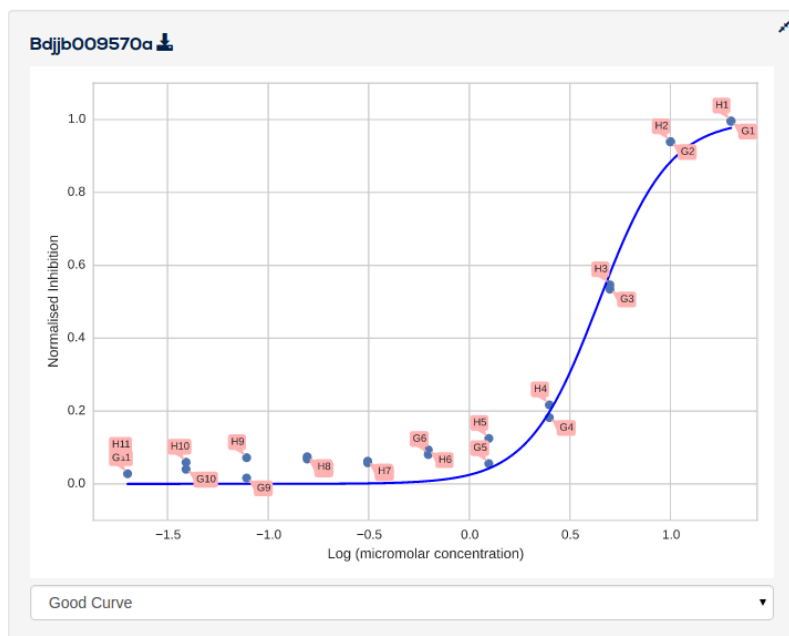
Export ▾

- 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	791k	790k	841k	852k	920k



The technology stack

What you see



HTML

CSS



Bootstrap



What that runs on



WebAuth



What we use to deploy and test it

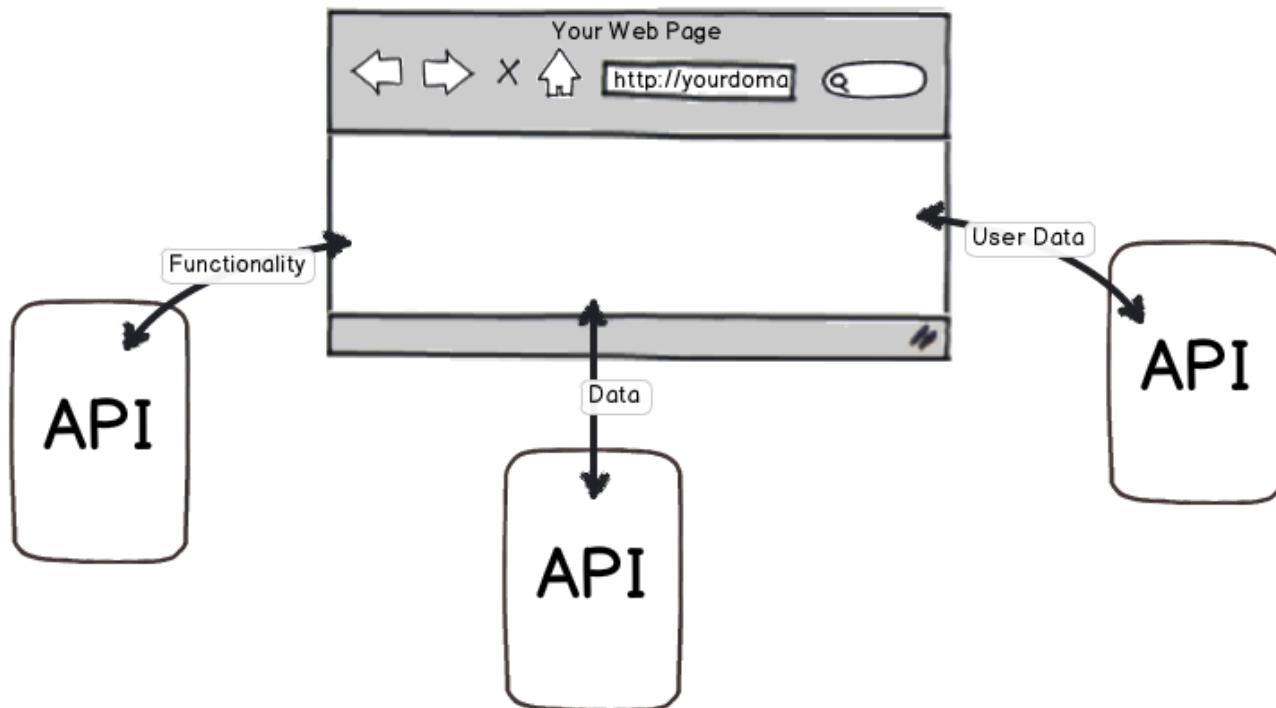


behave



Modular design for resilience and data sharing

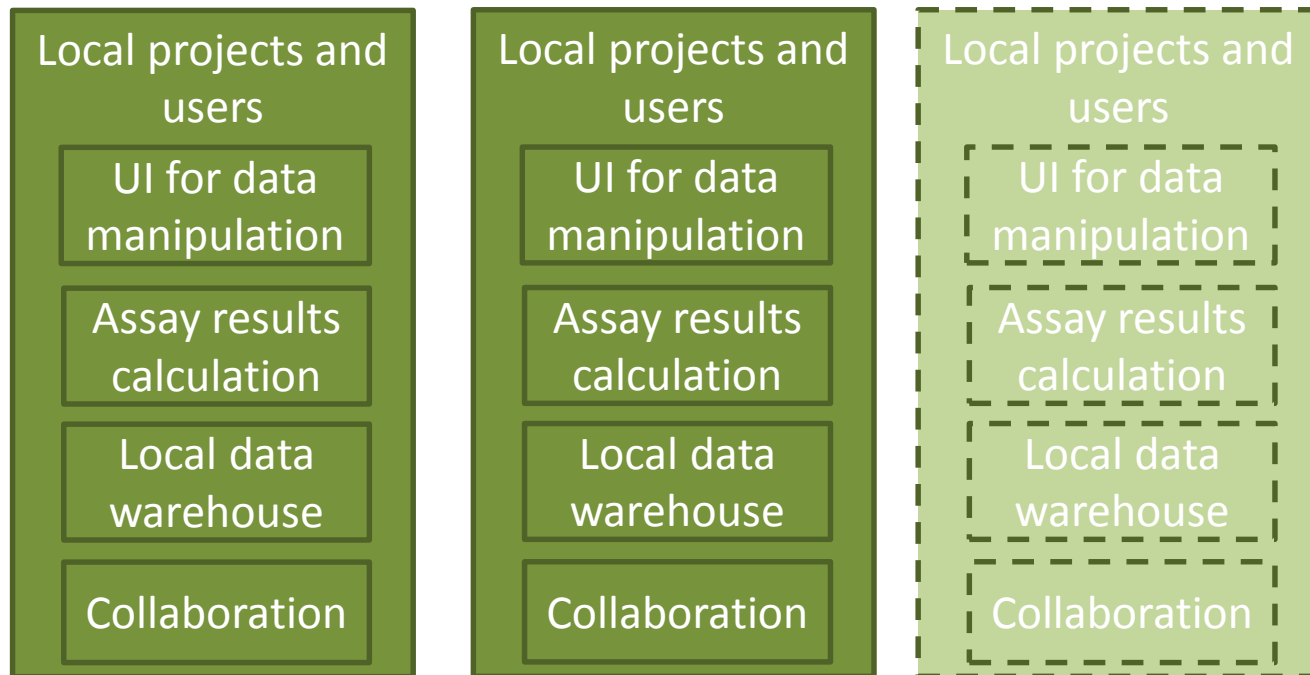
- System should “talk” to other installations
- Modules should be self-contained



High level architecture

Silos for groups who want to collaborate together

Instance level services



Global services that all can connect to

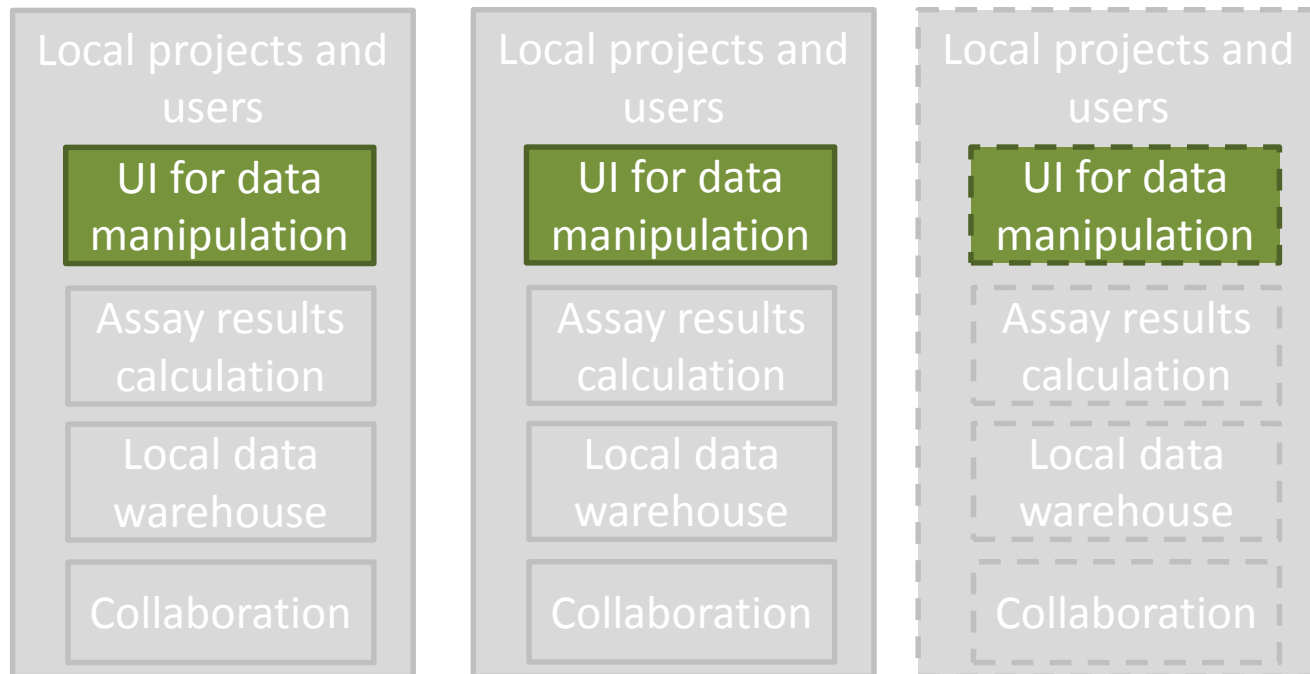
Global Services



Focus on Web Front End Technology

Silos for groups who want to collaborate together

Instance level services



Global services that all can connect to

Global Services



Websites vs apps = different front end

- Jekyll static site for website and blog
 - Easy deployment
 - SEO friendly
 - No database or security issues
 - Easy deployment



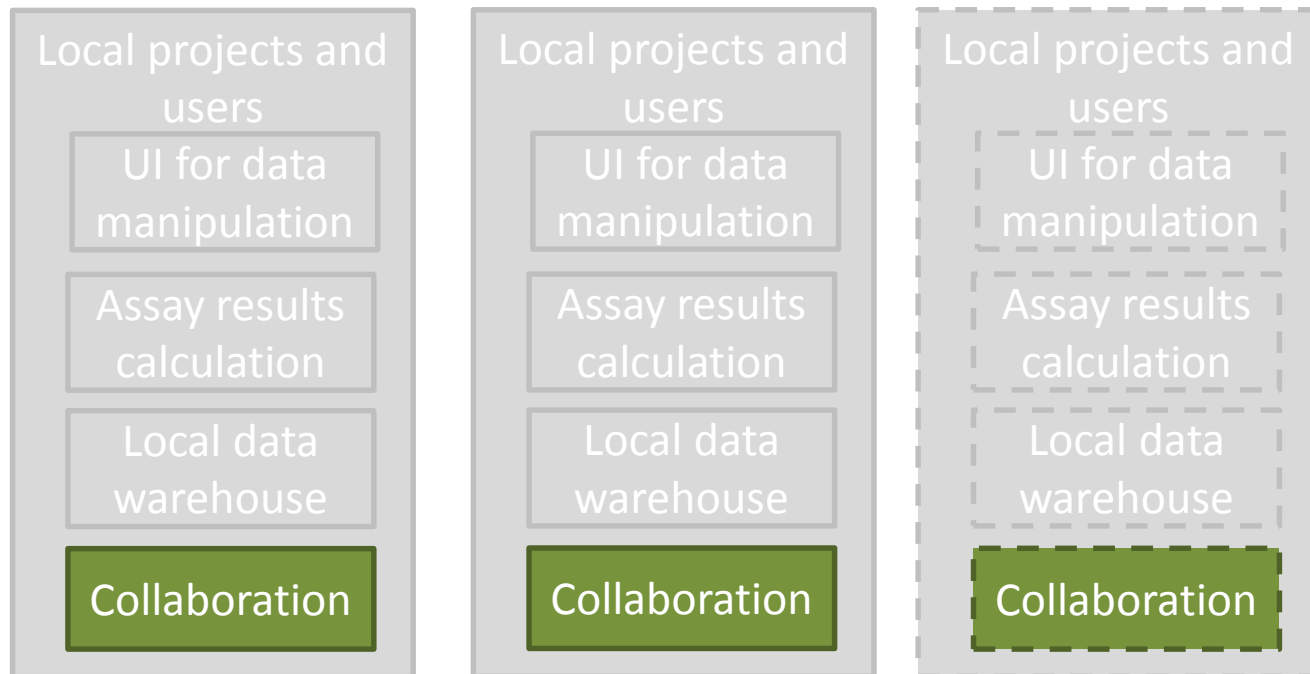
- AngularJS for apps
 - Easy-to-follow coding standards for javascript
 - Slick and customisable user experience without code repetition
 - Works with our API which we must develop anyway



Focus on Collaboration

Silos for groups who want to collaborate together

Instance level services



Global services that all can connect to

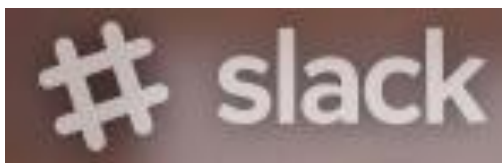
Global Services



Existing tool for user feedback and Q and A

- Many tools purport to help people collaborate
 - Success can depend on the people and the situation more than the tool
- We chose a simple team messaging tool

slack.com

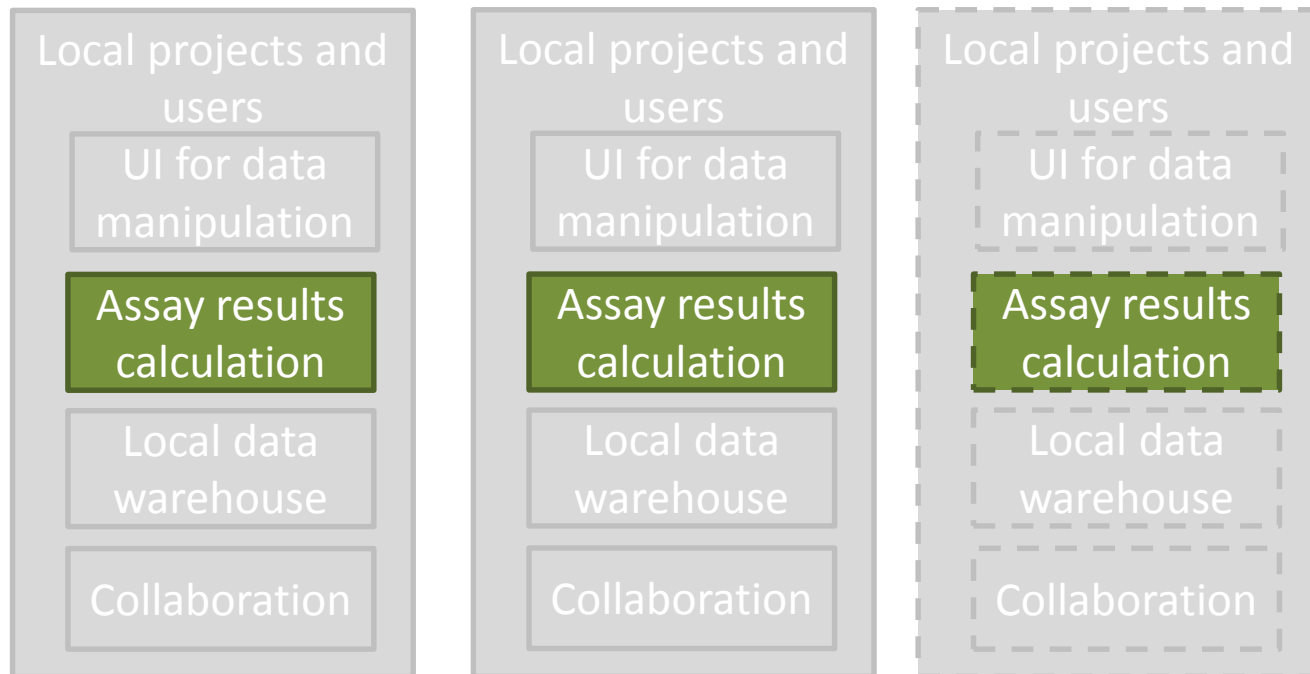


- Integrated feedback button
- Slack users will be able to message the team and ask each other questions

Focus on Calculations and Data Editing

Silos for groups who want to collaborate together

Instance level services



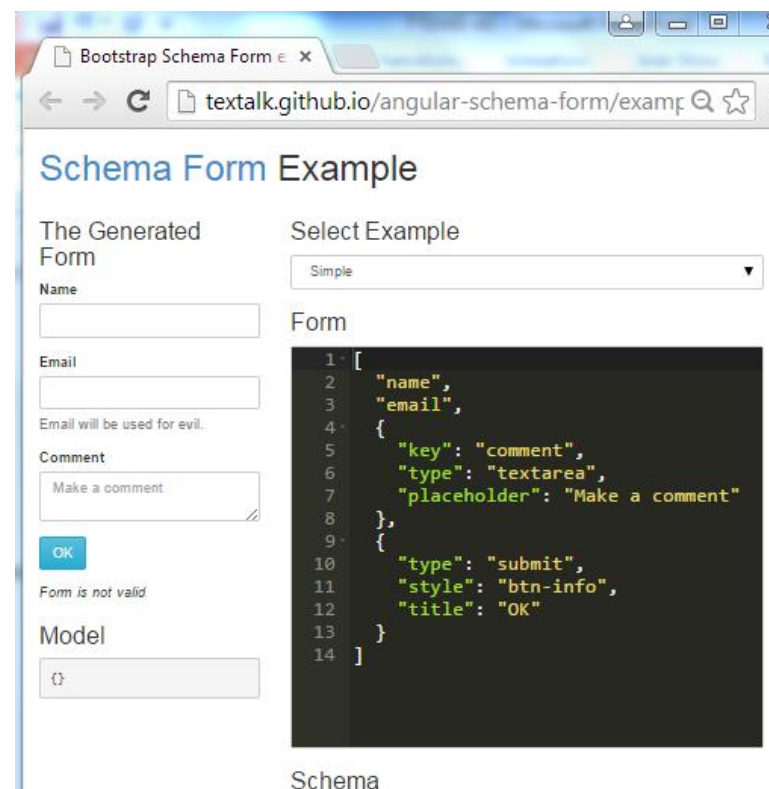
Global services that all can connect to

Global Services



ChemBio Crunch and future curation apps

- Work with raw data and custom fields
 - Build forms in browser with angularjs and JSON Schema
- Open source calculation software
 - Python Pandas to import
 - Python LMFit for curve fitting
 - Matplotlib for graphing
 - RDKit and Openbabel for chemistry calculations



Bootstrap Schema Form

textalk.github.io/angular-schema-form/example

Schema Form Example

The Generated Form

Select Example

Simple

Name

Email

Email will be used for evil.

Comment

Make a comment

OK

Form is not valid

Model

```

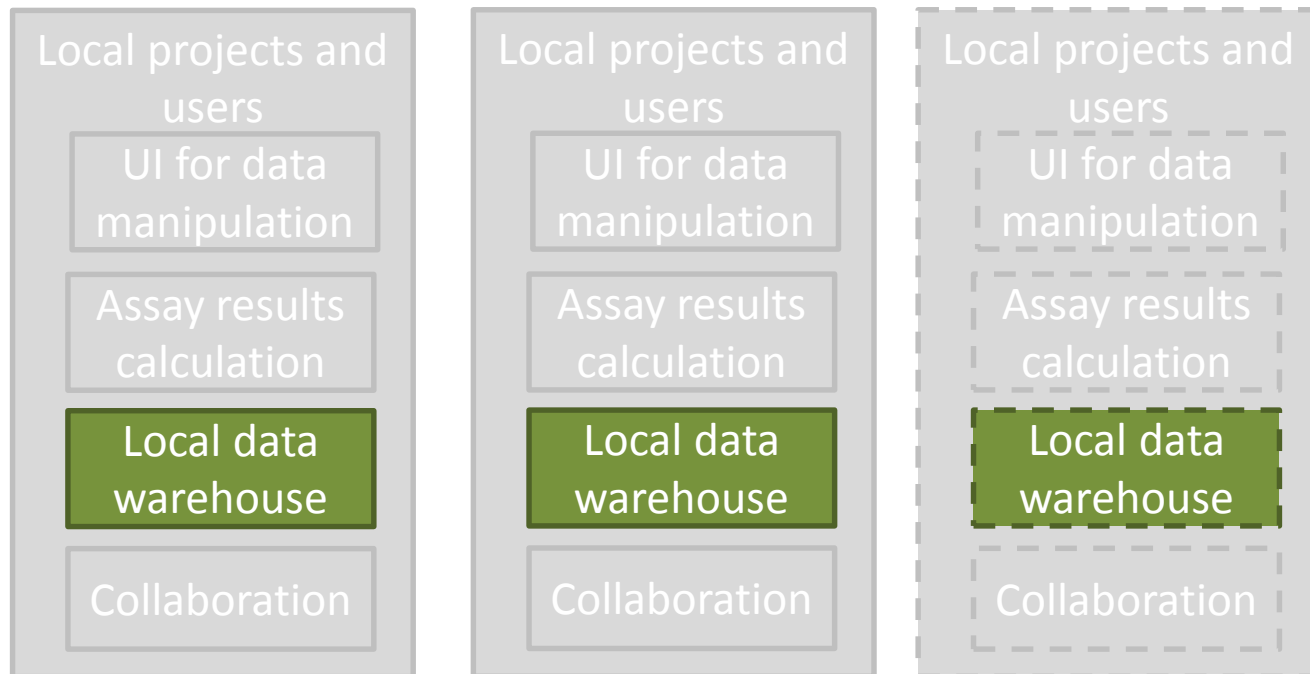
1 [
2   "name",
3   "email",
4   {
5     "key": "comment",
6     "type": "textarea",
7     "placeholder": "Make a comment"
8   },
9   {
10    "type": "submit",
11    "style": "btn-info",
12    "title": "OK"
13  }
14 ]
  
```

Schema

Focus on Data Storage and Transfer Formats

Silos for groups who want to collaborate together

Instance level services



Global services that all can connect to

Global Services



Backend Development and Databases

- PostgreSQL + HStore + RDKit extensions
 - Custom fields allowed
 - No custom nested objects
 - Too difficult to curate
- ChEMBL DB access code
 - Allows us to build the API we need
 - Avoids creating a new schema
 - Uses our preferred framework Django



Data Transfer and Deposition

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



- Where deposition APIs do not exist e.g. ChEMBL– we help build them
- Long term archiving in <https://databank.ora.ox.ac.uk/>
 - Single records will be packaged with files + json schema.
 - Private projects will be embargoed



ChemBio Hub

ChemBio Hub - Open Source is key

We are open source on Github

github.com/thesgc

GitHub

Important for reproducible, open science

Important for open collaboration

Important for open publications

Important for distributed deployment model

(virtual machines, less config to get software working)

Great!

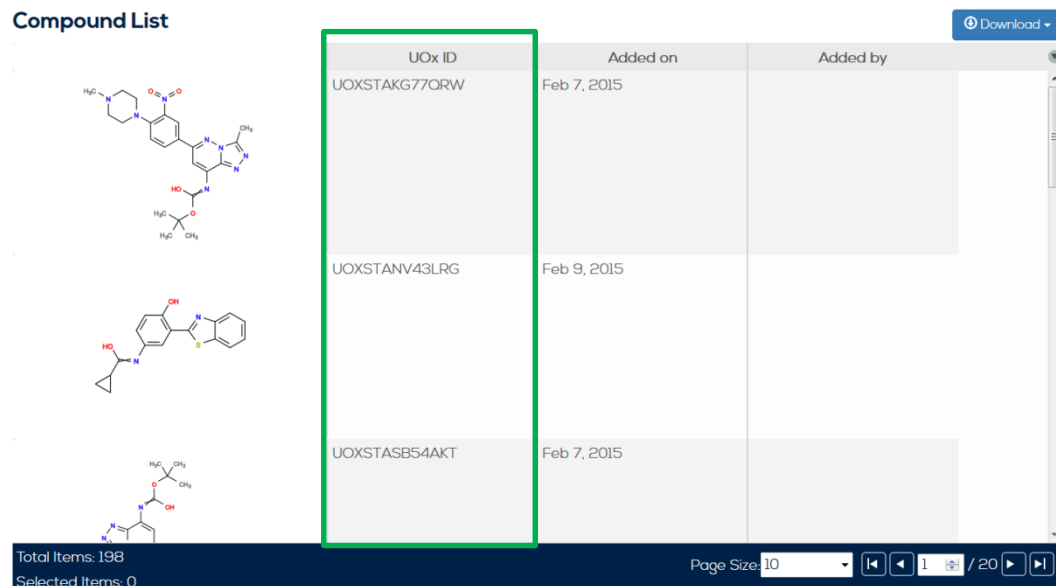
What else have you done?

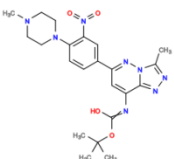
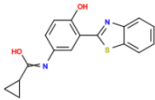

ChemBio Hub ChemReg

Compounds to be registered need to be uniquely identified. Users were concerned about sequential assignment inadvertently allowing others to deduce who had registered private compounds.

So we generate UOXIDs, in a number plate format:

Compound List



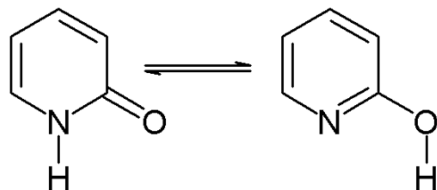
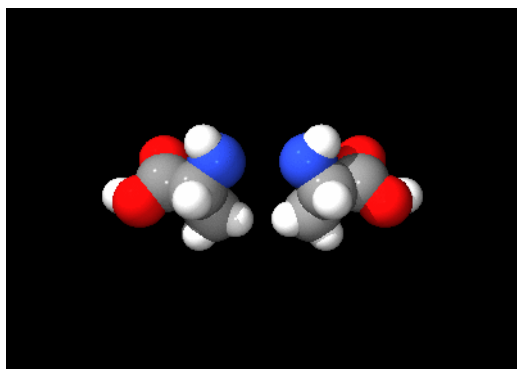
	UOX ID	Added on	Added by
	UOXSTAKG77QRW	Feb 7, 2015	
	UOXSTANV43LRG	Feb 9, 2015	
	UOXSTASB54AKT	Feb 7, 2015	

Total Items: 198
Selected Items: 0
Page Size: 10 / 20

Randomly generated, then checked to ensure they are unique.

The format UOX aa nn bbb will allow over 1bn combinations of relatively simple to remember identifiers

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
- 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.

Registration process

⚡ Wizard

Welcome to the ChemReg wizard

home



To upload multiple compounds you will need a list of SMILES; an InChi key or list of InChi keys; or either a SD or Excel file.



Add Single Compound



Add Multiple Compounds

Chemical input wizard walks chemists through the process of registration

Registration process

Wizard
Welcome to the ChemReg wizard ?

home add finish

1 of 2

Draw a molecule here
Use the sketcher window to draw your compound. Sketcher window provided by ChemDoodle.

Properties

- alogp: 0.00
- H-bond acceptors: 0
- H-bond donors: 0
- PSA: 0.00
- Rotatable bonds: 0

Stereochem options

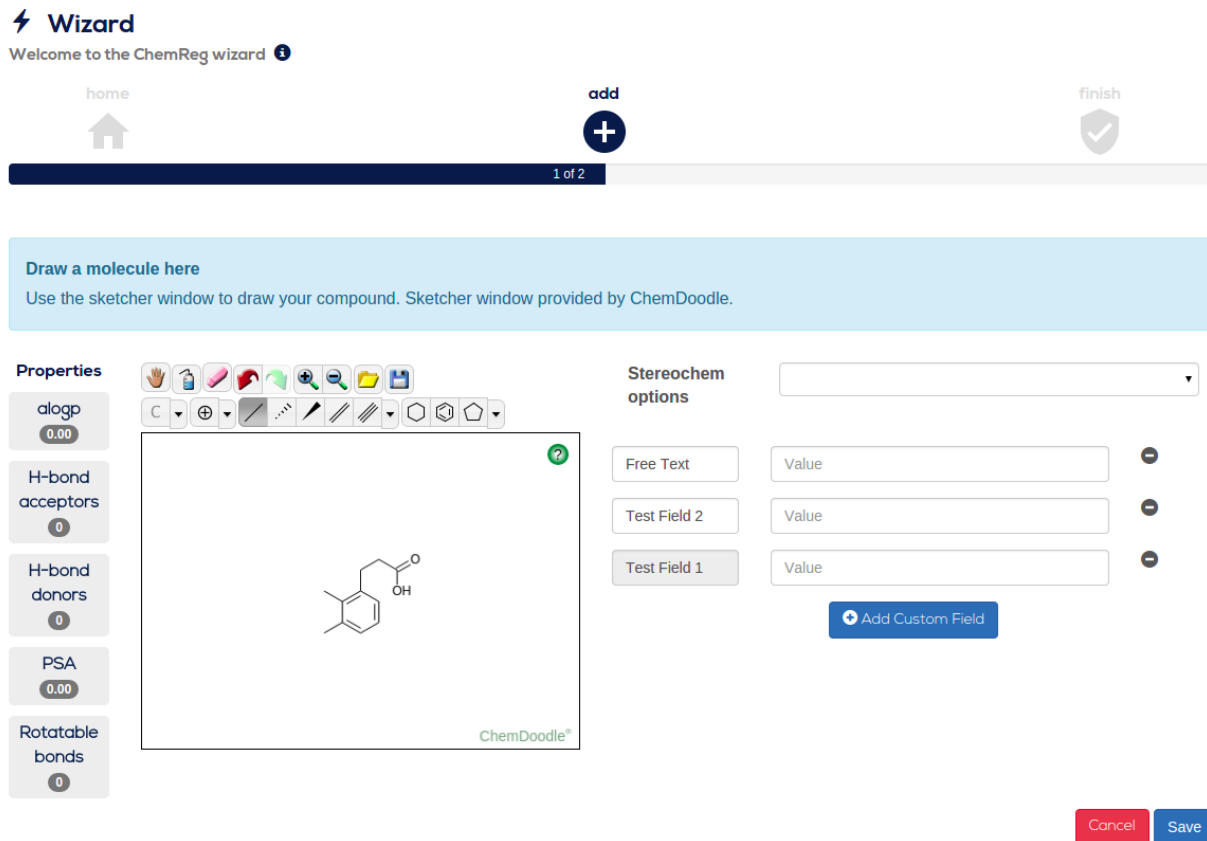
Free Text: Value

Test Field 2: Value

Test Field 1: Value

[Add Custom Field](#)

Cancel Save



Users can sketch their molecule.

Live calculation of chemical properties gives feedback about the nature of the structure.

Users can add extra custom data to their registration

Registration process

⚡ Wizard

Welcome to the ChemReg wizard ?

home



add



map



validate



finish



1 of 4

Add your compounds to be registered.

Paste a list of SMILES, Standard InChis

Paste your SMILES or IDs here

Select which type of Ids you have:

Auto-detect

Process Ids

...alternatively upload SD, Chemdraw or Excel file.

Upload your SD, Chemdraw or Excel files here

Upload File

Files added:

No files uploaded

Or Drag And Drop your file here

Cancel

Process File

Users can add SMILES strings or InChi keys (text-based chemical structure representation) or in standard format files (SDF, ChemDraw or Excel)

Registration process

Paste your SMILES or IDs here

```
Cn1cnc2c1c(=O)n(c(=O)n2C)C  
C(CS(=O)(=O)O)N
```

Select which type of IDs you have:

Auto-detect

2 substances were processed of which:

2 substances are not publically registered in ChemBio Hub ChemReg. Initial batches of these substances will be registered in this project for each of these.

No substances were already been registered in this project.

No substances have already been registered as public.

Process IDs

Active validation at every step

Information about what is currently in the database - overlaps

Problems and issues can be identified early in the process

Registration process

Wizard
Welcome to the ChemReg wizard

home add ✓ map

2 of 4

Add extra details to the list of compounds you have registered

M | Value

- Compound SMILES
- Valid SMILES
- Smile definition
- Smiles label
- Formula
- Mass_mg
- Mol_weight

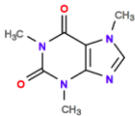
Wizard
Welcome to the ChemReg wizard

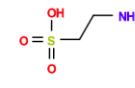
home add ✓ map ✓ validate ✓ finish ✓

4 of 4

Your prepared compound data is listed below. Click cancel to delete this molecule and start again.

UOx ID	Added on	Added by
UOXSTADA55WEK	Feb 16, 2015	
UOXSTAGV24ERP	Feb 16, 2015	





Users can add extra custom data relevant to their project
 Individual projects can make certain data fields mandatory
 Results can be exported as SD or Excel file for auditing etc.

But it isn't just about whizzy tools...

There's some cool technology involved...

But they don't and won't define a successful project.

The aim is to:

- Change **how** researchers work for the better.
- So let's talk about **Knowledge Exchange...**
 - Working with Business Development, ISIS, the KE community and Pharma/Biotech



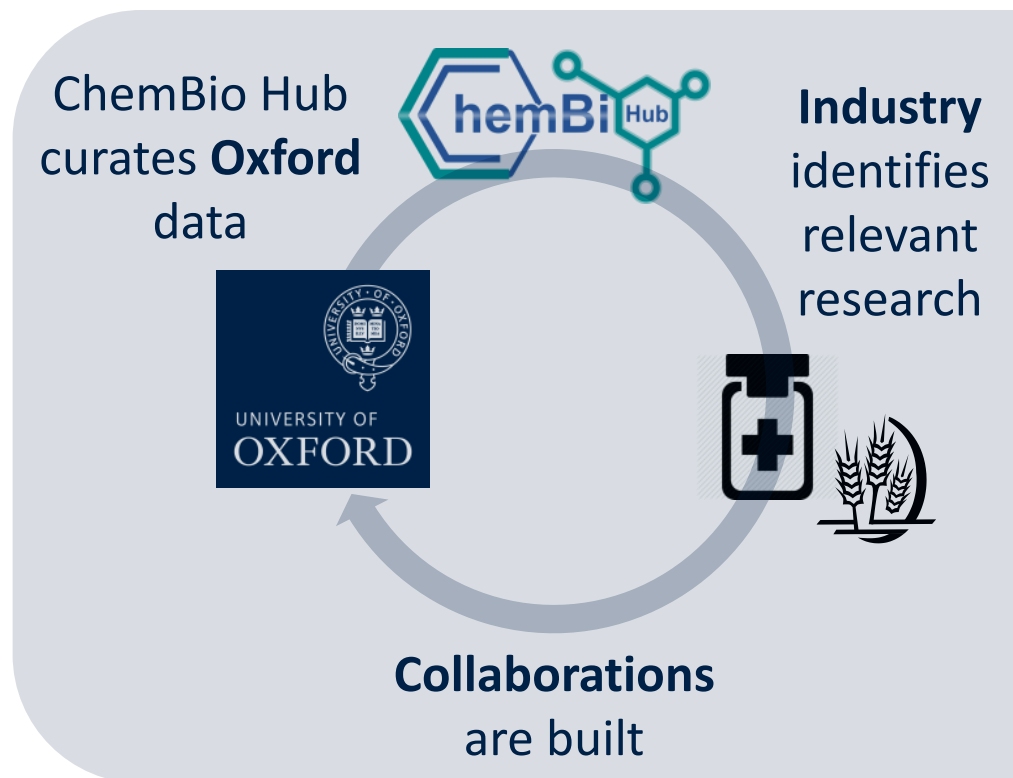
People *as well as* data

- Great data **management** is just the beginning...
 - Ultimately it is **people** we need to connect
 - So that existing data can be turned into new ideas



Industry outreach

- Critical to engage pharma and biotech companies now
- Learn industry-grade data management
- Find out how to best present data to industry
- Obtain industry buy-in to sustain ChemBio Hub
- Chicken & Egg: challenge of conveying vision early-on



People *as well as* data

- ChemBio Hub **Oxford** symposium – November 2014
- 100 Oxford researchers met and made new cross-departmental connections
- Generated 17 new collaboration ideas
- Live demos of ChemBio Hub data management tools



People *as well as* data



- ChemBio Hub **Oxford-industry** symposium – July 2015
- Showcase Oxford capability to pharma and biotech groups
- Spark connections, conversations and collaborations
- Drive further industry funding of Oxford research

What's next?

Right now

We're recruiting pilot testers from multiple departments to prove the value of ChemBio Reg

- Chemistry
- DPAG
- Biochem
- WIMM
- Physics
- SGC
- TDI
- Pharmacology

And we're thinking about how to tackle Assay Capture

What happens next?

- We have a year to develop code:
 - Registration of compounds
 - Assay definition
 - Assay results
- Keeping usability as a high priority
- Integrating with other systems
- Promoting collaboration and good data management
- Building the community

And then..

- This is a permanent fixture, so it needs a support and growth plan. Multiple options:
 - Professional open source
 - Industry angels
 - Licence
 - Public funding
 - Other flavours and combinations
- We will explore these possibilities and more...

Key Lessons

- You must fully understand the nuances of the data and its metadata before you dive in
- Engaging with users early is key – a technical solution that no one uses is a waste of time
- Keep your technology stack nimble, but don't box yourself into a corner with cool toys
- Cultural change is hard (no surprise!)
- Inspiration can come from very unlikely places
- You can build an awesome solution, but that does not guarantee sustainability

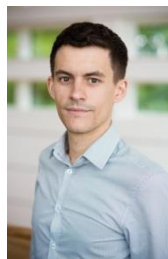
Acknowledgements

<http://chembiohub.ox.ac.uk>

The ChemBio Hub Team



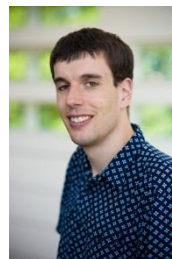
Karen



Michael



Adam



Andy



Paul

The SGC

Heads of Division

Heads of Departments

University Chemical Biologists



Nuffield Department of Medicine
Medical Sciences Division

Supported by
wellcometrust

HIGHER EDUCATION
hefce
FUNDING COUNCIL
FOR ENGLAND

