Practical cheminformatics workflows with mobile apps

Dr. Alex M. Clark

October 2012

http://molmatinf.com

References

- Redefining Cheminformatics with Intuitive Collaborative Mobile Apps, A.M. Clark, S. Ekins, A.J. Williams, *Molecular Informatics*, **31**, 569-584 (2012)

- Mobile apps for chemistry in the world of drug discovery, A.J. Williams, S. Ekins, A.M. Clark, J.J. Jack, R.L. Apodaca, *Drug Discovery Today*, **16**, 928-939 (2011).

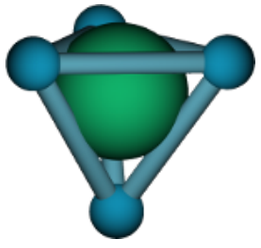
- Open Drug Discovery Teams: A Chemistry Mobile App for Collaboration, S. Ekins, A.M. Clark, A.J. Williams, *Molecular Informatics*, **31**, 585-597 (2012)

- Basic primitives for molecular diagram sketching, A.M. Clark, *Journal of Cheminformatics*, **2**:8 (2010).

Introduction

- Mobile apps in chemistry happening real fast: 3 years ago nothing but the most trivial of apps...
- ... now, there's a whole ecosystem. Complex workflows can be executed using just mobile apps & cloud-based infrastructure.
- Three core concepts:
  - **content consumption**
  - **content creation**
  - **communication and sharing**
- *Molecular Materials Informatics* founded with the goal of making mobile + cloud solutions into a viable replacement for cheminformatics infrastructure

info@molmatinf.com



App Functionality

Most chemistry apps concentrate on at least one of:

1) Information delivery
2) Catalog lookup
3) Data visualisation
4) Data creation
5) Lab notebook
6) Collaboration & sharing

For a detailed list, see **www.scimobileapps.com**

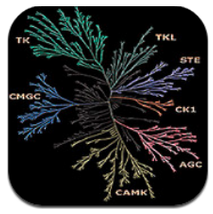
1. Information delivery

- Many apps designed for mainly one-way transmission of information, i.e. *content consumption*
  - **Green Solvents**
  - **Approved Drugs**
  - **The Elements**
  - **ACS Mobile, RSC Mobile**
  - **Reagents, Organic Named Reactions,**
- Typically a very simple user interface: close to zero learning curve
- Many for education, but some are relevant to professionals
- Mobile devices very well suited for these types of apps: quite easy to produce, and can be very useful, despite their limited capacity



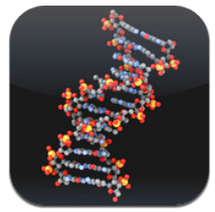

2. Catalog lookup

- Apps designed to input a search query, submit the query to a server, then display and utilise the results
- Examples include:
  - **ChemSpider Mobile**
  - **SPRESImobile**
  - **Mobile Reagents**
  - **iKinase**
  - **iProtein**
- These apps are moderately difficult to produce; most of the action takes place on remote servers, i.e. hosting the data and the search algorithms

3. Data visualisation

- Apps designed to visualise existing chemical data, e.g. PDB entries, docking results, structure-activity relationships and collections of molecules
- Examples include:
  - **Molecules**
  - **PyMol**
  - **SAR Table**
  - **iMolView**
  - **iSpartan**
  - **iMolecule Builder**





4. Data creation

- Some of the most sophisticated apps provide the ability to assemble chemical data within the app:
  - molecular structures
  - reactions
  - scalar data
  - higher order markup (SAR)
- Data can be created with the app UI, or pieces can be imported from other sources. Examples:
  - **Mobile Molecular DataSheet (MMDS)**
  - **MolPrime**
  - **SAR Table**
  - **Chirys, ChemDoodle Mobile, ChemJuice**




5. Lab notebooks

- Apps designed for entering information about experiments
- Ultimate objective is the paperless laboratory
- Mobile laboratory notebooks are mostly generic rather than chemically aware, i.e. not specifically for synthesis
- Prototypical example is **Yield101** (education focused)

- General purpose scientific lab notebook apps:
  - **LabGuru**
  - **irisnote**

- Many web-based lab notebooks that are vendor-endorsed for iPad use

6. Collaboration and sharing

- Apps have many options for sharing and importing data, e.g.
  - Email (personal, bidirectional)
  - Remote hosting (private, collaborative)
  - Web sharing and tweeting (global, open)
- Examples:
  - **MMDS**
  - **MolSync**
  - **Reaction101**
  - **ODDT**



Mobile app limitations

- Touchscreen is small, inaccurate, and doubles as a keyboard
- Modern mobile processors are fast (Moore's law), but respecting battery life requires careful conservation of CPU resources
- Limited memory and constricted multi-tasking: an app has a tightly defined lifespan, and is expected to be lightweight, low-state
- Persistent storage is limited and relatively inconvenient to access
- Functionality and making use of data is heavily network dependent
- Limited choice of development environments: *native* or *web*, stark choice of quality vs. portability

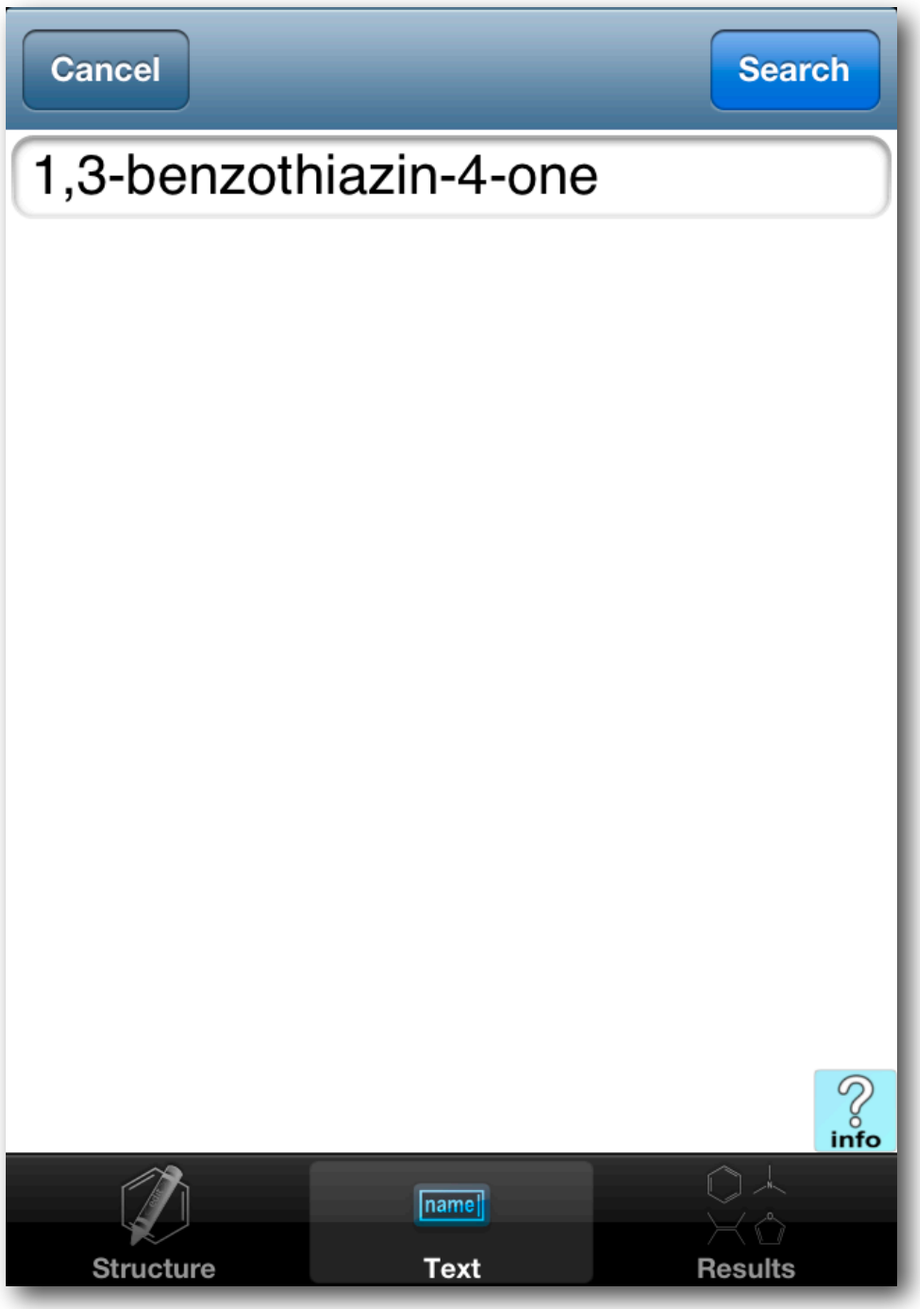
- When these limitations can be overcome, the mobile + cloud combination becomes a *full replacement for desktop computing*

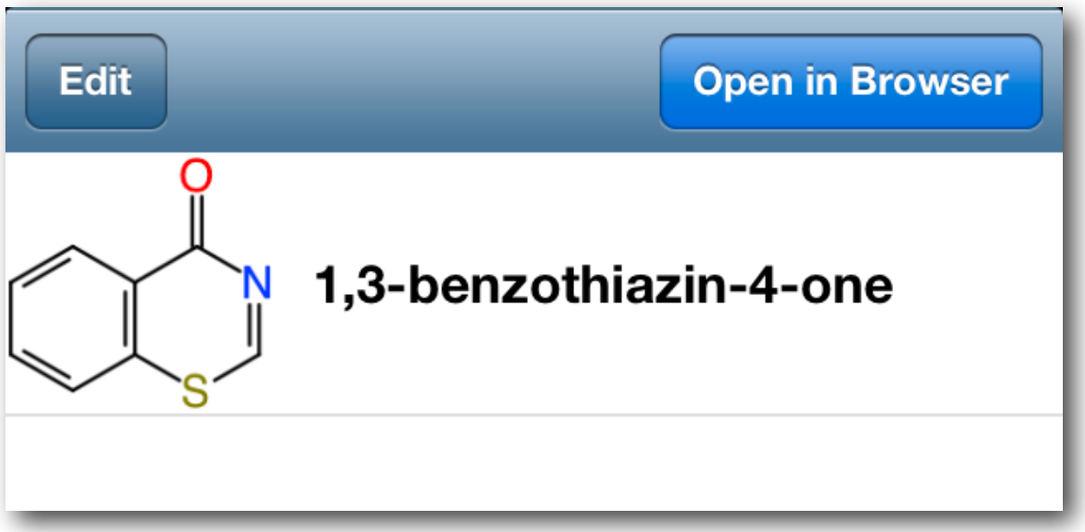
Workflow

- The following slides demonstrate a multipart workflow scenario using:
  - **iOS** apps (Apple iPhone/iPod/iPad)
  - cloud-hosted webservices
- Each step is carried out using technology that is currently available...
- ... and is representative of the state of the art.
- The sequential tasks are based on a medicinal chemistry exploration: investigating new *tuberculosis* drugs
- Any workflow with a similar level of technology is probably either already possible, or could be made possible.

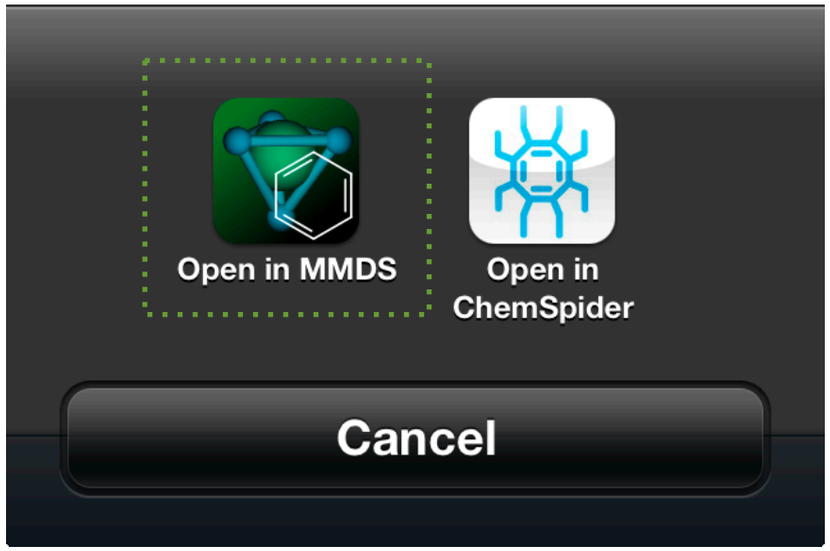
Step 1: Lookup

- Lookup a core scaffold for a series of known *tuberculosis* inhibitors using the *ChemSpider Mobile* app:





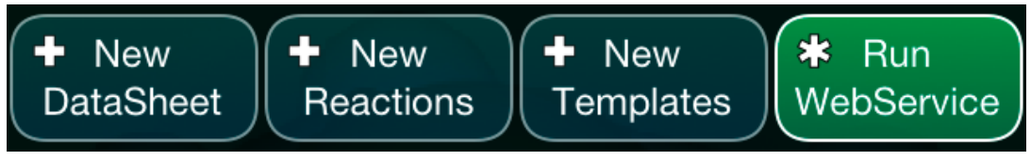
ChemSpider



- View *ChemSpider* record in mobile browser
- Structure can be downloaded as MDL Molfile
- Mofiles can be opened *directly in apps*
- Select *Mobile Molecular DataSheet* (**MMDS**)

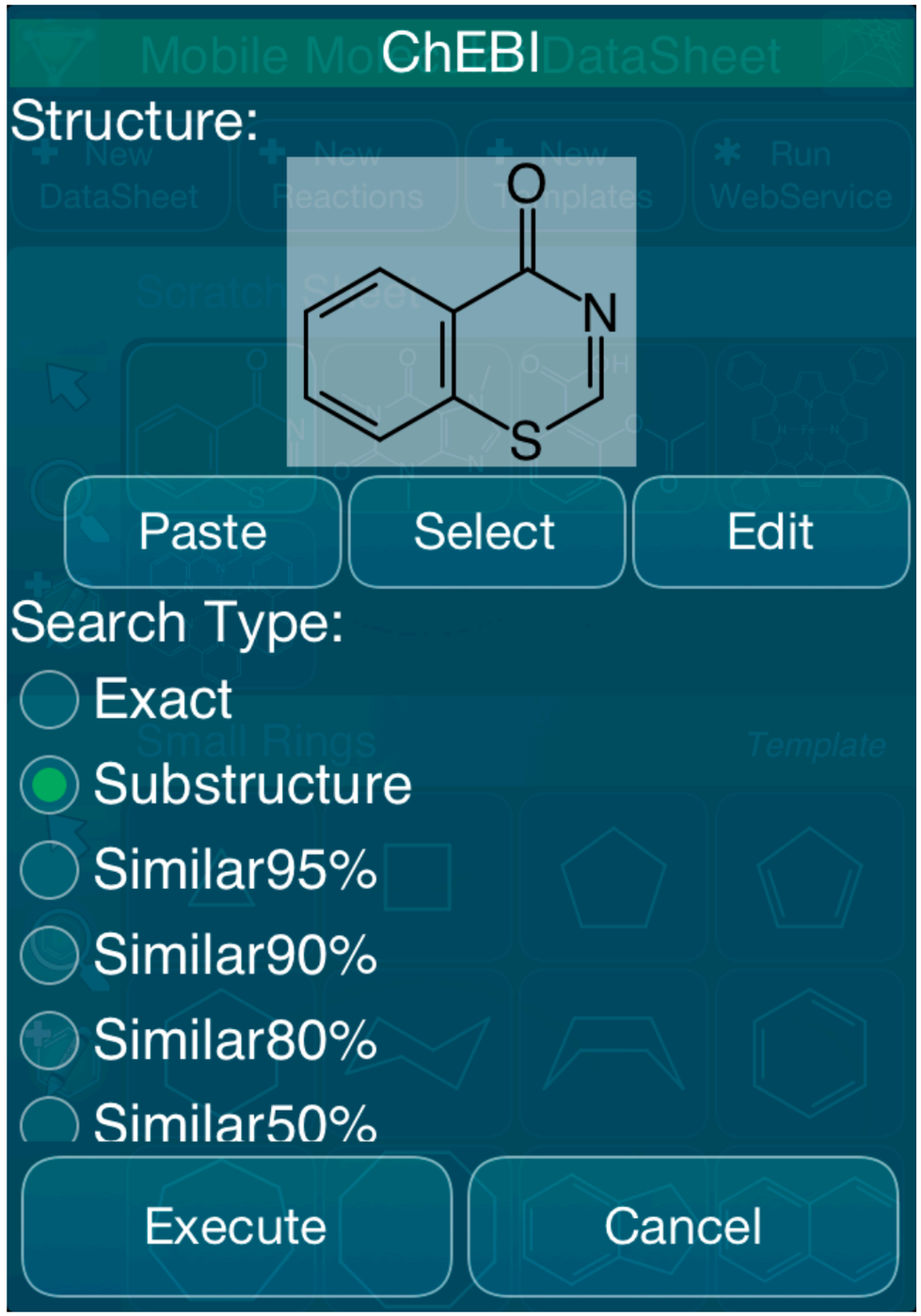
Step 2: Lookup a datasheet

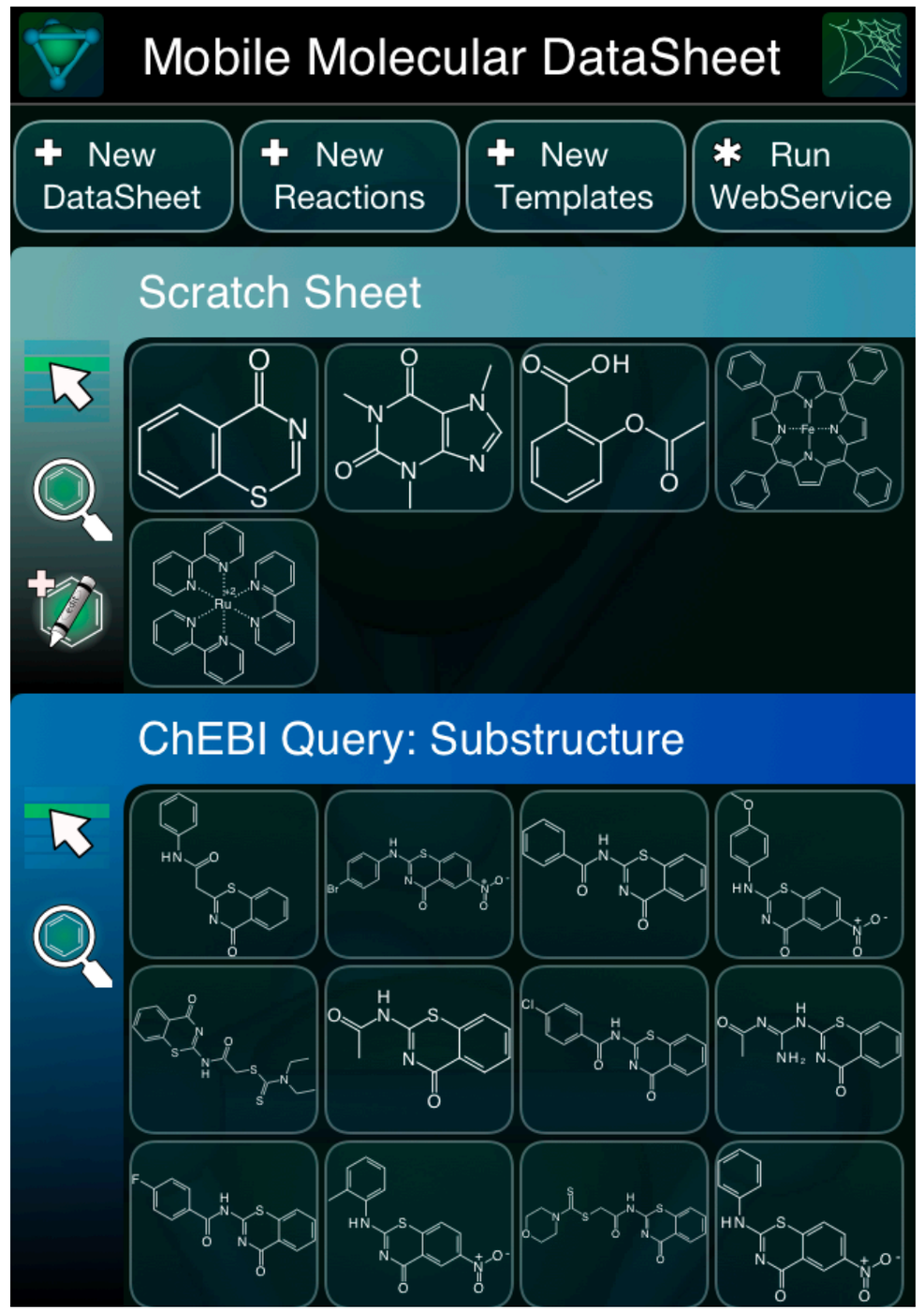
- Imported molecular structure into the *scratch sheet*



- Open the *Run WebService* feature...

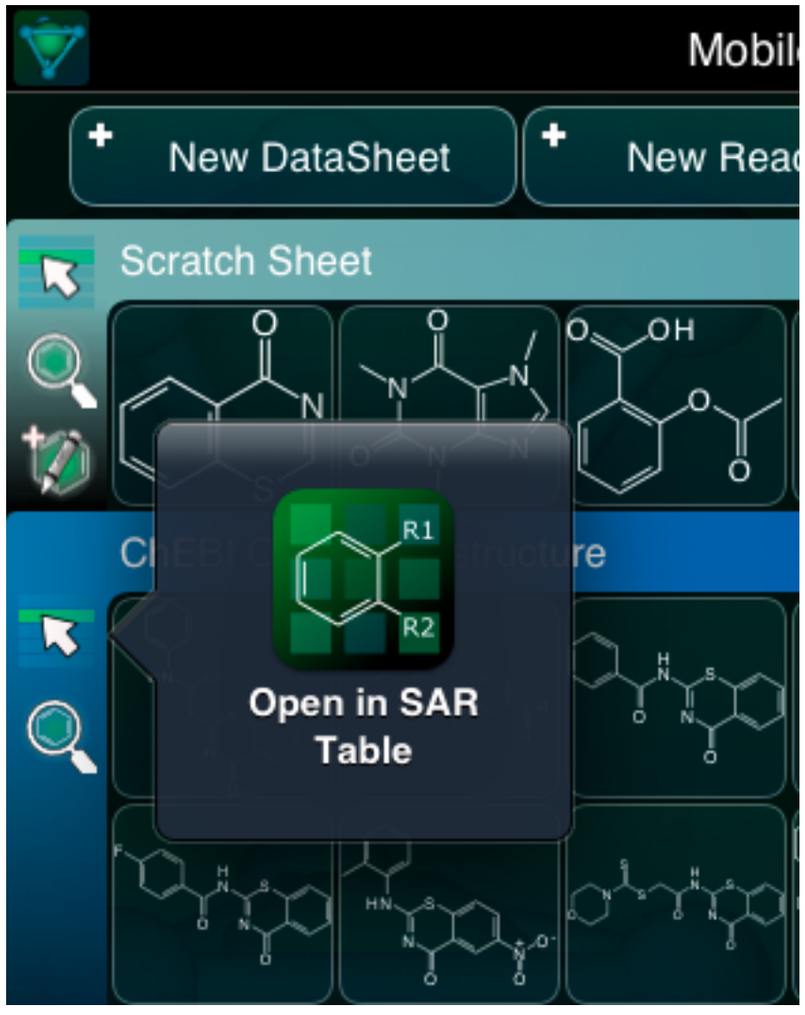
ChEBI Query

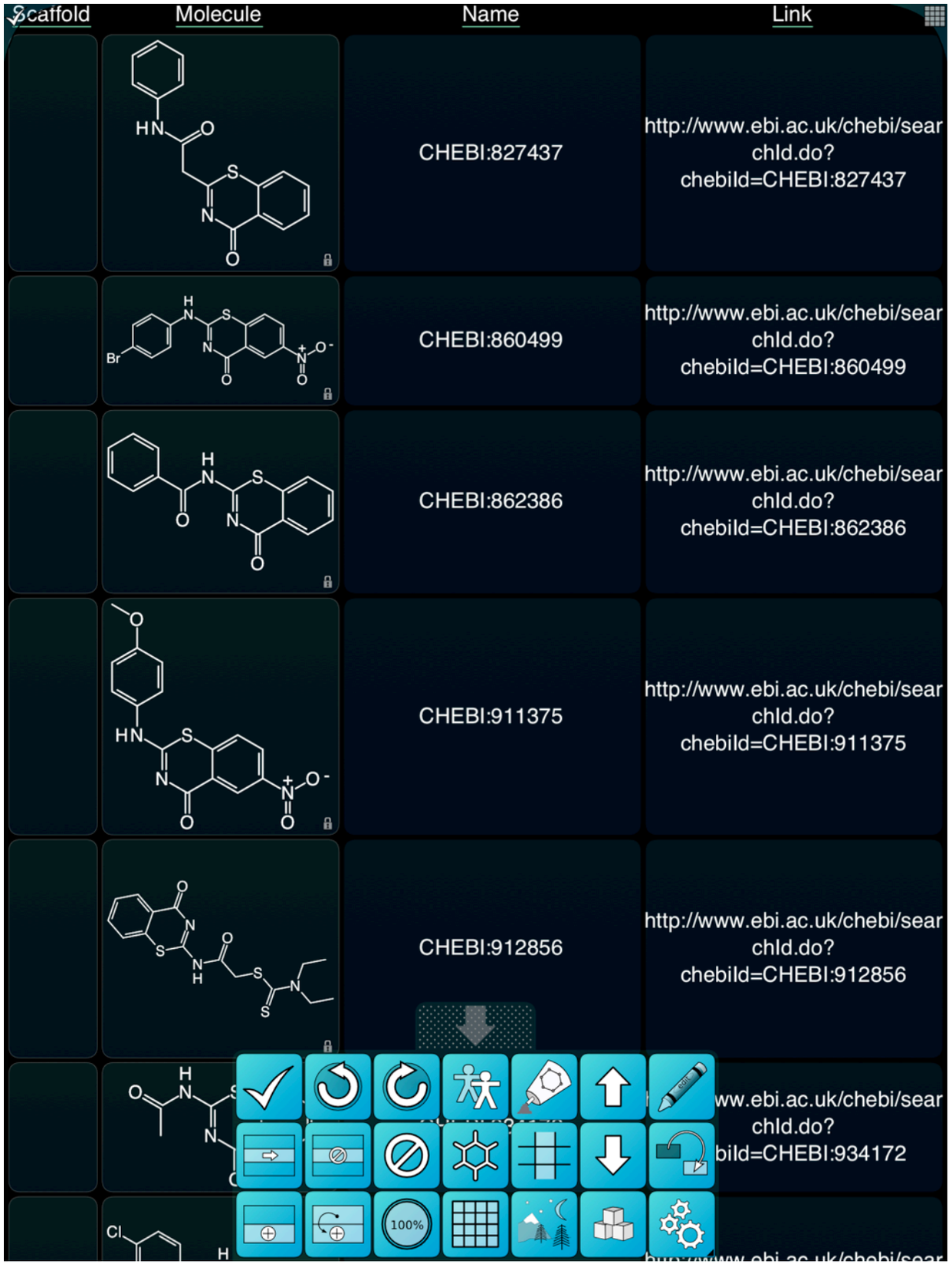




- Substructure search for the core scaffold produces a new datasheet

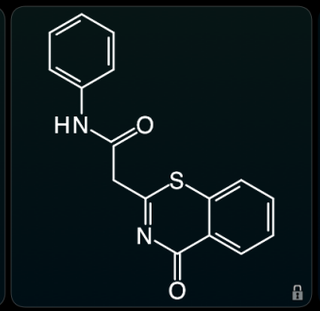
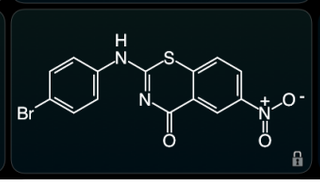
Step 3: Assign scaffolds



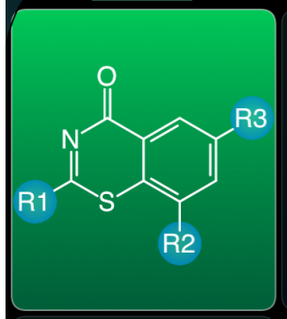
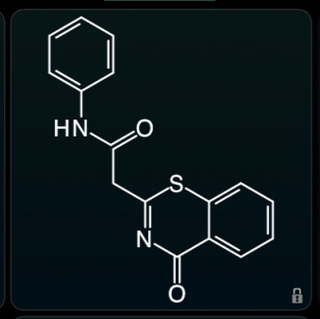
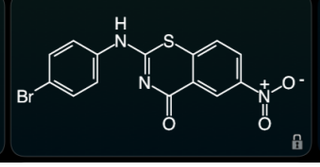


- Open in *SAR Table* app...
- ... which represents a table of compounds as **scaffolds** & **substituents**

1

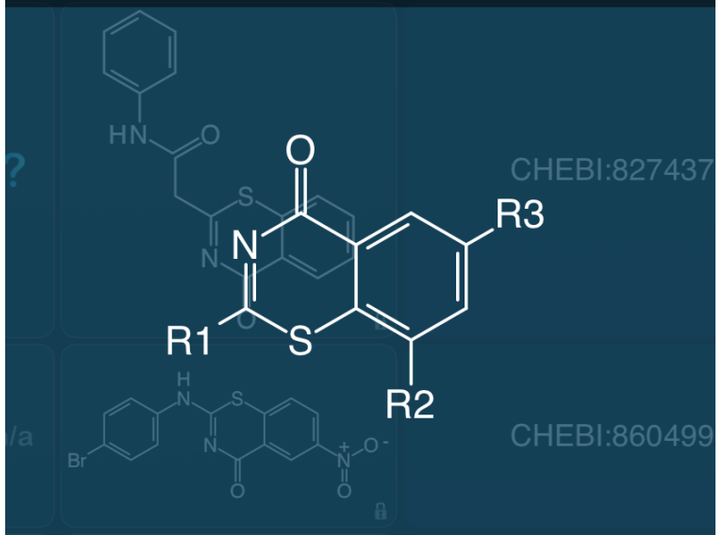
Scaffold	Molecule	Name	Link
		CHEBI:827437	http://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:827437
		CHEBI:860499	http://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:860499

2

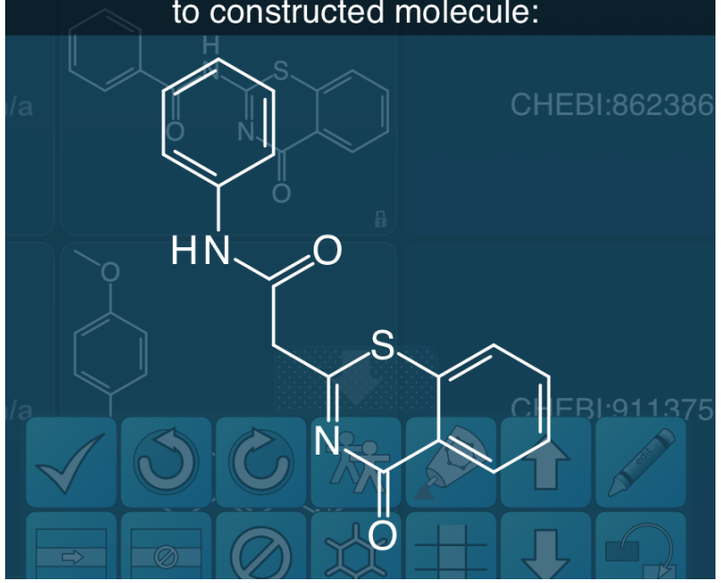
Scaffold	R1	R2	R3	Molecule	Name	
	?	?	?		CHEBI:827437	http:/ c
	n/a	n/a	n/a		CHEBI:860499	http:/ c

3

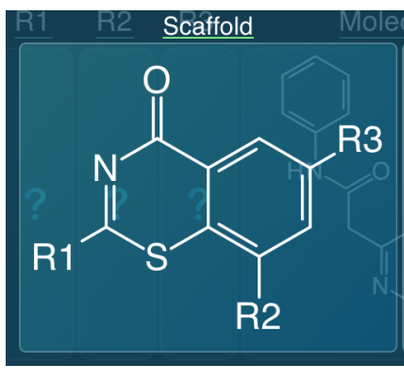
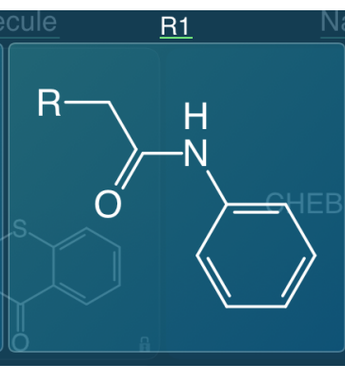
Match scaffold:



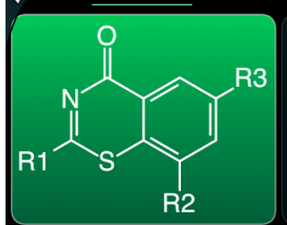
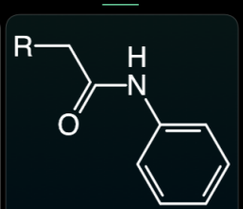
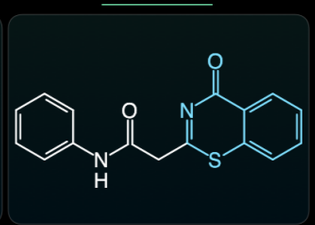
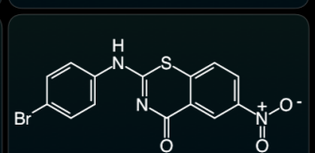
to constructed molecule:



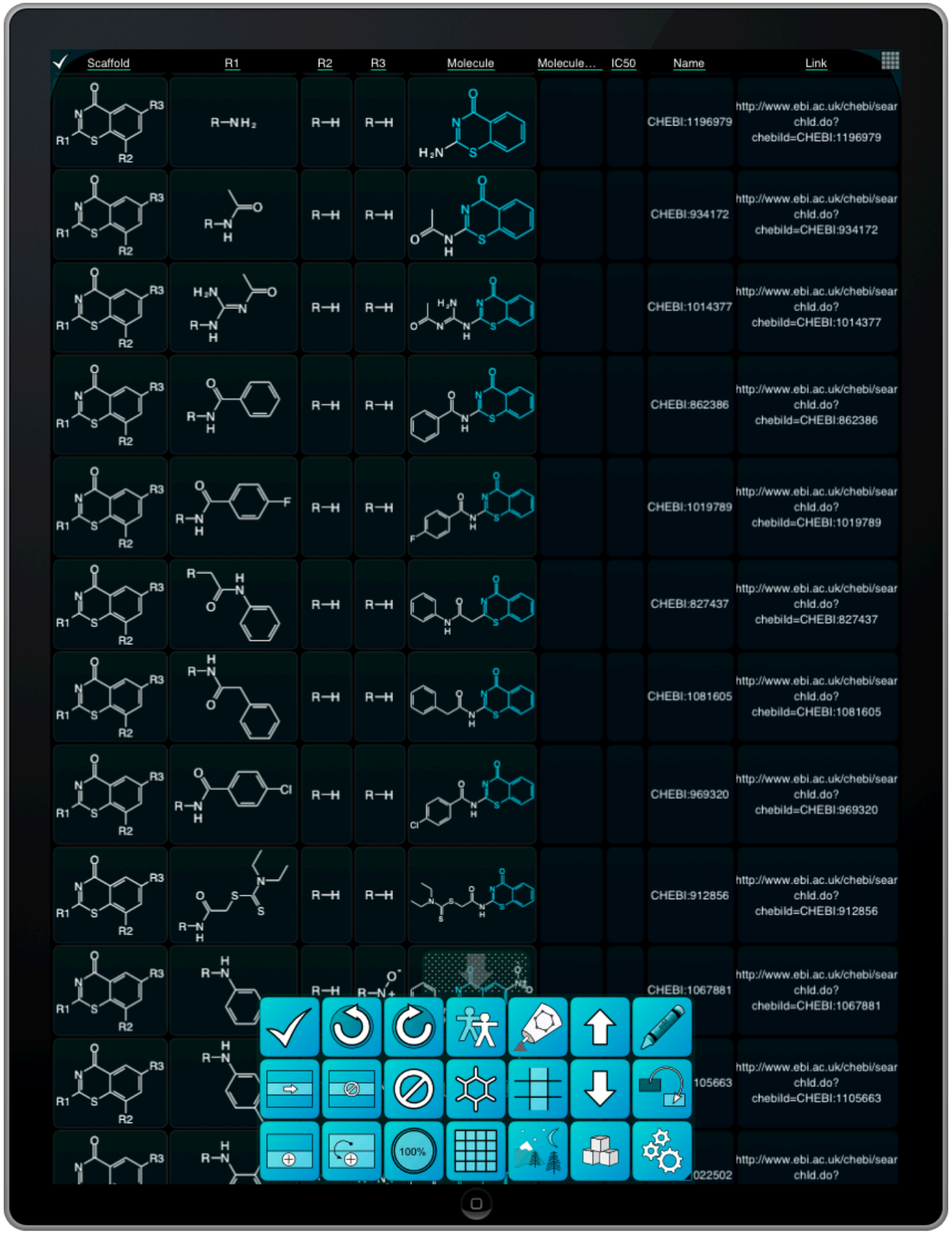
4

Scaffold	R1	R2	R3
		R—H	R—H

5

Scaffold	R1	R2	R3	Molecule	Name
		R—H	R—H		CHEBI:827437
	n/a	n/a	n/a		CHEBI:860499

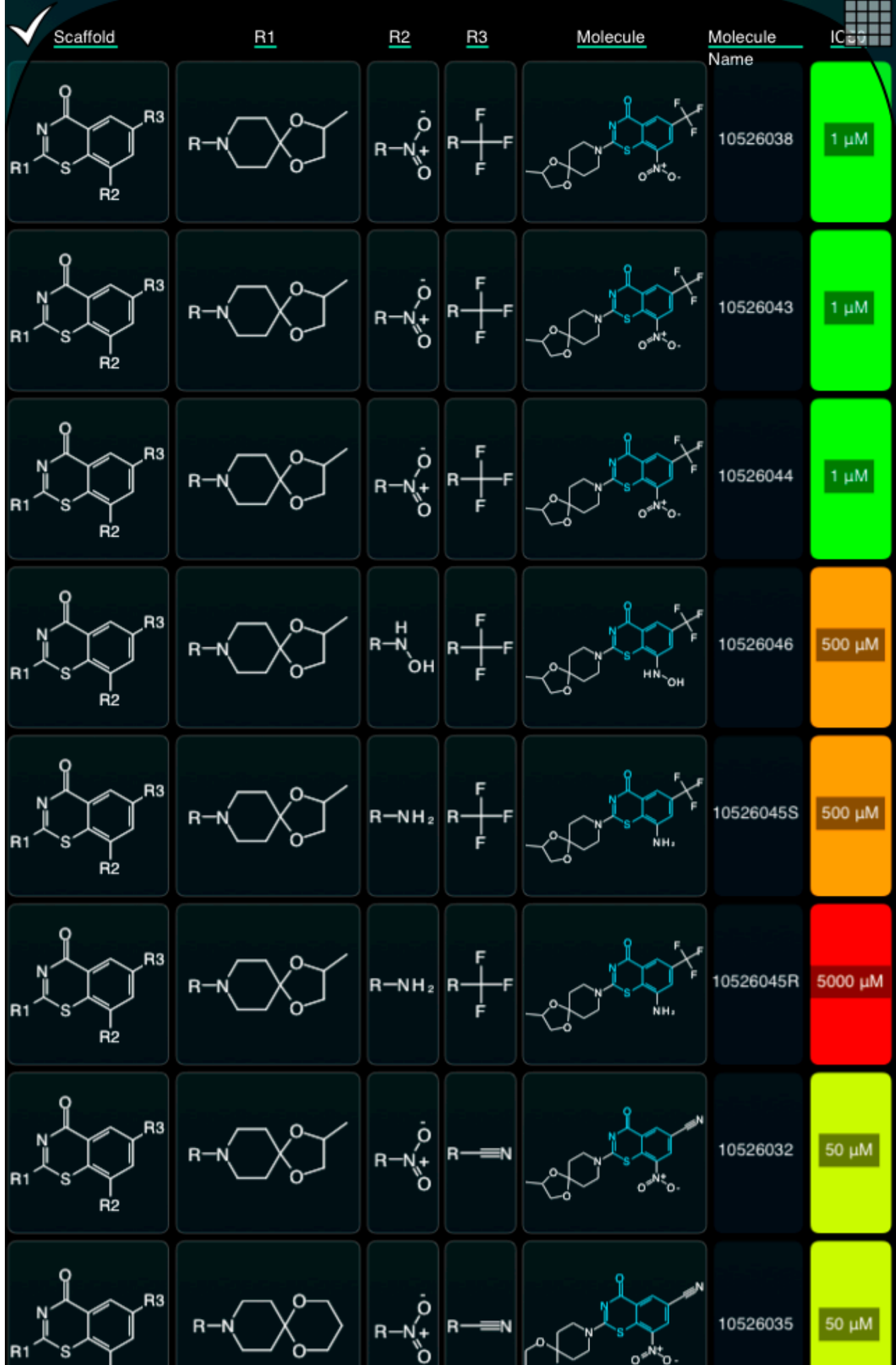
Scaffolds + Substituents



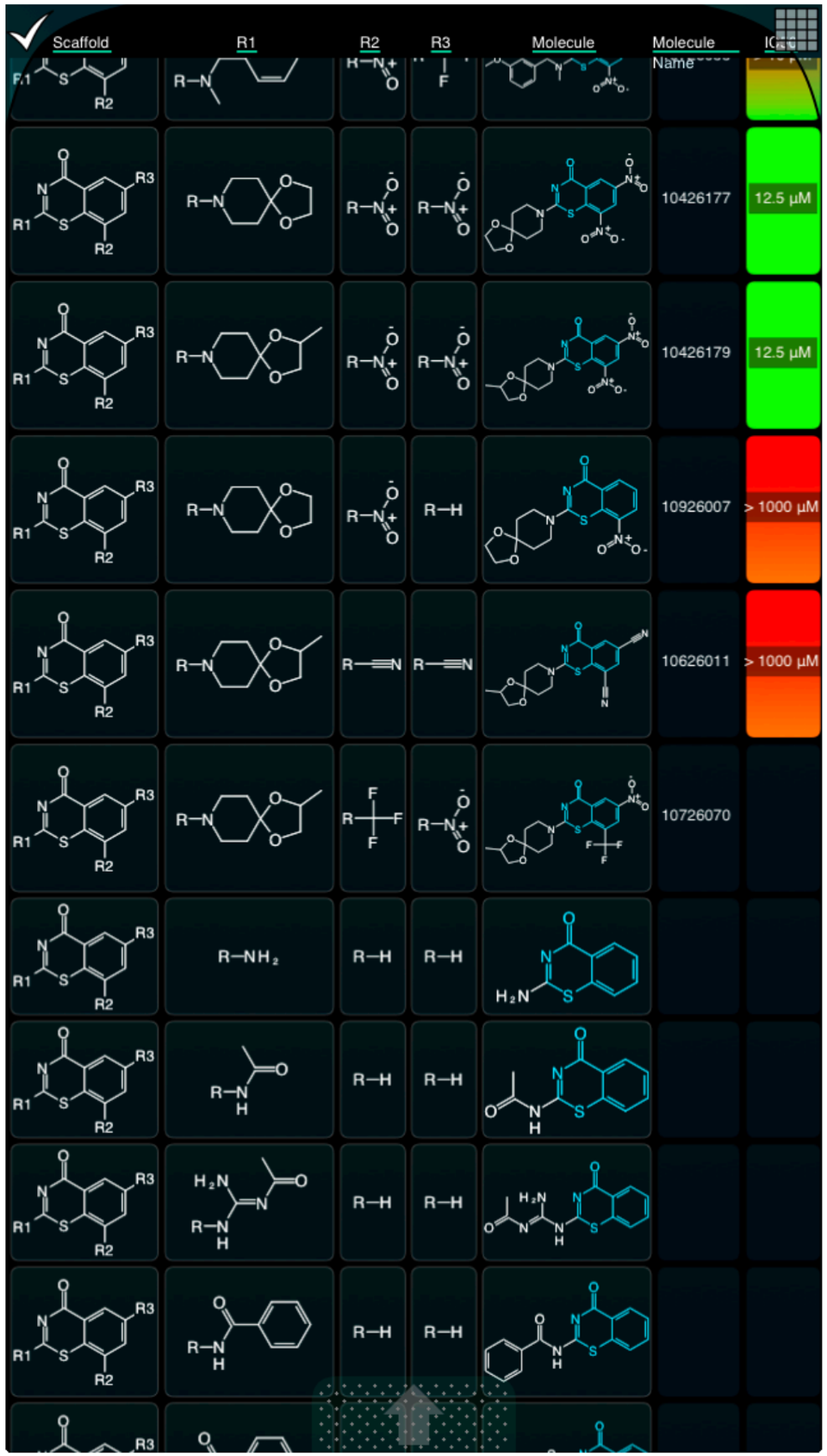
- Structure-activity relationship is aided by classifying **scaffold** and **substituents**
- Assignment process is partly manual, partly automated
- Structure fragments can be drawn within the app, copied from other cells, or computed by a webservice

Step 4: Merge

- Have existing table, created from literature data:
  - V. Makarov et al., Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis, *Science* **324**, pp. 801-804 (2009)
- Activity values colour-coded by defining a *scheme*
- Append the query data to the end...

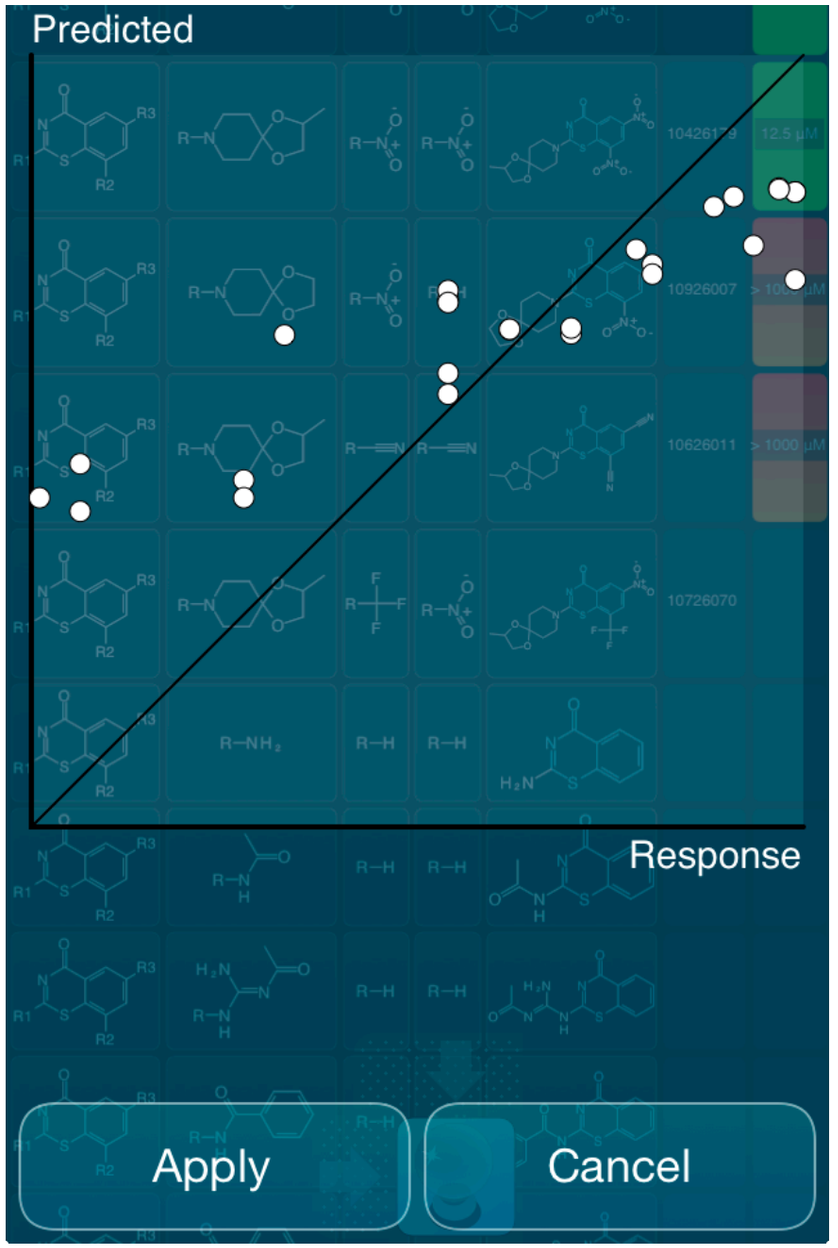


Appended rows

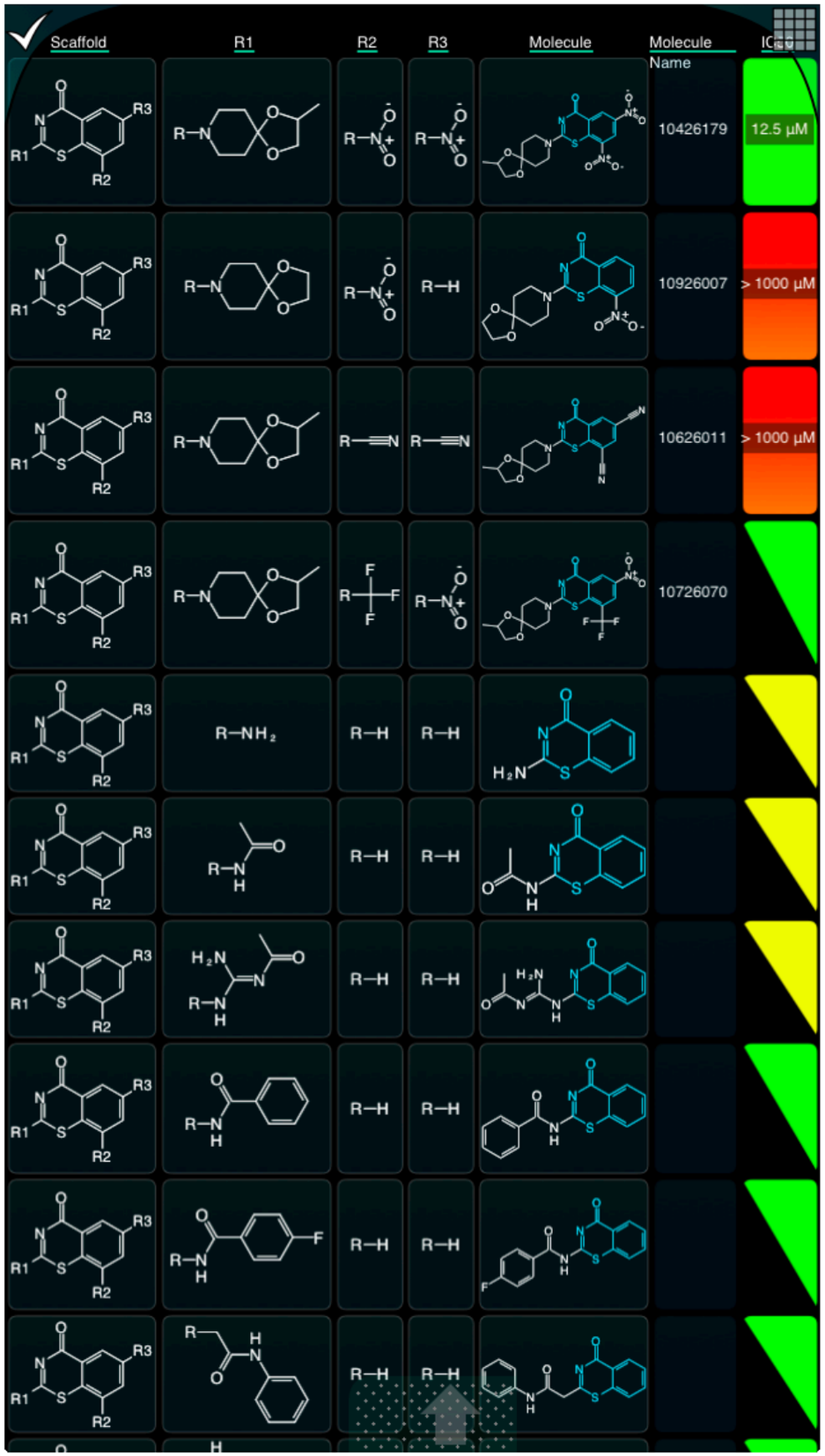


- Webservice builds a model based on existing activity values
- Applies predictions to missing values

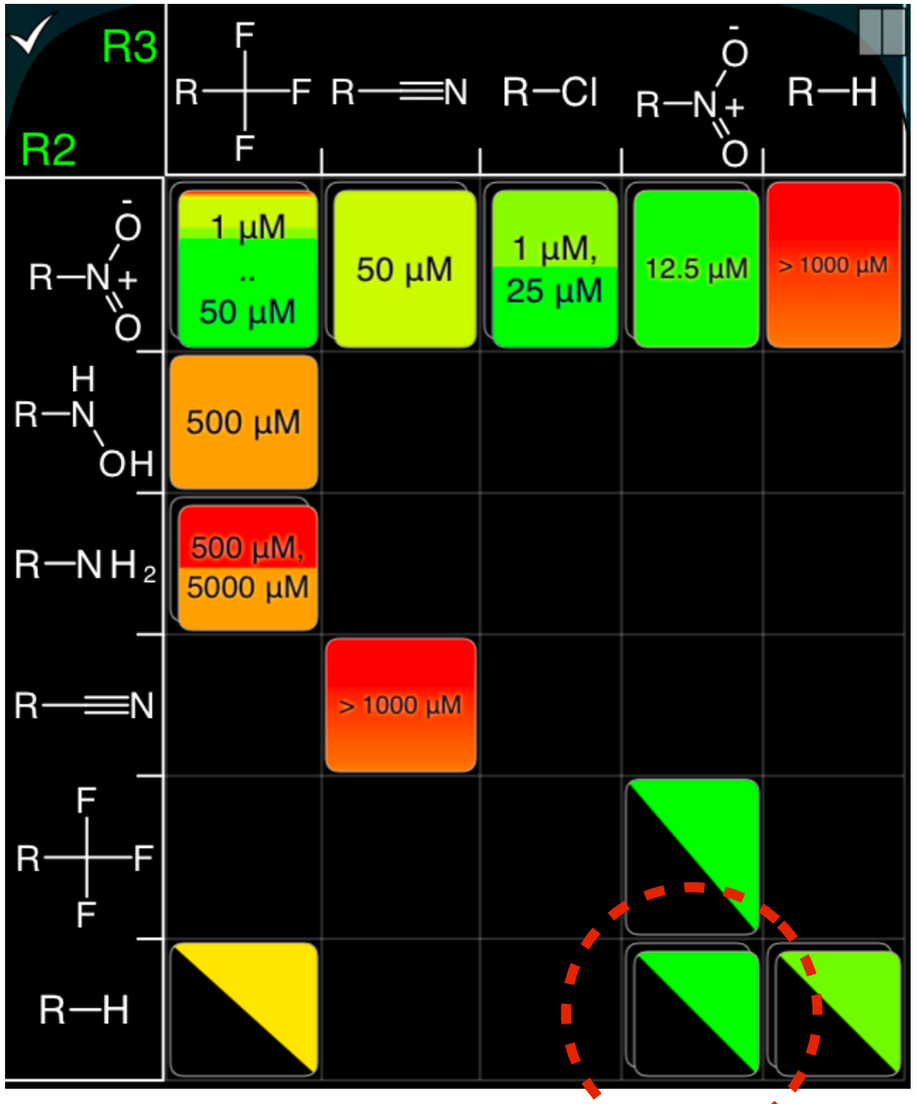
blank activity



Step 5: Examine SAR

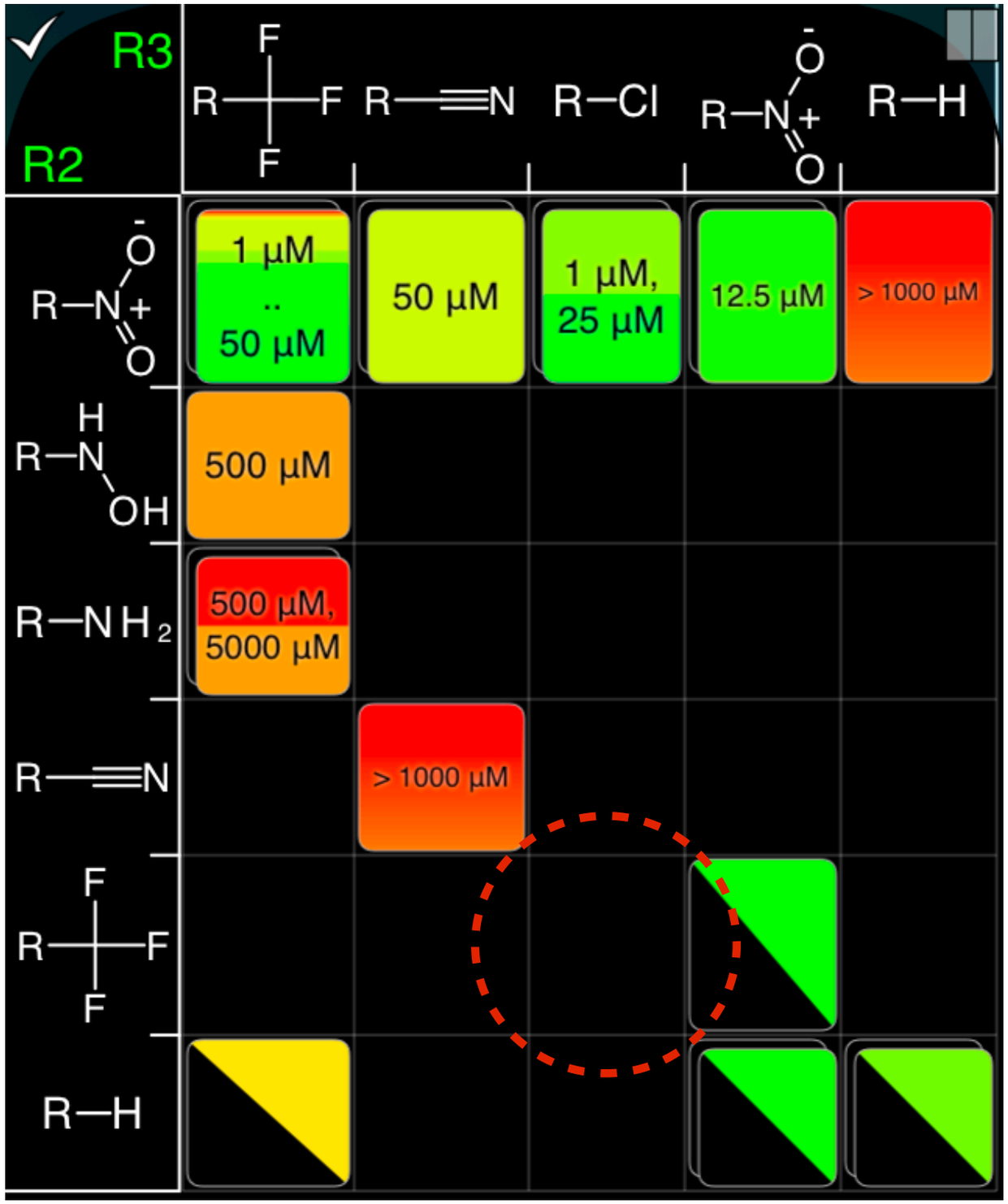


matrix view

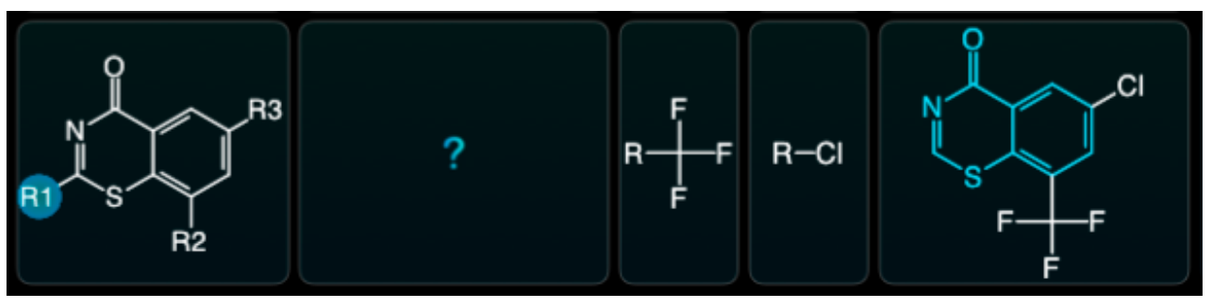


- Predicted values suggest some promising **R2** & **R3** substituents, based on known compounds
- These are from ChEBI, so can be looked up

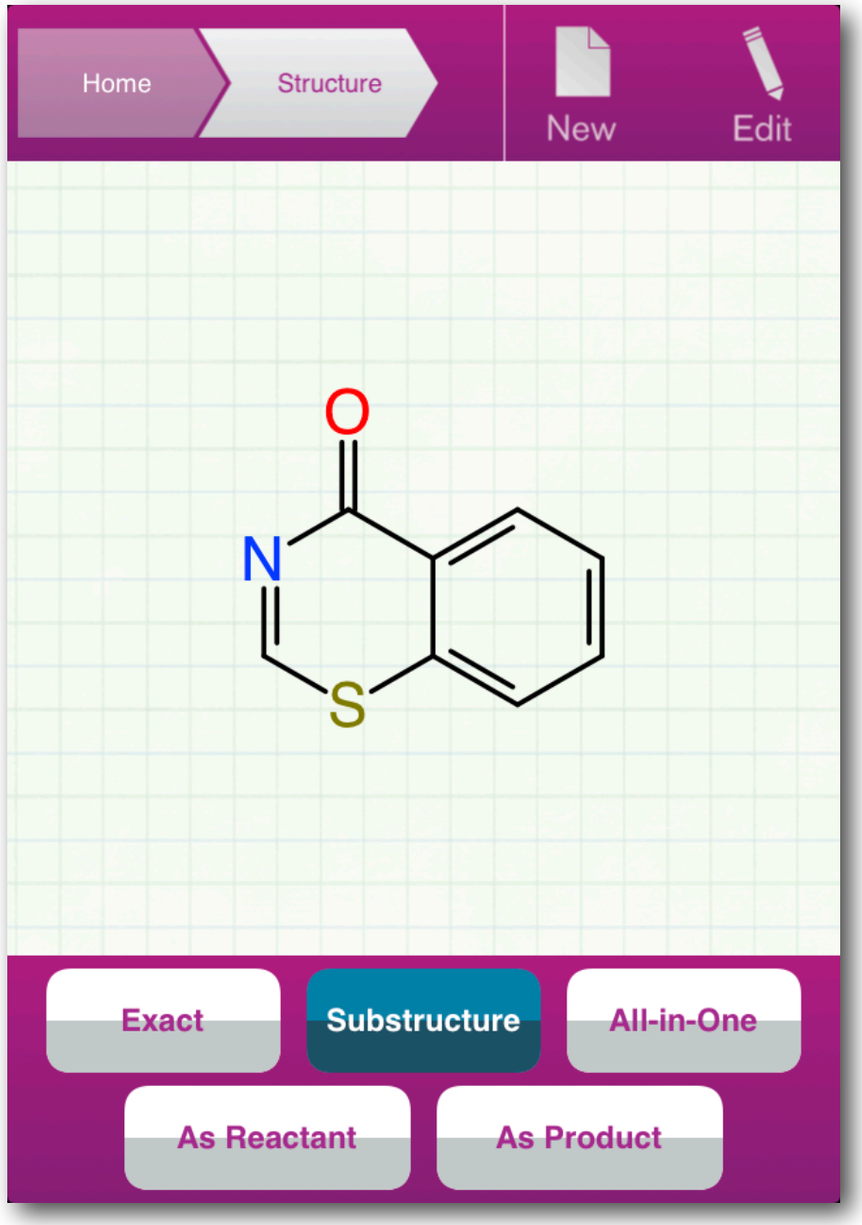
Step 6: Propose new lead

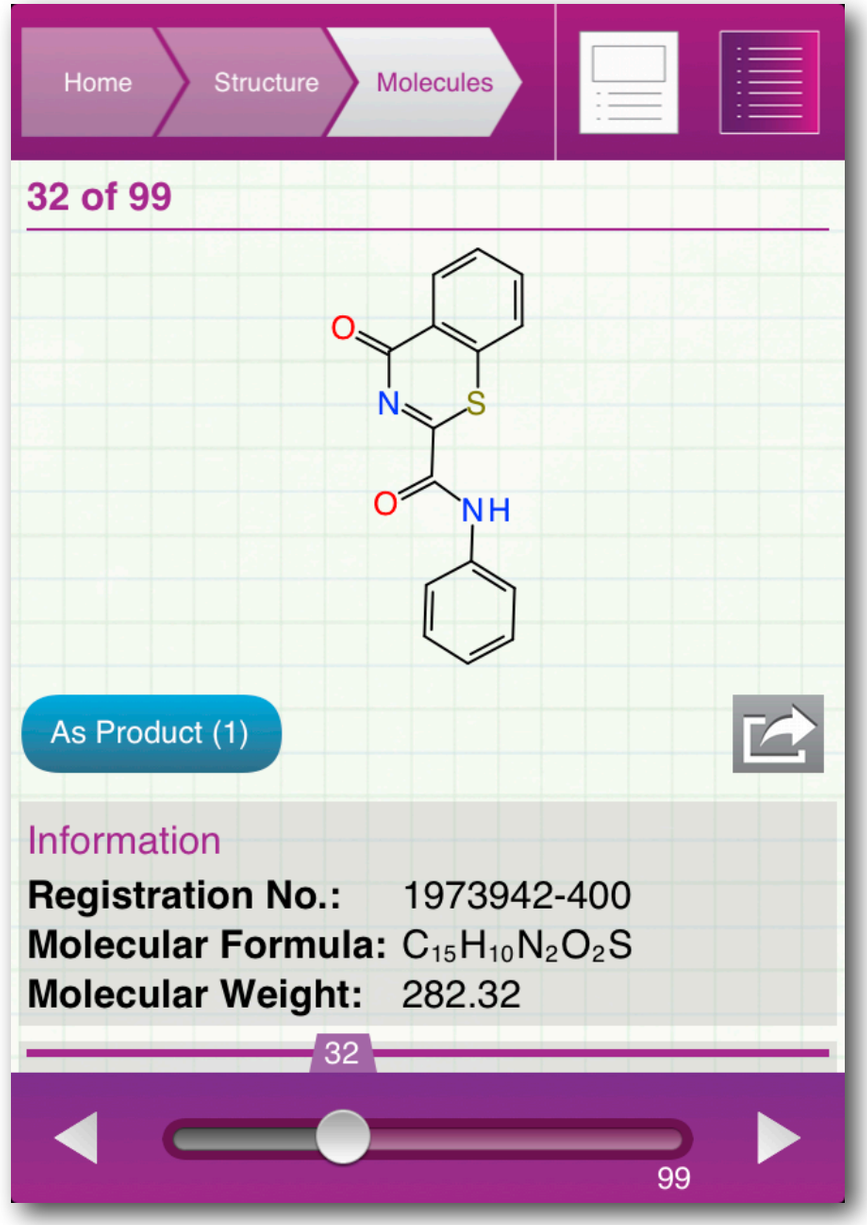


- Identify values for **R2** and **R3** by examining current *structure-activity relationship*
- Create a new row with the proposed substituent moieties:



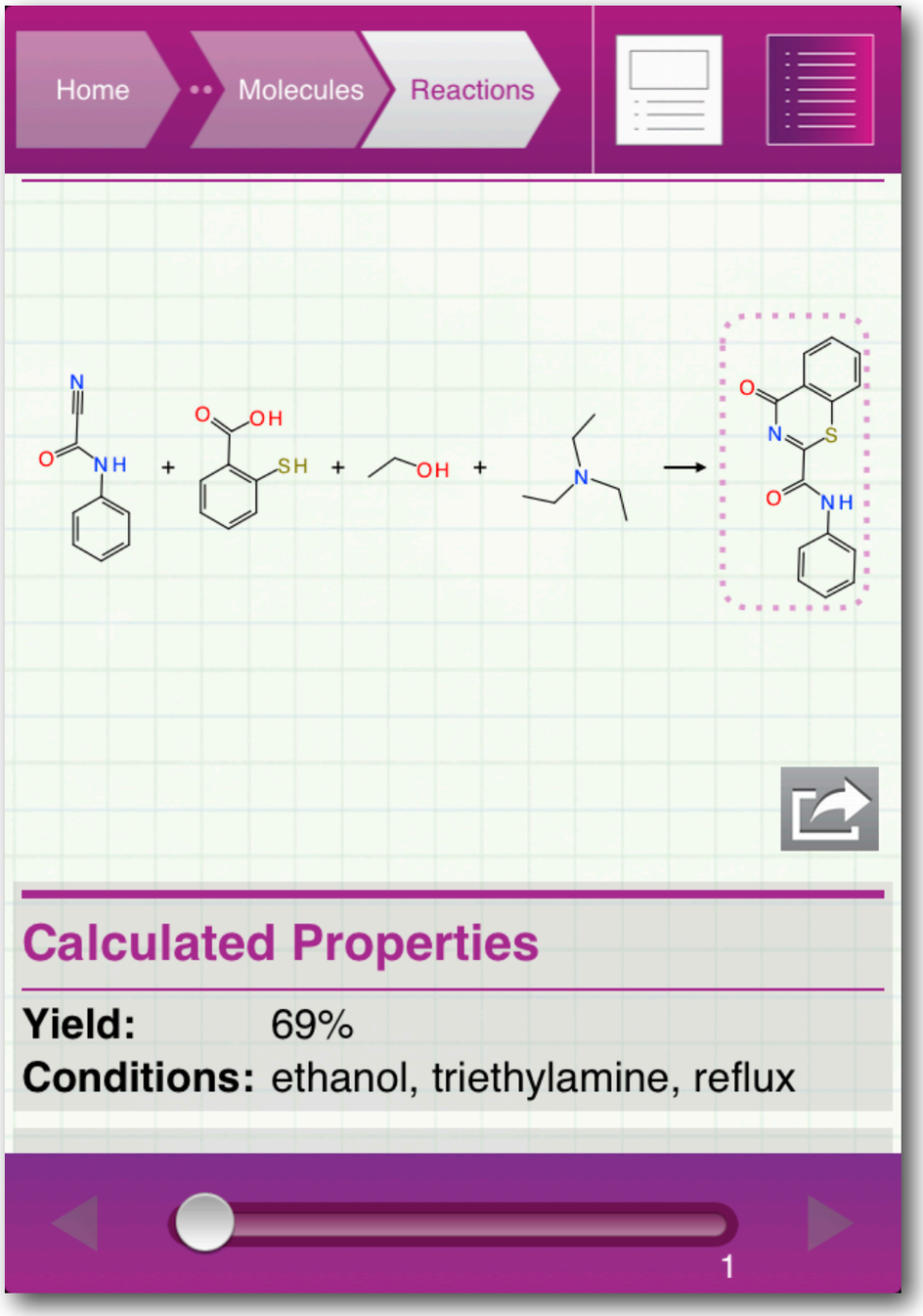
Step 7: Find a recipe





- Open the core structure in *SPRESImobile* app
- Search for similar structures, and search <u>As Product</u> to find a promising synthesis that can be adapted

Find reaction

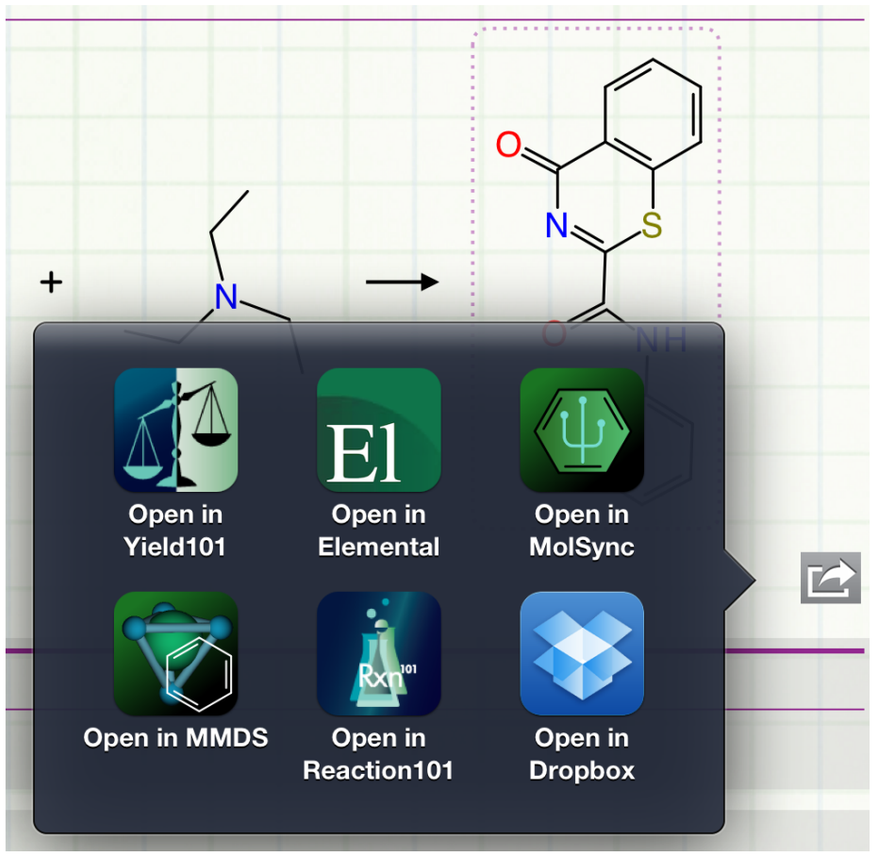


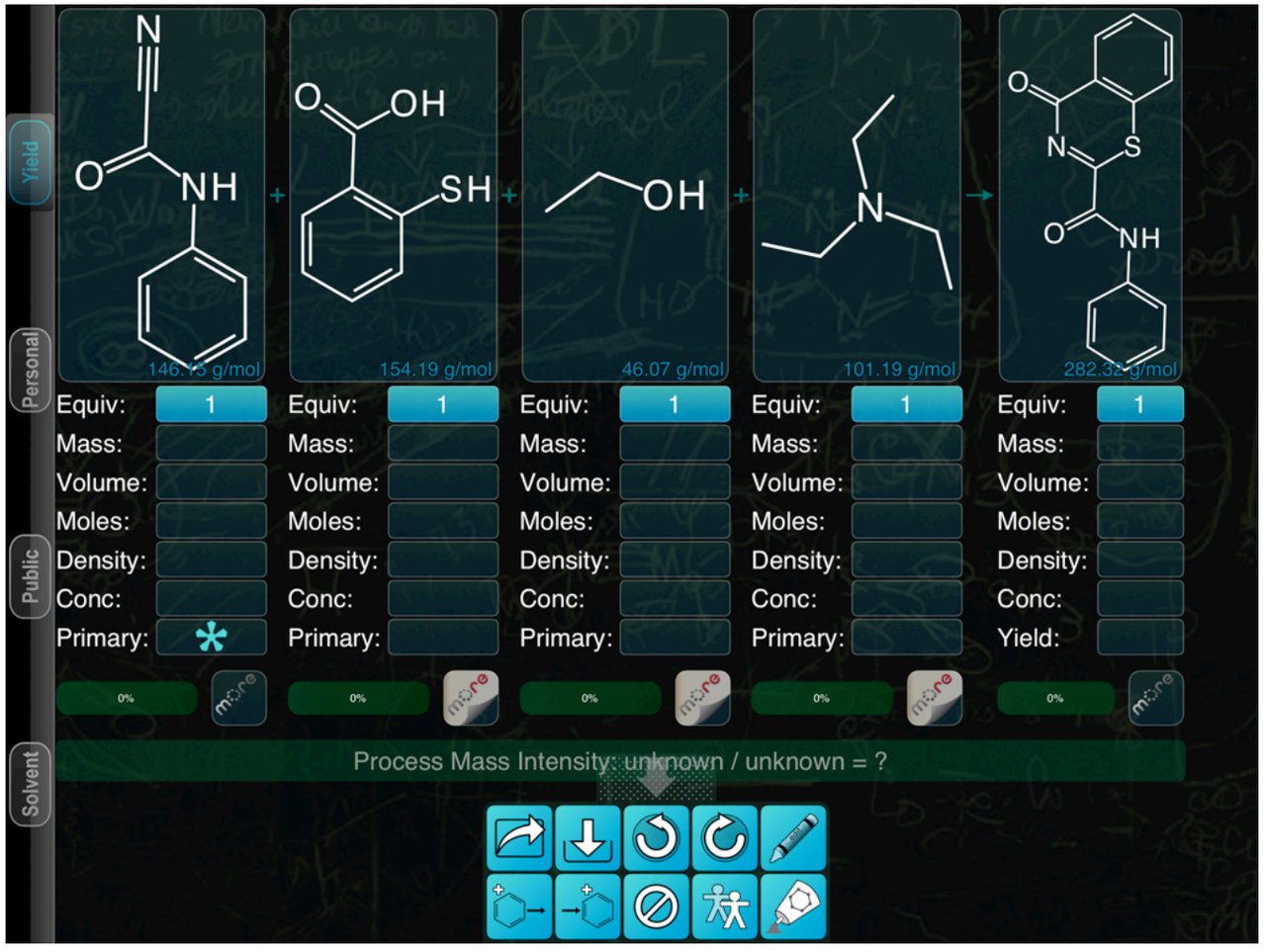
- Found a promising synthesis
- Starting materials suitable for desired functionalisation
- Link to literature provided within the app: can be opened in *Safari Mobile*

Journal Articles

SHERIF, SHERIF MOURAD;MOHAREB, RAFAAT MILAD;ELGEMEIE, GALAL ELDIN H., "NITRILES IN HETEROCYCLIC SYNTHESIS: 1-CYANOFORMANILIDE AS PRECURSOR FOR A VARIETY OF HETEROCYCLIC RING SYSTEMS", Heterocycles, vol. 27, pp. 1579-1583 (1988)
http://dx.doi.org/10.3987/COM-88-4559

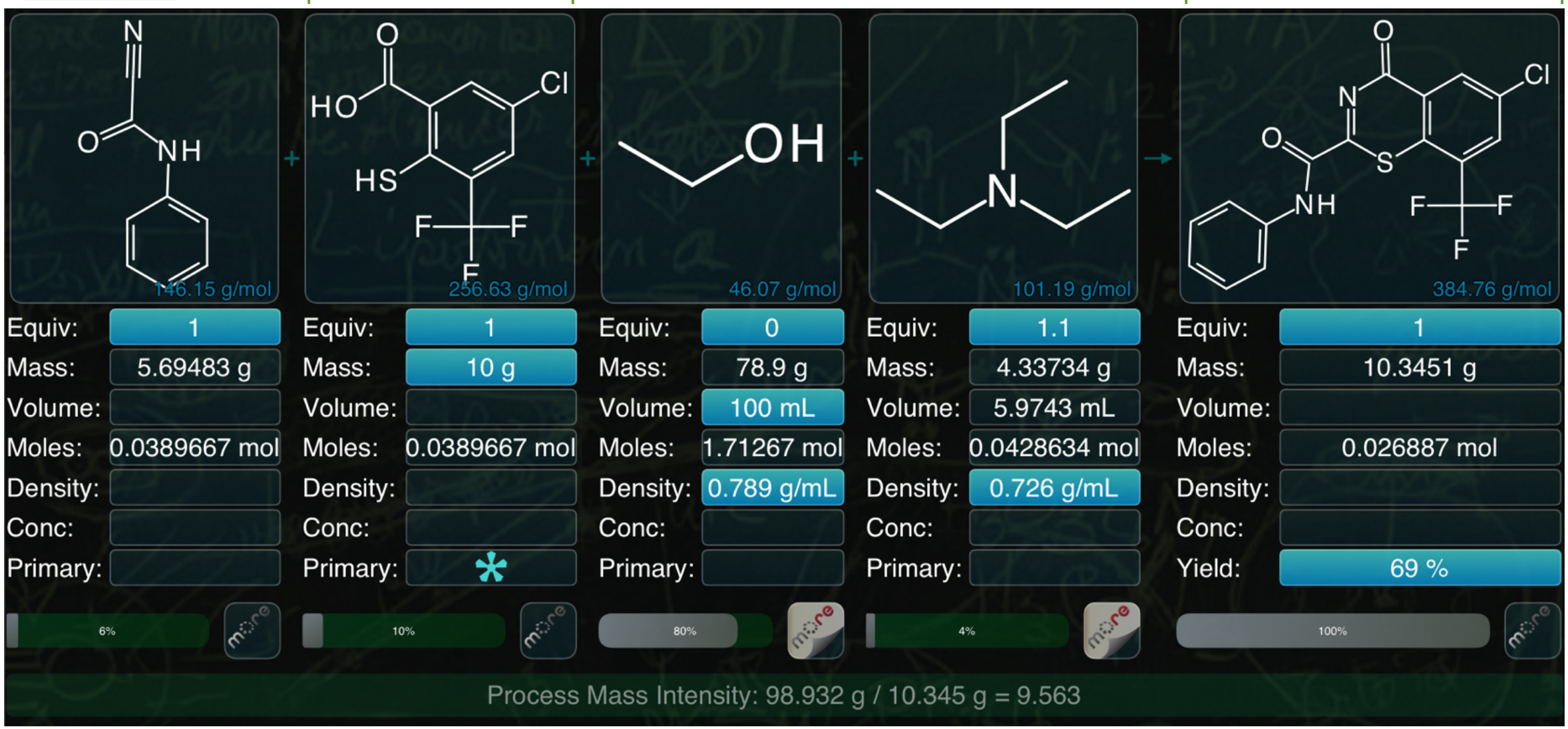
Step 8: Plan synthesis





- Can open the reaction with **Yield101** app: scheme is transferred
- Need to customise the reaction components and add quantities
- Have the literature reference on hand

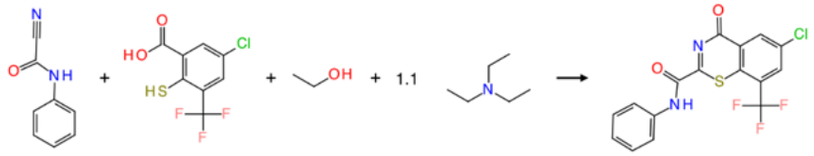
Adapt synthesis



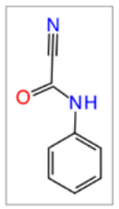
- Edit structure **reactant** and **product** to include functional groups
- Specify known quantities and expected yield: all implied quantities are calculated, as is the *process mass intensity*

Run the experiment

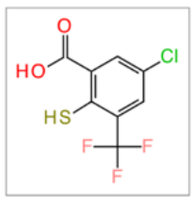
Done Print



Reactants

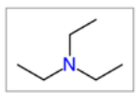


Molecular Formula:	$C_8H_6N_2O$
Molecular Weight:	146.146 g/mol
Equivalents:	1
Mass:	5.69483 g (calculated)
Moles:	0.0389667 mol (calculated)



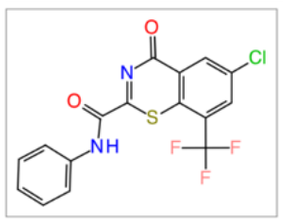
Molecular Formula:	$C_8H_4ClF_3O_2S$
Molecular Weight:	256.629 g/mol
Equivalents:	1
Mass:	10 g
Moles:	0.0389667 mol (calculated)
(Primary Reactant)	

Molecular Formula:	C_2H_6O
Molecular Weight:	46.0684 g/mol
Equivalents:	0
Mass:	78.9 g (calculated)
Volume:	100 mL
Moles:	1.71267 mol (calculated)
Density:	0.789 g/mL



Molecular Formula:	$C_6H_{15}N$
Molecular Weight:	101.19 g/mol
Equivalents:	1.1
Mass:	4.33734 g (calculated)
Volume:	5.9743 mL (calculated)
Moles:	0.0428634 mol (calculated)
Density:	0.726 g/mL

Products



Molecular Formula:	$C_{16}H_8ClF_3N_2O_2S$
Molecular Weight:	384.76 g/mol
Equivalents:	1
Mass:	10.3451 g (calculated)
Moles:	0.026887 mol (calculated)
Yield:	69 %

- Can use the mobile device in the lab, and record the results directly

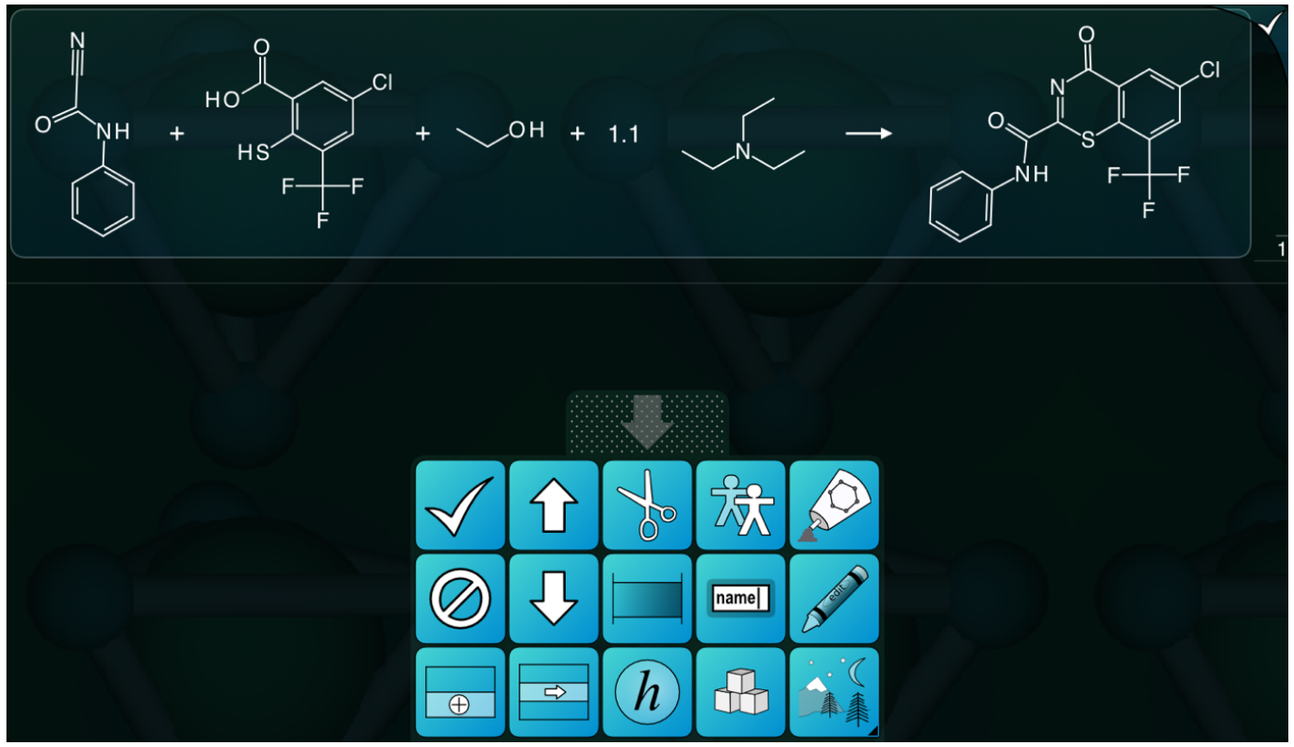
  OR:

- Create a PDF document, and print it directly from the device or send the file by email...

- ... then enter the results later.

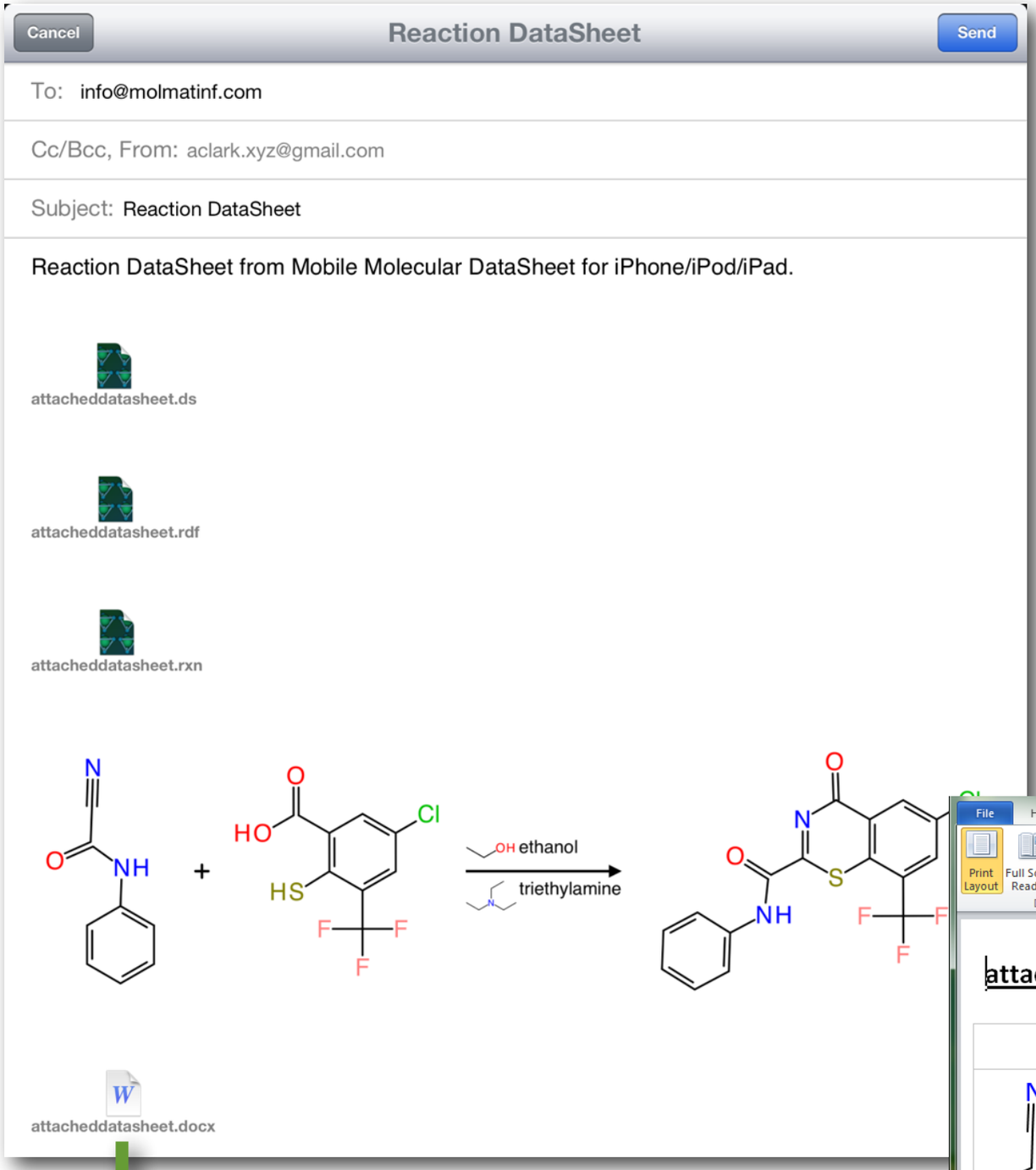
Step 9: Sharing data

- *Many* ways to share data: use Open-With to transfer to MMDS

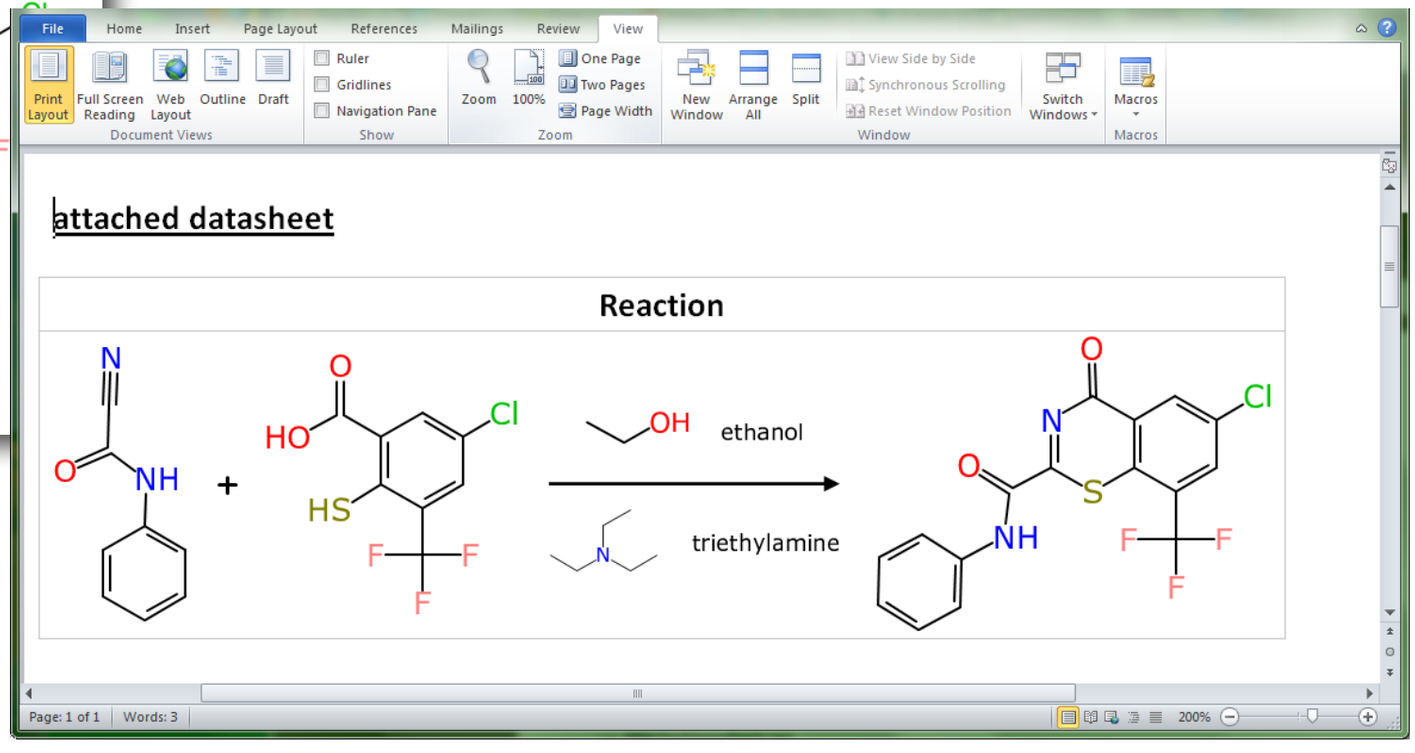


- From there, can:
  - send data by email
  - export presentation quality graphics
  - share on web & tweet
  - upload to repositories, e.g. Dropbox

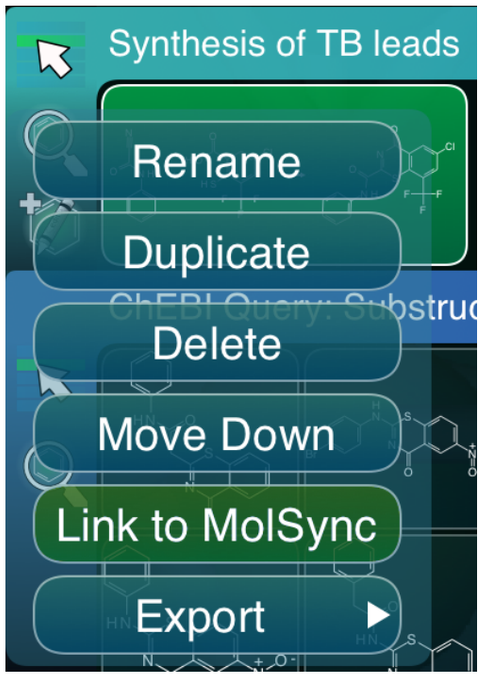
Microsoft Office documents

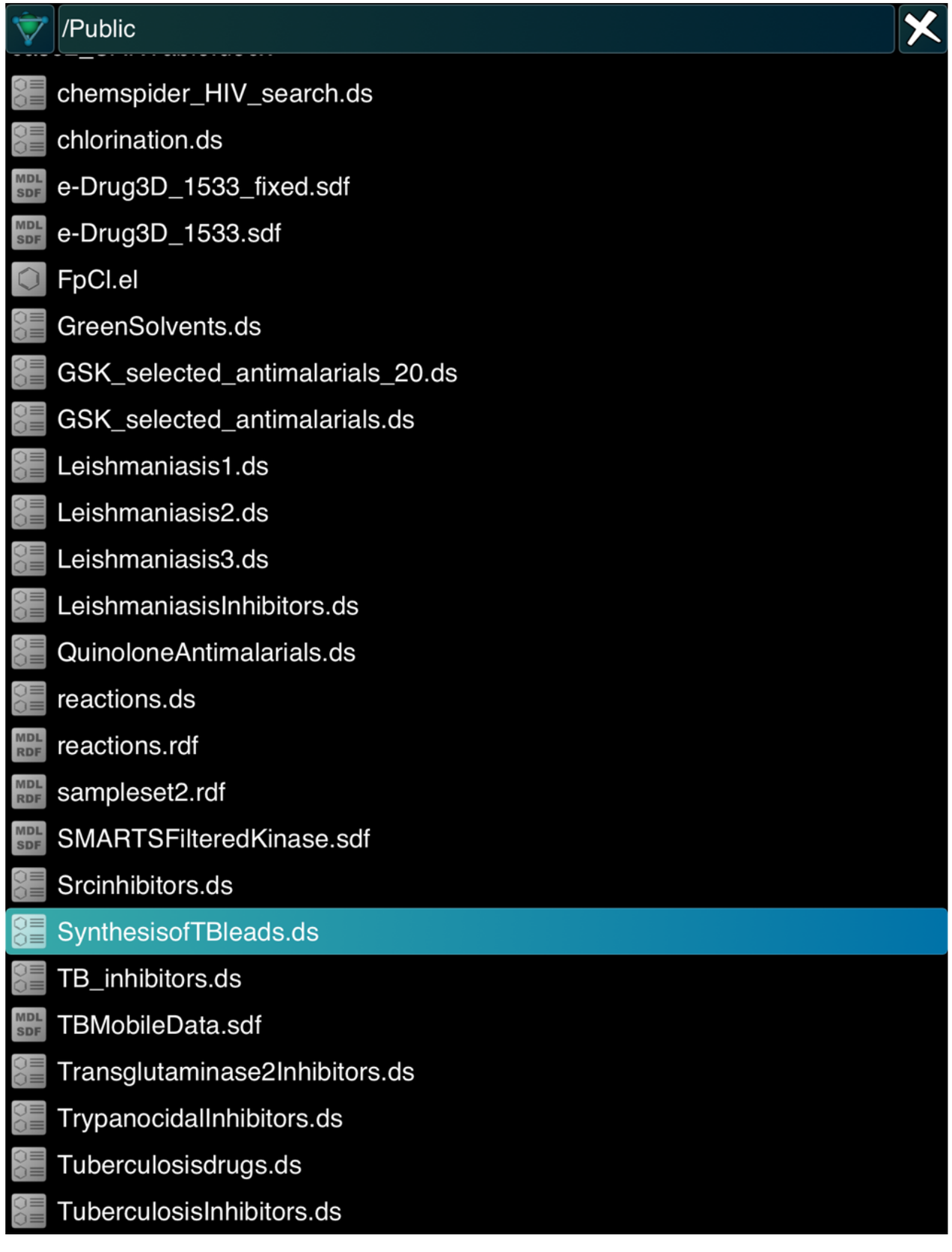


- Outgoing emails can include many attachments:
  - machine-readable data
  - graphics
- Graphics options include MS Word and Excel documents with embedded **vector** graphics



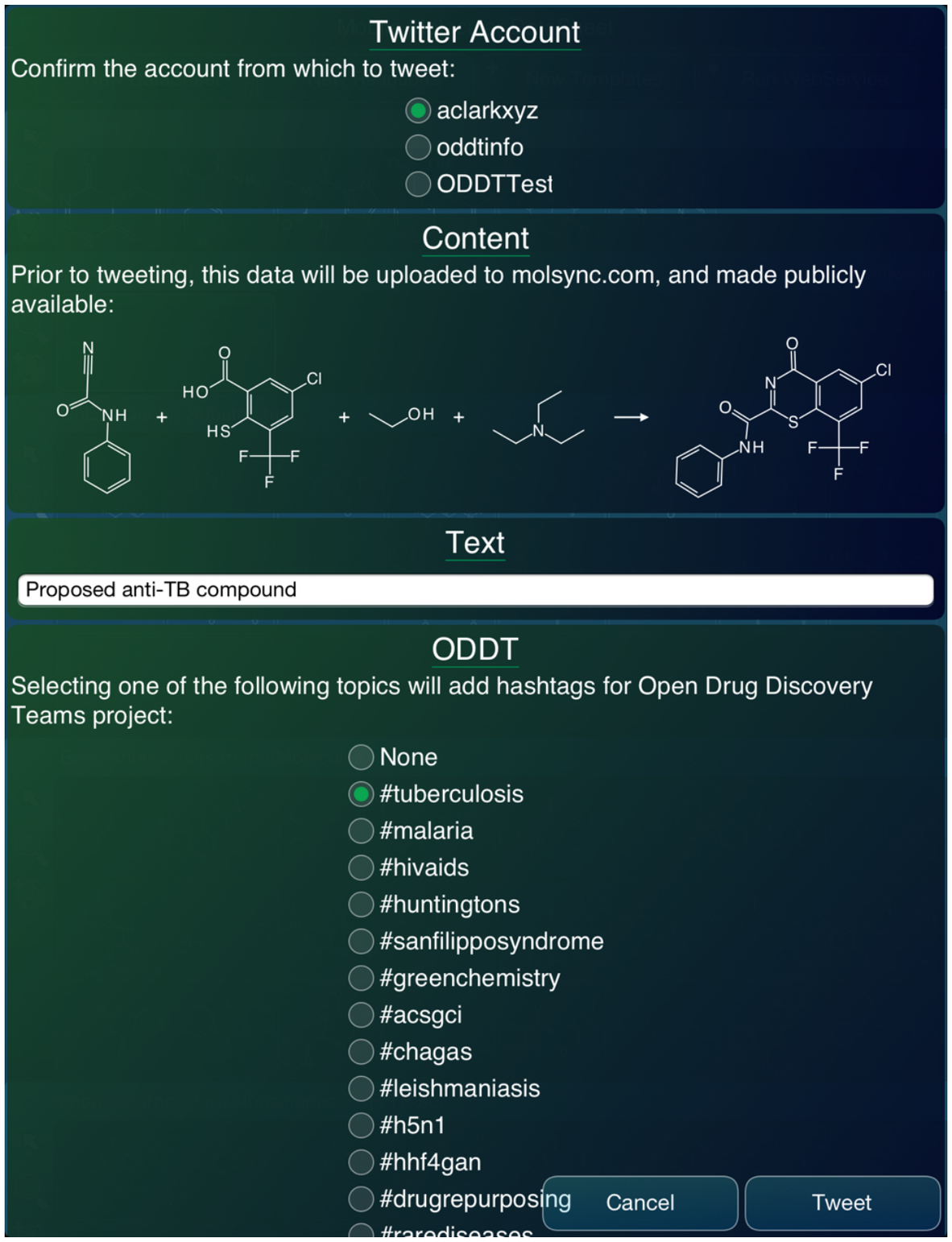
Sharing with Dropbox

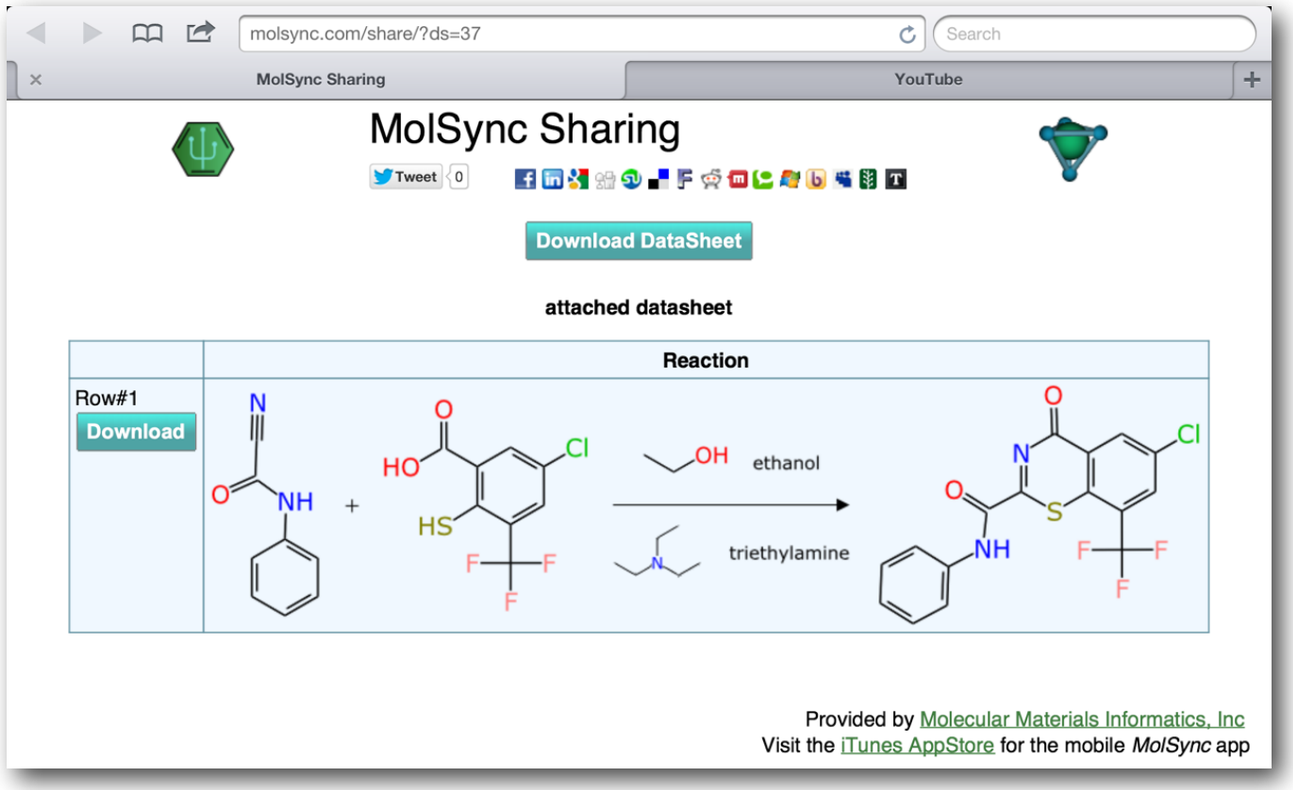


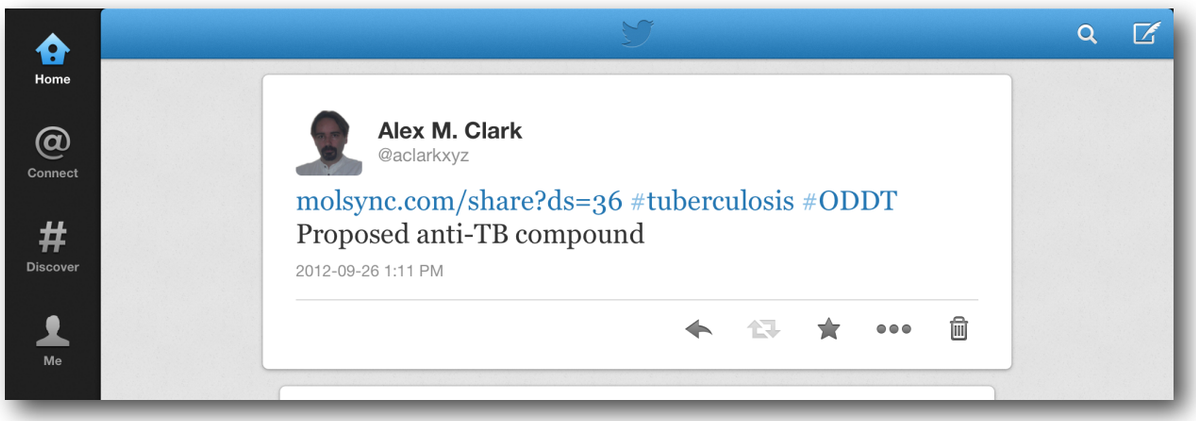


- Can link datasheets to *Dropbox* using the *MolSync* app
- Storing data on *Dropbox* opens up an entire landscape of collaboration and sharing options

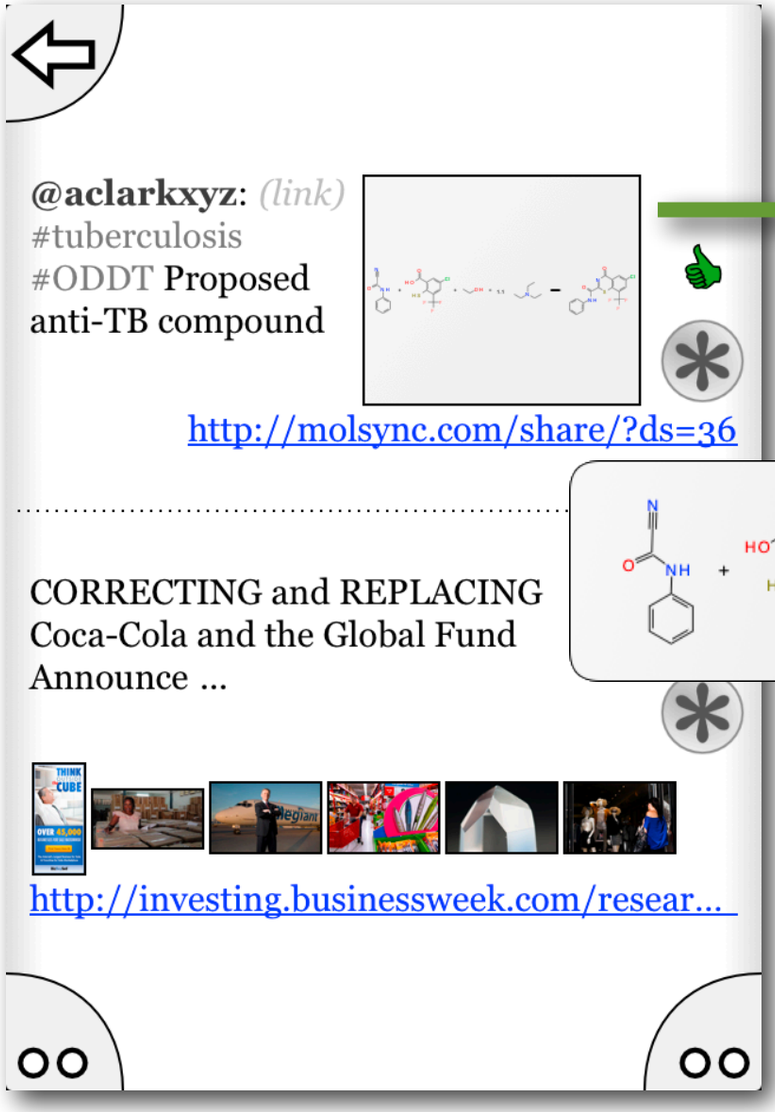
Step 10: Open sharing







Open Drug Discovery Teams



- Relevant tweets get picked up by the *Open Drug Discovery Teams* project, and made available for browsing with a free app
- Chemical data is parsed and can be accessed with other apps

Conclusion

- Some moderately sophisticated, and quite flexible, workflows are already possible using existing tools
- Workflows involve content creation, content consumption, network sharing and computation
- Native apps are currently quite good for visualisation and data entry, **but**: confined to small data; avoid complex calculations; need network access to offload
- Big data workflows need specialised server and protocols to packetise the data; database querying is a simple example, but more complex schemes are in the works, e.g. *Pistoia Alliance AppStore* infrastructure

Future work

- Evolution from proof-of-concept phase to apps designed to solve specific workflows: expect more customised apps
- Consolidation of platforms and APIs: supporting current platforms requires >5 incompatible codebases... need to pick one
- More integration with cloud services: apps cannot work with big data without back-end support
- *Pistoia Alliance* is currently working on an *AppStore*, which provides
  - an alternative storefront for mobile apps
  - standardised service infrastructure, for big data & calculations
- Gradual replacement of all mainstream chemical information software

Acknowledgments

- Antony Williams
- Sean Ekins
- Leah Rae McEwan, David Martinsen
- ACS CINF Division
- Inquiries to **info@molmatinf.com**

http://molmatinf.com
http://molsync.com
http://cheminf20.org

@aclarkxyz