Cheminformatics workflows using the mobile + cloud platform

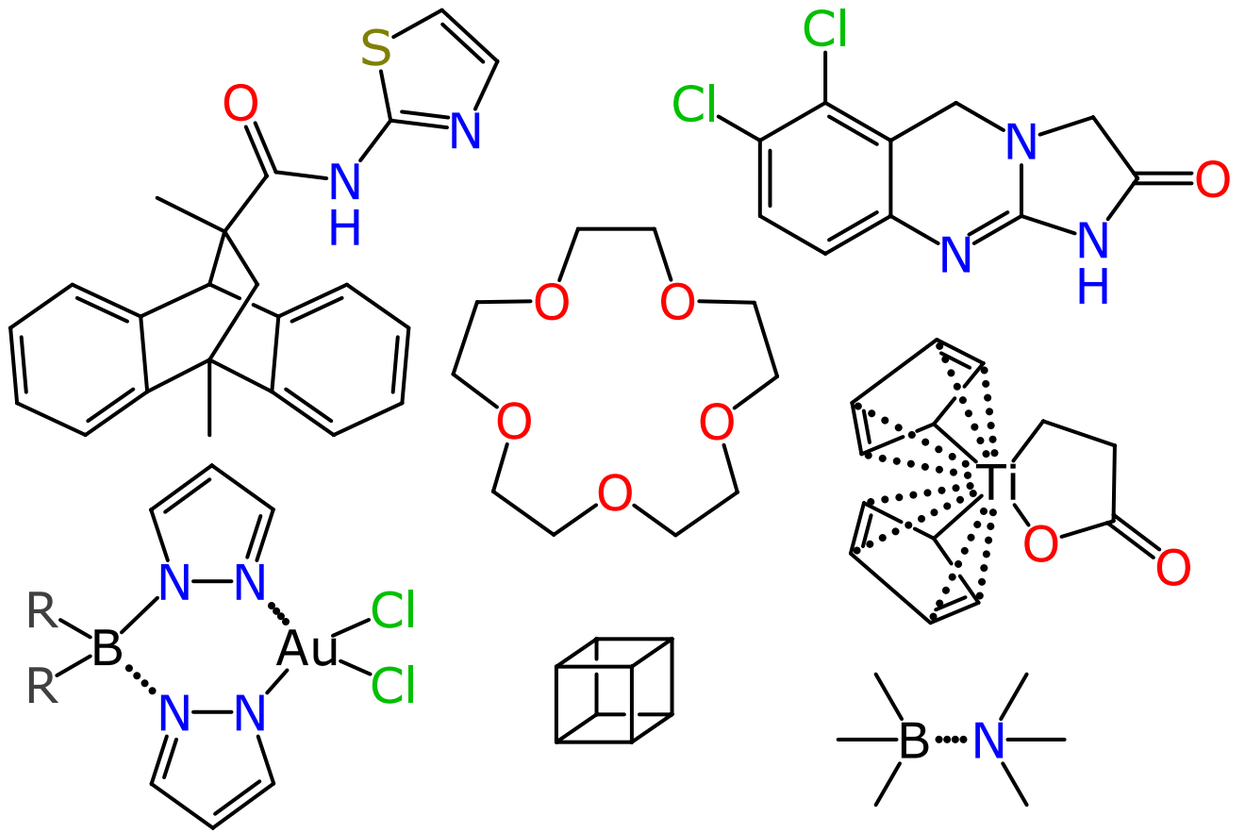
Dr. Alex M. Clark

October 2013

http://molmatinf.com

Overview

- Chemical structures
- Mobile apps
- Cloud computing



MOLECULAR MATERIALS INFORMATICS

3rd Revolution

mainframes
minicomputers

personal computers
portable laptops

mobile tablets
smartphones

- The mobile platform is revolutionary: **a clean break**

  - entirely new user interface
  - no backward compatibility
  - highly constrained resources
  - applicable to entirely new situations

Two platform stacks

desktop/
laptop

phone/
tablet

file/database
server

web
services

compute
cluster

cloud
resources

Delivery

Apps

- User experience
- Cached data
- Light calculations
- Sharing

Cloud

- Web API
- Heavy calculations
- Large data
- Centralised storage

Challenges

- Mobile user interfaces are **completely different** to desktop products
- Expectations: functionality, extreme simplicity
- Complex interfaces much harder: drawing chemical structures requires redesign
- Visualisation is more difficult: screen size
- Storage limitations: big data is a problem
- Calculations split *mobile* vs. *cloud*

App Catalog

Mobile Molecular DataSheet (MMDS)

MolSync

SAR Table

MolPrime+

MolPrime

Green Solvents

Lab Solvents

Approved Drugs

Reaction101

Yield101

Living Molecules

TB Mobile

ChemSpider

SPRESImobile

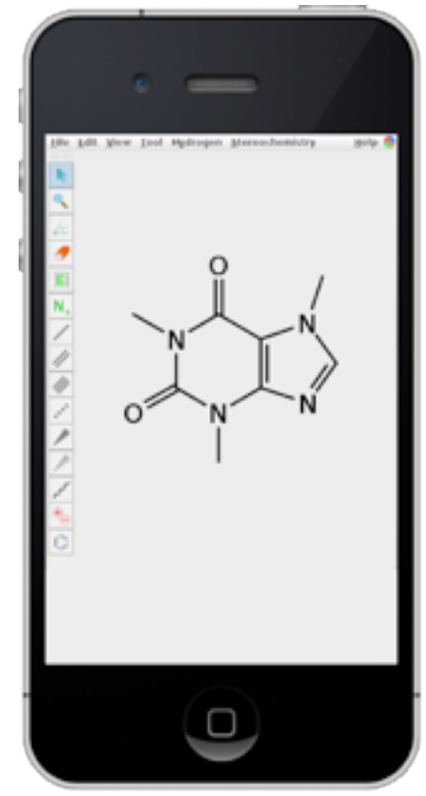
Open Drug Discovery Teams

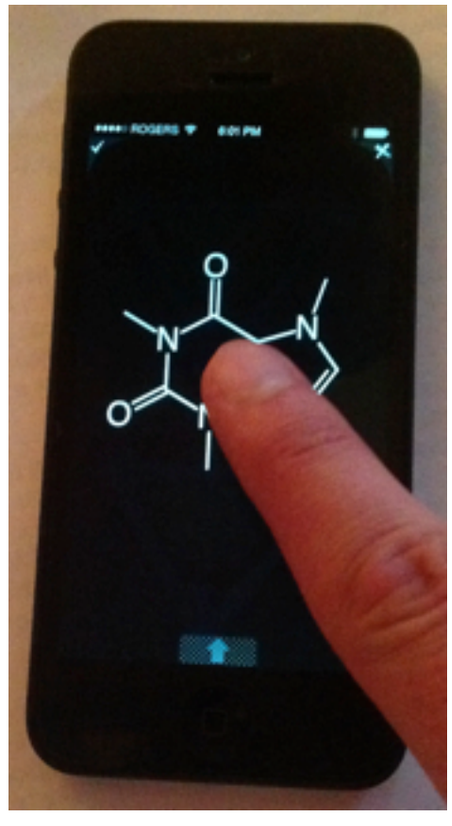
Drawing Structures

- Cheminformatics requires a user interface for drawing structures

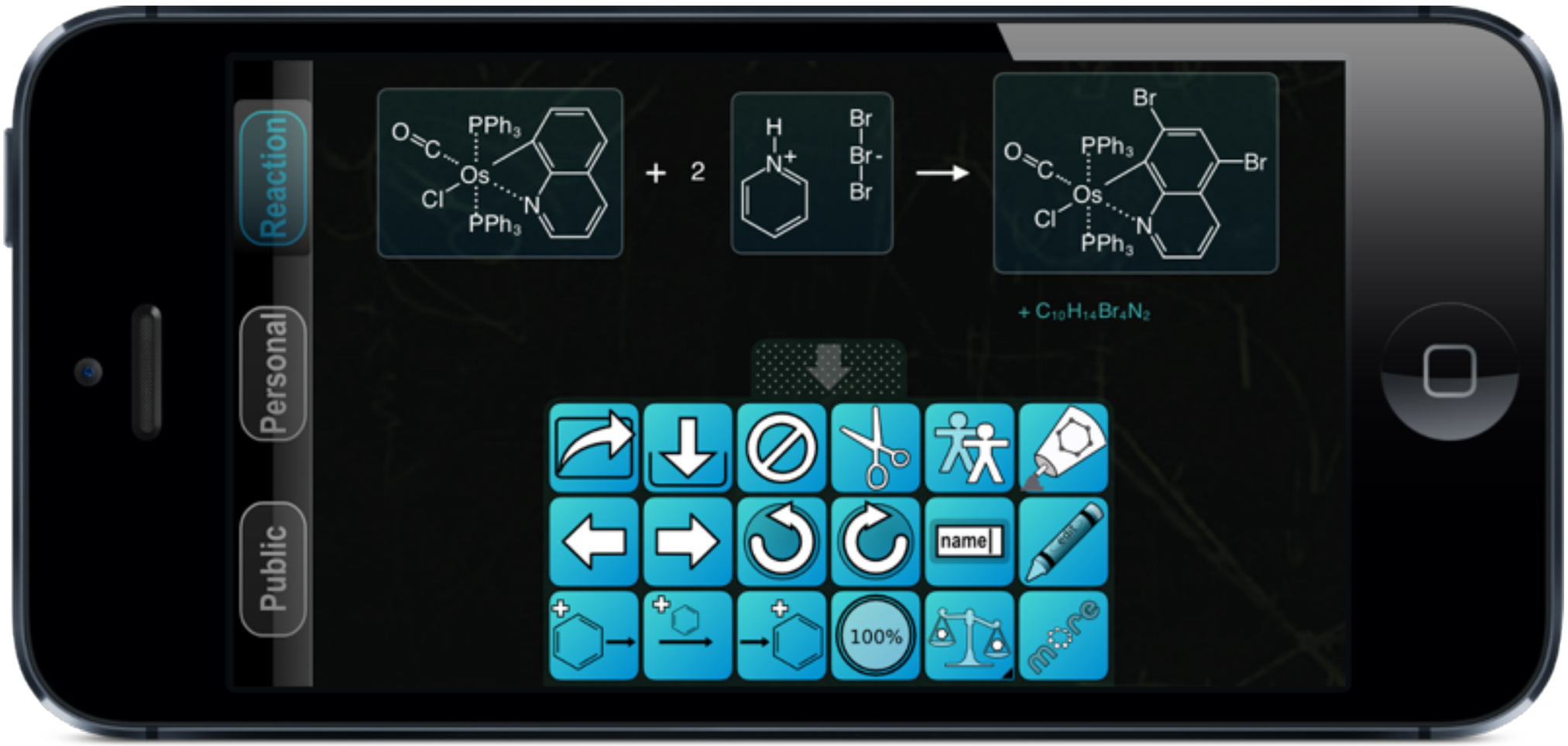
- Interface needs to be:
  - **capable** for complex structures
  - **publication** quality
  - **fast** to use
  - **phone**-size form factor

- Traditional interface paradigm on touchscreen: unusable...





Drawing Reactions



- Components: reuse sketcher, iPhone friendly
- Higher level of markup useful for lab notebooks
- Used by **MMDS**, **Reaction101**, **Yield101**

Assembling DataSheets

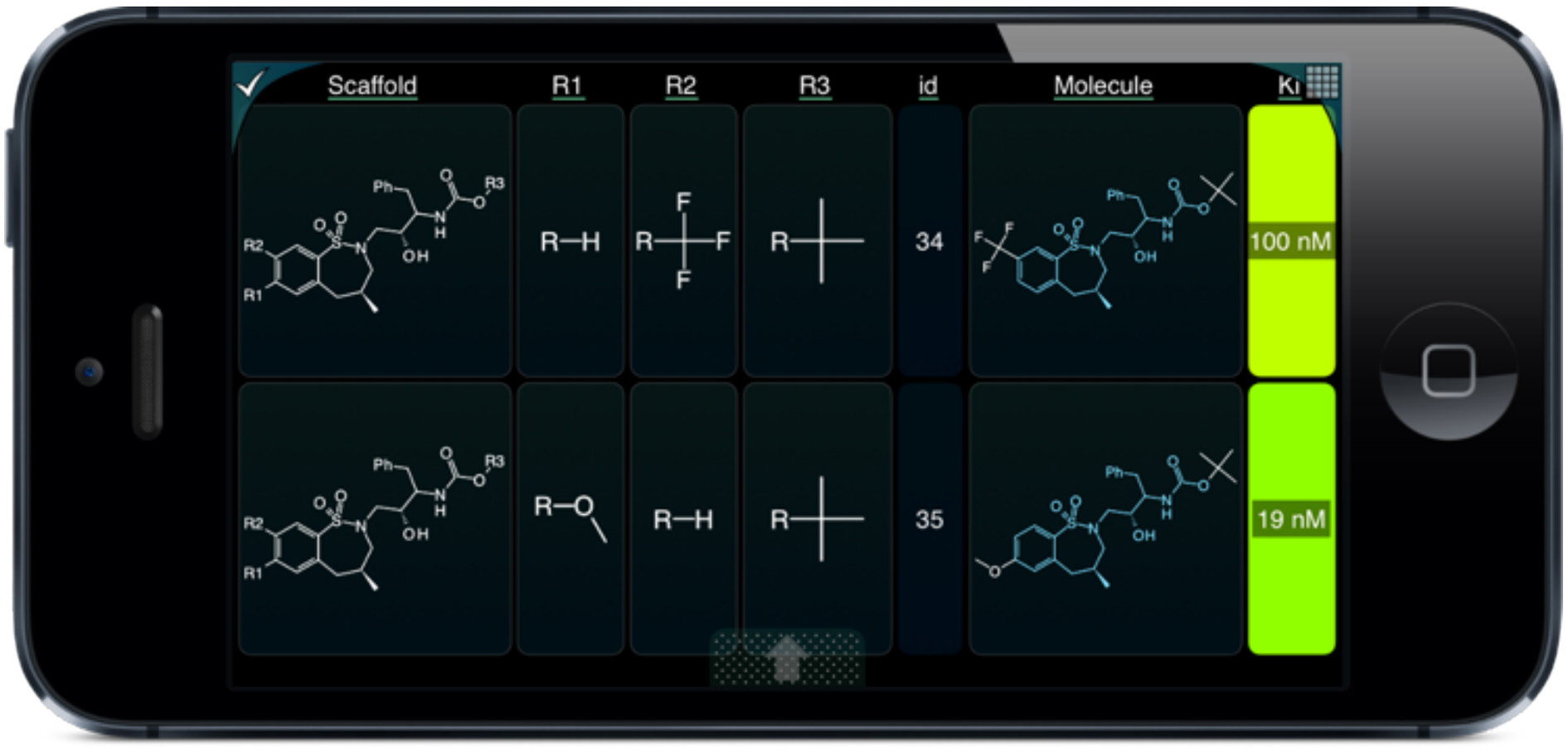
- **MMDS**: *molecular datasheet* unit

  - molecular structures
  - reaction schemes
  - scalar data (text, numbers)

- Table-like structure, roughly equivalent to SDfile/RDfile

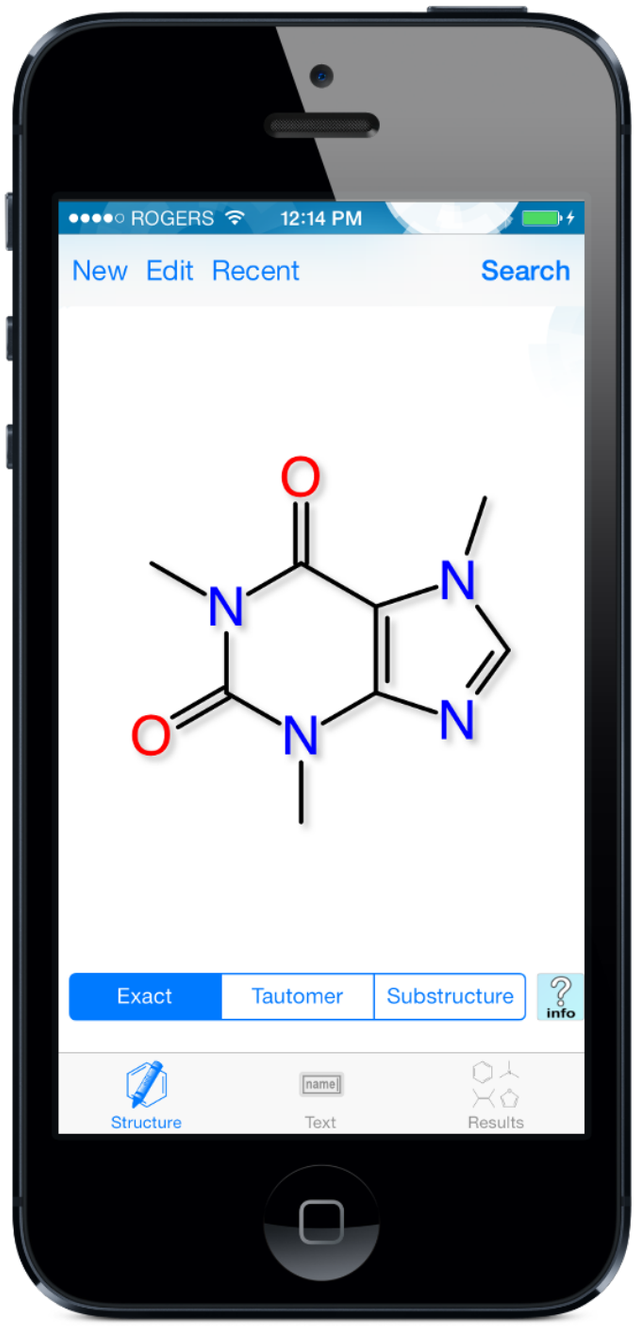
- Operate on individual rows or whole datasheets

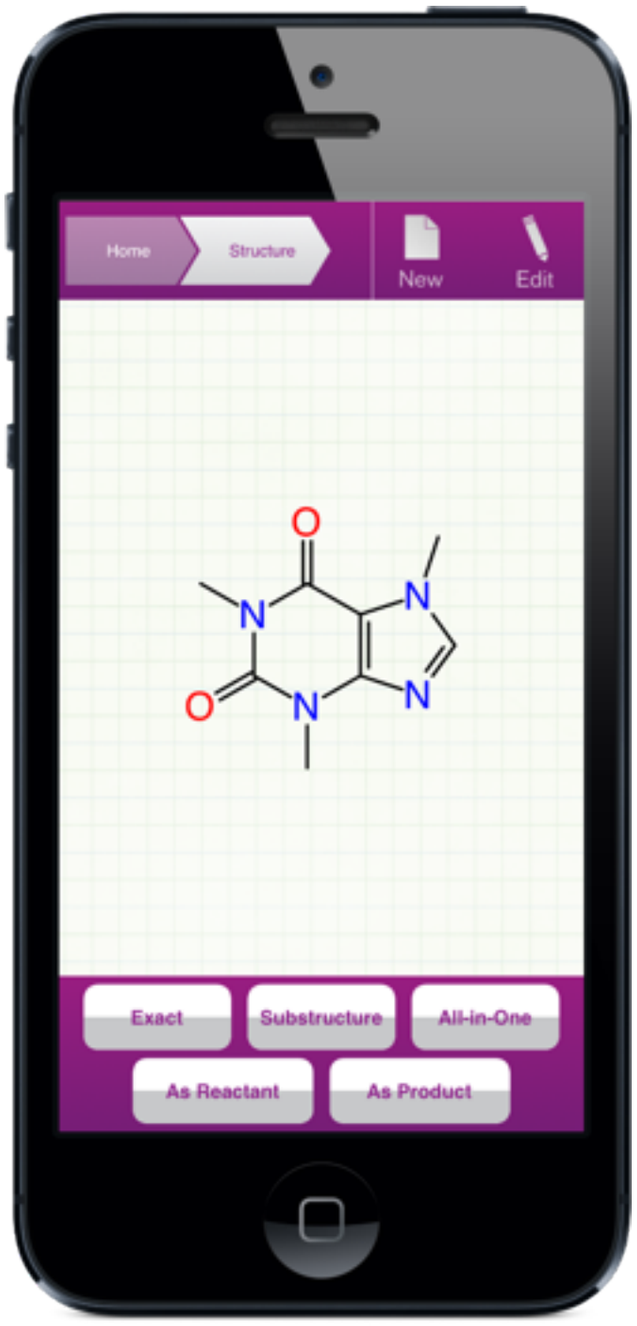
Scaffolds & Substituents



- Scaffold-centric model useful to drug discovery
- **SAR Table** app: optimised for data entry
- And: analysis, visualisation and calculation

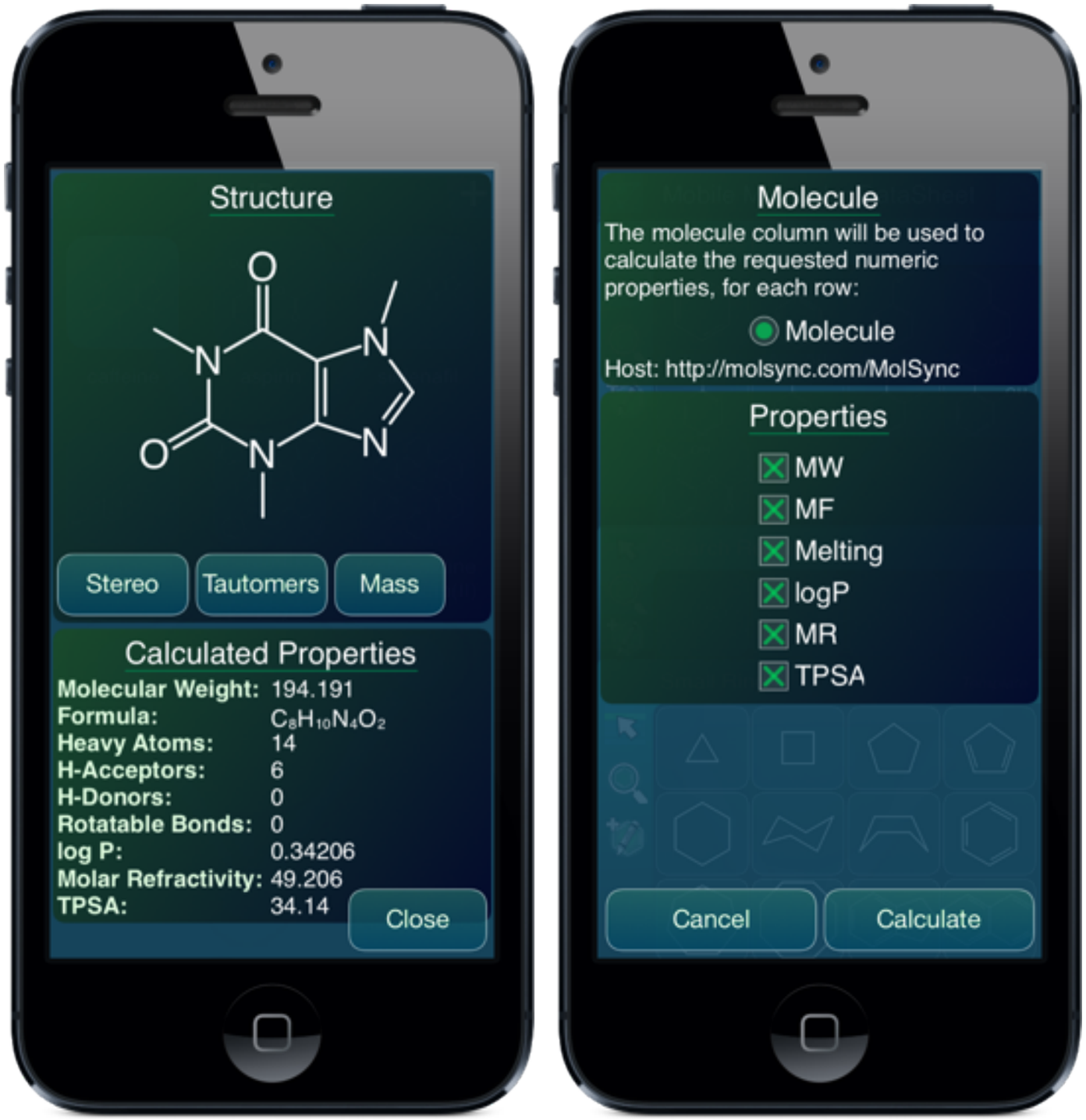
Searching





- App interfaces for structure searching

Calculation



- Structures used to calculate properties
- Single **molecules**, or whole **datasheets**
- Simple calculations: done by app
- Difficult calculations: use webservice

Importing

- Receive recognised filetypes
  - clipboard
  - email
  - web
  - other apps (IPC)
  - webservices
  - remote filesystems
- Documented formats, e.g. SketchEl, MDL, ChemDraw, CML

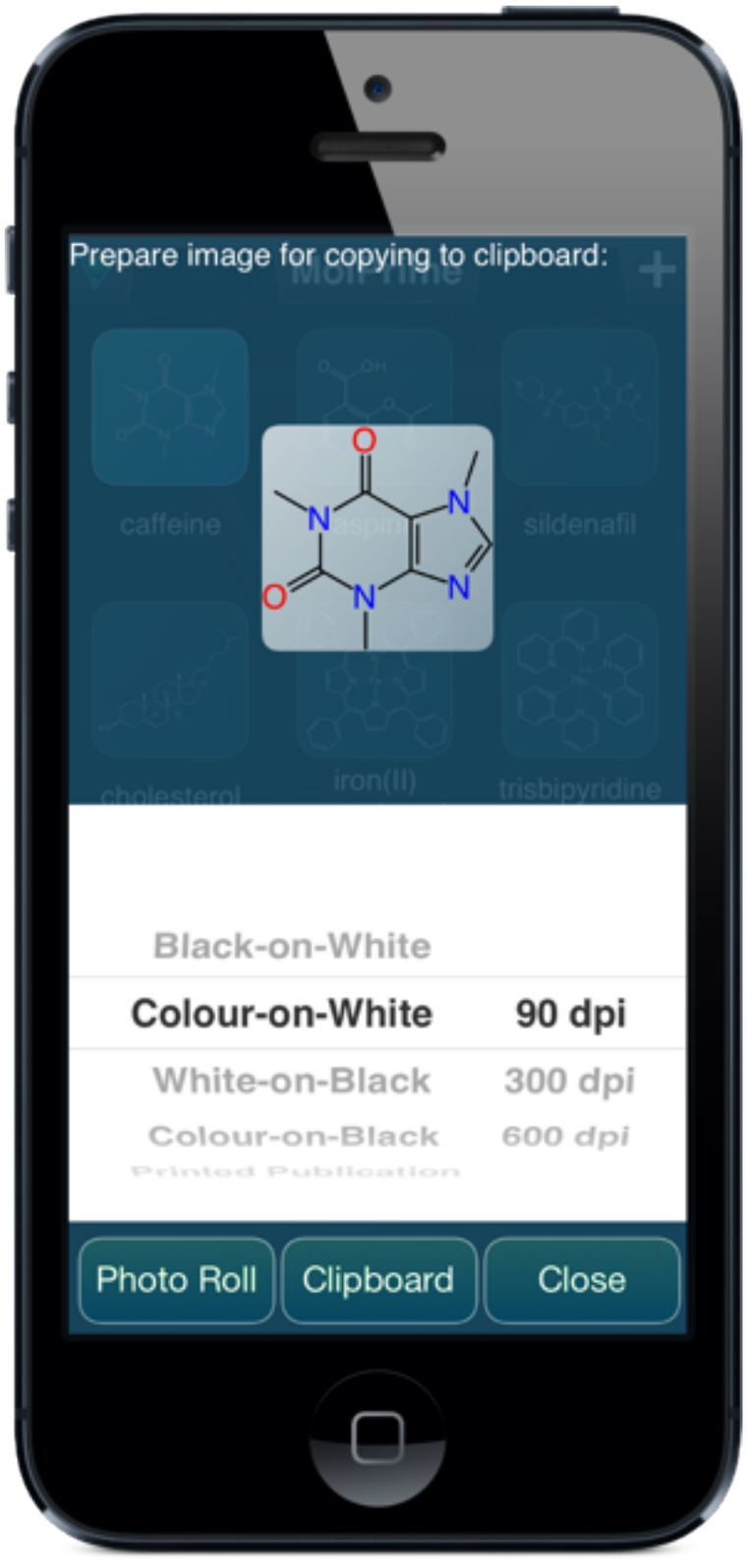
Exporting

- Apps are only as good as their ability to *use* the data

- Export actual data using cheminformatics formats

  - clipboard
  - email
  - web sharing
  - twitter
  - webservices
  - remote filesystems

Graphics



- Prepare bitmapped graphics for:
  - clipboard
  - photo roll
  - email
- Can incorporate into other apps on the device (e.g. *Keynote*)
- Vector graphics: SVG, EPS, Microsoft Word & Excel

Sharing

- Share data publicly or privately
- Upload to a webservice
  - store raw data
  - dynamically render
  - hosted by *molsync.com*
- Tweet directly from the app: share the link

Real world workflow: searching for new Tuberculosis drugs

Dr. Alex M. Clark

October 2013

http://molmatinf.com

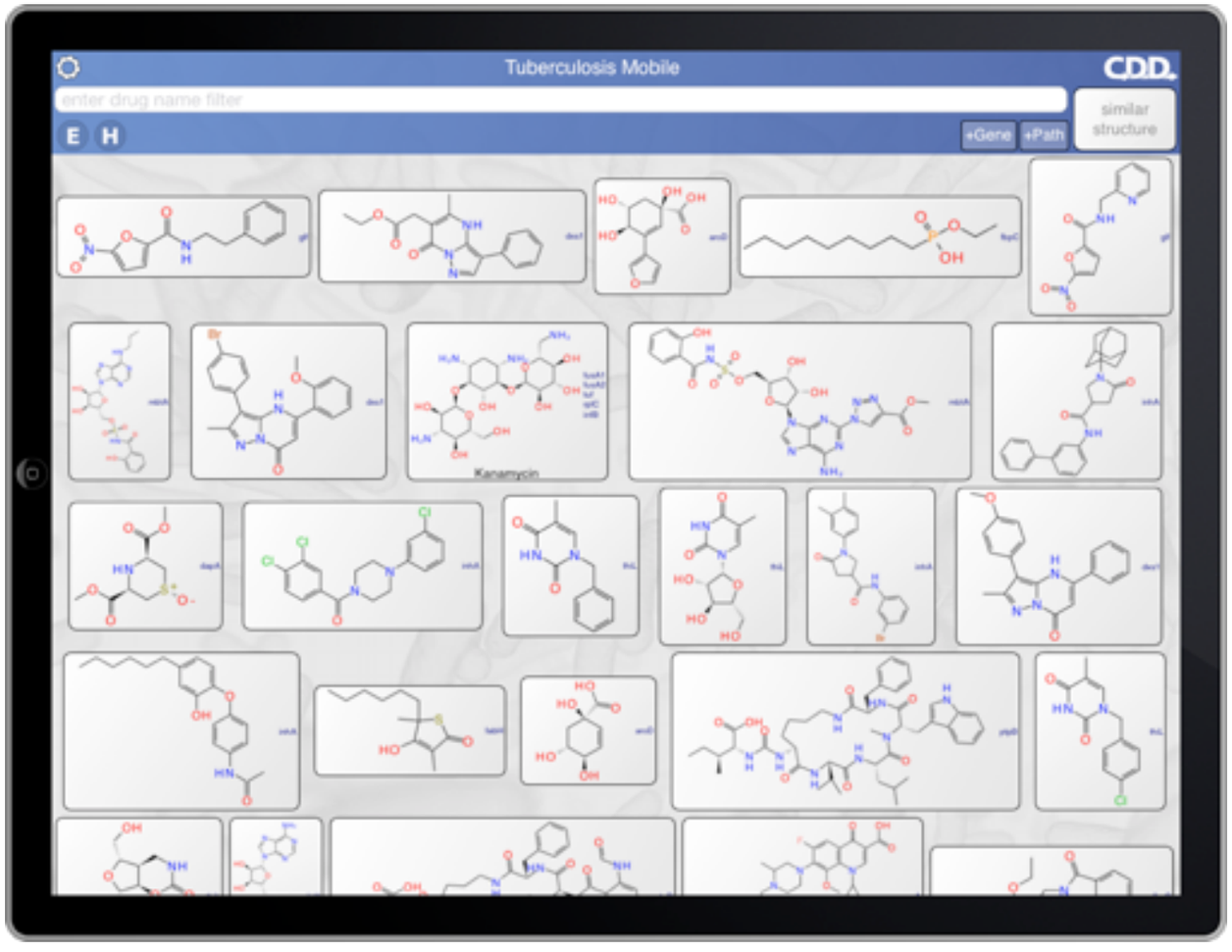
Outline

- Workflow is representative of lead discovery
- Searching for new *Mycobacterium tuberculosis* drugs
- Open data, open science, collaborating with experimentalists
- Workflow involves:
  - mobile apps
  - cloud-based webservices
  - traditional software

http://molmatinf.com/venice.html

TB Mobile

- Curated data for **tuberculosis** drugs
- iOS & Android
- CDD, NIAID grant, Sean Ekins

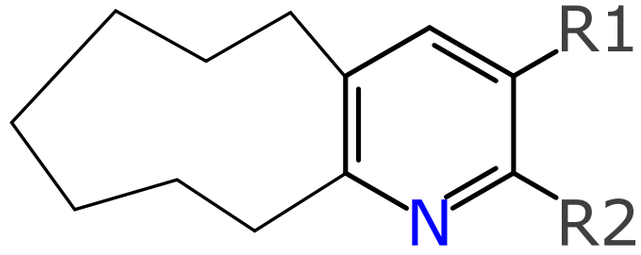


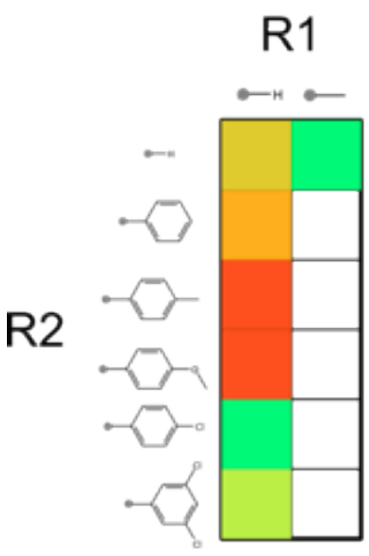
- App presents a browseable list of structures
- Can filter by target, sort by structural similarity
- Visual data exportable....

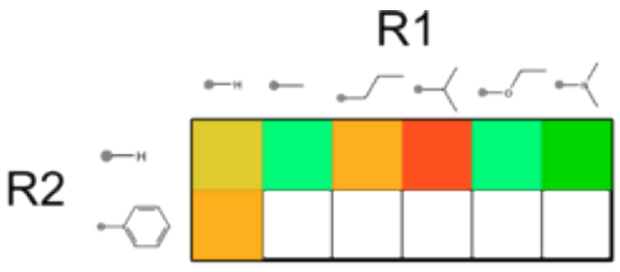
A Good Scaffold

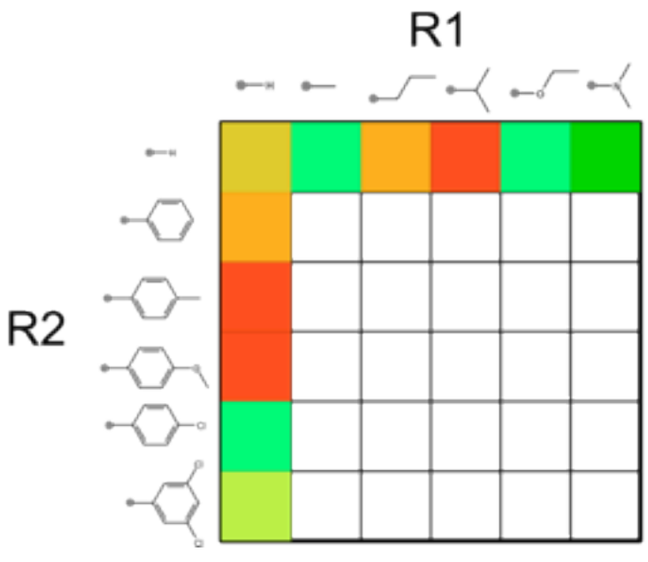
- Objective: use the data from **TB Mobile** to find a scaffold, then use to infer gene target from a larger collection of SAR data

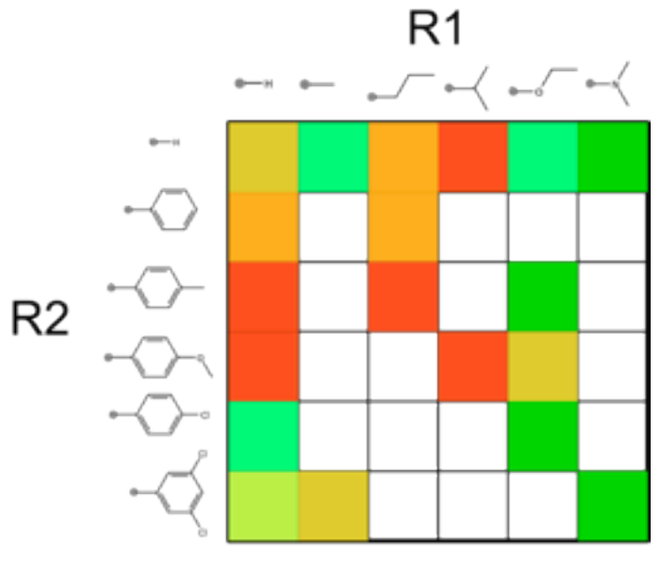
- What makes a good scaffold?





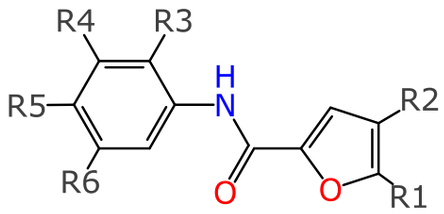


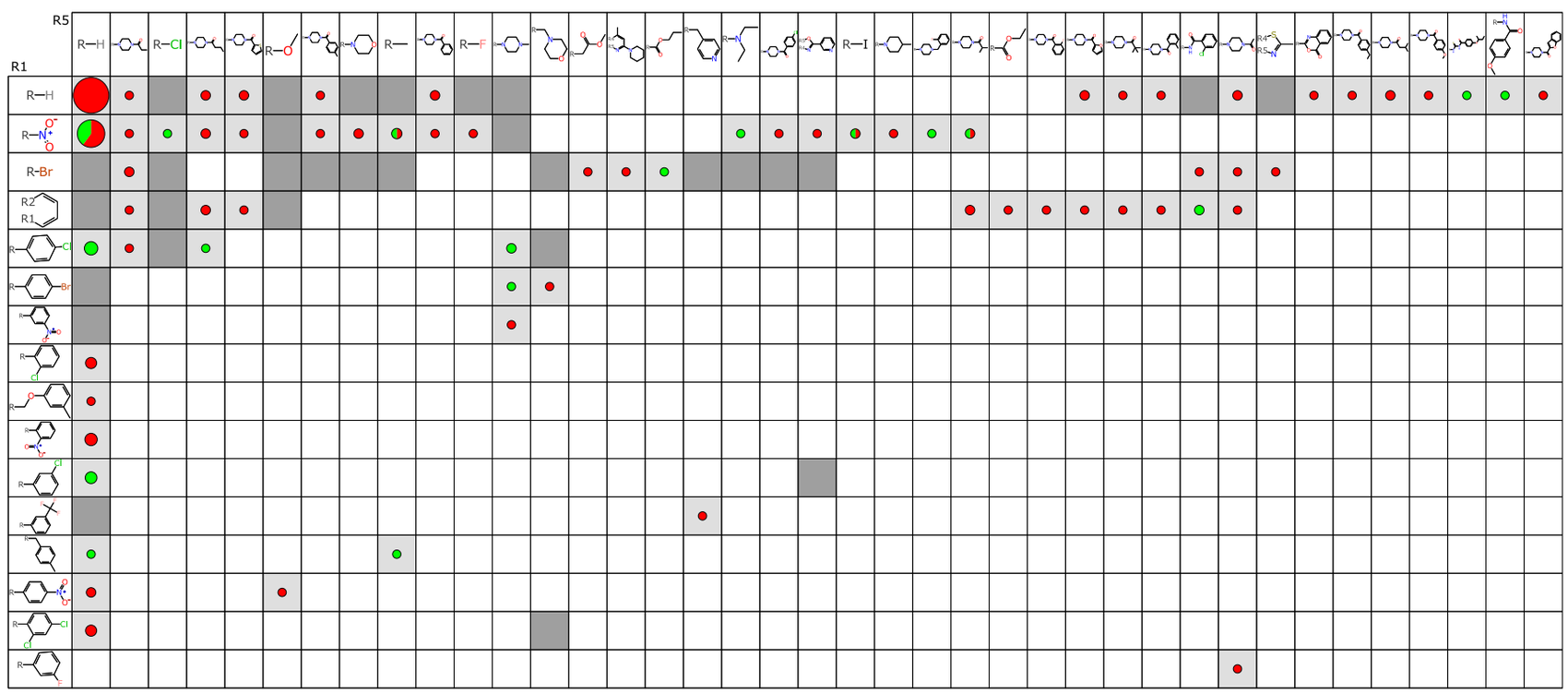




Methodology

- Preliminary work done with desktop software: *com.mmi*
- Fragment **TB Mobile** structures, scaffold-like
- Perform scaffold-substructure vs. 7000 *in vitro*
- Derive R-groups, tidy, present graphically, browse...

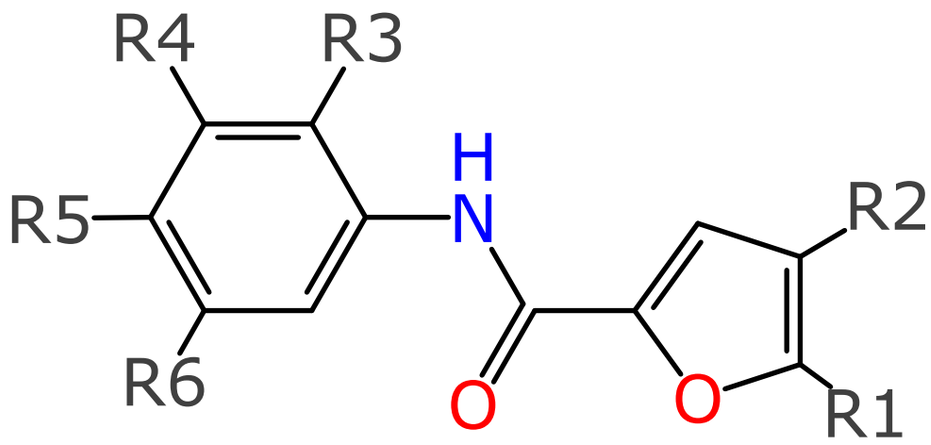




http://molmatinf.com/venice.html

Source Materials

- Scaffold:

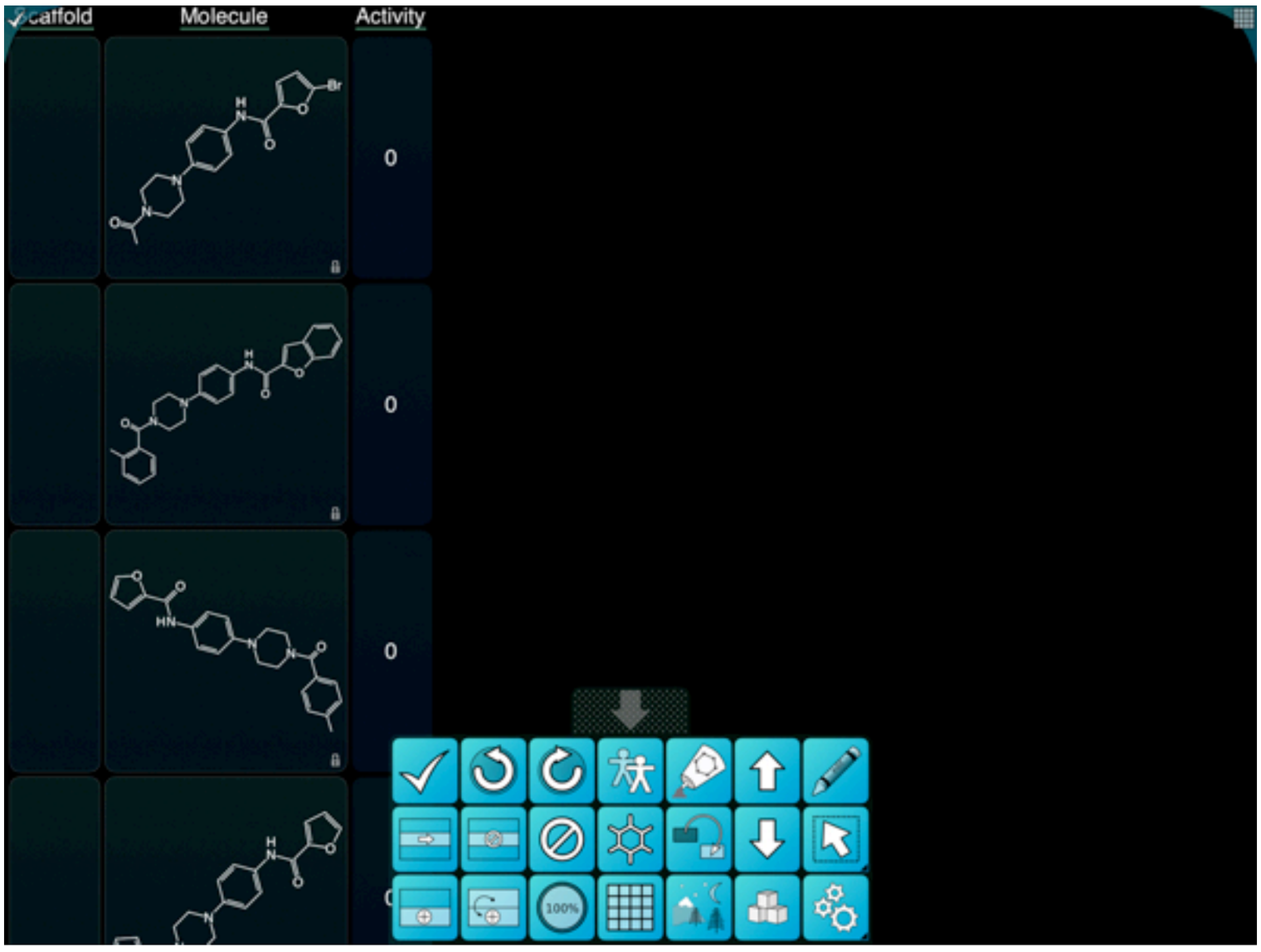


- Scaffold origin: inhibitor of **Glf** target
- 87 molecules with *in vitro* activity (yes/no)
- Scaffold seems to elicit an activity pattern
- Next step: load it into the app ecosystem...

Importing

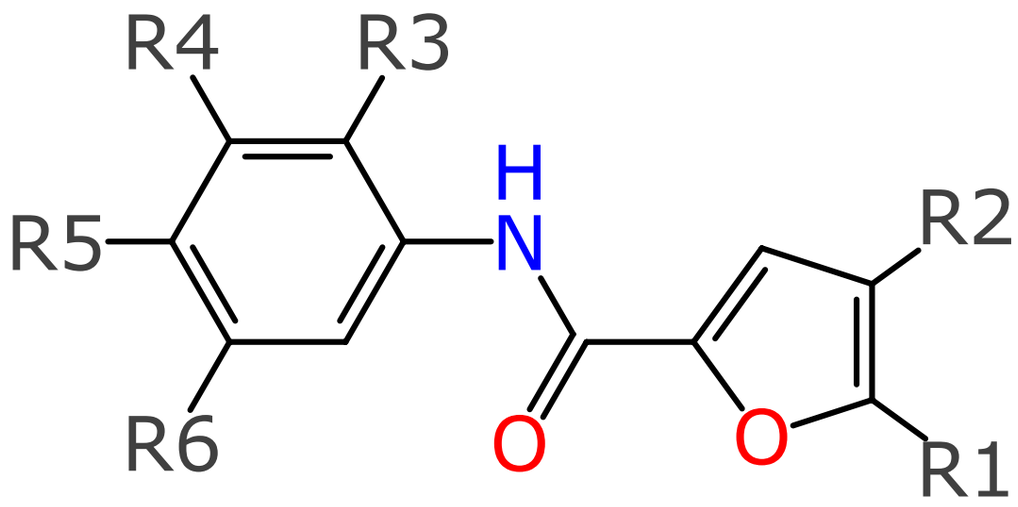
- Email attachment to self: open in **SAR Table**

Raw DataSheet



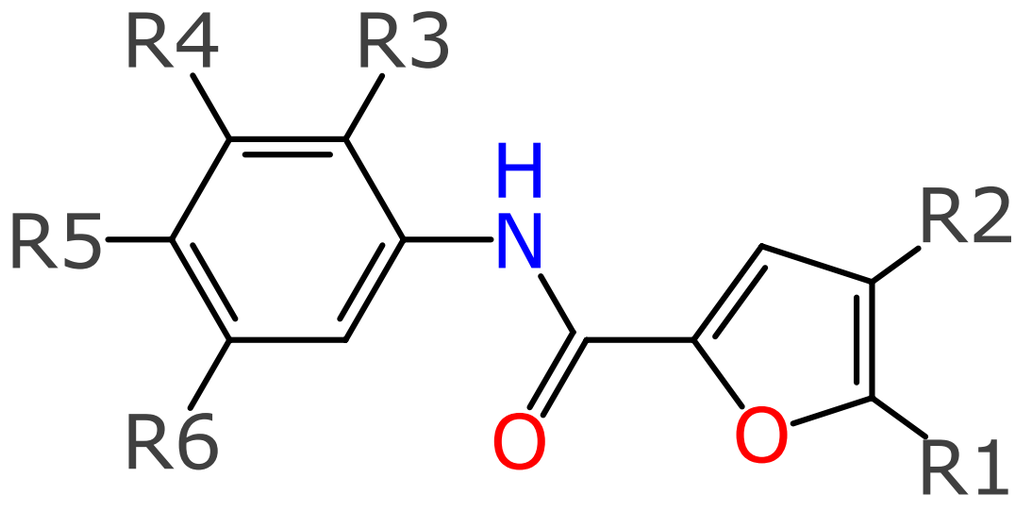
- Just **Molecule** and **Activity**...

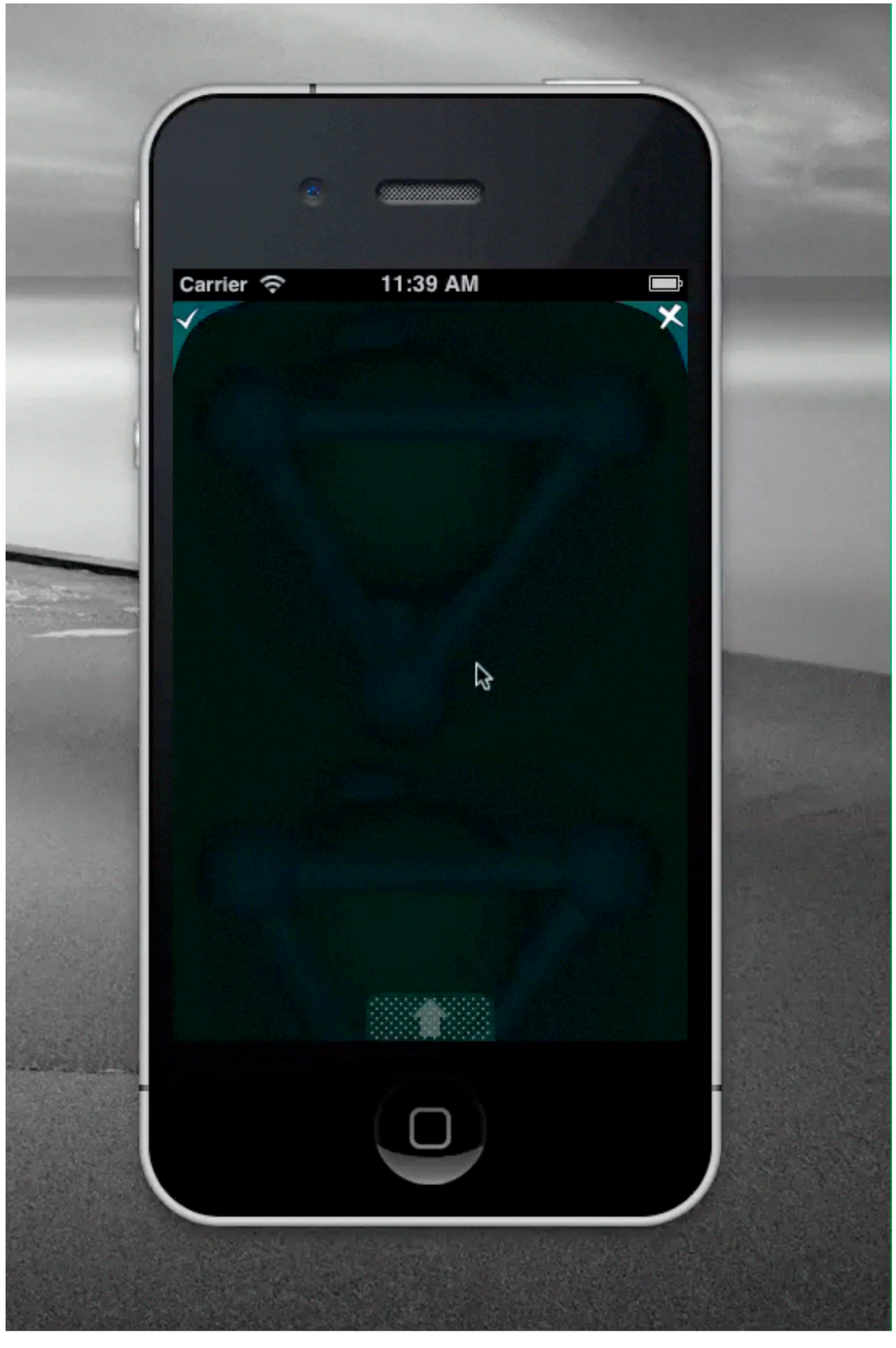
Draw Scaffold



- Select first row
- Use built in sketcher to draw scaffold

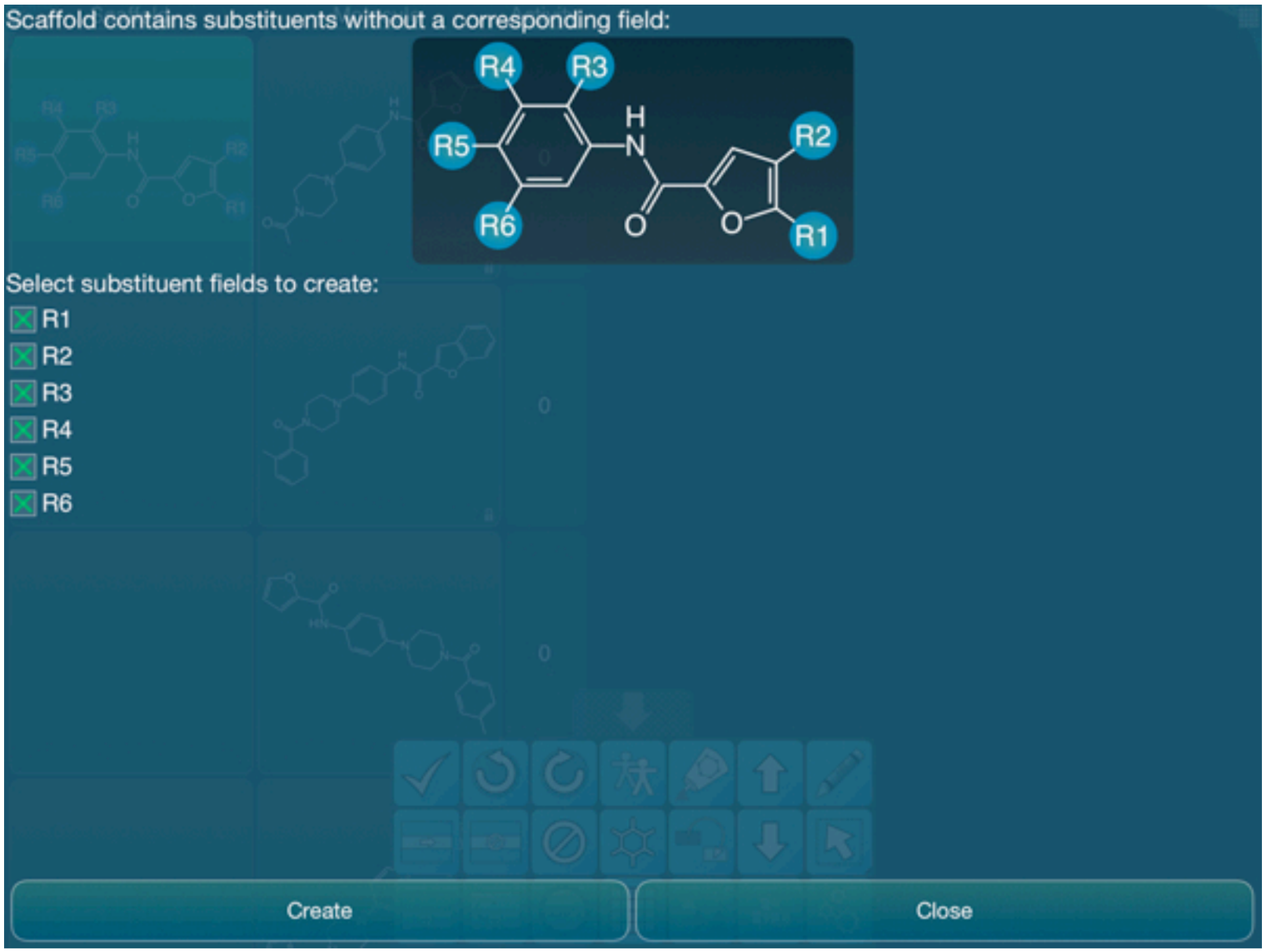
Draw Scaffold





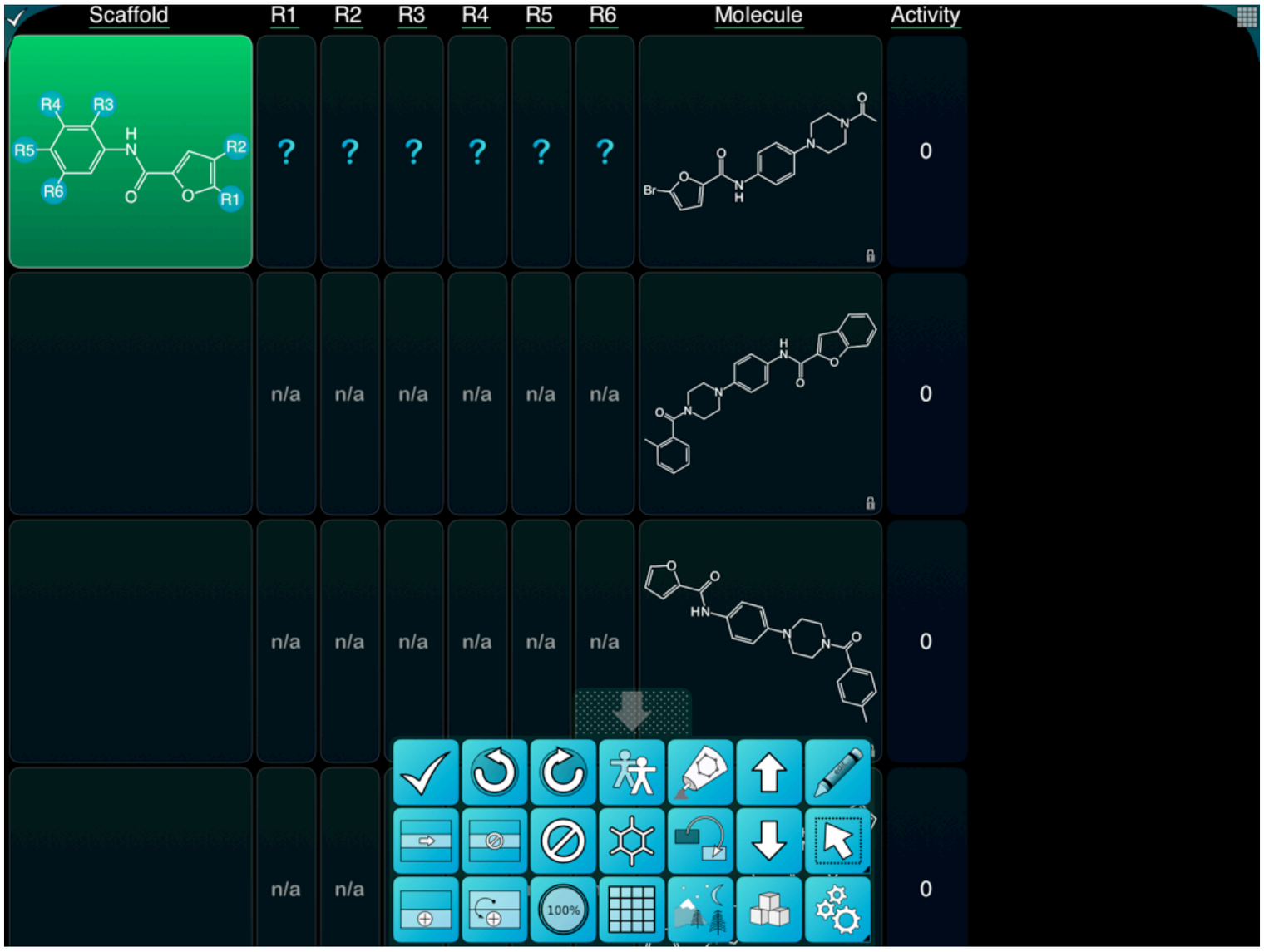
- Select first row
- Use built in sketcher to draw scaffold

Substituent Fields



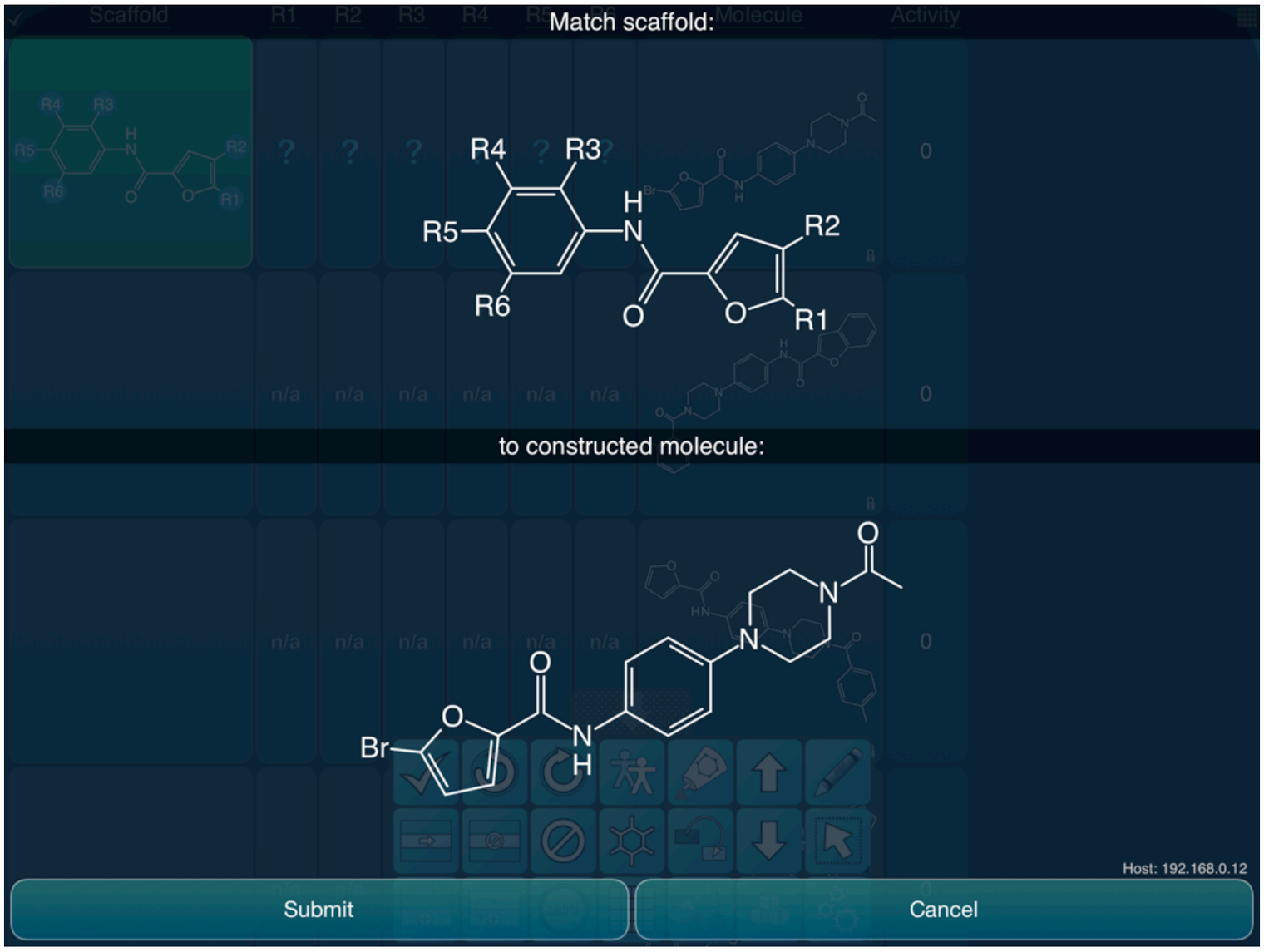
- R-group labels: offer to create fields

Scaffold & Molecule



- Need to assign substituents: automatically

Scaffold Matching

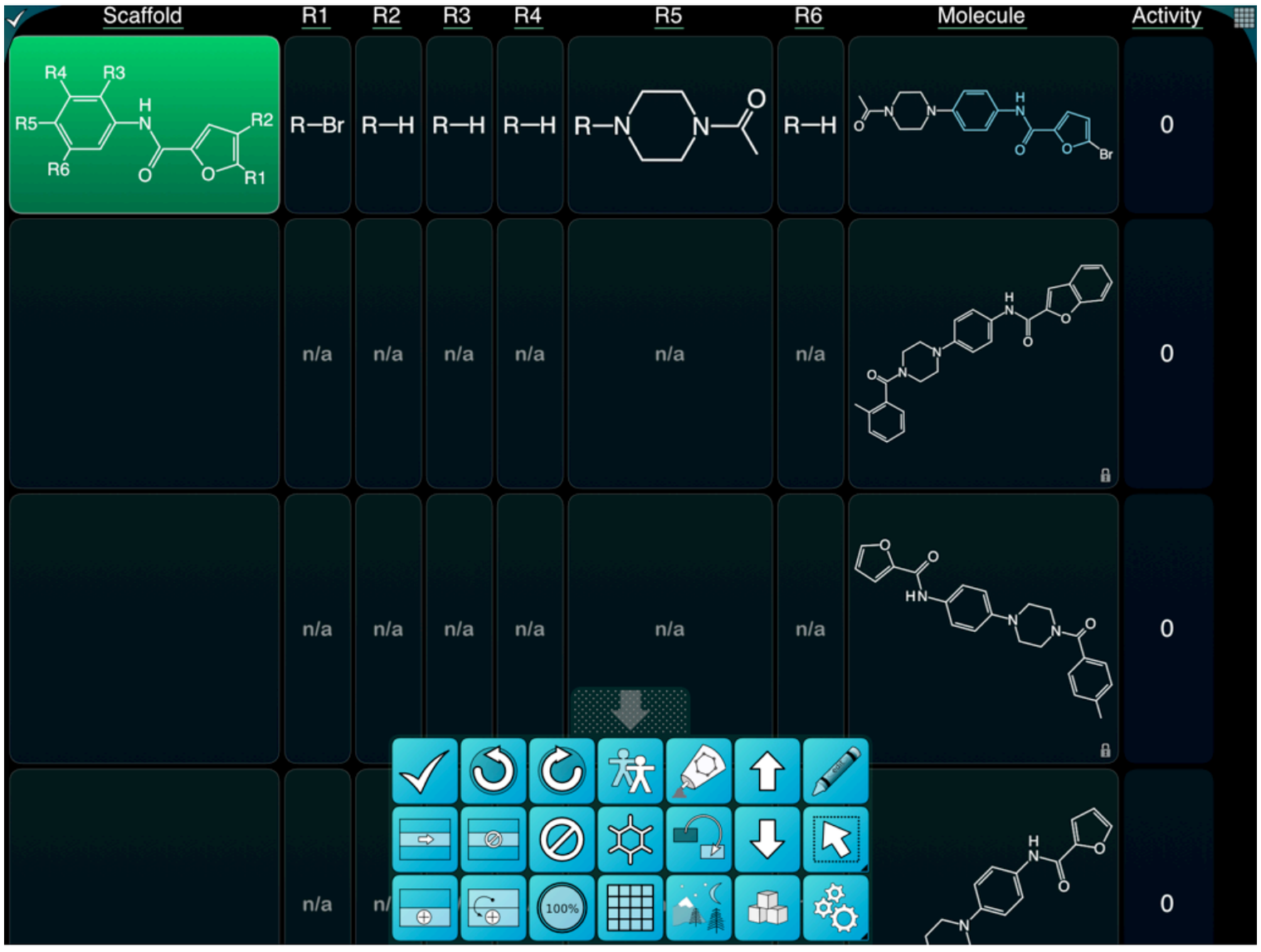


- Perform single **scaffold**:**molecule** match

Scaffold Matching

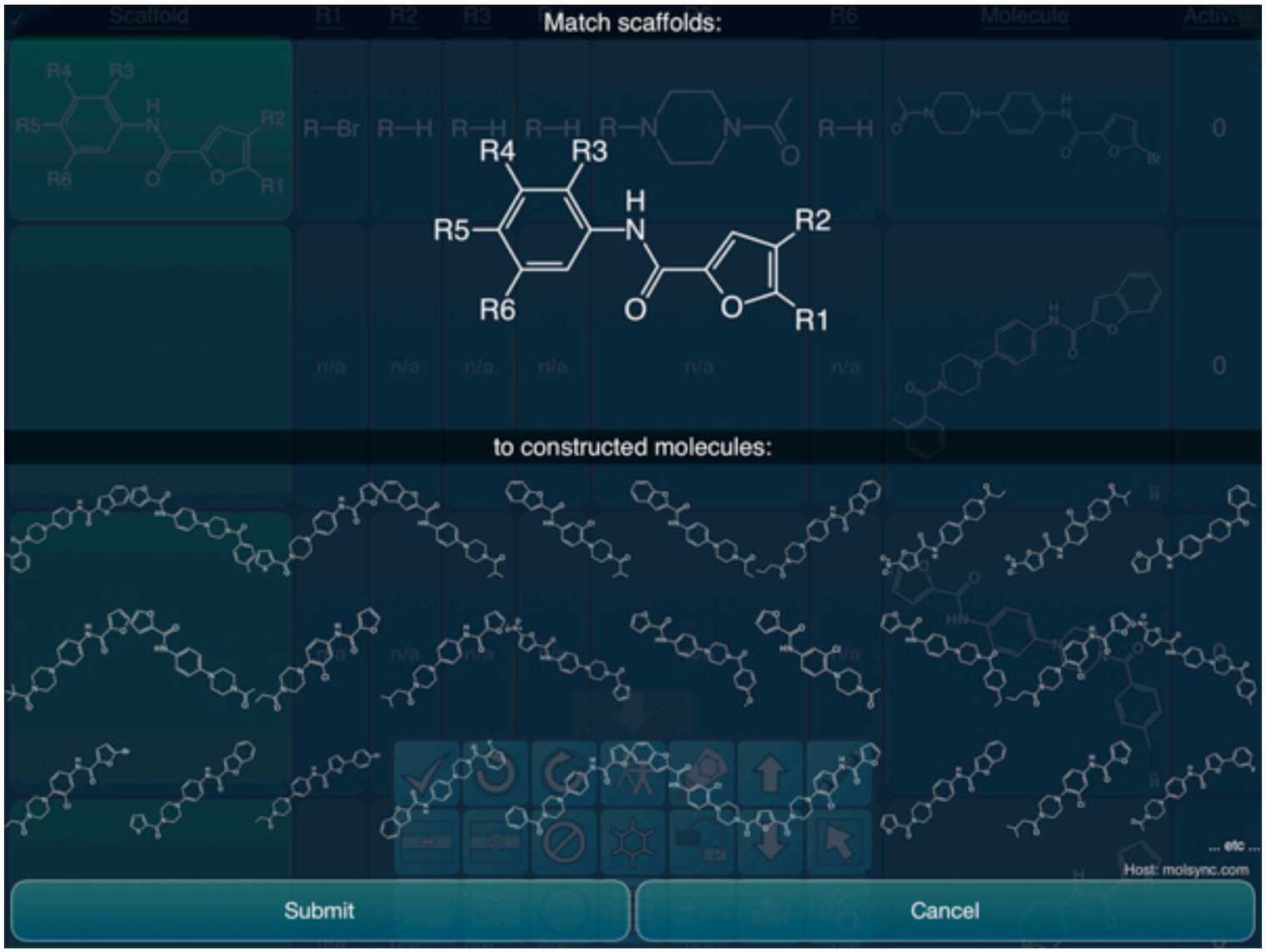
- Perform single **scaffold**:**molecule** match

Scaffold Matching



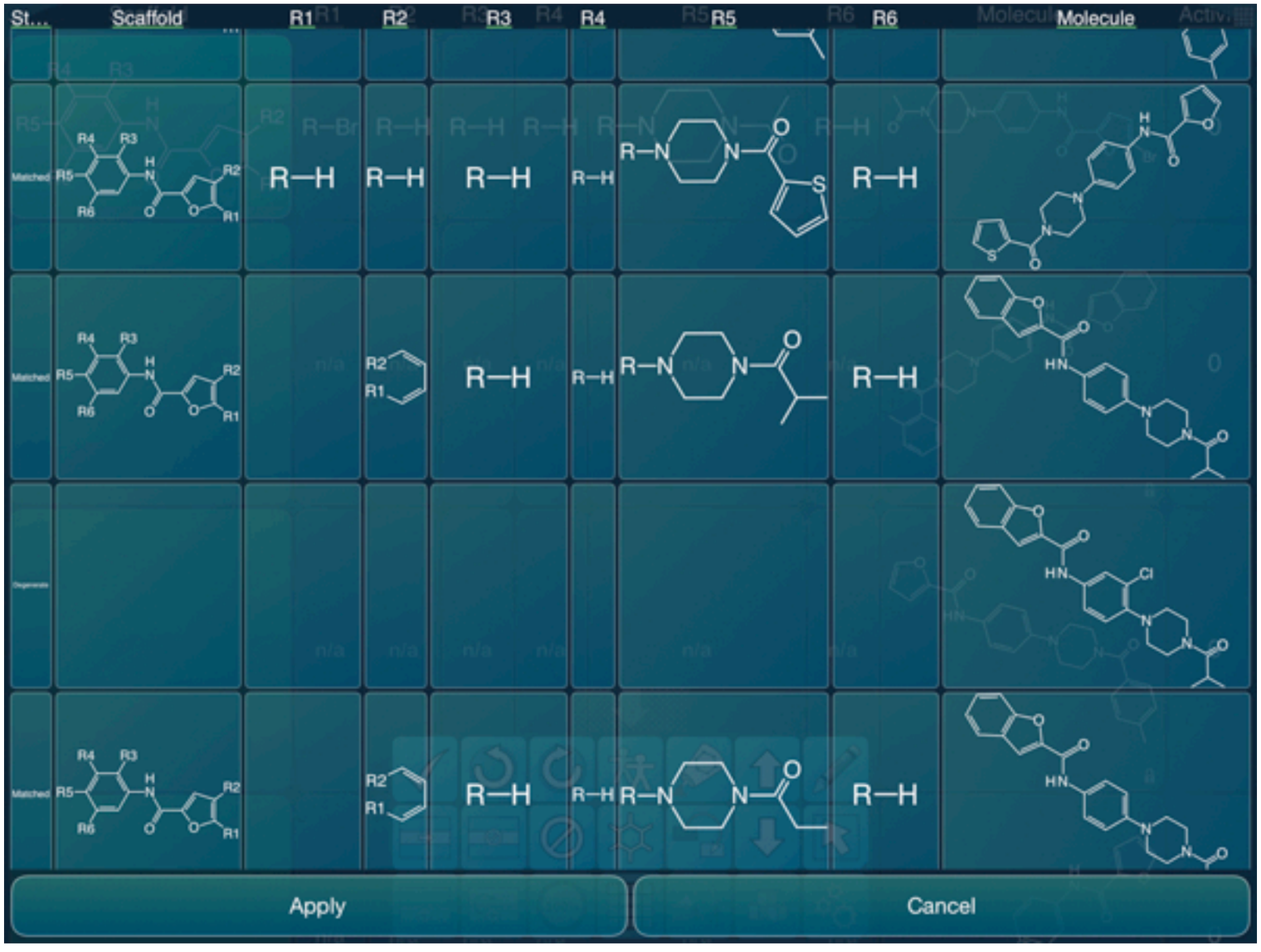
- Perform single **scaffold**:**molecule** match

Bulk Matching



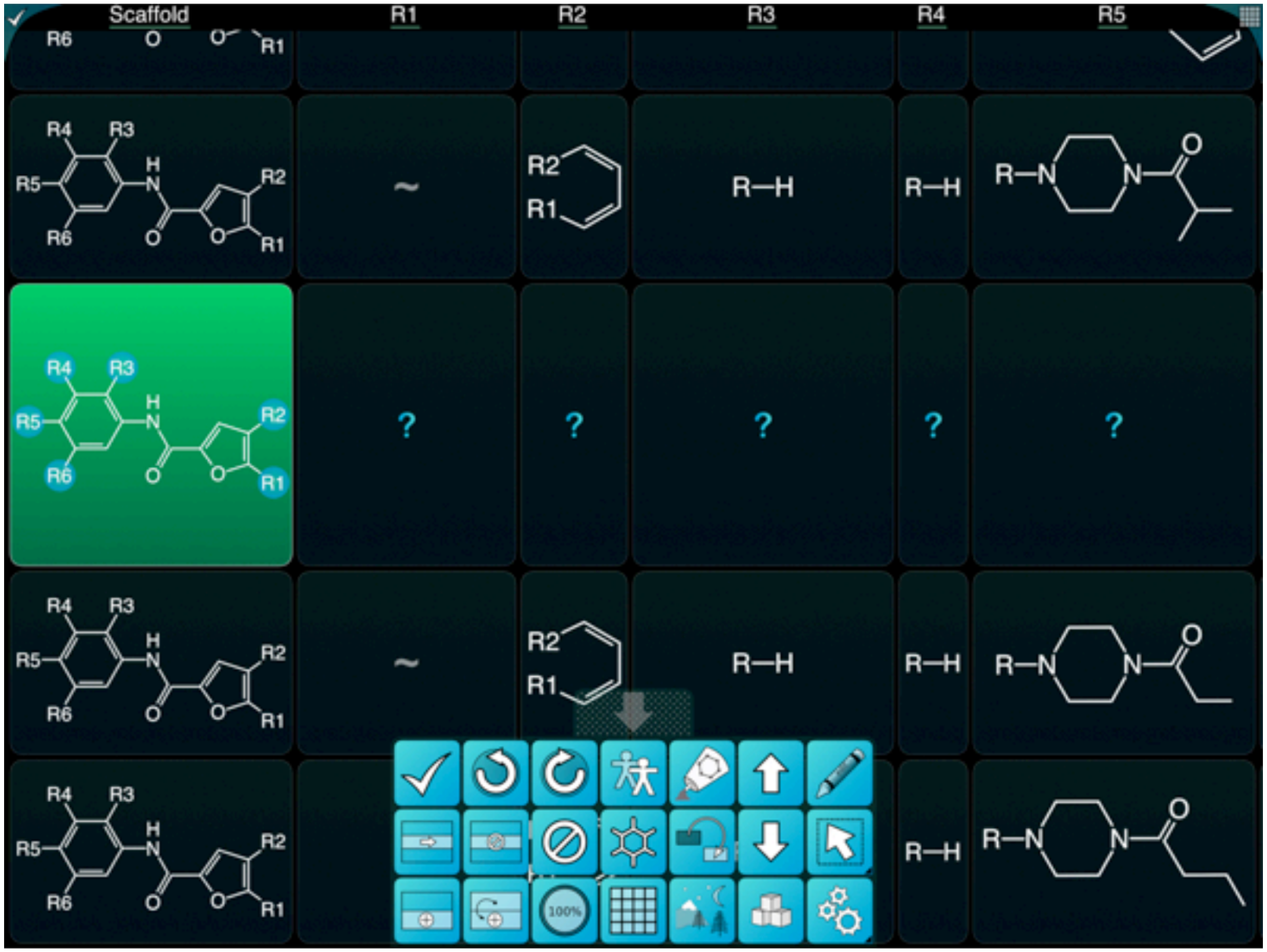
- Duplicate scaffold, match to all molecules

Bulk Results



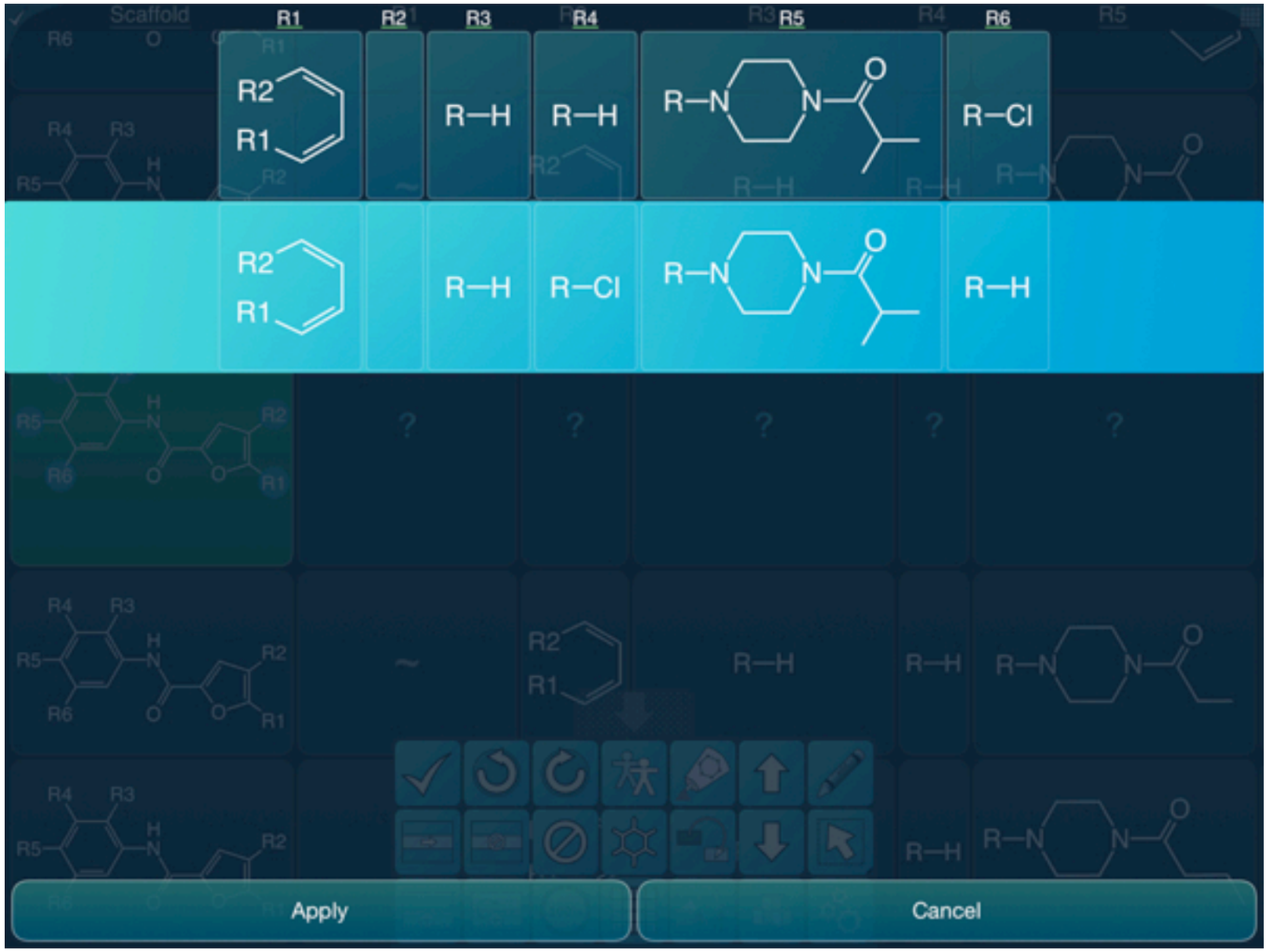
- Only unambiguous results are provided

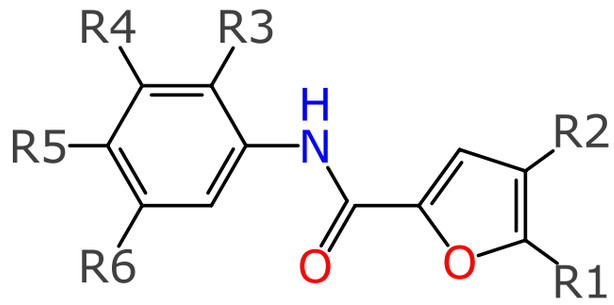
Bulk Results

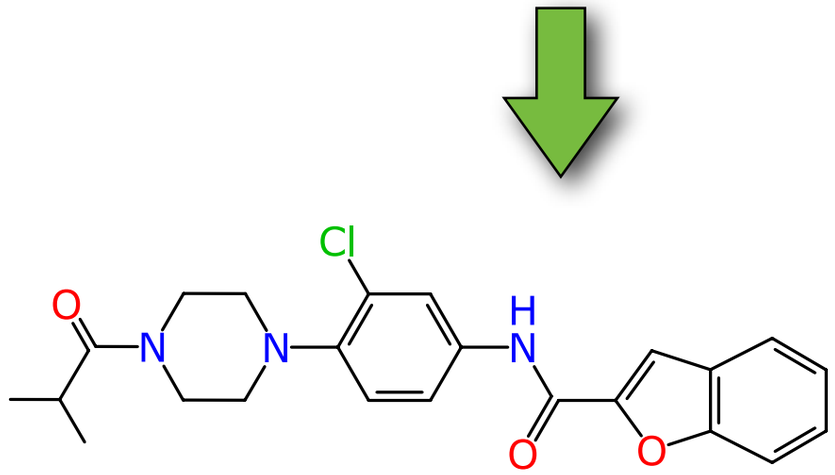


- Only unambiguous results are provided

Symmetry/Degeneracy

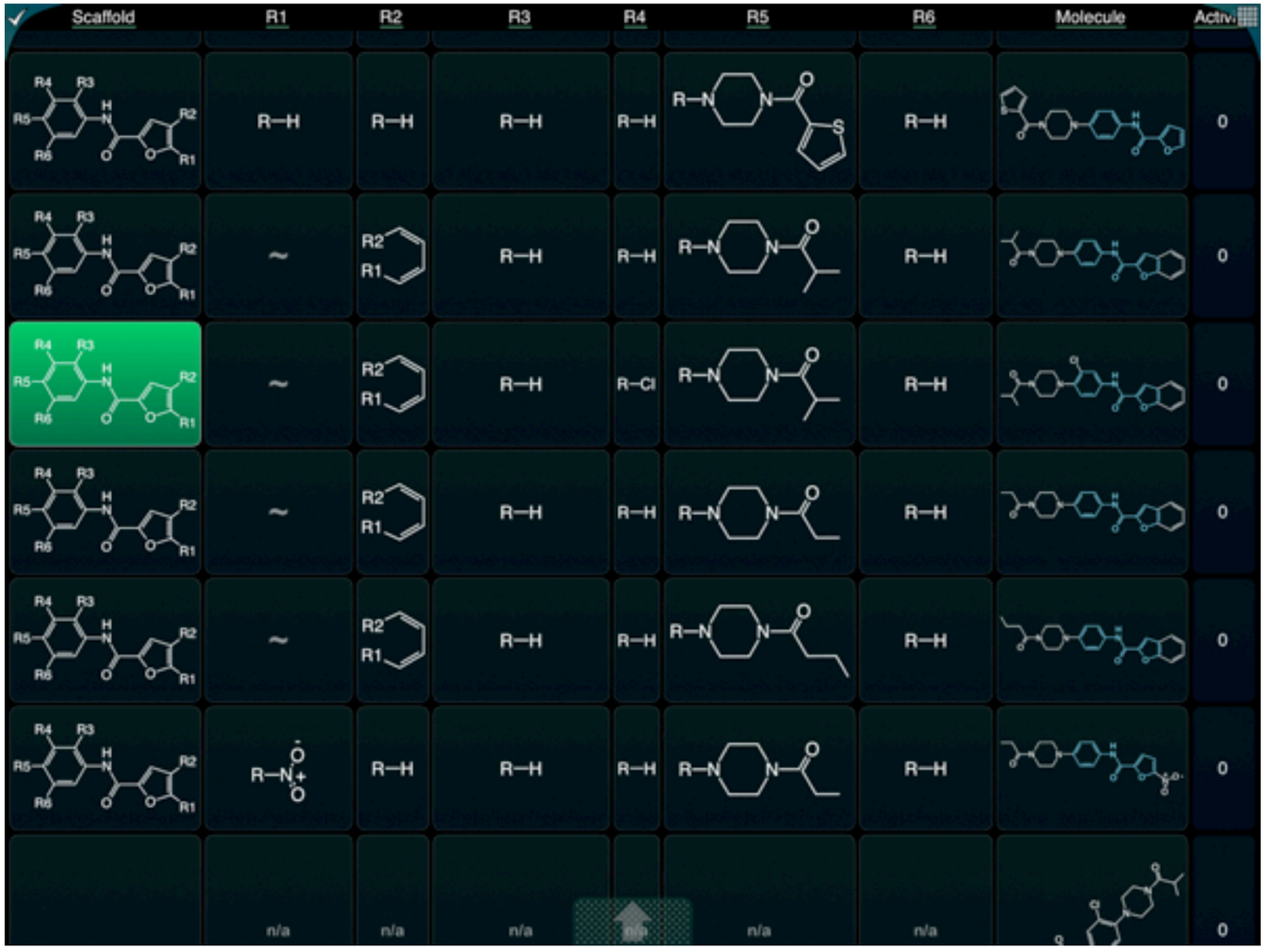


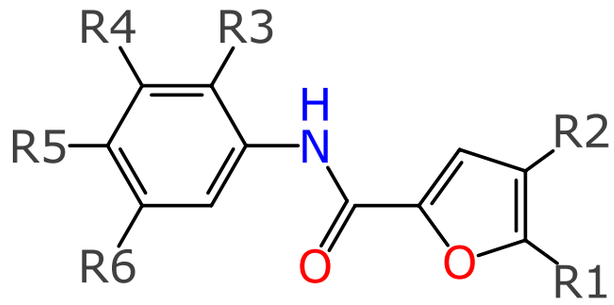


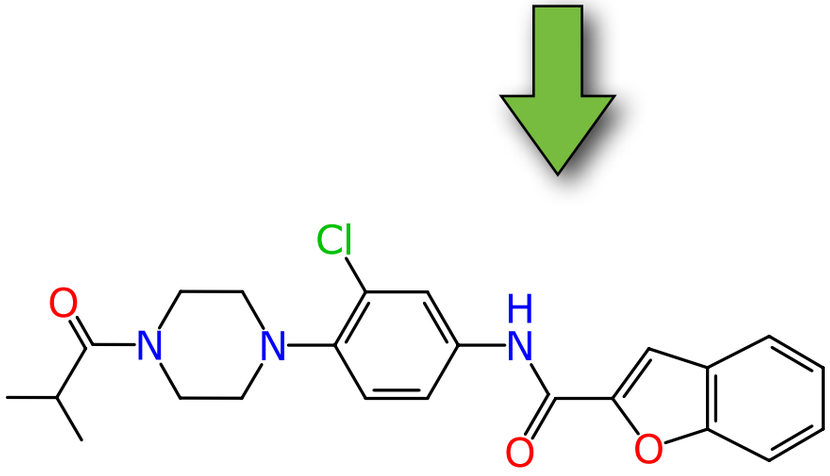


- Two possible results: note degeneracy **R4**,**R6**

Symmetry/Degeneracy





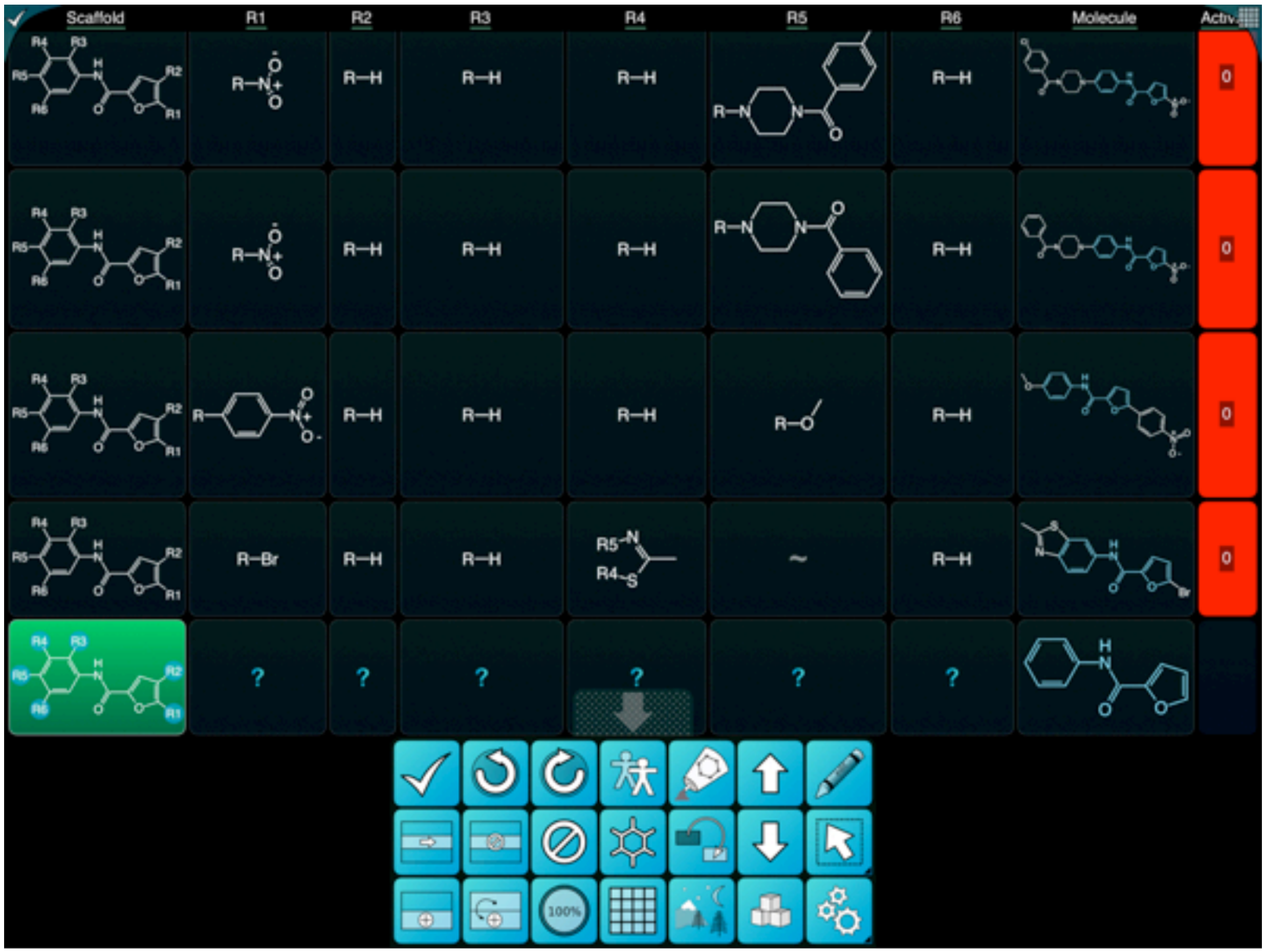


- Two possible results: note degeneracy **R4**,**R6**

Finishing Assignment

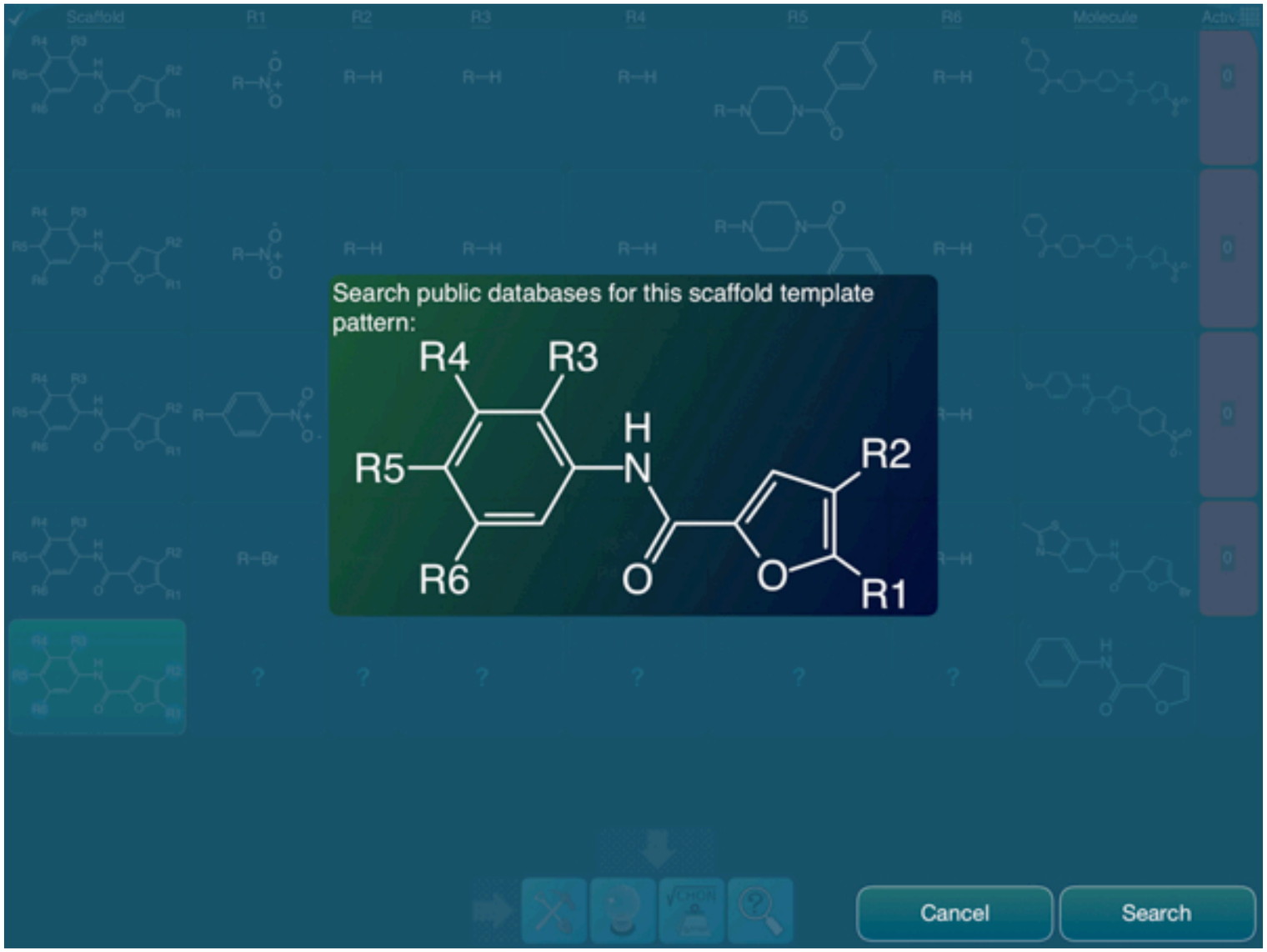
- Single & bulk scaffold matching: semi-automated assignment, human tiebreakers
- Have 87 compounds, 21 active against TB
- Based on scaffold from a binder of **Glf** enzyme (essential for mycobacterial growth)
- What next?
- Look for other compounds: same scaffold

Searching



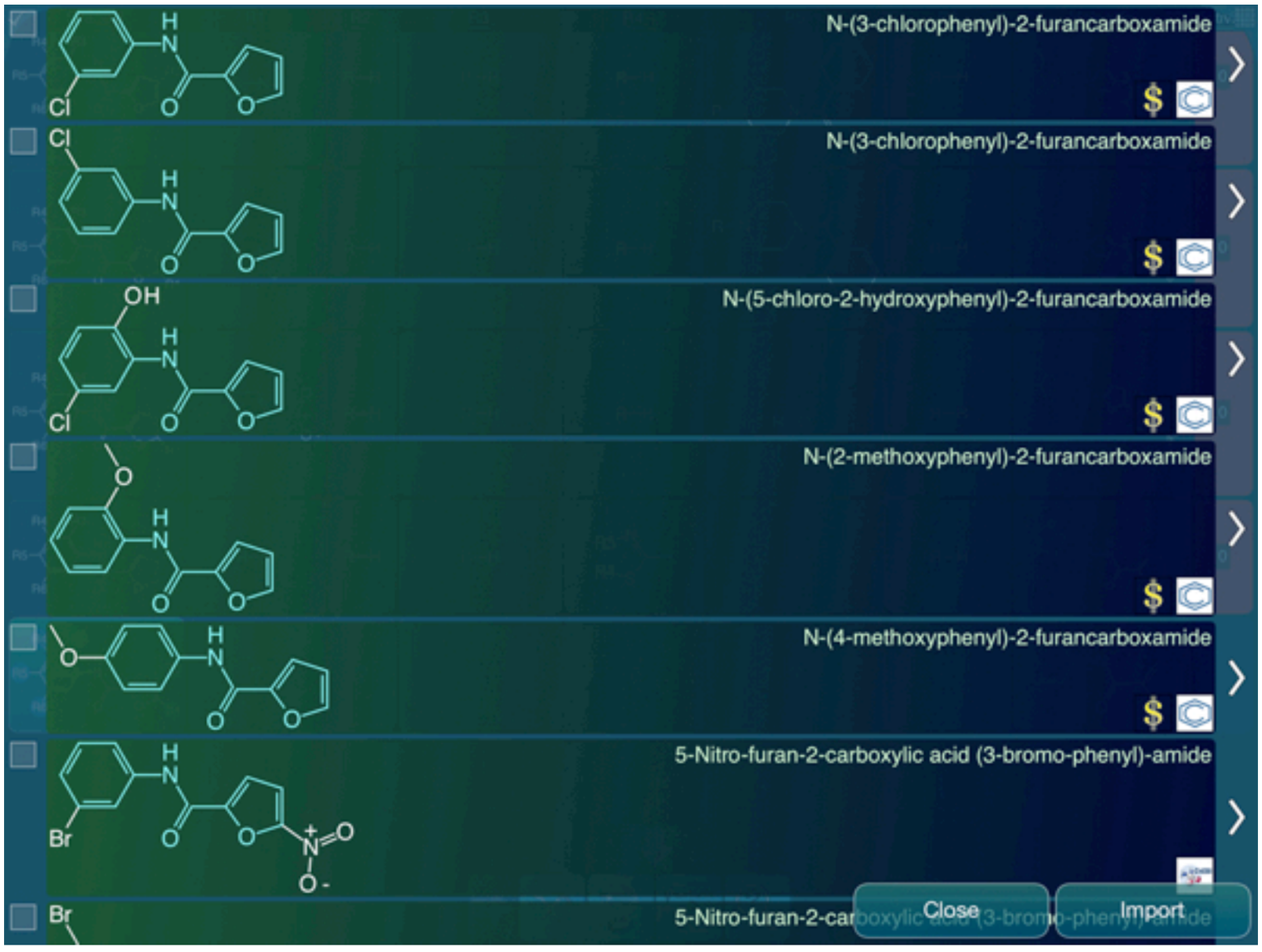
- Search public databases using template

Searching



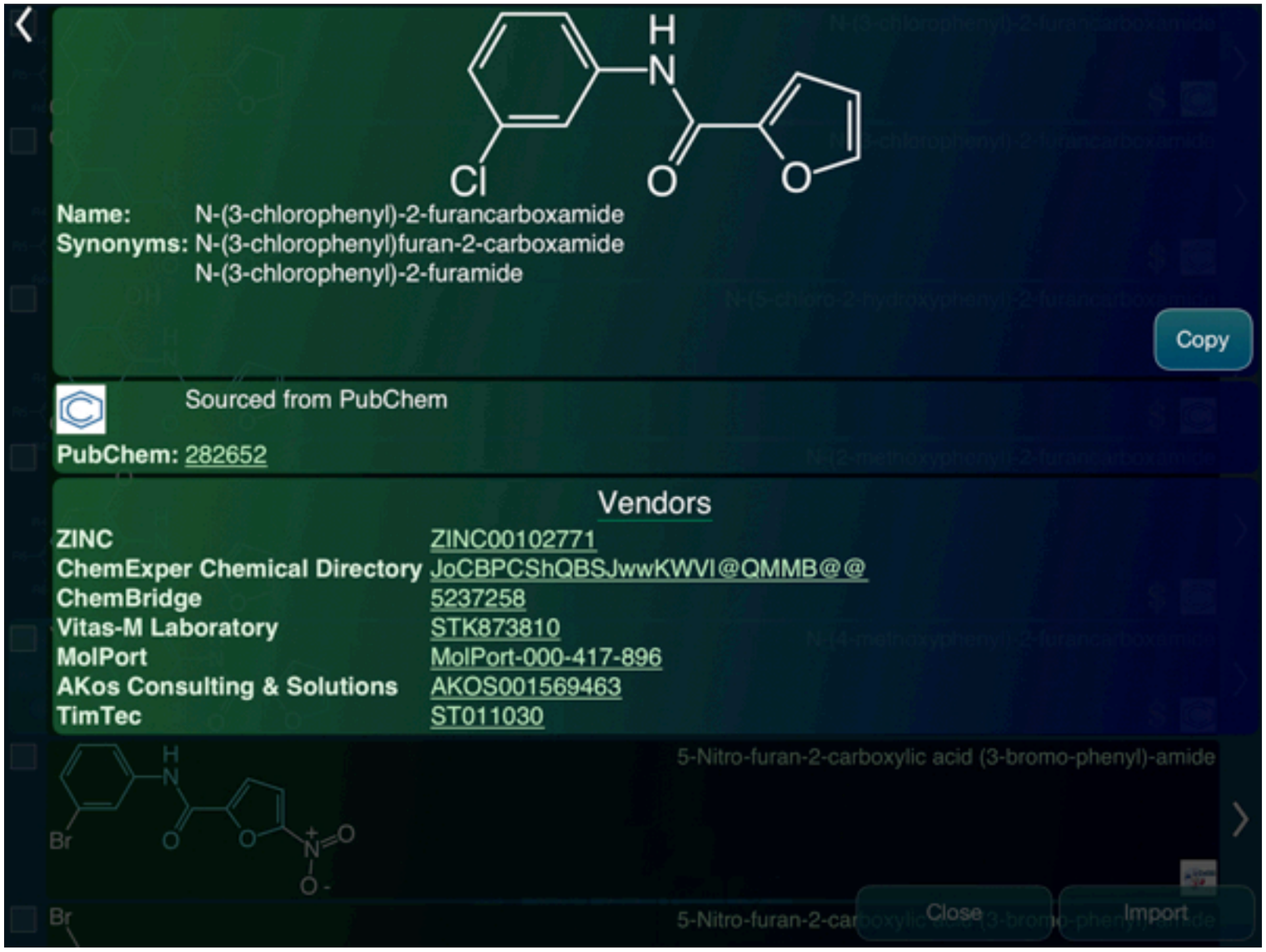
- Search public databases using template

Search Results



- Results come back with scaffolds assigned

Sources

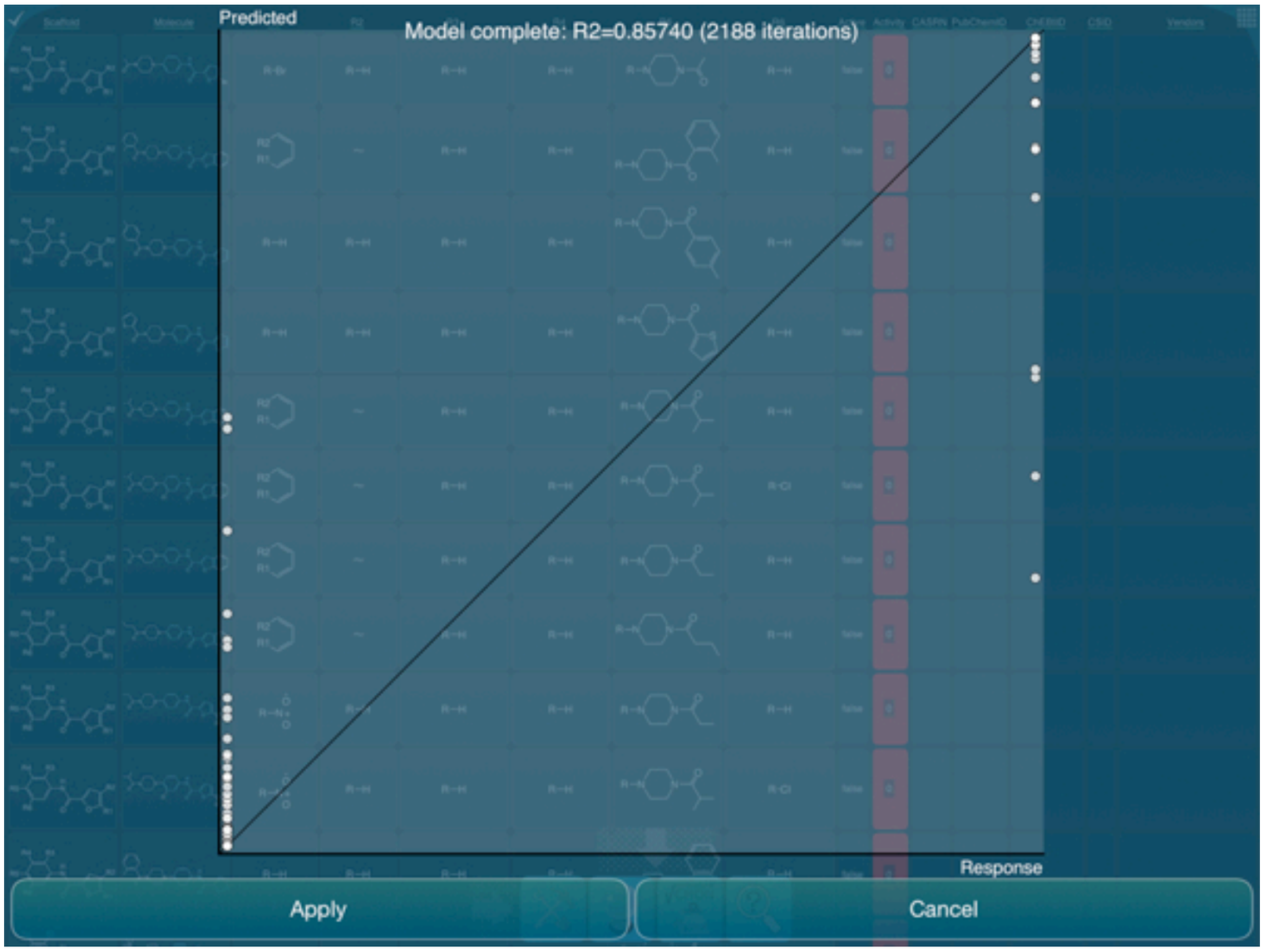


- Data from **ChemSpider**, **PubChem**, **ChEBI**
- Includes database links and vendor information

Importing

- Search results imported in marked-up form

Model Building



- Use known activities to build a SAR model

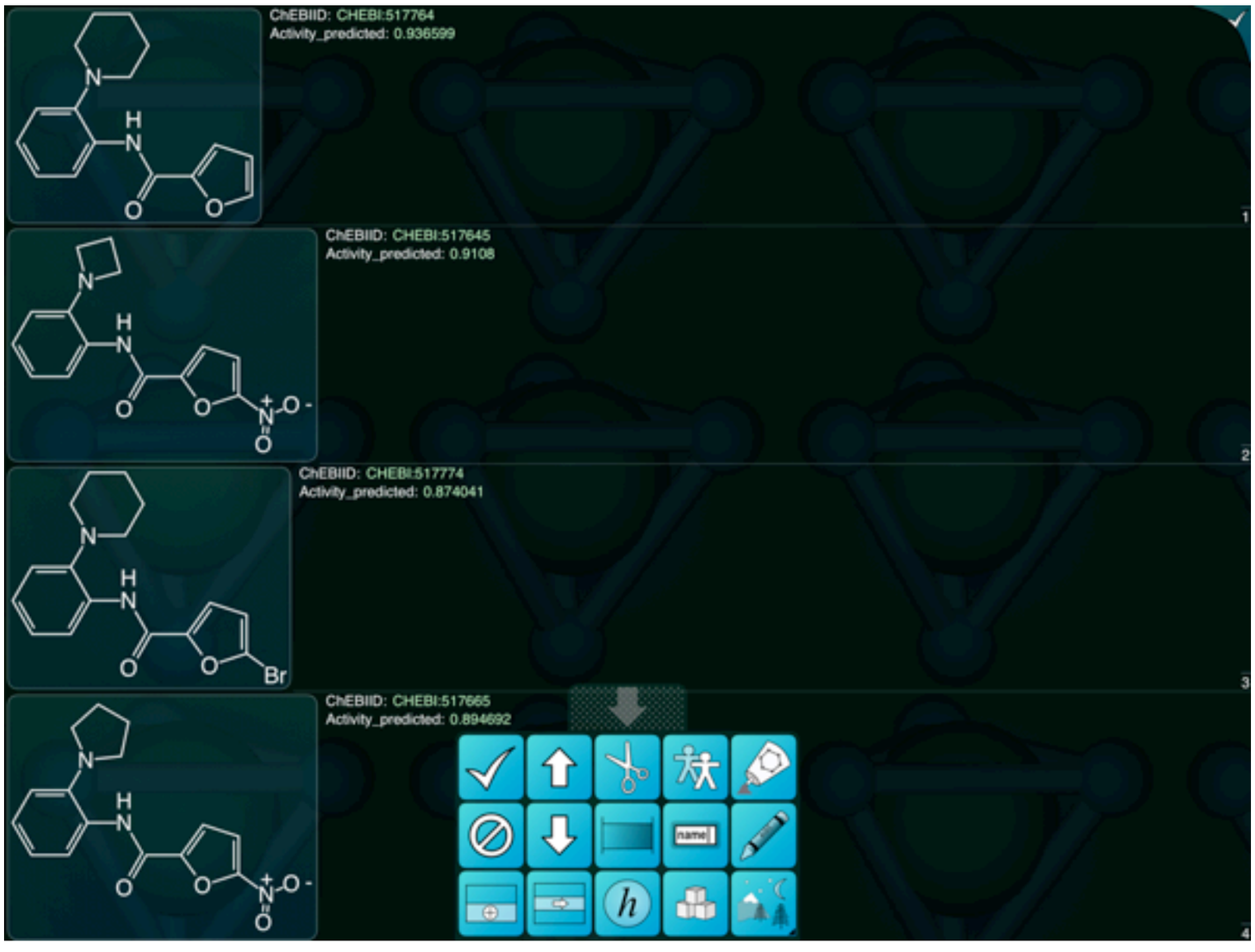
Model Application

- Predicted activities shown as coloured wedges

Selected Candidates

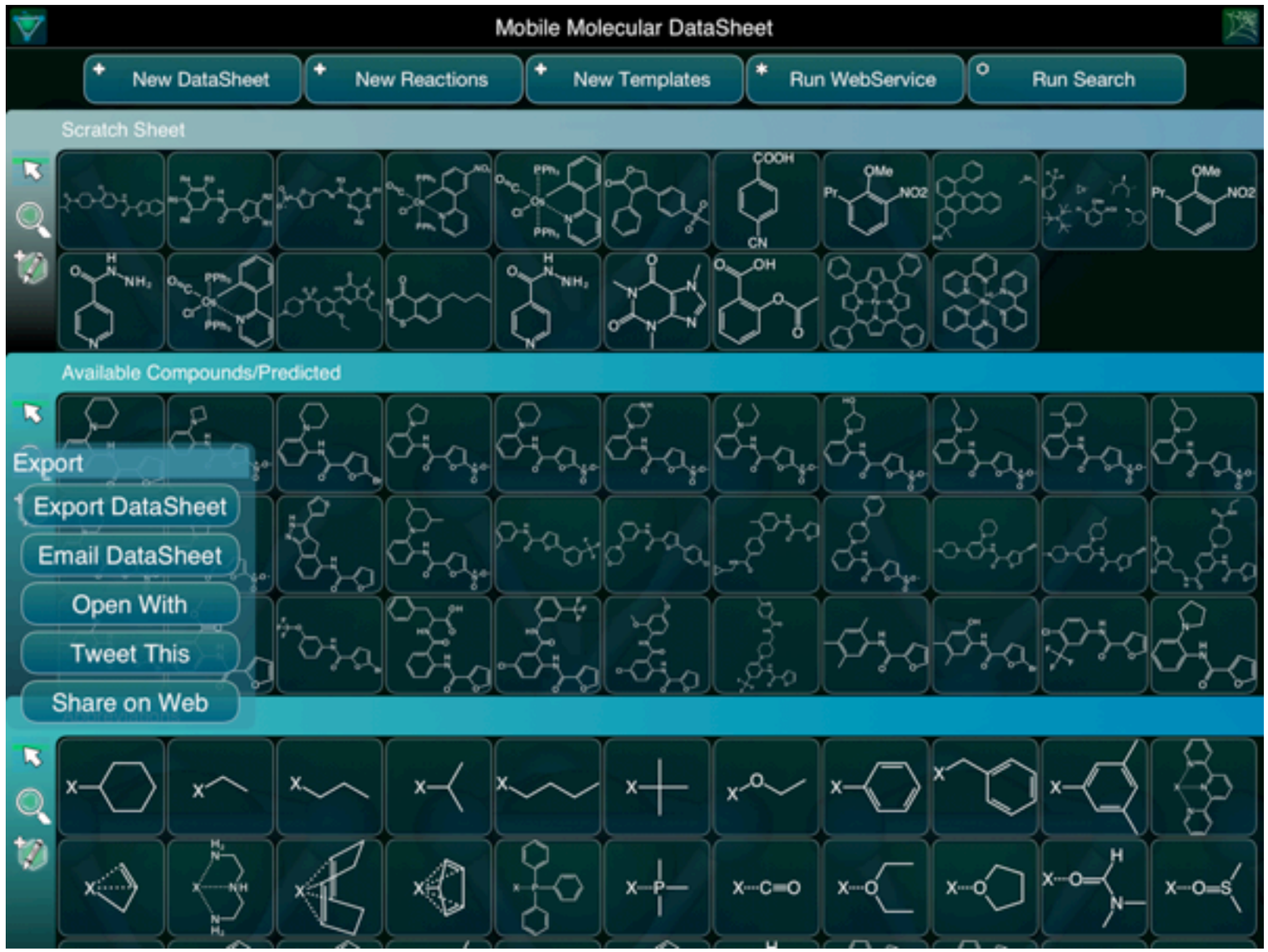
- Compile a list: predicted activity > 0.8
- Open in **Mobile Molecular DataSheet** (MMDS) app

Selected Candidates



- Compile a list: predicted activity > 0.8
- Open in **Mobile Molecular DataSheet** (MMDS) app

Sharing by Email

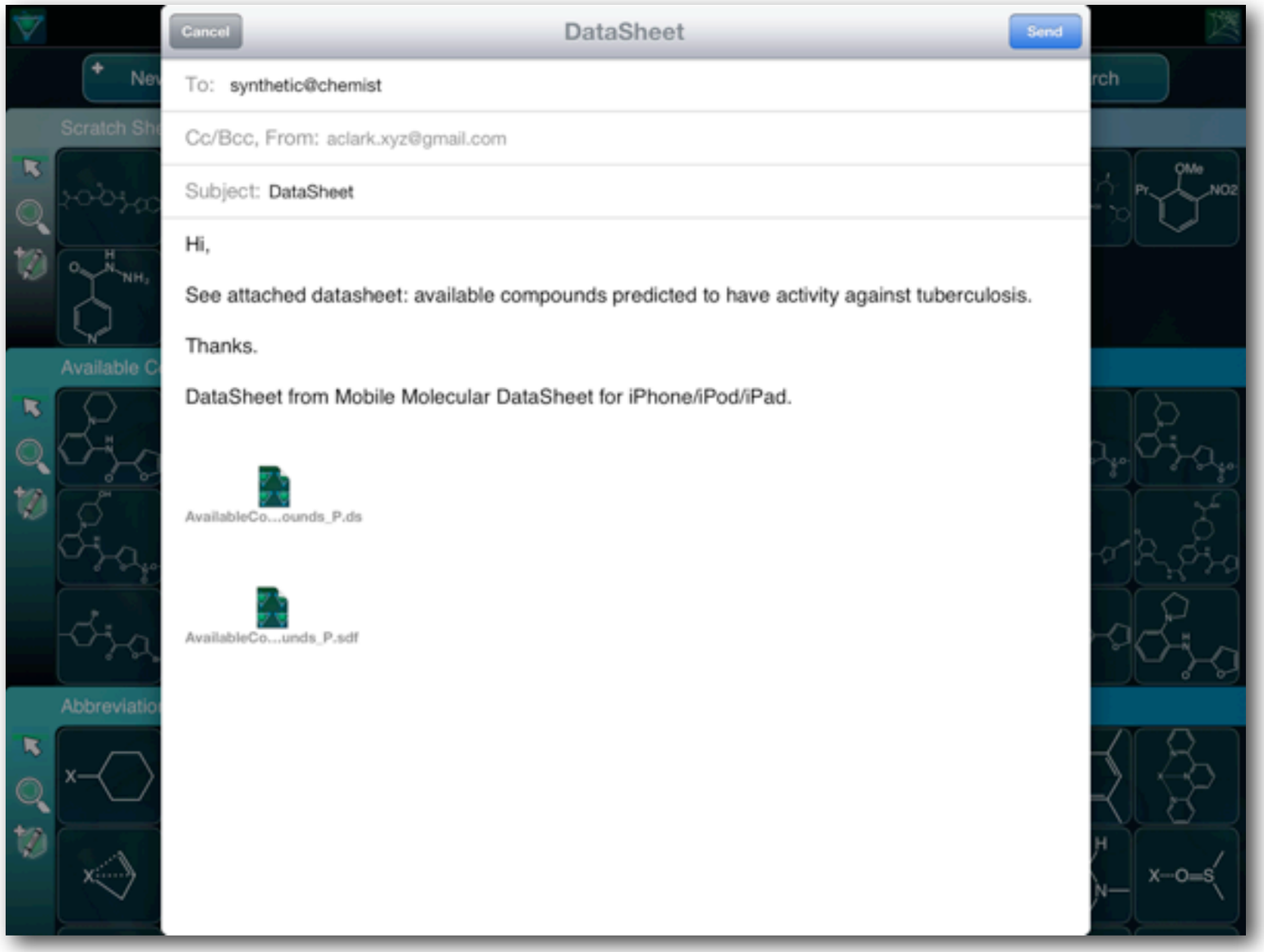


- Very easy to share using email attachments
- Recipient can open on any device, including mobile

Sharing by Email

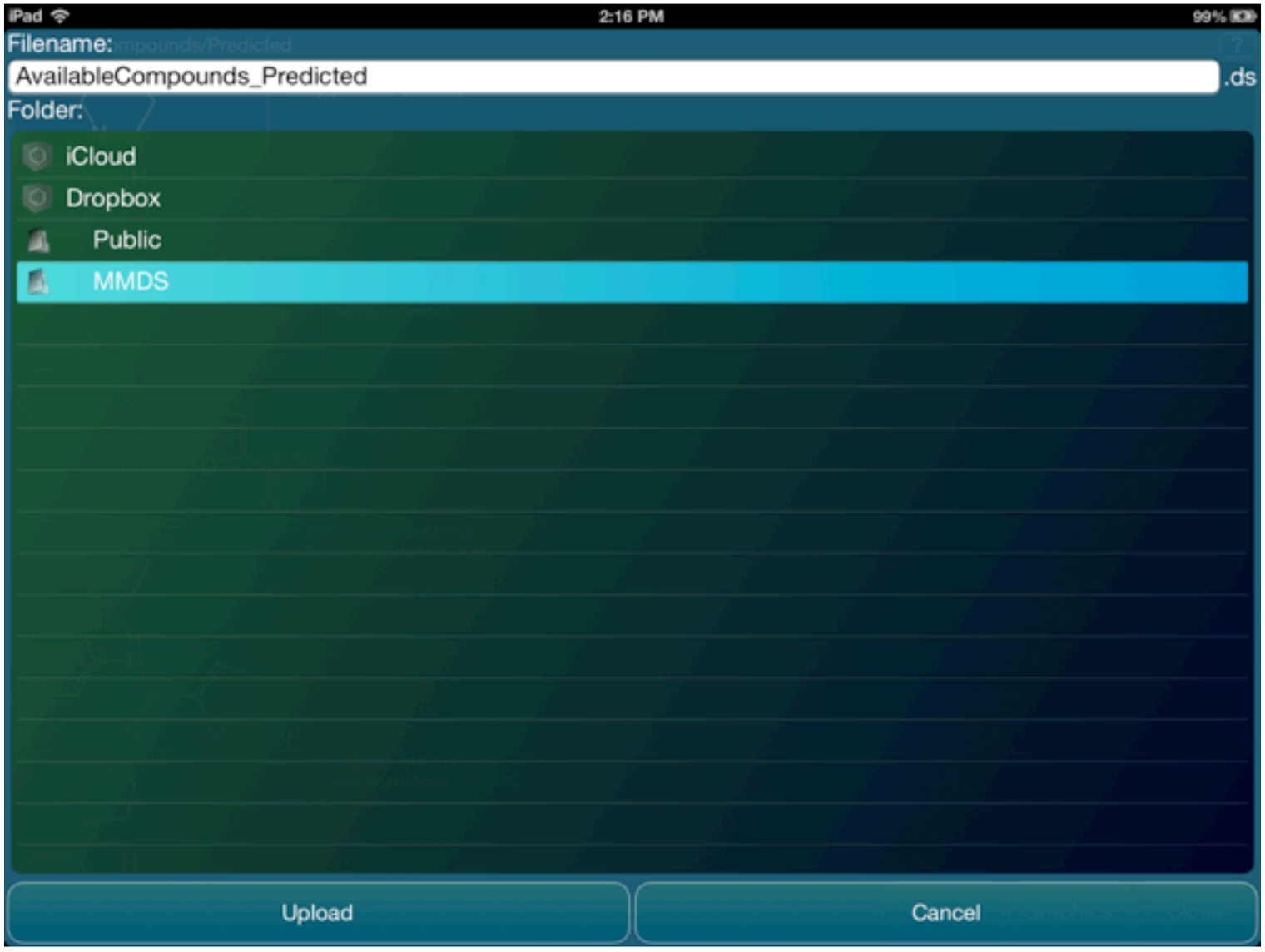
- Very easy to share using email attachments
- Recipient can open on any device, including mobile

Sharing by Email



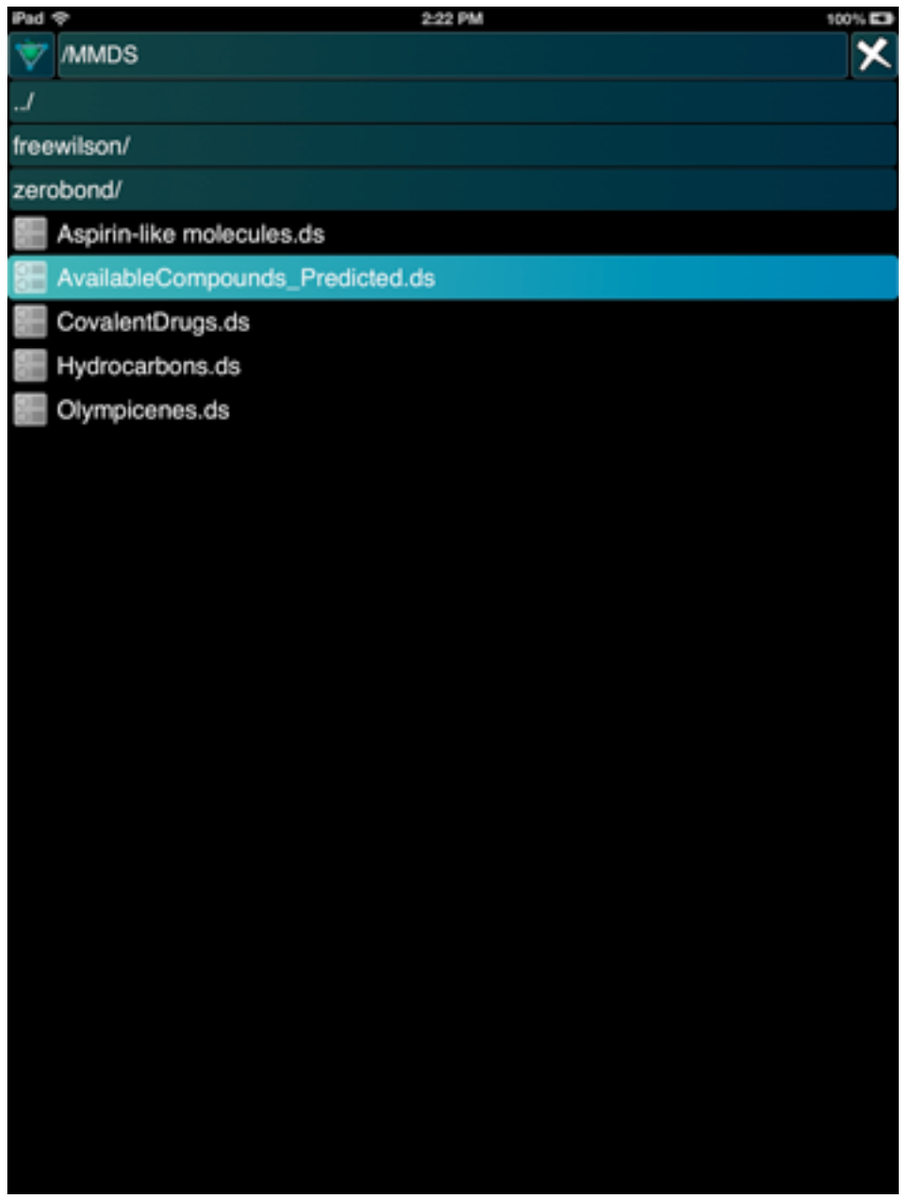
- Very easy to share using email attachments
- Recipient can open on any device, including mobile

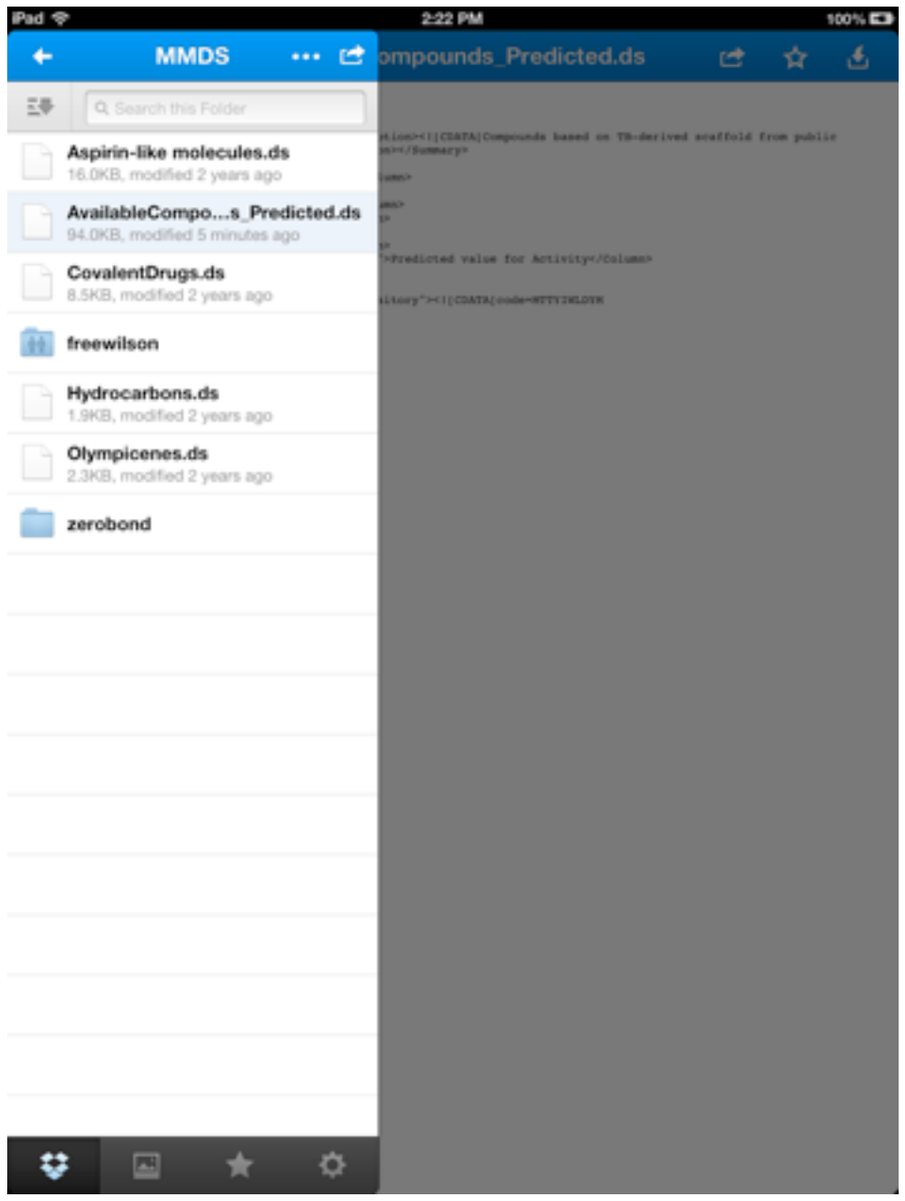
Sharing via Dropbox



- **MMDS** can interact with **MolSync** app
- DataSheets are uploaded & synchronised in **Dropbox**

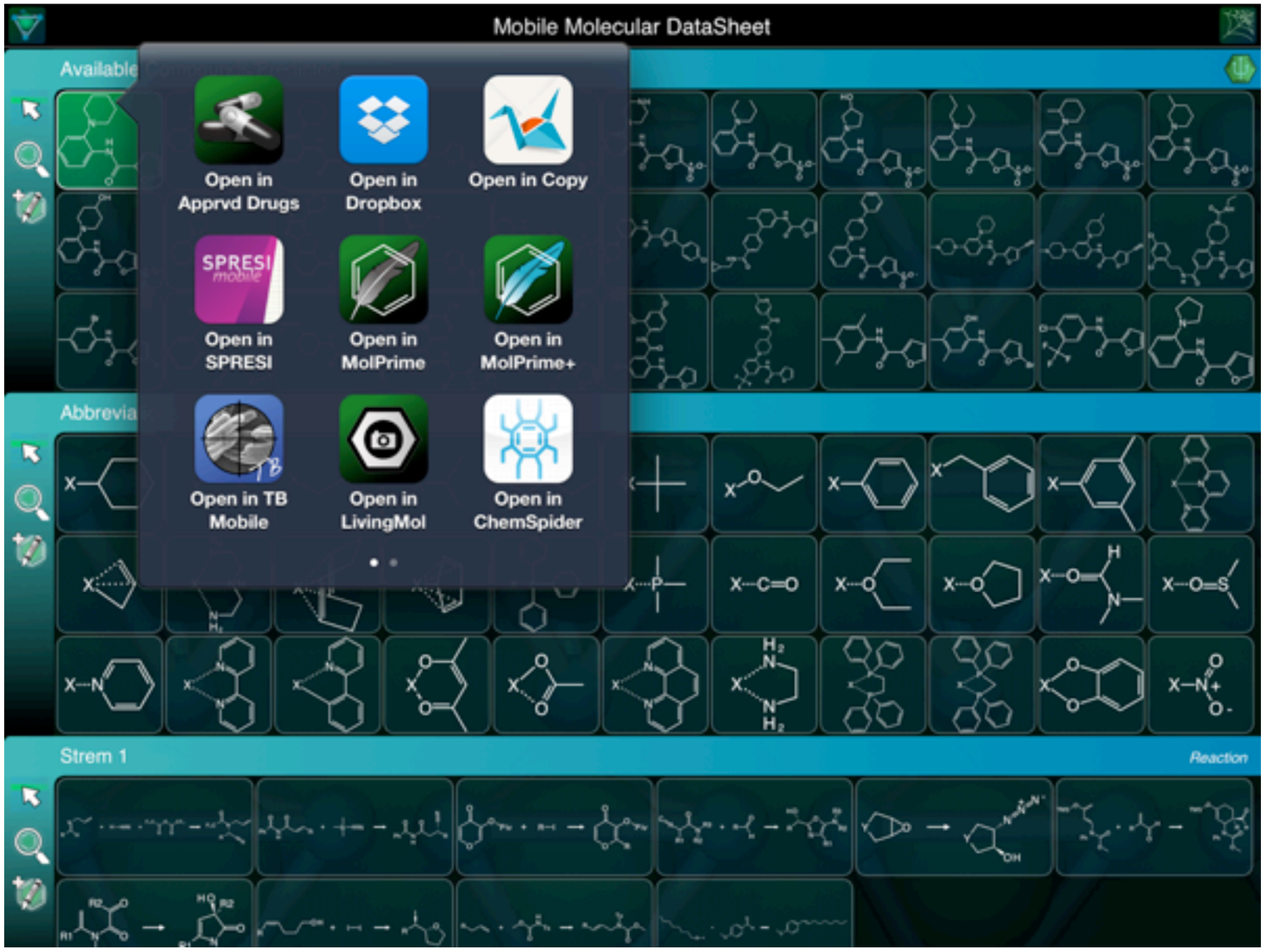
MolSync & Dropbox





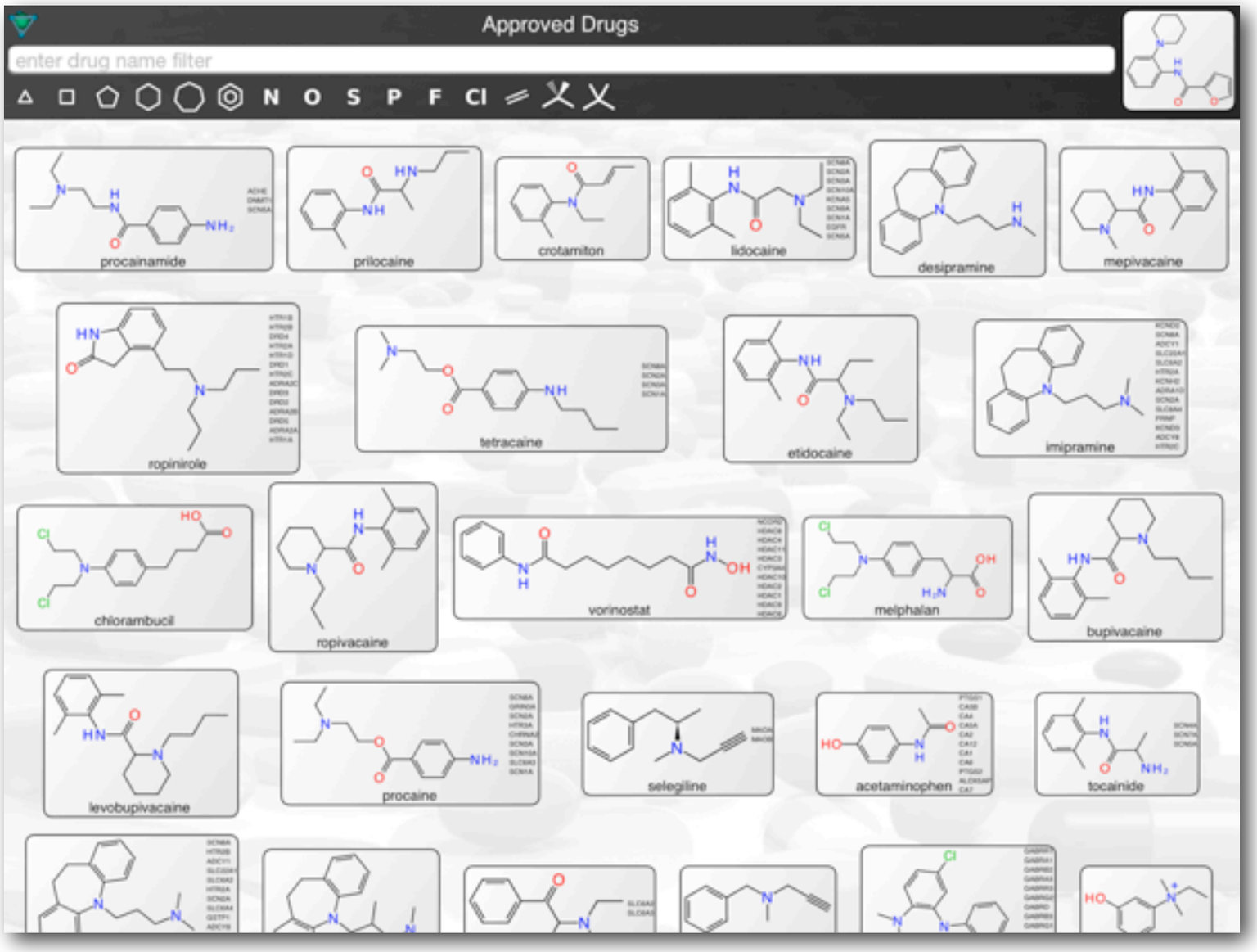
- **MolSync** app is a chemistry-aware file browser
- Can make use of **Dropbox**'s collaboration features

Lookup FDA Reference



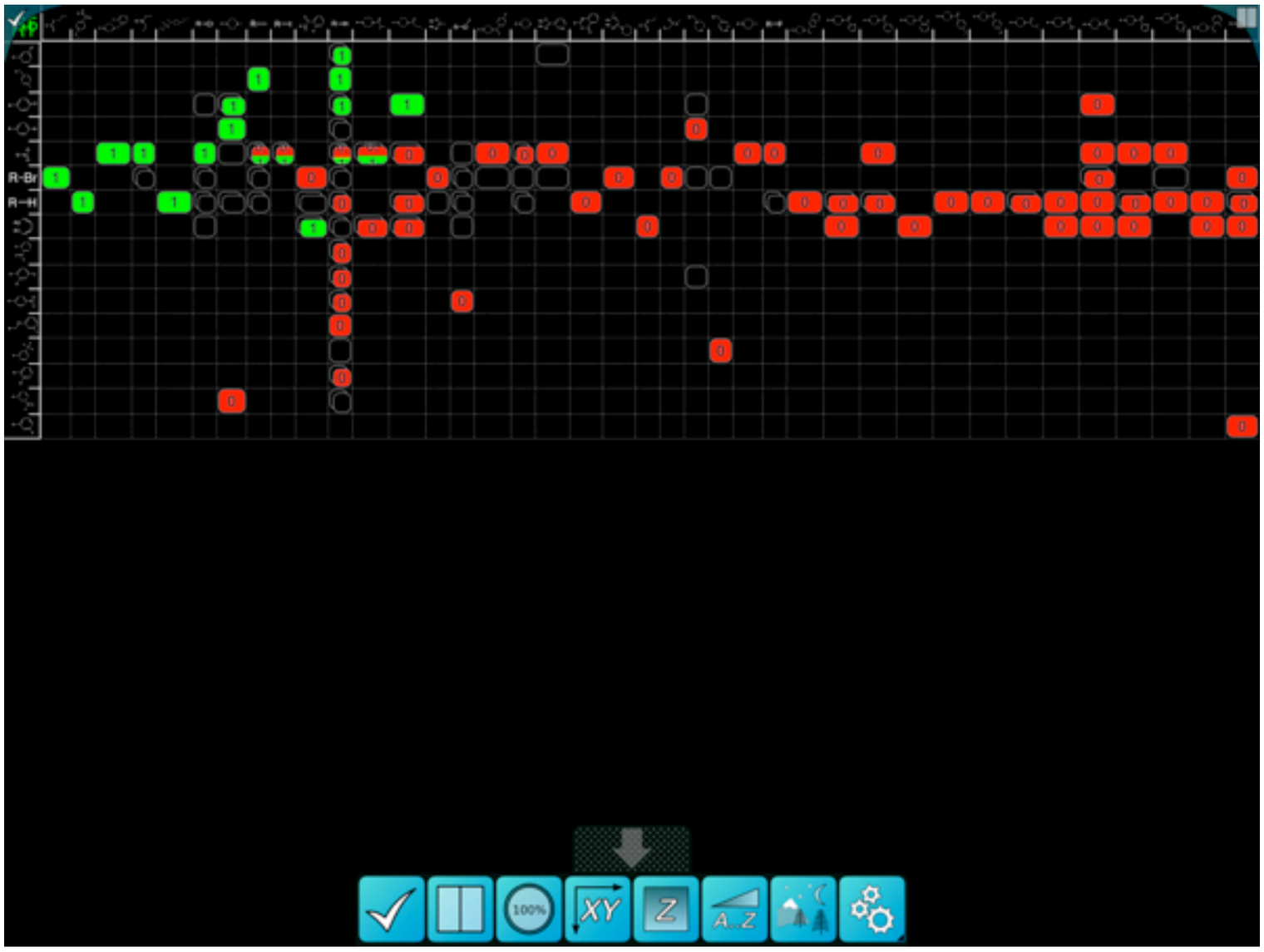
- *Open With* the **Approved Drugs** app

Approved Drugs

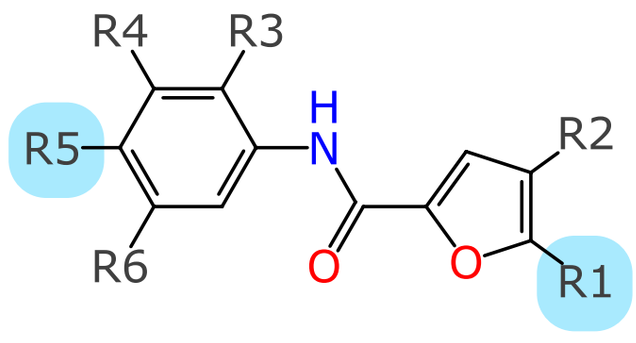


- 1300 structures for FDA-approved drugs
- Reference search by structural similarity

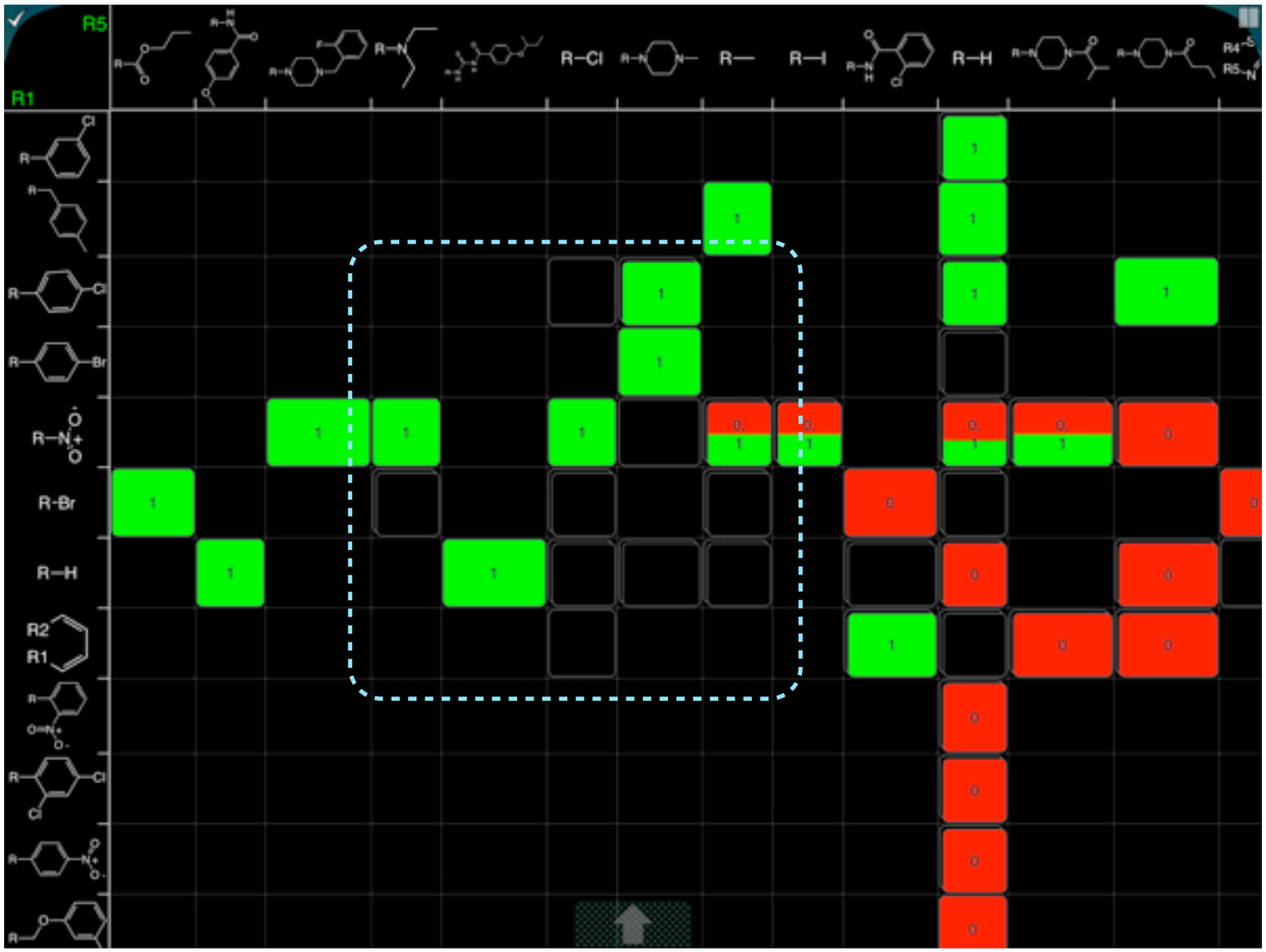
Matrix View



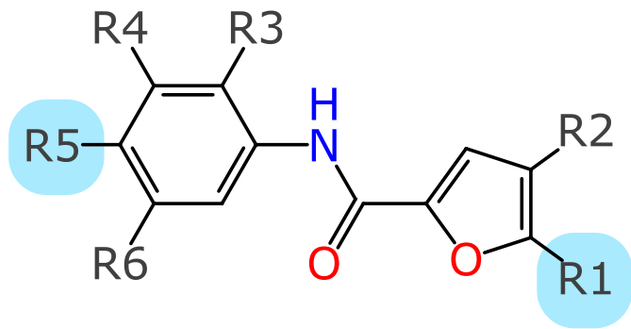
- Plot structures in a grid: **R1** vs **R5**



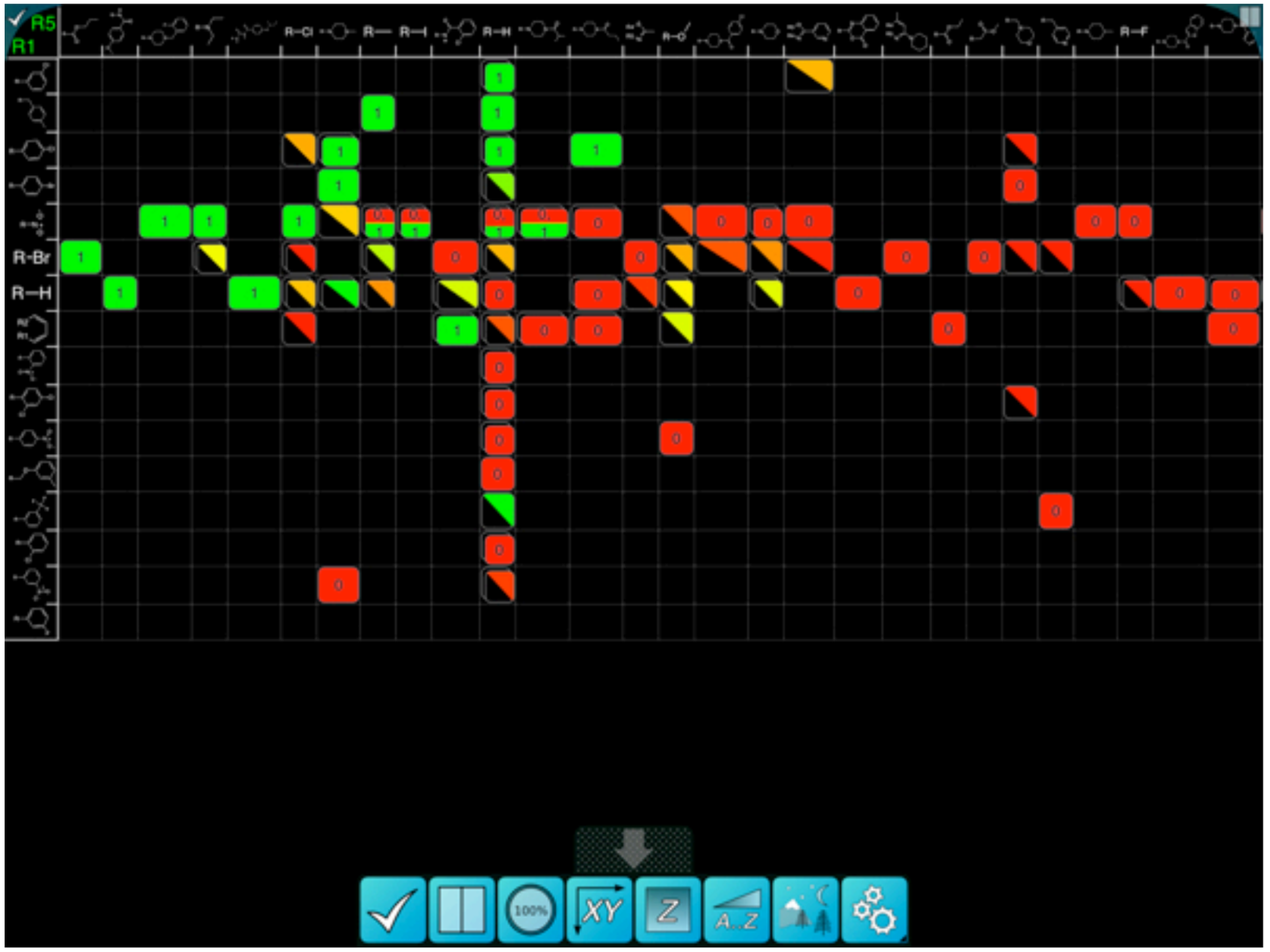
Matrix View



- Plot structures in a grid: **R1** vs **R5**

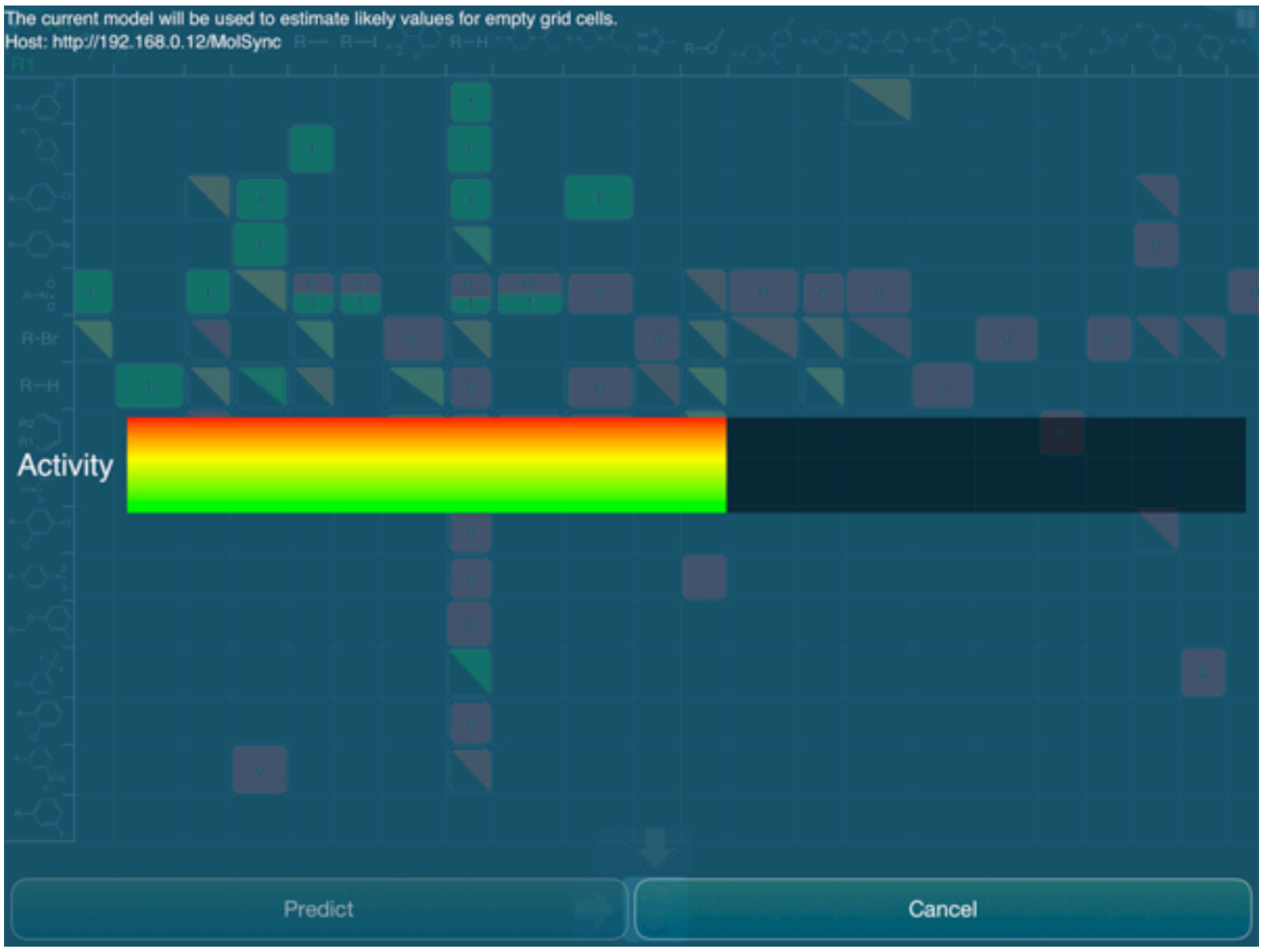


Matrix + Predictions



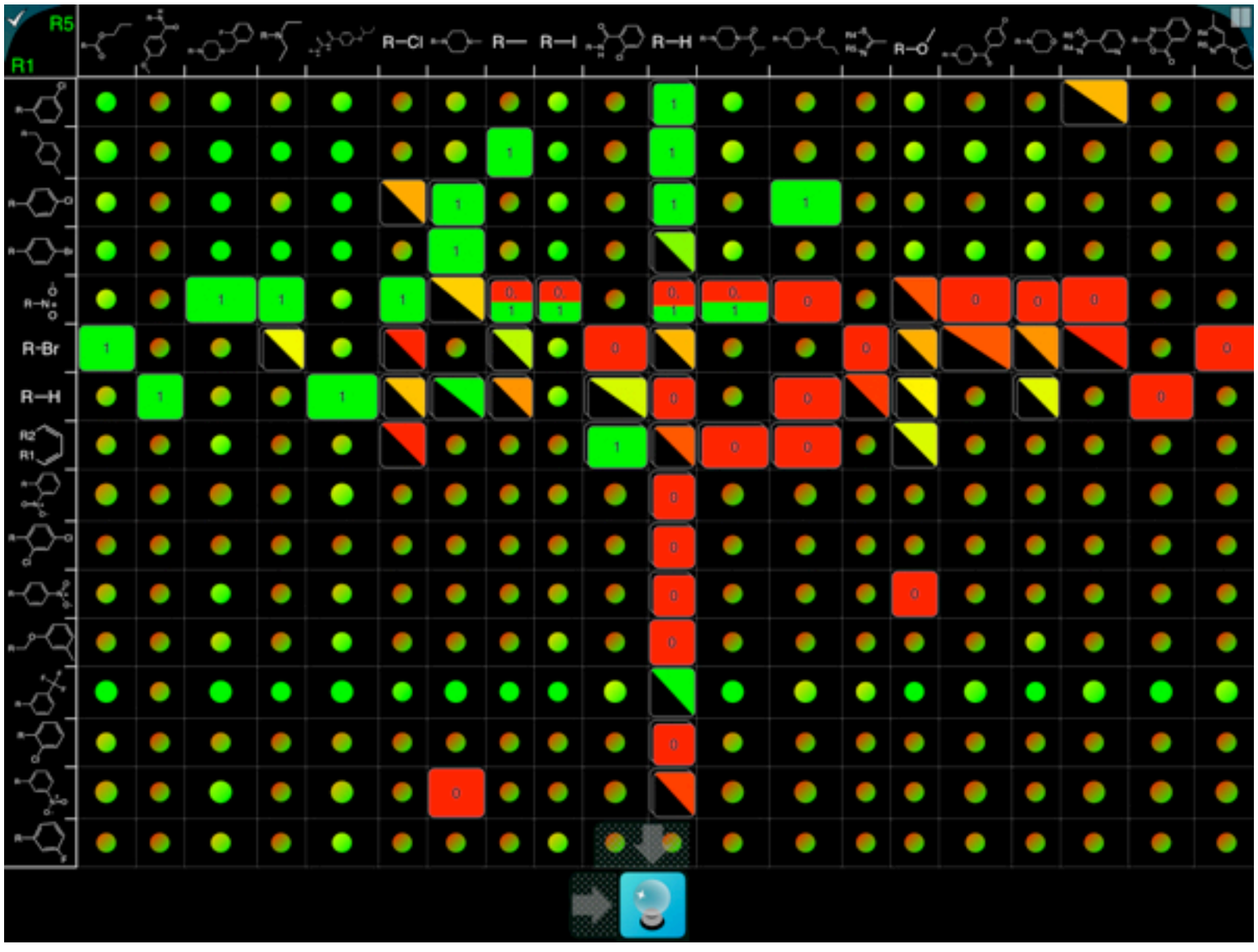
- Compounds with unknown activity predicted shown using wedge style

Matrix + Hypotheticals



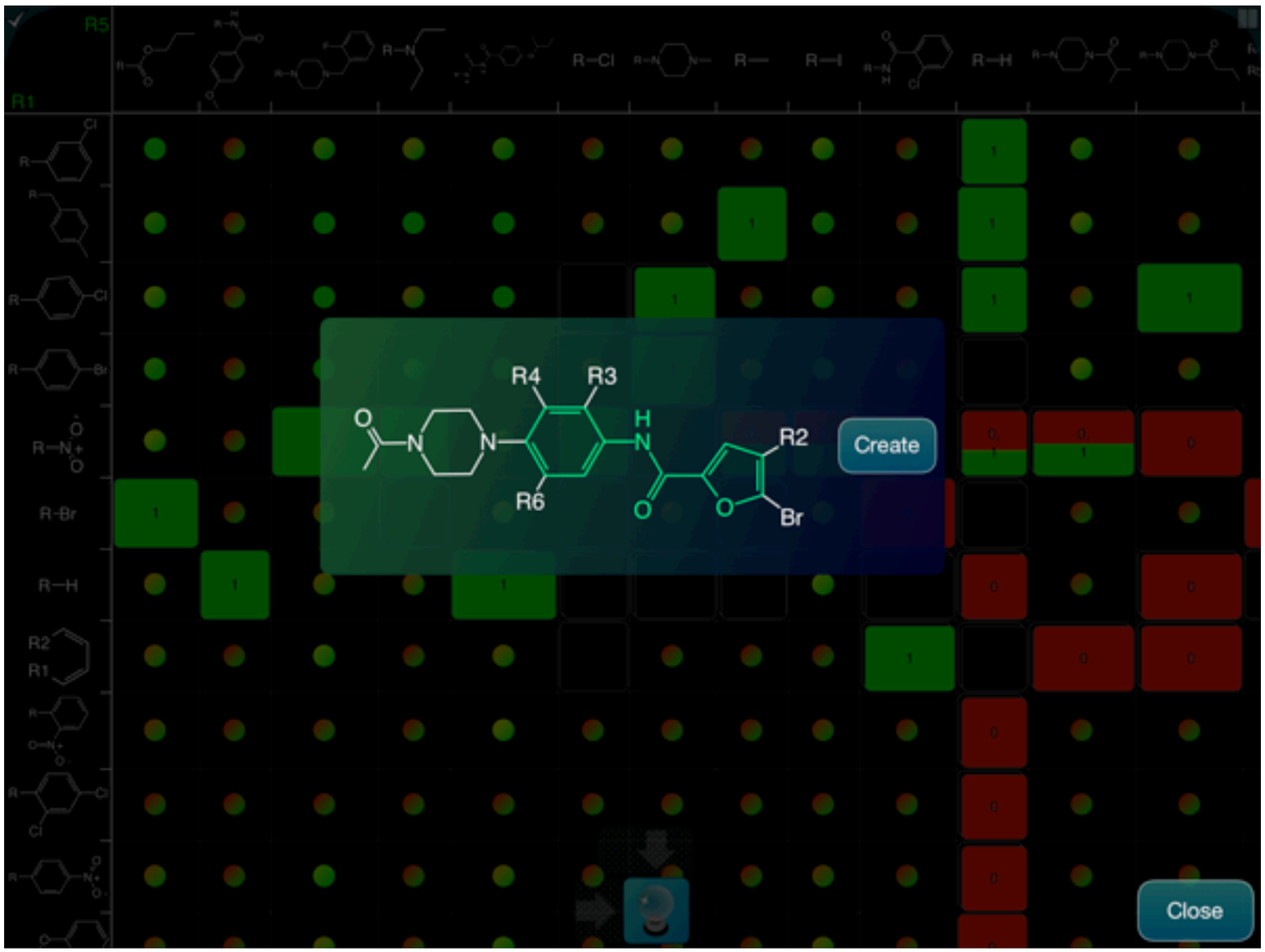
- Empty squares: propose compounds & predict

Matrix + Hypotheticals



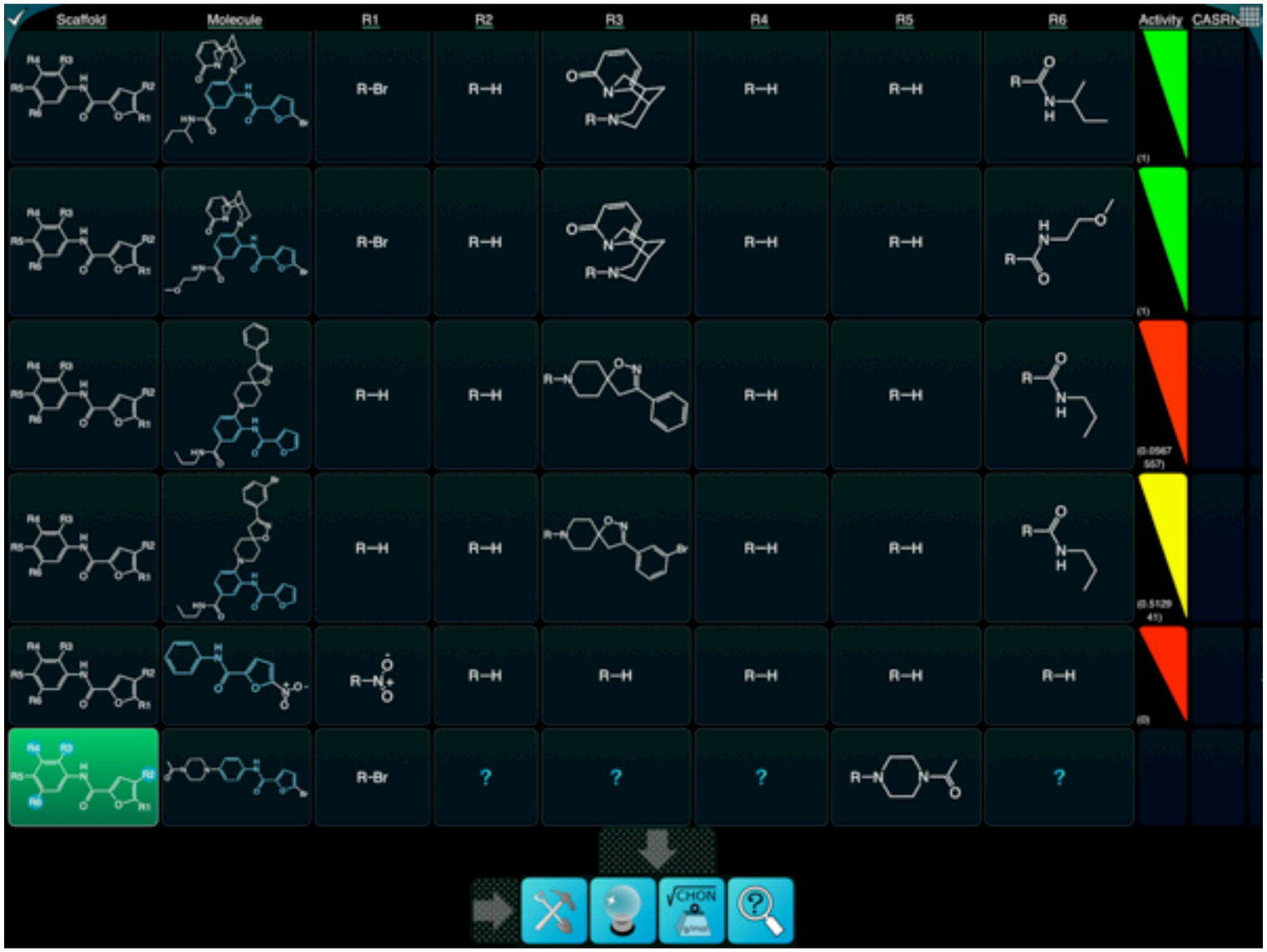
- Empty squares: propose compounds & predict

Proposing Compounds



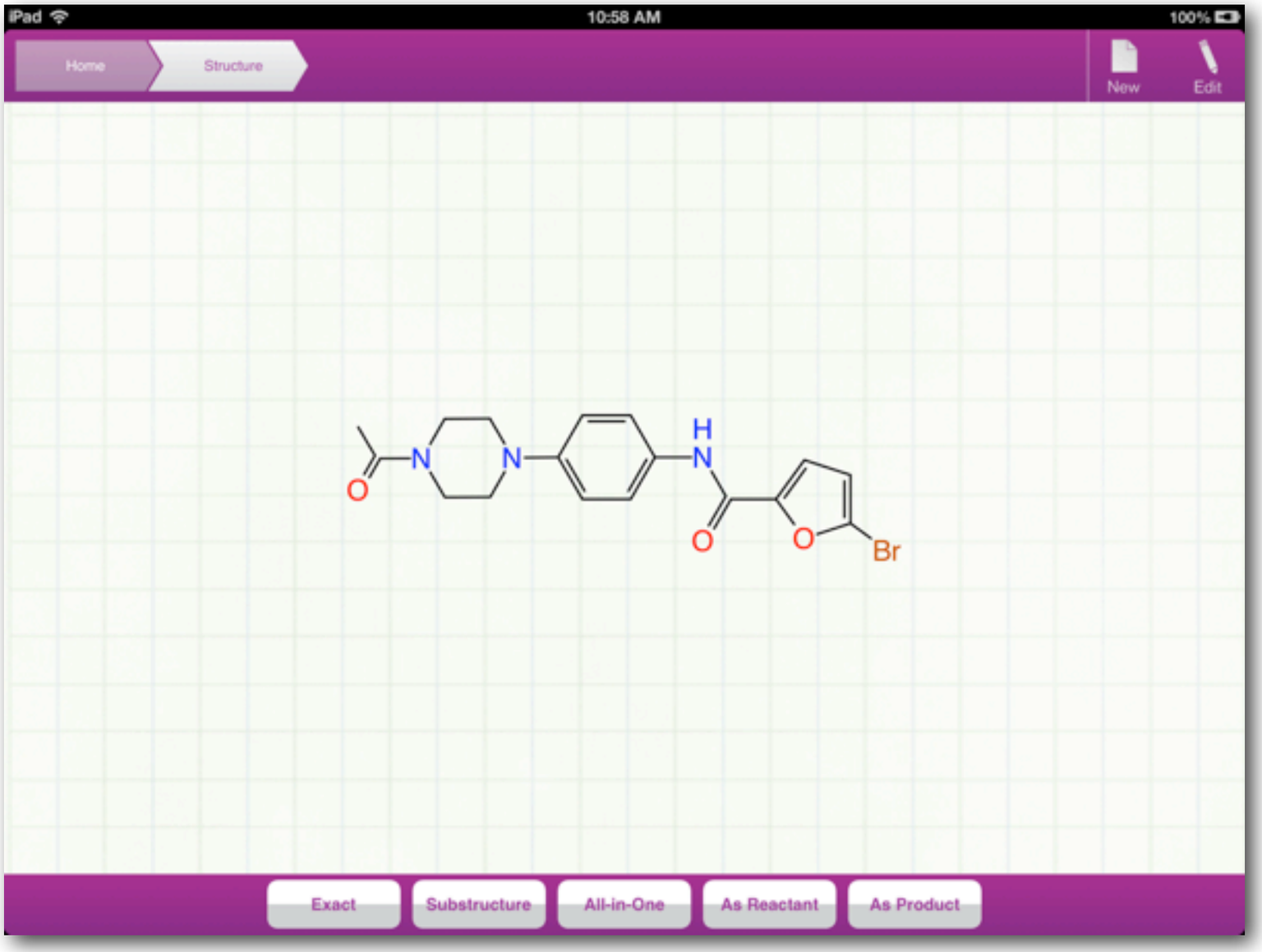
- Tap on a square: prompt to create partially defined entry, defining **R1** and **R5**

Hypothetical Compound



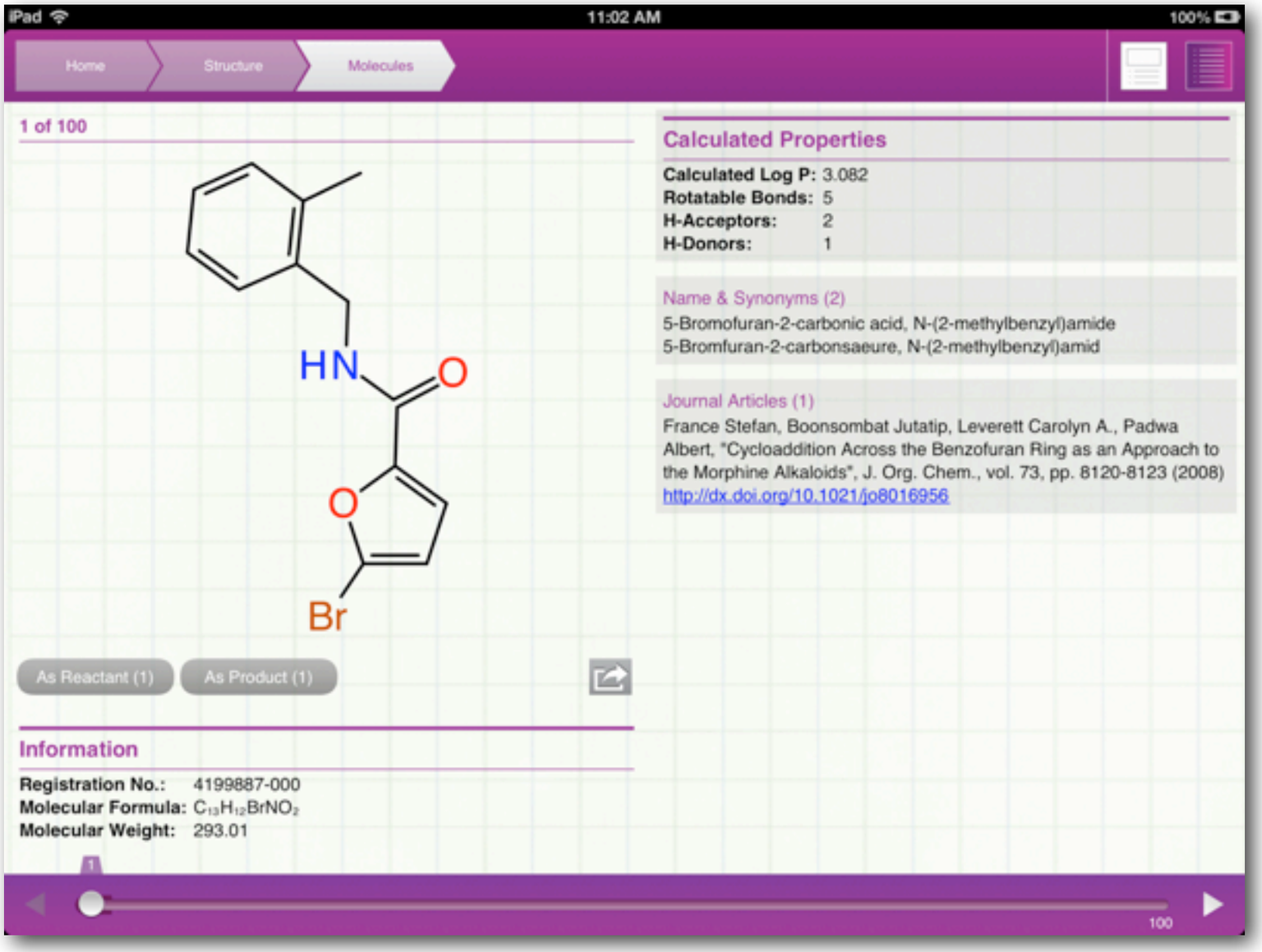
- New compound added: partial definition
- **R2**, **R3**, **R4** and **R6** unspecified

Lookup in SPRESI



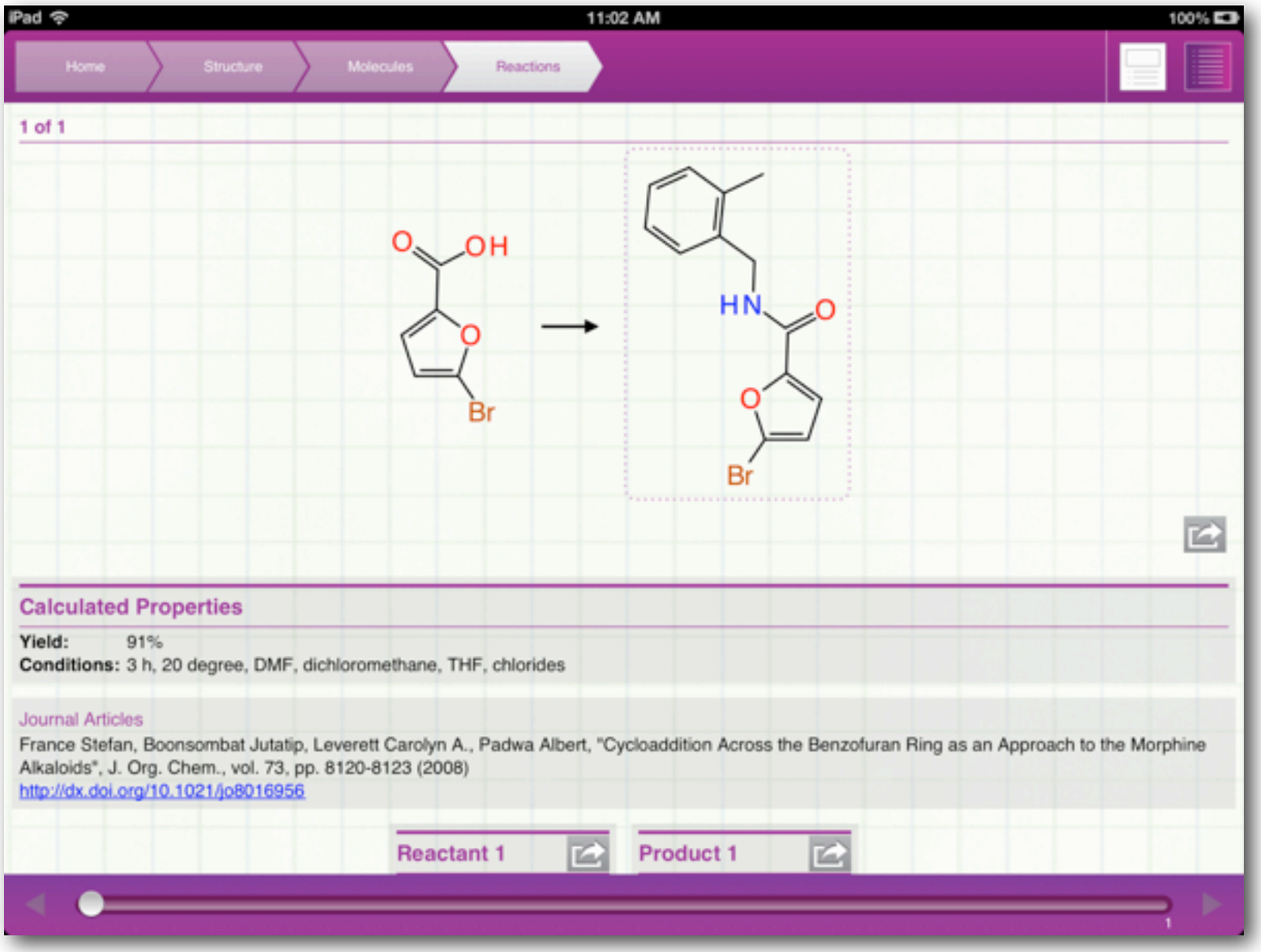
- Use **SPRESI*mobile*** app to find similar compounds: looking for a synthesis...

Similar Backbone



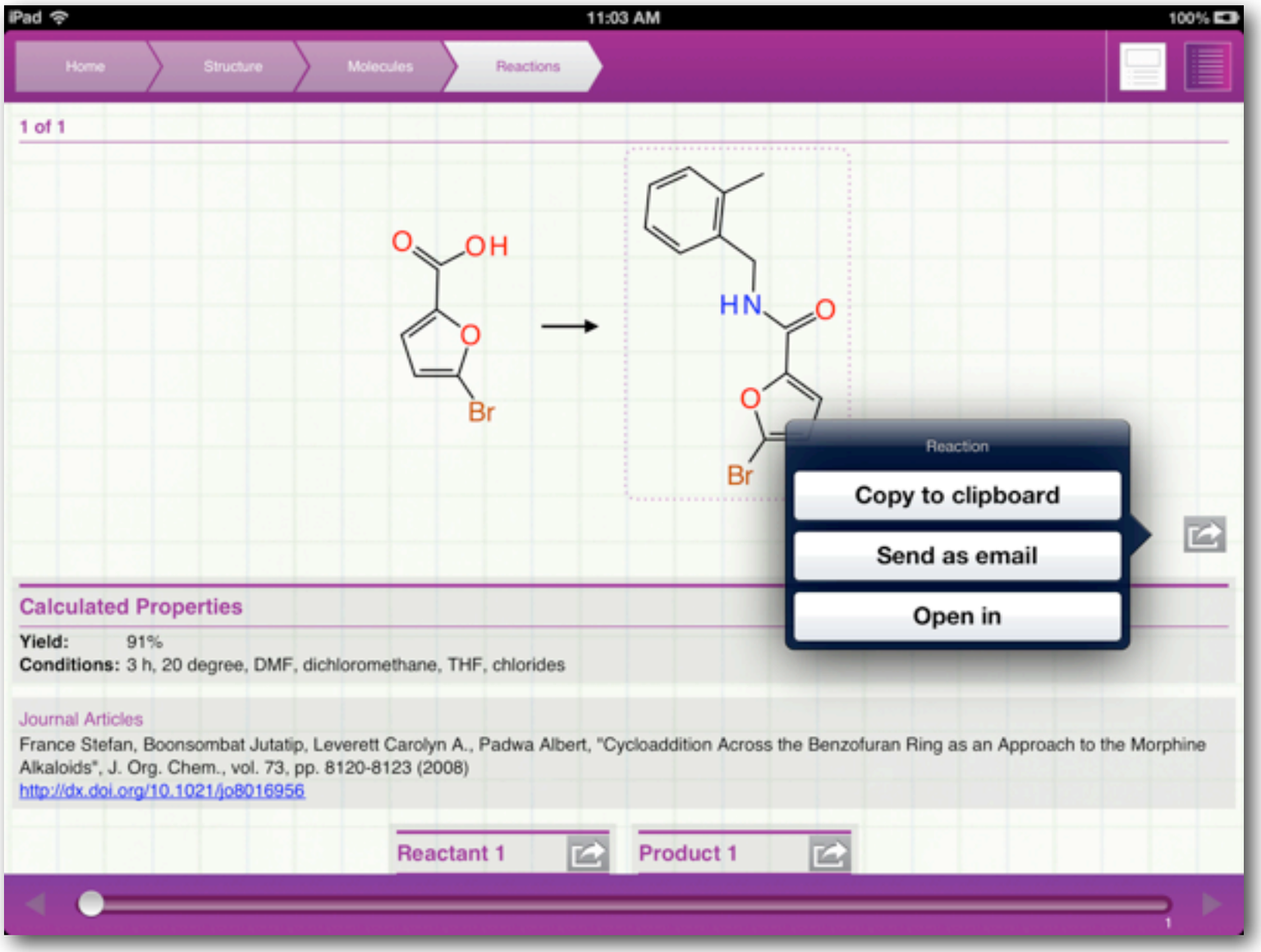
- Found a compound with similar core, possible synthesis template: consult literature

Find Reactions



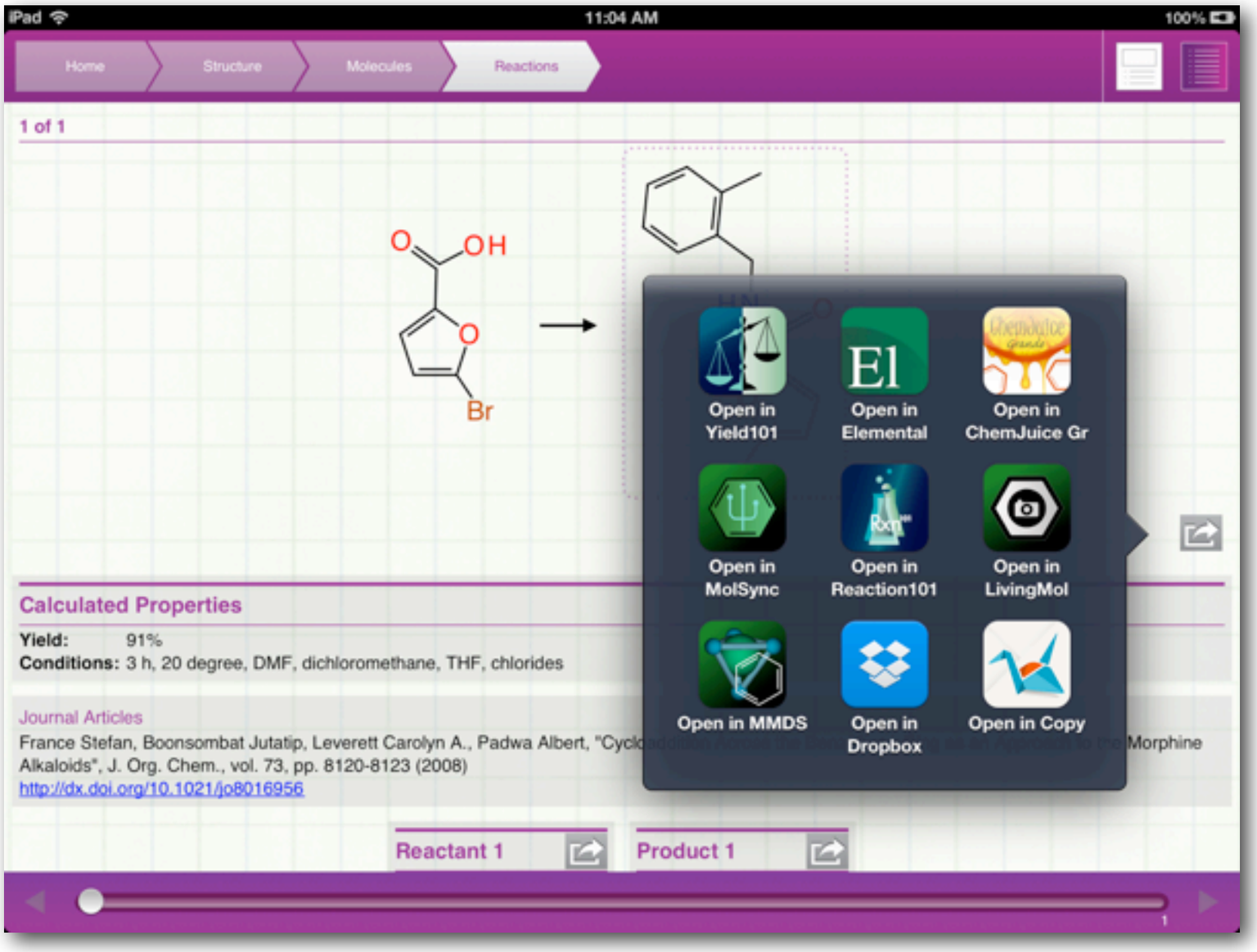
- Lookup synthetic routes: amide condensation
- Open the reaction data in **Yield101**

Find Reactions



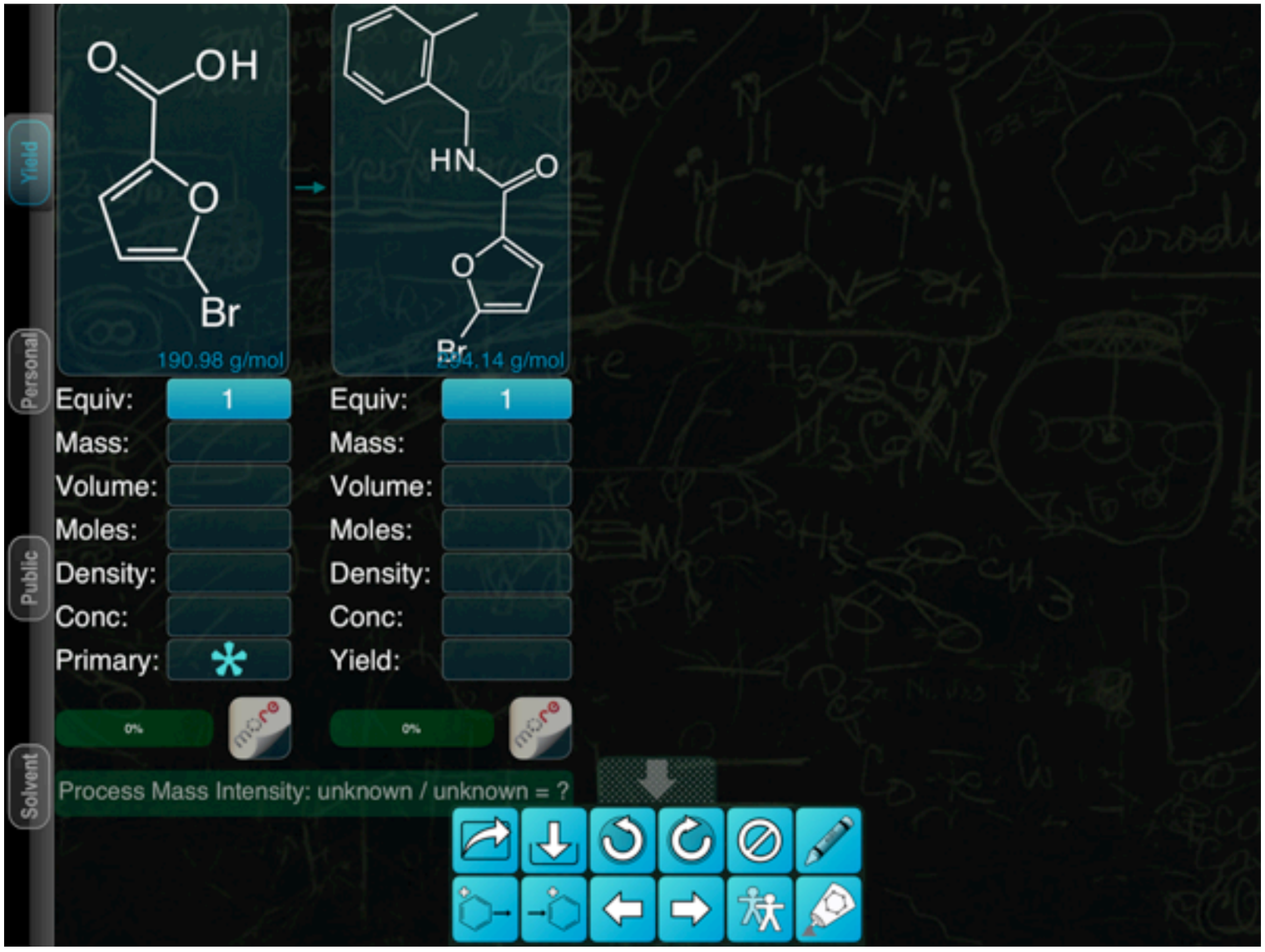
- Lookup synthetic routes: amide condensation
- Open the reaction data in **Yield101**

Find Reactions



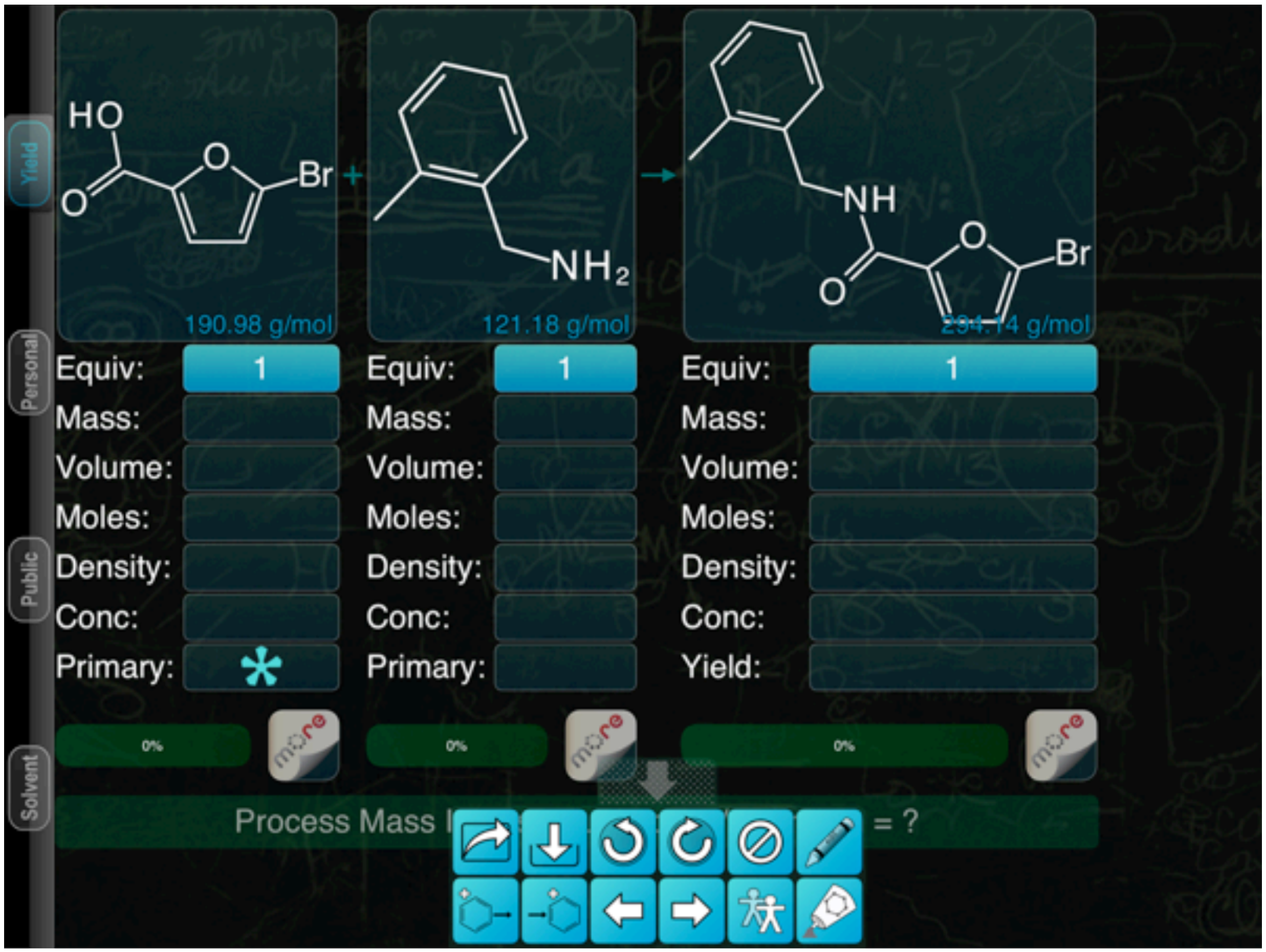
- Lookup synthetic routes: amide condensation
- Open the reaction data in **Yield101**

Yield101



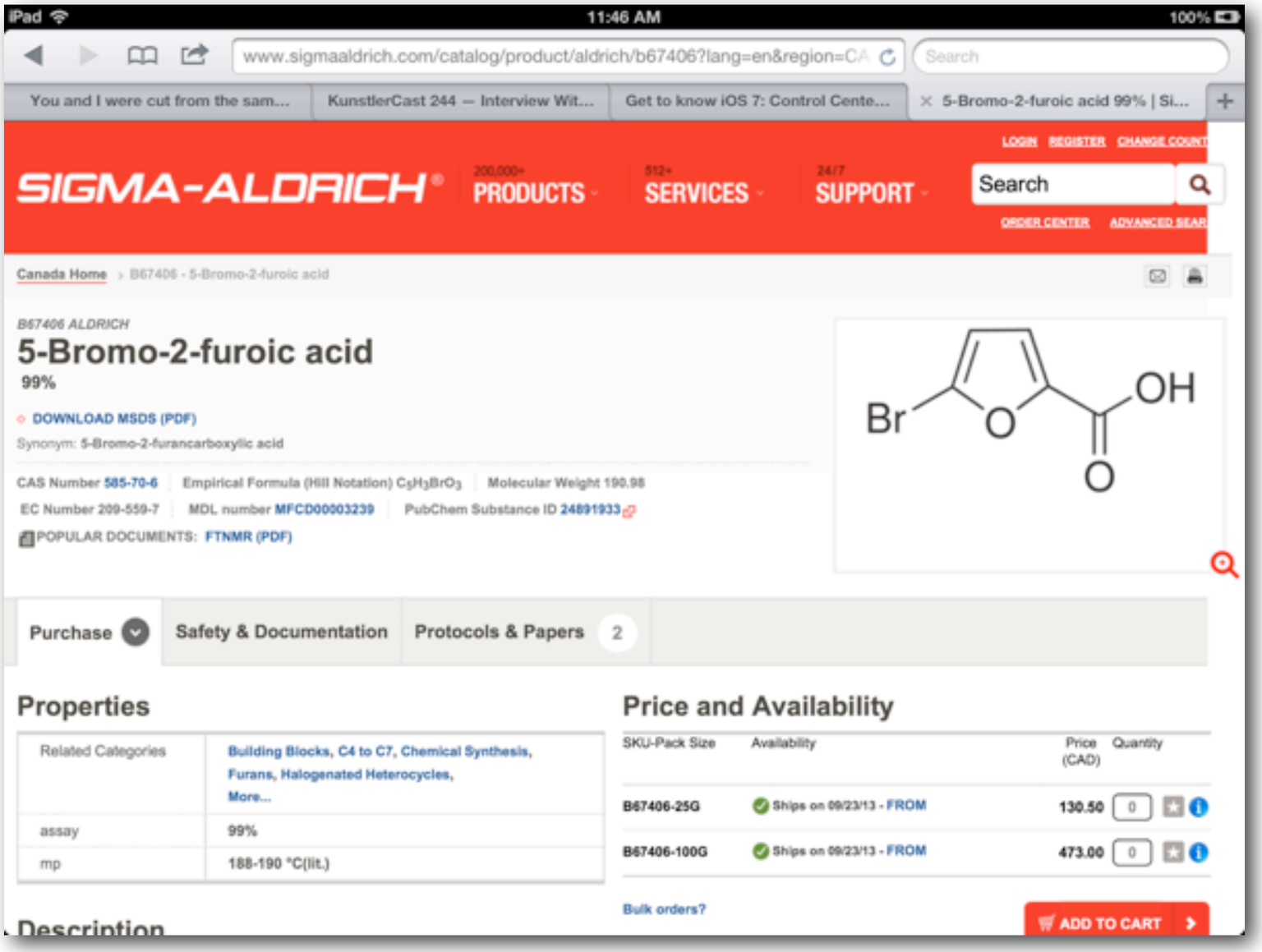
- Prototype synthetic chemistry lab notebook
- Reaction components imported

Reaction Scheme



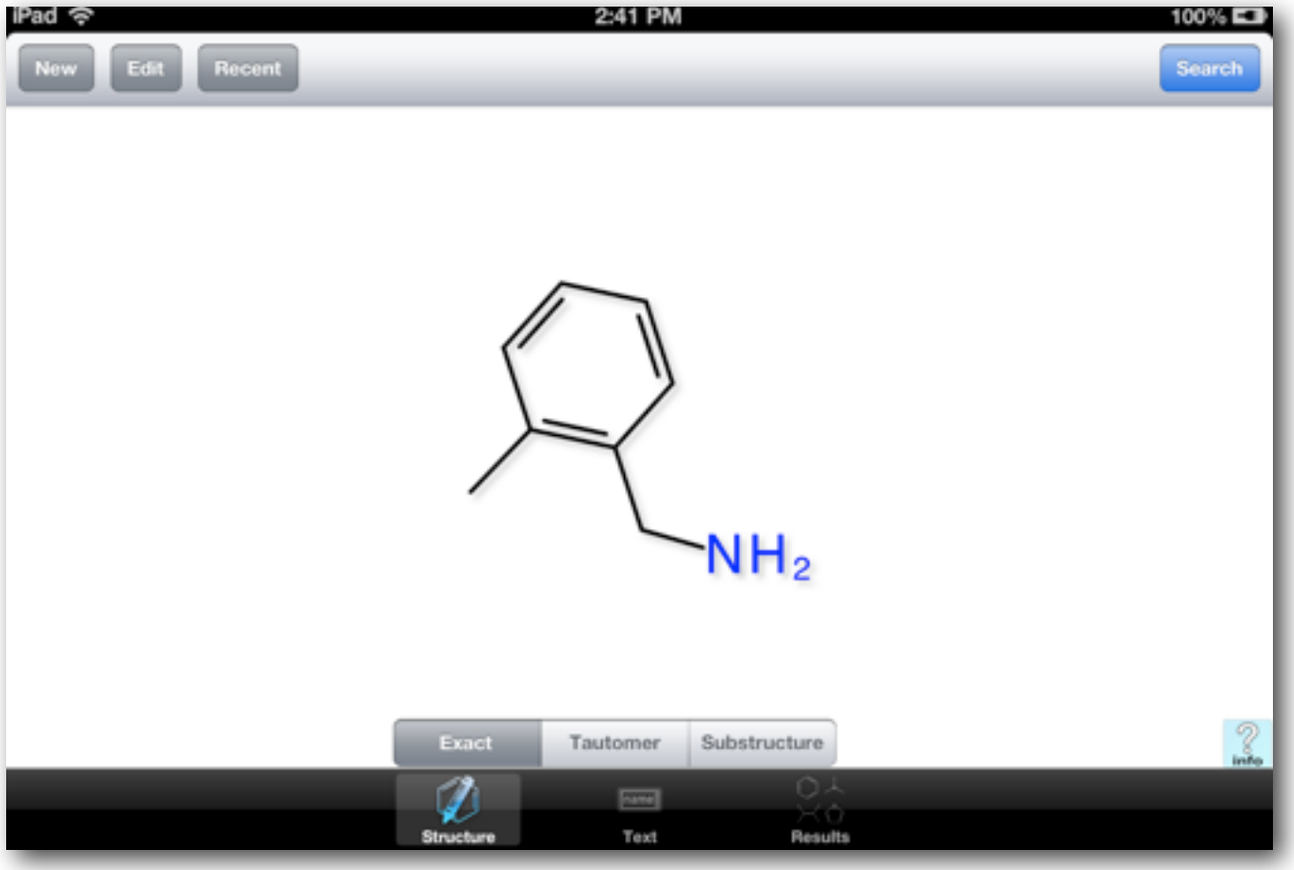
- Tidyup structures, add in reagent

Lookup Starting Material



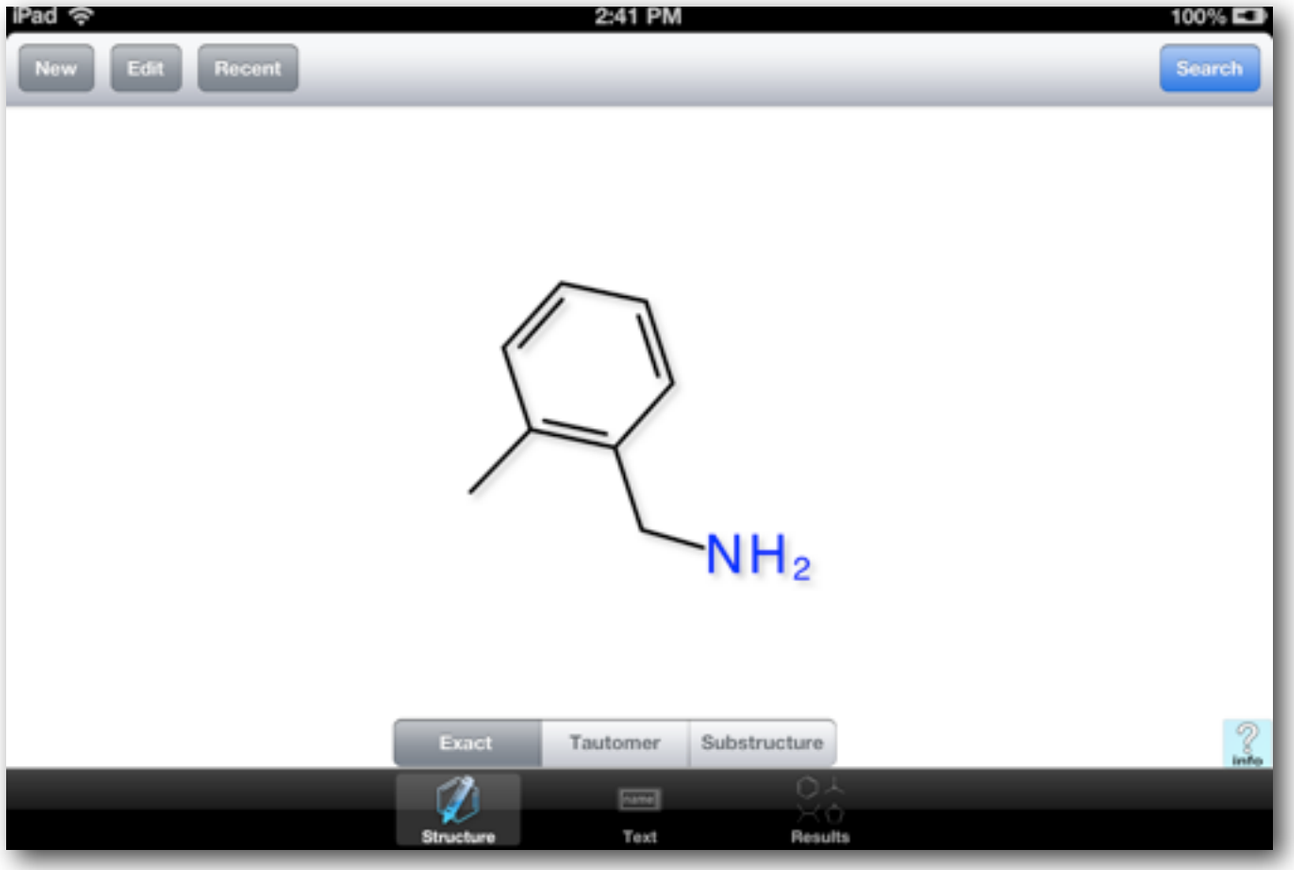
- Use **Mobile Reagents** integration to find commercial source for starting material

Lookup in ChemSpider



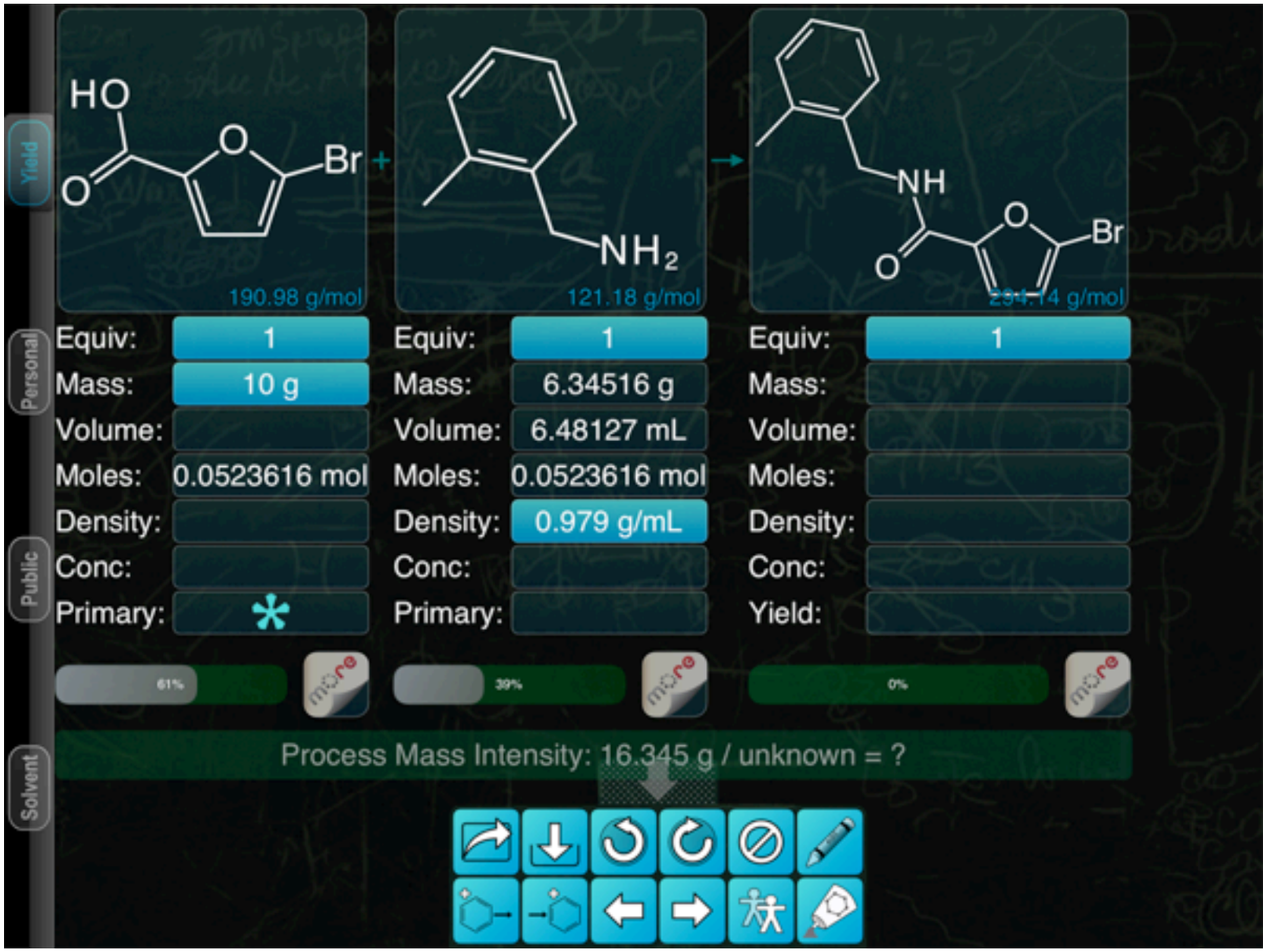
- Paste into **ChemSpider Mobile** app to do structure search
- Open matching **ChemSpider** compound information page

Lookup in ChemSpider



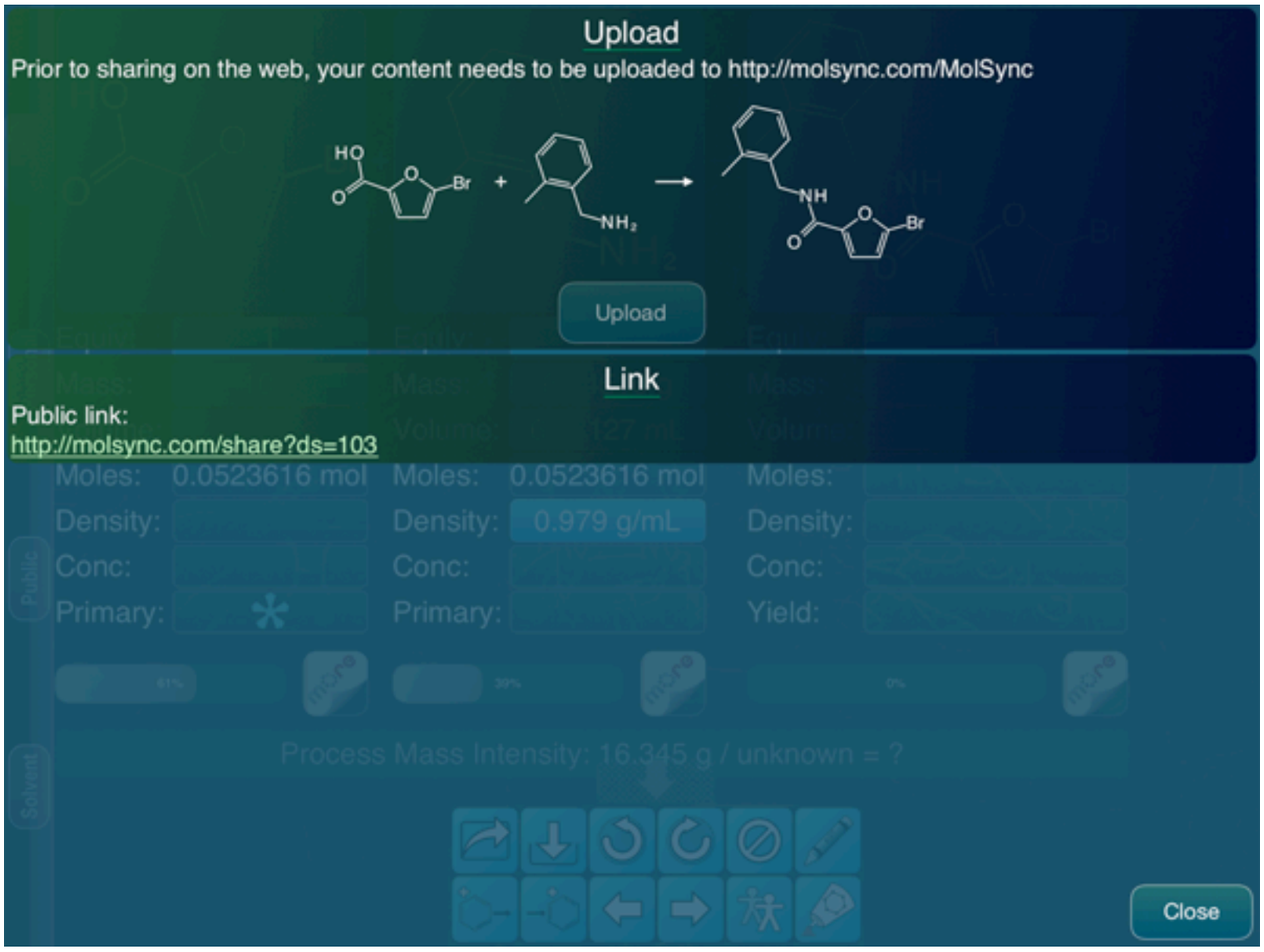
- Paste into **ChemSpider Mobile** app to do structure search
- Open matching **ChemSpider** compound information page

Quantitative Details



- Enter starting quantity, density: preliminary estimate of amounts used in synthesis

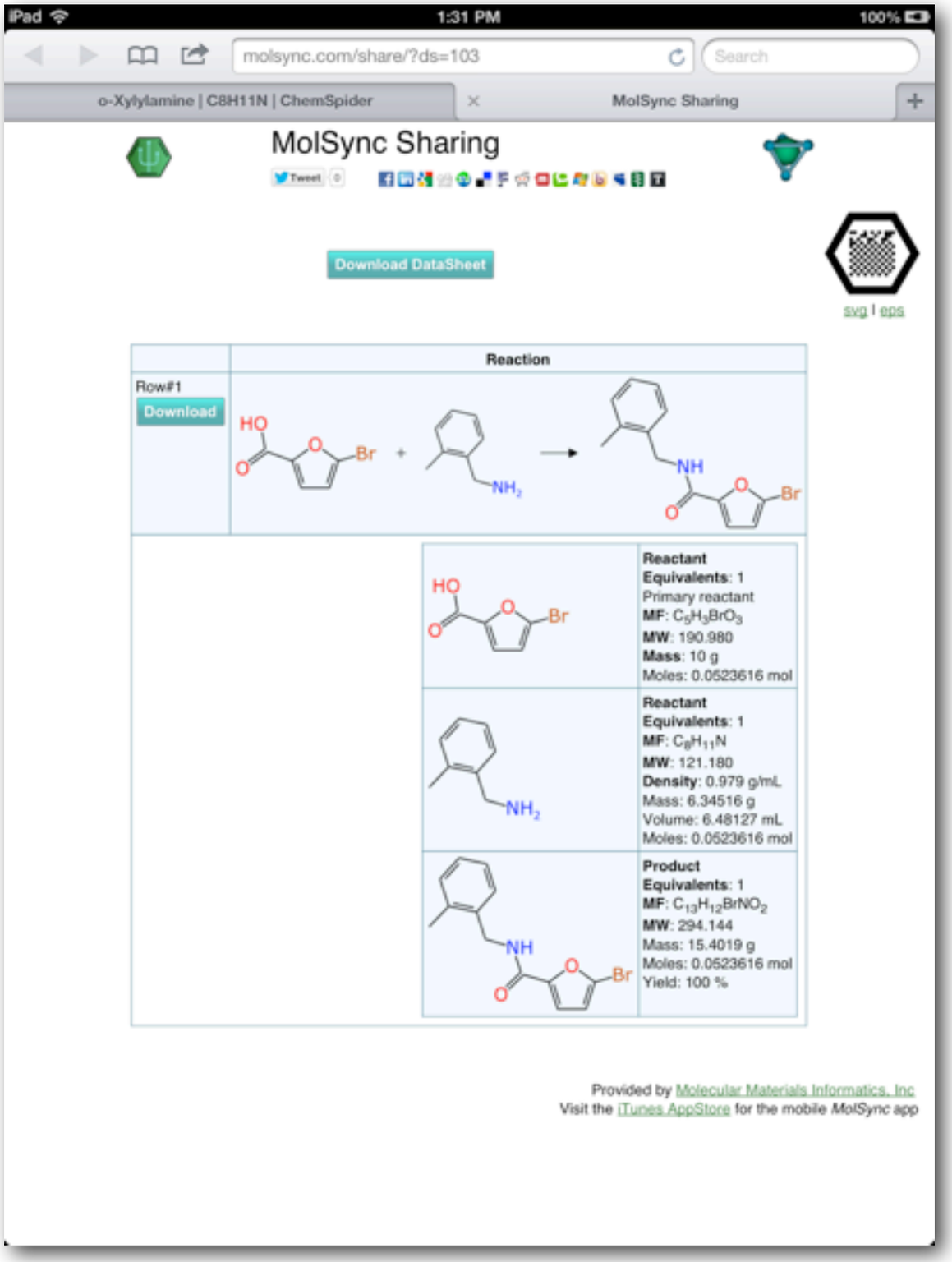
Public Sharing



- Upload scheme to *molsync.com*
- Generates publicly sharable link

Shared Content

- Stores chemical data in its native form:
  - molecules
  - reaction schemes
  - datasheets
- Can share, tweet, etc.
- Download data in many different forms...



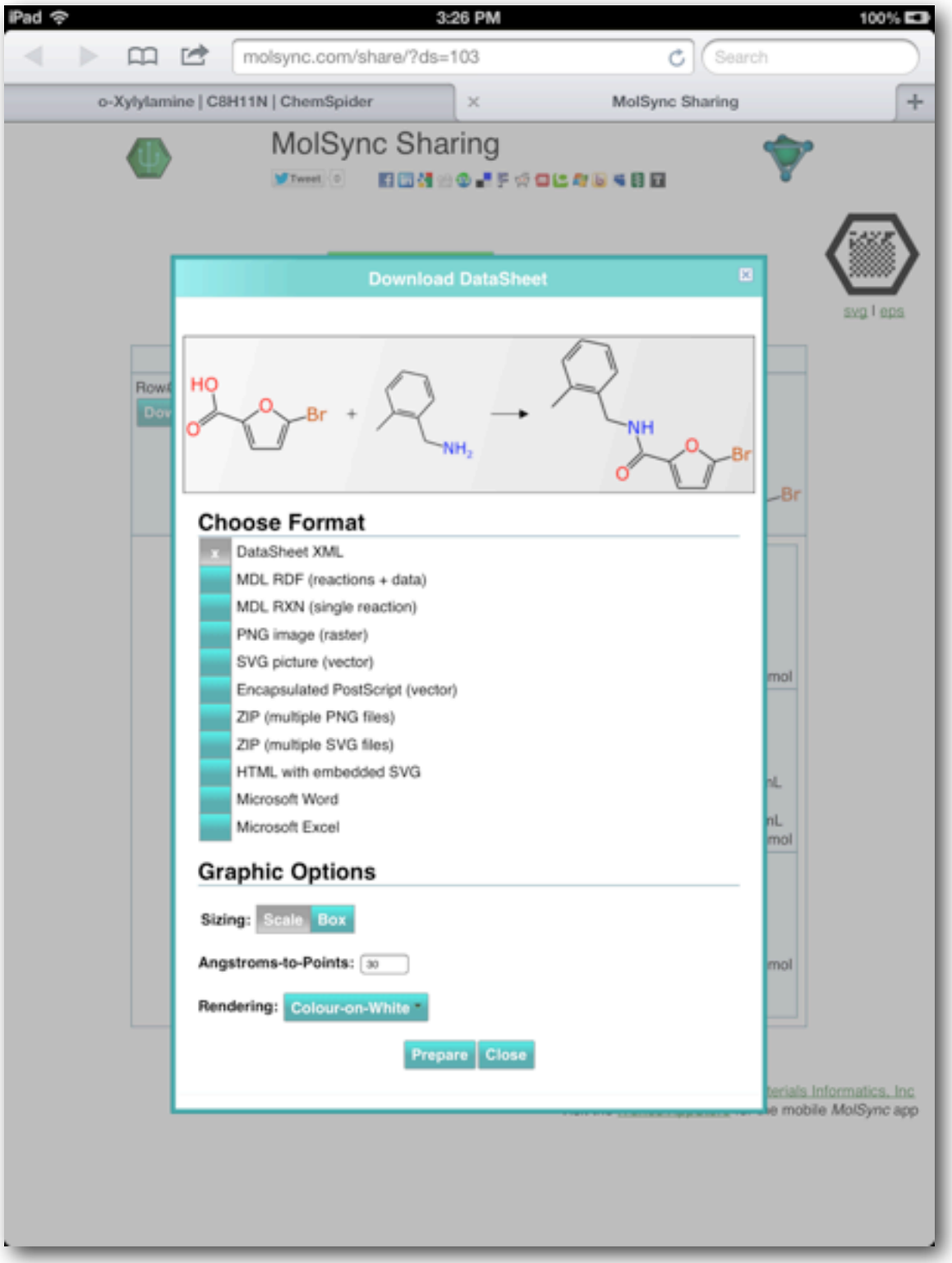
Shared Content

- Stores chemical data in its native form:

  - molecules
  - reaction schemes
  - datasheets

- Can share, tweet, etc.

- Download data in many different forms...



Numerous Formats

- Sharing site stores pure data, creates preview graphics dynamically... and interconversions

Molecules
- SketchEl
- MDL MOL
- CML

Reactions
- DataSheet XML
- MDL RXN
- MDL RDF

Collections
- DataSheet XML
- MDL SDfile

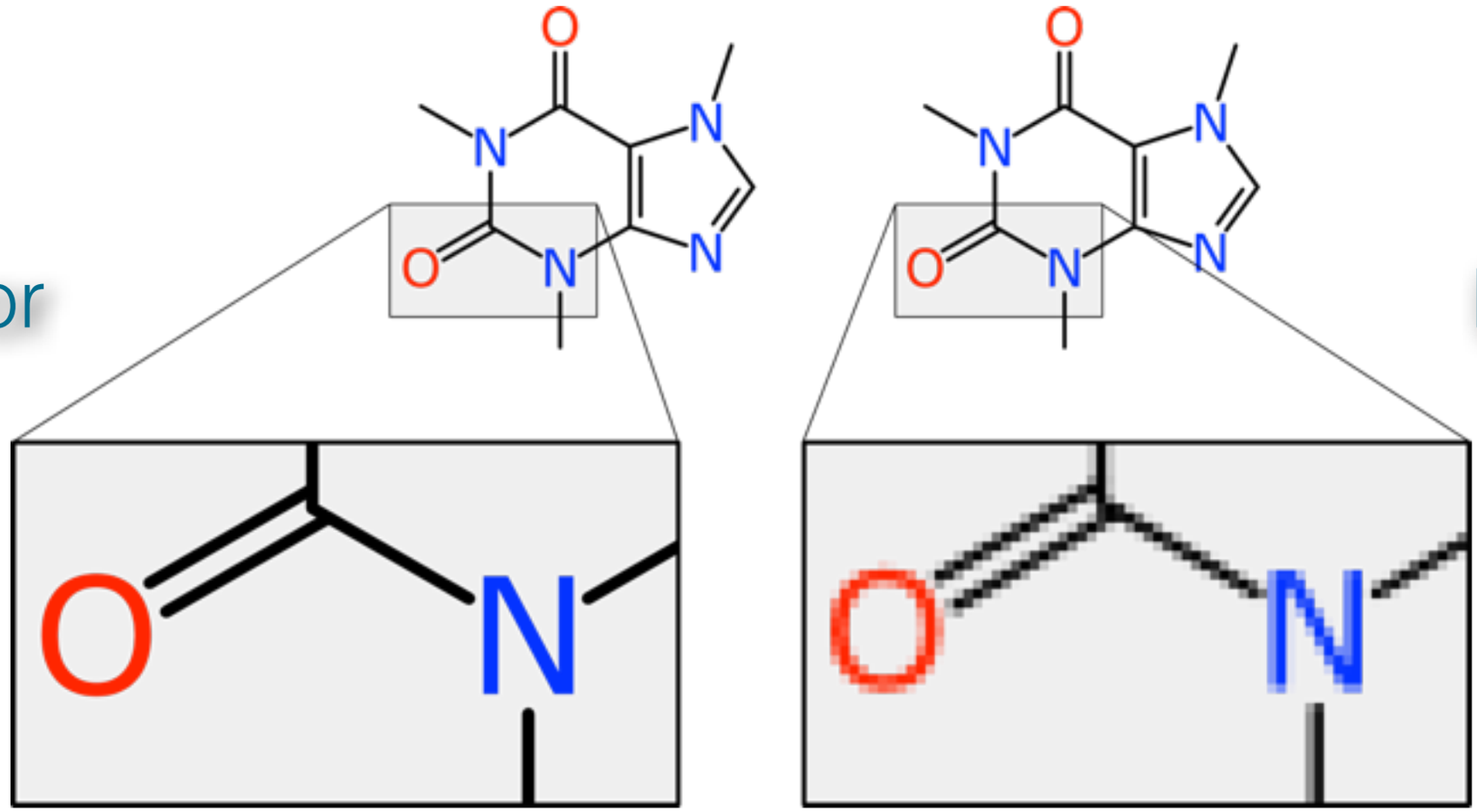
Raster Graphics
- PNG
- PNG ZIP

Vector Graphics
- SVG
- SVG ZIP
- HTML SVG
- EPS
- MS Word
- MS Excel

Vector Graphics

- Quality is perfect at all resolutions: screen, projector, web, printer, PDF...

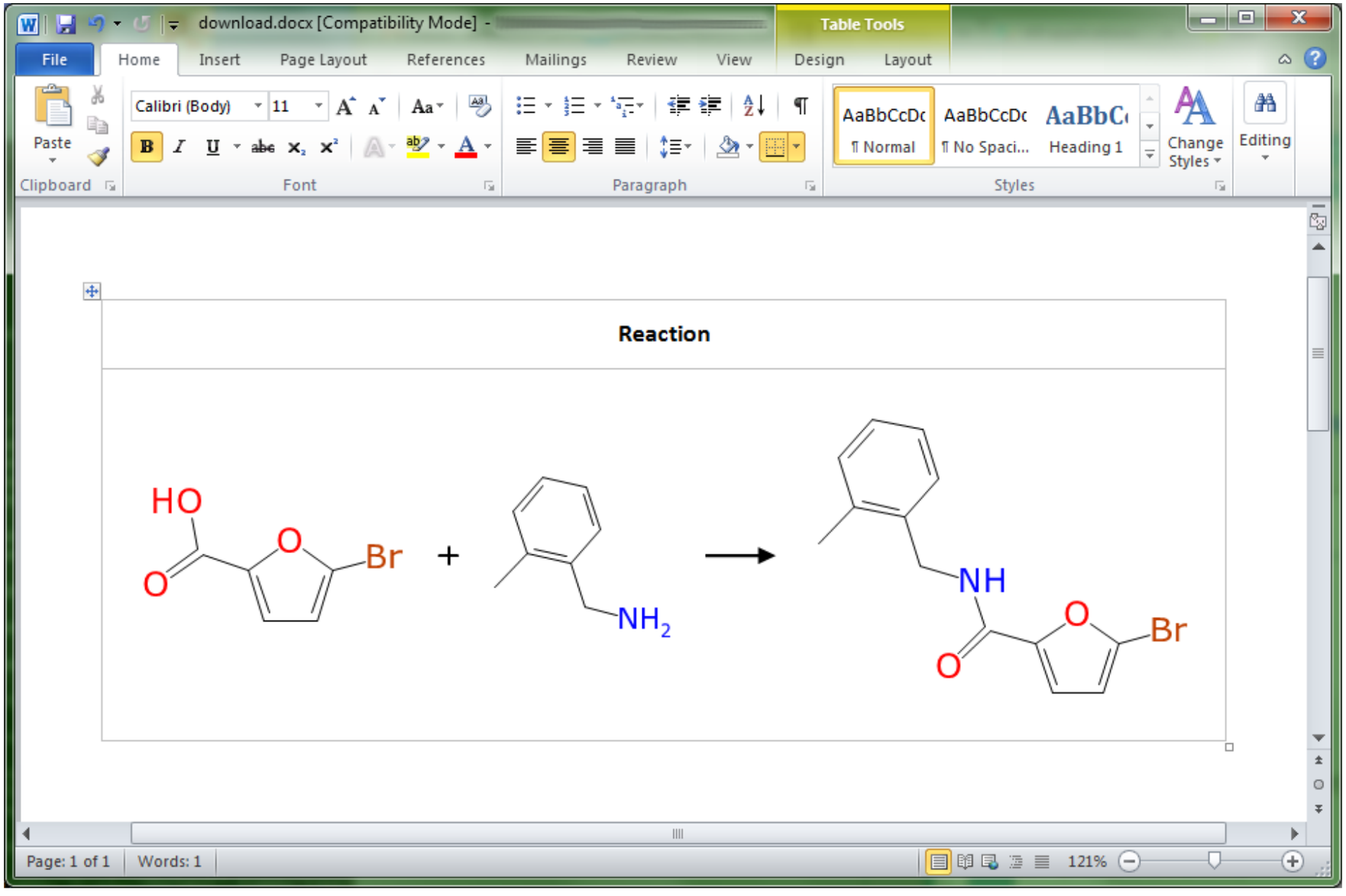
Vector



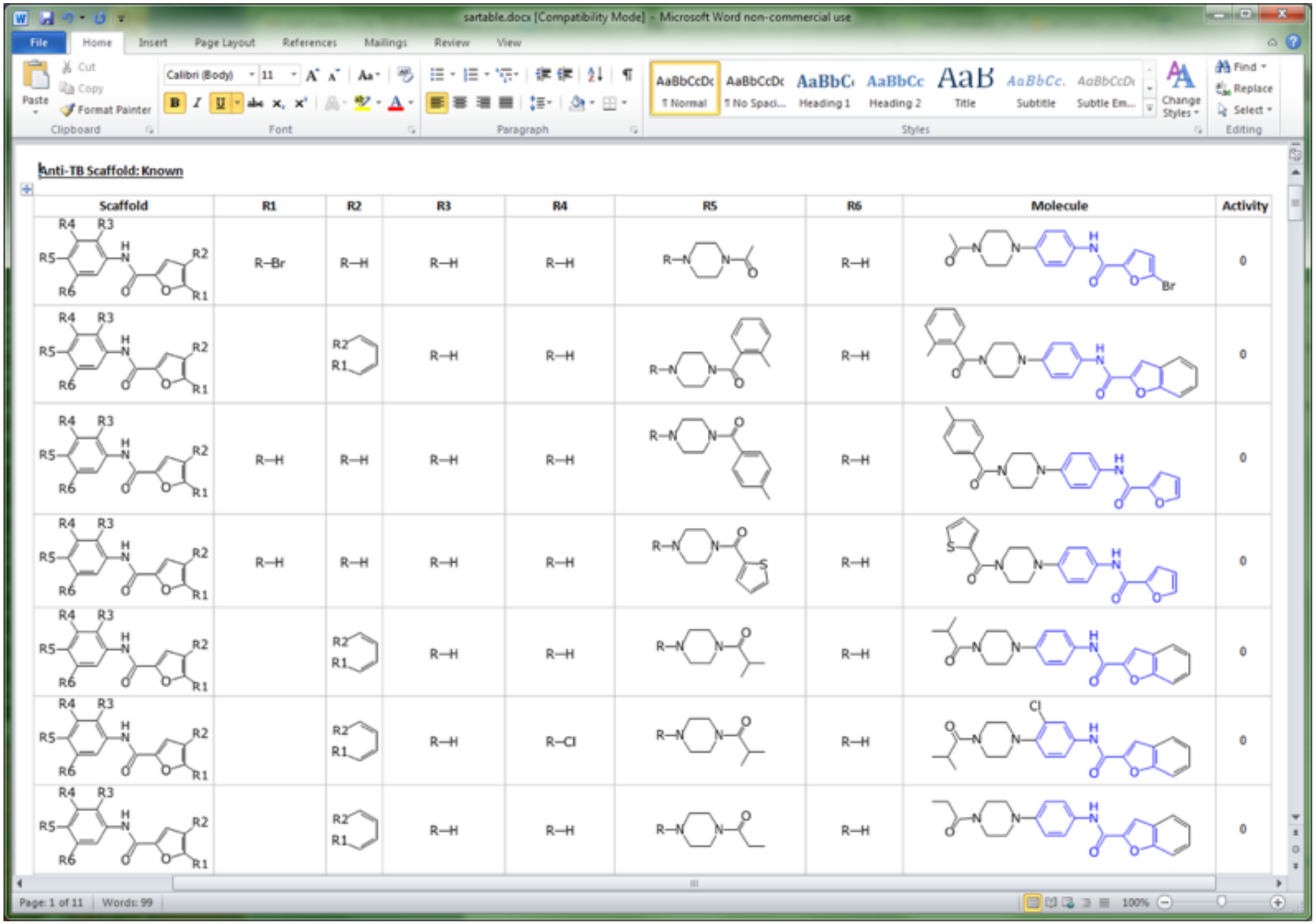
Bitmap

Microsoft Office Documents

- Independent implementation of **OOXML** spec
- Embedded graphics use vector **DrawingML**

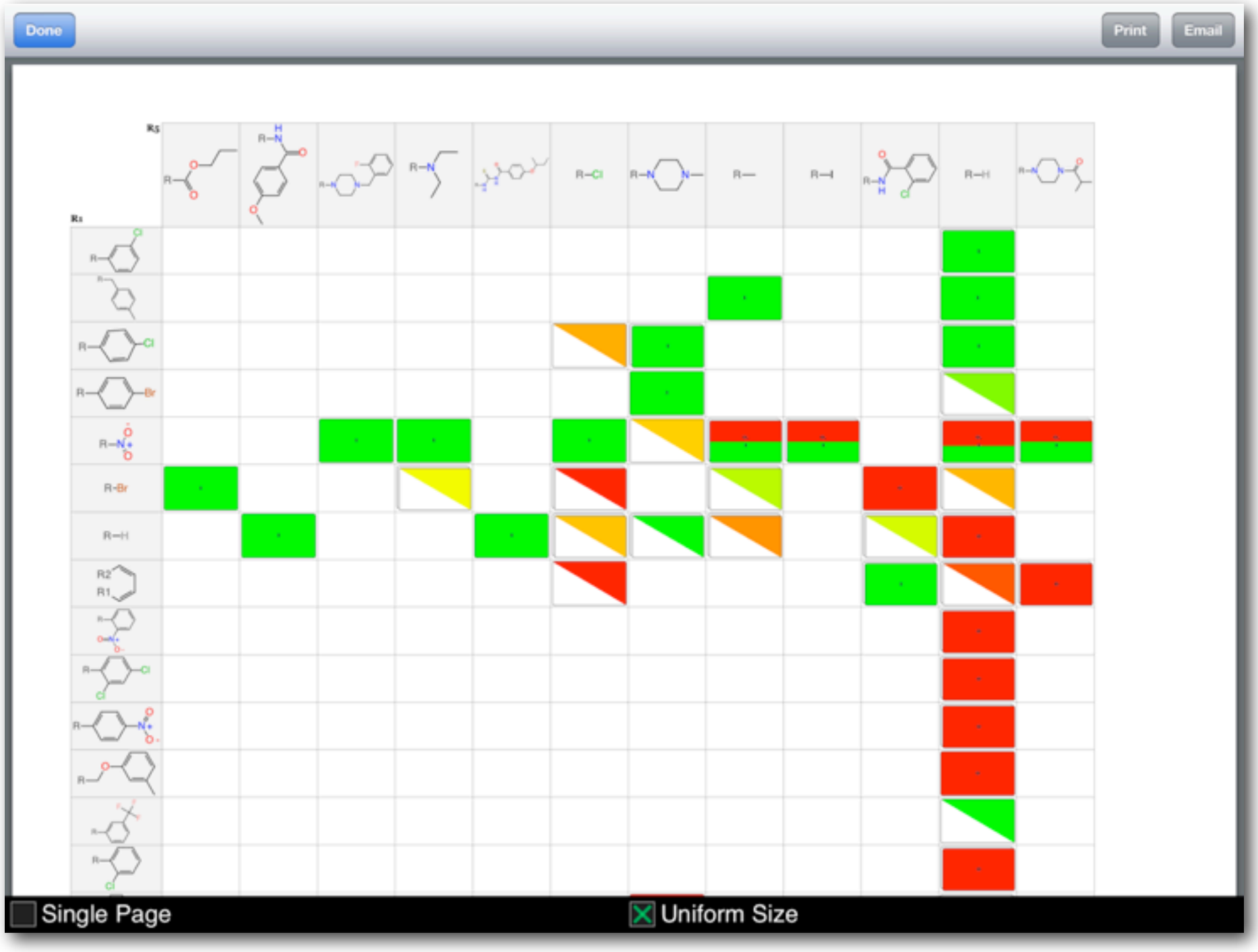


SAR Table Documents



- Scaffold/substituent tables for manuscripts

Matrix PDF

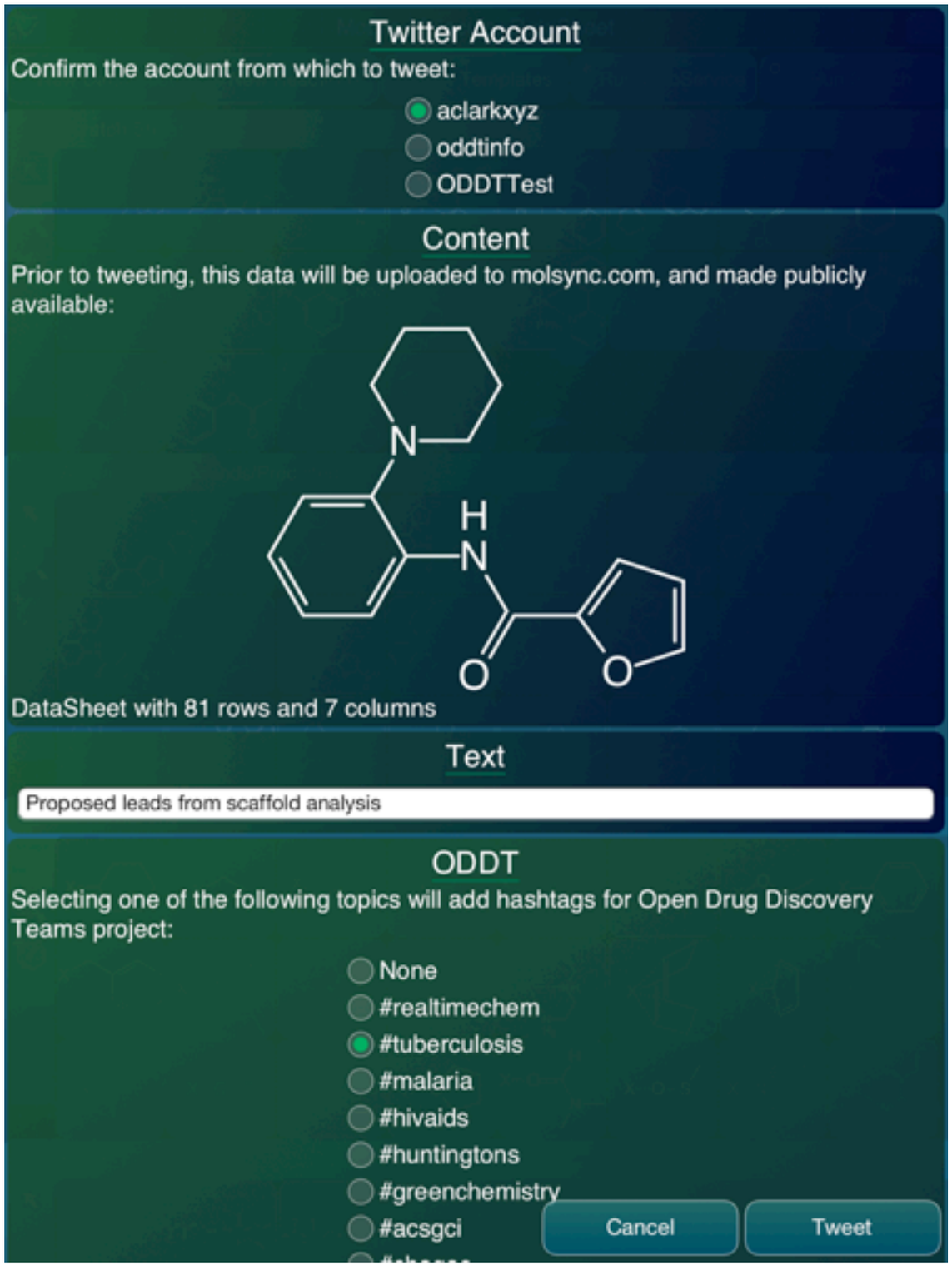


- Multi-page layout: print directly or email PDF

Tweeting

- Can tweet directly from apps like **MMDS**
- First upload data to **molsync.com**
- Then emit tweet with link
- Observe list of predefined hashtags:

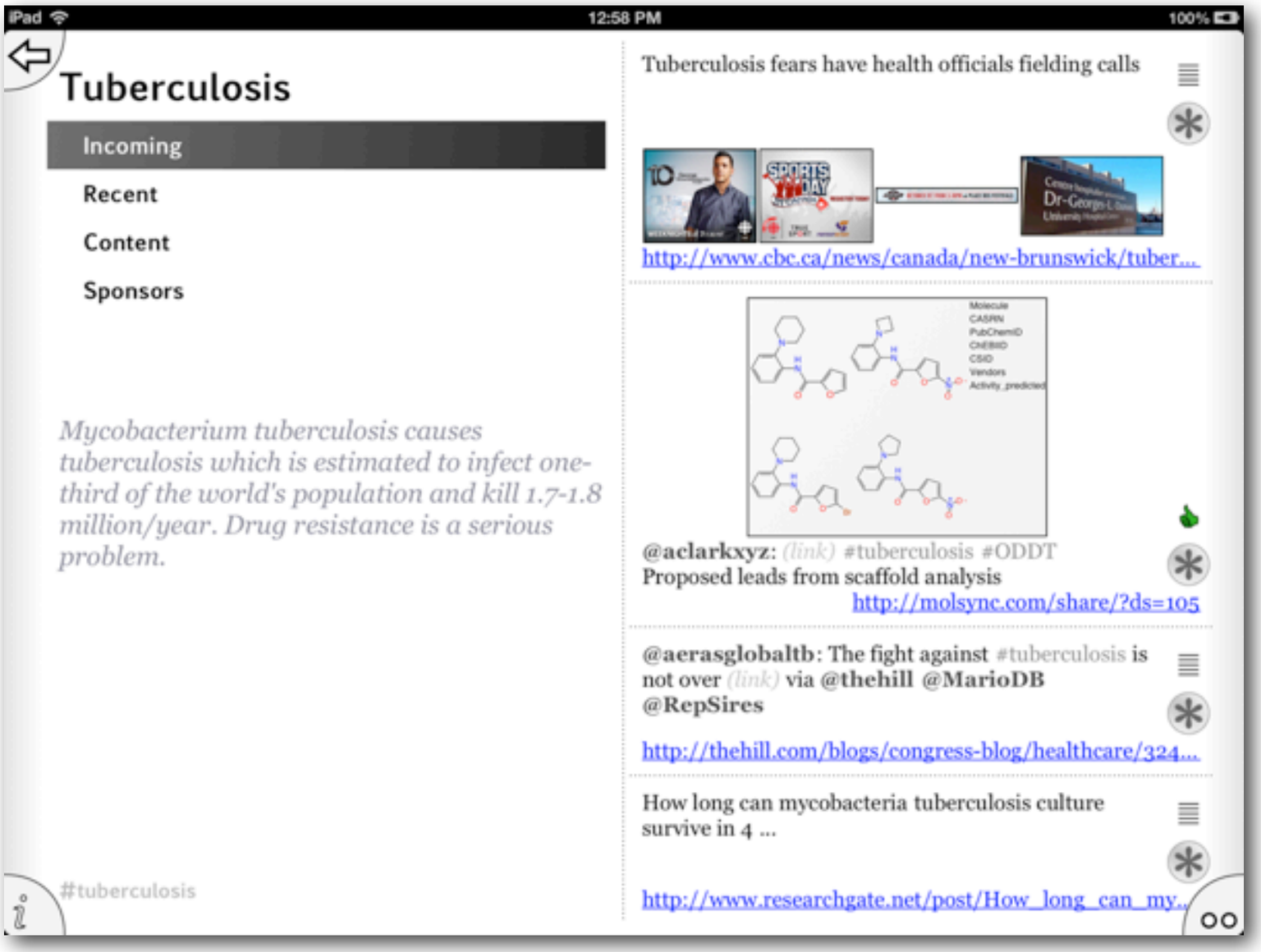
#tuberculosis



Open Drug Discovery Teams

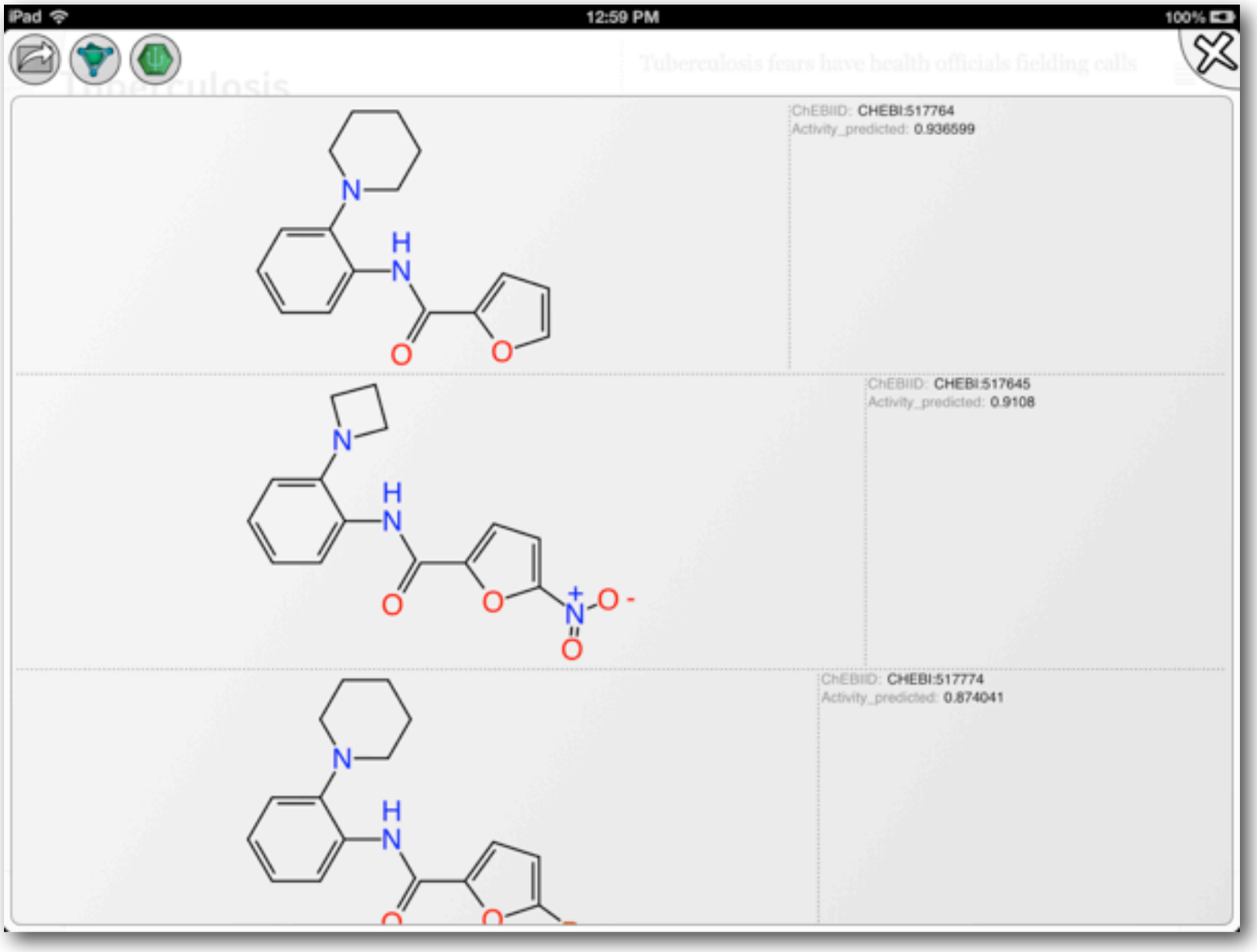
- Curation of open data, e.g. Twitter & RSS feeds
- Rare & neglected diseases, precompetitive areas

Harvested Tweet



- Tweet got *harvested* into Tuberculosis topic
- Inline preview browsed, with other thumbnails

Chemistry Aware



- The app understands chemical data
- Users can use the data, *open with* in apps...

Rinse & Repeat

- Defer to experimentalists: measure activity for selected compounds with predictions

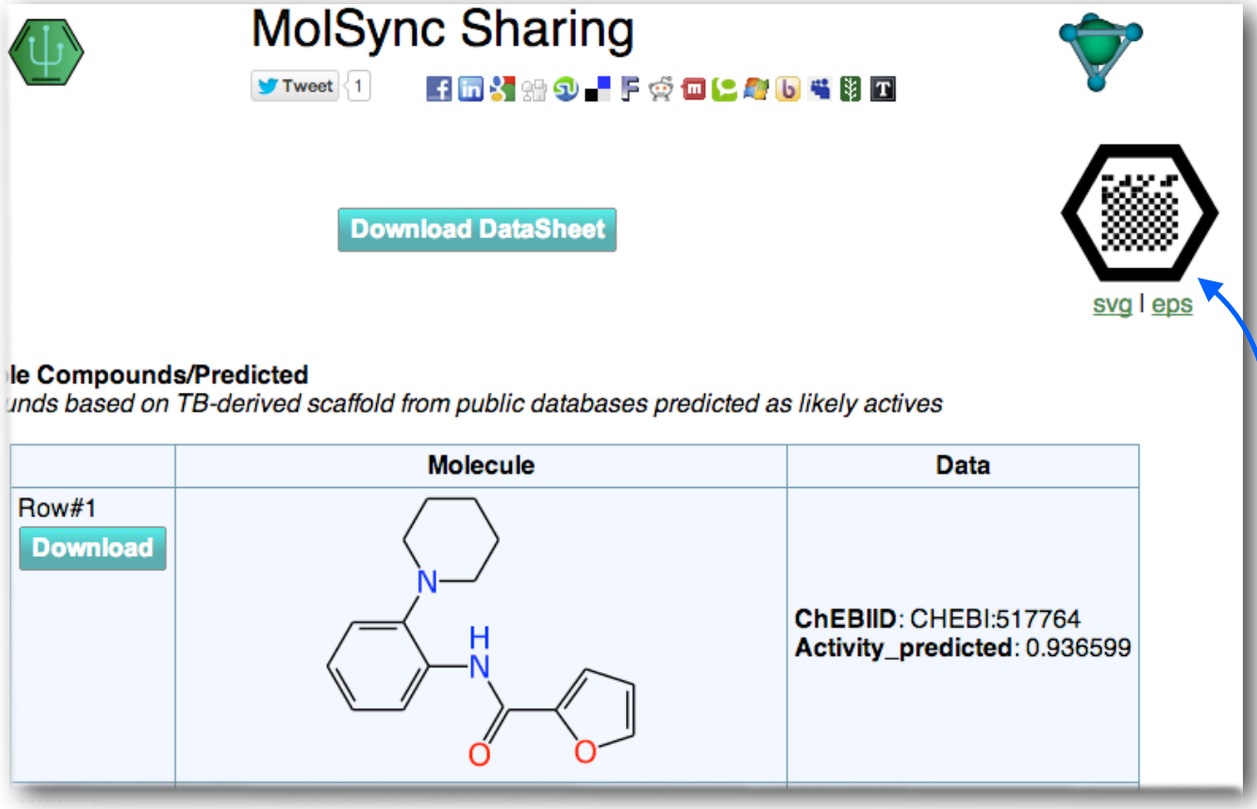
- Return to **SAR Table** app:
  - provide activity results
  - visualise matrix plot with actual data added
  - rebuild model, regenerate predictions
  - iteratively refine hypothesis

- Done using $50 worth of software and tablet

- Workflow is appropriate for use by experimentalists

Living Molecules

- Going beyond desktop-era functionality
- Devices have:
  - long battery life
  - bluetooth
  - location awareness
  - **camera!**



- Observe the strange symbol: **molecular glyph**

Molecular Glyphs

- Recognition sequence: point camera
- View content, import, manage, utilise, export

Conclusion

- Cheminformatics workflows historically the role of specialists: expensive and/or complex

- Mobile apps are much cheaper and much more accessible to experimentalists

- Mobile+cloud can:
  - **replace** simple-to-medium tasks
  - **coexist** with complex tasks run on desktop software

- Other advantages:
  - anywhere/anytime portability
  - excellent collaboration and sharing
  - non-existent installation or maintenance burden

Acknowledgments

- Antony Williams
- Sean Ekins
- Joel Freundlich
- NETTAB / Bioinformatics Italian Society
- InfoChem, Eidogen–Sertanty
- Inquiries to **info@molmatinf.com**

http://molmatinf.com/venice.html

MOLECULAR MATERIALS INFORMATICS

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

@aclarkxyz