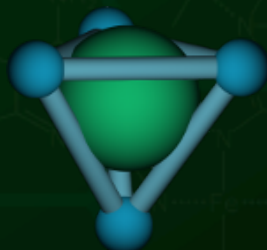


Mobile + cloud: a viable replacement for desktop cheminformatics?

Dr. Alex M. Clark

April 2013



© 2013 Molecular Materials Informatics, Inc.

<http://molmatinf.com>

Two platform stacks



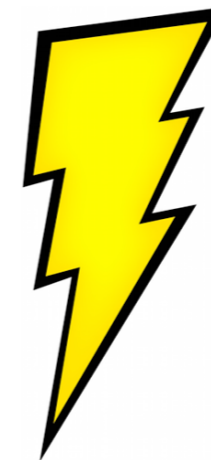
desktop/
laptop



phone/
tablet



file
server



web
services



compute
cluster



cloud
resources

What's the point?

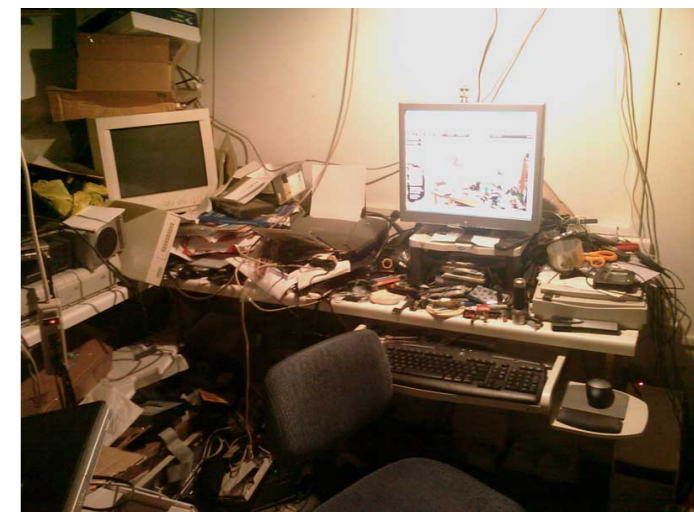
● Mobile computing

- a phone or tablet can follow you just about anywhere
- almost all other software is migrating to mobile...
- ... should chemistry be the last reason to use a PC?
- it's time for an overhaul anyway (the 1980s are over)



● Cloud computing

- economies of scale: pay as you go, as much as you need
- no air conditioned rooms full of servers
- no junk appliances under the desk
- no administrators



Mobile interface



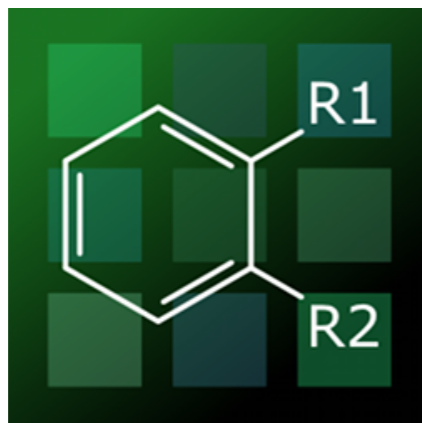
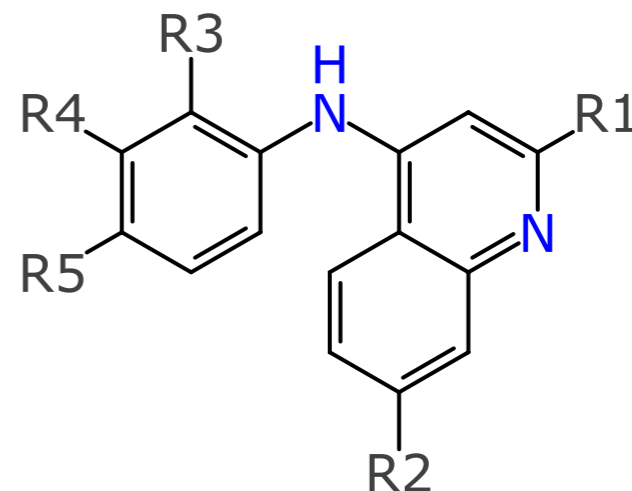
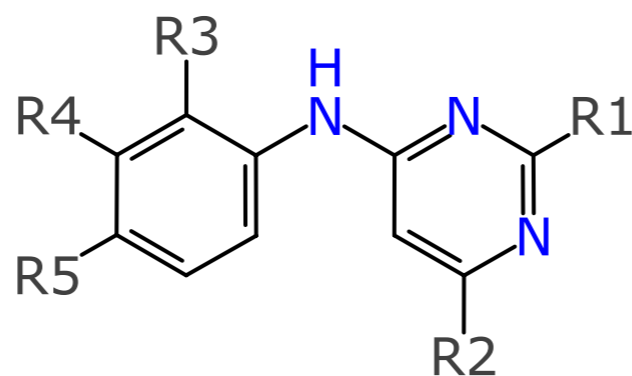
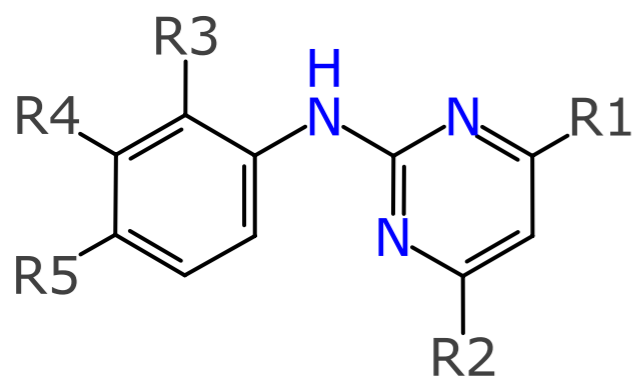
- **Less** compute power: just enough for interactivity
- **Limited** storage: think of it as a cache for network
- **Code** portability: zero legacy code reuse
- **Clumsy** touchscreen: very tight screen real estate planning
- **Different** style: UI paradigms must be rethought from scratch
 - e.g. conventional molecule sketchers don't work well...
 - ... different set of concepts, learning new interface

Cloud computation

- Offloading heavy calculations to a webservice can bring mobile interfaces up to parity with traditional stack
- Webservices either:
 - **stateless**: short tasks, finished before HTTP timeout
 - **asynchronous**: client maintains & resubmits state
 - **task state**: temporary state for duration of task
 - **persistent state**: login as a user to access content
- Workflow based on short exchanges: split evenly for efficiency and user experience
- Not easy to do, but worthwhile

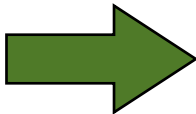
Workflow

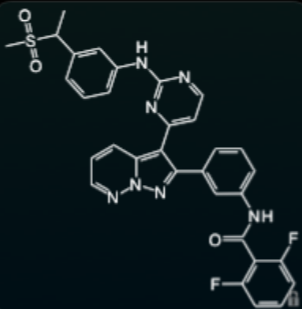
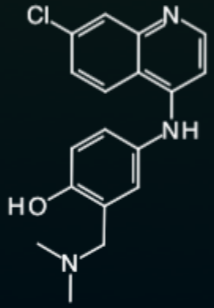
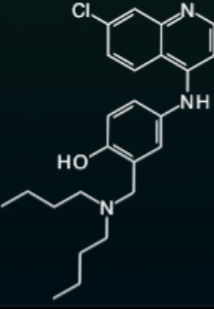
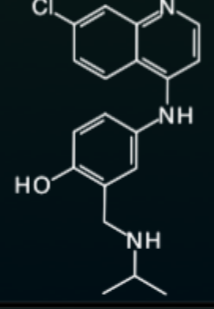
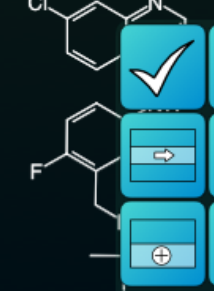
- Ingredients:
 - selected compounds + activity from the GSK malaria leads (248 compounds)
 - three known scaffolds:



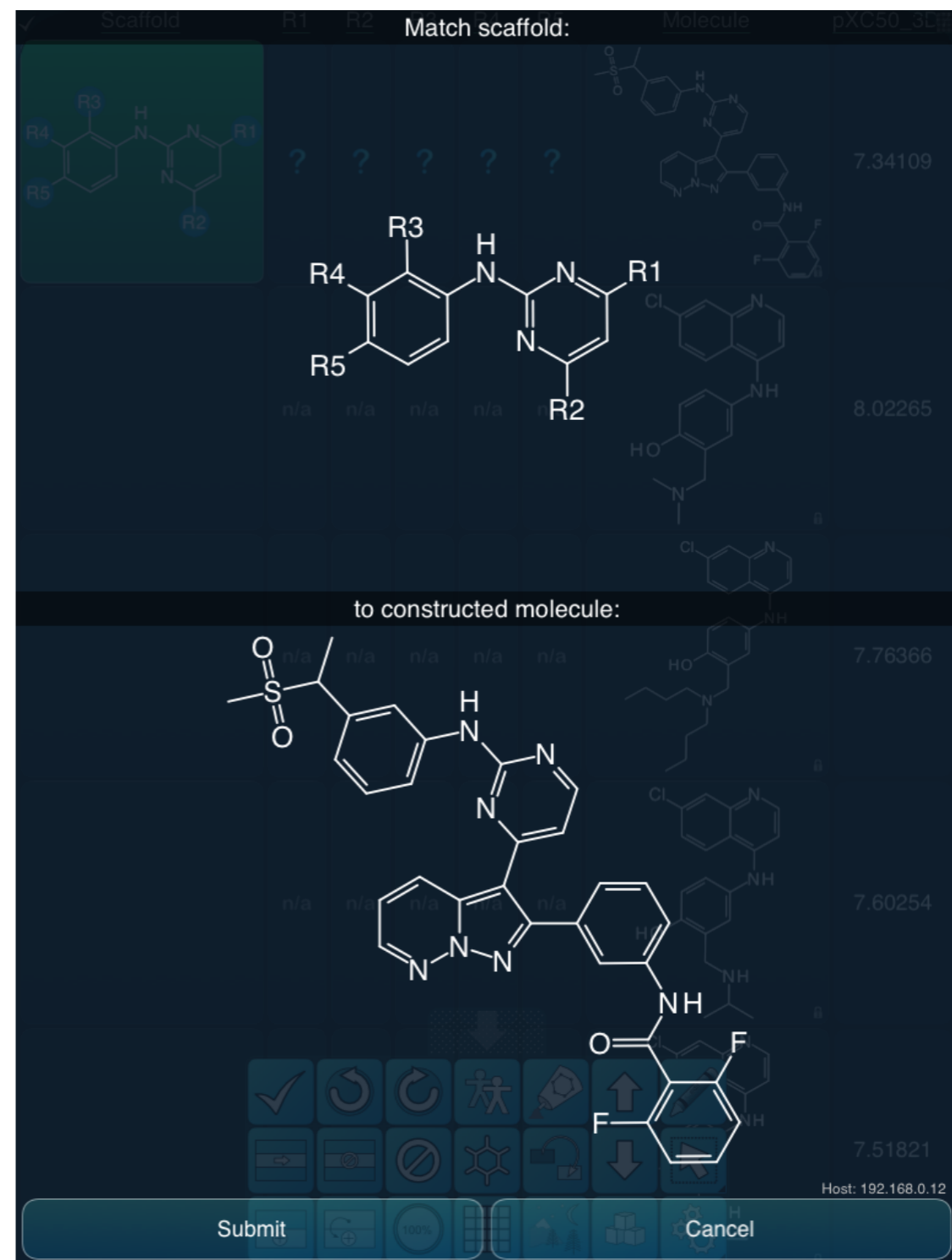
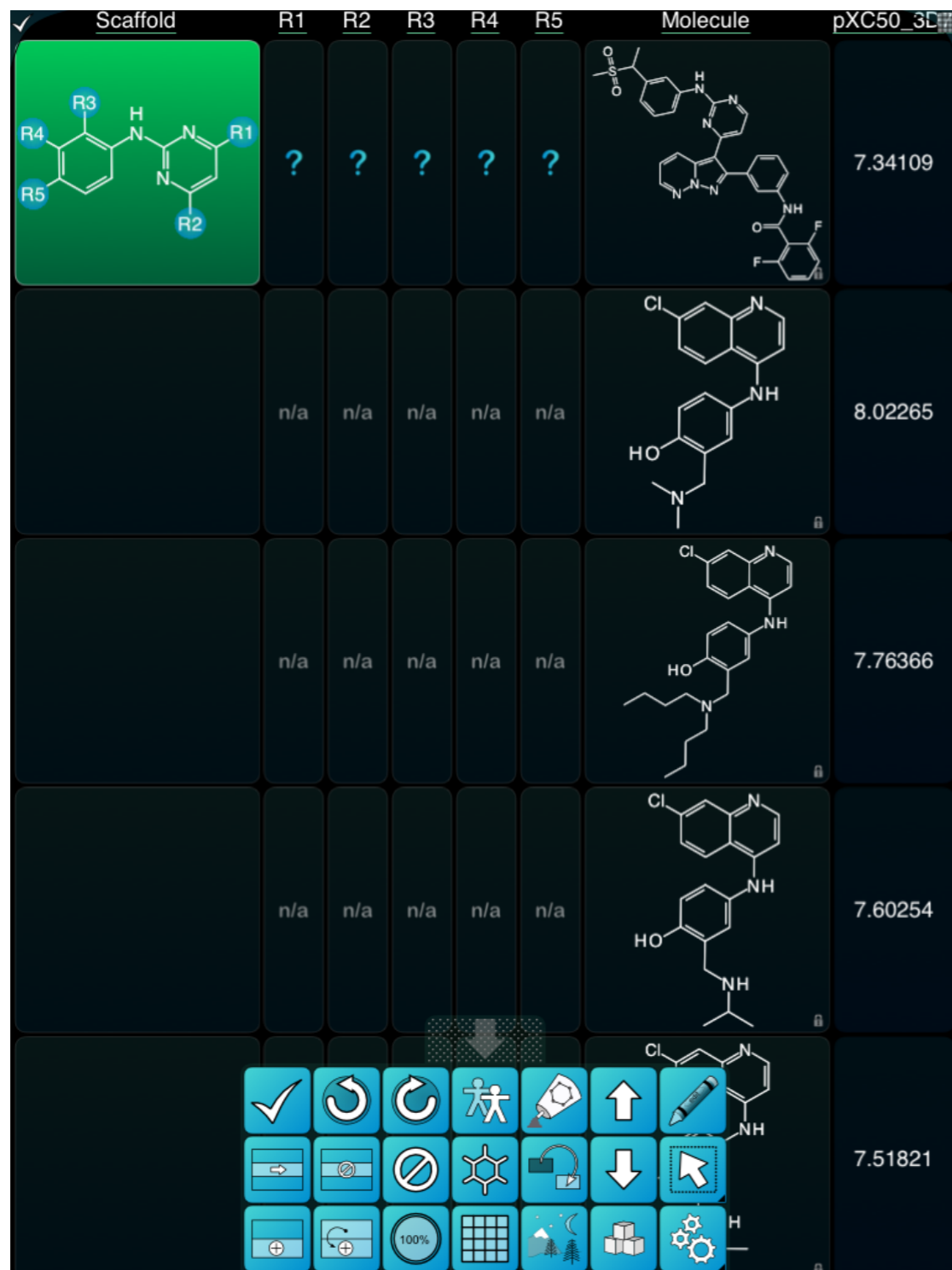
- The **SAR Table** app
 - running on an *iPad*
 - calling upon cloud-based resources

Data importing

- Download chemical structures from:
 - email attachments
 - arbitrary URL
 - network filesystem (e.g. *Dropbox*)
- Directly into **SAR Table** app 
- Passing small documents to/from apps is easy and effective

Scaffold	Molecule	pXC50_3D7
		7.34109
		8.02265
		7.76366
		7.60254
		

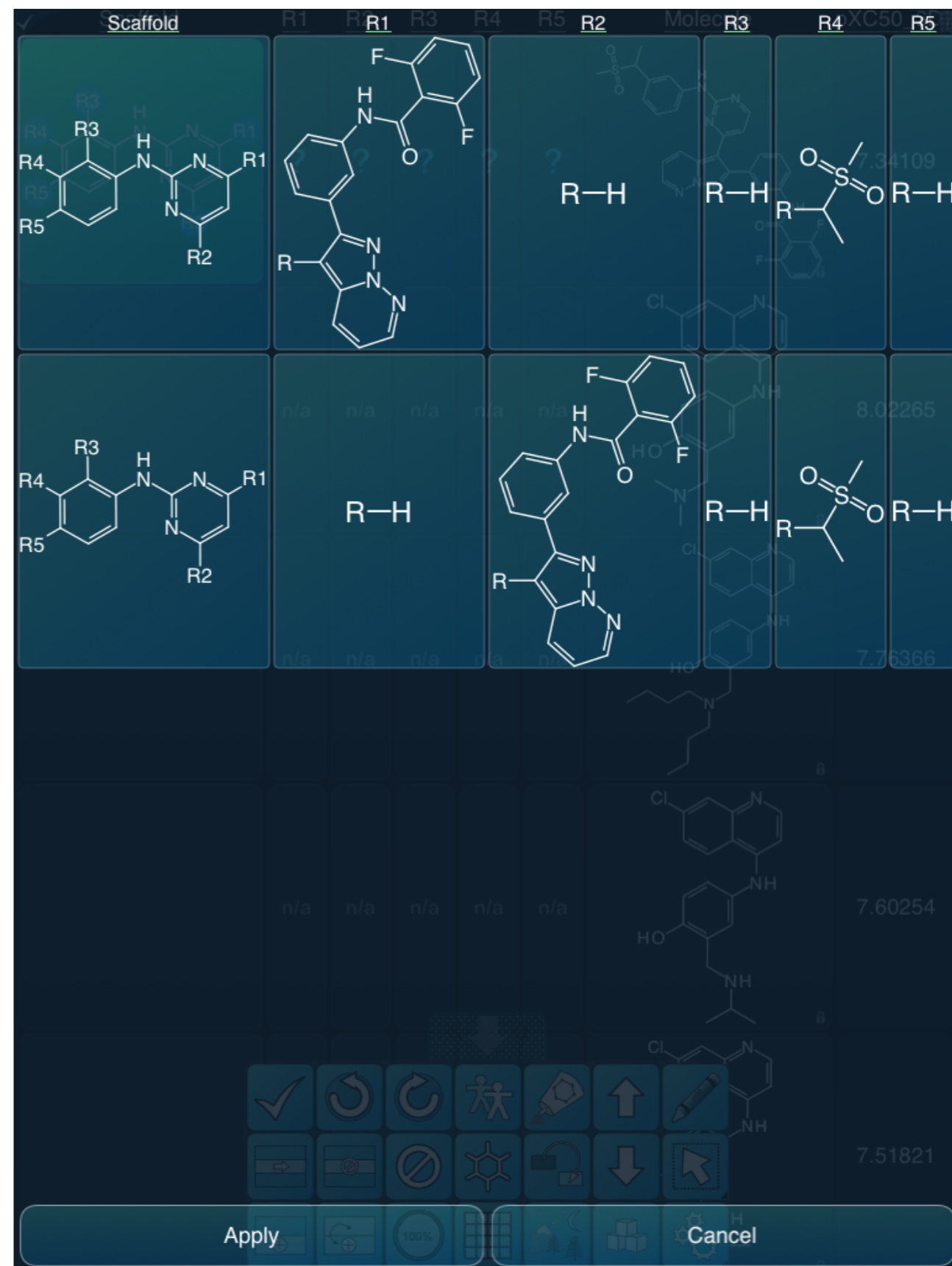
Specify scaffold



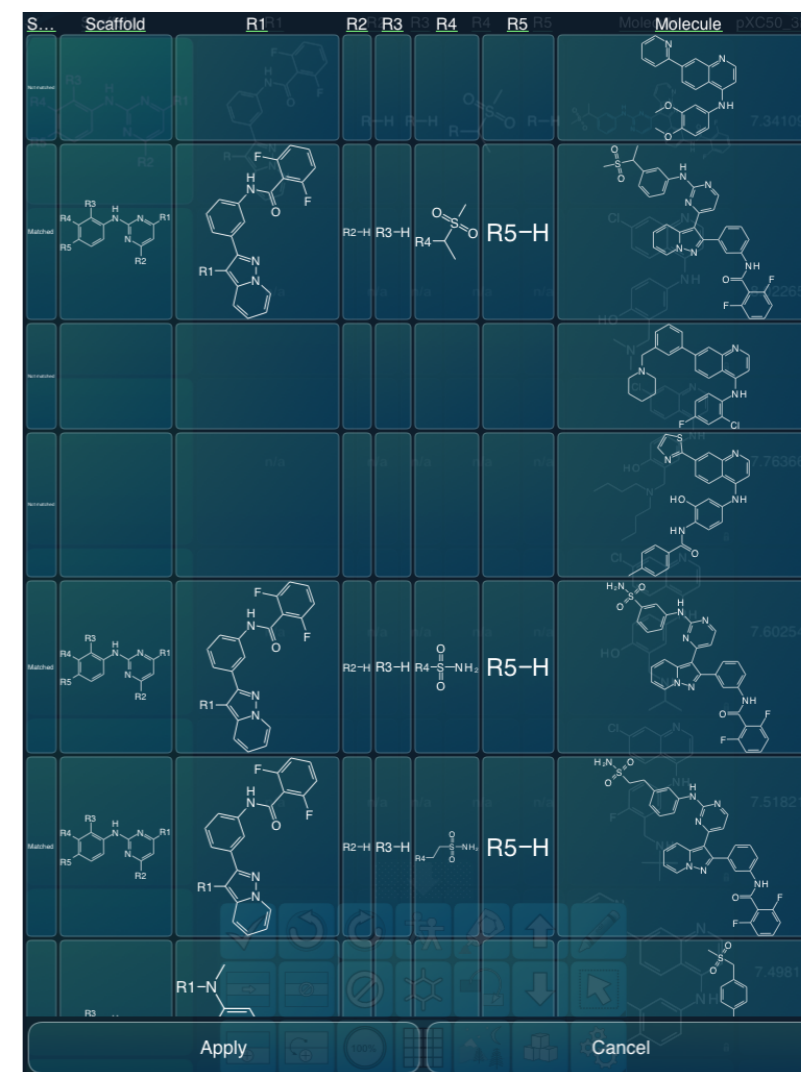
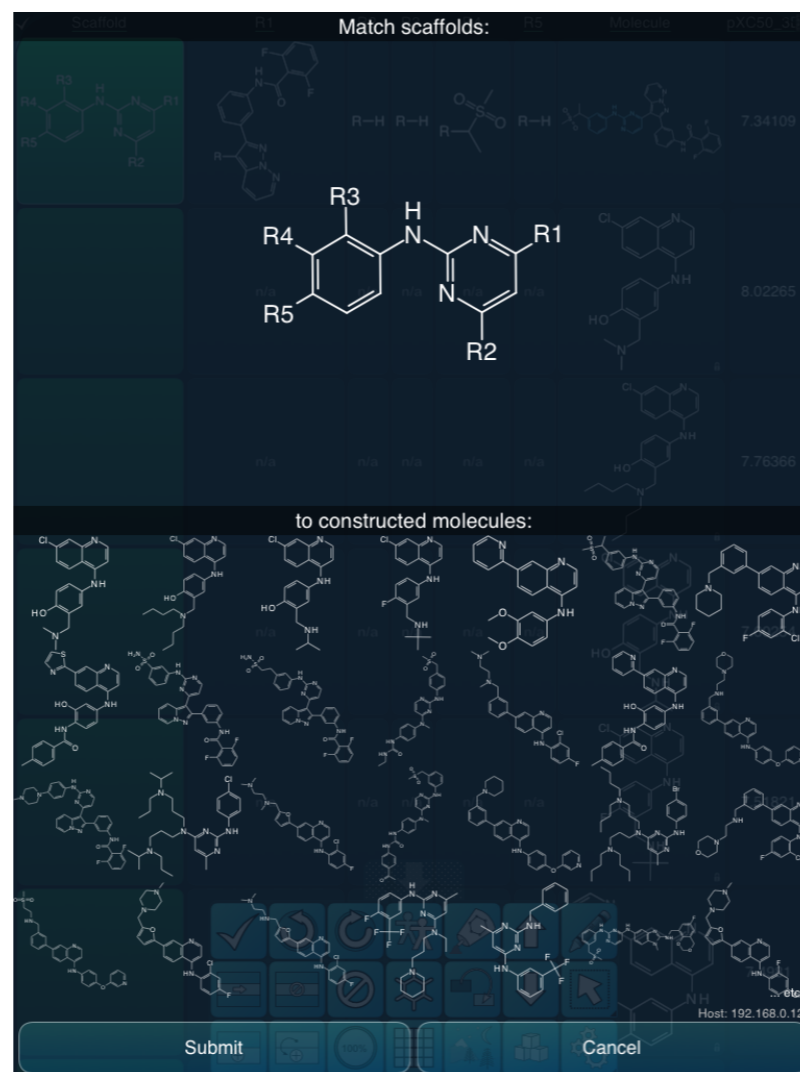
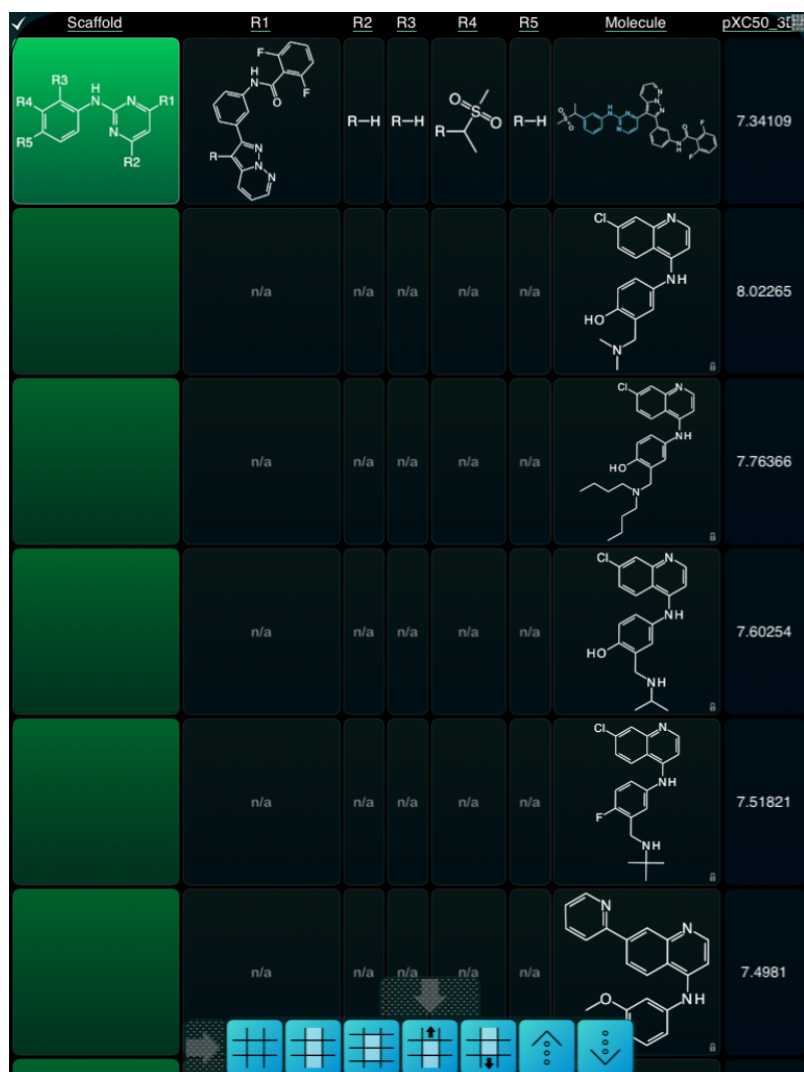
- Draw scaffold for first structure, select scaffold matching action

Scaffold matching

- Produces all useful combinations of scaffold + R-groups
- User selects from the degenerate possibilities
- Defers to a webservice: *molsync.com*
- Service properties:
 - **complex** calculation
 - **short** (<< 30 seconds)
 - **stateless**
 - **small** input, **small** output

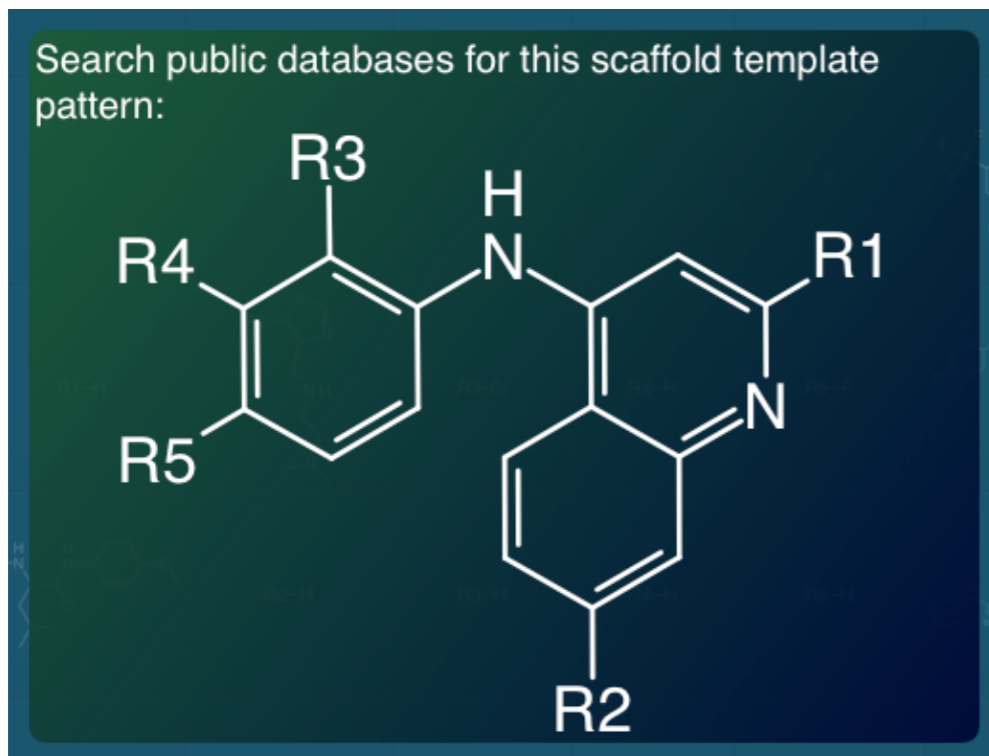


Multiple matching

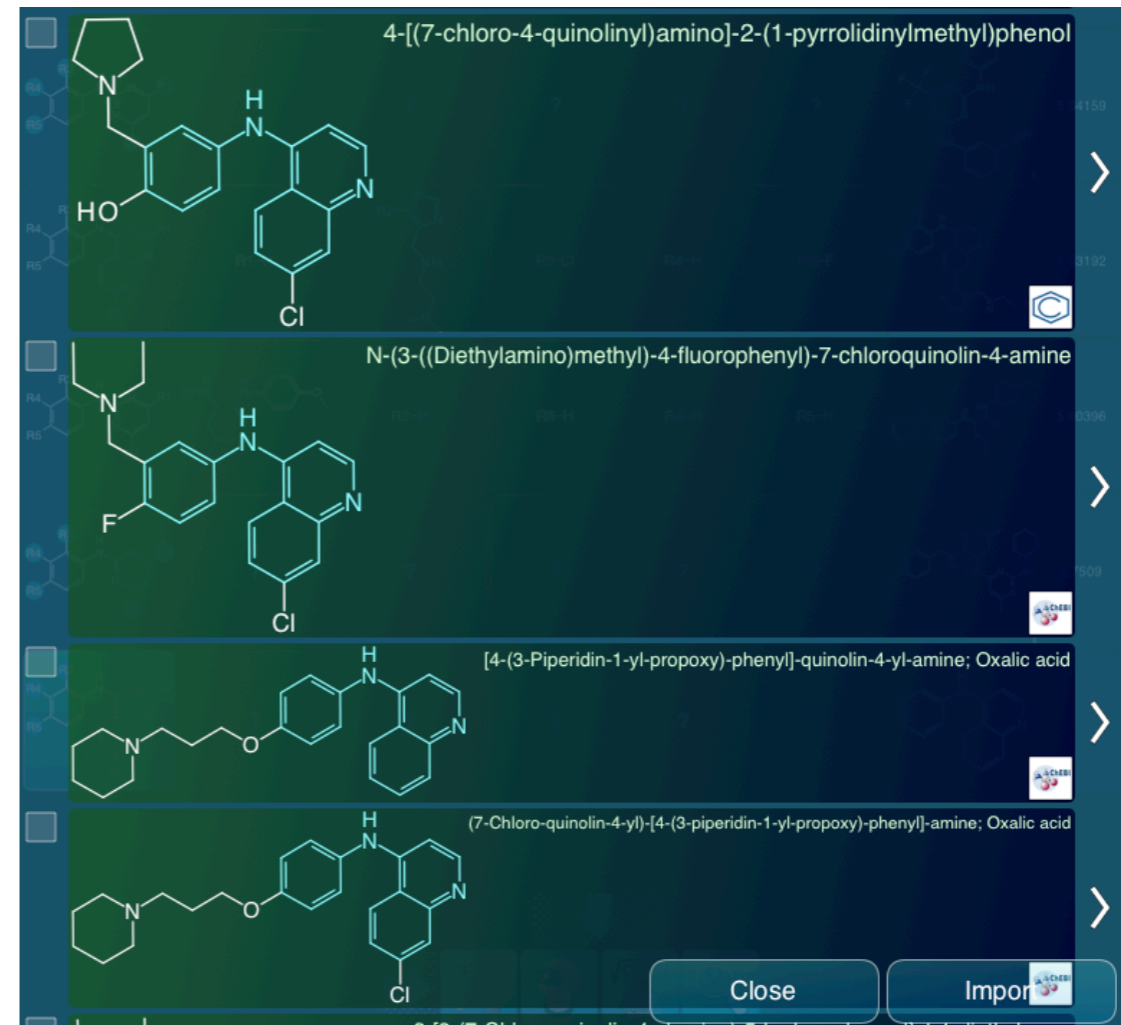


- Can try to match many scaffolds to many molecules
- Similar webservice request (*complex, fast, stateless, small data*)
- Iterative dance between **app** interface + **web** service:
 - algorithm applies chemical structure logic, user resolves degeneracy

Searching for more

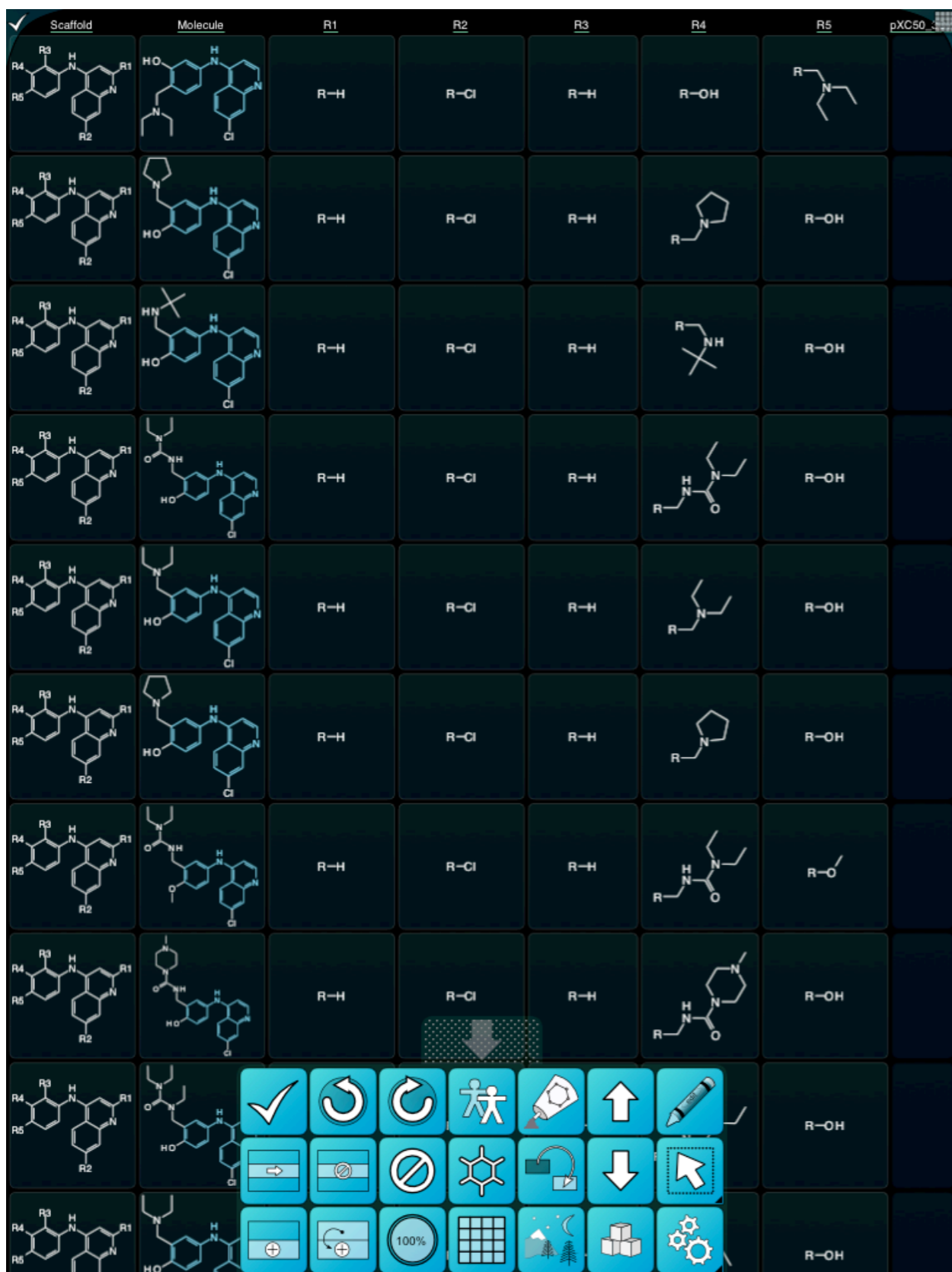


- Search for templates:
 - **ChEBI**
 - **PubChem**
- Intermediate webservice layered on top of publicly available search capabilities



- Service properties:
 - **complex** labour-intensive
 - **medium** speed (< 3 minutes)
 - **stateful** reference data
 - **small** input, **small** output

Template results



- Scaffold & R-group decomposition handled automatically by the meta-layer
- Now have numerous additional compounds in the table
- Duplicates excluded
- Additional compounds are all known:
 - lookup recipe
 - purchase from vendor
- Only thing missing is activity...

Activity visualisation

- Setup a **scheme**...

- Units
- Transform
- Colours
- Range

Edit Property Scheme

Name
pXC50

Units
Field name:

Transform-to-linear
 None
 Log

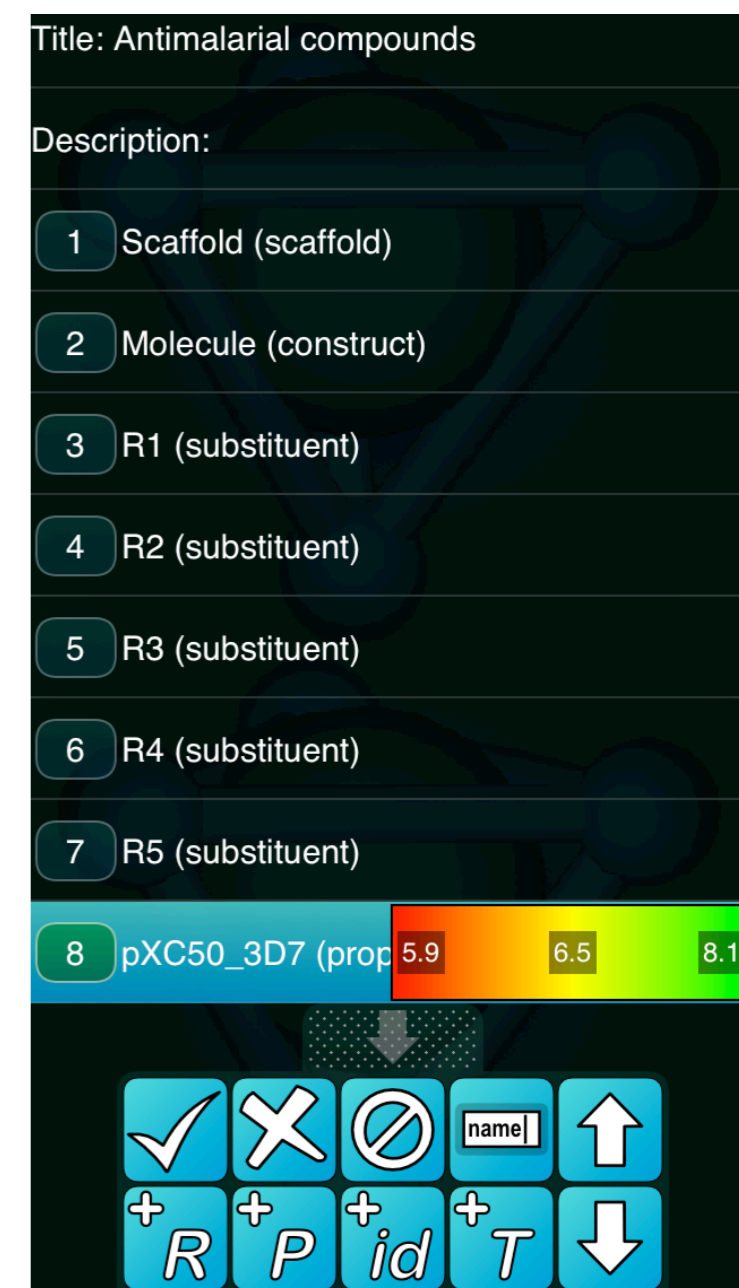
Colour triplet
HeatMap

Low
5.9

Medium
6.5

High
8.1

Accept Reject



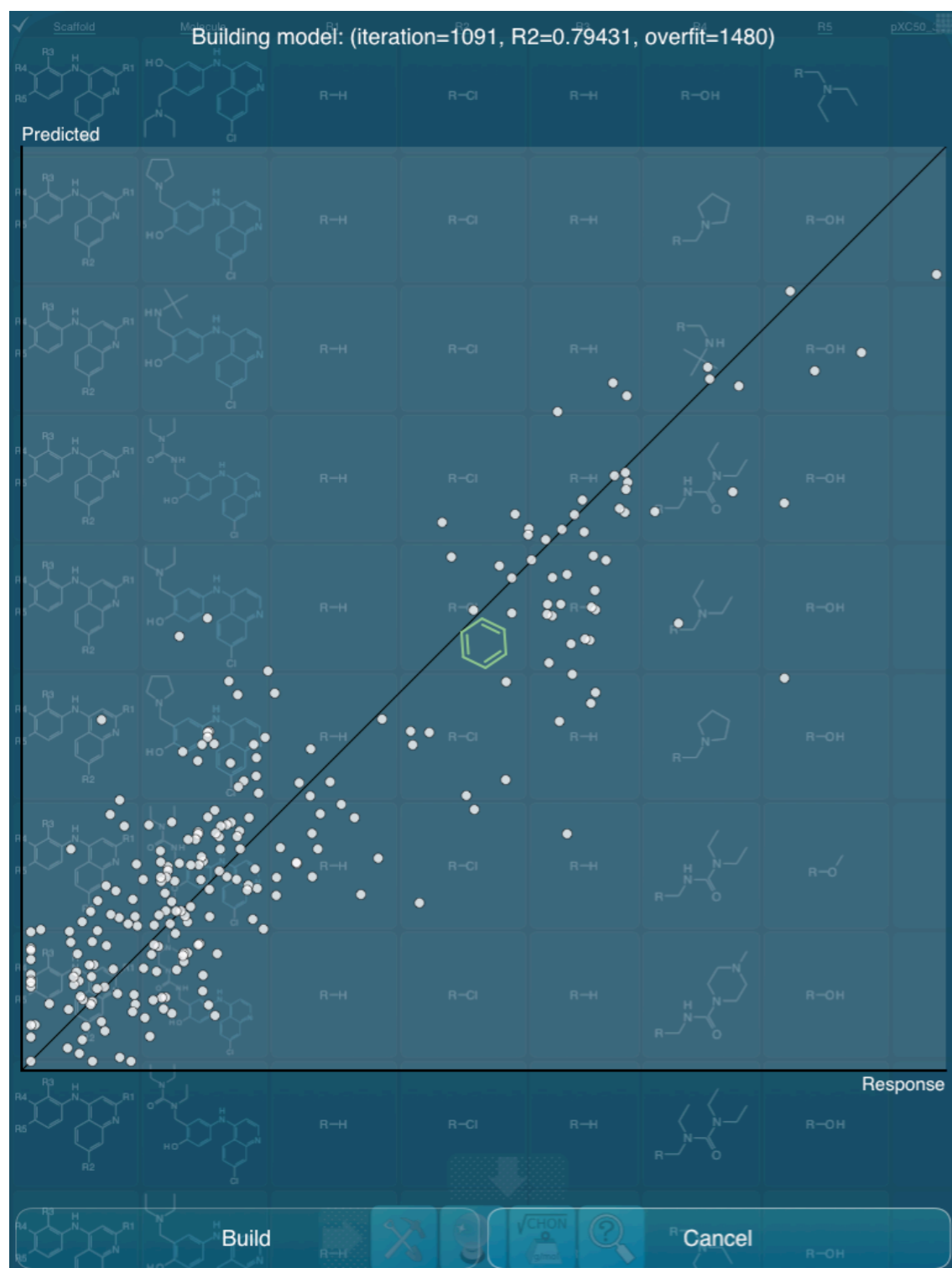
- Customise for the data

- All these leads quite active

- Define 5.9 to 6.5 as the low activity range (red to yellow)

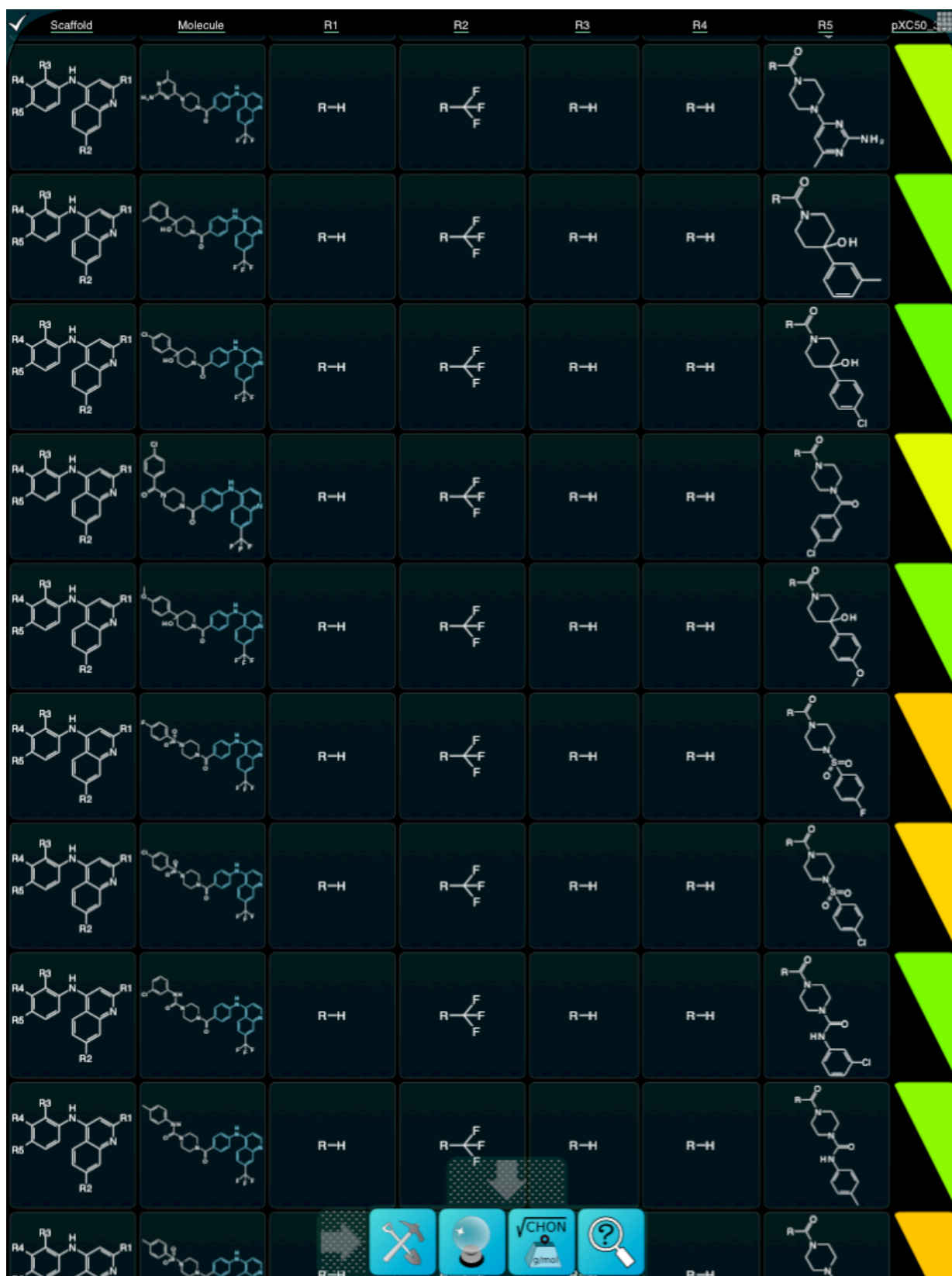
- Define 6.5 to 8.1 as the high activity range (yellow to green)

Model building



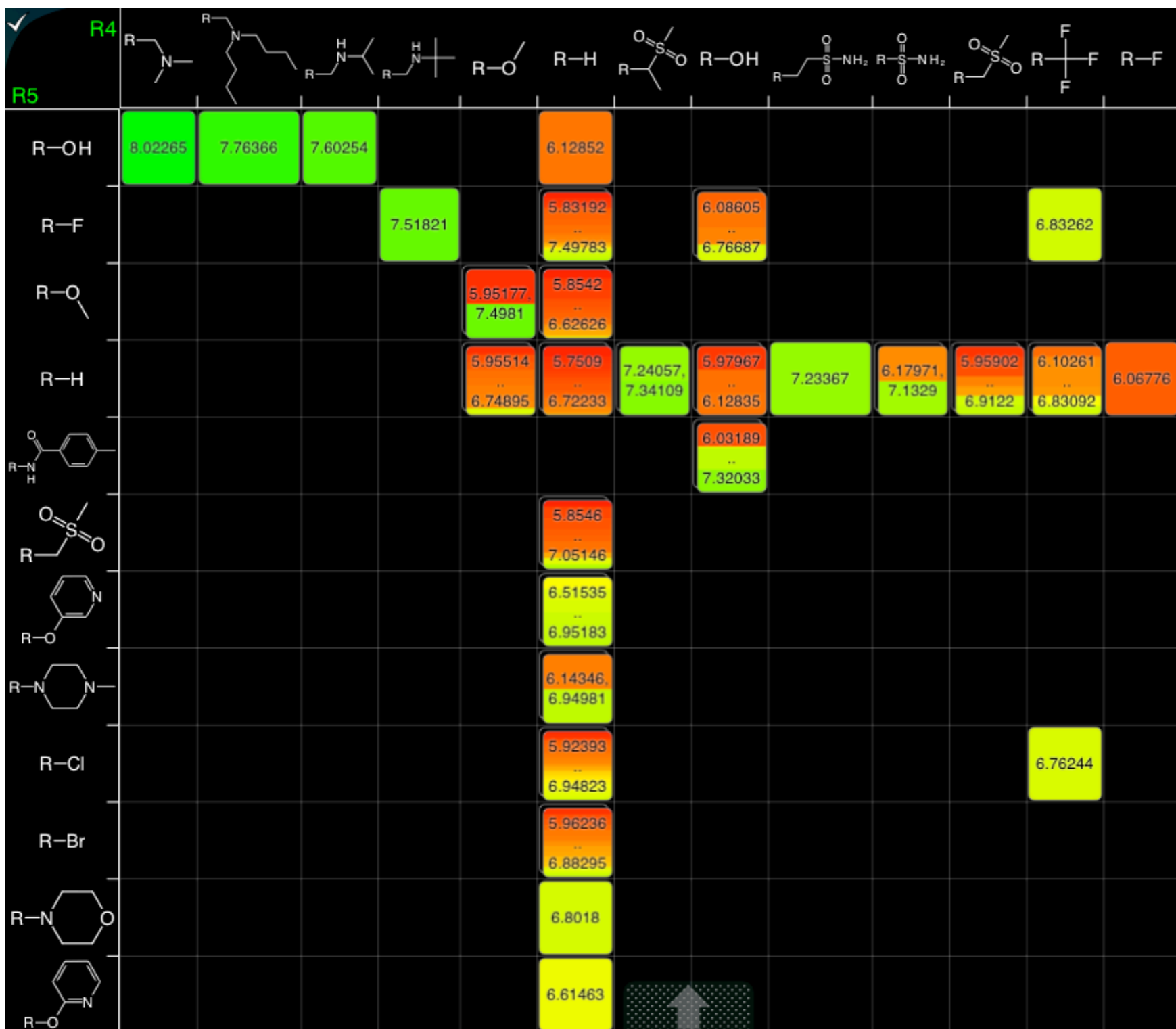
- Have some data, but not all?
 - No problem, build a model...
- Based on a generic method: descriptors based on substructures
- Webservice calls iterate over refinement steps
- Service properties:
 - **complex** labour-intensive calculation
 - **medium** speed (each cycle < 30 seconds)
 - **state** managed by client
 - **small** input, **small** output

Model application



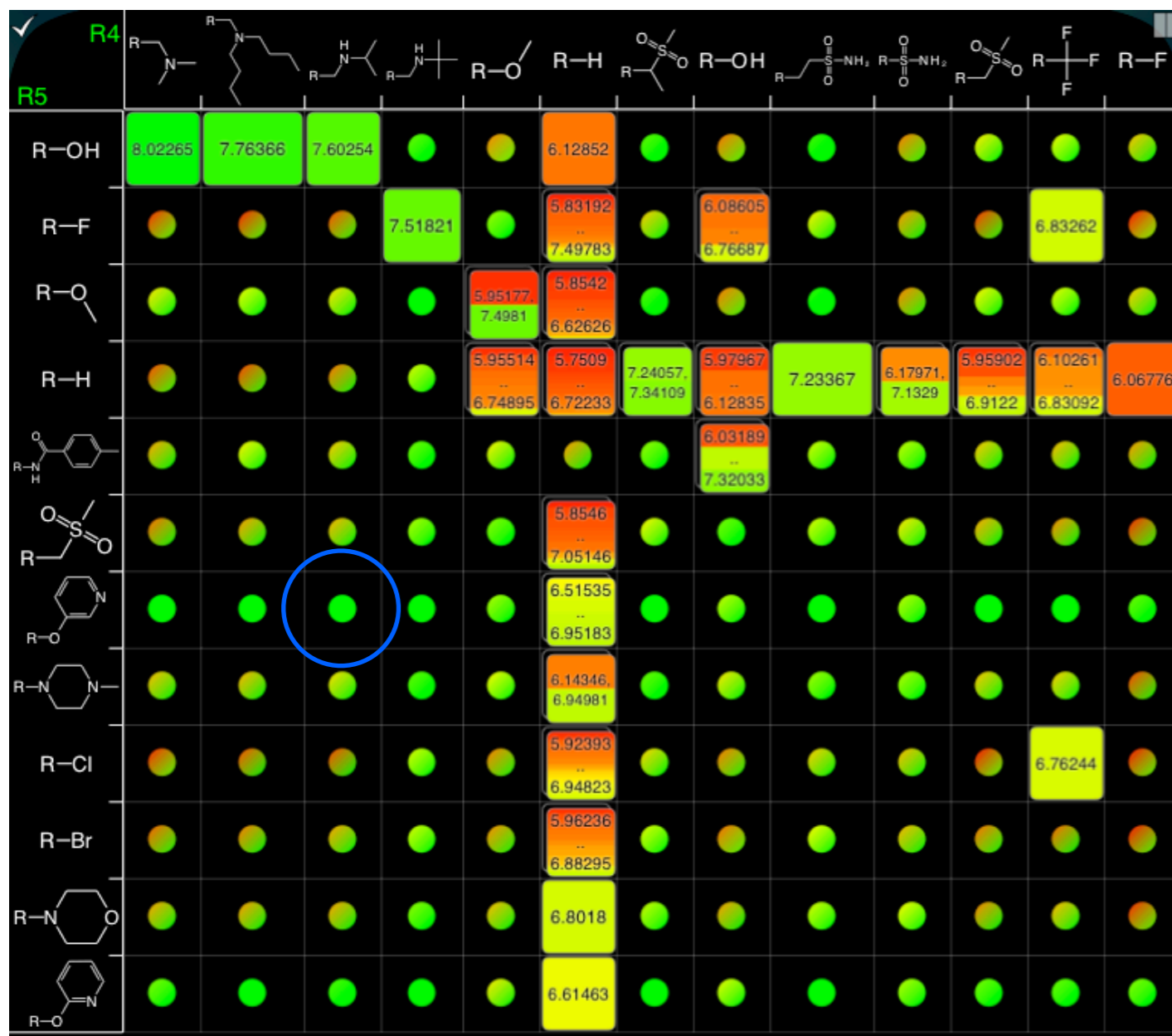
- The model is then applied to compounds with unknown activity
- Compounds with **green** wedges can be
 - prepared/purchased
 - measured
 - real data included
- A measurements are added, predictivity improves

Matrix view



- Select axes, e.g.
 - R4
 - R5
- Distinct fragments plotted on axes
- Activity summarised by colour coding

Empty cells



- Can use the previously built model to estimate activity for empty cells
- The webservice creates hypothetical compounds for each cell:
 - R4 and R5 are implied
 - scaffold, R1, R2 and R3 are selected from currently known cases
- Service properties:
 - **complex** labour-intensive calculation
 - **fast** speed (< 30 seconds)
 - **batch** state
 - **small** input, **small** output

Proposing compounds

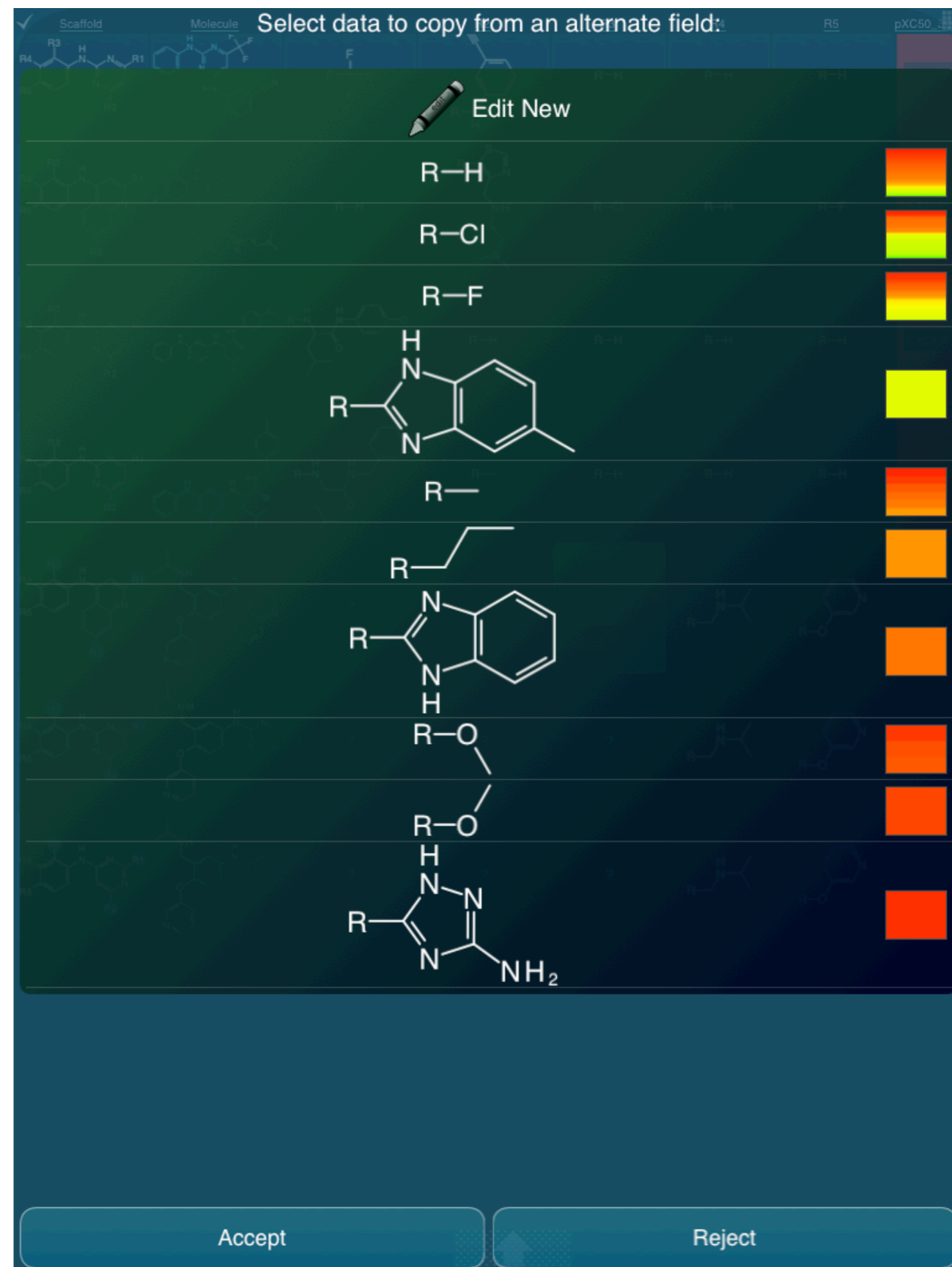
- Any of the cells that look good: tap on one, create a new partial specification...
- ... define the missing variables from the row view.

Scaffold	Molecule	R1	R2	R3	R4	R5
		?	?	?		
		?	?	?		
		?	?	?		

New compounds



- Fill in missing substituents from a list of existing fragments
- Rainbow hues denote activity distribution



Presenting data

- Can output data, bitmaps for structures, and PDFs for the whole matrix view (arts'n'crafts)
- More specialised output is possible:
 - vector graphics (SVG, EPS)
 - Microsoft Word & Excel documents with tables of vector graphics, designed for manuscript preparation
- Tricky formatting done by webservice
 - **complex** calculation
 - **fast** speed (< 1 second)
 - **stateless**
 - **small** input, **small** output

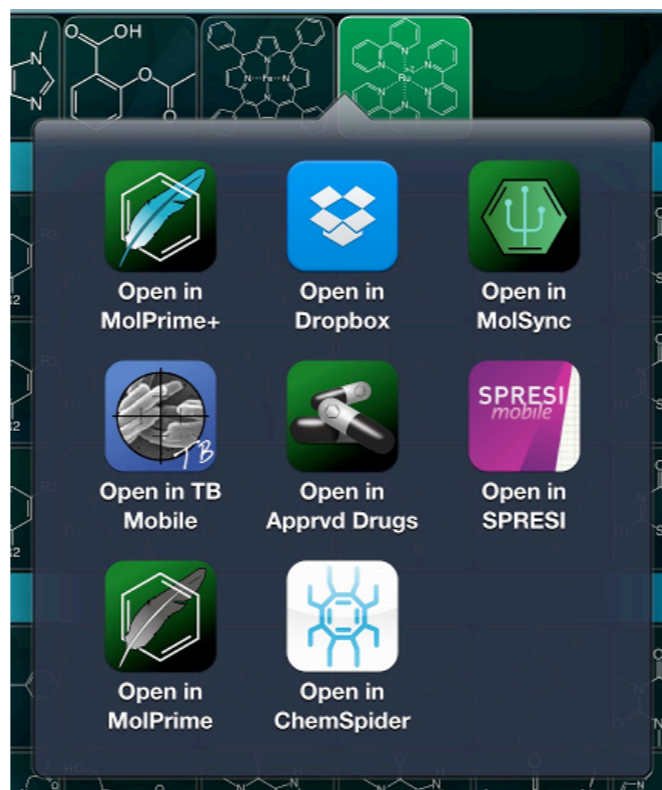
HIV-1 Protease Inhibitors
J. Med. Chem., vol. 54, pp. 7176-7183 (2011)

Scaffold	R1	R2	R3	Id	Molecule	Ki
				29		27 nM
				30		34 nM
				31		350 nM
				32		20 nM

Sharing data

- Can pass the *SAR Table* document, or individual structures, to other apps via:

- open in
- clipboard



- Send by email with attachments
- Store/retrieve on remote filesystems (e.g. *Dropbox* via *MolSync* app)
- Share on the web, by uploading to an open repository
- Tweet the web link, directly from the app

Conclusion

- Major workflow chunks are already possible without using a PC
- Mobile devices are powerful enough to drive a great user experience with complex data
- Cloud-based infrastructure can boost functionality, that is hard or impossible to provide on the device
- **BUT** notice the trend:
 - small **inputs**, small **outputs**, restricted **state**
- Apps are great for working on **small documents** that are easily passed around (app-to-app or via network)

Future work

- The largest remaining problem is **big data**
 - requires complex serverside infrastructure
 - with security
 - and interoperability
- The *Pistoia Alliance* app strategy intends to address this

Prognostication

- Apps will become increasingly prevalent in chemistry
- So will cloud computing

Acknowledgments

- Sean Ekins
- Antony Williams
- Evan Bolton
- The Pistoia Alliance
- Inquiries to
info@molmatinf.com

<http://molmatinf.com>

<http://molsync.com>

<http://cheminf20.org>

@aclarkxyz

