

***Compact models for  
compact devices:  
Visualisation of SAR  
using mobile apps***

**Alex M. Clark, Ph.D.**

August 2015



© 2015 Molecular Materials Informatics, Inc.

<http://molmatinf.com>

# Modern User Experience

- Continuity from 1980s
- Mostly open architecture



- Deploy on anything
- Very open
- Tied to server



- Small touchscreen
- Ultimate mobility

- Unitary functionality
- Company store



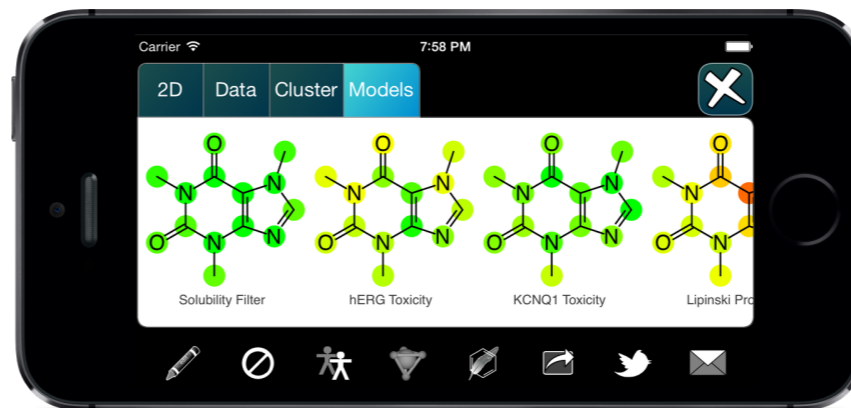
# Mobile Particularly



**Small Data**

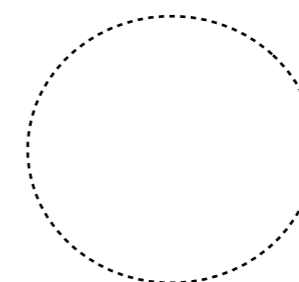
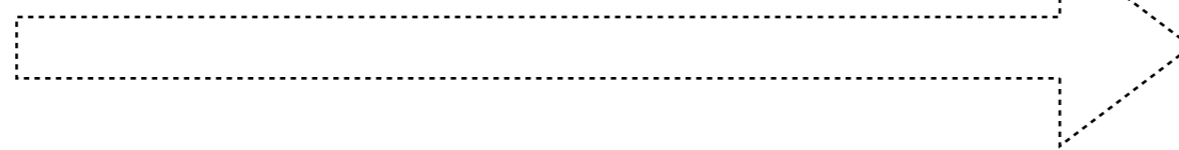


Device Local



- presentation
- visualisation
- calculation
- sharing

**Medium Data**



- custom content
- hard to appify

**Large Data**

Remote API



- searching
- registration
- online sharing
- heavy calculations



# The Remoteness Problem



- Very tempting to use servers for heavy lifting
- Requests via Internet API

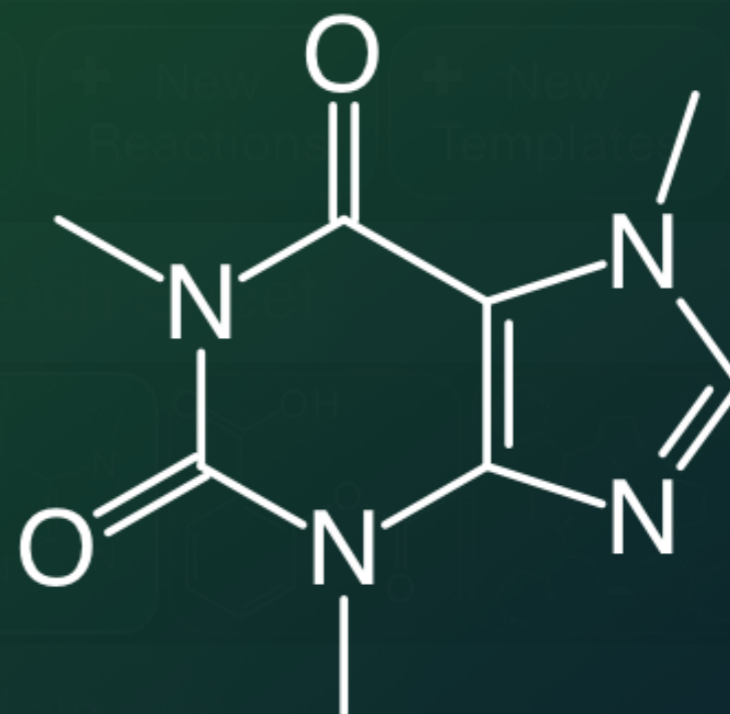
- User identity
- Security issues
- Maintenance burden
- Hostile user experience
  - no signal
  - foreign countries
  - planes
  - wilderness



# The Local Solution

- Structure → Property calculation algorithms
- Start with scalars:
  - some are easy
  - others less so

## Calculated Properties



<b>Molecular Formula:</b>	C <sub>8</sub> H <sub>10</sub> N <sub>4</sub> O <sub>2</sub>
<b>Molecular Weight:</b>	194.1906
<b>Heavy Atoms:</b>	14
<b>H-Acceptors:</b>	6
<b>H-Donors:</b>	0
<b>Rotatable Bonds:</b>	0
<b>Log P:</b>	-1.0293
<b>Molar Refractivity:</b>	51.196
<b>Stereoambiguity:</b>	n/a

# Substructure Filters

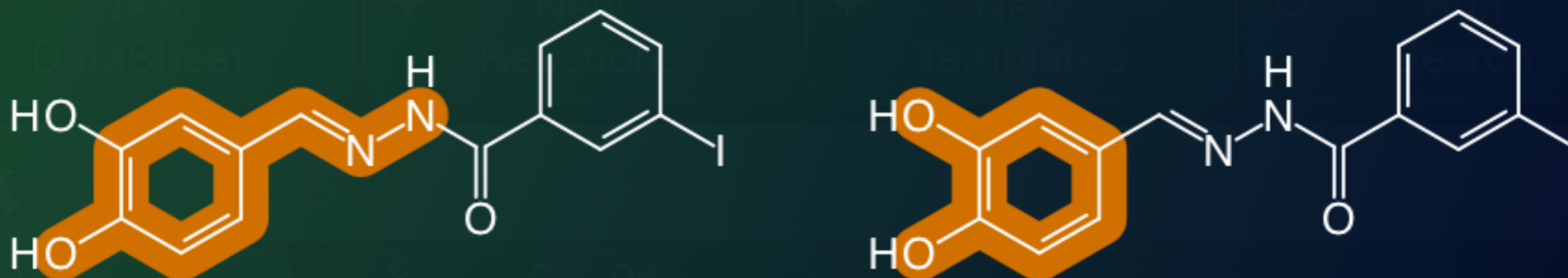
- **PAINS** and **AZ Filters**: both require complex recursive substructure queries
- Overlay visualisation

GClogP, LLE and compound quality calculations

Novice Expert

Formula: C<sub>17</sub>H<sub>14</sub>Cl<sub>4</sub>FN<sub>3</sub>O<sub>4</sub>  
 Weight: 485.12  
 Potency pIC<sub>50</sub>: 6.7 Calculate  
 log P: 3.08  
 Ligand Efficiency: 0.32  
 Lipophilic Efficiency: 3.62  
 Filter 2.10: alkyl halides  
 Filter 2.16: peroxide  
 Filter 7.1: too many halogens

## PAINS Filters

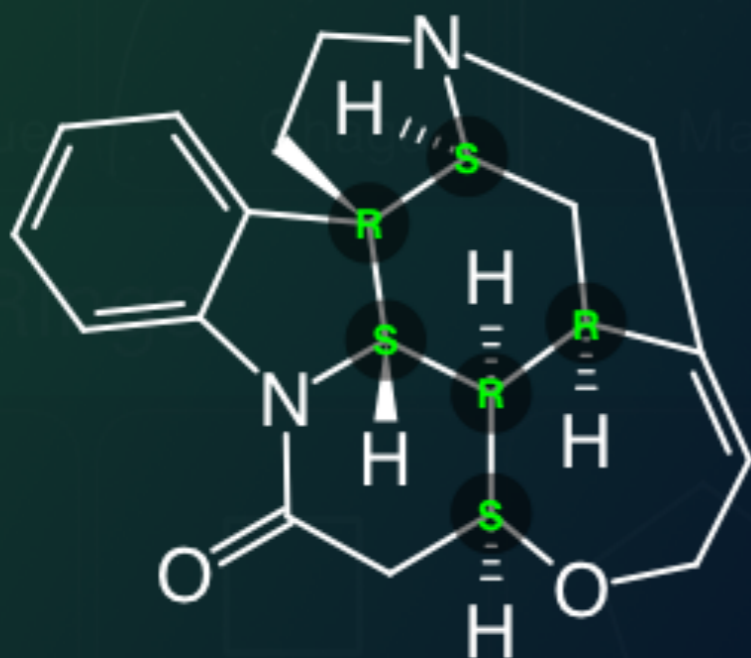




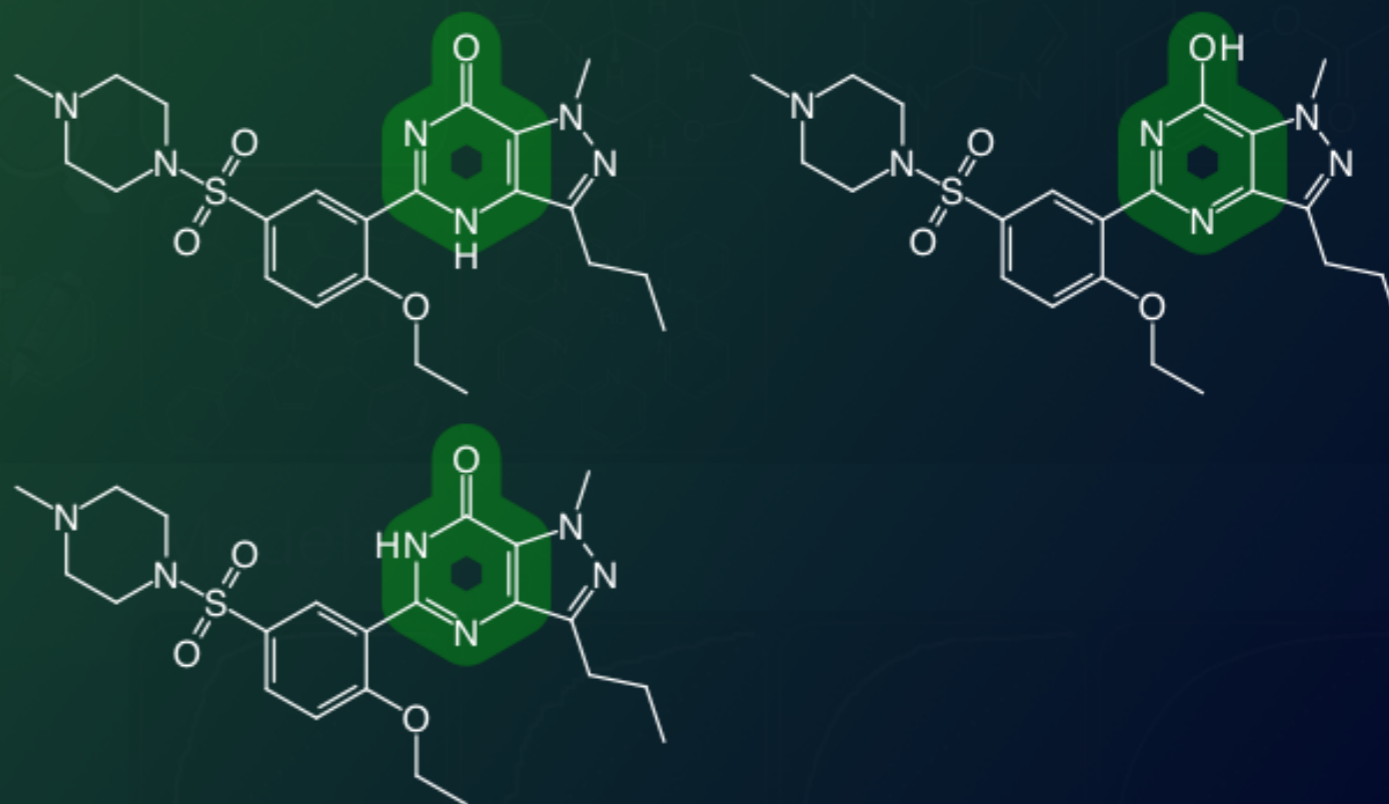
# Molecule Perception

- Hardcoded algorithms for **stereochemistry** and **tautomer enumeration**
- Displayed by structure annotation

## Stereochemistry



## Tautomers



# Prepare & Shrink

Molecular Datasheet - bubonic.ds

Header Aspects Assays Bayesian New View Export Graphics

Model: Model 1: Value [Add] [Delete]

Molecule: Molecule [Value]

Value: Value [Value]

Threshold: < > ≤ ≥ 0.5

Bit Folding: 2048 32768 [None]

Field: ThePlague

Title: Activity against Bubonic Plague

Origin: Assembled by Collaborations in Chemistry

Comment: Data originally from PubChem

[Build] [Reject] [Accept]

SourceURI	ECFP6
28156189,-14608	86007,-13033083
58,-1255426533,-1225008491,-11774	77479836,-11774
39689,-10978690	93,-1093659781,
-2125195270,-19	07119120,-12609
09563,-11081158	55,-618366217,-159315253,7624
56762,13462767	07,1397850344,1432771425,1477
-2066149583,-19	28686588,-18332
38808,-14161320	94,-1186265318,-1122599409,-96
4879417,-716561	880,-697889088,-690611791,-507
-2069831617,-19	28123969,-19088
33380,-17409659	59,-1698628981,-1642473537,-16
10362824,-11488	27238,-10354424
77,-1027418143,	-2114446392,-20
15207159,-19785	83541,-19159697
55,-1668352467,-1456182266,-13	12635493,-11817
60498,-10620925	76,-995448429,-
-2030939070,-19	



Carrier 4:48 PM 100%

Bayesian Model

Title: Activity against Bubonic Plague

Origin: Molecular Materials Informatics: <http://molmatinf.com>

Field: Plague

[Edit]

Comments: Originally built for use with the Mobile Molecular DataSheet mobile app

Details

Fingerprints: ECFP6

Folding: 32768 (15 bits)

Training Size: 139933 compounds

Actives: 223 (0.2%)

ROC Curve

X-Validation: five-fold

Integral: 0.8098

Receiver-operator-characteristic: higher integrals are better. 1 is perfect, 0.5 is random.

[Close] [Save]

- Prepare data using desktop software (and/or of scripts)
- Build **Bayesian** (ECFP6) model: concise, portable
- Transfer model to mobile app...

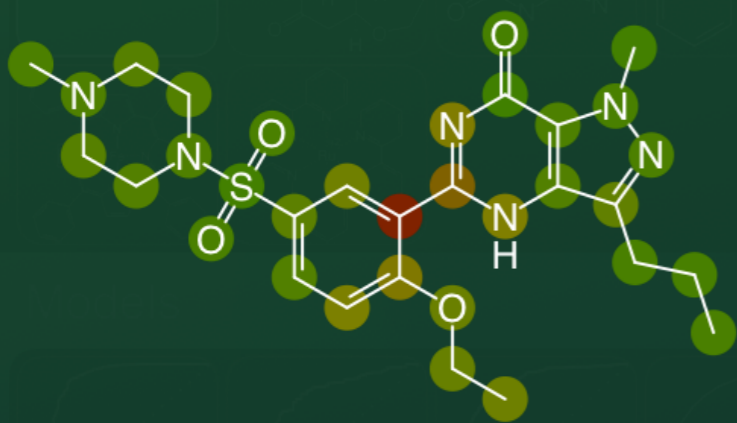


# Predict & Visualise

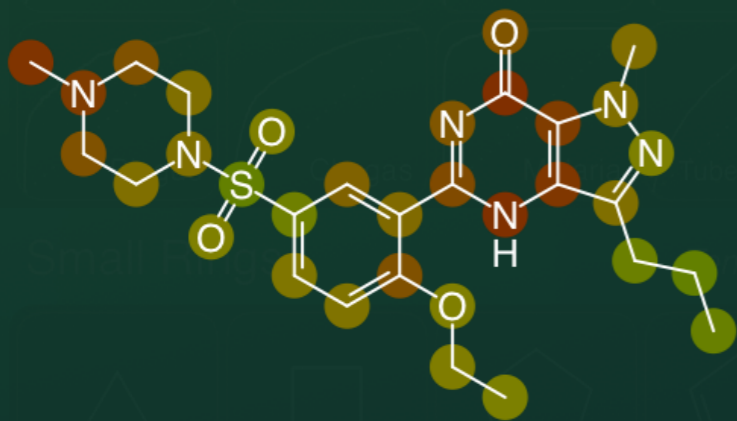
## Bayesian Models

Predicted numbers are scaled to probability-like predictions: most molecules within the domain are in the range 0..1.

**Solubility:** 0.630



**Probelike:** 0.409

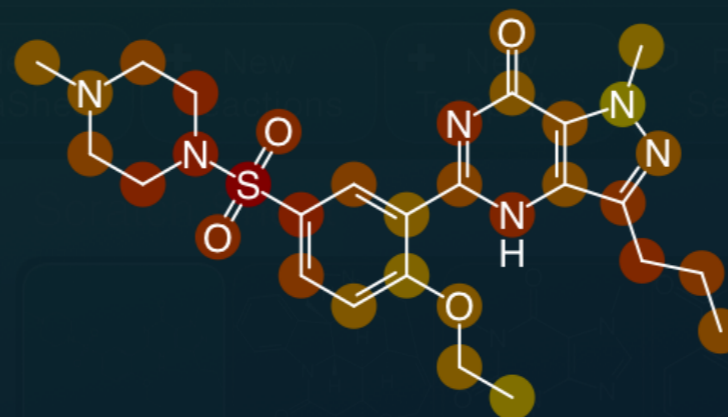


**hERG:** 0.473

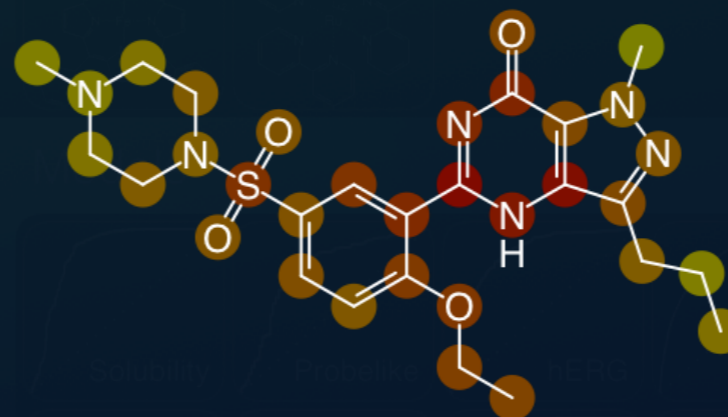


Close

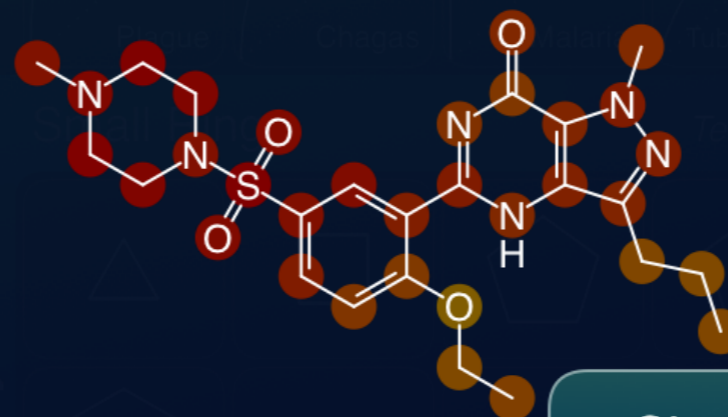
**Plague:** 0.247



**Chagas:** 0.302



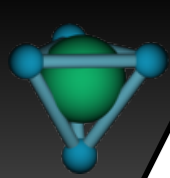
**Malaria:** 0.123



**Tuberculosis:** 0.293

Close

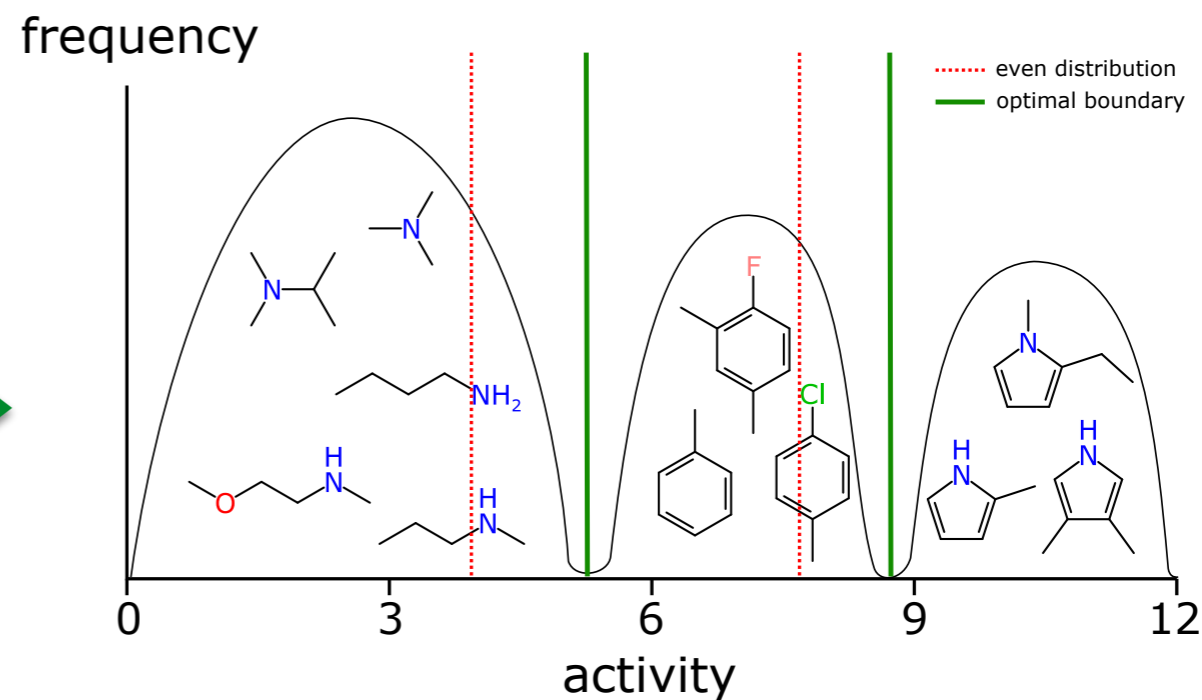
- Fast to apply
- Numeric or molecular visualisation
- Overlay uses ECFP6 fingerprints
- With one molecule or many



# More is Better

ChEMBL	Target
	$\beta$ -Hydroxysteroid dehydrogenase
	1-Acylglycerol-3-phosphate
	Acetylcholine esterase
	ADAM10
	Adenosine A1 receptor
	Adrenergic receptor $\alpha$ 1
	Aldoketo reductase
	Alkaline phosphatase
	Aminopeptidase N
	Angiotensin converting enzyme
	Anthrax lethal factor Bcl2
Arachidonate 12	

~2000



*JCIM* **55**, 1246-1260 (2015)

- Extract target:activity collections from ChEMBL
- Compute active/inactive boundaries
- Automatically build Bayesian models...

# Extreme Appification

- PolyPharma app: work in progress
- Visualise a lot of SAR data... with 2 taps

Carrier 7:02 PM 100%

### PolyPharma

**Profiles**

Tuberculosis Malaria Ebola The Plague

Chagas Disease

**Predictions**

Select a target profile  
Pick a molecule

**Molecules**

Carrier 7:02 PM 100%

### PolyPharma

**Profiles**

Tuberculosis Malaria Ebola The Plague

Chagas Disease

**Predictions**

**Tuberculosis**  
(*Mycobacterium tuberculosis*)

**Solubility**  
(Water)

**P-glycoprotein 1**  
(*Homo sapiens*)

**HERG**  
(*Homo sapiens*)

**Cytochrome P450 1A2**  
(*Homo sapiens*)

**Tuberculosis** *Mycobacterium tuberculosis* 0.744

**Solubility** Water 0.242

**Cytochrome P450 1A2** Homo... 0.247

**HERG** *Homo sapiens* 0.432

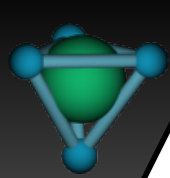
**P-glycoprotein 1** *Homo sapiens* 0.294

Carrier 0.432 0.294 7:04 PM 100%

**Tuberculosis** (*Mycobacterium tuberculosis*)

**Solubility** (Water)

**Cytochrome P450 1A2** (*Homo sapiens*)



# Demo

- **PolyPharma** app demonstrated live...
- Since the meeting, the app has become available on the iTunes AppStore:

**<http://itunes.apple.com/app/polypharma/id1025327772>**

- For more information, see:

**<http://molmatinf.com/polypharma.html>**

# Acknowledgments

- Sean Ekins, Antony Williams, Andy Davis
- Royal Society of Chemistry
- Collaborative Drug Discovery
- Inquiries to **info@molmatinf.com**

**MOLECULAR  
MATERIALS  
INFORMATICS**

<http://molmatinf.com>

<http://molsync.com>

<http://cheminf20.org>

@aclarkxyz

