Putting together the pieces:

Building a reaction-centric electronic lab notebook for mobile devices

Alex M. Clark, Ph.D.

November 2013

http://molmatinf.com

Introduction

mainframes
minicomputers

personal computers
portable laptops

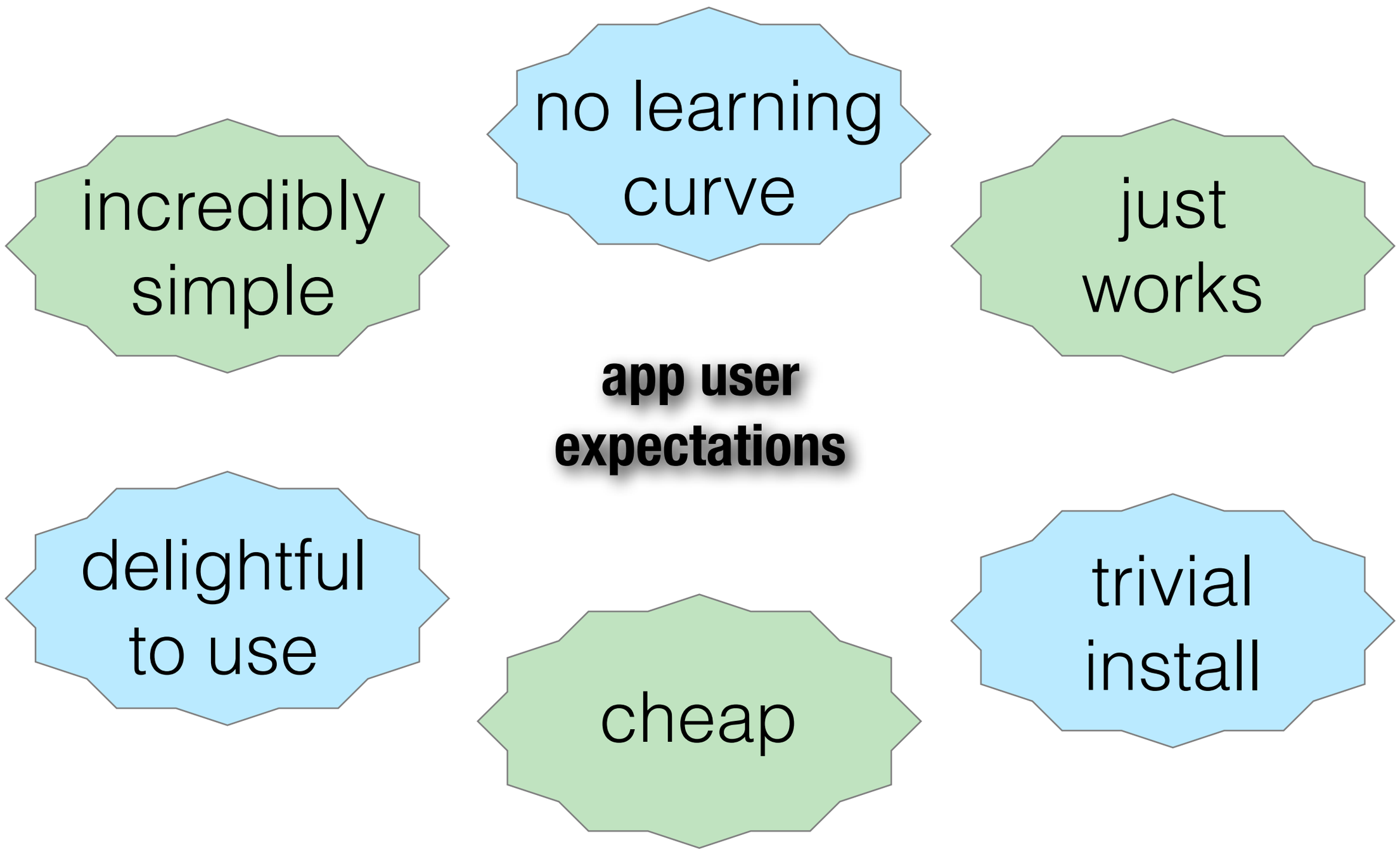
mobile tablets
smartphones

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

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

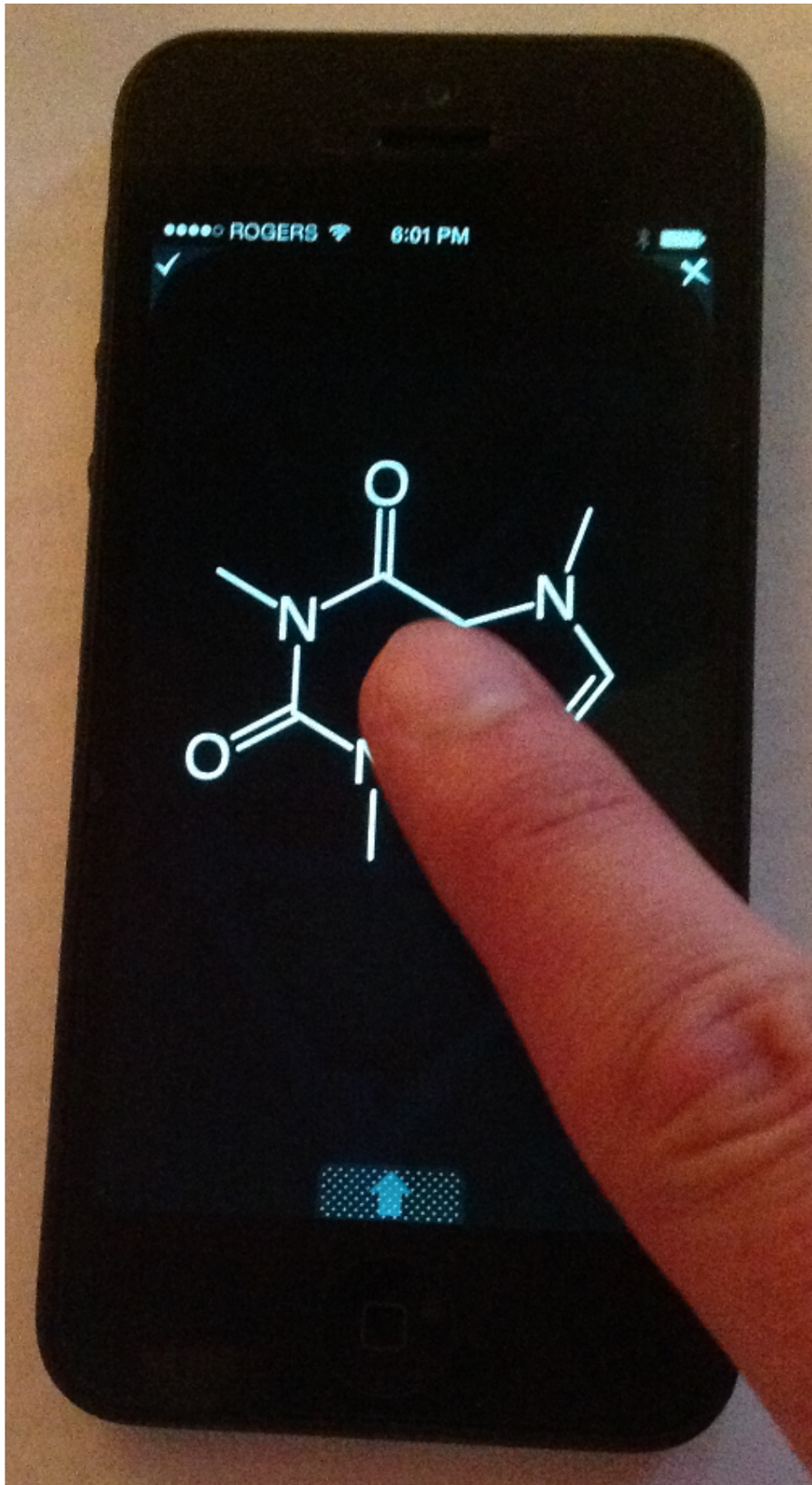
Disruption

- Entire software industry ➠ mobile, like it or not



Technical Challenges

- Tiny screens (palm-sized)
- Fingers instead of mouse
- Reduced computing power
- Limited storage
- Delegate tasks to cloud servers: split between UI & API
- Multiple platforms



Chemistry App Ecosystem

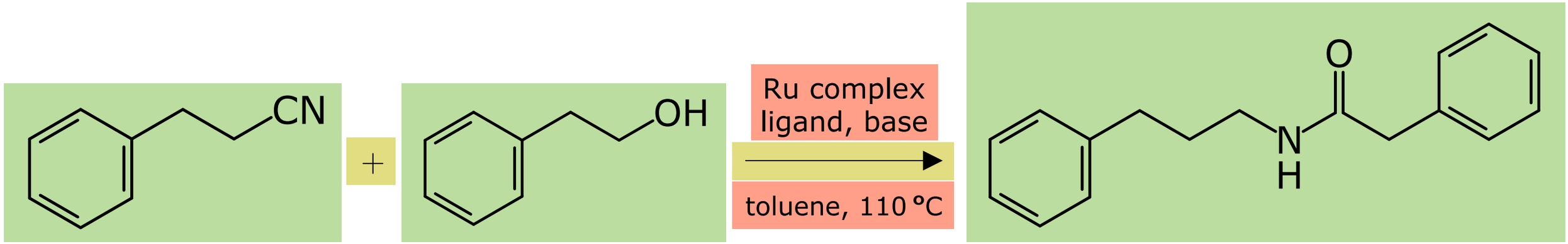
- Reference data
- Education
- Structure drawing
- Database searching
- 3D viewing
- Reactions & collections
- Property calculation
- Model building
- Graphical presentation
- Data sharing

Reaction Lab Notebook?

- **Electronic Lab Notebooks**: huge business
- Some products targeting *iPad* or mobile web
- Reaction-centric lab book app:
  - should understand chemistry
  - easy access to reference information
  - incorporate green chemistry metrics
  - make it easy to share or centralise data
  - appeal to academia & industry

Reaction Canvas

- Reactants and products: atoms & bonds (mostly)

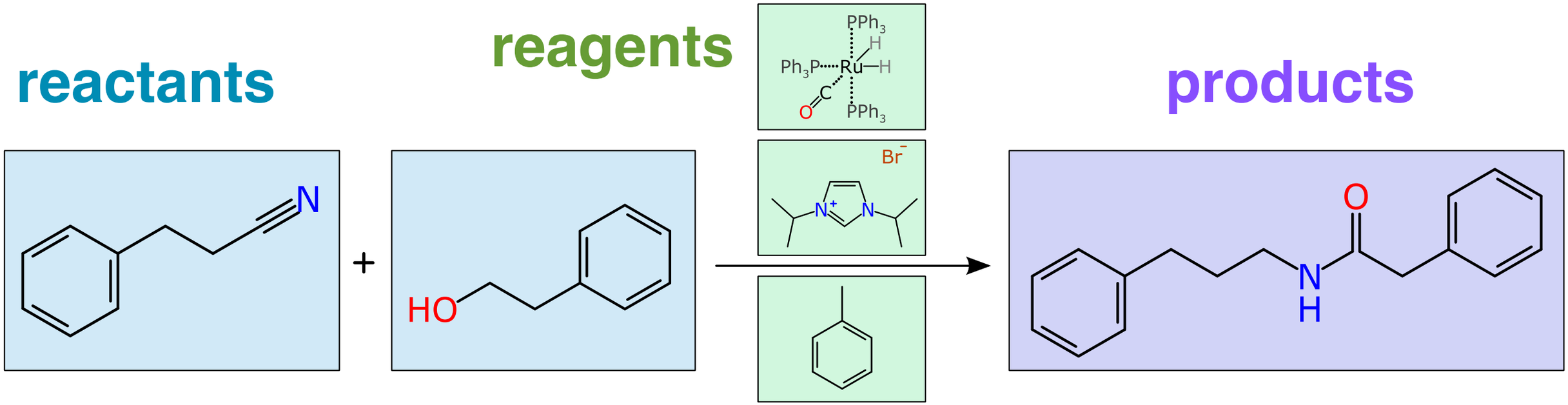


- Component separation: layout dependent
- Reagents, solvents, conditions: essentially meaningless

Hong et al, *JACS*, **135**, 11704 (2013)

Reaction Representation

- Markup as component structures:

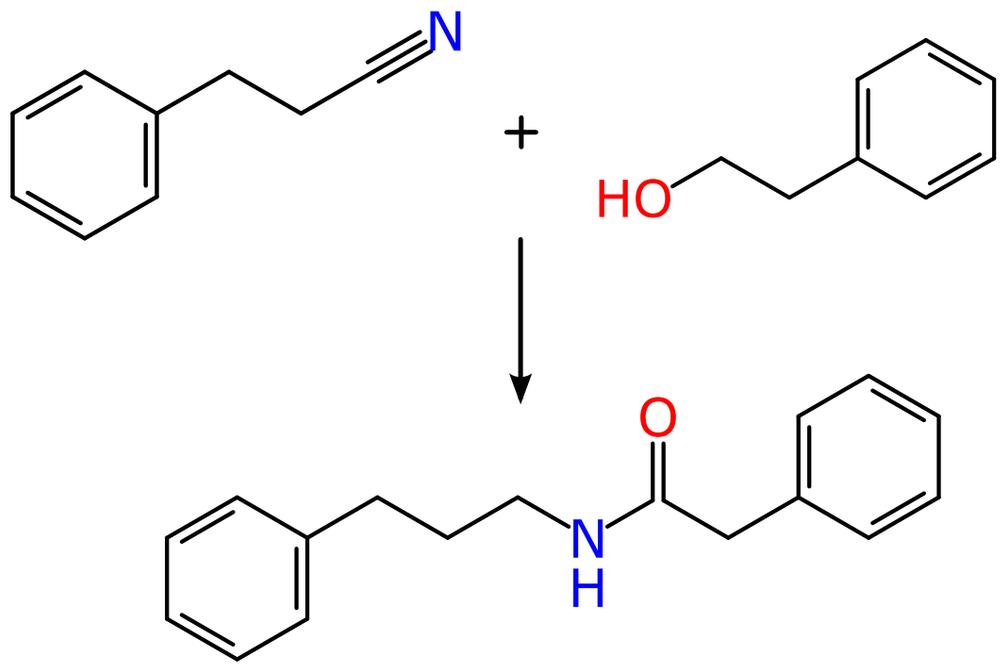


- Properties can be associated with each component, e.g. **name**, **stoichiometry**

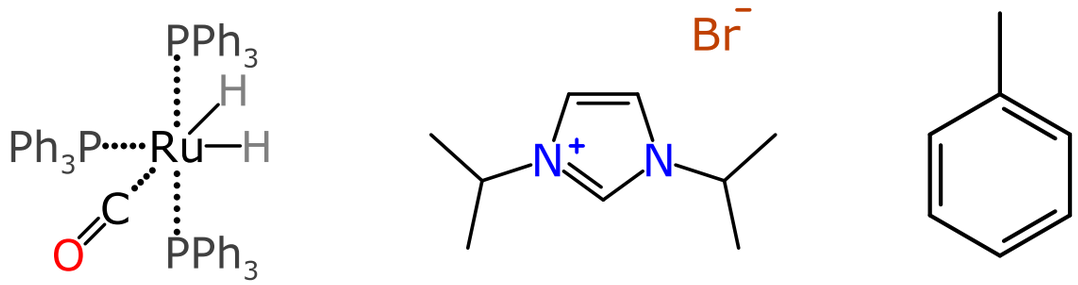
- Draw each structure separately

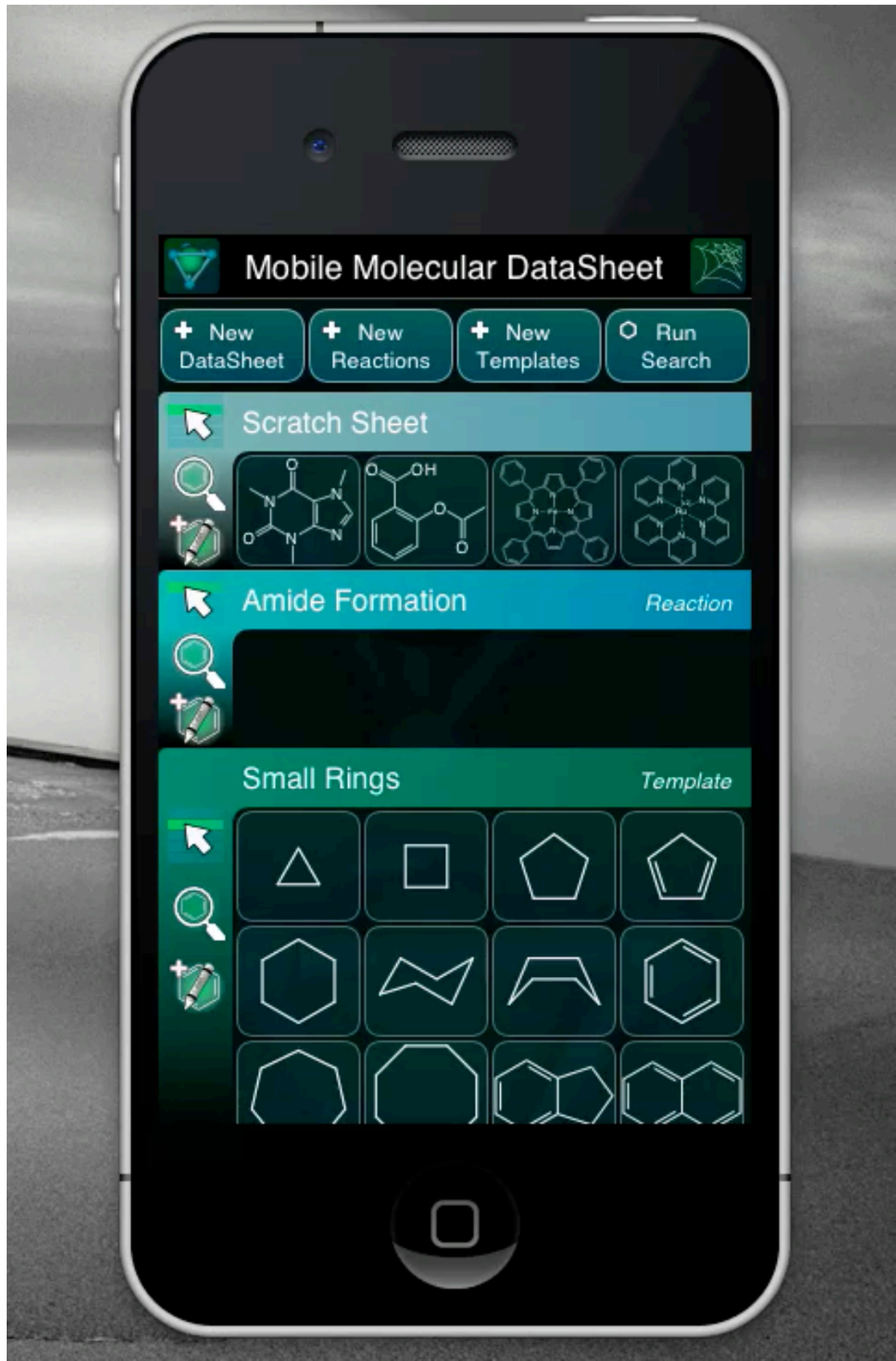
Sketcher

- Drawing reactants & products on an iPhone is quick



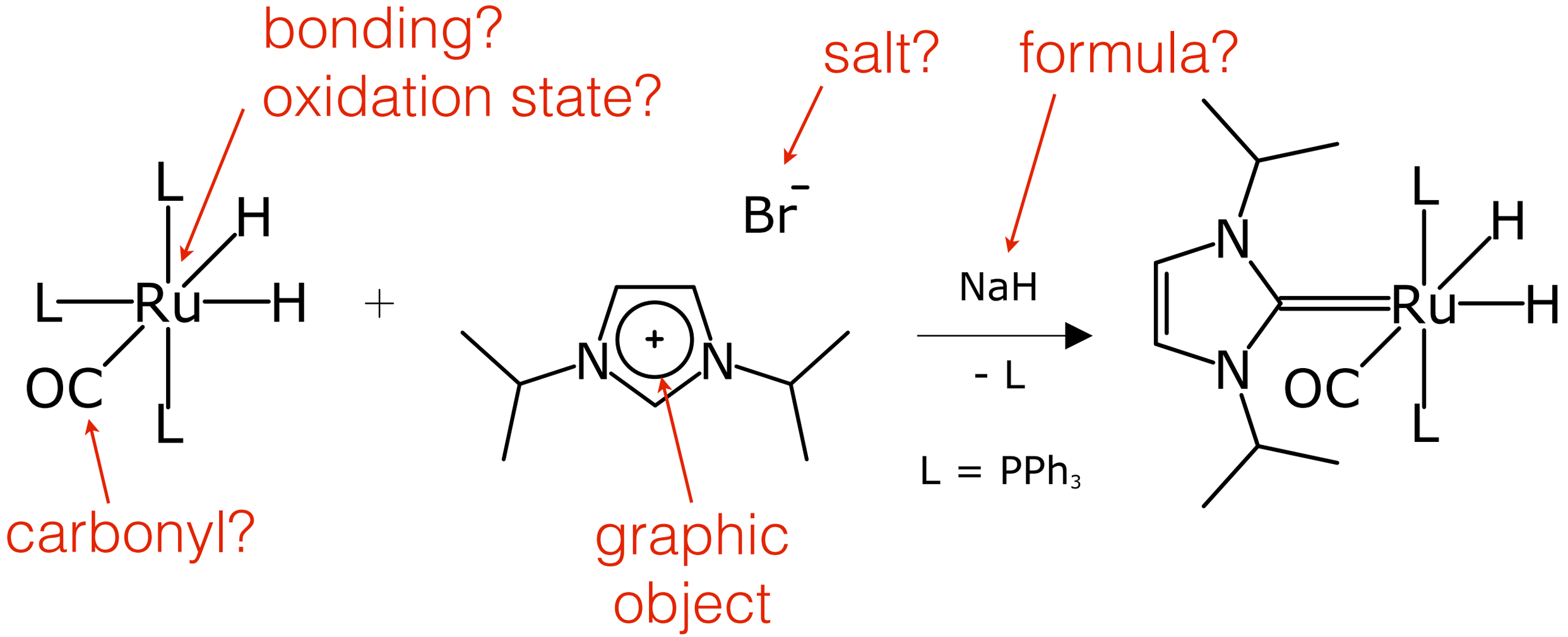
- Reagents & solvents added subsequently





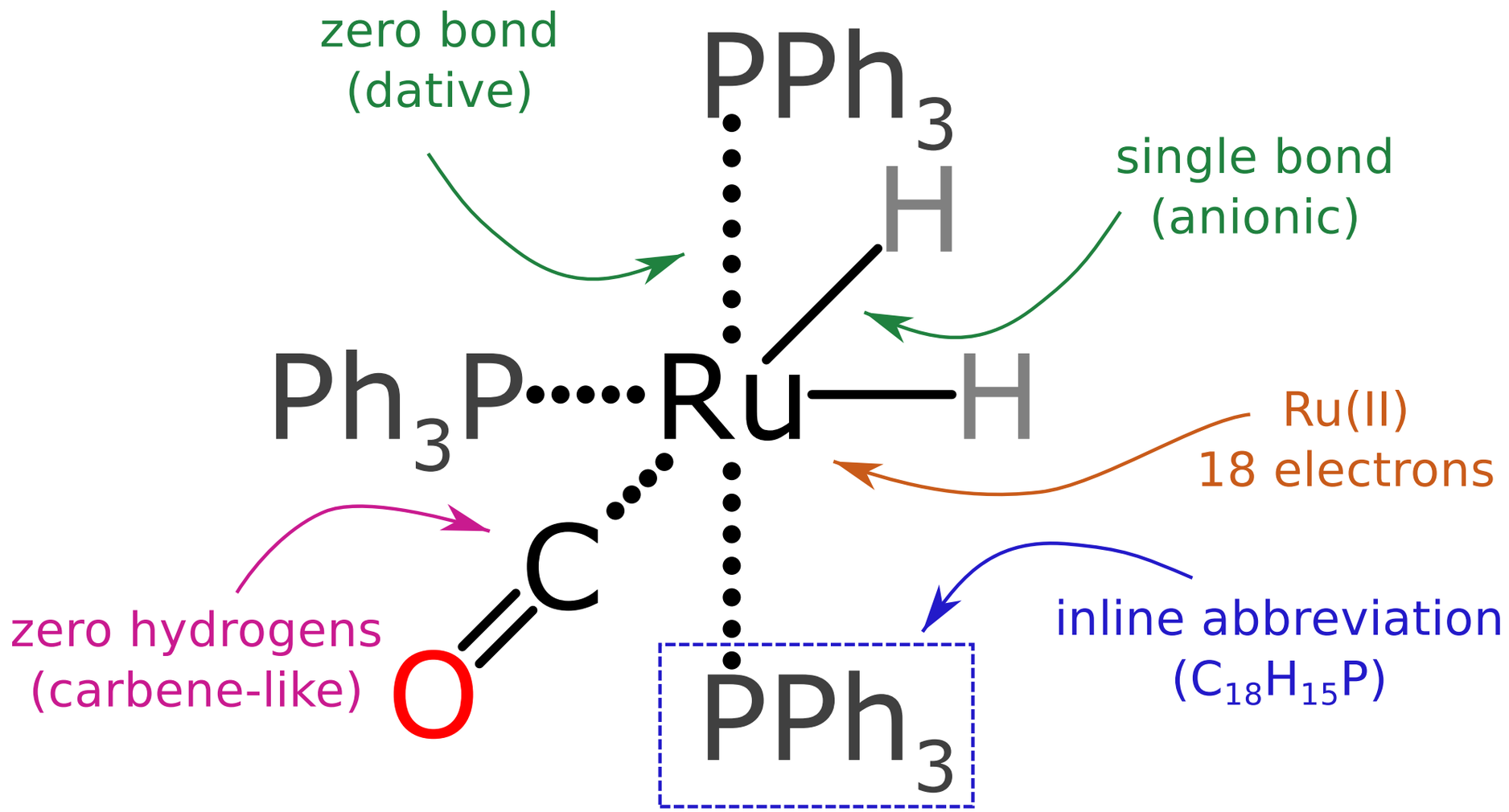
Organometallics

- Global & local style conventions anathema to machine-readability...



Organometallic Anatomy

- Commonly used connection table formats are unable to provide aesthetics *or* accuracy



SketchEl Format

```
SketchEl!(8,7)
Ru=0.0143,-1.3143;0,0,i0
H=2.2104,-1.3143;0,0,i0
H=1.5672,0.2386;0,0,i0
P|Ph{3}=0.0143,-3.9763;0,0,i0,
    ➥aSketchEl!(20\002C22)\000A*
P|Ph{3}=0.0143,1.3477;0,0,i0,
    ➥aSketchEl!(20\002C22)\000A*
P|Ph{3}=-1.9857,-1.3143;0,0,i0,
    ➥aSketchEl!(20\002C22)\000A*
C=-1.3975,-2.7260;0,0,e0
O=-2.4581,-3.7867;0,0,i0
1-2=1,0
1-3=1,0
1-4=0,0
1-5=0,0
1-6=0,0
1-7=0,0
7-8=2,0
!End
```

$C_{55}H_{47}OP_3Ru$
917.9517 g/mol

- Used by open source *SketchEl* & mobile apps
- Core principles:
  - extreme minimalism
  - extensibility (forward)
  - zero-order bonds
  - virtual hydrogen control
  - inline abbreviations
- Lowest common denominator:

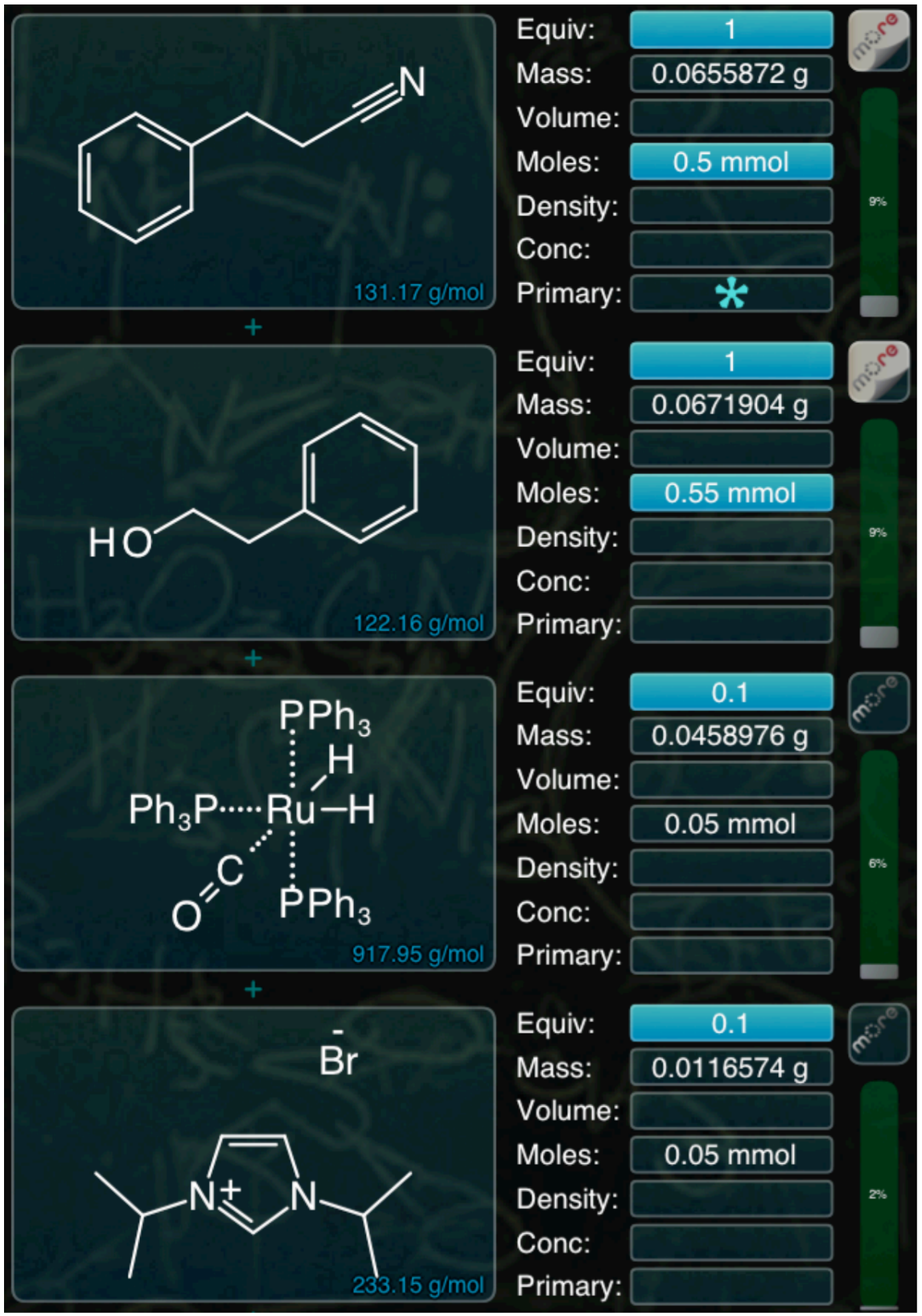
  SketchEl ⇋ MDL Molfile

J. Chem. Inf. Model., **52**, 3149 (2011)

Yield101

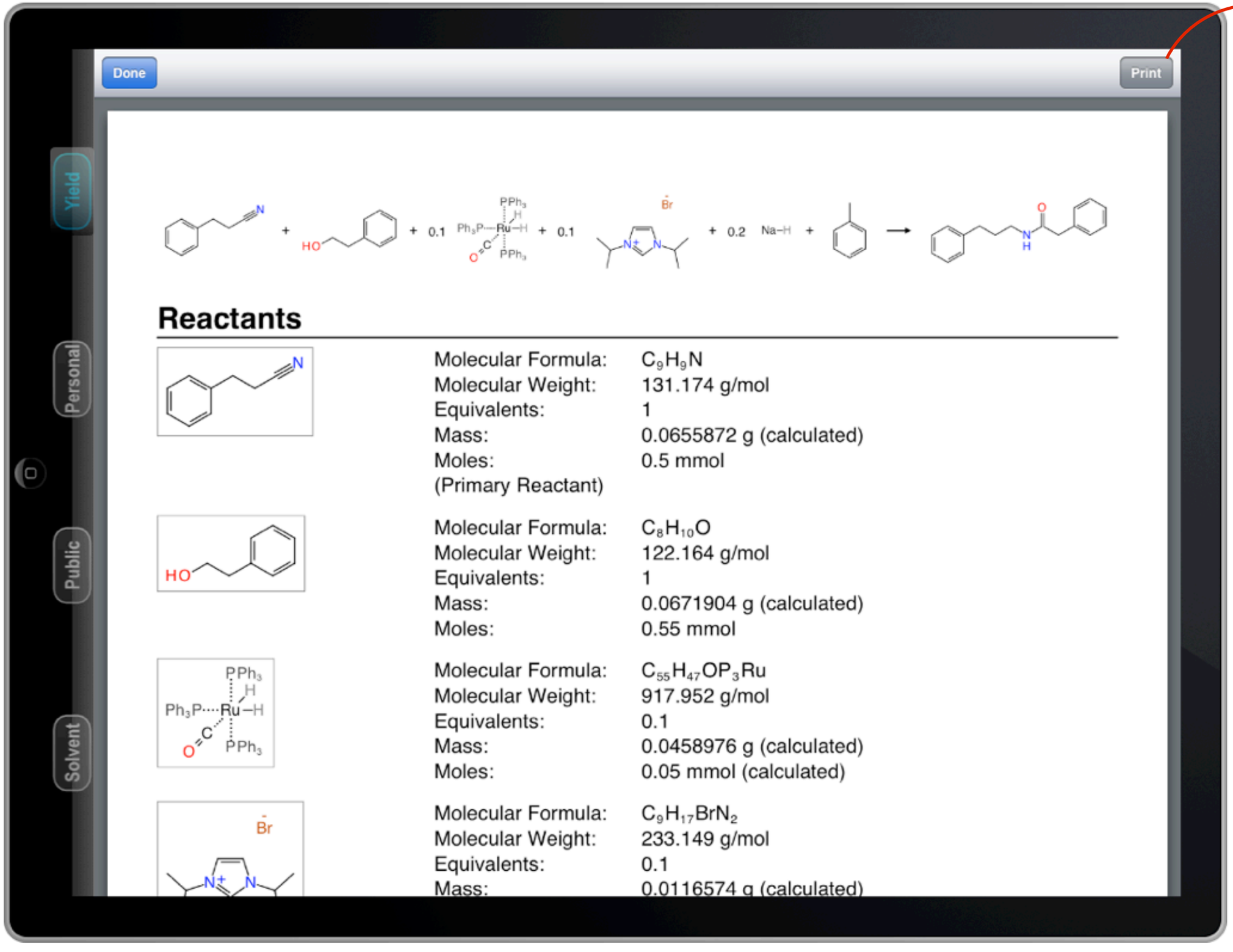
- Prototype lab notebook already exists on the market
- Designed for undergraduates, more general use

Quantities



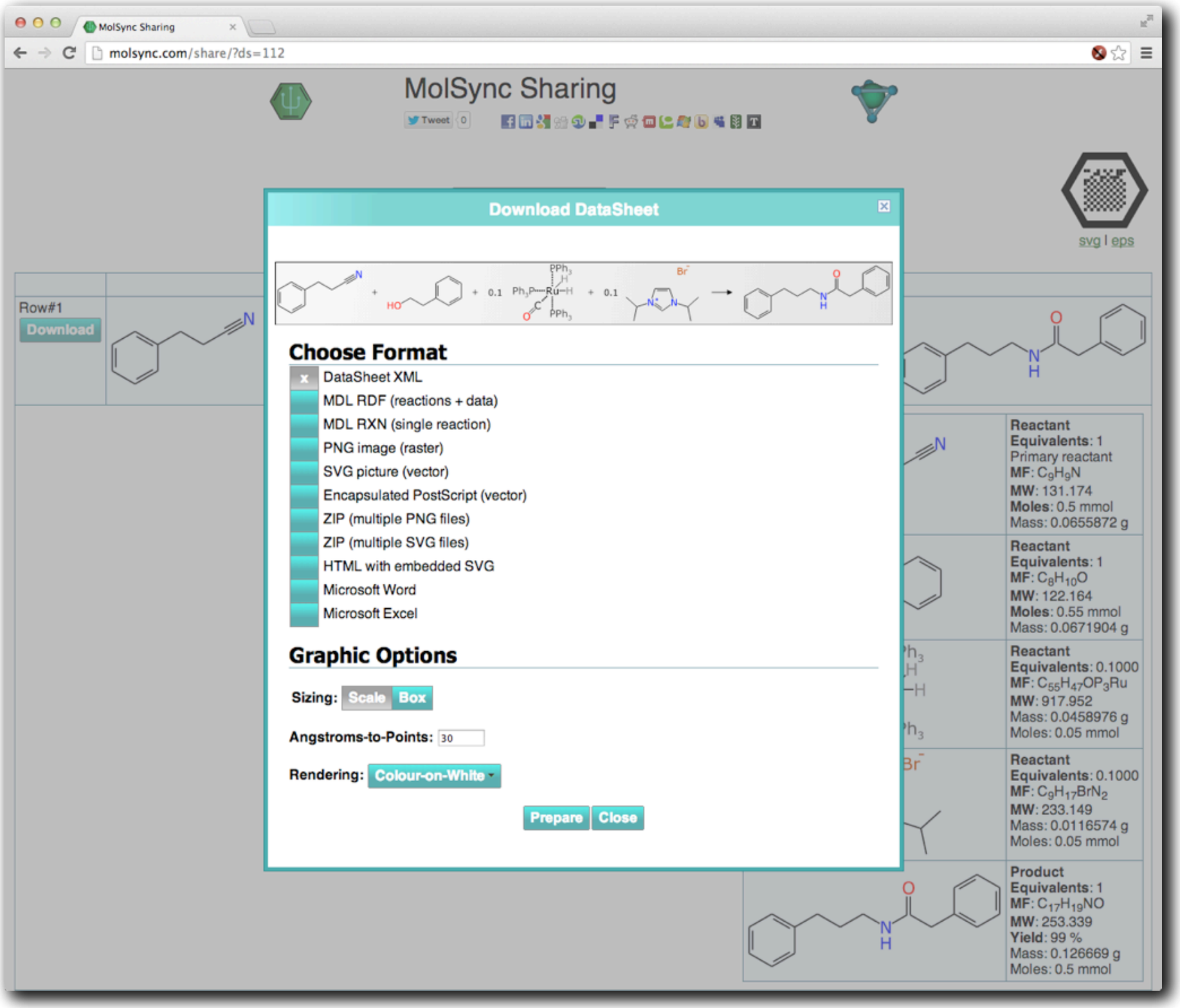
- Adds *quantities* to the Reaction aspect: auto-interconverted
- Green chemistry:
  - solvent reference
  - process mass intensity calculation
- Automatic lookup of structures, cross reference to *Mobile Reagents*
- Graphics creation
- Private & public sharing of data

Graphics Creation



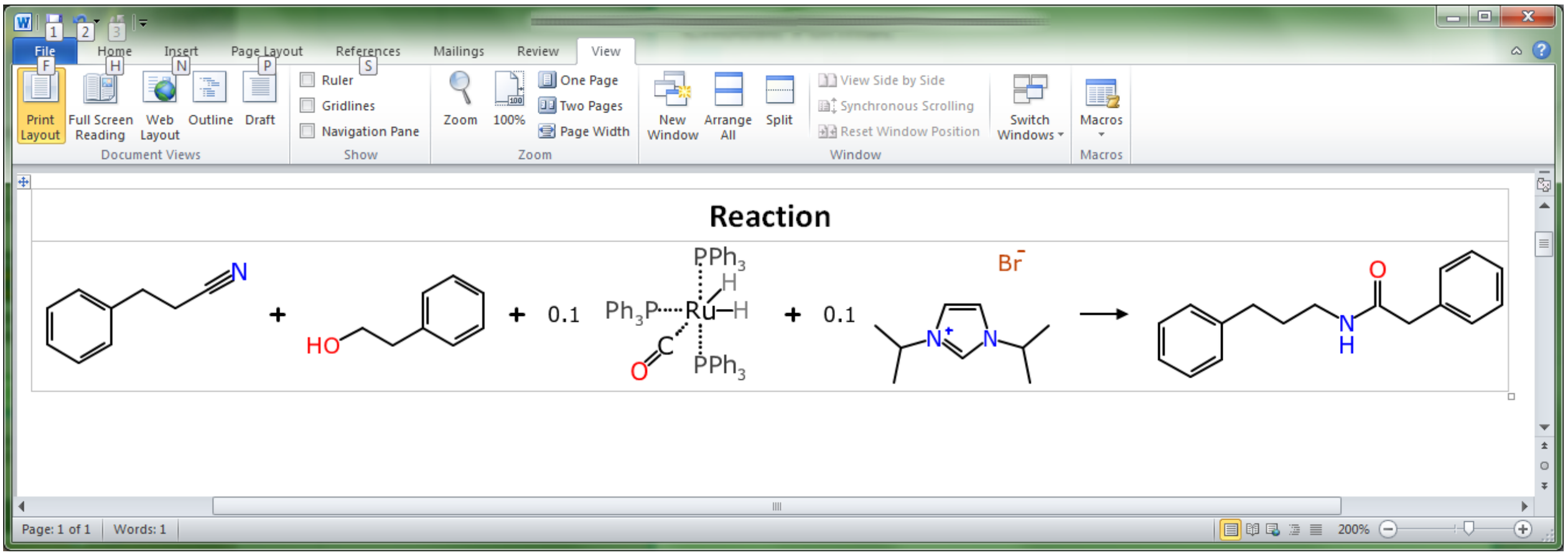
hardcopy:
print or
export
PDF

Public Sharing



- Scheme uploaded
- Public
- Persistent
- URL access
- Formats
- Graphics

Manuscripts



- Export MS Word/Excel documents: DrawingML
- Encapsulated PostScript (EPS)
- Scalable Vector Graphics (SVG)

Reaction Lab Notebook

Core Technology	Enhancements
• structure drawing • reaction drawing • quantity calculation • green metrics • graphics • public databases • data sharing	• professional chemists • calculation services • reference data • green chemistry • usability • data centralisation • sustainable business

The Green Lab Notebook (GLN)

Transforms

Tutorial

Feedstocks

Solvents

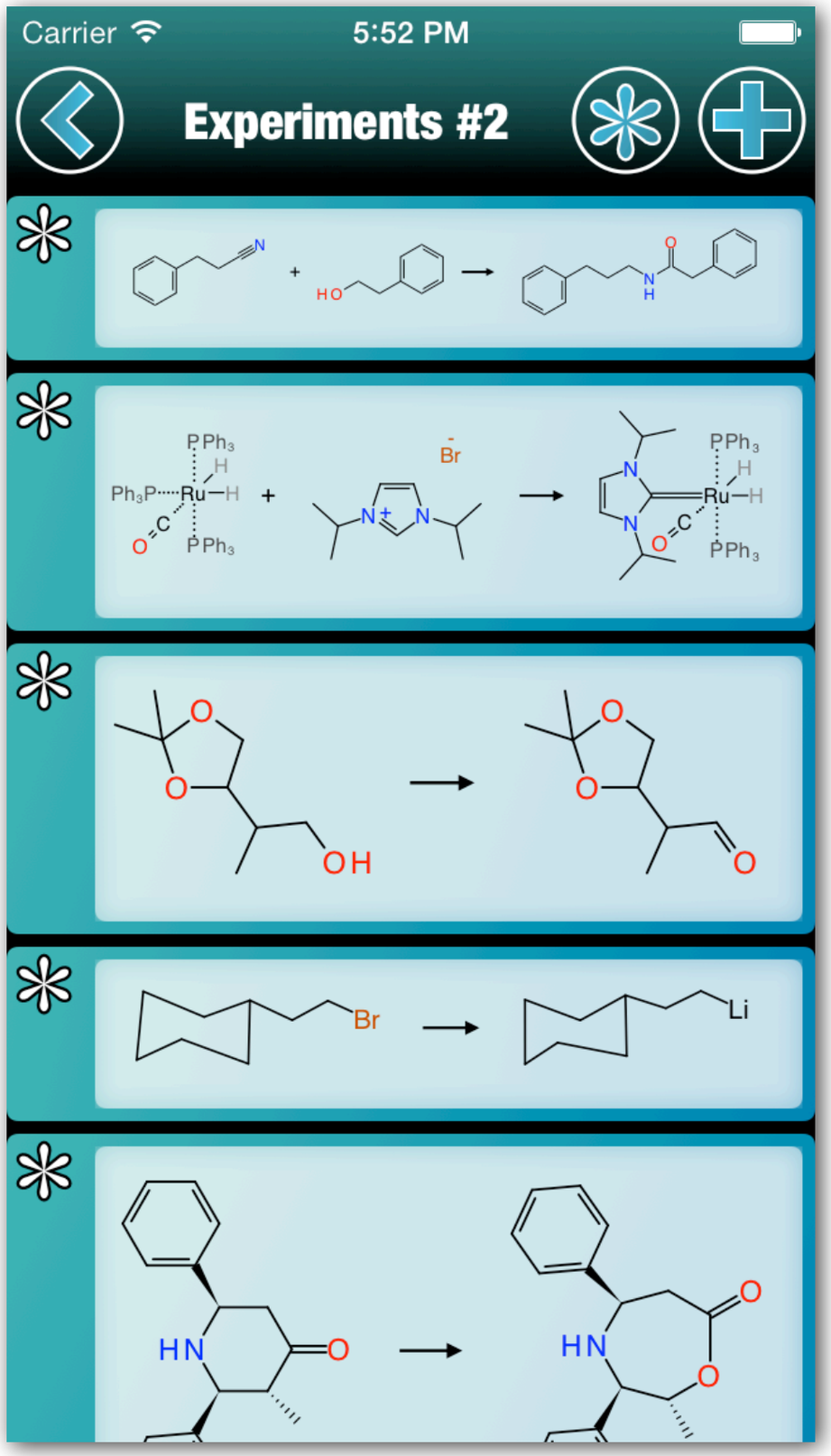
Green Chemistry Principles

Experiment Folders

Scratch Sheet

Templates

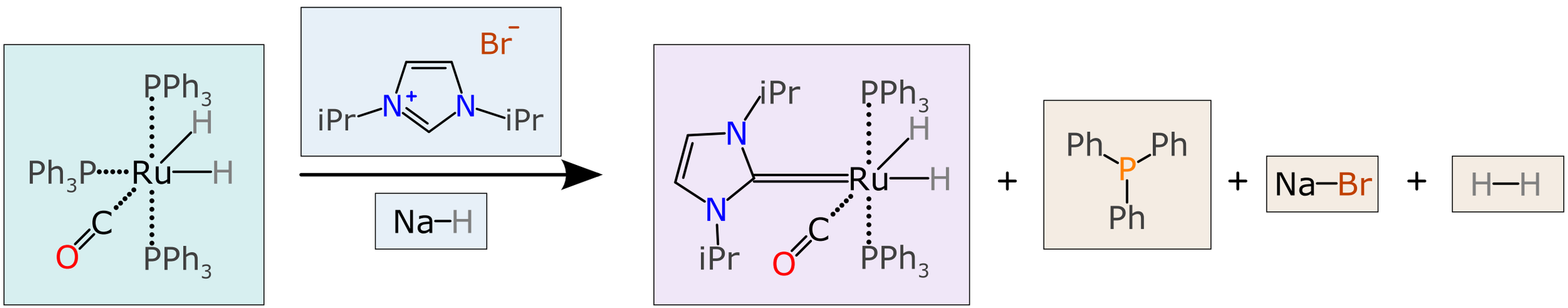
Folders of Experiments



- Groups of experiments
- Stored locally on device: network optional
- Synchronised to central location
- Expand on the basic reaction definition...

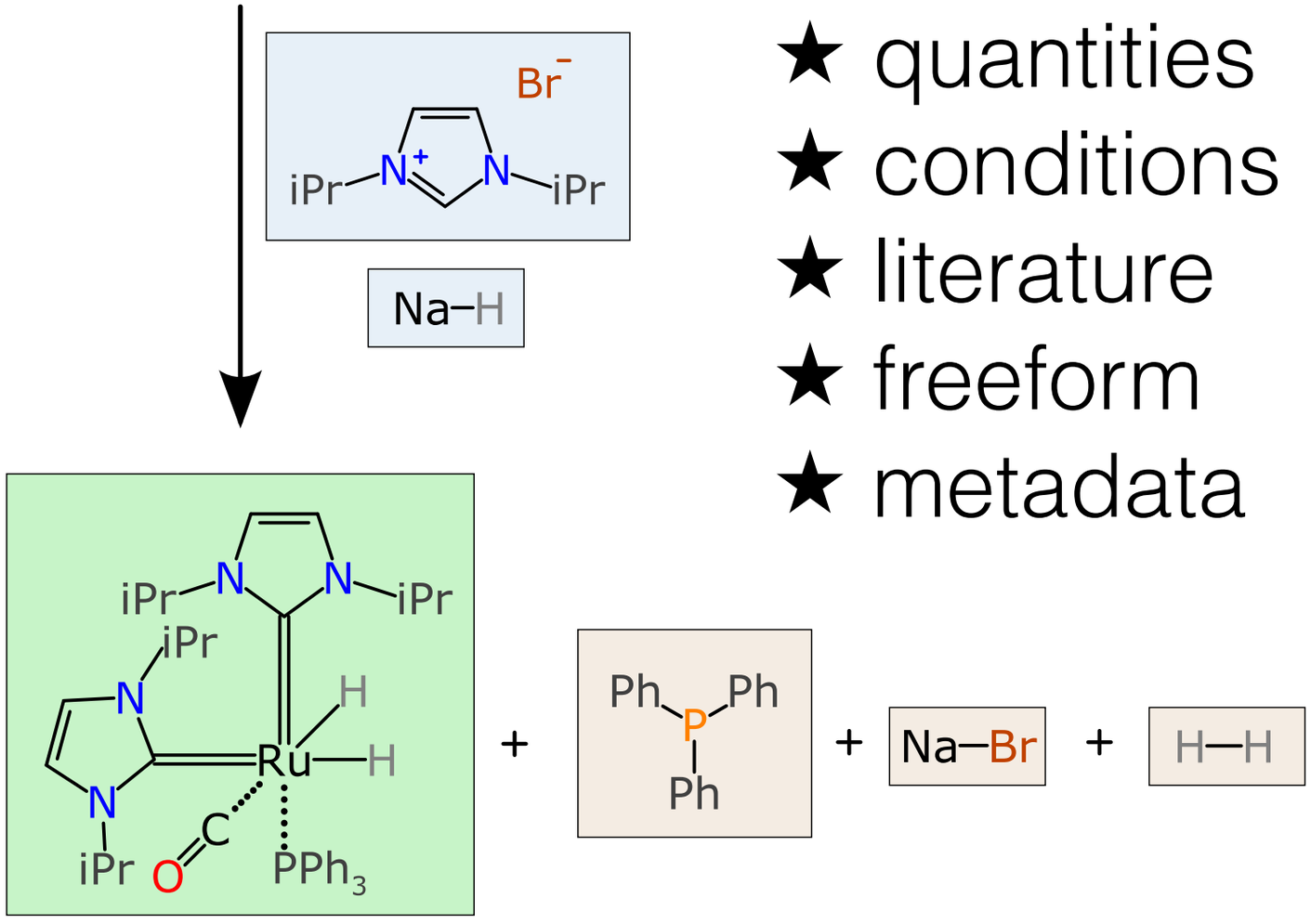
Reaction Experiments

- Multistep reaction editor, more supporting fields



Multistep:

★ reactant
★ reagent
★ intermediate
★ product
★ waste



★ quantities
★ conditions
★ literature
★ freeform
★ metadata

Green Metrics

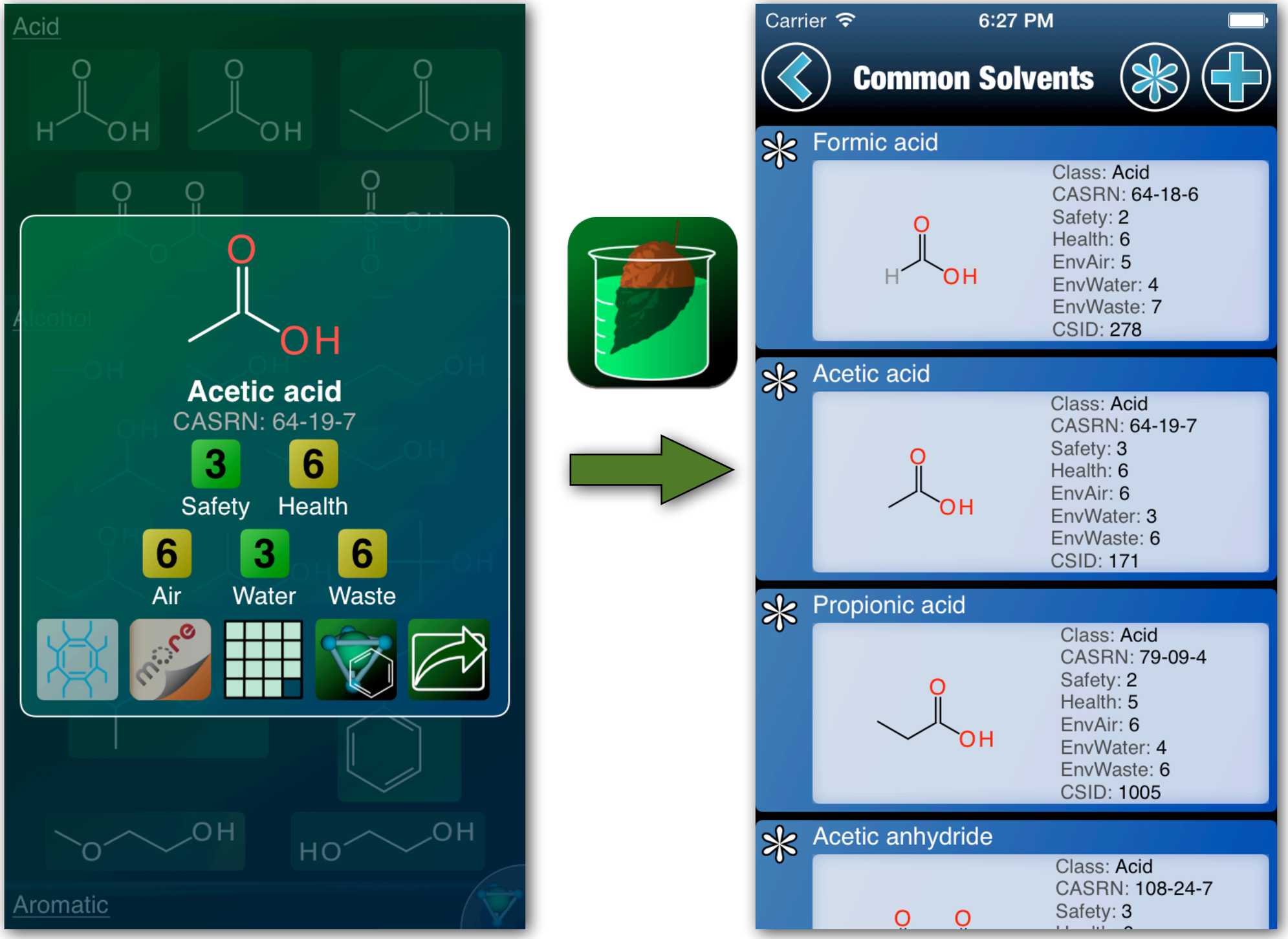
$$\text{Process Mass Intensity (}\mathbf{PMI}\text{)} = \frac{\text{mass of all reactants}}{\text{mass of products}}$$

$$\mathbf{E\text{-}factor} = \frac{\text{mass of waste}}{\text{mass of products}}$$

$$\mathbf{Atom\ Economy} = \frac{\Sigma\ \text{molecular weight reactants}}{\Sigma\ \text{molecular weight products}}$$

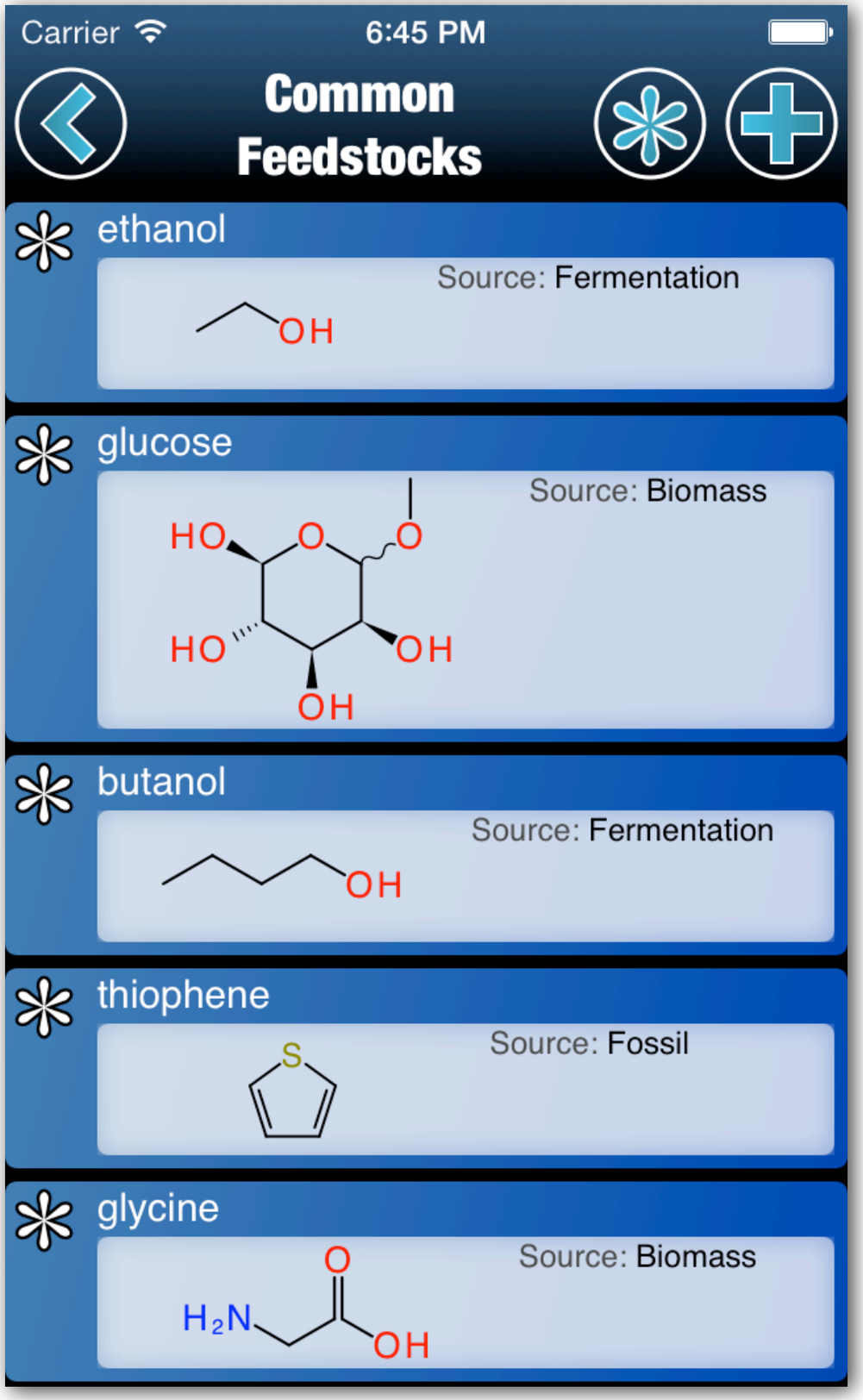
- Databases: e.g. Toxics Release Inventory (**TRI**), Persistent Bioaccumulative Toxic Chemicals (**PBT**)
- Curate as structure-searchable database, automatic recall

Green Solvents



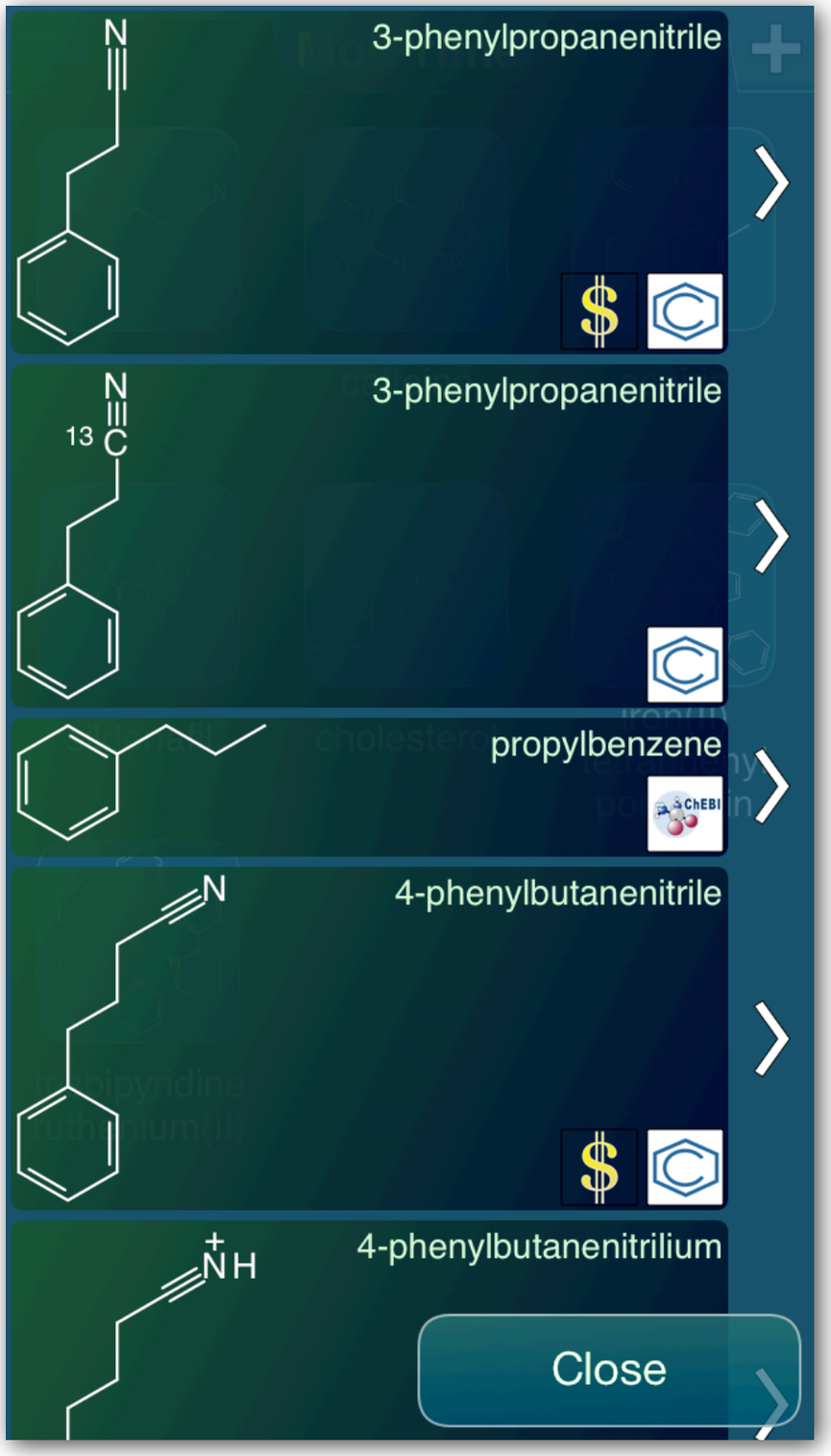
- Environmental data from ACS GCI & GSK

Feedstocks



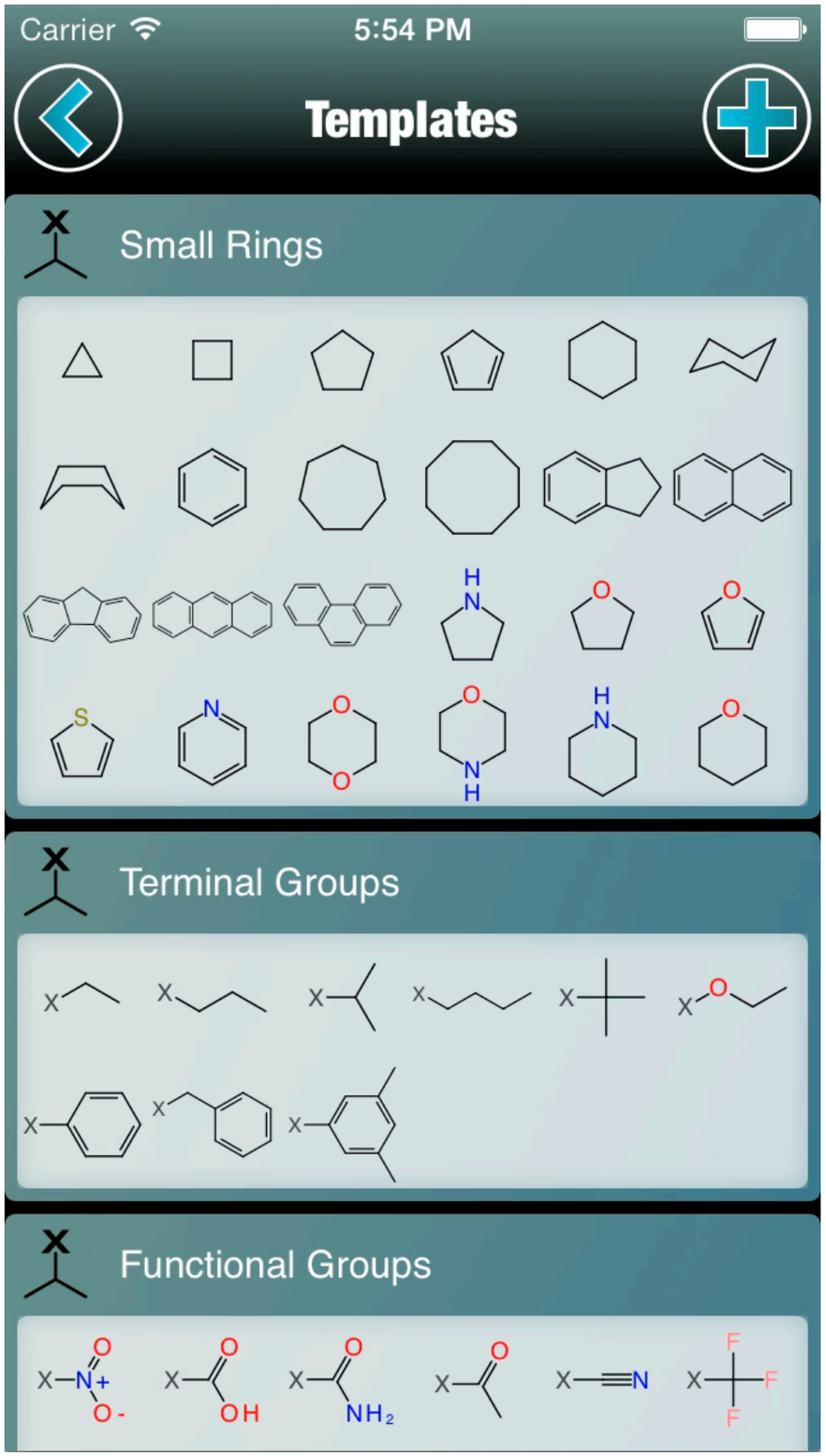
- Reference collection:
  - info about supply chain
  - encourage renewable use
- Link lab-available quantities to experiment records
- Could synchronise with inventory software
- Lookup in vendor catalogs...

Vendor Lookup



- MetaSearch engine:
  - PubChem
  - ChEBI
  - ChemSpider
- Vendor links available
- Can integrate more services

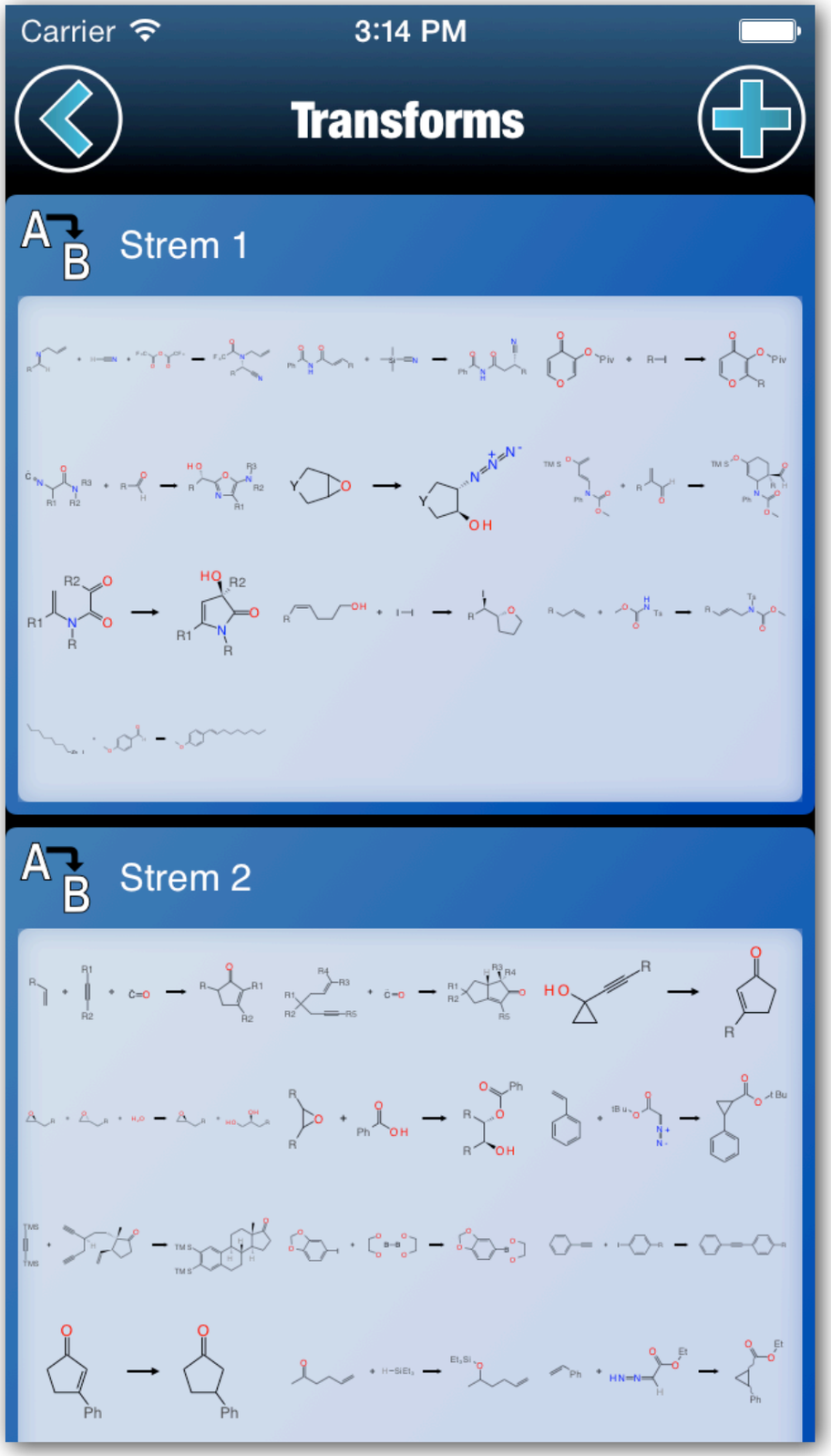
Templates

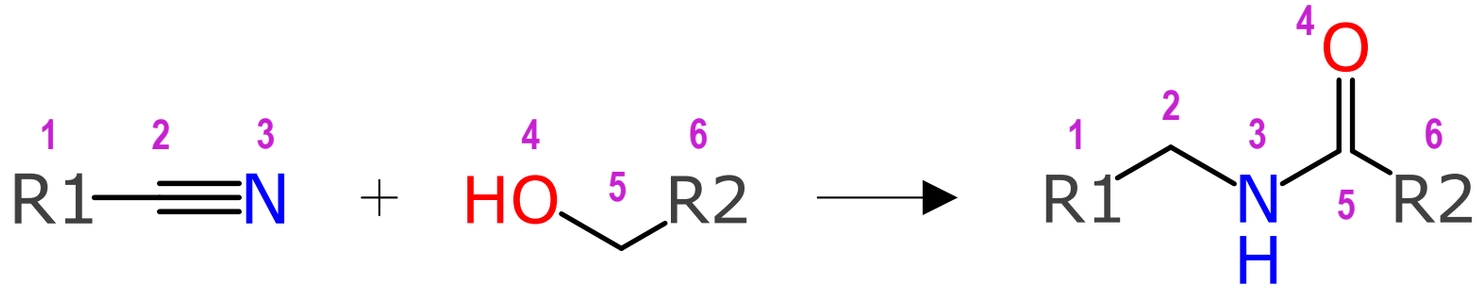


- Editable list
- Advanced placement algorithms
- Complicated ring systems
- Inorganic ligands

Journal of Cheminformatics, **2**:8 (2010)

Reaction Transforms





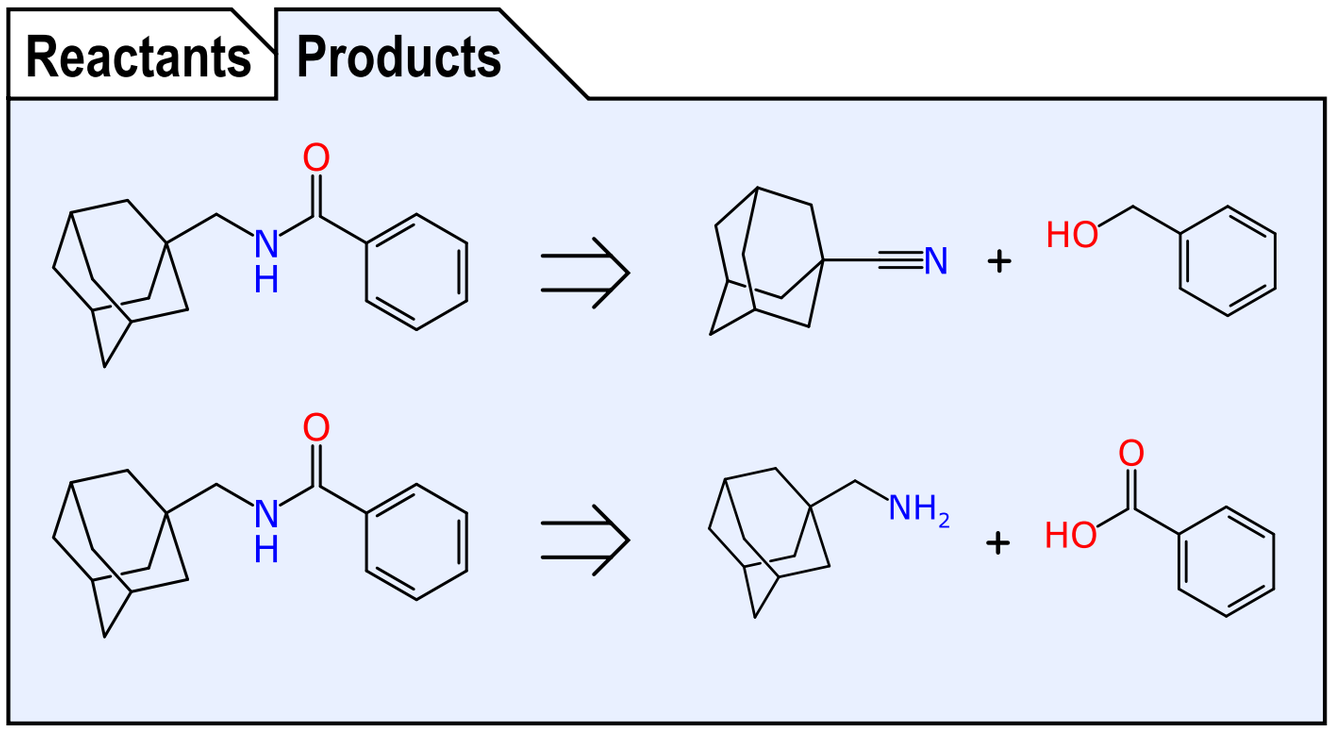
- Provide a pre-curated list of "green" reaction transforms

- Promote user entered experiments into transforms (numbering, clipping)

- Associate with:

  - reagents, catalysts & solvents
  - stoichiometry & quantities
  - yield & experimental conditions
  - literature & green reference data

Matching Transforms



- Search by Reactants or Products
- Defer to webservice
- User-defined transforms & server collection

- Use sketch coordinates of query & transform molecules to produce consistent orientations
- Allow filtering & sorting by yield, green metrics, availability of ingredients, etc.

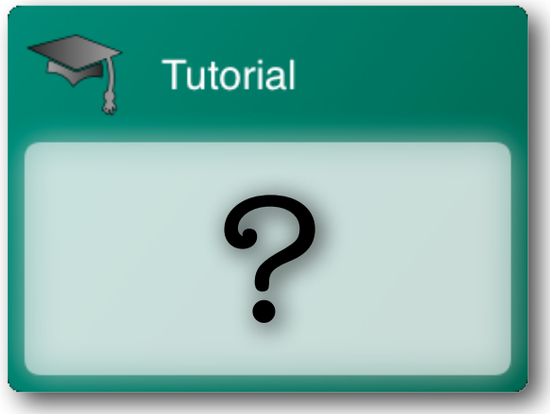
Tutorials

- Learned the hard way: intolerance for steep learning curves

- Scientific software is inherently complicated

- Two strategies:

  - workthrough tutorials
  - feature level-up: unlock



- Careful to associate features with value

Consumer vs. Enterprise

- Business paradox:
  - apps expected to be low cost
  - not all markets are high volume

Consumer	Enterprise
• App is cheap	• App cost negligible
• Webservices are free	• Services licensed
• Data access unsecured	• Secure in-house data
• No integration	• Hybrid infrastructure

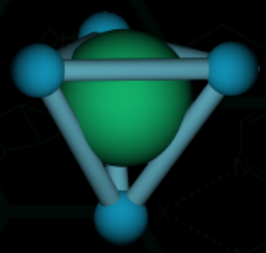
Acknowledgments

- Guido Kirsten
- Thomas Engel
- GDCh organisers
- RSC & ChemSpider, Eidogen-Sertanty, InfoChem, CDD, PubChem, ChEBI
- Inquiries to **info@molmatinf.com**

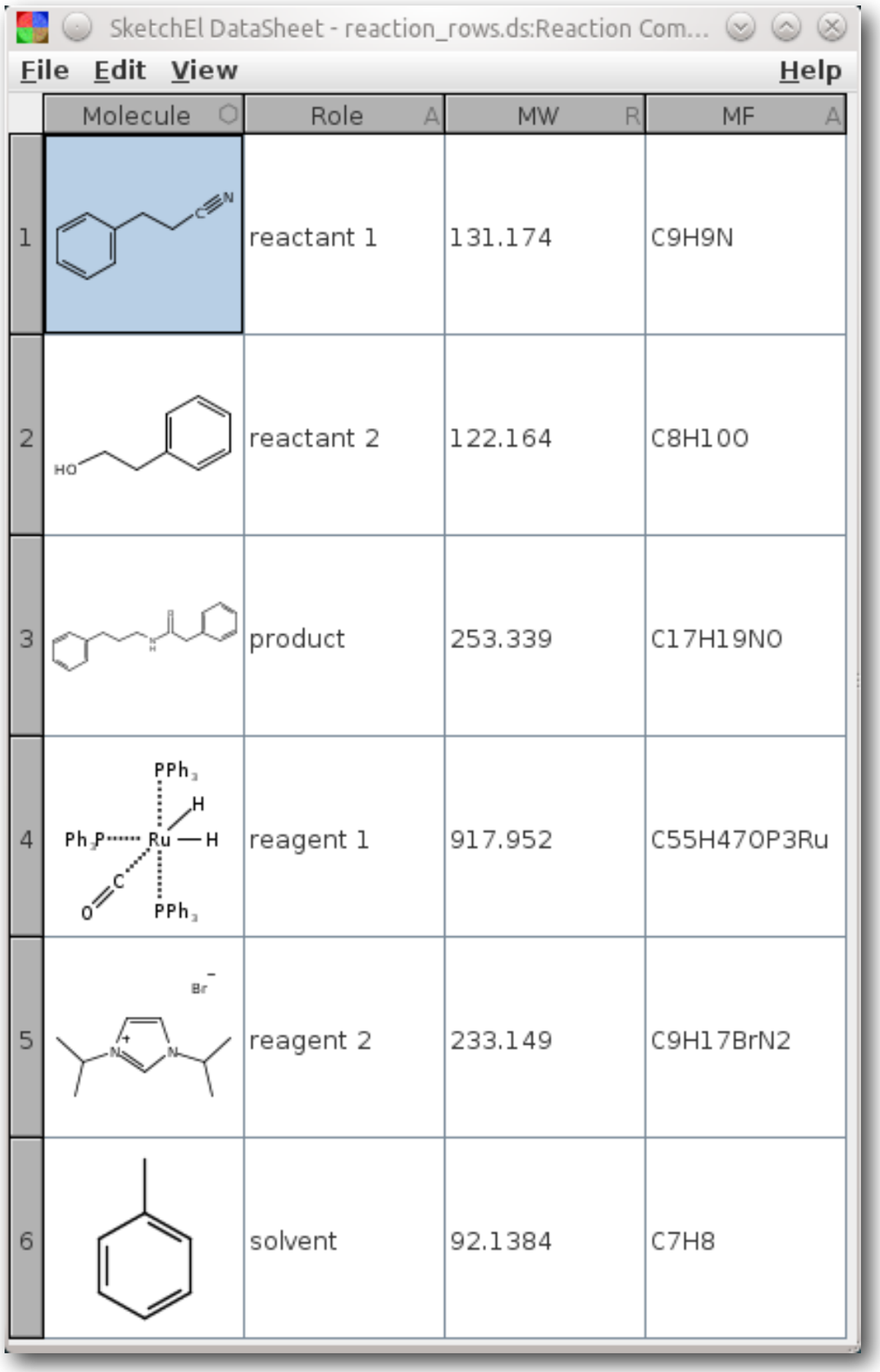
MOLECULAR MATERIALS INFORMATICS

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

@aclarkxyz



Extensible DataSheet



- Need a well designed format
- Surprisingly few widespread options, all inadequate:
  - MDL SDfile/RDfile
  - tab/comma-separated text
  - Excel
- Tabular, typed columns
- Native molecules
- Extensible meta-layers

DataSheet XML

```
<?xml version="1.0" encoding="UTF-8"?>
<DataSheet>
 <Summary>
  <Title>Reaction Components</Title>
  <Description><![CDATA[Compounds involved in amide formation reaction]]></Description>
 </Summary>
 <Header nrows="6" ncols="4">
  <Column name="Molecule" type="molecule" id="1">Molecular structure</Column>
  <Column name="Role" type="string" id="2">Role in reaction</Column>
  <Column name="MW" type="real" id="3">Molecular weight (g/mol)</Column>
  <Column name="MF" type="string" id="4">Molecular formula</Column>
 </Header>
 <Content>
  <Row id="1">
   <Cell id="1"><![CDATA[SketchEl!(10,10)
C=-0.8571,-0.7714;0,0,i0
C=0.4419,-0.0214;0,0,i2
C=1.7409,-0.7714;0,0,i2
C=-2.1562,-0.0214;0,0,i1
C=-3.4552,-0.7714;0,0,i1
C=-3.4552,-2.2714;0,0,i1
C=-2.1562,-3.0214;0,0,i1
C=-0.8571,-2.2714;0,0,i1
C=3.0400,-0.0214;0,0,i0
N=4.3391,0.7286;0,0,i0
1-2=1,0
2-3=1,0
1-4=1,0
4-5=2,0
5-6=1,0
6-7=2,0
7-8=1,0
8-1=2,0
3-9=1,0
9-10=3,0
!End]]></Cell>
   <Cell id="2"><![CDATA[reactant 1]]></Cell>
   <Cell id="3">131.174</Cell>
   <Cell id="4"><![CDATA[C9H9N]]></Cell>
  </Row>
  <Row id="2">
...
```

- Minimalistic baseline
- Data types:
  - **molecule**
  - **string**
  - **boolean**
  - **integer**
  - **real**
  - **extend**
- Header, column definitions
- Content: row-major
- Streamable

Reaction Aspect

```
<?xml version="1.0" encoding="UTF-8"?>
<DataSheet>
 <Summary>
  <Title>Amide Formation</Title>
  <Description><![CDATA[]]></Description>
 </Summary>
 <Header nrows="1" ncols="15">
  <Column name="ReactantMol1" type="molecule" id="1"/>
  <Column name="ReactantName1" type="string" id="2"/>
  <Column name="ReactantStoich1" type="string" id="3"/>
  <Column name="ReactantMol2" type="molecule" id="4"/>
  <Column name="ReactantName2" type="string" id="5"/>
  <Column name="ReactantStoich2" type="string" id="6"/>
  <Column name="ProductMol1" type="molecule" id="7"/>
  <Column name="ProductName1" type="string" id="8"/>
  <Column name="ProductStoich1" type="string" id="9"/>
  <Column name="ReagentMol1" type="molecule" id="10"/>
  <Column name="ReagentName1" type="string" id="11"/>
  <Column name="ReagentMol2" type="molecule" id="12"/>
  <Column name="ReagentName2" type="string" id="13"/>
  <Column name="ReagentMol3" type="molecule" id="14"/>
  <Column name="ReagentName3" type="string" id="15"/>
 </Header>
 <Extension>
  <Ext type="org.mmi.aspect.Reaction" name="Reaction">
   <![CDATA[nreactants=2
nproducts=1
nreagents=3
]]>
  </Ext>
 </Extension>
 <Content>
  <Row id="1">
   <Cell id="1"><![CDATA[SketchEl!(10,10)
C=-0.8571,-0.7714;0,0,i0
C=0.4419,-0.0214;0,0,i2
C=1.7409,-0.7714;0,0,i2
C=-2.1562,-0.0214;0,0,i1
C=-3.4552,-0.7714;0,0,i1
C=-3.4552,-2.2714;0,0,i1
C=-2.1562,-3.0214;0,0,i1
...
```

- An *aspect* is an extension field
- Implies additional behaviour & visualisation features
- Parser recognition optional:
  - fallback to baseline specification
  - datasheet still highly editable
  - backward *and* forward compatible
- **org.mmi.aspect.Reaction**:
  - comparable to MDL RDfile
  - includes reagent structures, stoichiometry
  - readily extendable