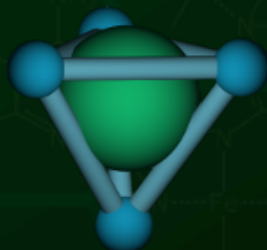


Putting together the pieces:

***Building a reaction-centric
electronic lab notebook for
mobile devices***

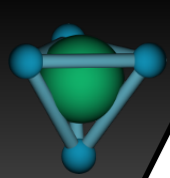
Alex M. Clark, Ph.D.

November 2013



© 2013 Molecular Materials Informatics, Inc.

<http://molmatinf.com>



Introduction



mainframes
minicomputers

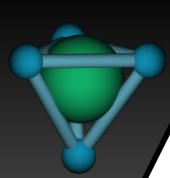


personal computers
portable laptops



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

incredibly
simple

no learning
curve

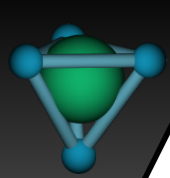
just
works

**app user
expectations**

delightful
to use

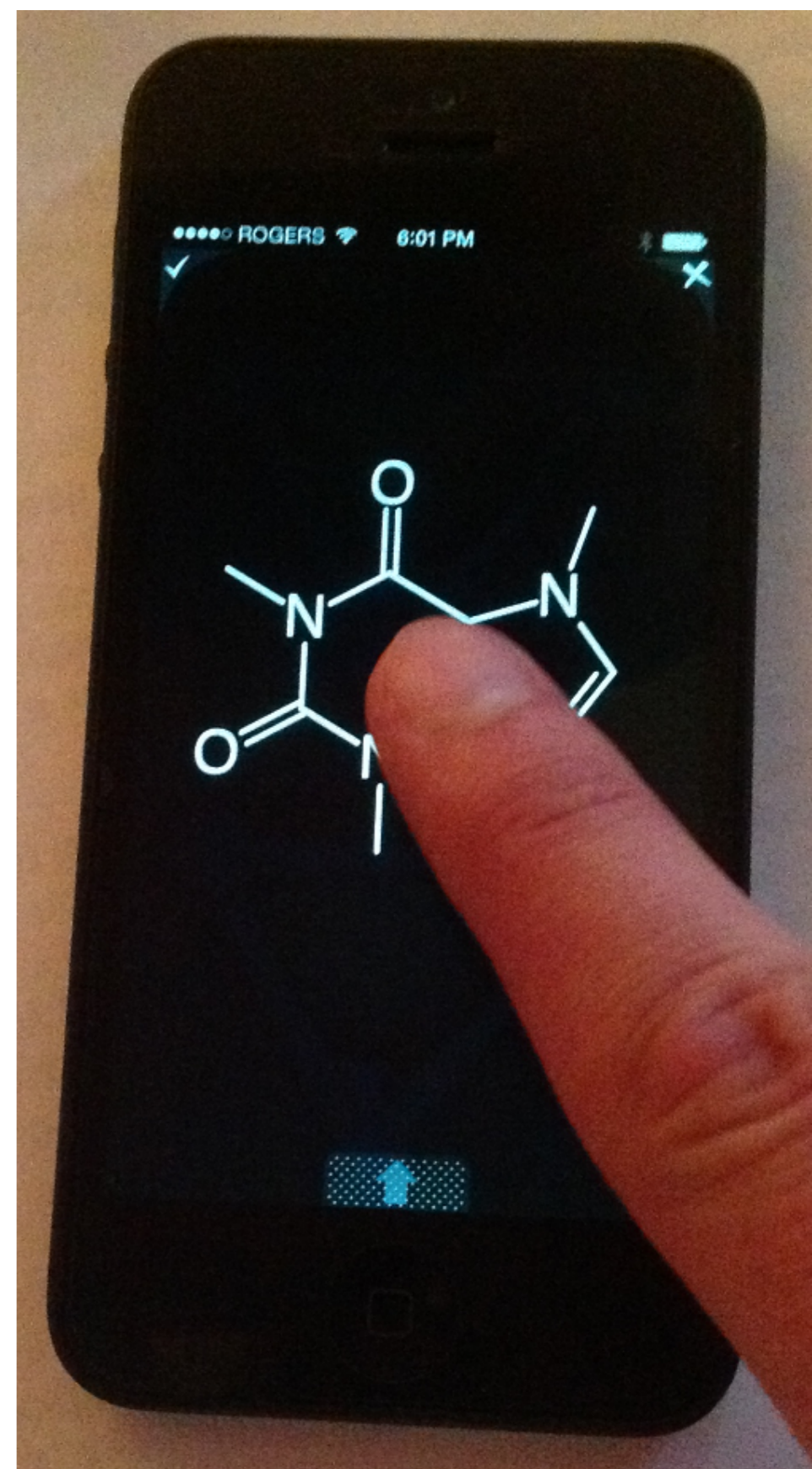
cheap

trivial
install



Technical Challenges

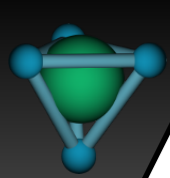
- Tiny screens (palm-sized)
- Fingers instead of mouse
- Reduced computing power
- Limited storage
- Delegate tasks to cloud servers: split between UI & API
- Multiple platforms



Chemistry App Ecosystem

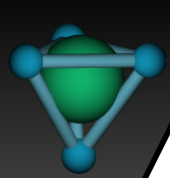


- Reference data
- Education
- Structure drawing
- Database searching
- 3D viewing
- Reactions & collections
- Property calculation
- Model building
- Graphical presentation
- Data sharing



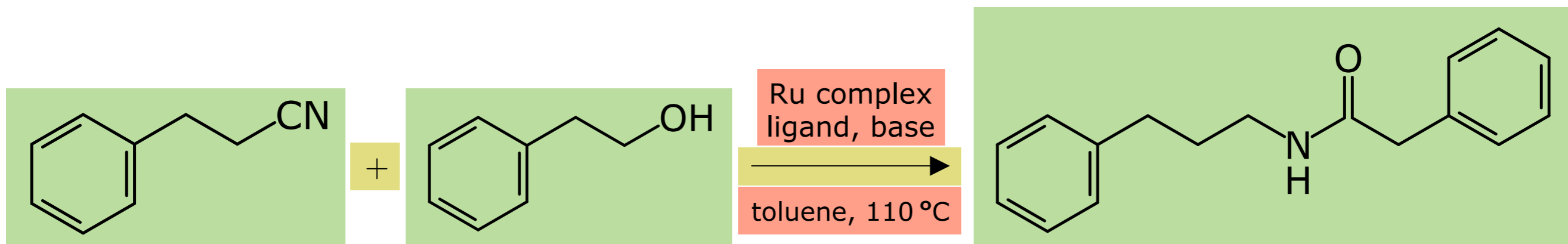
Reaction Lab Notebook?

- **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



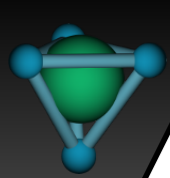
Reaction Canvas

- Reactants and products: atoms & bonds (mostly)



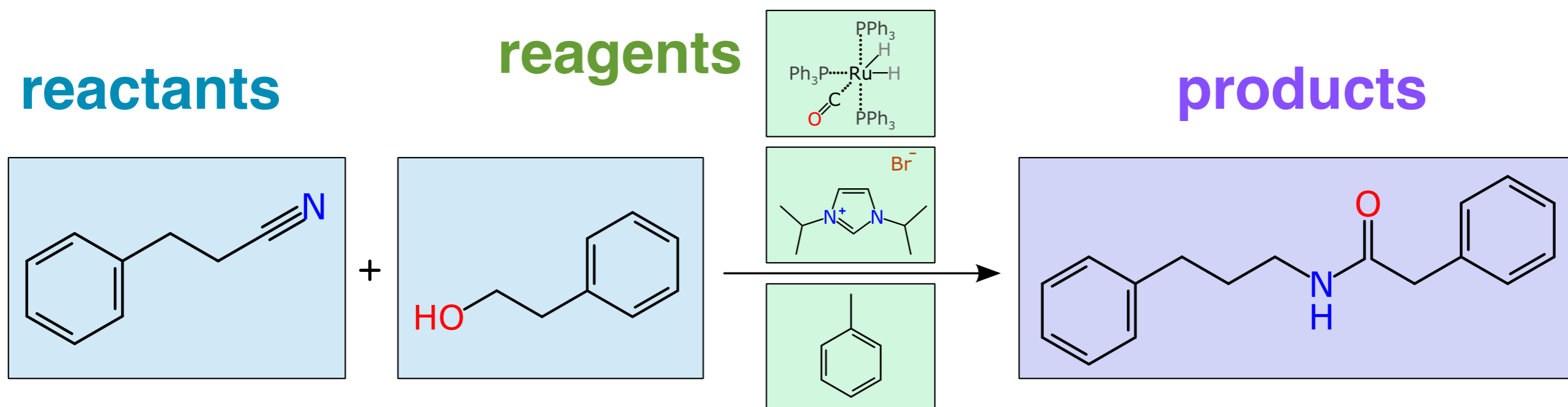
- Component separation: layout dependent
- Reagents, solvents, conditions: essentially meaningless

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

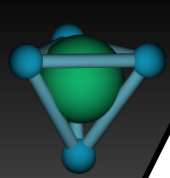


Reaction Representation

- Markup as component structures:

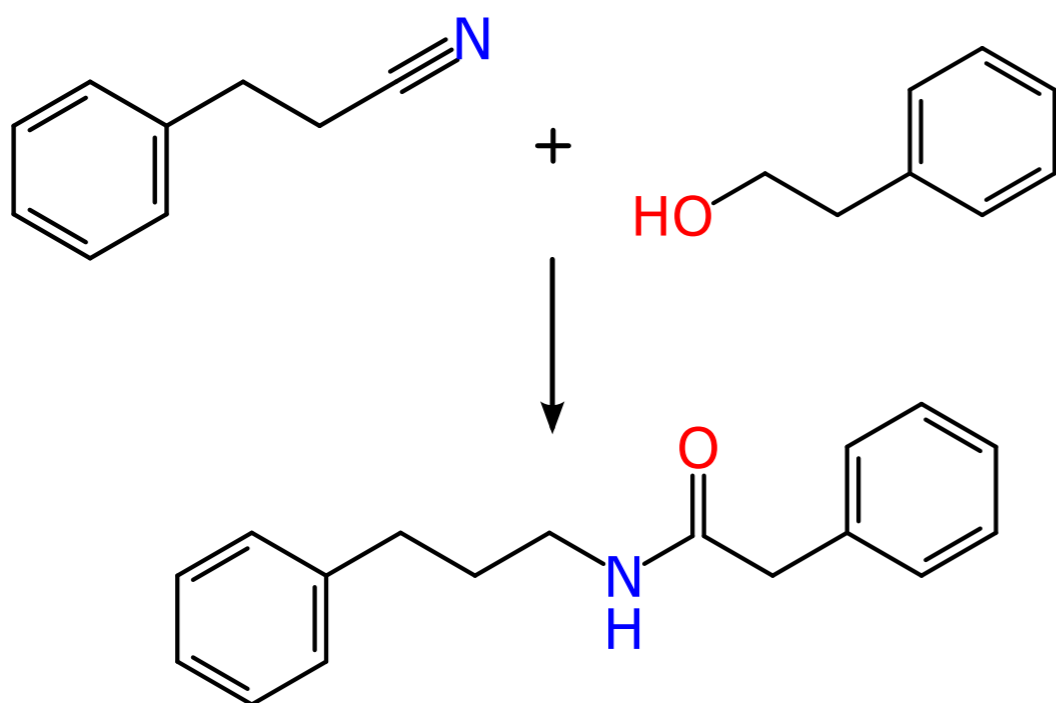


- Properties can be associated with each component, e.g. **name**, **stoichiometry**
- Draw each structure separately

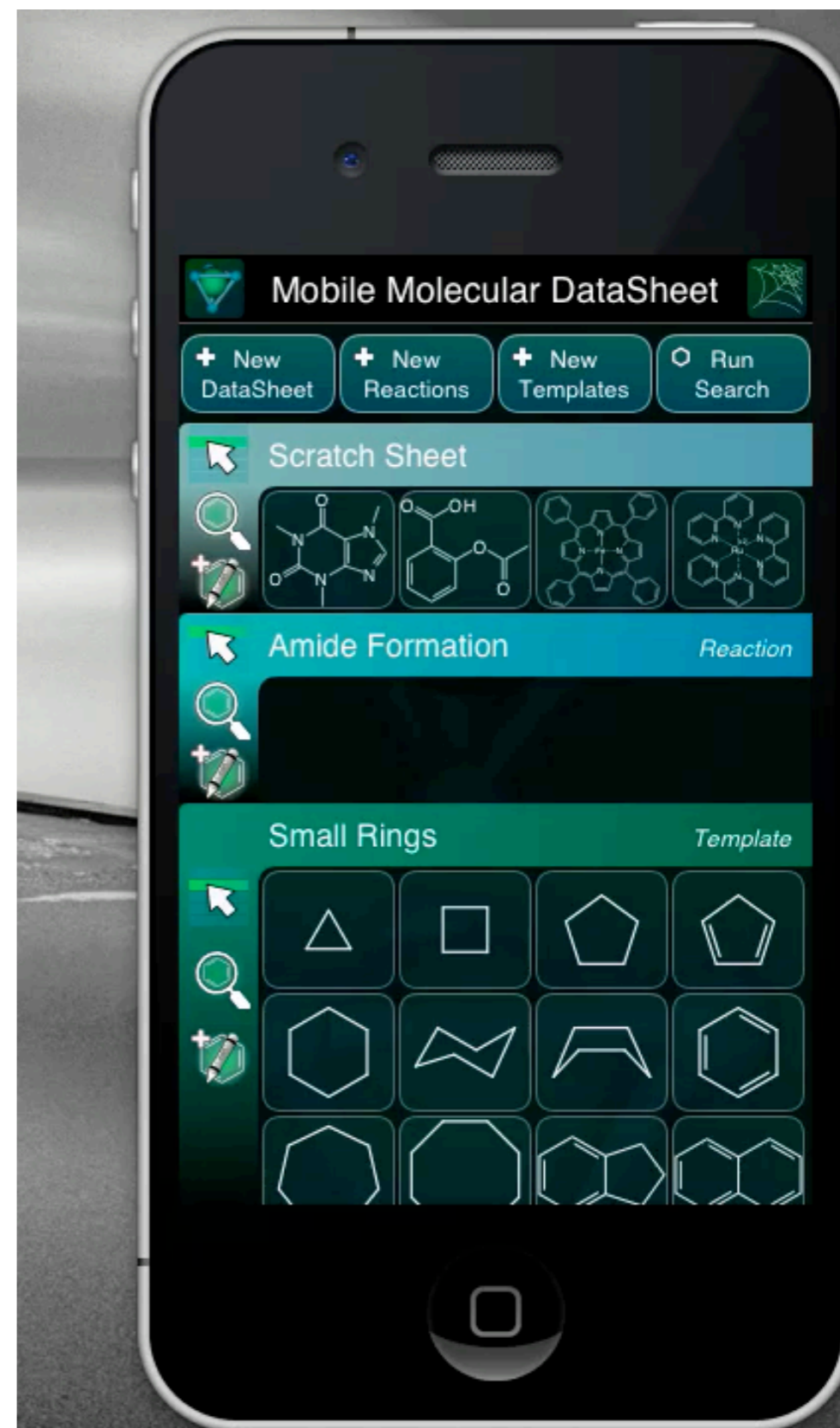
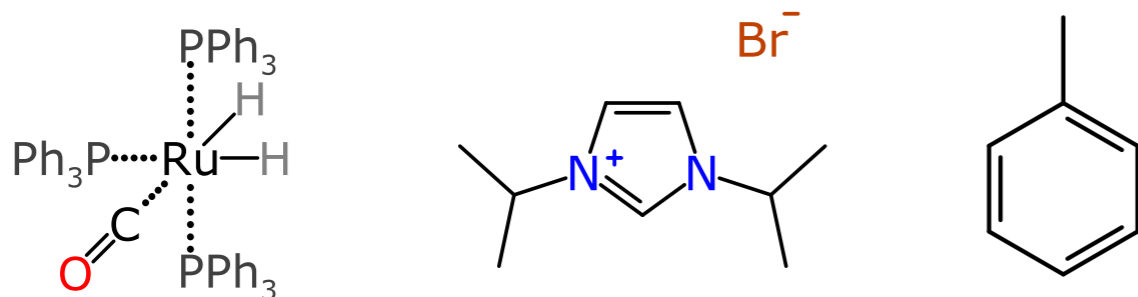


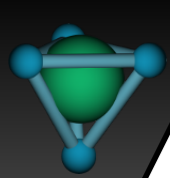
Sketcher

- Drawing reactants & products on an iPhone is quick



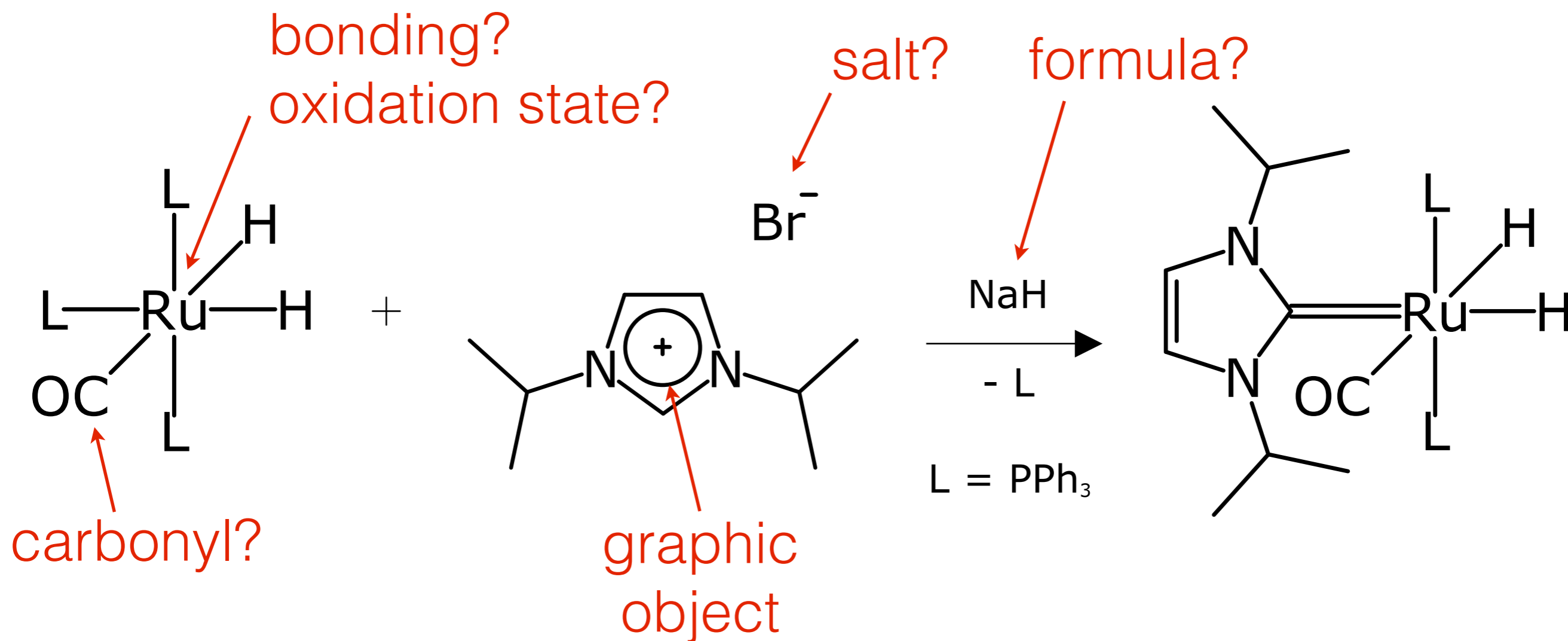
- Reagents & solvents added subsequently





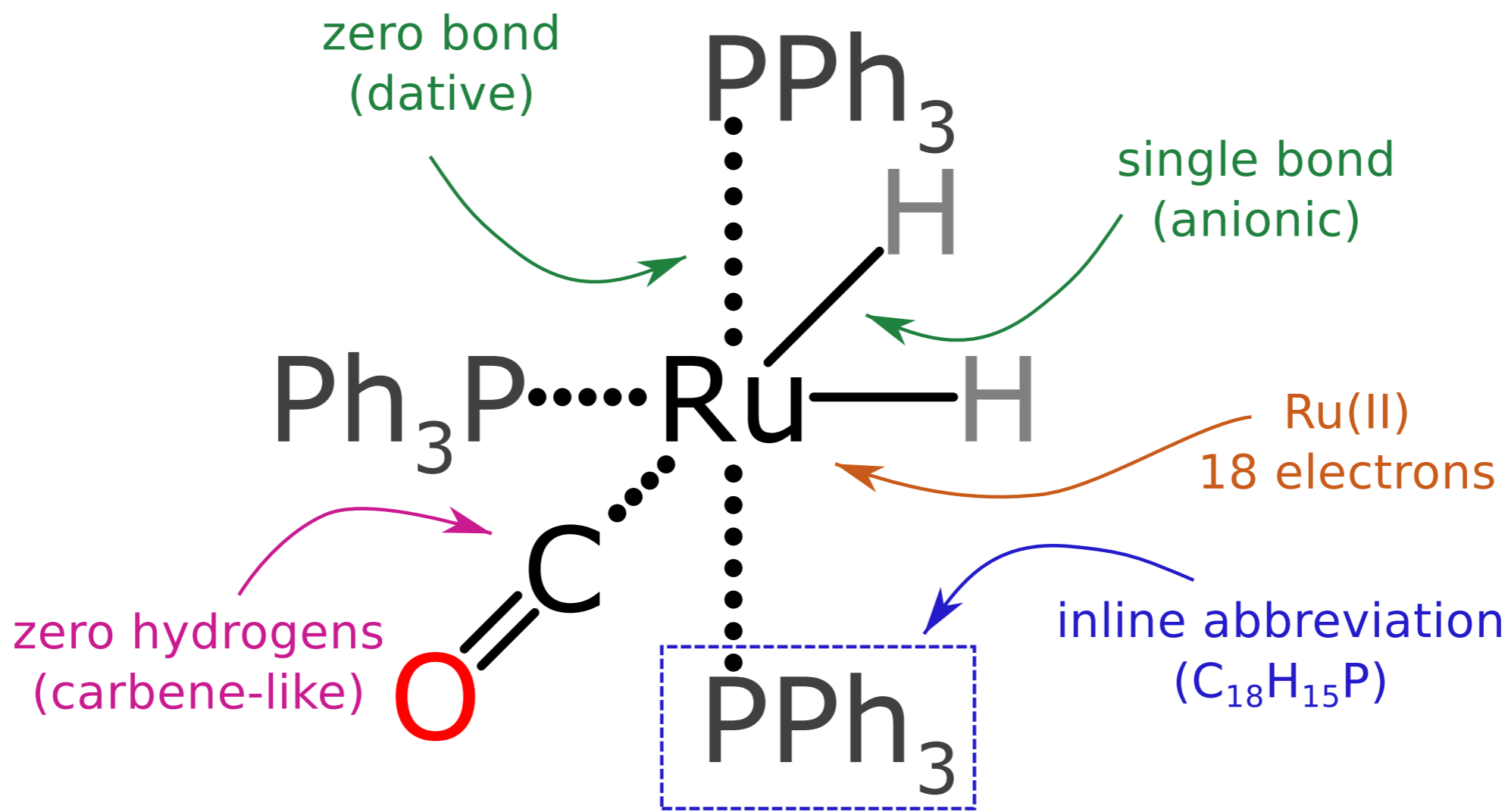
Organometallics

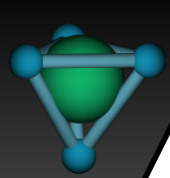
- Global & local style conventions anathema to machine-readability...



Organometallic Anatomy

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





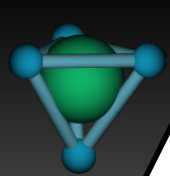
SketchE1 Format

```
SketchE1! (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,
    ↪aSketchE1!(20\002C22)\000A*
P|Ph{3}=0.0143,1.3477;0,0,i0,
    ↪aSketchE1!(20\002C22)\000A*
P|Ph{3}=-1.9857,-1.3143;0,0,i0,
    ↪aSketchE1!(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
```

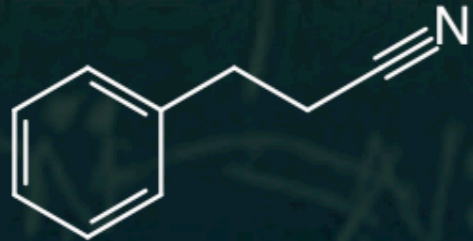

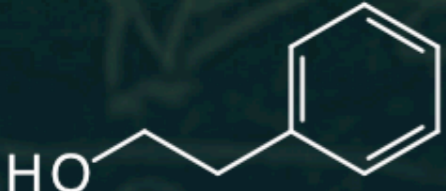

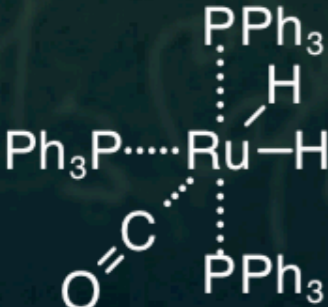

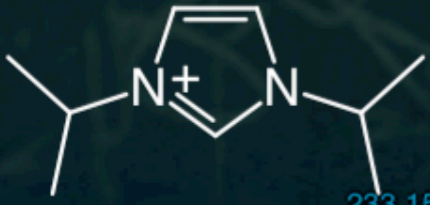

C₅₅H₄₇OP₃Ru
917.9517 g/mol

- Used by open source *SketchE1* & mobile apps
- Core principles:
 - extreme minimalism
 - extensibility (forward)
 - zero-order bonds
 - virtual hydrogen control
 - inline abbreviations
- Lowest common denominator:

SketchE1 \Leftrightarrow MDL Molfile



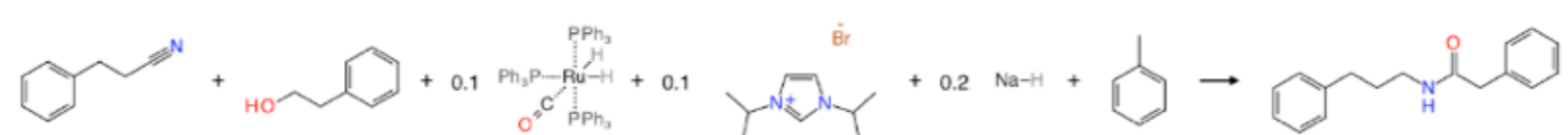
Quantities

 131.17 g/mol	Equiv: 1 Mass: 0.0655872 g Volume: <input type="text"/> Moles: 0.5 mmol Density: <input type="text"/> Conc: <input type="text"/> Primary: *	 9%
+		
 122.16 g/mol	Equiv: 1 Mass: 0.0671904 g Volume: <input type="text"/> Moles: 0.55 mmol Density: <input type="text"/> Conc: <input type="text"/> Primary: <input type="text"/>	 9%
+		
 917.95 g/mol	Equiv: 0.1 Mass: 0.0458976 g Volume: <input type="text"/> Moles: 0.05 mmol Density: <input type="text"/> Conc: <input type="text"/> Primary: <input type="text"/>	 6%
+		
 233.15 g/mol	Equiv: 0.1 Mass: 0.0116574 g Volume: <input type="text"/> Moles: 0.05 mmol Density: <input type="text"/> Conc: <input type="text"/> Primary: <input type="text"/>	 2%

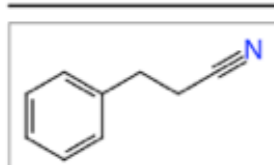
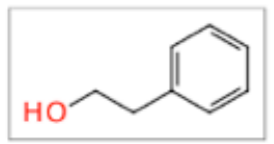
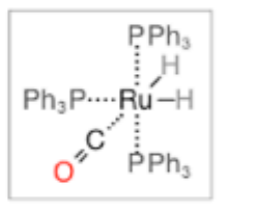

- Adds *quantities* to the Reaction aspect: auto-interconverted
- Green chemistry:
 - solvent reference
 - process mass intensity calculation
- Automatic lookup of structures, cross reference to *Mobile Reagents*
- Graphics creation
- Private & public sharing of data

Graphics Creation

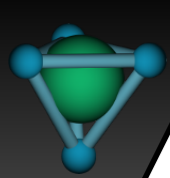
Done Print



Reactants

	Molecular Formula:	C ₉ H ₉ N
	Molecular Weight:	131.174 g/mol
	Equivalents:	1
	Mass:	0.0655872 g (calculated)
	Moles:	0.5 mmol
	(Primary Reactant)	
	Molecular Formula:	C ₈ H ₁₀ O
	Molecular Weight:	122.164 g/mol
	Equivalents:	1
	Mass:	0.0671904 g (calculated)
	Moles:	0.55 mmol
	Molecular Formula:	C ₅₅ H ₄₇ OP ₃ Ru
	Molecular Weight:	917.952 g/mol
	Equivalents:	0.1
	Mass:	0.0458976 g (calculated)
	Moles:	0.05 mmol (calculated)
	Molecular Formula:	C ₉ H ₁₇ BrN ₂
	Molecular Weight:	233.149 g/mol
	Equivalents:	0.1
	Mass:	0.0116574 g (calculated)

hardcopy:
print or
export
PDF



Public Sharing

The screenshot shows the MolSync Sharing interface. A chemical reaction scheme is displayed, showing the synthesis of a product from several reactants. A 'Download DataSheet' dialog box is open, allowing the user to choose a format for the data and adjust graphic options.

Download DataSheet

Choose Format

- DataSheet XML
- MDL RDF (reactions + data)
- MDL RXN (single reaction)
- PNG image (raster)
- SVG picture (vector)
- Encapsulated PostScript (vector)
- ZIP (multiple PNG files)
- ZIP (multiple SVG files)
- HTML with embedded SVG
- Microsoft Word
- Microsoft Excel

Graphic Options

Sizing:

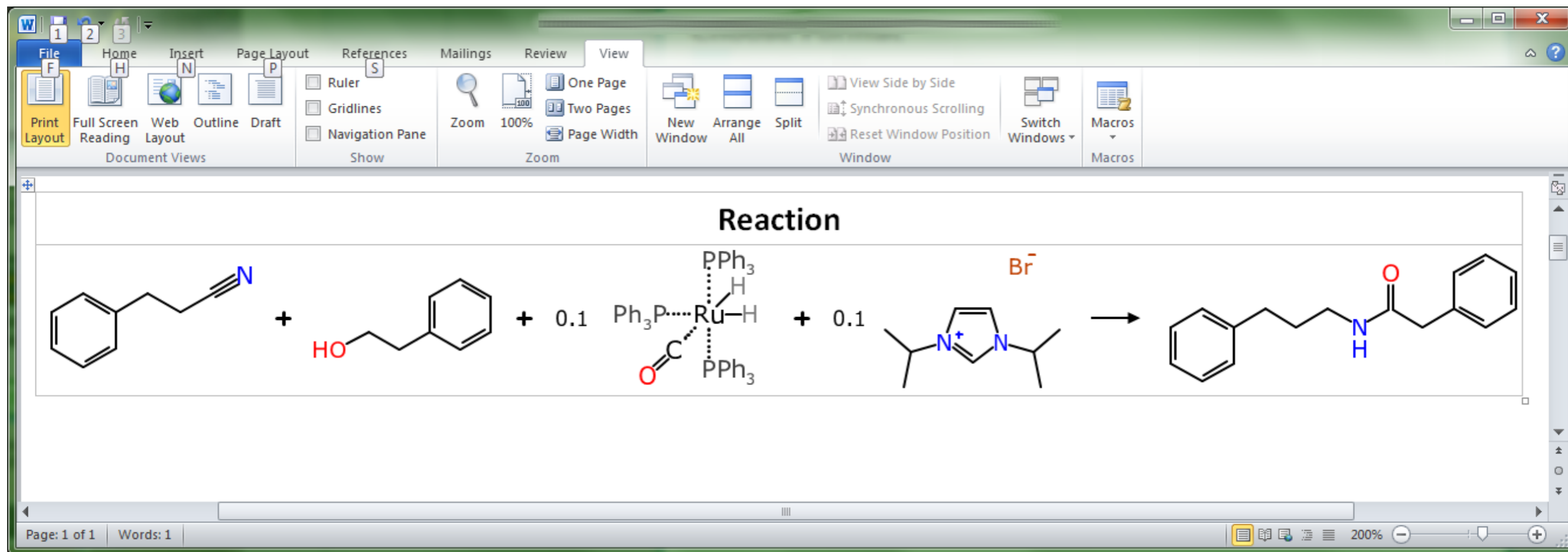
Angstroms-to-Points:

Rendering:

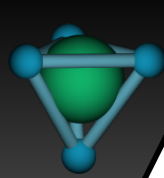
Reactant	Equivalents	MF	MW	Mass	Moles
Primary reactant	1	C ₉ H ₉ N	131.174	0.0655872 g	0.5 mmol
Reactant	1	C ₈ H ₁₀ O	122.164	0.0671904 g	0.55 mmol
Reactant	0.1000	C ₅₅ H ₄₇ OP ₃ Ru	917.952	0.0458976 g	0.05 mmol
Reactant	0.1000	C ₉ H ₁₇ BrN ₂	233.149	0.0116574 g	0.05 mmol
Product	1	C ₁₇ H ₁₉ NO	253.339	0.126669 g	0.5 mmol

- 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
<ul style="list-style-type: none">• structure drawing• reaction drawing• quantity calculation• green metrics• graphics• public databases• data sharing	<ul style="list-style-type: none">• professional chemists• calculation services• reference data• green chemistry• usability• data centralisation• sustainable business

The Green Lab Notebook (GLN)

Transforms

Tutorial

Feedstocks

Solvents

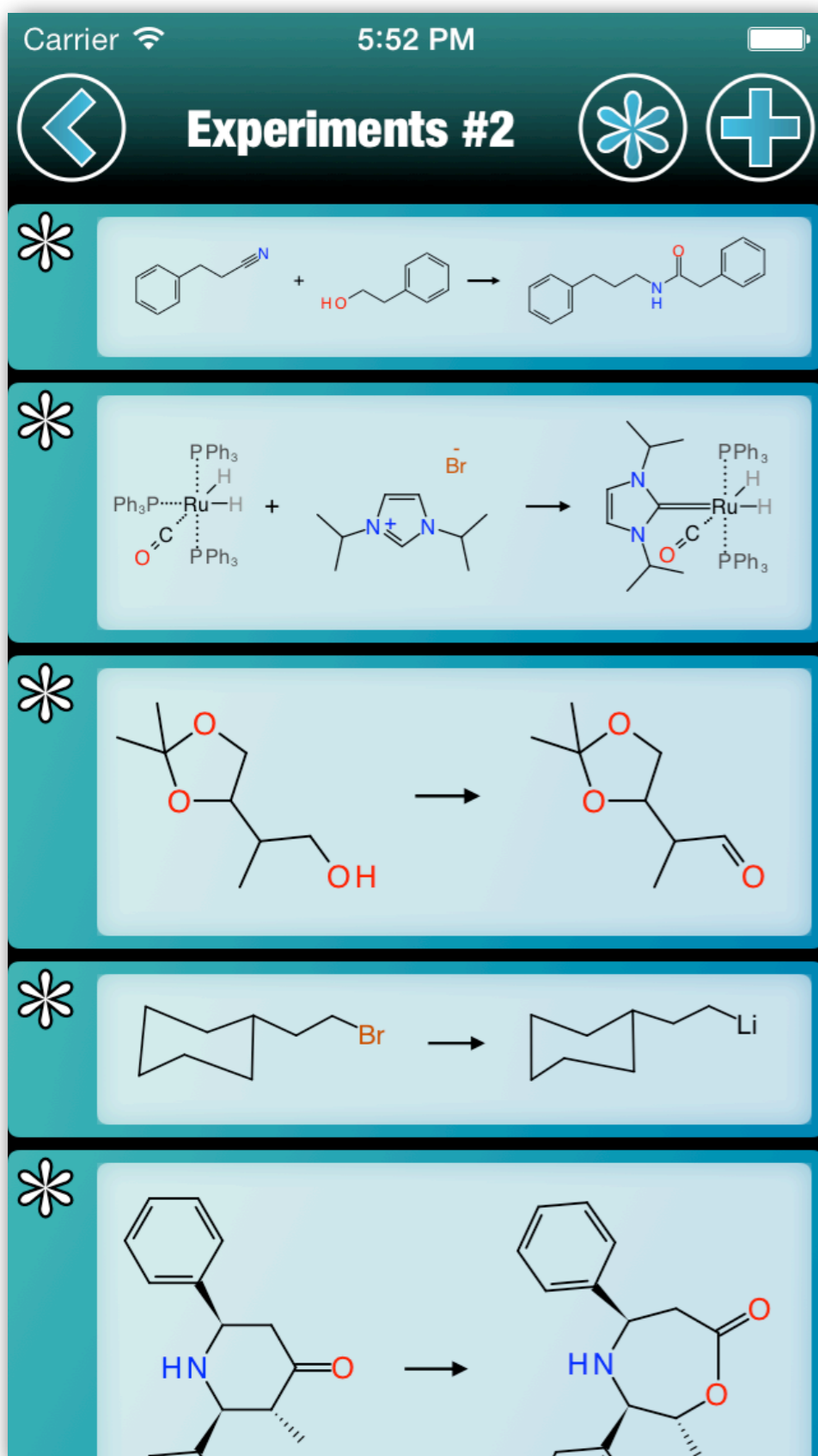
Green
Chemistry
Principles

Experiment
Folders

Scratch
Sheet

Templates

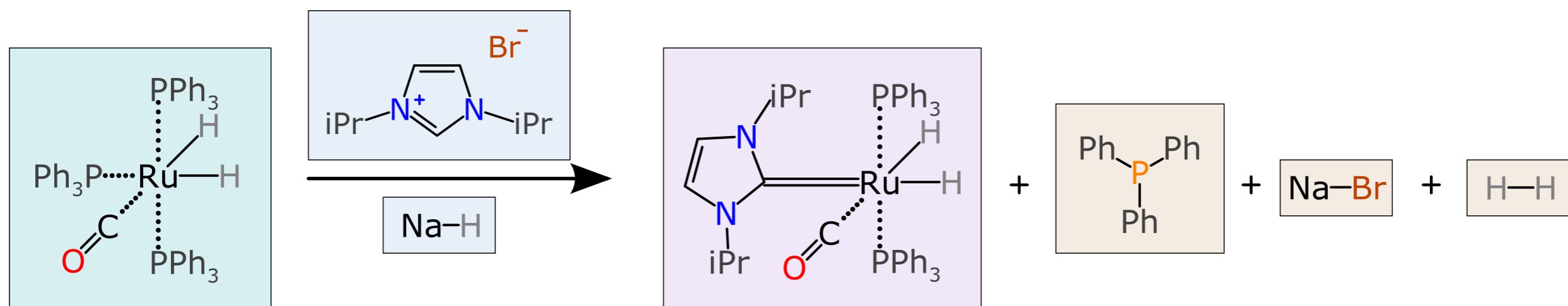
Folders of Experiments



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

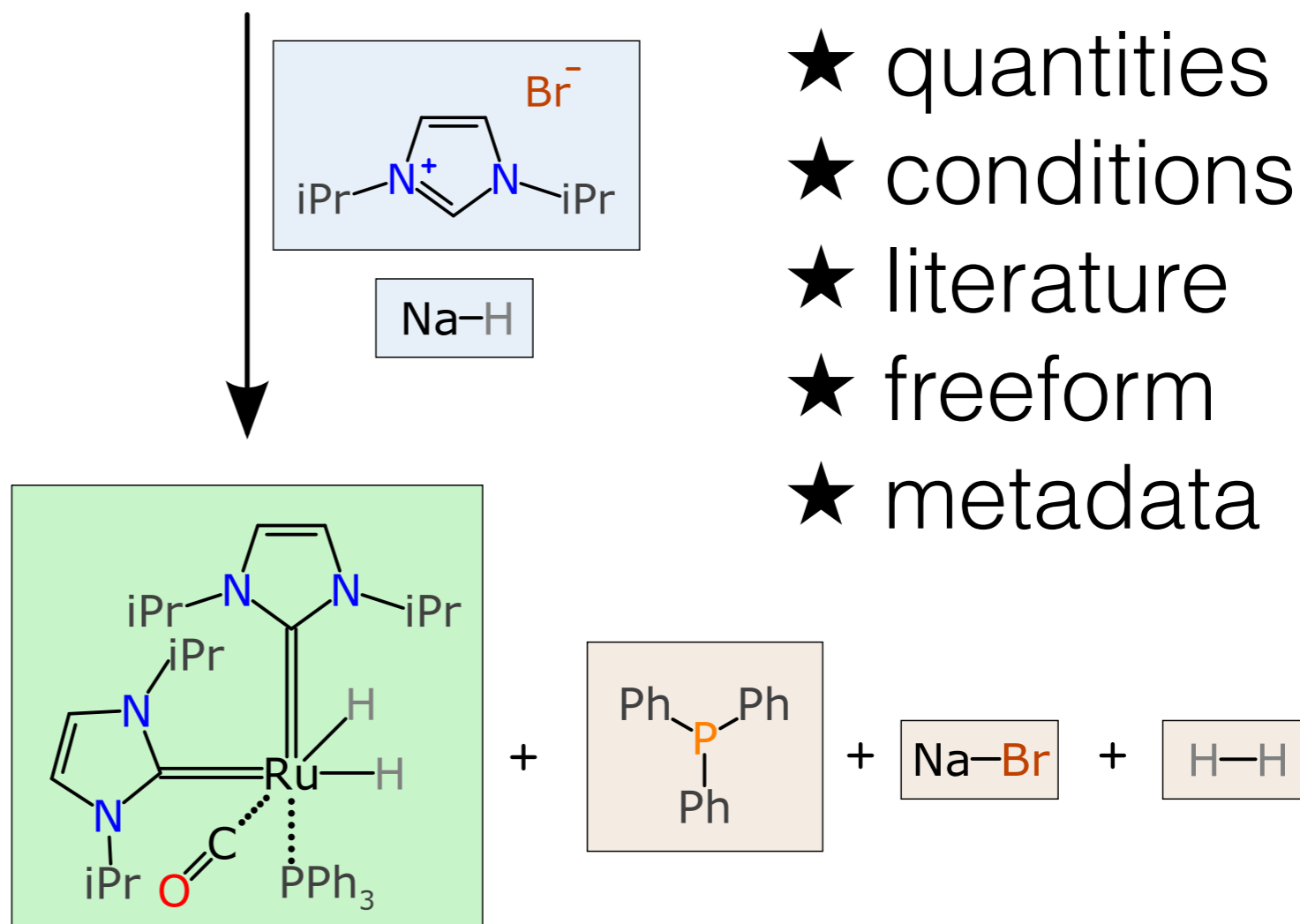
Reaction Experiments

- Multistep reaction editor, more supporting fields

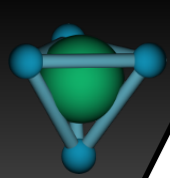


Multistep:

- ★ reactant
- ★ reagent
- ★ intermediate
- ★ product
- ★ waste



- ★ quantities
- ★ conditions
- ★ literature
- ★ freeform
- ★ metadata



Green Metrics

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



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

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

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

Green Solvents

Acid

HCOOH CC(=O)O CCC(=O)O

CC(=O)O

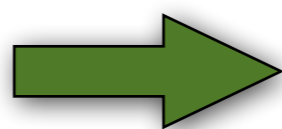
Acetic acid
CASRN: 64-19-7

3 **6**
Safety Health

6 **3** **6**
Air Water Waste

more

Aromatic



Carrier 6:27 PM

Common Solvents

* Formic acid

HCOOH

Class: Acid
CASRN: 64-18-6
Safety: 2
Health: 6
EnvAir: 5
EnvWater: 4
EnvWaste: 7
CSID: 278

* Acetic acid

CC(=O)O

Class: Acid
CASRN: 64-19-7
Safety: 3
Health: 6
EnvAir: 6
EnvWater: 3
EnvWaste: 6
CSID: 171

* Propionic acid

CCC(=O)O

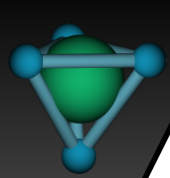
Class: Acid
CASRN: 79-09-4
Safety: 2
Health: 5
EnvAir: 6
EnvWater: 4
EnvWaste: 6
CSID: 1005

* Acetic anhydride

CC(=O)OC(=O)C

Class: Acid
CASRN: 108-24-7
Safety: 3

- Environmental data from ACS GCI & GSK



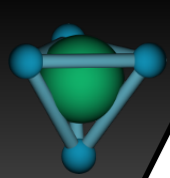
Feedstocks

Carrier 6:45 PM



Common Feedstocks


- * ethanol
Source: Fermentation
CCO
- * glucose
Source: Biomass
C1C(C(C(C(C(O1)O)O)O)O)O
- * butanol
Source: Fermentation
CCCCO
- * thiophene
Source: Fossil
C1=CC=C(S1)
- * glycine
Source: Biomass
NC(=O)O


- 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

3-phenylpropanenitrile  

3-phenylpropanenitrile 

propylbenzene 

4-phenylbutanenitrile  

4-phenylbutanenitrilium 

Close

Name: 3-phenylpropanenitrile
Synonyms: 3-phenylpropionitrile

Copy Import

Sourced from PubChem

 PubChem: [12581](#)

Vendors

ChemExper Chemical Directory
[HeV@@@RUITYZ'@@@@](#)

ZINC
[ZINC01747831](#)

Sigma-Aldrich
[171573_ALDRICH](#)

MolPort
[MolPort-003-927-093](#)

AKos Consulting & Solutions
[AKOS000121591](#)

ChemFrog
[888-584-707](#)

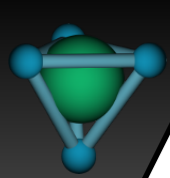
Chemical Synthesis Database
[645-59-0](#)

IS Chemical Technology
[114-50114](#)

Amadis Chemical
[206252](#)

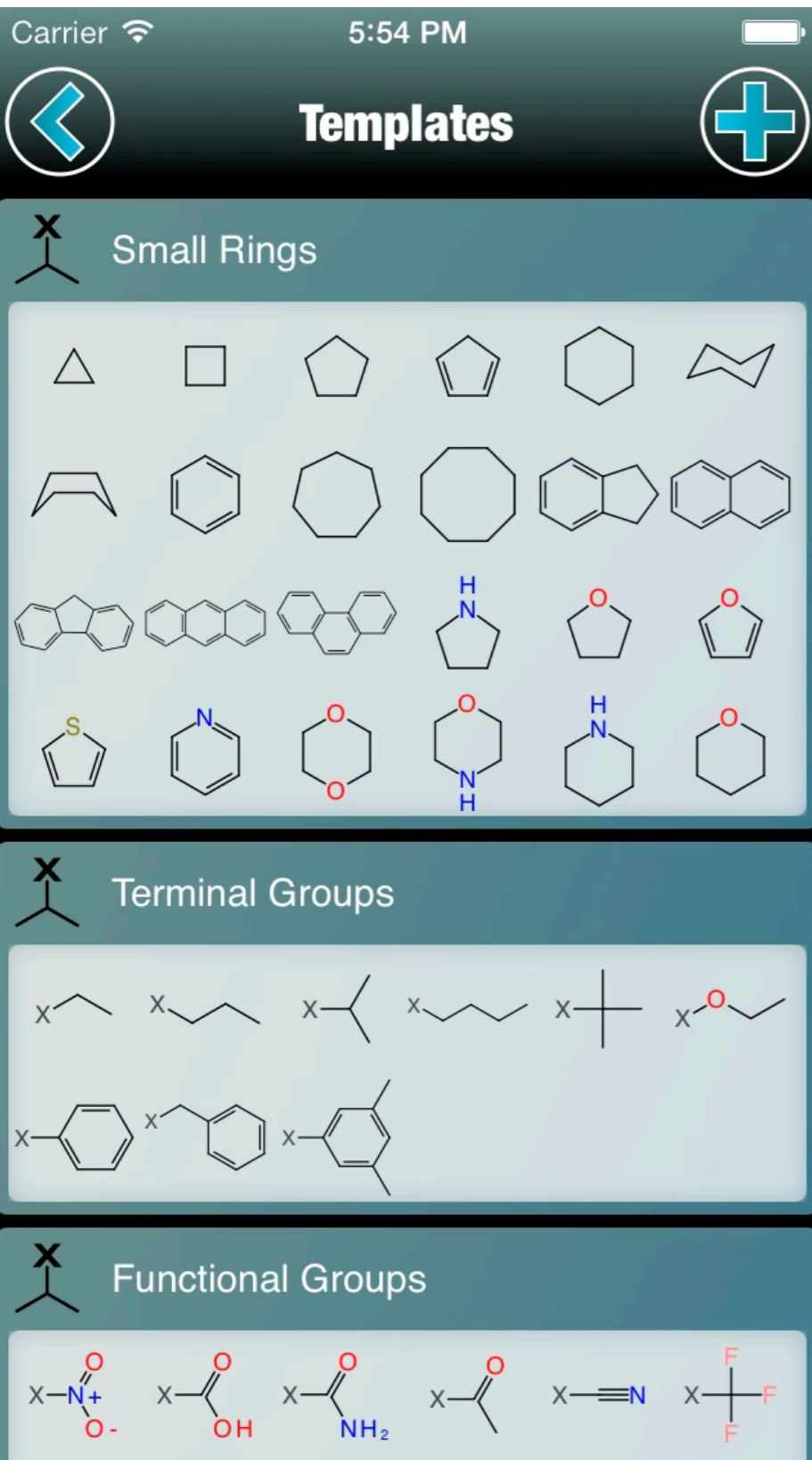
Close

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



Templates

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



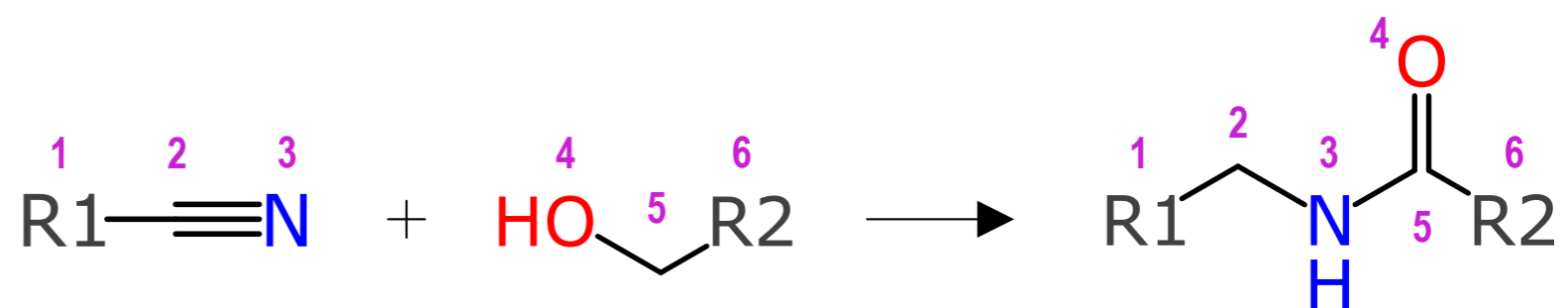
Reaction Transforms

Carrier 3:14 PM

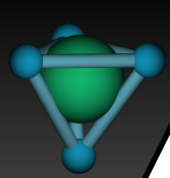
Transforms

Strem 1

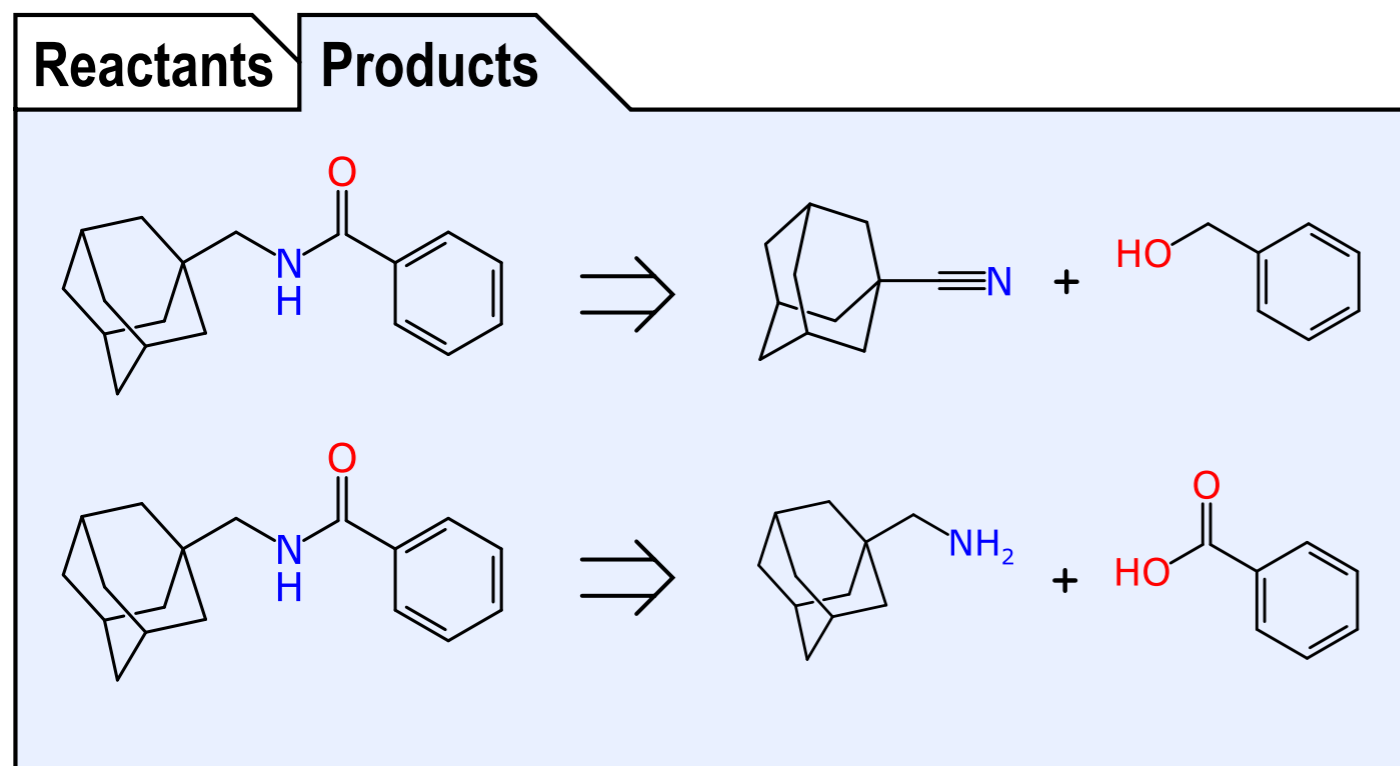
Strem 2



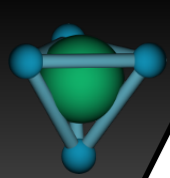
- Provide a pre-curated list of "green" reaction transforms
- Promote user entered experiments into transforms (numbering, clipping)
- Associate with:
 - reagents, catalysts & solvents
 - stoichiometry & quantities
 - yield & experimental conditions
 - literature & green reference data



Matching Transforms

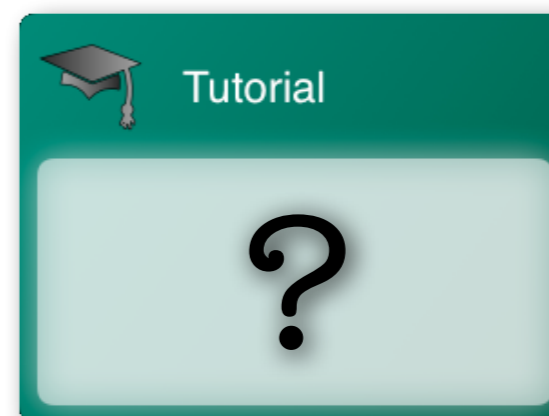


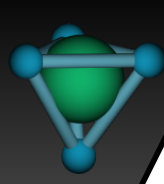
- Search by Reactants or Products
 - Defer to webservice
 - User-defined transforms & server collection
-
- Use sketch coordinates of query & transform molecules to produce consistent orientations
 - Allow filtering & sorting by yield, green metrics, availability of ingredients, etc.



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
<ul style="list-style-type: none">• App is cheap• Webservices are free• Data access unsecured• No integration	<ul style="list-style-type: none">• App cost negligible• Services licensed• Secure in-house data• Hybrid infrastructure

Acknowledgments

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

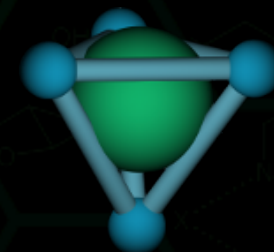
MOLECULAR MATERIALS INFORMATICS

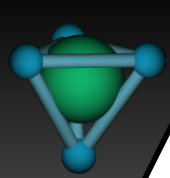
<http://molmatinf.com>

<http://molsync.com>

<http://cheminf20.org>

@aclarkxyz

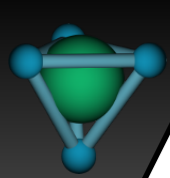




Extensible DataSheet

	Molecule	Role	MW	MF
1		reactant 1	131.174	C9H9N
2		reactant 2	122.164	C8H10O
3		product	253.339	C17H19NO
4		reagent 1	917.952	C55H47OP3Ru
5		reagent 2	233.149	C9H17BrN2
6		solvent	92.1384	C7H8

- 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>
```

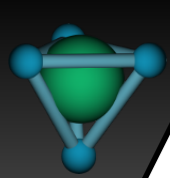
```
  <Cell id="4"><![CDATA[C9H9N]]></Cell>
```

```
</Row>
```

```
<Row id="2">
```

```
...
```

- Minimalistic baseline
- Data types:
 - molecule
 - integer
 - string
 - real
 - boolean
 - 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
...
]]></Cell>
    </Row>
  </Content>
</DataSheet>
```

- 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