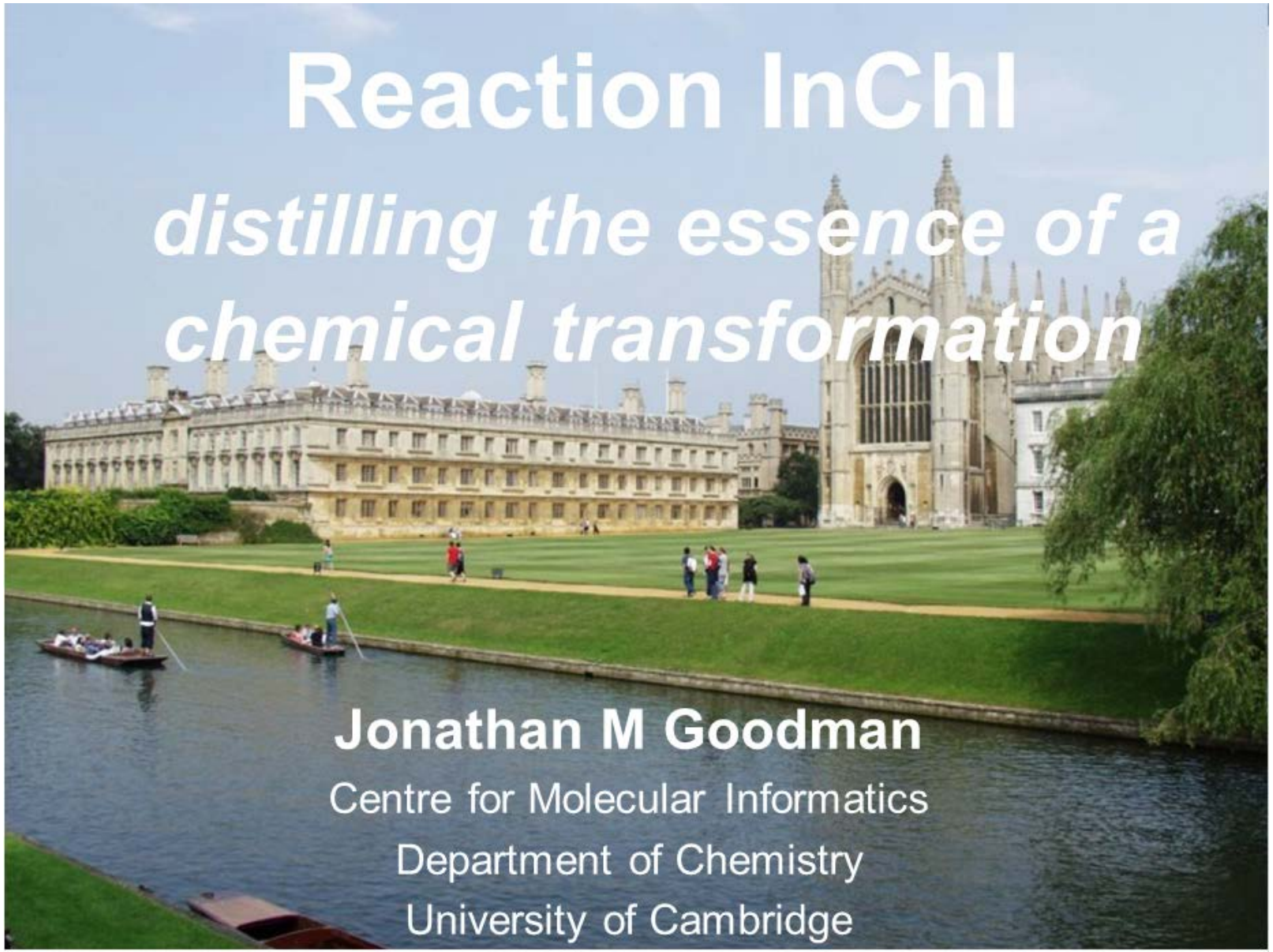


Reaction InChI

distilling the essence of a chemical transformation

Jonathan M Goodman

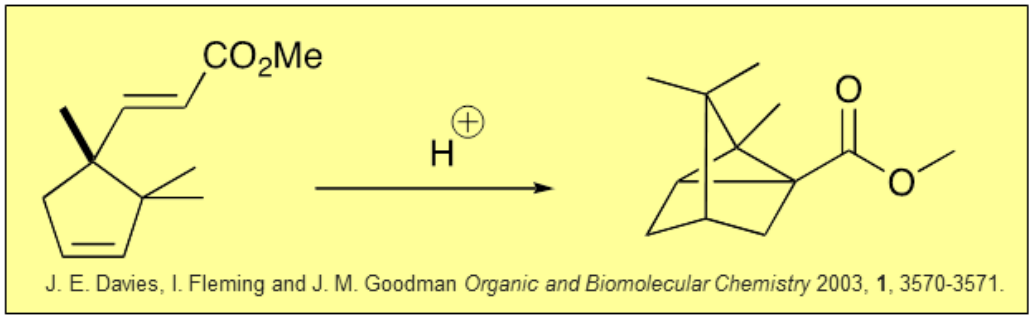
Centre for Molecular Informatics

Department of Chemistry

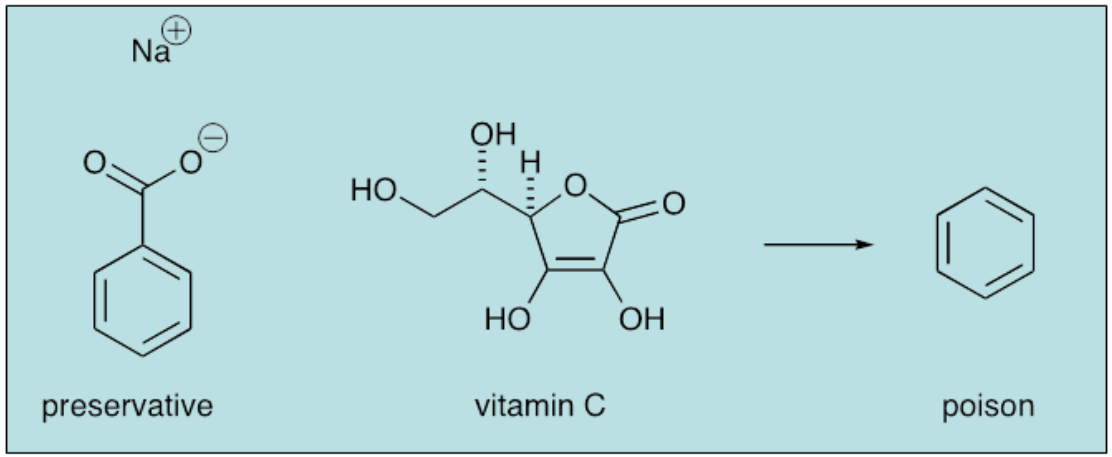
University of Cambridge

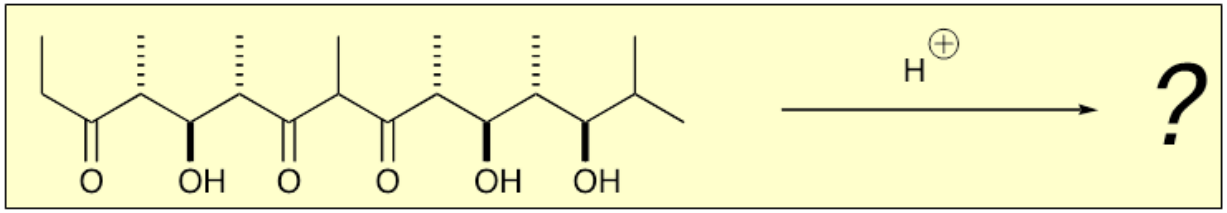
Extend InChI to reactions?

- “The IUPAC International Chemical Identifier (InChI™) is a non-proprietary identifier for chemical substances that can be used in printed and electronic data sources thus enabling easier linking of diverse data compilations”



J. E. Davies, I. Fleming and J. M. Goodman *Organic and Biomolecular Chemistry* 2003, **1**, 3570-3571.

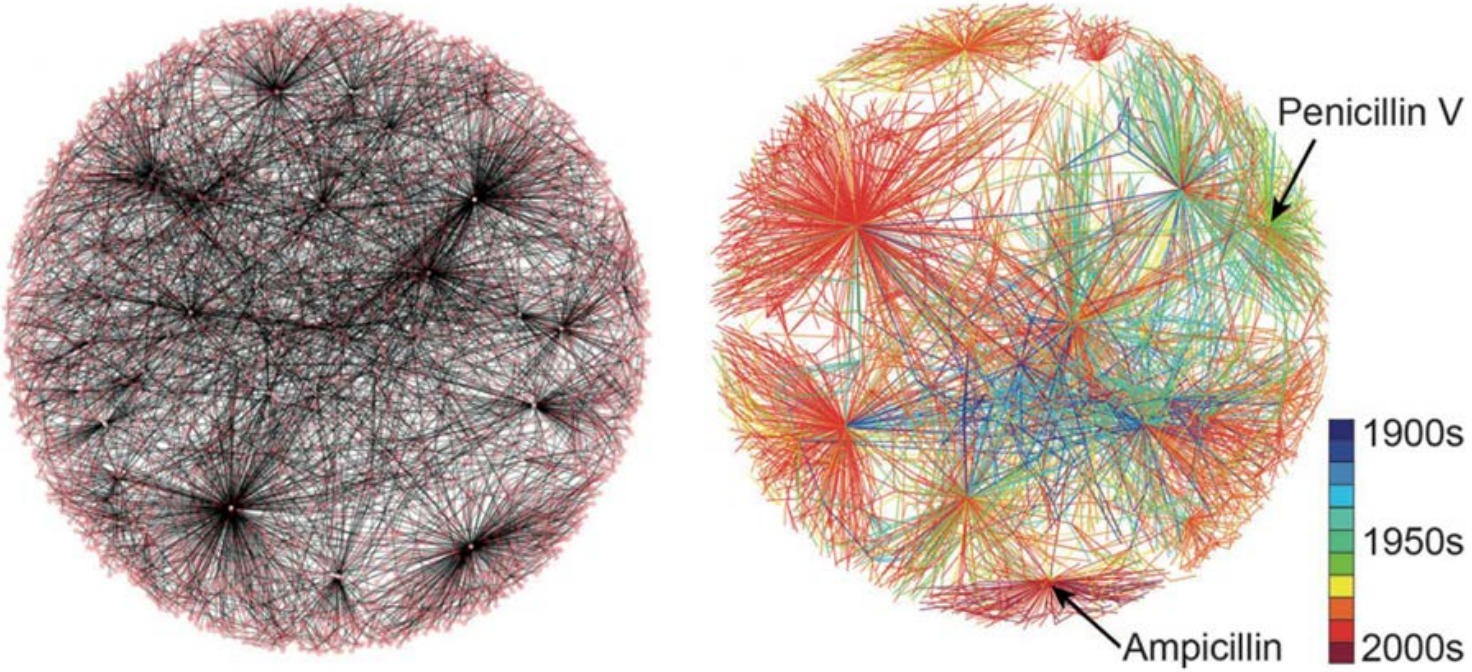




Chemical Reactions

Individual nodes represent the molecules and arrows represent reactions: a small fragment (0.1 %) of the organic chemistry in the Beilstein database

Colored by date reaction was first reported



Rewiring Chemistry: Algorithmic Discovery and Experimental Validation of One-Pot Reactions in the Network of Organic Chemistry

Chris M. Gothard, Siowling Soh, Nosheen A. Gothard, Bartlomiej Kowalczyk, Yanhu Wei, Bilge Baytekin, and Bartosz A. Grzybowski

Angew. Chem. Int. Ed. 2012, **51**, 7922–7927. DOI: 10.1002/anie.201202155

Reaction InChI *(RInChI)*

1) Is it possible to extend InChIs to reactions…
2) …and do something useful…
3) …that no one else is doing…
4) …and make it easy to use…
5) …and get people to use it…
6) …?????

A Reaction Identifier ?
possibilities

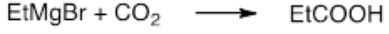
- Could different people working on the same reaction could produce the same identifier without communicating?
- Find duplication without sharing data (InChIKey)
- Could different reaction lists be combined easily?
- CAS, Beilstein *et al.* are not open and are slow to be updated
- How are Electronic Lab Notebooks going?
- Toxicologists, nano-scientists, biologists, *etc*, need chemistry (both molecules and reactions)

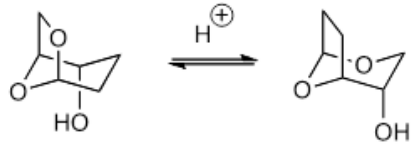
Extend InChI to reactions?

- The Royal Society of Chemistry has sponsored summer students
- IUPAC is sponsoring Reaction InChIs

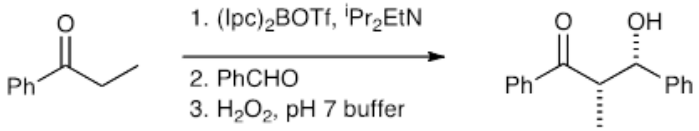
- Many people: *Günter Grethe, Steve Heller, Alan McNaught, Sandy Lawson, Keith Taylor, Martin Schmidt, Colin Batchelor, Henry Matuszczyk, Hans Kraut, Igor Pletnev, Tony Williams, Richard Kidd, Rudi Pisa, James Davies, Nick Parker, Chad Allen, Ben Hammond …*

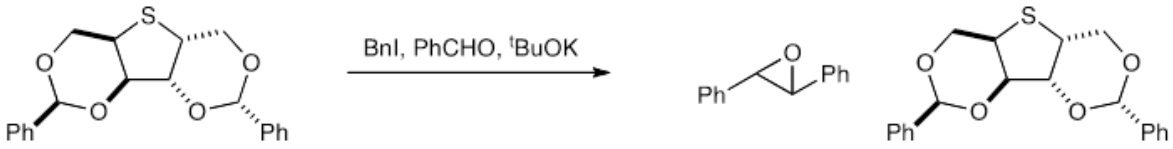
Reactions are hard to describe

A 

B 

What should be included?
What should be omitted?

C 

D 

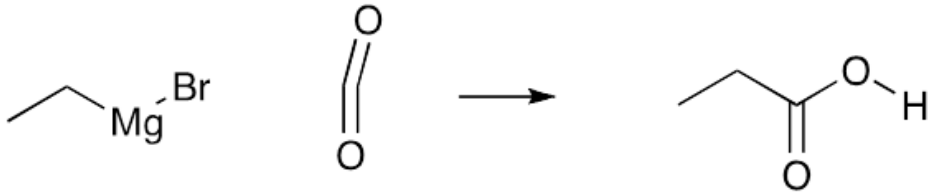
E lots of things
1.
2.
3.
4.
5.
.
... lots of steps ...
.
palyotoxin

RInChI version 0.1

RInChI = version #
/ LHS InChI 1 // LHS InChI 2 ...
/// RHS InChI 1 // LHS InChI 2 ...
/// both sides InChI 1 // both sides InChI 2 ...
/// additional information

RInChI version 0.01

RInChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3///C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)/d+



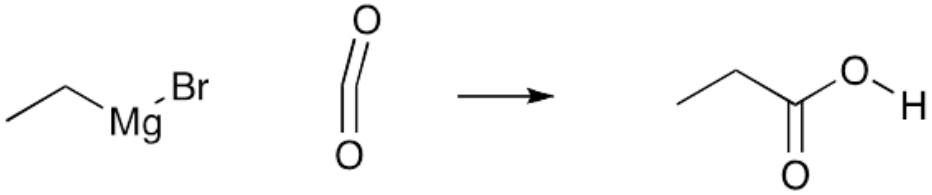
RInChI= version (RinChI.InChI)

/ reactants: InChI#1 // InChI#2

/// product: InChI#1 / reaction direction

RInChI version 0.01

RInChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3///C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)/d+



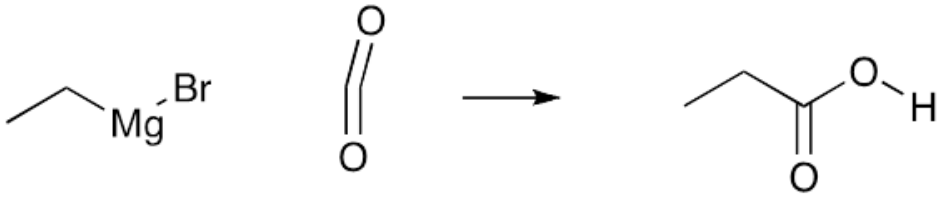
RInChI= version (RinChI.InChI)

/ reactants: InChI#1 // InChI#2

/// product: InChI#1 / reaction direction

RInChI version 0.01

RInChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3///C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)/d+



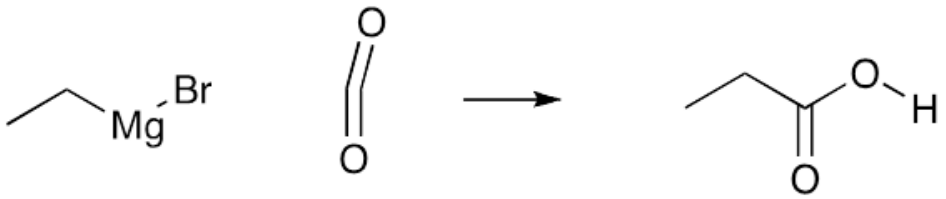
RInChI= version (RinChI.InChI)

/ reactants: InChI#1 // InChI#2

/// product: InChI#1 / reaction direction

RInChI version 0.01

RInChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3///C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)/d+



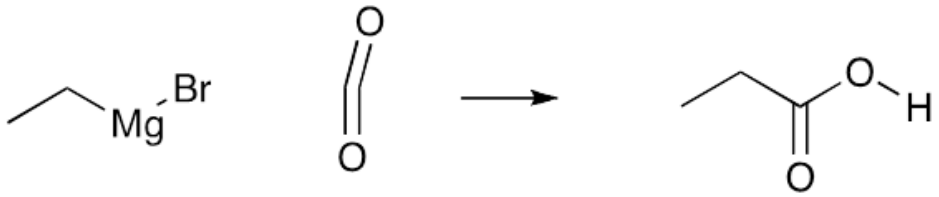
RInChI= version (RinChI.InChI)

/ reactants: InChI#1 // InChI#2

/// product: InChI#1 / reaction direction

RInChI version 0.01

RInChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3///C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)/d+



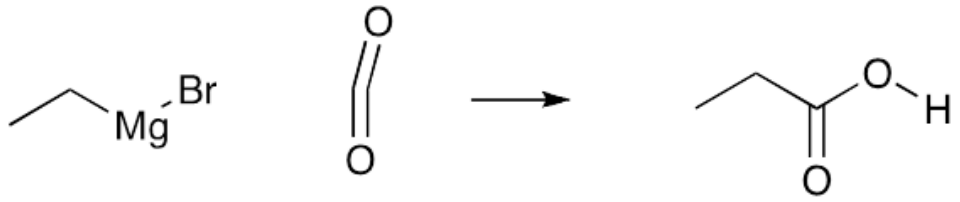
RInChI= version (RinChI.InChI)

/ reactants: InChI#1 // InChI#2

/// product: InChI#1 / reaction direction

RInChI version 0.01

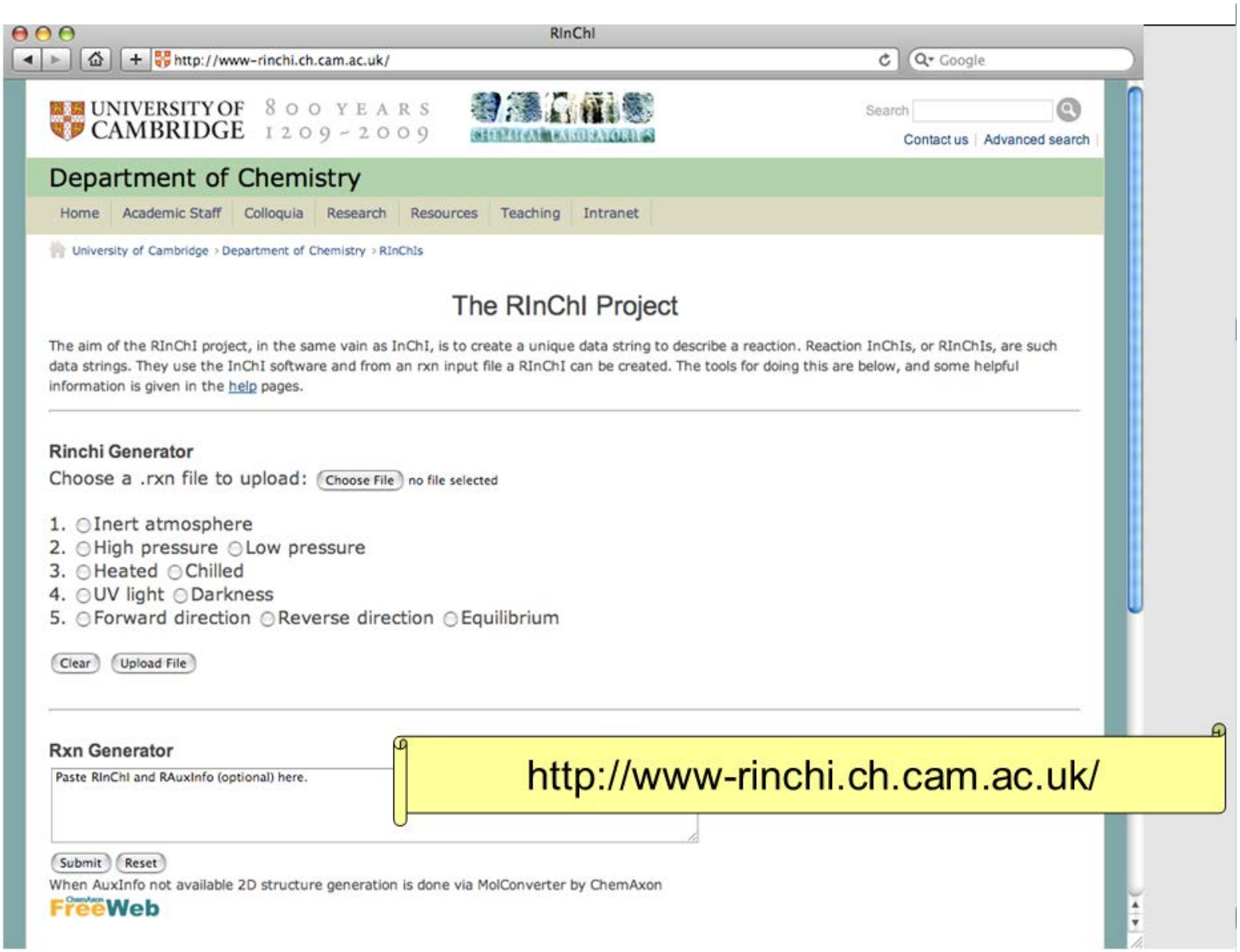
RInChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3///C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)/d+

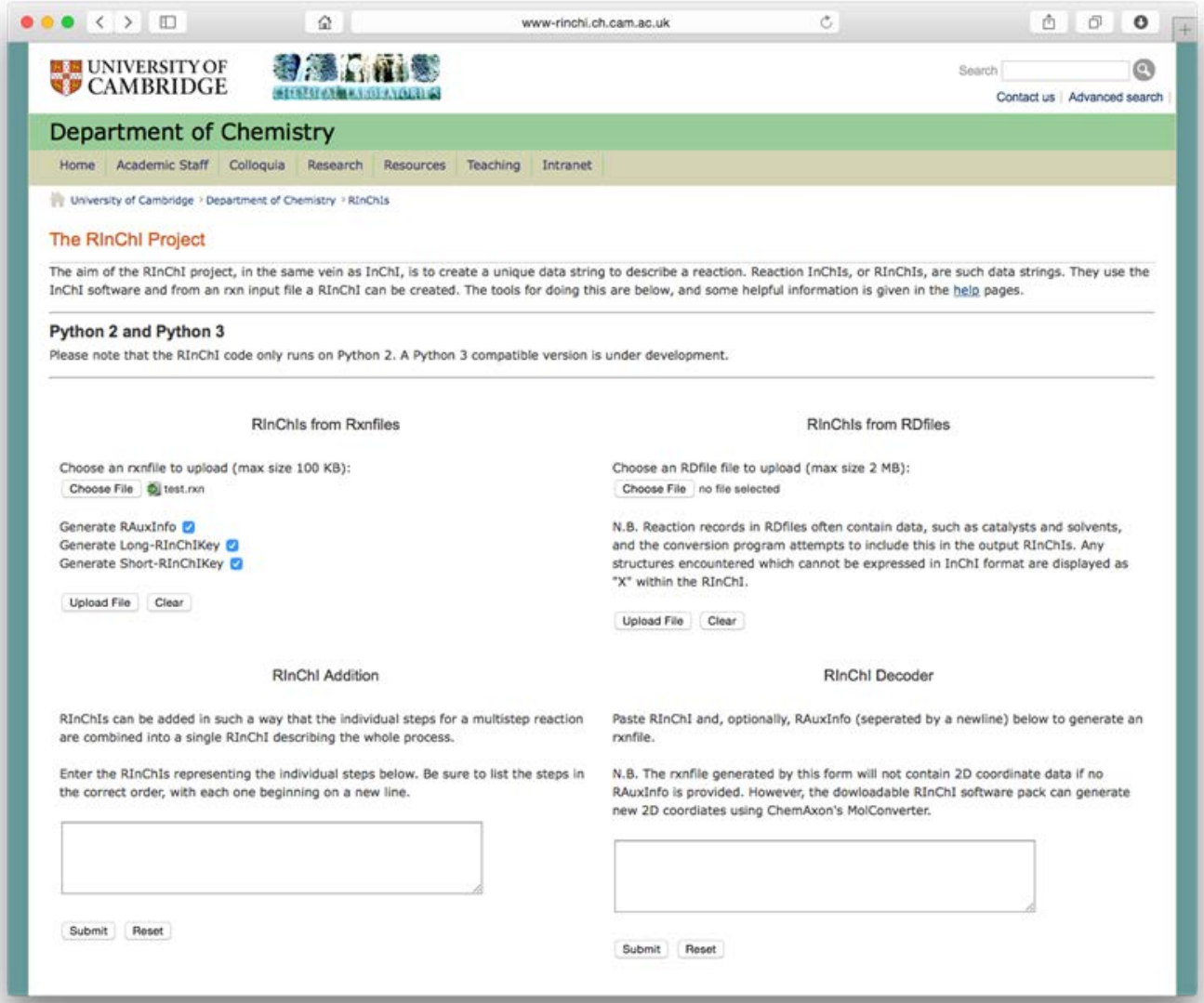


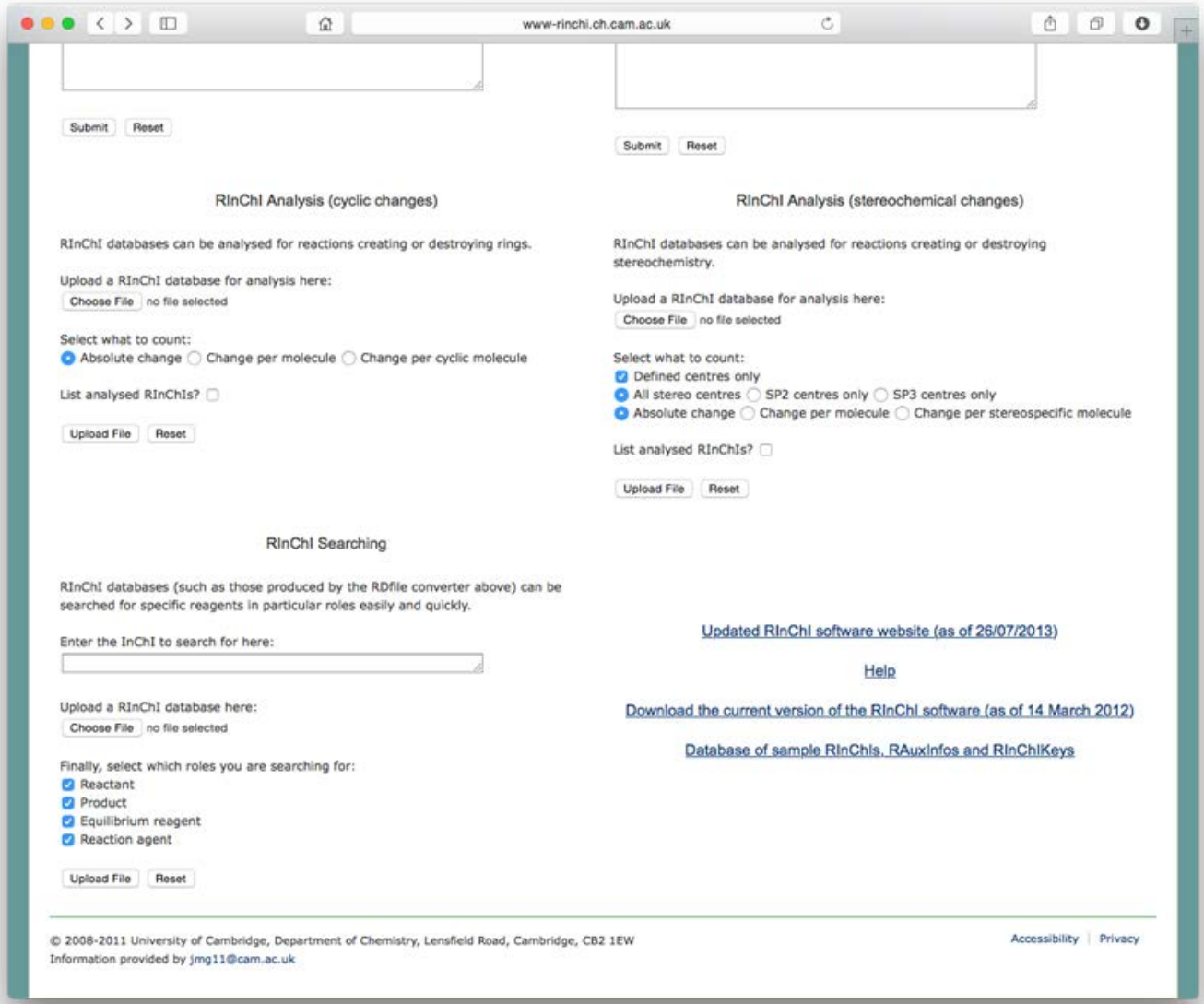
RInChI= version (RinChI.InChI)

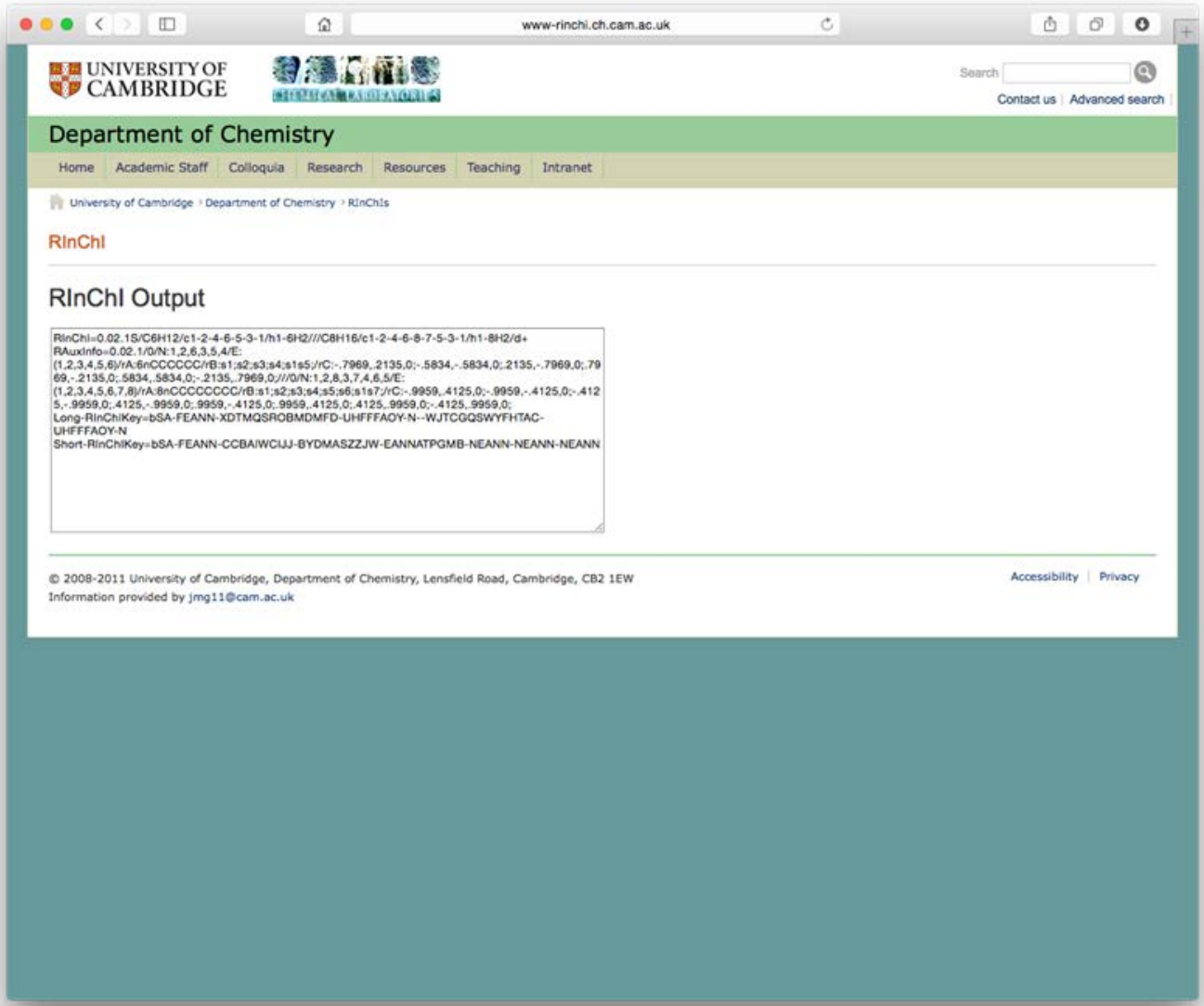
/ reactants: InChI#1 // InChI#2

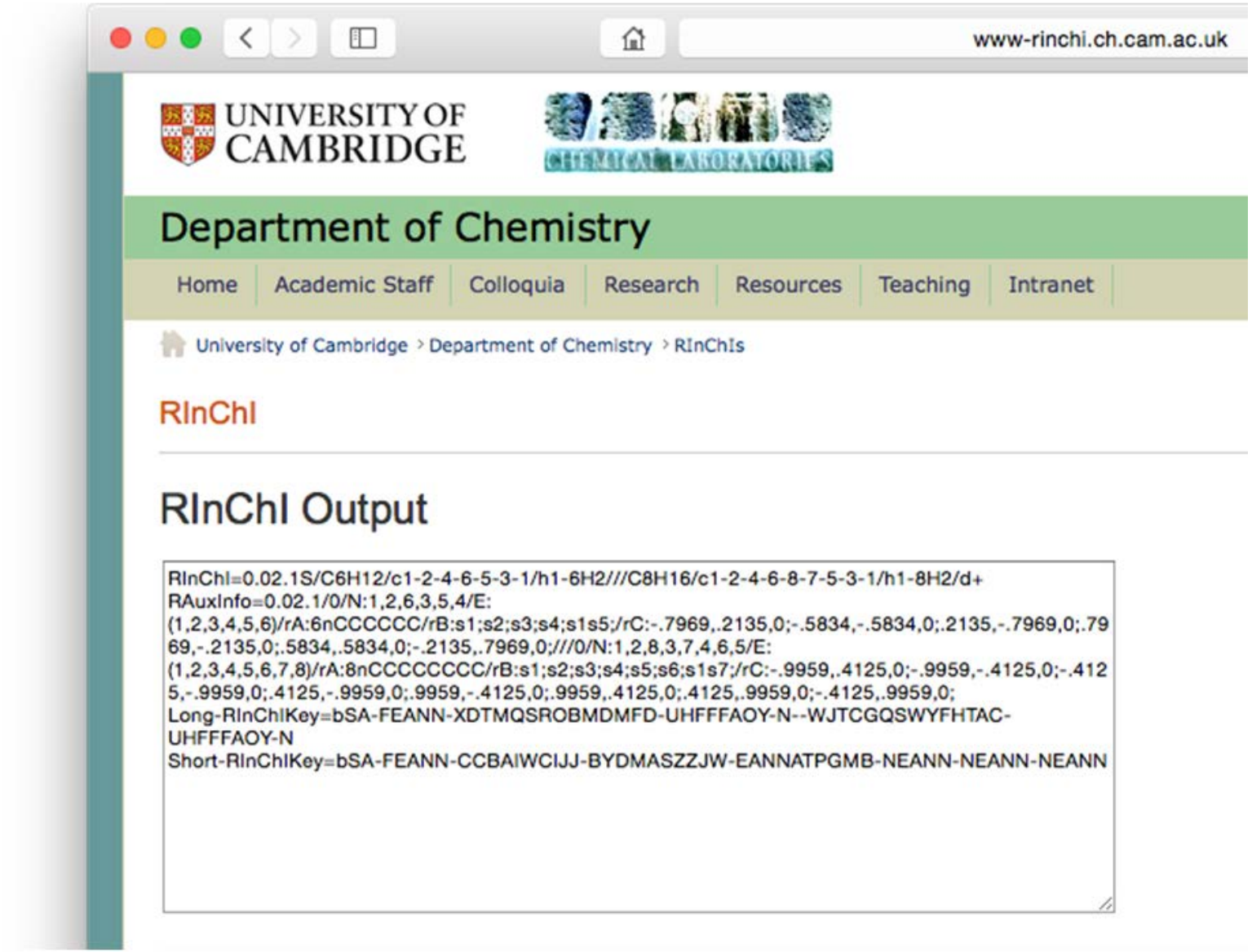
/// product: InChI#1 / reaction direction


RInChI
http://www-rinchi.ch.cam.ac.uk/
Google
UNIVERSITY OF CAMBRIDGE
800 YEARS 1209–2009
Search
Contact us | Advanced search
Department of Chemistry
Home | Academic Staff | Colloquia | Research | Resources | Teaching | Intranet
University of Cambridge › Department of Chemistry › RInChIs
The RInChI Project
The aim of the RInChI project, in the same vain as InChI, is to create a unique data string to describe a reaction. Reaction InChIs, or RInChIs, are such data strings. They use the InChI software and from an rxn input file a RInChI can be created. The tools for doing this are below, and some helpful information is given in the help pages.
Rinchi Generator
Choose a .rxn file to upload: Choose File no file selected
1. Inert atmosphere
2. High pressure Low pressure
3. Heated Chilled
4. UV light Darkness
5. Forward direction Reverse direction Equilibrium
Clear
Upload File
Rxn Generator
Paste RInChI and RAuxInfo (optional) here.
Submit
Reset
When AuxInfo not available 2D structure generation is done via MolConverter by ChemAxon
FreeWeb
http://www-rinchi.ch.cam.ac.uk/

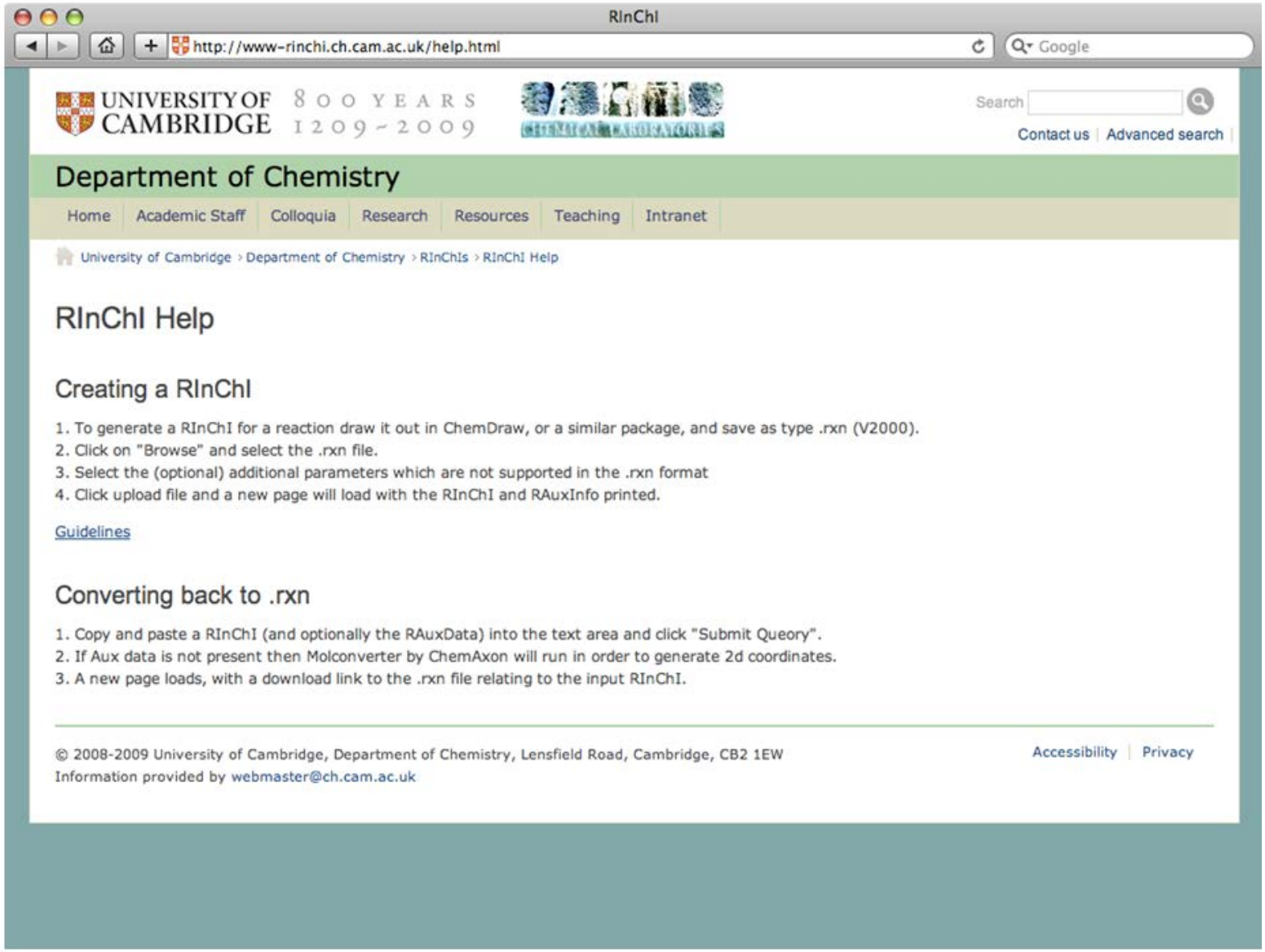

www-rinchi.ch.cam.ac.uk
UNIVERSITY OF CAMBRIDGE
Search
Contact us | Advanced search
Department of Chemistry
Home | Academic Staff | Colloquia | Research | Resources | Teaching | Intranet
University of Cambridge › Department of Chemistry › RInChIs
The RInChI Project
The aim of the RInChI project, in the same vein as InChI, is to create a unique data string to describe a reaction. Reaction InChIs, or RInChIs, are such data strings. They use the InChI software and from an rxn input file a RInChI can be created. The tools for doing this are below, and some helpful information is given in the help pages.
Python 2 and Python 3
Please note that the RInChI code only runs on Python 2. A Python 3 compatible version is under development.
RInChIs from Rxnfiles
Choose an rxnfile to upload (max size 100 KB):
Choose File test.rxn
Generate RAuxInfo
Generate Long-RInChIKey
Generate Short-RInChIKey
Upload File Clear
RInChIs from RDfiles
Choose an RDfile file to upload (max size 2 MB):
Choose File no file selected
N.B. Reaction records in RDfiles often contain data, such as catalysts and solvents, and the conversion program attempts to include this in the output RInChIs. Any structures encountered which cannot be expressed in InChI format are displayed as "X" within the RInChI.
Upload File Clear
RInChI Addition
RInChIs can be added in such a way that the individual steps for a multistep reaction are combined into a single RInChI describing the whole process.
Enter the RInChIs representing the individual steps below. Be sure to list the steps in the correct order, with each one beginning on a new line.
Submit Reset
RInChI Decoder
Paste RInChI and, optionally, RAuxInfo (seperated by a newline) below to generate an rxnfile.
N.B. The rxnfile generated by this form will not contain 2D coordinate data if no RAuxInfo is provided. However, the dowloadable RInChI software pack can generate new 2D coordiates using ChemAxon's MolConverter.
Submit Reset


www-rinchi.ch.cam.ac.uk
Submit
Reset
Submit
Reset
RInChI Analysis (cyclic changes)
RInChI databases can be analysed for reactions creating or destroying rings.
Upload a RInChI database for analysis here:
Choose File
no file selected
Select what to count:
Absolute change
Change per molecule
Change per cyclic molecule
List analysed RInChIs?
Upload File
Reset
RInChI Analysis (stereochemical changes)
RInChI databases can be analysed for reactions creating or destroying stereochemistry.
Upload a RInChI database for analysis here:
Choose File
no file selected
Select what to count:
Defined centres only
All stereo centres
SP2 centres only
SP3 centres only
Absolute change
Change per molecule
Change per stereospecific molecule
List analysed RInChIs?
Upload File
Reset
RInChI Searching
RInChI databases (such as those produced by the RDfile converter above) can be searched for specific reagents in particular roles easily and quickly.
Enter the InChI to search for here:
Upload a RInChI database here:
Choose File
no file selected
Finally, select which roles you are searching for:
Reactant
Product
Equilibrium reagent
Reaction agent
Upload File
Reset
Updated RInChI software website (as of 26/07/2013)
Help
Download the current version of the RInChI software (as of 14 March 2012)
Database of sample RInChIs, RAuxInfos and RInChIKeys
© 2008-2011 University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW
Information provided by jmg11@cam.ac.uk
Accessibility
Privacy

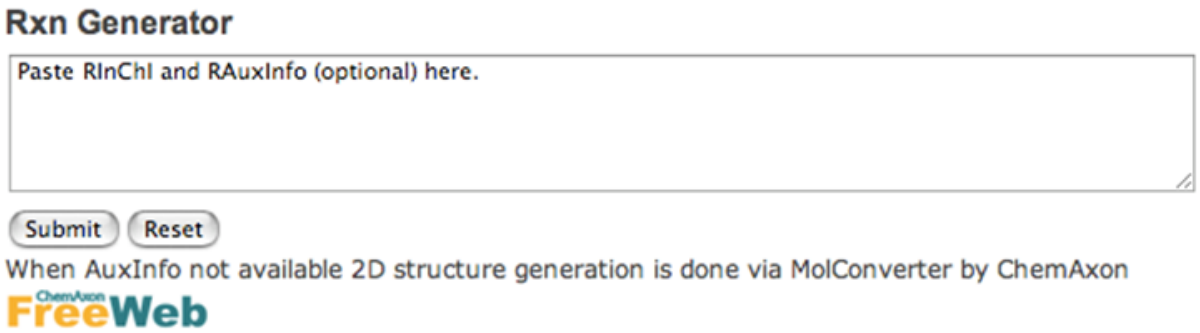

www-rinchi.ch.cam.ac.uk
UNIVERSITY OF CAMBRIDGE
Search
Contact us | Advanced search
Department of Chemistry
Home | Academic Staff | Colloquia | Research | Resources | Teaching | Intranet
University of Cambridge › Department of Chemistry › RInChIs
RInChI
RInChI Output
RInChI=0.02.1S/C6H12/c1-2-4-6-5-3-1/h1-6H2///C8H16/c1-2-4-6-8-7-5-3-1/h1-8H2/d+
RAuxInfo=0.02.1/0/N:1,2,6,3,5,4/E:
(1,2,3,4,5,6)/rA:6nCCCCCC/rB:s1;s2;s3;s4;s1s5;/rC:-.7969,.2135,0;-.5834,-.5834,0;.2135,-.7969,0;.7969,-.2135,0;.5834,.5834,0;-.2135,.7969,0;///0/N:1,2,8,3,7,4,6,5/E:
(1,2,3,4,5,6,7,8)/rA:8nCCCCCCCC/rB:s1;s2;s3;s4;s5;s6;s1s7;/rC:-.9959,.4125,0;-.9959,-.4125,0;-.4125,-.9959,0;.4125,-.9959,0;.9959,-.4125,0;.9959,.4125,0;.4125,.9959,0;-.4125,.9959,0;
Long-RInChIKey=bSA-FEANN-XDTMQSROBMDMFD-UHFFFAOY-N--WJTCGQSWYFHTAC-UHFFFAOY-N
Short-RInChIKey=bSA-FEANN-CCBAIWCIJJ-BYDMASZZJW-EANNATPGMB-NEANN-NEANN-NEANN
© 2008-2011 University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW
Information provided by jmg11@cam.ac.uk
Accessibility | Privacy


www-rinchi.ch.cam.ac.uk
UNIVERSITY OF CAMBRIDGE
CHEMICAL LABORATORIES
Department of Chemistry
Home
Academic Staff
Colloquia
Research
Resources
Teaching
Intranet
University of Cambridge › Department of Chemistry › RInChIs
RInChI
RInChI Output
RInChI=0.02.1S/C6H12/c1-2-4-6-5-3-1/h1-6H2///C8H16/c1-2-4-6-8-7-5-3-1/h1-8H2/d+
RAuxInfo=0.02.1/0/N:1,2,6,3,5,4/E:
(1,2,3,4,5,6)/rA:6nCCCCCC/rB:s1;s2;s3;s4;s1s5;/rC:-.7969,.2135,0;-.5834,-.5834,0;.2135,-.7969,0;.79
69,-.2135,0;.5834,.5834,0;-.2135,.7969,0;///0/N:1,2,8,3,7,4,6,5/E:
(1,2,3,4,5,6,7,8)/rA:8nCCCCCCCC/rB:s1;s2;s3;s4;s5;s6;s1s7;/rC:-.9959,.4125,0;-.9959,-.4125,0;-.412
5,-.9959,0;.4125,-.9959,0;.9959,-.4125,0;.9959,.4125,0;.4125,.9959,0;-.4125,.9959,0;
Long-RInChIKey=bSA-FEANN-XDTMQSROBMDMFD-UHFFFAOY-N--WJTCGQSWYFHTAC-
UHFFFAOY-N
Short-RInChIKey=bSA-FEANN-CCBAIWCIJJ-BYDMASZZJW-EANNATPGMB-NEANN-NEANN-NEANN

InChI=0.01.1S/2C2H4O2.Hg/c2*1-2(3)4;/h2*1H3,(H,3,4);/q;;+2/p-2//C4H6/c1-3-4-2/h1H,4H2,2H3//H2O/h1H2///C4H8O/c1-3-4(2)5/h3H2,1-2H3//p+1
InChI=0.01.1S/2C2H4O2.Hg/c2*1-2(3)4;/h2*1H3,(H,3,4);/q;;+2/p-2//BH4.Na/h1H4;/q-1;+1//C4H6/c1-3-4-2/h1H,4H2,2H3//H2O/h1H2///C4H10O/c1-3-4(2)5/h4-5H,3H2,1-2H3//p+1
InChI=0.01.1S/4O.Os//C8H12/c1-4-8(3)5-2/h4H,5H2,1-3H3/b6-4+///C8H14O2/c1-4-8(3,8)5(2)7/h5,7-8H,4H2,1-3H3/t5-,6-/m0/s1
InChI=0.01.1S/Al.Li.4H/q-1;+1;;;;//C8H10O/c1-3-5-6(7)4-2/h6-7H,4H2,1-2H3///C8H12O/c1-3-5-6(7)4-2/h3,5-7H,4H2,1-2H3/b5-3+
InChI=0.01.1S/BH2.Na/h1H2;/q+1;-1//C8H14O/c1-8(2)5-8(9)7(3)4/h5,7H,1-4H3//C8H18O3/c1-5-9-8(4,10-6-2)11-7-3/h5-7H2,1-4H3//p+1///C12H22O2/c1-6-14-11(13)9-12(4,5)8-7-10(2)3/h7-8,10H,6,9H2,1-5H3/b8-7+
InChI=0.01.1S/BH3/h1H3//C8H12/c1-4-6(3)5-2/h4H,5H2,1-3H3/b6-4+///C8H15B/c1-4-5(2)6(3)7/h5-6H,4,7H2,1-3H3/t5-,6+/m1/s1
InChI=0.01.1S/BH4/h1H4/q-1//C10H12O/c1-8(9(2)11)10-6-4-3-5-7-10/h3-8H,1-2H3/t8-/m1/s1//Na/q+1///C10H14O/c1-8(9(2)11)10-6-4-3-5-7-10/h3-9,11H,1-2H3/t8-,9+/m1/s1//C10H14O/c1-8(9(2)11)10-6-4-3-5-7-10/h3-9,11H,1-2H3/t8-,9-/m1/s1
InChI=0.01.1S/BH4/h1H4/q-1//C6H15N/c1-4-7(5-2)6-3/h4-6H2,1-3H3//C8H14O2/c1-5-6(2)8(10)3-7(5)4-9/h5-7,9H,3-4H2,1-2H3/t5-,6-,7+/m1/s1//CH3ClO2S/c1-5(2,3)4/h1H3//H3N/h1H3//Na/q+1///C8H15N/c1-5-6(2)8-3-7(5)4-9-8/h5-9H,3-4H2,1-2H3/t5-,6-,7?,8?/m1/s1
InChI=0.01.1S/Br2/c1-2//C3H7NO/c1-2-3(4)5/h2H2,1H3,(H2,4,5)//Na.H2O/h;1H2/q+1;/p-1///C2H7N/c1-2-3/h2-3H2,1H3
InChI=0.01.1S/Br2/c1-2//C4H8/c1-3-4-2/h3-4H,1-2H3/b4-3+///C4H8Br2/c1-3(5)4(2)6/h3-4H,1-2H3/t3-,4+
InChI=0.01.1S/Br2/c1-2//C4H8/c1-3-4-2/h3-4H,1-2H3/b4-3+//H2O/h1H2///C4H9BrO/c1-3(5)4(2)6/h3-4,6H,1-2H3/t3-,4+/m1/s1
InChI=0.01.1S/Br2/c1-2//C5H8/c1-3-5-4-2/h3H2,1-2H3///C5H8Br2/c1-3-5(7)4(2)6/h3H2,1-2H3/b5-4+
InChI=0.01.1S/Br2/c1-2//C6H12/c1-4-6(3)5-2/h4H,5H2,1-3H3/b6-4+/i4D///C6H12Br2/c1-4-6(3,8)5(2)7/h5H,4H2,1-3H3/t5-,6+/m1/s1/i5D
InChI=0.01.1S/Br2/c1-2//C9H8/c1-2-6-9-7-4-3-5-8-9/h3-5,7-8H,1H3///C9H8Br2/c1-7(10)9(11)8-5-3-2-4-6-8/h2-6H,1H3
InChI=0.01.1S/BrH/h1H//C2H4/c1-2/h1-2H2///C2H5Br/c1-2-3/h2H2,1H3
InChI=0.01.1S/BrH/h1H//C3H6/c1-3-2/h3H,1H2,2H3///C3H7Br/c1-3(2)4/h3H,1-2H3
InChI=0.01.1S/BrH/h1H//C8H12/c1-4-6(3)5-2/h4H,5H2,1-3H3/b6-4+/i4D///C6H13Br/c1-4-6(3,7)5-2/h4-5H2,1-3H3/i4D/t4-,6?/m0/s1
InChI=0.01.1S/C11H14FNO/c1-3-13(4-2)11(14)9-6-5-7-10(12)8-9/h5-8H,3-4H2,1-2H3//C4H9.Li/c1-3-4-2;/h1,3-4H2,2H3;//C7H6O/c8-6-7-4-2-1-3-5-7/h1-6H//C18H20FNO2/c1-3-20(4-2)18(22)14-11-8-12-15(19)16(14)17(21)13-9-6-5-7-10-13/h5-12,17,21H,3-4H2,1-2H3/t17-/m0/s1
InChI=0.01.1S/C12H14O2/c1-8(2)11-10(12(13)14-11)9-6-4-3-5-7-9/h3-8,10-11H,1-2H3//C8H14N/c1-5(2)7-6(3)4/h5-6H,1-4H3/q-1//C7H7Br/c8-6-7-4-2-1-3-5-7/h1-5H,6H2//Li/q+1///C19H20O2/c1-14(2)17-19(18(20)21-17,16-11-7-4-8-12-16)13-15-9-5-3-6-10-15/h3-12,14,17H,13H2,1-2H3/t17-,19+/m1/s1
InChI=0.01.1S/C12H22BClI/c14-13(11-7-3-1-4-8-11)12-9-5-2-6-10-12/h11-12H,1-10H2//C6H15N/c1-4-7(5-2)6-3/h4-6H2,1-3H3//C7H6O/c8-6-7-4-2-1-3-5-7/h1-6H//C9H16O/c1-2-9(10)8-6-4-3-5-7-8/h8H,2-7H2,1H3///C16H22O2/c1-12(15(17)13-8-4-2-5-9-13)16(18)14-10-6-3-7-11-14/h2,4-5,8-9,12,14-15,17H,3,6-7,10-11H2,1H3/t12-,15+/m1/s1
InChI=0.01.1S/C16H24O5SSi/c1-11-15(23(4,5)13-9-7-6-8-10-13)14(20-16(11)17)12(2)21-22(3,18)19/h6-12,14-15H,1-5H3/t11-,12+,14?,15+/m0/s1//N3/c1-3-2/q-1///C15H21N3O2Si/c1-10-14(13(20-15(10)19)11(2)17-18-16)21(3,4)12-8-6-5-7-9-12/h5-11,13-14H,1-4H3/t10-,11-,13?,14+/m0/s1
InChI=0.01.1S/C18H15P/c1-4-10-16(11-5-1)19(17-12-6-2-7-13-17)18-14-8-3-9-15-18/h1-15H//C2H6O/c1-2-3/h3H,2H2,1H3//CCl4/c2-1(3,4)5///C2H5Cl/c1-2-3/h2H2,1H3
InChI=0.01.1S/C21H34BF3O3S/c1-11-15-7-13(19(15,3)4)9-17(11)22(28-29(26,27)21(23,24)25)18-10-14-8-16(12(18)2)20(14,5)6/h11-18H,7-10H2,1-6H3//C7H6O/c8-6-7-4-2-1-3-5-7/h1-6H//C8H19N/c1-6-9(7(2)3)8(4)5/h7-8H,6H2,1-5H3//C9H10O/c1-2-9(10)8-6-4-3-5-7-8/h3-7H,2H2,1H3//H2O2/c1-2/h1-2H///C16H16O2.CH4/c1-12(15(17)13-8-4-2-5-9-13)16(18)14-10-6-3-7-11-14;/h2-12,15,17H,1H3;1H4
InChI=0.01.1S/C2H2/c1-2/h1-2H//OO2/c2-1-3//H2N.Na/h1H2;/q-1;+1///C3H2O2/c1-2-3(4)5/h1H,(H,4,5)/p-1
InChI=0.01.1S/C2H4O3/c1-2(3)5-4/h4H,1H3//C5H8/c1-2-4-5-3-1/h1-2H,3-5H2///C5H8O/c1-2-4-5(3-1)6-4/h4-5H,1-3H2
InChI=0.01.1S/C2H4O3/c1-2(3)5-4/h4H,1H3//C5H8O/c6-5-3-1-2-4-5/h1-4H2///C5H8O2/c6-5-3-1-2-4-7-5/h1-4H2
InChI=0.01.1S/C2H5.BrH.Mg/c1-2;;/h1H2,2H3;1H;/q;;+1/p-1//CO2/c2-1-3//C3H6O2/c1-2-3(4)5/h2H2,1H3,(H,4,5)
InChI=0.01.1S/C2H6S.BrH.Cu/c1-3-2;;/h1-2H3;1H;/q;;+1/p-1//C4H7.BrH.Mg/c1-3-4-2;;/h3H,1-2,4H2;1H;/q;;+1/p-1//C8H8O3/c1-8-3-2-5(9)6(4-8)11-7(8)10/h2-3,6H,4H2,1H3/t6-,8-/m0/s1///C12H16O3/c1-3-4-5-8-6-9(13)10-7-12(8,2)11(14)15-10/h3,8,10H,1,4-7H2,2H3/t8-,10+,12-/m1/s1
InChI=0.01.1S/C2H6S/c1-3-2/h1-2H3//C8H12/c1-3-5-6-4-2/h5-6H,3-4H2,1-2H3/b6-5+///C8H14N/c1-5(2)7-6(3)4/h5-6H,1-4H3/q-1//Li/q+1///C2H4O/c1-2-3/h2H,1H3
InChI=0.01.1S/C3H6O/c1-3(2)4/h1-2H3//C6H10O/c1-5(2)4-6(3)7/h4H,1-3H3///p+1
InChI=0.01.1S/C4H10O/c1-4(2,3)5/h5H,1-3H3//C4H8/c1-3-4-2/h3-4H,1-2H3/b4-3+//C5H11NO2/c1-6(7)2-4-8-5-3-6/h2-5H2,1H3//H2O/h1H2///C4H10O2/c1-3(5)4(2)6/h3-6H,1-2H3//4O.Os
InChI=0.01.1S/C4H10O/c1-4(2,3)5/h5H,1-3H3//C5H8/c1-3-5-4-2/h3H2,1-2H3//Na.H//C4H9O/c1-4(2,3)5/h1-3H3/q-1//C5H10/c1-3-5-4-2/h3,5H,4H2,1-2H3/b5-3+///H3N/h1H3
InChI=0.01.1S/C4H2/c1-3-4-2/h1-2H//CH3O/c1-2/h1H3/q-1//CH4O/c1-2/h2H,1H3///C5H6O/c1-3-4-5-6-2/h1,4-5H,2H3/b5-4-
InChI=0.01.1S/C4H7Br/c1-2-3-4-5/h2-3H,4H2,1H3//C4H9.Li/c1-3-4-2;/h1,3-4H2,2H3;//C7H5ClO/c8-7(9)6-4-2-1-3-5-6/h1-5H///C15H19NO/c1-4-5-8-12-9-6-7-10-13(12)14-16-15(2,3)11-17-14/h4-7,9-10H,8,11H2,1-3H3
InChI=0.01.1S/C4H8O/c1-3-4(2)5-3/h3-4H,1-2H3/t3-,4?/m0/s1//CH3.BrH.Mg/h1H3;1H;/q;;+1/p-1//ClH.Cu/h1H;/q;;+1/p-1///C5H12O/c1-4(2)5(3)6/h4-6H,1-3H3
InChI=0.01.1S/C4H9BrO/c1-3(5)4(2)6/h3-4,6H,1-2H3/t3-,4+/m1/s1//Na.H2O/h;1H2/q+1;/p-1//C4H8O/c1-3-4(2)5-3/h3-4H,1-2H3/t3-,4?/m0/s1
InChI=0.01.1S/C4H9NO/c1-3-4(2)5-6/h6H,3H2,1-2H3/b5-4+//p+1///C4H9NO/c1-3-5-4(2)6/h3H2,1-2H3,(H,5,6)
InChI=0.01.1S/C4H9O.K/c1-4(2,3)5/h1-3H3;/q-1;+1///C7H13Cl/c1-7(8)5-3-2-4-6-7/h2-6H2,1H3///C7H12/c1-7-5-3-2-4-6-7/h1-6H2
InChI=0.01.1S/C4H9O/c1-4(2,3)5/h1-3H3/q-1//C7H6O/c8-6-7-4-2-1-3-5-7/h1-6H//C7H7I/c8-6-7-4-2-1-3-5-7/h1-5H,6H2//K/q+1///C14H12O/c1-3-7-11(8-4-1)13-14(15-13)12-9-5-2-6-10-12/h1-10,13-14H/t13-,14-/m1/s1///C20H20O4S/c1-3-7-13(8-4-1)19-21-11-15-17(23-19)18-16(25-15)12-22-20(24-18)14-9-5-2-6-10-14/h1-10,15-20H,11-12H2/t15-,16-,17+,18?,19+,20+/m1/s1
InChI=0.01.1S/C5H10O2/c1-3-5(6)7-4-2/h3-4H2,1-2H3/CH3O.Na/c1-2;/h1H3;/q-1;+1//p+1///C8H14O3/c1-4-7(9)6(3)8(10)11-5-2/h6H,4-5H2,1-3H3
InChI=0.01.1S/C5H8/c1-3-5-4-2/h3H2,1-2H3//H2/h1H//C5H10/c1-3-5-4-2/h3,5H,4H2,1-2H3/b5-3-
InChI=0.01.1S/C6H10O/c7-6-4-2-1-3-5-6/h2,4,6-7H,1,3,5H2//C7H5ClO3/c8-6-3-1-2-5(4-6)7(9)11-10/h1-4,10H///C8H12O3/c1-5(9)10-6-3-2-4-7-8(6)11-7/h6-8H,2-4H2,1H3/t6-,7?,8?/m1/s1
InChI=0.01.1S/C6H10O3/c7-4-1-2-6-8-3-5(4)9-6/h4-7H,1-3H2///C6H10O3/c7-4-3-8-6-2-1-5(4)9-6/h4-7H,1-3H2///p+1
InChI=0.01.1S/C6H12/c1-4-6(3)5-2/h4H,5H2,1-3H3/b6-4+//C7H5ClO3/c8-6-3-1-2-5(4-6)7(9)11-10/h1-4,10H///C8H12O/c1-4-6(3)5(2)7-6/h5H,4H2,1-3H3/t5-,6-/m0/s1
InChI=0.01.1S/C6H12/c1-4-6(3)5-2/h4H,5H2,1-3H3/b6-4+//H2O/h1H2///C6H14O/c1-4-6(3,7)5-2/h7H,4-5H2,1-3H3/p+1
InChI=0.01.1S/C6H12O/c1-4-6(3)5(2)7-6/h5H,4H2,1-3H3/t5-,6-/m0/s1//H2O/h1H2/p-1///C6H14O2/c1-4-6(3,8)5(2)7/h5,7-8H,4H2,1-3H3/t5-,6+/m1/s1
InChI=0.01.1S/C6H15B/c1-4-5(2)6(3)7/h5-6H,4,7H2,1-3H3/t5-,6+/m1/s1//H2O/h1H2/p-1///H2O2/c1-2/h1-2H///C6H14O/c1-4-5(2)6(3)7/h5-7H,4H2,1-3H3/t5-,6+/m1/s1
InChI=0.01.1S/C6H5Br/c7-6-4-2-1-3-5-6/h1-5H//H2N.Na/h1H2;/q-1;+1///C6H7N/c7-6-4-2-1-3-5-6/h1-5H,7H2
InChI=0.01.1S/C7H14O/c1-7(8)5-3-2-4-6-7/h8H,2-6H2,1H3///C7H12/c1-7-5-3-2-4-6-7/h5H,2-4,6H2,1H3///p+1
InChI=0.01.1S/C7H5ClO3/c8-6-3-1-2-5(4-6)7(9)11-10/h1-4,10H//C8H12O2/c1-7(9)10-8-5-3-2-4-6-8/h3,5,8H,2,4,6H2,1H3//C8H12O3/c1-5(9)10-6-3-2-4-7-8(6)11-7/h6-8H,2-4H2,1H3/t6-,7?,8?/m1/s1
InChI=0.01.1S/C7H6O/c8-6-7-4-2-1-3-5-7/h1-6H//C8H19N/c1-6-9(7(2)3)8(4)5/h7-8H,6H2,1-5H3//C9H16O/c1-2-9(10)8-6-4-3-5-7-8/h8H,2-7H2,1H3//C9H18BF3O3S/c1-3-5-7-10(8-6-4-2)16-17(14,15)9(11,12)13/h3-8H2,1-2H3///C16H22O2/c1-12(15(17)13-8-4-2-5-9-13)16(18)14-10-6-3-7-11-14/h2,4-5,8-9,12,14-15,17H,3,6-7,10-11H2,1H3/t12-,15-/m1/s1
InChI=0.01.1S/C8H17Br/c1-3-4-5-6-7-8(2)9/h8H,3-7H2,1-2H3/t8-/m0/s1//H2O/h1H2/p-1///C8H18O/c1-3-4-5-6-7-8(2)9/h8-9H,3-7H2,1-2H3/t8-/m1/s1


RInChI
http://www-rinchi.ch.cam.ac.uk/help.html
Google
UNIVERSITY OF CAMBRIDGE
800 YEARS 1209-2009
CHEMICAL LABORATORIES
Search
Contact us | Advanced search
Department of Chemistry
Home
Academic Staff
Colloquia
Research
Resources
Teaching
Intranet
University of Cambridge › Department of Chemistry › RInChIs › RInChI Help
RInChI Help
Creating a RInChI
1. To generate a RInChI for a reaction draw it out in ChemDraw, or a similar package, and save as type .rxn (V2000).
2. Click on "Browse" and select the .rxn file.
3. Select the (optional) additional parameters which are not supported in the .rxn format
4. Click upload file and a new page will load with the RInChI and RAuxInfo printed.
Guidelines
Converting back to .rxn
1. Copy and paste a RInChI (and optionally the RAuxData) into the text area and click "Submit Queory".
2. If Aux data is not present then Molconverter by ChemAxon will run in order to generate 2d coordinates.
3. A new page loads, with a download link to the .rxn file relating to the input RInChI.
© 2008-2009 University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW
Information provided by webmaster@ch.cam.ac.uk
Accessibility | Privacy

Reading, Writing and Arithmetic



RInChI can be converted into .rxn files.

2D geometries may be in the Auxiliary Information, or else generated by ChemAxon's MolConverter

Reading, **Writing** and Arithmetic

1. ○ Inert atmosphere
2. ○ High pressure ○ Low pressure
3. ○ Heated ○ Chilled
4. ○ UV light ○ Darkness
5. ◉ Forward direction ○ Reverse direction ○ Equilibrium

RInChI are generated from .rxn files.

Some additional information is added as extra layers

Too much information is bad

- different RInChI for closely related processes

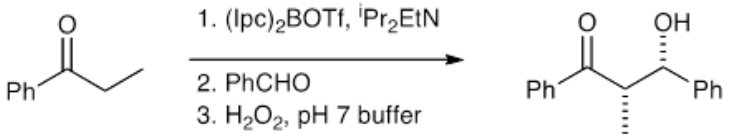
How much is too much? Is this list the right list?

Reading, Writing and **Arithmetic**

Rinchi Addition

Paste RInChIs (and RAuxInfo) here. Do not include additional information.
eg:
RInChI 1
(RAuxInfo 1)
RInChI 2

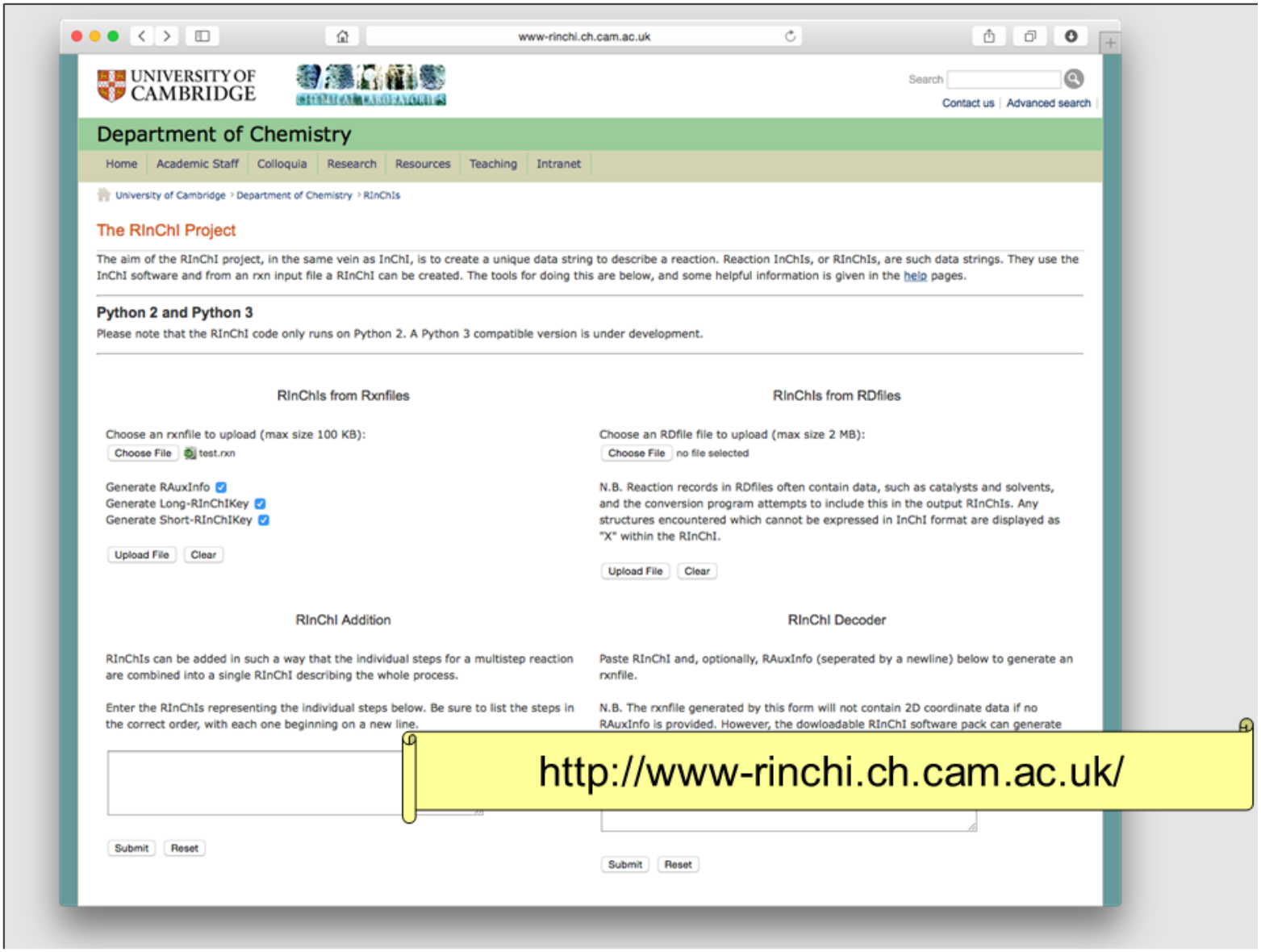
Submit Reset



One RInChI or Three RInChI ?

The three separate RInChIs for each step can be added up to form a single RInChI for the whole process.

If some authors consider a reaction to be one step, and others the same reaction to be two steps, they should have the same RInChI for the whole transformation


www-rinchi.ch.cam.ac.uk
UNIVERSITY OF CAMBRIDGE
Search
Contact us | Advanced search
Department of Chemistry
Home | Academic Staff | Colloquia | Research | Resources | Teaching | Intranet
University of Cambridge › Department of Chemistry › RInChIs
The RInChI Project
The aim of the RInChI project, in the same vein as InChI, is to create a unique data string to describe a reaction. Reaction InChIs, or RInChIs, are such data strings. They use the InChI software and from an rxn input file a RInChI can be created. The tools for doing this are below, and some helpful information is given in the help pages.
Python 2 and Python 3
Please note that the RInChI code only runs on Python 2. A Python 3 compatible version is under development.
RInChIs from Rxnfiles
Choose an rxnfile to upload (max size 100 KB):
Choose File test.rxn
Generate RAuxInfo
Generate Long-RInChIKey
Generate Short-RInChIKey
Upload File Clear
RInChIs from RDfiles
Choose an RDfile file to upload (max size 2 MB):
Choose File no file selected
N.B. Reaction records in RDfiles often contain data, such as catalysts and solvents, and the conversion program attempts to include this in the output RInChIs. Any structures encountered which cannot be expressed in InChI format are displayed as "X" within the RInChI.
Upload File Clear
RInChI Addition
RInChIs can be added in such a way that the individual steps for a multistep reaction are combined into a single RInChI describing the whole process.
Enter the RInChIs representing the individual steps below. Be sure to list the steps in the correct order, with each one beginning on a new line.
Submit Reset
RInChI Decoder
Paste RInChI and, optionally, RAuxInfo (seperated by a newline) below to generate an rxnfile.
N.B. The rxnfile generated by this form will not contain 2D coordinate data if no RAuxInfo is provided. However, the dowloadable RInChI software pack can generate
Submit Reset
http://www-rinchi.ch.cam.ac.uk/

RInChI

http://www-rinchi.ch.cam.ac.uk/database.html

ICXJVZHDZFXYQC-MSLOCZ

UNIVERSITY OF CAMBRIDGE 800 YEARS 1209–2009

CHEMICAL LABORATORIES

Search

Contact us | Advanced search

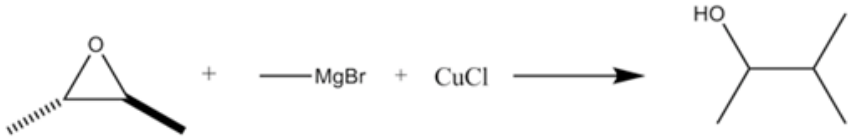
Department of Chemistry

Home | Academic Staff | Colloquia | Research | Resources | Teaching | Intranet

University of Cambridge › Department of Chemistry › RInChIs

Example RInChIs

RInChI=0.01.1S/C4H9BrO/c1-3(5)4(2)6/h3-4,6H,1-2H3/t3-,4+/m1/s1//Na.H2O/h;1H2/q+1;/p-1///C4H8O/c1-3-4(2)5-3/h3-4H,1-2H3/t3-,4?/m0/s1
RAuxInfo=0.01.1S/0/N:4,1,3,2,6,5/it:im/rA:6nCCCCOBr/rB:s1;s2;s3;N2;P3;/rC:-.825,-.7557,0;-.4125,-.0412,0;.4125,-.0412,0;.825,.6733,0;-.626,.7557,0;.825,-.7557,0;//1/N:1;2/rA:2nNaO/rB:s1;/rC:-.4125,0,0;.4125,0,0;///0/N:4,1,3,2,5/E:(1,2)(3,4)/it:im/rA:5nCCCCO/rB:N1;s2;P3;s2s3;/rC:-1.127,-.5635,0;-.4125,-.151,0;.4125,-.151,0;1.127,-.5635,0;0,.5635,0;



RInChI=0.01.1S/C4H8O/c1-3-4(2)5-3/h3-4H,1-2H3/t3-,4?/m0/s1//CH3.BrH.Mg/h1H3;1H;/q;;+1/p-1//ClH.Cu/h1H;/q;+1/p-1///C5H12O/c1-4(2)5(3)6/h4-6H,1-3H3
RAuxInfo=0.01.1S/0/N:4,1,3,2,5/E:(1,2)(3,4)/it:im/rA:5nCCCCO/rB:N1;s2;P3;s2s3;/rC:-1.127,-.5635,0;-.4125,-.151,0;.4125,-.151,0;1.127,-.5635,0;0,.5635,0;//1/N:1;3;2/CRV:1.3;;/rA:3nCMgBr/rB:s1;s2;/rC:-.6188,-

InChIKey

- http://www.iupac.org/inchi/release102.html
- A fixed-length (25-character) condensed digital representation of an InChI
- There is a finite, but very small probability of finding two structures with the same InChIKey … equivalent to a single collision in one of 75 databases of a billion compounds each
- Caffeine:
  - InChIKey=RYYVLZVUVIJVGH-UHFFFAOYAW

InChIKey

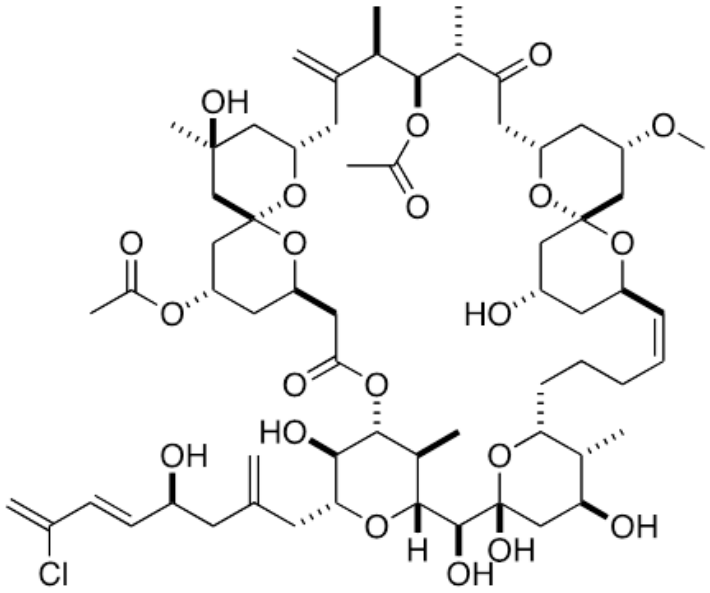
- http://en.wikipedia.org/wiki/International_Chemical_Identifier
  - With all databases currently having below 50 million structures, such duplication appears unlikely at present.

- http://inchi.info/inchikey_overview_en.html
  - At the time or writing of this article no such collisions are known, but they are unavoidable in the future. On the other hand it is possible that the first collision will not be found in the next 100 or 1000 years.

- http://depth-first.com/articles/2009/04/24/the-first-inchikey-collision
  - Those are pretty slim odds indeed. But they're not zero, either. As far as I know, not one InChI Key collision has been reported to date. But that doesn't mean one won't be found later on today. *(April 2009)*

Spongistatin 1

(or altohyrtin or cinachyrolide)



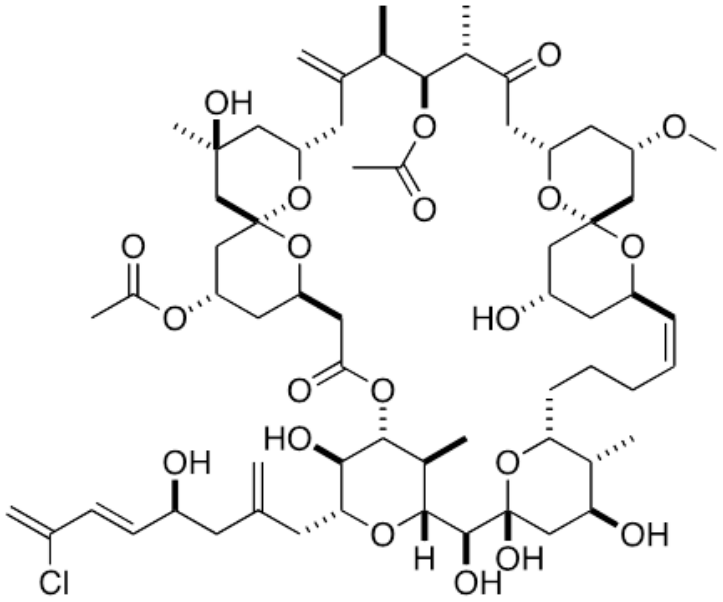
A unique family of antimitotic macrolides …display exceptional potency against a wide variety of human cancer cell lines... Initial discrepancies in the configurational assignments were resolved in 1997 by the first total synthesis of altohyrtin C by the Evans group, and soon after altohyrtin A by the Kishi group

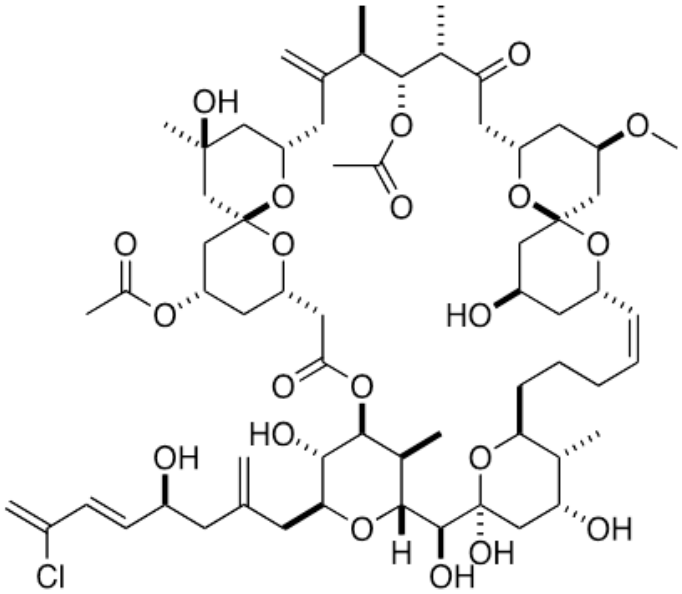
Paterson, *et al.*
Angew. Chem. Int. Ed. 2001, **40**, 4055

InChIKey=ICXJVZHDZFXYQC-MSLOCZROBI

InChI=1/C63H95ClO21/c1-33(19-42(67)18-17-35(3)64)20-53-55(72)57-39(7)58(79-53)59(73)63(75)31-51(70)37(5)52(85-63)16-14-12-13-15-44-22-43(68)27-61(81-44)29-47(76-11)23-45(82-61)25-50(69)38(6)56(78-41(9)66)36(4)34(2)21-49-28-60(10,74)32-62(84-49)30-48(77-40(8)65)24-46(83-62)26-54(71)80-57/h13,15,17-18,36-39,42-49,51-53,55-59,67-68,70,72-75H,1-3,12,14,16,19-32H2,4-11H3/b15-13-,18-17+/t36-,37-,38-,39+,42+,43+,44-,45-,46+,47+,48+,49+,51+,52+,53+,55+,56-,57-,58-,59+,60+,61?,62?,63-/m1/s1

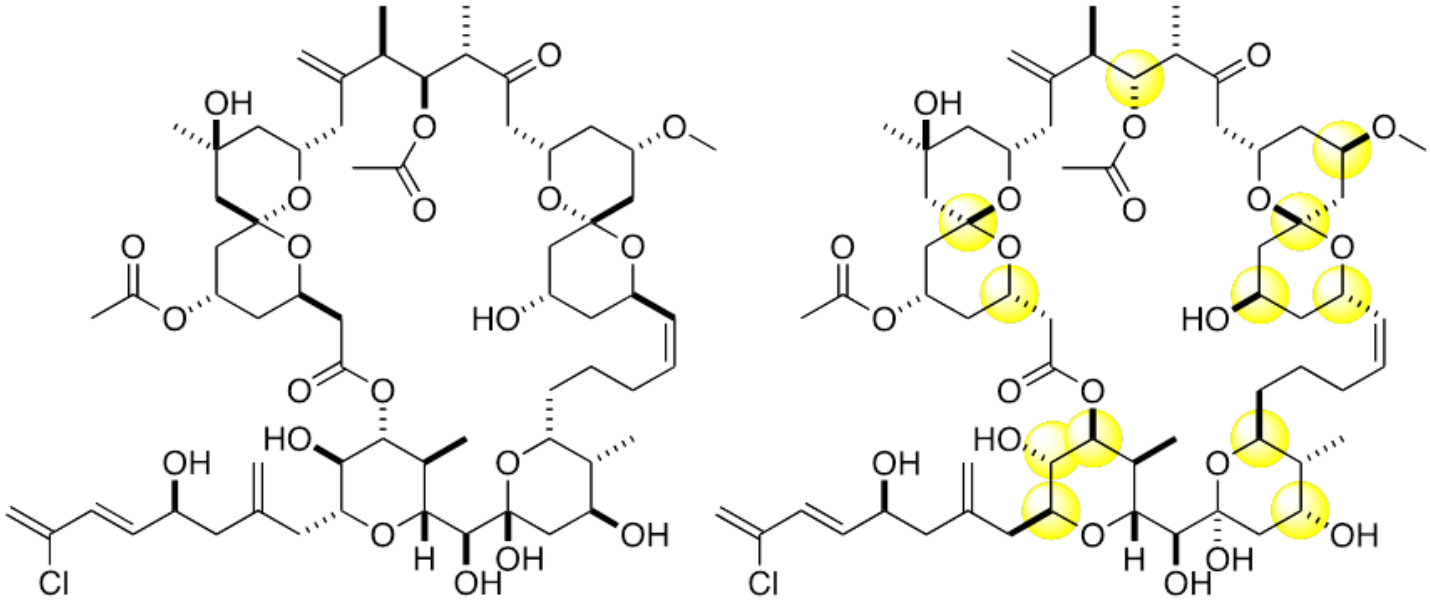
Spongistatin 1





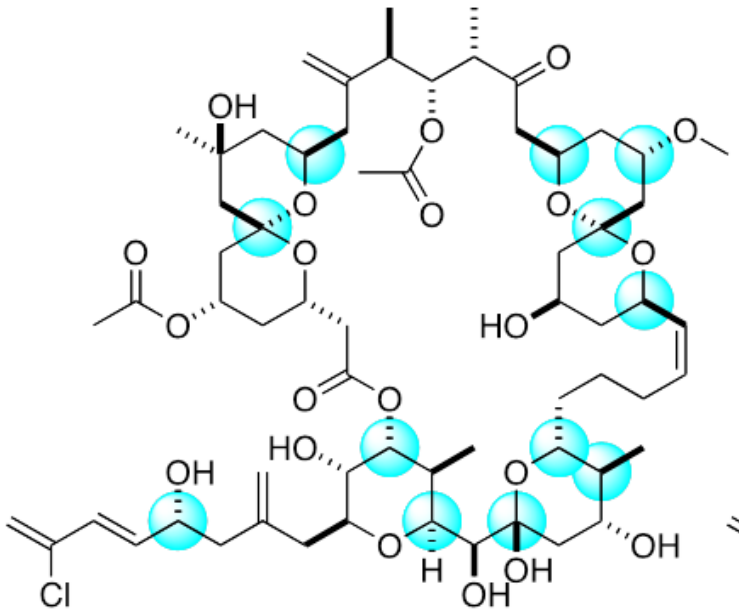
Not spongistatin

Spongistatin 1

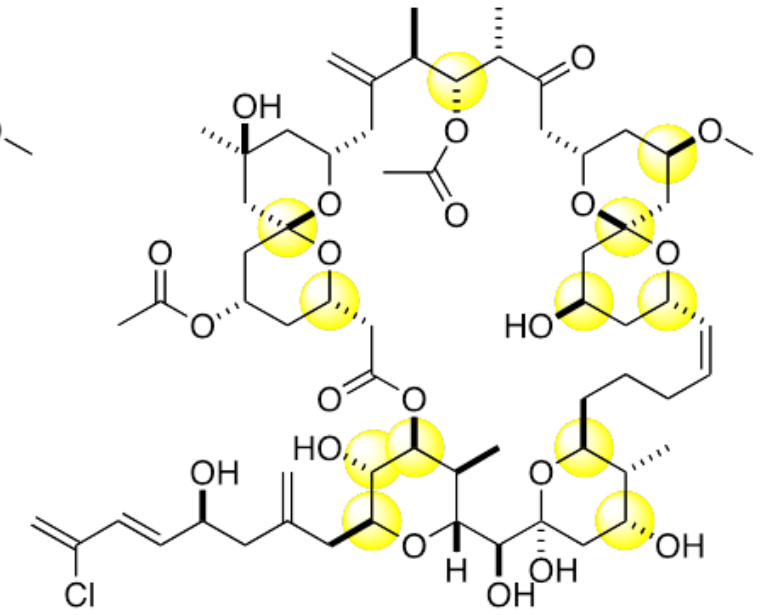


Not spongistatin

Spongistatin 1

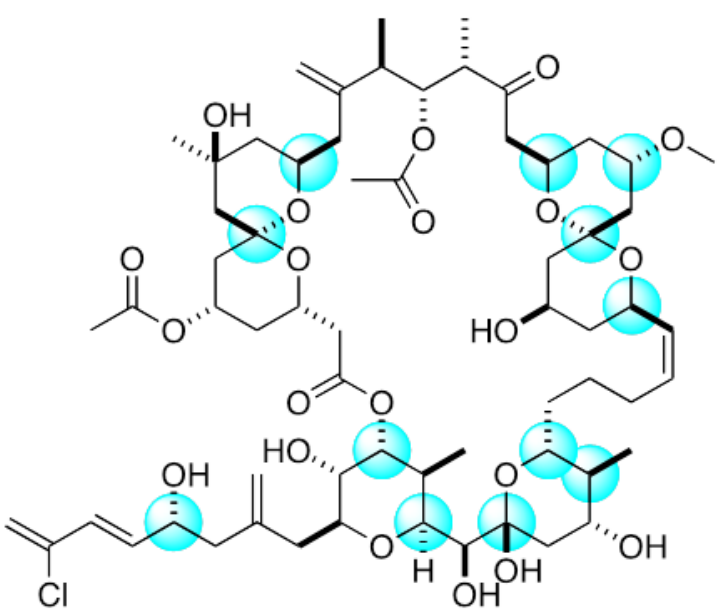


Not spongistatin

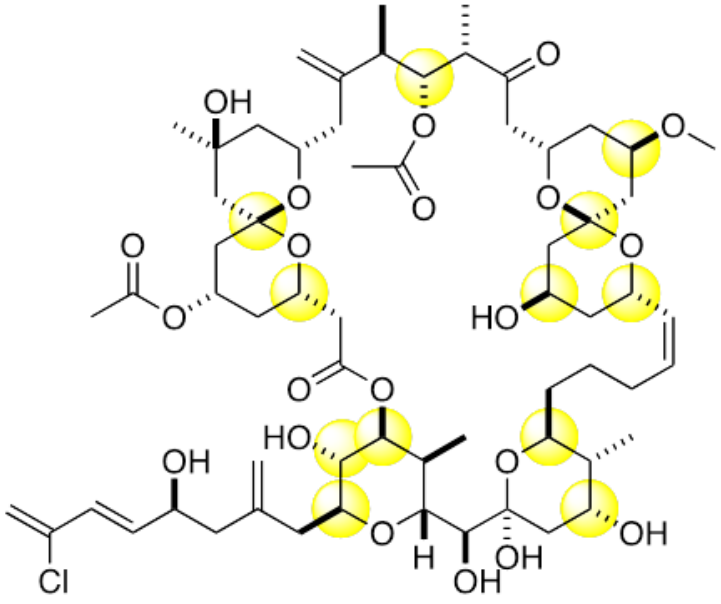


Not spongistatin

Spongistatin 1



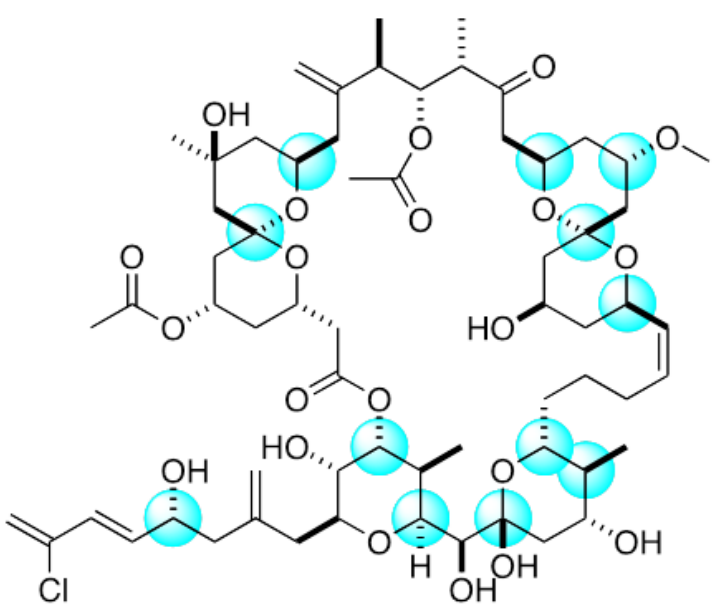
Not spongistatin



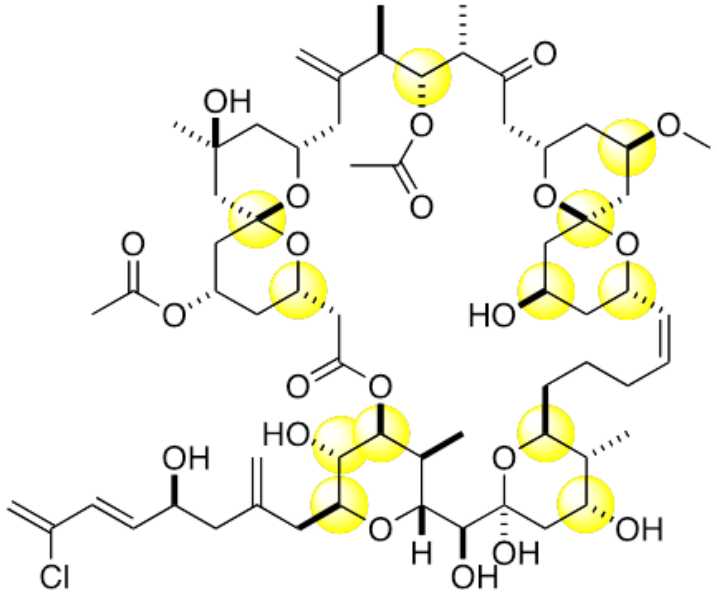
Not spongistatin either

InChIKey=ICXJVZHDZFXYQC-AAAYXFPSBK

Spongistatin 1



Not spongistatin

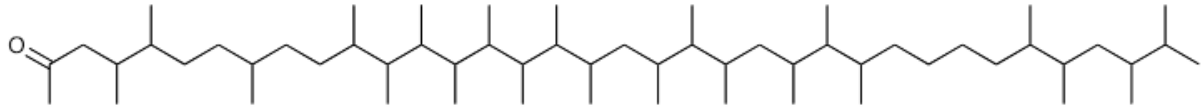


Not spongistatin either

InChIKey=ICXJVZHDZFXYQC-AAAYXFPSBK

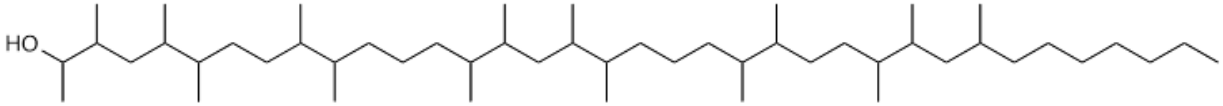
InChIKey=ICXJVZHDZFXYQC-AAAYXFPSBK

4,5,8,11,12,13,14,15,16,17,18,20,21,22,24,25,26,31,32,34,35-henicosamethylhexatriacontan-2-one



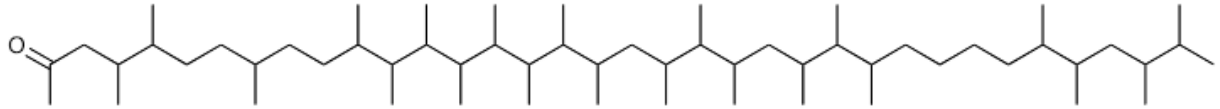
Chemical Formula: $C_{57}H_{114}O$
Exact Mass: 814.89
Molecular Weight: 815.51m/z: 814.89 (100.0%), 815.89 (63.0%), 816.89 (18.9%), 817.90 (4.0%)
Elemental Analysis: C, 83.95; H, 14.09; O, 1.96

3,5,6,9,10,14,15,17,18,22,23,26,27,29-tetradecamethylhexatriacontan-2-ol



Chemical Formula: $C_{50}H_{102}O$
Exact Mass: 718.79
Molecular Weight: 719.34m/z: 718.79 (100.0%), 719.80 (55.3%), 720.80 (15.2%), 721.80 (2.6%)
Elemental Analysis: C, 83.48; H, 14.29; O, 2.22

4,5,8,11,12,13,14,15,16,17,18,20,21,22,24,25,26,31,32,34,35-henicosamethylhexatriacontan-2-one

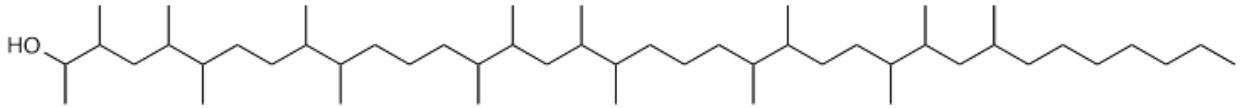


Chemical Formula: $C_{57}H_{114}O$
Exact Mass: 814.89
Molecular Weight: 815.51m/z: 814.89 (100.0%), 815.89 (63.0%), 816.89 (18.9%), 817.90 (4.0%)
Elemental Analysis: C, 83.95; H, 14.09; O, 1.96

InChI Key

OCPAUTFLLNMYSX-UHFFFAOYSA-N

3,5,6,9,10,14,15,17,18,22,23,26,27,29-tetradecamethylhexatriacontan-2-ol



Chemical Formula: $C_{50}H_{102}O$
Exact Mass: 718.79
Molecular Weight: 719.34m/z: 718.79 (100.0%), 719.80 (55.3%), 720.80 (15.2%), 721.80 (2.6%)
Elemental Analysis: C, 83.48; H, 14.29; O, 2.22

Testing RInChI

- Lots of reaction information is omitted:
  - Mechanism
  - Temperatures
  - Yields
  - Quantities
- Different people doing the same reaction should generate the same RInChI without consulting each other
  - Not a registration service
  - Can generate a RInChI without knowing much about the reaction
- Barrier to creating RInChIs should be low
- Misleading RInChIs will inevitably be developed

Testing RInChI

- .rdf files from three companies have been translated into RInChIs and combined
  - This only took a few minutes for about 3 000 reactions
- The RInChI files is 3 % of the size of the .rdf files
- Contains much of the useful information
- The file of RInChIs can easily be manipulated with unix commands
  - Files from different companies readily joined together
  - Find duplicates (sort | uniq -c)
  - Same starting materials, different products
  - Same products, different starting materials
  - Product of one reaction is a starting material for another
  - How many different molecules involved?

Testing RInChI

- Current work:
  - Easy questions, without processing InChI
    - What atoms are different in starting materials and products?
    - Has the number of rings changed?
    - Has the number of chiral centres changed?
    - What bonds have changed?
  - Atom mapping
    - How accurate does it have to be?
      - Should the scientist be allowed to get it wrong, or should this be left to the computer?
    - Mechanism?
    - Verification
    - Machine learning
  - Substructure searching

RInChI

- Have we made the right decisions?

- Future work:
  - more layers or fewer?
  - Polymers - encode as the polymerisation conditions?
  - Key reactions - encode with minimal atoms and list permitted substitution positions in optional layer
  - Non-molecular reagents (Raney Nickel, pH7 buffer, *etc*). List through controlled vocabulary and database?
  - Automated checking of RInChI validity?

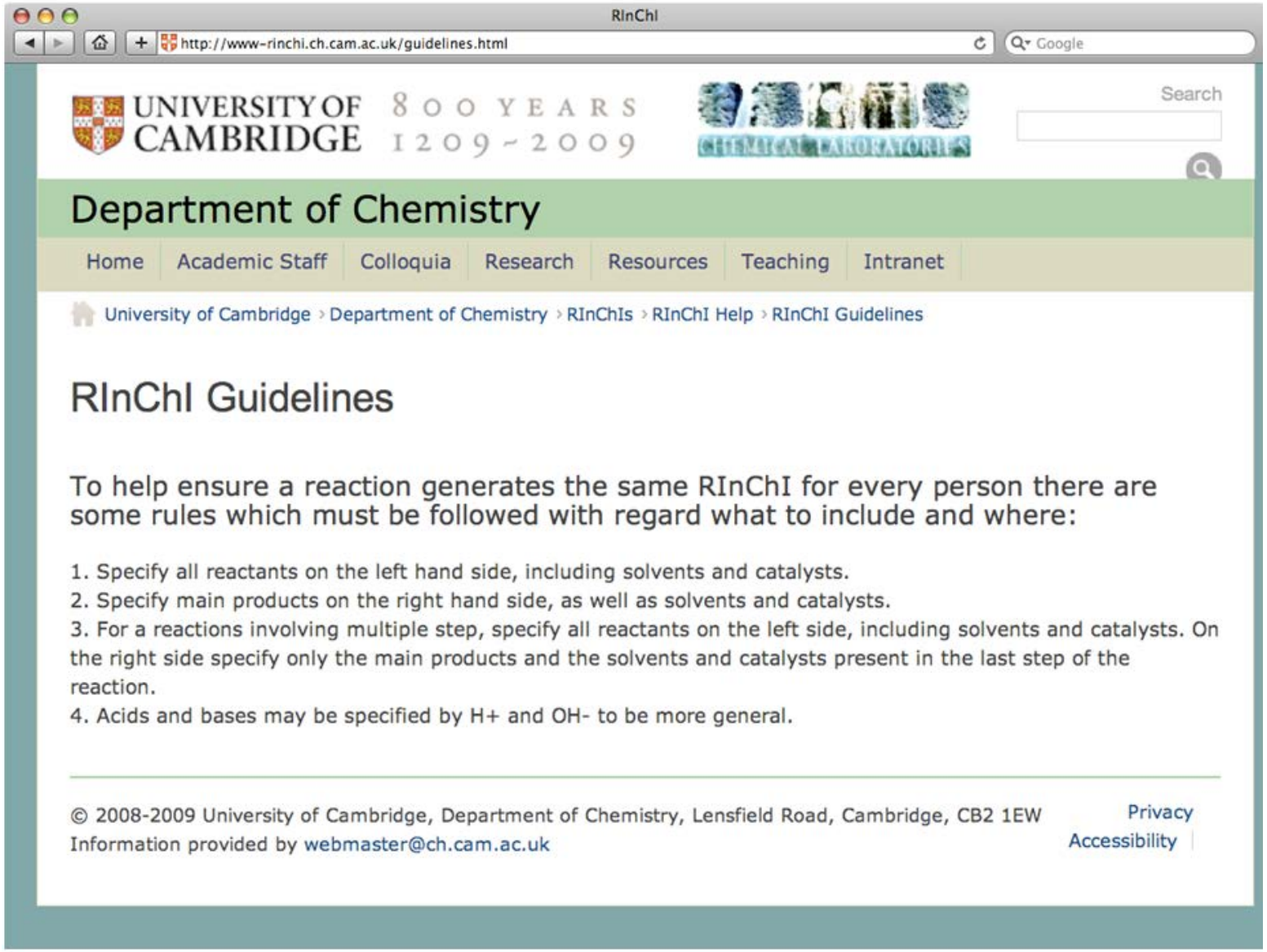
RInChI Summary

- RInChI - *draft version now available*
  - Different groups working independently on one reaction should produce the same RInChI, without communicating
- RInChI:
  - Compact
  - Easy to generate
  - Easy to handle
  - Omits unnecessary information
  - Omits hard-to-find information
  - Omits subjective information
  - Contains the right information
  - Provides a consistent outline of reaction information
  - *A valuable resource for open chemistry*

Reaction InChI *(RInChI)*

1) Is it possible to extend InChIs to reactions…
2) …and do something useful…
3) …that no one else is doing…
4) …and make it easy to use…
5) …and get people to use it…
6) …?????

International chemical identifier for reactions (RInChI)
G. Grethe, J. M. Goodman and C. H. G. Allen
Journal of Cheminformatics 2013, **5**, 45.
DOI: 10.1186/1758-2946-5-45


RInChI
http://www-rinchi.ch.cam.ac.uk/guidelines.html
Google
UNIVERSITY OF CAMBRIDGE
800 YEARS
1209-2009
CHEMICAL LABORATORIES
Search
Department of Chemistry
Home
Academic Staff
Colloquia
Research
Resources
Teaching
Intranet
University of Cambridge › Department of Chemistry › RInChIs › RInChI Help › RInChI Guidelines
RInChI Guidelines
To help ensure a reaction generates the same RInChI for every person there are some rules which must be followed with regard what to include and where:
1. Specify all reactants on the left hand side, including solvents and catalysts.
2. Specify main products on the right hand side, as well as solvents and catalysts.
3. For a reactions involving multiple step, specify all reactants on the left side, including solvents and catalysts. On the right side specify only the main products and the solvents and catalysts present in the last step of the reaction.
4. Acids and bases may be specified by H+ and OH- to be more general.
© 2008-2009 University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW
Information provided by webmaster@ch.cam.ac.uk
Privacy
Accessibility

A Reaction Identifier ?

- Reactions make molecules seem easy
- Reaction space is much bigger than chemical space
- CAS, Beilstein *et al.* provide excellent resources
- Do Electronic Lab Notebooks (ELNs) make this unnecessary?
- Chemistry is too hard to communicate to toxicologists, nano-scientists, biologists, *etc*