

# Open Notebook Science: Does Transparency Work?

## HUBzero Conference

Jean-Claude Bradley

Department of Chemistry  
Drexel University

April 6, 2011

# **The current state of transparency in scientific communication**

## **Case study of melting point data**

# The Chemical Information Validation Sheet

567 curated and referenced measurements from  
Fall 2010 Chemical Information Retrieval course

	CSID	data source	data source type	experimental or predicted	property	source value	source units	common value	common units	link
DDT	2928	MSDS	database - government	experimental	boiling point	260	C	260	C	<a href="http://msds.chem.ox.ac.uk/">http://msds.chem.ox.ac.uk/</a>
DDT	2928	CDC	database - government	experimental	boiling point	230	F	110	C	<a href="http://www.cdc.gov/niosh/np">http://www.cdc.gov/niosh/np</a>
DDT	2928	Wikipedia	website - commercial	experimental	boiling point	260	C	260	C	<a href="http://en.wikipedia.org/wiki/D">http://en.wikipedia.org/wiki/D</a>
DDT	2928	HBCP	website - commercial	experimental	boiling point	260	C	260	C	<a href="http://www.hbcpNetbase.com/id=03_01_91&amp;DocId=11738">http://www.hbcpNetbase.com/id=03_01_91&amp;DocId=11738</a>
DDT	2928	Chemspider	website - commercial	predicted	boiling point	416.2	C	416.2	C	<a href="http://www.chemspider.com/Structure.2928.html">http://www.chemspider.com/Structure.2928.html</a>

cheminfo2010.wikispaces.com

<http://usefulchem.blogspot.com/2011/01/chemical-information-validation-results.html>

# The Chemical Information Validation Explorer

## Chemical Information Retrieval

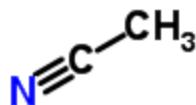
Drexel University Course CHEM 367-767

### Property Explorer

Common Name

Property

acetonitrile



### Statistics

Number of Results: 5

Average Value: -47.1 C

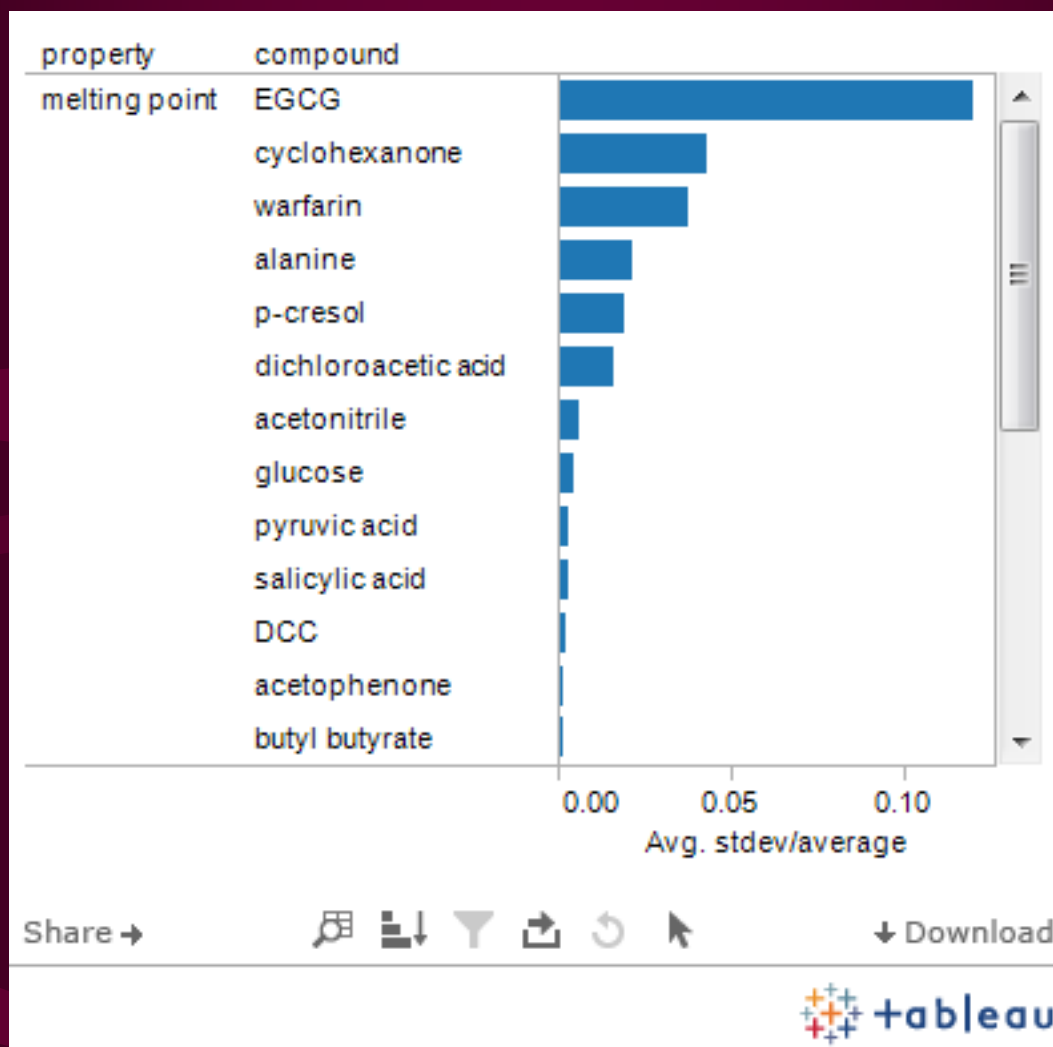
95% Confidence Interval: (-48.152,-46.048)

Standard Deviation: 1.200

Coefficient of Variation: -0.025

Value (C)	Source	Credit	Type
-45.000	Wikipedia	Mirlinda Biba	experimental
-48.000	Sigma Aldrich	Mirlinda Biba	experimental
-48.000	Wolfram Alpha	Mirlinda Biba	experimental
-46.500	Alfa Aesar	Mirlinda Biba	experimental
-48.000	Chemical Book	Mirlinda Biba	experimental

# Discovering outliers for melting points (stdev/average)



# Investigating the m.p. inconsistencies of EGCG

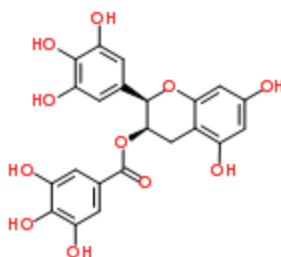
## Chemical Information Retrieval

Drexel University Course CHEM 367-767

### Property Explorer

Common Name   
Property

### EGCG



### Statistics



Number of Results: 2

Average Value: 452.650 K

95% Confidence Interval: (399.292,506.008)

Standard Deviation: 38.500

Coefficient of Variation: 0.085

Value (K)	Source	Image	Credit	Source Type	Value Type	Notes
414.150	Chemistry of Natural Compounds		Jose Rafael Quejada	peer reviewed journal	experimental	
491.150	Merck Index		Jose Rafael Quejada	commercial database	experimental	

(-)-Epigallocatechin-3-O-gallate,  $C_{22}H_{18}O_{11}$ ,  $[M]^+$  458, mp 140-142°C.

CAS Name	synonyms	tradename(manufac	CAS Registry No.	mol. formula	mol. weight	melting point (°C)
3,4,5-Trihydroxybe	<a href="#">view synonyms</a>		989-51-5	$C_{22}H_{18}O_{11}$	458.37	mp 218°

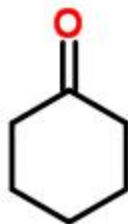
# Investigating the m.p. inconsistencies of cyclohexanone

## Property Explorer

Common Name

Property

cyclohexanone



### Statistics


Number of Results: 5

Average Value: 233.350 K

95% Confidence Interval: (225.444,241.256)

Standard Deviation: 9.020

Coefficient of Variation: 0.039

Value (K)	Source	Image	Credit	Source Type	Value Type	Notes
226.150	Acros		Byron Forte	chemical vendor	experimental	
247.150	Alfa Aesar		Byron Forte	chemical vendor	experimental	
241.150	Oxford		Byron Forte	academic website	experimental	
226.150	Sigma-Aldrich		Byron Forte	chemical vendor	experimental	
226.150	Wolfram Alpha		Byron Forte	free database	experimental	

# Sigma-Aldrich, Acros and Wolfram Alpha apparently use the same sources for melting points

property	compound	data source
melting point	acetonitrile	Sigma-Aldrich
		Wolfram Alpha
	cyclohexanone	Acros
		Sigma-Aldrich
		Wolfram Alpha
	DCC	Sigma-Aldrich
		Wolfram Alpha
	dichloroacetic acid	Acros
		Sigma-Aldrich
	methimazole	Sigma-Aldrich
		Wolfram Alpha
	p-cresol	Sigma-Aldrich
		Wolfram Alpha
	thiophene	Acros
		Sigma-Aldrich



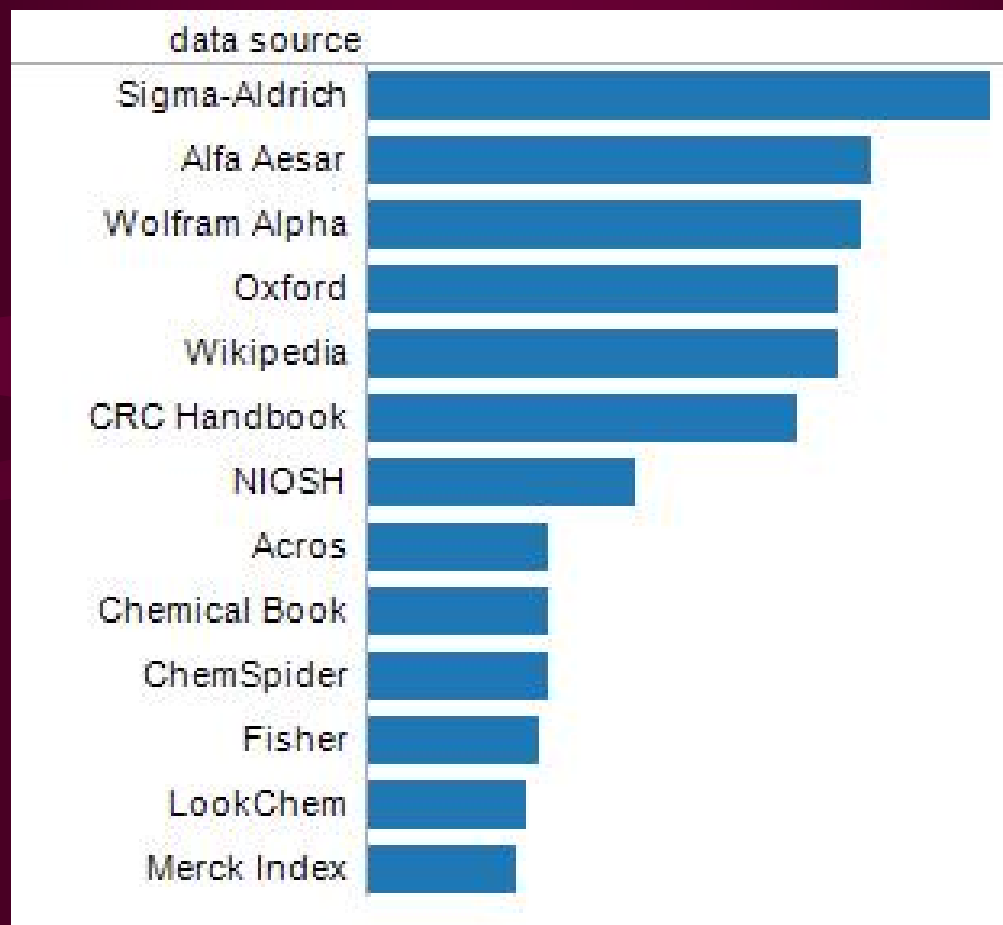
# Sigma-Aldrich, Acros and Wolfram Alpha apparently use the same sources for boiling points

property	compound	data source
boiling point	acetonitrile	Sigma-Aldrich
		Wolfram Alpha
	cyclohexanone	Acros
		Sigma-Aldrich
		Wolfram Alpha
dichloroacetic acid		Acros
		Sigma-Aldrich
p-cresol		Sigma-Aldrich
		Wolfram Alpha
salicylic acid		Sigma-Aldrich
		Wolfram Alpha
thiophene		Acros
		Wolfram Alpha

# Sigma-Aldrich, Acros and Wolfram Alpha apparently DO NOT use the same sources for flash points

property	compound	data source
flash point	acetophenone	Sigma-Aldrich
		Wolfram Alpha
	butyl butyrate	Acros
		Sigma-Aldrich
	cyclohexanone	Acros
		Sigma-Aldrich
		Wolfram Alpha
	DCC	Sigma-Aldrich
		Wolfram Alpha
	dichloroacetic acid	Acros
		Sigma-Aldrich
	methanol	Sigma-Aldrich
		Wolfram Alpha
	p-cresol	Sigma-Aldrich
		Wolfram Alpha
	salicylic acid	Sigma-Aldrich
		Wolfram Alpha
	thiophene	Acros
		Sigma-Aldrich
		Wolfram Alpha

# Most popular data sources



# Alfa Aesar donates melting points to the public



DSSTK	DSDESC	DSPURE	DCMELT
H26903	(-)-1,4-Di-O-tosyl-2,3-O-isopropylidene-L-threitol	98%	90°
B24134	(-)-2,3-O-Isopropylidene-D-threitol	98%	45-49°
L04759	(-)-alpha-Pinene	98%, cont. variat	-64°
A12684	(-)-Borneol	98%	205-208°
A18040	(-)-Camphene	tech. 80%	ca 35°
A18796	(-)-Cinchonidine	99% (total base),	201-206°
A16180	(-)-Dibenzoyl-L-tartaric acid monohydrate	98+%	90-92°
A16181	(-)-Dibenzoyl-L-tartaric acid, anhydrous	99%	154-156°
A17992	(-)-Diethyl D-tartrate	99%	17°
B21029	(-)-DIOP	98%	88-90°
L15151	(-)-Fenchone	98+%	3-5°
L18485	(-)-Lupinine	97%	62-65°
41570	(-)-N-Methylephedrine	98+%	86°-88°
L04848	(-)-Shikimic acid	98%	184-188°
B23090	(+)-2,3-O-Isopropylidene-L-threitol	98%	45-49°
A11542	(+)-5-Iodo-2'-deoxyuridine	98%	ca 190° dec.

# Open Melting Point Explorer

## Open Melting Point Data Explorer

### Melting Point Explorer

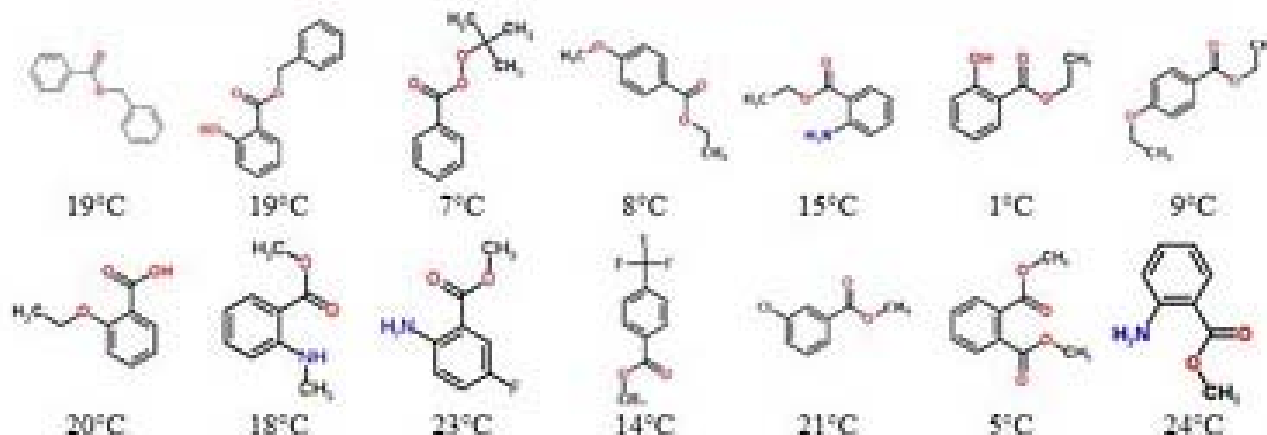
Common Name

- OR -

Find Melting Points via Substructure Search

SMARTS

Filter Min MP:  °C (optional) Max MP:  °C (optional)



14 compounds found

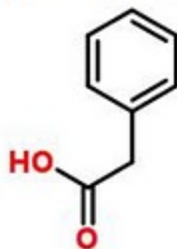
# Outliers

## MDPI dataset



name	mp °C	source	SMILES
phenylacetic acid	77.50	Alfa Aesar	<chem>c1ccc(cc1)CC(=O)O</chem>
Phenylacetic acid	150.00	peer reviewed journal	<chem>O=C(O)Cc1ccccc1</chem>
phenylacetic acid	76.70	government database	<chem>O=C(O)Cc1ccccc1</chem>
phenylacetic acid	77.00	commercial database	<chem>O=C(O)Cc1ccccc1</chem>
phenylacetic acid	77.50	commercial database	<chem>O=C(O)Cc1ccccc1</chem>

The average melting point of phenylacetic acid is 91.74 °C

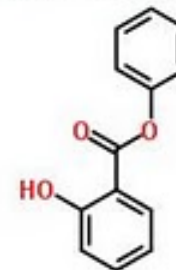


## EPI (via ChemSpider)



name	mp °C	source	SMILES
phenyl salicylate	43.00	Alfa Aesar	<chem>c1ccc(cc1)OC(=O)c2ccccc2O</chem>
phenyl salicylate	130.50	government database	<chem>O=C(Oc1ccccc1)c2ccccc2O</chem>
phenyl salicylate	42.00	chemical vendor	<chem>O=C(Oc1ccccc1)c2ccccc2O</chem>

The average melting point of phenyl salicylate is 71.83 °C



# Outliers

## Alfa Aesar



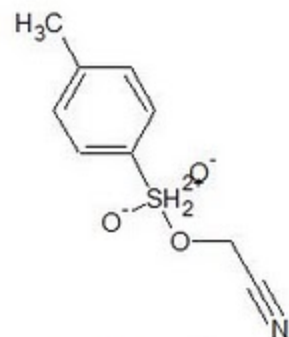
name	mp °C	source	SMILES
ethanol	-130.00	Alfa Aesar	CCO
ethanol	-114.40	commercial database	CCO
ethanol	-114.30	crowdsourced database	CCO
ethanol	-114.14	commercial database	CCO
ethanol	-114.00	academic website	CCO
ethanol	-114.00	chemical vendor	CCO

The average melting point of ethanol is -116.81 °C

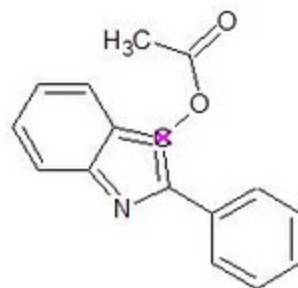


# Inconsistencies and SMILES problems within MDPI dataset

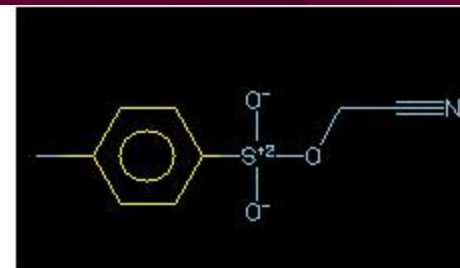
name	mp C	Error C
androstane-3,17-dione	132	47.5
androstane-3,17-dione	179.5	47.5
Mephénytoin	136	74
Mephénytoin	210	74
N-cyclohexylbenzamide	145	17
N-cyclohexylbenzamide	162	17
9-Fluoro-6,7-dihydro-5,8-dimethyl-1-oxo-1H,5H-benzo[ij]quinoli:	194	75
9-Fluoro-6,7-dihydro-5,8-dimethyl-1-oxo-1H,5H-benzo[ij]quinoli:	269	75
5,6,7,8-tetrahydronaphthalene-1-carboxylic acid	148.5	11
5,6,7,8-tetrahydronaphthalene-1-carboxylic acid	159.5	11
N1-phenyl-3,3,3-trifluoro-2-(trifluoromethyl)propanamide	166	25
N1-phenyl-3,3,3-trifluoro-2-(trifluoromethyl)propanamide	191	25
N,N'-diphenylethanebis(thioamide)	134.5	16.5
N,N'-diphenylethanebis(thioamide)	151	16.5



[S+2]([O-])([O-])(OCC#N)c1ccc(C)cc1



O=C(OC=1=C2C=CC=CC2=NC=1c1ccccc1)C



[S+2]([O-])([O-])(OCC#N)c1ccc(C)cc1



# MDPI Dataset labeled with High Trust Level

www.cheminformatics.org

QSPR datasets [top](#)

Name	Activity/Property	Reference	Dataset/Linkout	Trust Level
Karthikeyan Melting Point Dataset	Melting Points for 4173 Training Set Molecules and 277 Test Set Compounds (Drug-Like)	Karthikeyan, M.; Glen, R.C.; Bender, A. General melting point prediction based on a dense compound dataset and artificial neural networks. <i>J Chem. Inf. Model.</i> 2005, 45(3), 581-590. - <a href="#">linkout</a>	<a href="#">Dataset</a> (7MB: Excel File with Structures in SMILES Format, Melting Points and MOE 2D and 3D Descriptors)	High - Original Author Data

# Open Melting Point Datasets

**ONSMP000:** ([ONSCwiki](#)) 15591 full raw entries from Alfa Aesar containing duplicates and non numerical values

**ONSMP001:** ([ORU](#)) 12986 measurements as simple numeric values converted from mp ranges and other entries with non-numeric characters from Alfa Aesar (ONSMP000).

**ONSMP002:** ([ORU](#)) 8739 measurements derived from ONSMP001 with redundancies, salts, inorganics and organometallics removed. Silicon, phosphorus and boron containing organic compounds were retained. SMILES, CSIDs and links to the Alfa Aesar catalog are included.

**ONSMP003:** ([ORU](#)) 4450 measurements from [Karthikeyan 2005](#). Includes SMILES and many descriptors.

**ONSMP004:** ([ORU](#)) 4084 measurements derived from ONSMP003 - includes compound names and CSIDs - excludes SMILES that did not properly render with OpenEye. 48 compounds were missing from ONSMP004 that were in ONSMP003 - these have been recovered but they do not have associated names or CSIDs: [ONSMP004a](#)

**ONSMP005:** ([ORU](#)) 277 measurements from [Bergstrom 2003](#). Drug molecules separated as training and validation sheets. SMILES provided.

**ONSMP006** ([ORU](#)) 277 measurements derived from ONSMP005 compiled into one sheet and both SMILES and CSIDs provided.

**ONSMP007** (pending) curated Karthikeyan dataset ONS004 further curated by removal of all duplicate entries (with very different melting points)

**ONSMP008** ([ONSCwiki](#)) 33 Duplicates (66 measurements) with a difference with more than 10C from the Karthikeyan dataset ONSMP003

**ONSMP009** ([ONSCwiki](#)) 311 SMILES which could not be rendered correctly on ChemSketch from Karthikeyan dataset ONSMP003

**ONSMP010** ([ONSCwiki](#)) 150 SMILES consisting of all EPI melting point data (via ChemSpider) from a 2011-03-04 snapshot of Cheminfo Validation sheet. 106 of these have at least one MP from another source. 10 of the 106 show a difference of at least 5C between the EPI and the other sources.

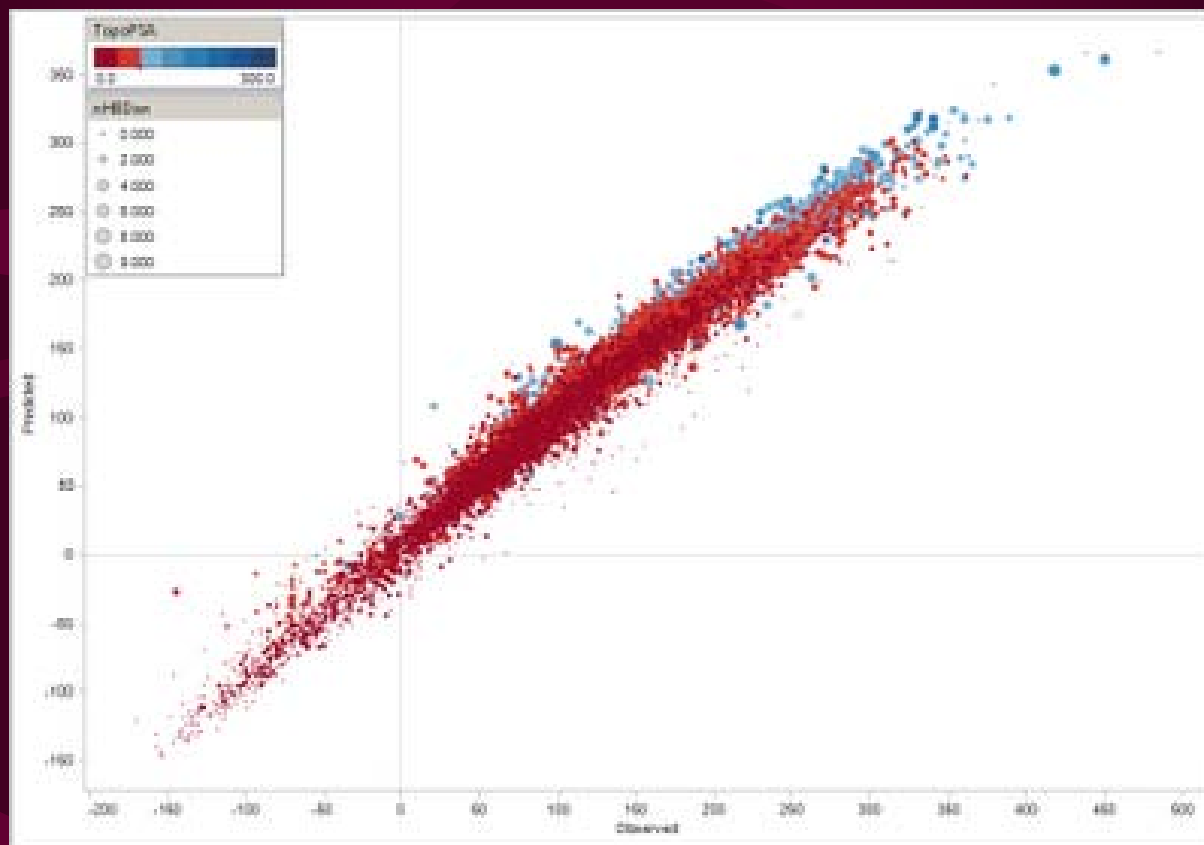
**ONSMP011** ([ONSCwiki](#)) 335 measurements. A snapshot taken 2011-02-20 of the crowdsourced melting point data in the [ChemInfo Validation Sheet](#).

**ONSMP012** ([ONSCwiki](#)) 1286 measurements removed from the union of ONSMP002, ONSMP003, ONSMP006, and ONSMP011. Data were removed because they were either salts, had a large discrepancy in measurements (greater than 10C), were suspected erroneous measurements, were unneeded duplicates, or failed to produce CDK descriptors, see [meltingpointmodel001](#).

**ONSMP013** ([ORU](#)) 12634 highly curated (see ONSMP012 above) unique melting point measurements with CDK descriptor values based upon the union of ONSMP002, ONSMP003, ONSMP006, and ONSMP011.

# Open Random Forest modeling of Open Melting Point data using CDK descriptors (Andrew Lang)

R2 = 0.78, TPSA and nHdon most important

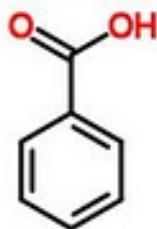


# Melting point prediction service



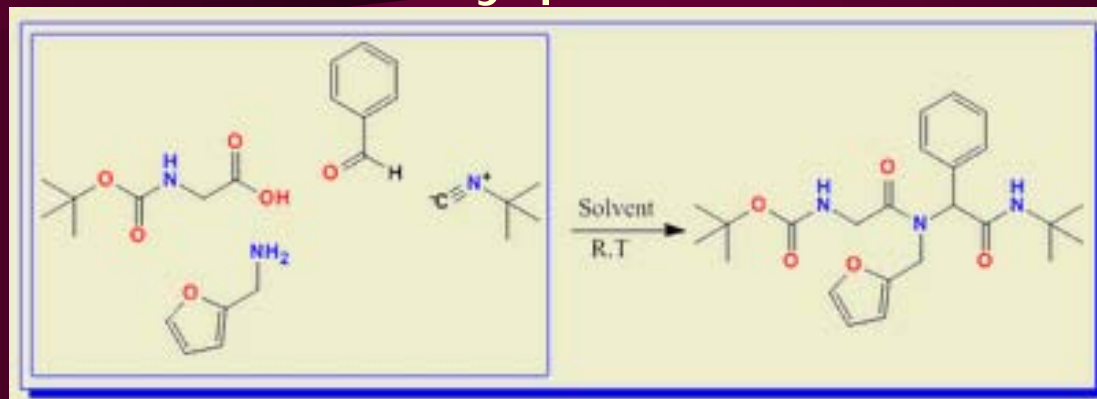
name	mp °C	source	SMILES
benzoic acid	123.00	Alfa Aesar	<chem>c1ccc(cc1)C(=O)O</chem>
benzoic acid	122.00	peer reviewed journal (sup. data)	<chem>OC(=O)c1ccccc1</chem>
Benzoic Acid	122.00	peer reviewed journal (sup. data)	<chem>O=C(O)c1ccccc1</chem>
benzoic acid	122.40	government database	<chem>c1ccc(cc1)C(=O)O</chem>
benzoic acid	122.00	peer reviewed journal	<chem>c1ccc(cc1)C(=O)O</chem>
benzoic acid	122.00	peer reviewed journal	<chem>c1ccc(cc1)C(=O)O</chem>
benzoic acid	123.00	chemical vendor	<chem>c1ccc(cc1)C(=O)O</chem>
benzoic acid	122.35	commercial database	<chem>c1ccc(cc1)C(=O)O</chem>
benzoic acid	122.40	commercial database	<chem>c1ccc(cc1)C(=O)O</chem>

The average melting point of benzoic acid is 122.35 °C



Predicted melting point °C.  
107.91

# Using melting point for temperature dependent solubility prediction



## Optimal Solvent Prediction

Solvents have then been ranked by theoretical percent yield and the number of criteria met.

Criteria: lim. reactants (M) 0.3    lim. product (M) 0.03    bp °C 100

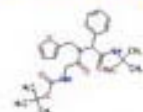
Options: washes 0    (alters % yield)   

all criteria met

two criteria met

one criterion met

no criteria met



rank solvent	AD model101	AD measured experimental	AD model101	AD measured experimental	BP °C	% Yield	SME*
1 2-propanol	0.067	*	3.625	0.021	*	0.000	73.04 100
2 diethyl ether	0.010	*	2.650	0.002	*	0.000	33.24 100 3.9
3 ethanol	0.101	*	4.110	0.040	*	0.017	72.64 98 2.5
4 acetonitrile	0.019	*	2.460	0.006	*	0.022	63.54 98 4.6
5 water	0.055	0.830	0.830	0.000	*	*	100.04 100
6 hexane	0.000	*	0.250	0.000	*	0.000	68.54 100 4.1
7 benzene	0.002	*	0.255	0.020	*	0.006	78.84 98 4.2
8 methanol	0.198	*	4.725	0.088	*	0.046	64.71 95 2.6
9 methyl tert-butyl ether	0.008	*	*	0.001	*	*	55.24 84
10 2-methyl-2-propanol	0.060	*	*	0.012	*	*	84.64 80
11 dichloromethane	0.020	*	4.140	0.606	*	0.238	39.64 76 3.6
12 THF	0.062	*	3.716	0.084	*	0.258	68.34 74 3.8

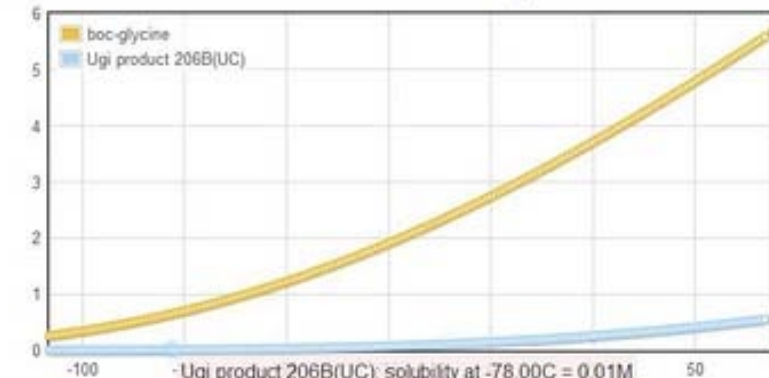
## Solubility (M) in THF from -108.4C (mp) to 68.3C (bp)

### boc-glycine

miscible point 6.62M  
melting point 88C  
temp at 0.03M -152.05C  
temp at 0.3M -104.11C

### Ugi product 206B(UC)

miscible point 2.58M  
melting point 203C  
temp at 0.03M -52.34C  
temp at 0.3M 32.33C





WIKIPEDIA  
The Free Encyclopedia

## Open Notebook Science

From Wikipedia, the free encyclopedia

**Open Notebook Science** is the practice of making the entire primary record of a research project publicly available online as it is recorded. This involves placing the personal, or laboratory, notebook of the researcher online along with all raw and processed data, and any associated material, as this material is generated. The

## References

- <sup>a</sup> <sup>b</sup> Goetz, T. Freeing the Dark Data of Failed Scientific Experiments *Wired Magazine*, Sept.25, 2007. [↗](#)
- <sup>a</sup> Sanderson, K (September 2008). "Data on display". *Nature*. doi:10.1038/455273a [↗](#).
- <sup>a</sup> Singh, S. (April 2008). "Data on display". *Cell*. doi:10.1016/j.cell.2008.04.003 [↗](#).
- <sup>a</sup> Lloyd, R. Era of Scientific Secrecy Near End *Live Science*, Sept 2, 2008. [↗](#)
- <sup>a</sup> Williams, A. J. Internet-based tools for communication and collaboration in chemistry *Drug Discovery Today*, vol 13, p. 502 (2008). [↗](#)
- <sup>a</sup> Everts, S. Open Source Science, *Chemical & Engineering News*, July 2006, 84 (30) p. 34. [↗](#)

Motivation: Faster Science, Better Science

There are NO FACTS,  
only measurements embedded  
within assumptions

Open Notebook Science maintains  
the integrity of data provenance by  
making assumptions explicit

TRUST



PROOF



# Strategy for an Open Notebook:

First record then abstract structure

In order to be discoverable use Google friendly formats (simple HTML, no login)

In order to be replicable use free hosted tools (Wikispaces, Google Spreadsheets)

# Crowdsourcing Solubility Data

Open Notebook Science Challenge



**Alfa Aesar**<sup>®</sup>  
*A Johnson Matthey Company*

**RSC** | Advancing the  
Chemical Sciences

Sponsors of previous challenges



submeta

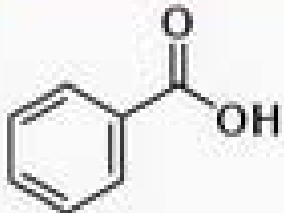
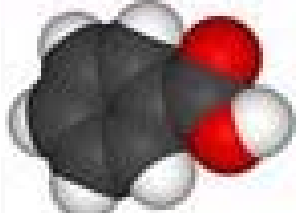


nature



<http://onschallenge.wikispaces.com/>

# Data provenance: From Wikipedia to...


Benzoic acid	
	


  

Properties	
Molecular formula	$C_6H_5COOH$
Molar mass	122.12 g/mol
Appearance	Colourless crystalline solid
Density	1.32 g/cm <sup>3</sup> , solid
Melting point	122.4 °C (395 K)
Boiling point	249 °C (522 K)
Solubility in water	Soluble (hot water) 3.4 g/l (25 °C)
Solubility in THF, ethanol, methanol	THF 3.646 M, ethanol 2.435 M, methanol 2.904 M <sup>[1]</sup>

### References

- <sup>\*</sup> <http://oru.edu/ccoda/s/solubility/s/solvents.php?solute=benzoic%20acid>  Open Notebook Science Challenge Data.

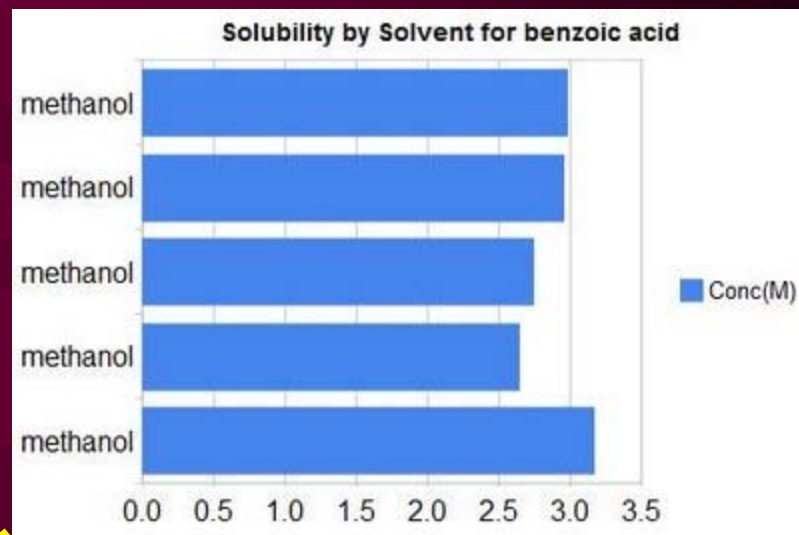


# ...the lab notebook and raw data

OPEN NOTEBOOK SCIENCE CHALLENGE

Solubility of benzoic acid in non-aqueous solvents.  
Total Number of Results: 13  
Total Number of Solvents: 5

Solvent	Hits	Mean Conc. (M)	SD	Link to Detailed Results
1. THF	2	3.646	0.073	<a href="#">Solubility of benzoic acid in THF</a>
2. acetonitrile	2	0.757	0.015	<a href="#">Solubility of benzoic acid in acetonitrile</a>
3. ethanol	2	2.435	0.002	<a href="#">Solubility of benzoic acid in ethanol</a>
4. methanol	5	2.904	0.182	<a href="#">Solubility of benzoic acid in methanol</a>
5. toluene	2	0.63	0	<a href="#">Solubility of benzoic acid in toluene</a>



Exp005 Edit This Page page discussion history notify me

### Objective

To measure the solubility of benzoic acid ([Chemspider](#)), in methanol ([Chemspider](#)). For project see [here](#).

### Procedure

Saturated solutions are made of benzoic acid ([Chemspider](#)), in methanol ([Chemspider](#)) dram vial with 700 $\mu$ l of the solvent.  
The solute is then added in subsequent amounts until the solution has reached a clear point of saturation. Each vial is vortexed for 30 seconds. After the process of vortexing is complete, the solution is centrifuged for one minute. 300 $\mu$ l of the mother liquor is then pipetted into pre-weighed half-dram vials. The solution is weighed and entered into the speedvac, noting the pressure. The solution is then dried until the liquid has evaporated off. The results can be found on the following [Spreadsheet](#).

Solute	Solvent	Wt of empty vials (g)
Benzoic acid	Methanol	2.55842
Benzoic acid	Methanol	2.56923

# Calculations Made Public on Google Spreadsheets

**EXP208-WS1**

File Edit Format Insert Tools Form Help

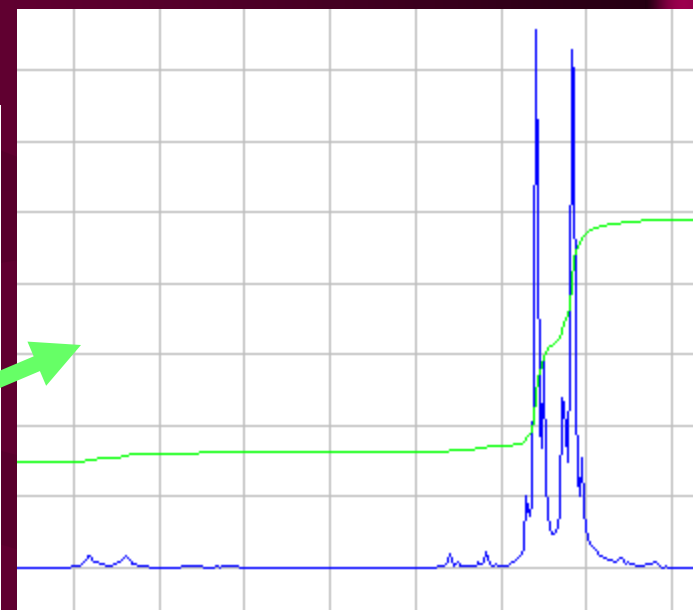
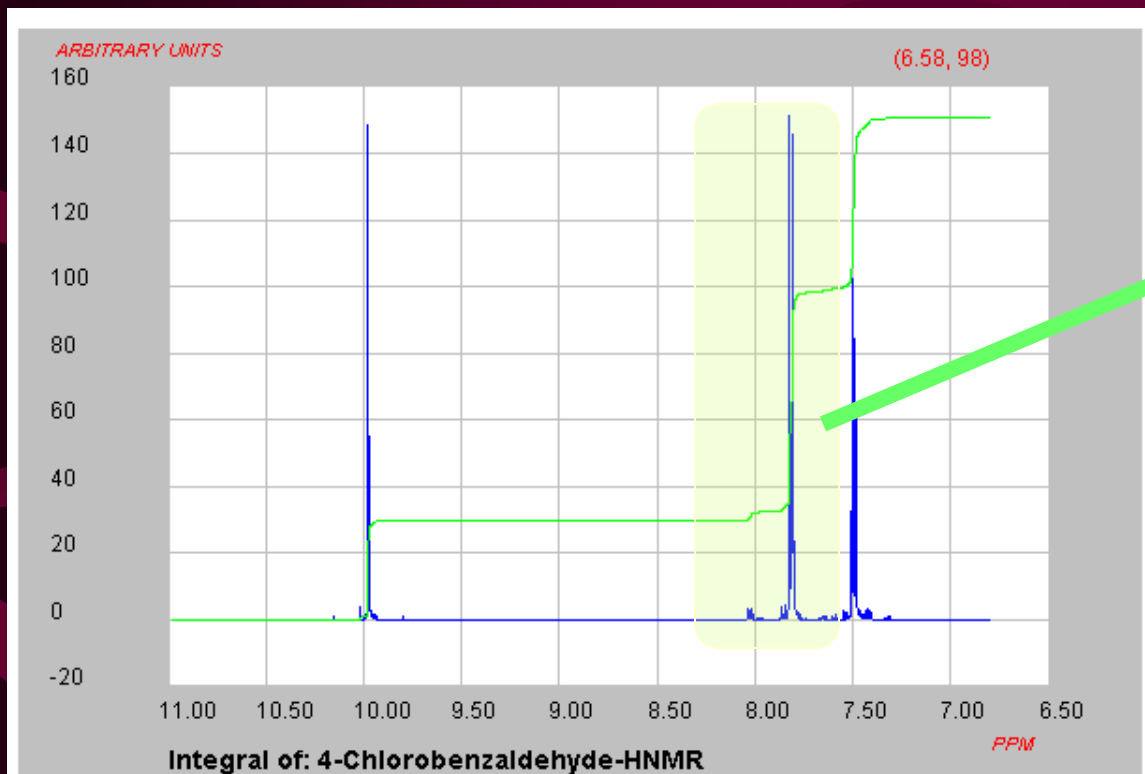
\$ % 123 10pt B Abc A [Color] [Grid] [List] [Link] [Unlink] Σ

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Si	ID	Solute	Solvent	Wt of empty vials (g)	Wt of vial w/solution (300uL)	Wt of sample solution (300uL)	Wt of vial after SpeedVac (g)	Solid dissolved in 300 uL of solvent (mg)	molecular weight of solid	Amt dissolved in 300 uL solvent (mmol)	Saturated Soln Conc (M)	Saturated soln-Density
2	1	1b	3,4-dimethoxyb	THF	2.57997	2.89851	0.31854	2.8292	249.23	166.17	1.50	5.00	1.0618
3	2	2b	3,5-dimethoxyb	THF	2.56974	2.8651	0.29536	2.77057	200.83	166.17	1.21	4.03	0.9845
4	3	3b	0-vanillin	THF	2.58903	2.92017	0.33114	2.83427	245.24	152.15	1.61	5.37	1.1038
5	4	4b	4-nitrobenzaldh	THF	2.56176	Aborted	Aborted	Aborted	Aborted	151.12	Aborted	Aborted	Aborted
6	5	5b	p-Toluenesulfon isocyanide	THF	2.54702	2.7989	0.25188	2.65069	103.67	186.21	0.56	1.86	0.8396



<http://usefulchem.wikispaces.com/Exp208>

# Interactive NMR spectra using JSpecView and JCAMP-DX



JSpecView version number is 1.0.20071009-2130

Created and maintained by Prof Robert John Lancashire



<http://usefulchem.wikispaces.com/Exp208>

# Raw Data As Images

flickr®

## Usefulchem EXP208 41b-50b

ALL SIZES



Splatter?

Some liquid

### Comments



[Chris Bohinski](#) pro says:

Hi, I'm an admin for a group called [Flickr Envy – LOVE TO SHARE PHOTOS](#), and we'd love to have this added to the group!

Posted 5 weeks ago. ( [permalink](#) )

# YouTube for demonstrating experimental set-up

Exp 009



0:17 / 0:35

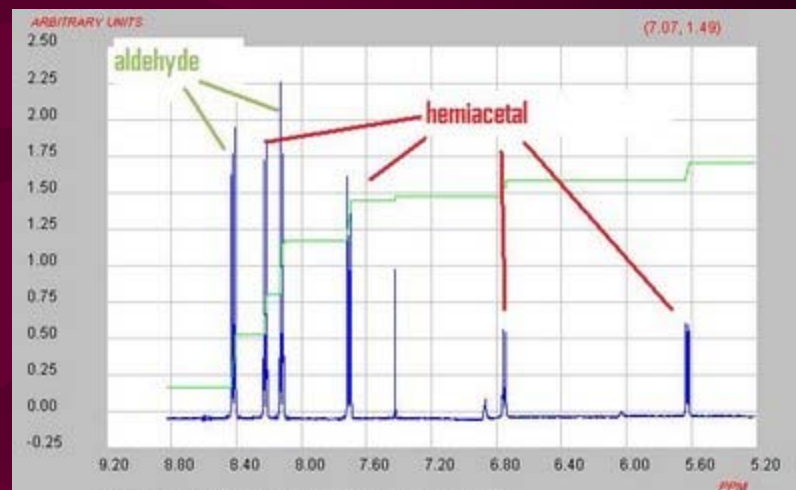
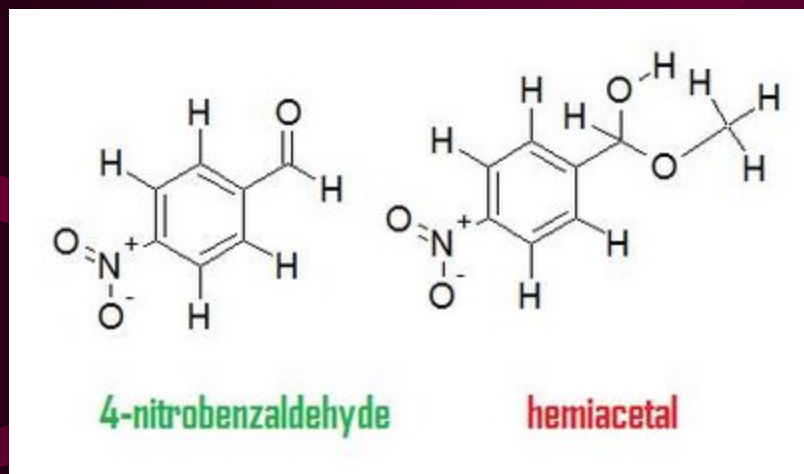


Rate: ★★☆☆☆ 15 ratings

Views: 5,631



# The importance of raw data availability



Missed in a prior publication on solubility for this compound



## DETERMINATION OF THE SOLUBILITIES

An excess of finely powdered PNBA, BMTTP, TPPO, and PNPPPO was added to 5 ml of the appropriate solvent and shaken in a thermostatted bath for 24 h. Aliquots of 0.5 ml were withdrawn and analyzed after appropriate dilution. PNBA and TPPO were analyzed by GLC with internal standard methods (*p*-nitroacetophenone and *p,p'*-dimethoxybenzophenone, respectively). BMTTP and PNPPPO were analyzed by UV spectrophotometry in ethanol at 318 nm for the ylid ( $\epsilon=11,000$ ) and at 316 nm for the chalcone ( $\epsilon=29,000$ ).

# Solubilities collected in a Google Spreadsheet

Google Docs BETA

## SolubilitiesSum

File Edit Format Insert Tools Form Help

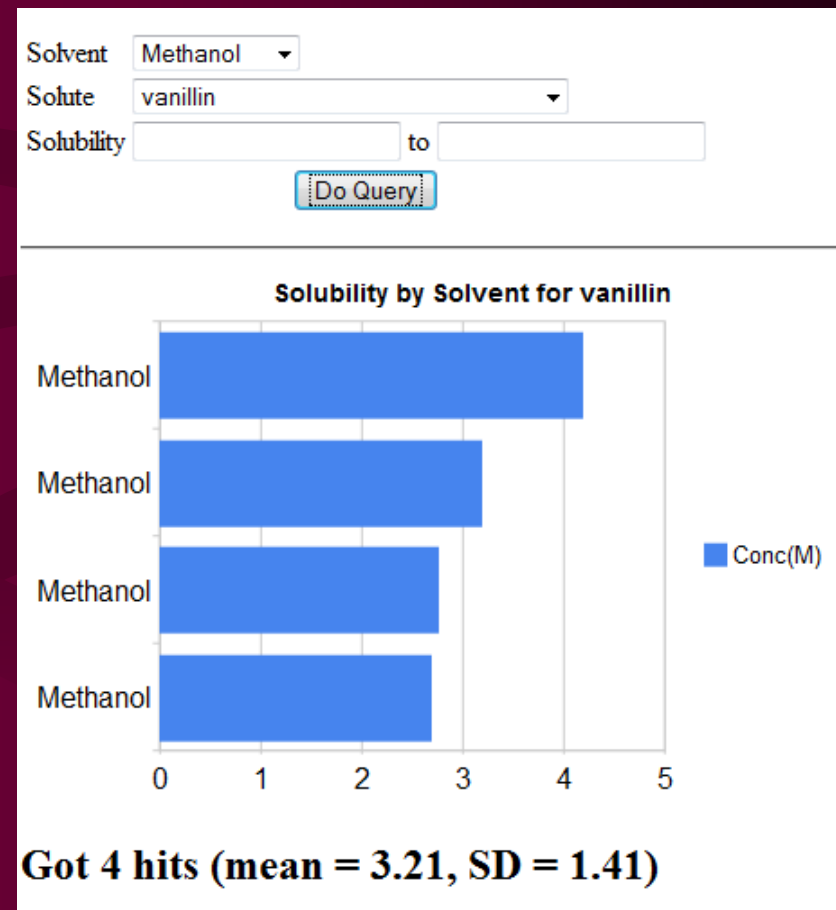
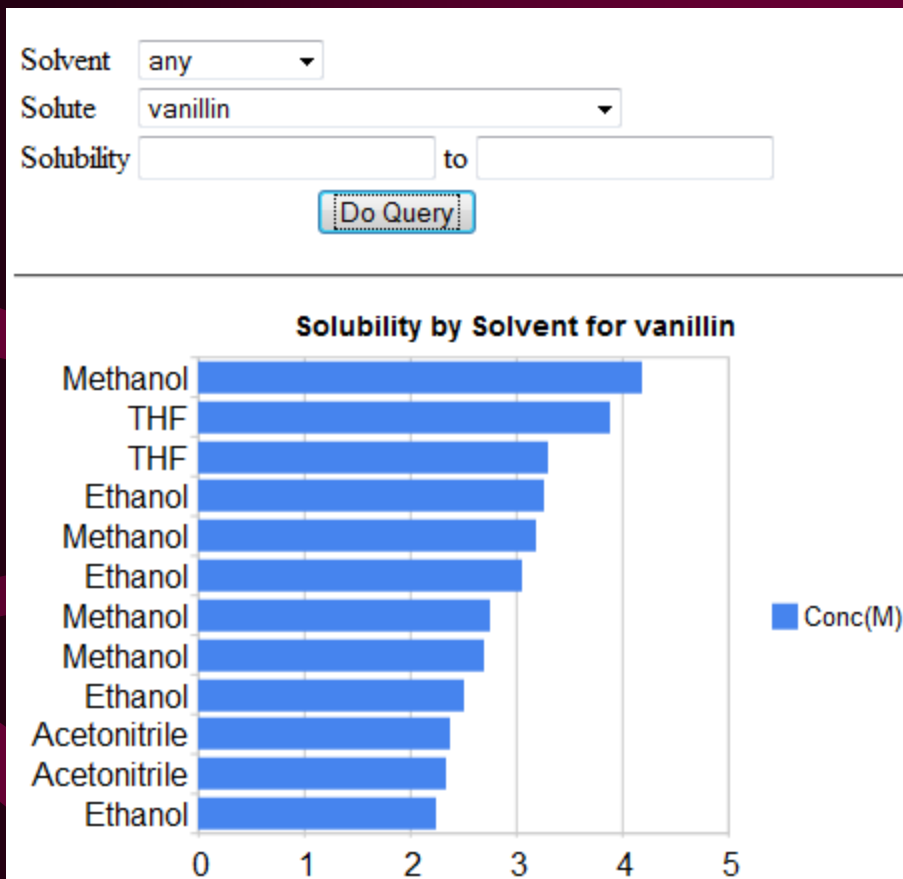
File Edit Format Insert Tools Form Help

10pt B A

	A	B	C	D
	EXP	sample	ref	solute
2	208	11b	<a href="http://usefulchem.wikispaces.com/exp208">http://usefulchem.wikispaces.com/exp208</a>	3,4-dimethoxybenzaldehyde
3	208	41b	<a href="http://usefulchem.wikispaces.com/exp208">http://usefulchem.wikispaces.com/exp208</a>	3,4-dimethoxybenzaldehyde
4	208	21b	<a href="http://usefulchem.wikispaces.com/exp208">http://usefulchem.wikispaces.com/exp208</a>	3,4-dimethoxybenzaldehyde
5	208	3b	<a href="http://usefulchem.wikispaces.com/exp208">http://usefulchem.wikispaces.com/exp208</a>	o-vanillin
6	208	1b	<a href="http://usefulchem.wikispaces.com/exp208">http://usefulchem.wikispaces.com/exp208</a>	3,4-dimethoxybenzaldehyde
7	205		1 <a href="http://usefulchem.wikispaces.com/exp205">http://usefulchem.wikispaces.com/exp205</a>	3,4-dimethoxybenzaldehyde
8	210		40 <a href="http://usefulchem.wikispaces.com/exp210">http://usefulchem.wikispaces.com/exp210</a>	crotonic acid

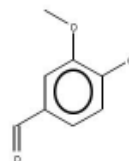
solute SMILES	solvent	solvent SMILES	concentration (M)	wiki page	Added to ChemSpider SDF
<chem>COc1cc(ccc1OC)C=O</chem>	Acetonitrile	<chem>N#CC</chem>	5.57	UC	
<chem>COc1cc(ccc1OC)C=O</chem>	Ethanol	<chem>OCC</chem>	5.55	UC	
<chem>COc1cc(ccc1OC)C=O</chem>	Chloroform	<chem>ClC(Cl)Cl</chem>	5.44	UC	
<chem>Oc1c(cccc1OC)C=O</chem>	THF	<chem>O1CCCC1</chem>	5.37	UC	
<chem>COc1cc(ccc1OC)C=O</chem>	THF	<chem>O1CCCC1</chem>	5.00	UC	
<chem>COc1cc(ccc1OC)C=O</chem>	Methanol	<chem>OC</chem>	4.92	UC	
<chem>O=C(O)/C=C/C</chem>	Ethanol	<chem>OCC</chem>	4.65	UC	
<chem>O=C(O)/C=C/C</chem>	Methanol	<chem>OC</chem>	4.56	UC	
<chem>O=C(OC(C)(C)C)NCC(=O)O</chem>	Methanol	<chem>OC</chem>	4.40	UC	YES

# Rajarshi Guha's Live Web Query using Google Viz API



Serial	EXP	Sample	Ref	Solute	Solute SMILES	Solvent	Solvent SMILES	Conc (M)
--------	-----	--------	-----	--------	---------------	---------	----------------	----------

1	207	3	<a href="http://usefulchem.wikispaces.com/exp207">http://usefulchem.wikispaces.com/exp207</a>	vanillin	<chem>O=Cc1ccc(OC)cc1</chem>	Methanol	OC	4.19
---	-----	---	---	----------	------------------------------	----------	----	------

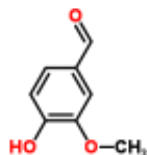


# Web services for summary data



These results are from the [Open Notebook Science Solubility Challenge](#) as of January 15, 2011

A compilation of the results in book form can be obtained from [Nature Precedings](#)



**Solubility of vanillin in organic solvents.**

**Total Number of Results: 17**

**Total Number of Solvents: 9**

Solvent	Ave. (M)	Hits	SD	Link to Detailed Results
1. 1,2-dichloroethane	1.175	1	0.000	<a href="#">Solubility of vanillin in 1,2-dichloroethane is 1.175 M</a>
2. 1-propanol	1.820	1	0.000	<a href="#">Solubility of vanillin in 1-propanol is 1.820 M</a>
3. THF	3.594	2	0.419	<a href="#">Solubility of vanillin in THF is 3.594 M</a>
4. acetonitrile	2.360	2	0.028	<a href="#">Solubility of vanillin in acetonitrile is 2.360 M</a>
5. butanone	2.138	1	0.000	<a href="#">Solubility of vanillin in butanone is 2.138 M</a>
6. ethanol	2.470	5	0.152	<a href="#">Solubility of vanillin in ethanol is 2.470 M</a>
7. methanol	4.160	3	0.026	<a href="#">Solubility of vanillin in methanol is 4.160 M</a>
8. toluene	0.302	1	0.000	<a href="#">Solubility of vanillin in toluene is 0.302 M</a>
9. water	0.070	1	0.000	<a href="#">Solubility of vanillin in water is 0.070 M</a>

Permalink: <http://old.oru.edu/ccda/sl/solubility/allsolvents.php?solute=vanillin>

(Andrew Lang)

# Web service calls from within a Google Spreadsheet for solubility measurement and prediction

Google docs PredictionDemo Public on the web

File Edit View Insert Format Form Tools Help

10pt B Abc

	A	B	C	D	E
1	solute (name, CSID or SMILES)	solvent	predicted solubility (M)	measured solubility (M)	%error
2	benzoic acid	methanol	1.39	2.831	51
3	cinnamic acid	ethanol	0.843	1.03	18
4	ibuprofen	methanol	2.639	0.858	208
5	phenanthrene-9-carboxaldehyde	ethyl acetate	0.018	0.44	96
6	<chem>O=C(O)c1ccccc1</chem>	acetonitrile	0.464	No Records	#VALUE!
7	phenanthrene-9-carboxaldehyde	benzene	0.066	0.66	90

<http://onswebservices.wikispaces.com/>

(Andrew Lang)

# Integration of Multiple Web Services to Recommend Solvents for Reactions

## Optimal Solvent Prediction

Solvents have then been ranked by **theoretical percent yield** and the number of criteria met.

Criteria: lim. reactants (M)  lim. product (M)  bp °C

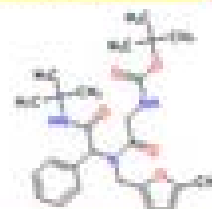
Options: washes  (alters % yield)

all criteria met

two criteria met

one criterion met

no criteria met



Green Solvent Metric



rank	solvent	AD model1001	AD measured experimental	AD model1001	AD measured experimental	BP °C	% Yield	SHE	
1	methyl tert-butyl ether	0.006	1.798	*	0.004	*	55.2	100	
2	diethyl ether	0.011	2.169	2.650	0.005	*	33.2	99	3.9
3	carbon tetrachloride	0.000	0.731	*	0.006	*	76.0	99	3.1
4	1-chlorobutane	0.001	0.981	*	0.010	*	78.2	99	3.8
5	acetonitrile	0.019	1.285	2.460	0.013	*	63.5	99	4.6
6	ethyl acetate	0.016	2.386	*	0.015	*	73.9	99	2.9

## Open Notebook Science Challenge: Solubilities of Organic Compounds in Organic Solvents (3RD)

by [Andrew Lang et al.](#)

★★★★★ (1 Rating)

Paperback, 130 pages

**\$8.20**

Ships in 3–5 business days

Solubilities of organic compounds in organic solvents compiled and measured during the Open Notebook Science Challenge

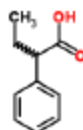
Open Notebook Science  
Challenge: Solubilities of  
Organic Compounds in  
Organic Solvents (3RD)



Jean-Claude Bradley, Cameron Neylon, Andrew Williams, Rajeshi Gaha,  
Bill Hooper, Andrew Lang, Tim Holinski/David Balguy, Matthew  
Federici, Jennifer Hale, Jenna Mascitelli, Khalid Miza, Marshall Moritz,  
Daniel Reiss, Cedric Tchakounte, Hal Wooning, Brent Frazier  
Editors: Jean-Claude Bradley and Andrew Lang

### Product Details

<b>ISBN</b>	978-0-557-31801-8
<b>Copyright</b>	©2010 Jean-Claude Bradley, et al.
<b>Language</b>	English
<b>Country</b>	United States
<b>Publication Date</b>	February 11, 2010
<b>Page Count</b>	130 pages
<b>Size</b>	U.S. Trade
<b>Binding</b>	Perfect Bound
<b>Interior Color</b>	Black And White



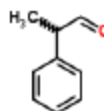
#### Compound Data

Molecular weight	164.201	H bond acceptors	2	Rule of 5 violations	0
Compound type	carboxylic acid	H bond donors	1	ACD/ALogP	2.38
Phase 25°C	solid	Rotatable bonds	3	Predicted density	1.09 g/cm <sup>3</sup>
SMILES	O=C(O)C(c1ccccc1)CC				
InChIKey	OF7WFSNDPCAWDK-UHFFFAOYSA-N				

#### Solubility Data

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
acetonitrile	6.18	0.83	1968.04
DMSO	5.97	0.81	886.12
ethanol	6.20	0.85	1977.18
THF	5.96	0.82	1059.60
toluene	5.55	0.74	519.73

### 2-phenylpropanal C<sub>9</sub>H<sub>10</sub>O<sup>21</sup>



#### Compound Data

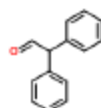
Molecular weight	134.175	H bond acceptors	1	Rule of 5 violations	0
Compound type	aldehyde	H bond donors	0	ACD/ALogP	2.13
Phase 25°C	liquid	Rotatable bonds	2	Predicted density	0.98 g/cm <sup>3</sup>
SMILES	c1ccccc1C(C)C=O				
InChIKey	IQVAERLDLAZARL-UHFFFAOYSA-N				

#### Solubility Data

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
methanol	7.53	0	0

0 Solute is very soluble/miscible, conversion fail.

### 2,2-diphenylacetaldehyde C<sub>14</sub>H<sub>12</sub>O<sup>21</sup>



#### Compound Data

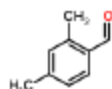
Molecular weight	196.245	H bond acceptors	1	Rule of 5 violations	0
Compound type	aldehyde	H bond donors	0	ACD/ALogP	3.67
Phase 25°C	liquid	Rotatable bonds	3	Predicted density	1.069 g/cm <sup>3</sup>
SMILES	c1ccccc1C(c2ccccc2)C(=O)				
InChIKey	HLLGFBGLKQIZOM-UHFFFAOYSA-N				

#### Solubility Data

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
methanol	5.64	0	0

0 Solute is very soluble/miscible, conversion fail.

### 2,4-dimethylbenzaldehyde C<sub>9</sub>H<sub>10</sub>O<sup>21</sup>



#### Compound Data

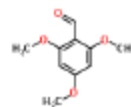
Molecular weight	134.175	H bond acceptors	1	Rule of 5 violations	0
Compound type	aldehyde	H bond donors	0	ACD/ALogP	2.56
Phase 25°C	liquid	Rotatable bonds	1	Predicted density	1.003 g/cm <sup>3</sup>
SMILES	O=Cc1ccc(C)cc1C				
InChIKey	GISVICWQYMLUPF-UHFFFAOYSA-N				

#### Solubility Data

15

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
methanol	7.17	0.88	3128.10

### 2,4,6-trimethoxybenzaldehyde C<sub>10</sub>H<sub>12</sub>O<sub>4</sub><sup>22</sup>



#### Compound Data

Molecular weight	196.2	H bond acceptors	4	Rule of 5 violations	0
Compound type	aldehyde	H bond donors	0	ACD/ALogP	1.49
Phase 25°C	solid	Rotatable bonds	4	Predicted density	1.133 g/cm <sup>3</sup>
SMILES	O=Cc1c(OC)c(OC)c(OC)c1OC				
InChIKey	CRBZVDLXAIFERF-UHFFFAOYSA-N				

#### Solubility Data

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
THF	0.14	0.01	3.11

### 2,6-dichlorobenzaldehyde C<sub>7</sub>H<sub>4</sub>Cl<sub>2</sub>O<sup>20, 205</sup>



#### Compound Data

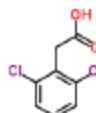
Molecular weight	175.012	H bond acceptors	1	Rule of 5 violations	0
Compound type	aldehyde	H bond donors	0	ACD/ALogP	3.03
Phase 25°C	solid	Rotatable bonds	1	Predicted density	1.4 g/cm <sup>3</sup>
SMILES	O=Cc1c(Cl)cccc1Cl				
InChIKey	DMYKWFEPFPTPY-UHFFFAOYSA-N				

#### Solubility Data

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
acetonitrile	1.35	0.08	38.05
chloroform	3.41	0.32	69.35
ethanol	*	*	*
methanol	*	*	*
THF	2.48	0.22	69.58
toluene	1.74	0.19	44.68

\* This aldehyde reacts with alcohols to form a hemiacetal.

### 2,6-dichlorophenylacetic acid C<sub>8</sub>H<sub>6</sub>Cl<sub>2</sub>O<sub>2</sub><sup>85, 82</sup>



#### Compound Data

Molecular weight	205.038	H bond acceptors	2	Rule of 5 violations	0
Compound type	carboxylic acid	H bond donors	1	ACD/ALogP	2.71
Phase 25°C	solid	Rotatable bonds	2	Predicted density	1.456 g/cm <sup>3</sup>
SMILES	Clc1ccc(Cl)cc1CC(=O)O				
InChIKey	SFAILOQFZNOAU-UHFFFAOYSA-N				

#### Solubility Data

Solvent	Concentration (M)	Mole Fraction (X)	pph (g/100g)
THF	3.03	0.30	120.05

### 3-mercaptopropionic acid C<sub>3</sub>H<sub>4</sub>O<sub>2</sub>S<sup>22</sup>

#### Compound Data

Molecular weight	106.144	H bond acceptors	2	Rule of 5 violations	0
Compound type	carboxylic acid	H bond donors	1	ACD/ALogP	0.43

16



# Reaction Attempts

The UsefulChem Project

## Find Reactions via Dropdown Filters

Component: 1-pyrenebutyric acid

Component: 2,4-dimethylbenzaldehyde

Component: Select Reactant or Named Product

## Find Reactions via Substructure Search

SMARTS

everything  reactants only  products only

Substructure Search

## 1 Reactions Found

Reaction ID UCEXP181-087V2

Researcher: Khalid Mirza

Reaction Type: Ugi

Solvent: methanol

Limiting Reactant: 0.222 M

Precipitate: Reactants Insoluble

Comments: Protocol from Explan005 followed

Reference: <http://usefulchem.wikispaces.com/Exp181>

Solvent Selection: [Optimal Solvent Prediction](#)

OPEN  
NOTEBOOK  
SCIENCE



## Reaction Attempts: Advanced Search

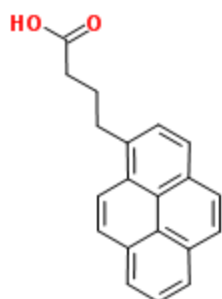
Compound: Any compound

Solvent: Any Solvent

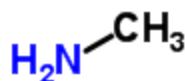
Reaction Type: Any Reaction Type

Researcher: Any Researcher

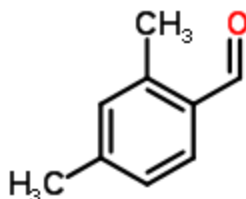
Search



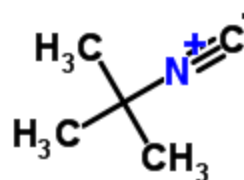
1-pyrenebutyric acid



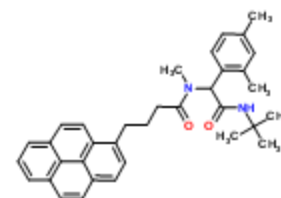
methylamine



2,4-dimethylbenzaldehyde



t-butyl isocyanide



21106206

# Reaction Attempts Book

## Reaction Attempts

By Andrew Lang, By Jean-Claude Bradley, By Tim Bohinski, By David Bulger, By Khalid Mirza, By Marshall Moritz, By Alicia Merchant, By Emily Messner, By Shannon Oseback, By Rikesh Parikh, By Mitesh Shah

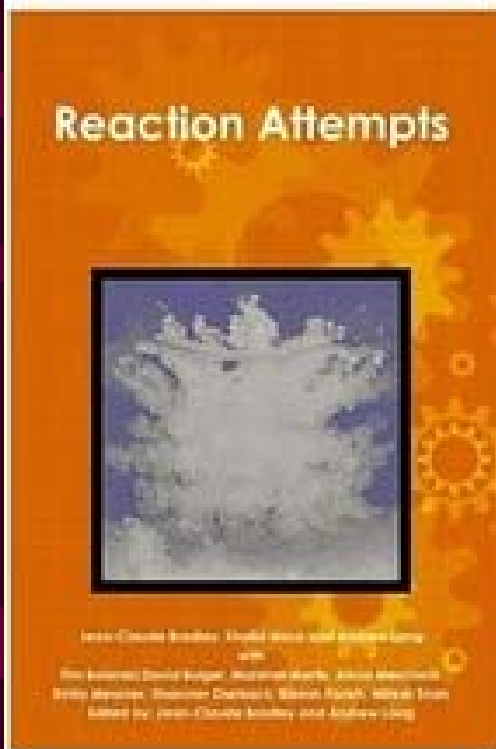
★★★★★ (1 Rating)

Paperback, 582 pages

**\$26.28**

Ships in 3-5 business days

A compilation of reaction attempts from the UsefulChem project.



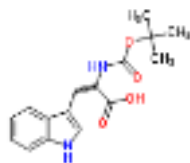
# Reaction Attempts Book: Reactants listed Alphabetically

Limiting Reactant 0.61 M

Precipitate No

Comments Protocol from Explan005 followed

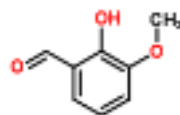
Reference <http://usefulchem.wikispaces.com/Exp131>



Boc-Tip-OH



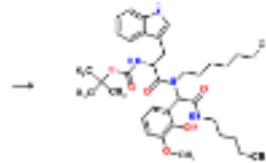
1-heptylamine



o-vanillin



1-pentyl isocyanide



21106066

Reaction ID UCEXP148A-V1B153

Researcher Khalid Mirza/ Emily Messner

Reaction Type Ugi

Solvent methanol

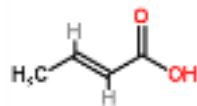
Limiting Reactant 0.398 M

Precipitate Yes but the aldehyde alone is not soluble in methanol

Yield 50.1 %

Comments Protocol from Explan005 followed

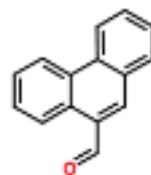
Reference <http://usefulchem.wikispaces.com/Exp148>



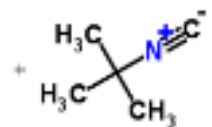
crotonic acid



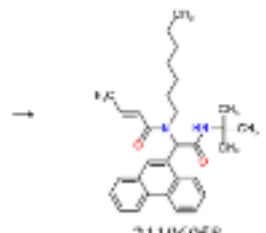
1-heptylamine



phenanthrene-9-carboxaldehyde



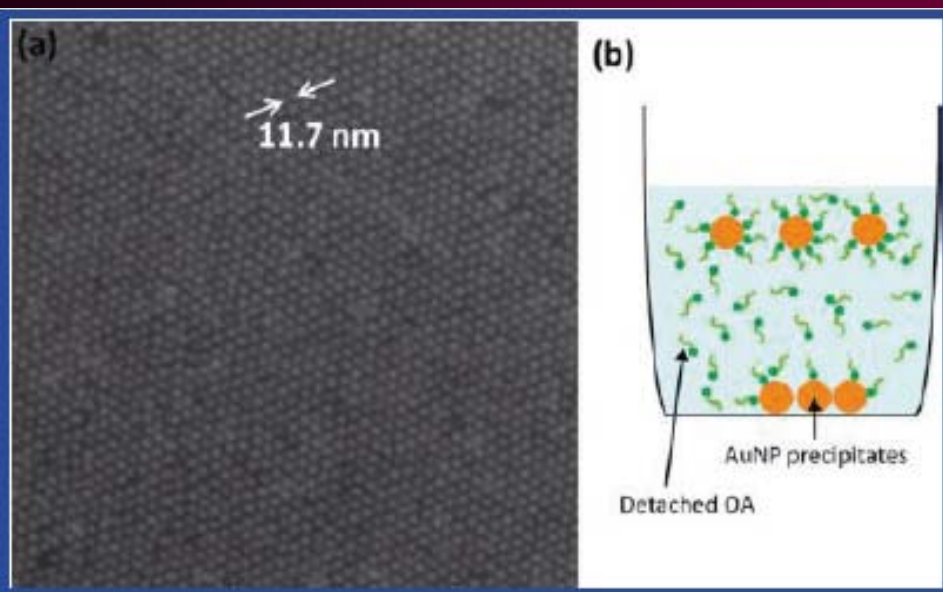
t-butyl isocyanide



21106058



# ONS Challenge Solubility Book cited for nanotechnology application



Although to our best knowledge there lacks literature value of OA solubility in the two solvents, the 10-fold better solubility of 1-octadecylamine (*sic*), the saturated version of oleylamine, in toluene than hexane is in line with our hypothesis. (33) This increased solubility caused the OA molecules that were originally attached to the AuNPs to gradually detach from the AuNPs, which is supported by our observations in poor AuNP stability and surface-pressure isotherms.

## Langmuir

### Enhanced Ordering in Gold Nanoparticles Self-Assembly through Excess Free Ligands

[dx.doi.org/10.1021/la104786z](https://doi.org/10.1021/la104786z)

Cindy Y. Lau,<sup>†</sup> Huigao Duan,<sup>†</sup> Fuke Wang,<sup>†</sup> Chao Bin He,<sup>†</sup> Hong Yee Low,<sup>†</sup> and Joel K. W. Yang<sup>\*,†</sup>

# Lulu.com Data Disks

<http://www.lulu.com/content/compact-disc/8337307>

Help Need help? Visit our support page.



Welcome back, Andrew Lang

Log Out



Publish

Buy

Services

Community

My Lulu

Support

Lulu Demo >

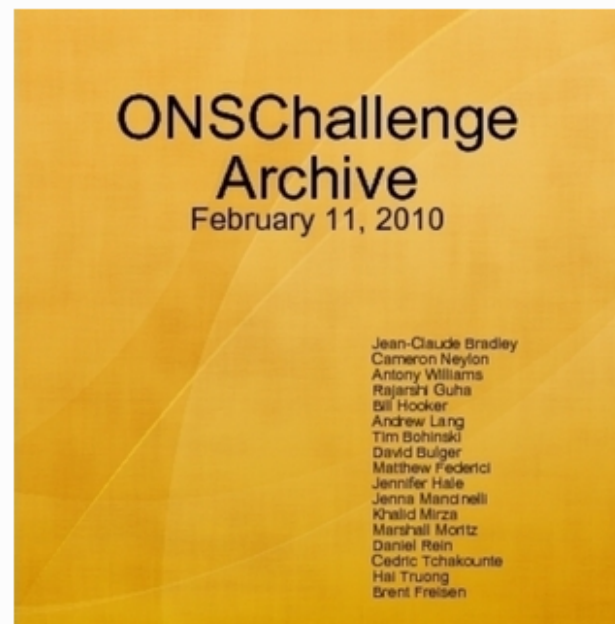
## Lulu Marketplace

### ONSchallenge Archive

by Andrew Lang et al.

★★★★★ (1 Rating)

CD



Jean-Claude Bradley  
Cameron Neylon  
Antony Williams  
Rajarshi Guha  
Bill Hooker  
Andrew Lang  
Tim Bohinski  
David Bulger  
Matthew Federici  
Jennifer Hale  
Jenna Mandrelli  
Khalid Mirza  
Marshall Moritz  
Daniel Rein  
Cedric Tchakourte  
Hai Truong  
Brent Freisen

**\$5.50**

Ships in 3-5 business days

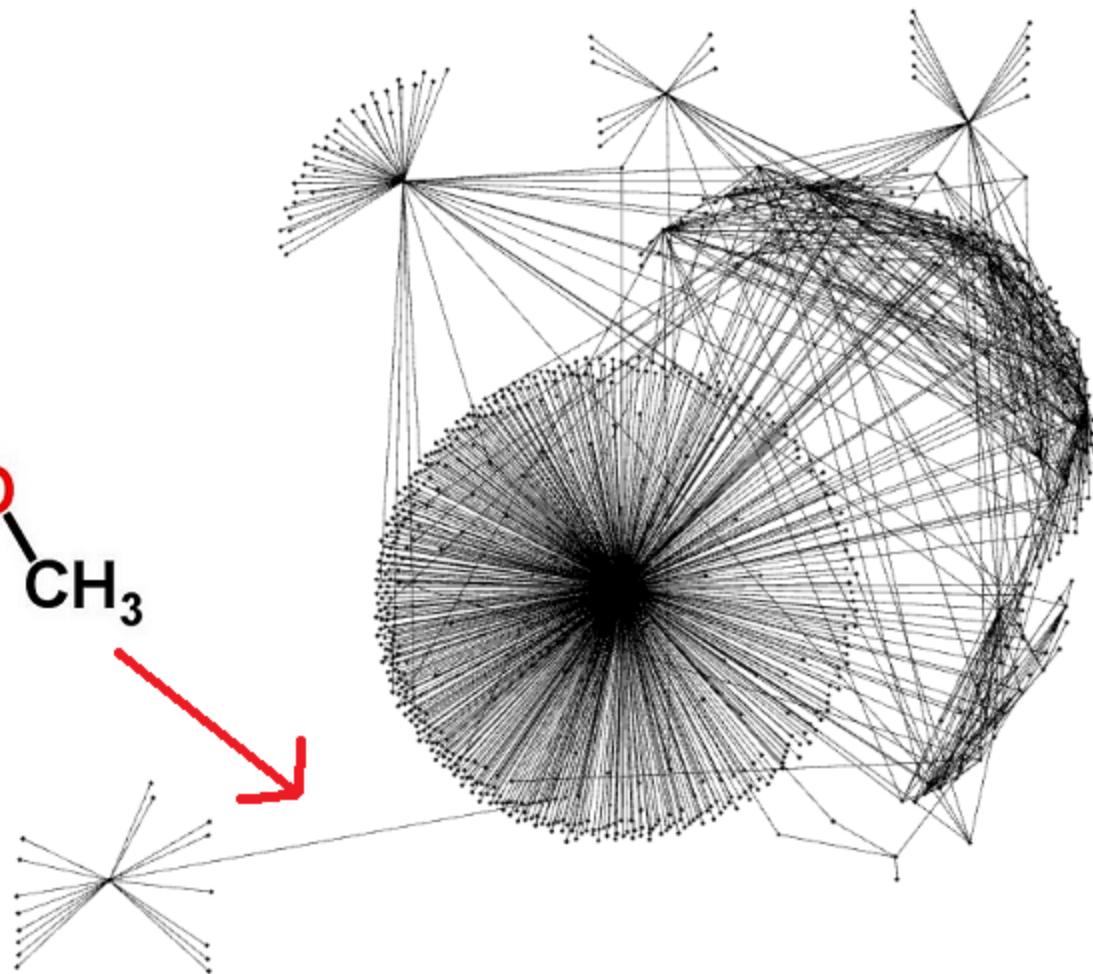
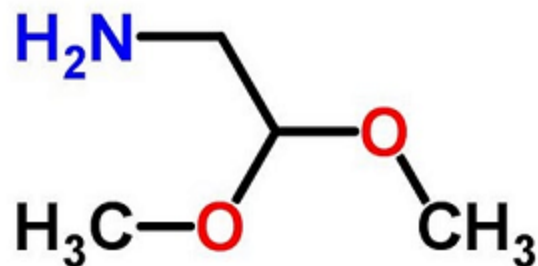
Open Notebook Science Challenge Data and Lab Notebooks with supporting Media, Excel Spreadsheets and Spectral Data.

Add to Cart

Email



# Visualizing molecule-researcher connection maps reveals link between 2 Open Notebooks (Todd and Bradley)



<http://usefulchem.blogspot.com/2010/12/visualizing-social-networks-in-open.html>

(Don Pellegrino)

# All ONS web services

## Open Notebook Science Web Services

maintained by [Jean-Claude Bradley](#) and [Andrew Lang](#)

[Solubility](#)

[Reactions](#)

[Chemical Information Validation](#)

[Melting Points](#)

[MyExperiment](#)

[Mendeley](#)

[NMR](#)

[Miscellaneous](#)

<http://onswebservices.wikispaces.com/>

# For all Formats of ONS Projects

## Open Notebook Science Solubility Challenge

### Live version

Book - HTML ([entire](#) [predictions only](#))

Spreadsheet - ([SolubilitiesSum](#))

OData XML feed - ([Solubilities](#))

### Edition 3 (2010-02-11)

Book - PDF and shippable physical copy ([LuLu](#))

Book - PDF only ([Nature Precedings](#))

Data Archive - CD ([LuLu](#))

Data Archive - ZIP ([Drexel Chem Server](#)) ([Drexel Library iDEA Repository](#))

Data Archive - hosted interactive ([Drexel Chem Server](#))

[UsefulChem post](#)

## Reaction Attempts

### Live version

Book - HTML ([entire](#) [selected reactants](#))

Spreadsheets - linked via RXIDs ([Reactions](#) [ReactionCompounds](#))

OData XML feeds - linked via RXIDs ([Reactions](#) [ReactionCompounds](#))

### Edition 1 (2010-04-27) Data Source: UsefulChem Project

Book - PDF and shippable physical copy ([LuLu](#))

Book - PDF only ([Scribd](#)) ([Drexel Library iDEA repository](#)) ([Nature Precedings](#))

UsefulChem Data Archive - DVD ([LuLu](#))

UsefulChem Data Archive - ZIP ([Drexel Chem Server](#))

UsefulChem Data Archive - hosted interactive ([Drexel Chem Server](#))

Reaction Attempts RXIDs spreadsheet - XLS ([WebCite](#))

Reaction Attempts Reactants and Products spreadsheet XLS ([WebCite](#))

[UsefulChem Blog post](#)

<http://onsbooks.wikispaces.com>



# Conclusions

- Our current system of publication is not as transparent as it could be
- Open Notebook Science offers an efficient way to make research transparent and discoverable