February 2021

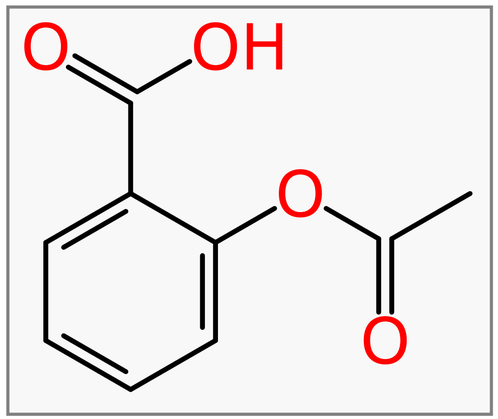
Mixtures

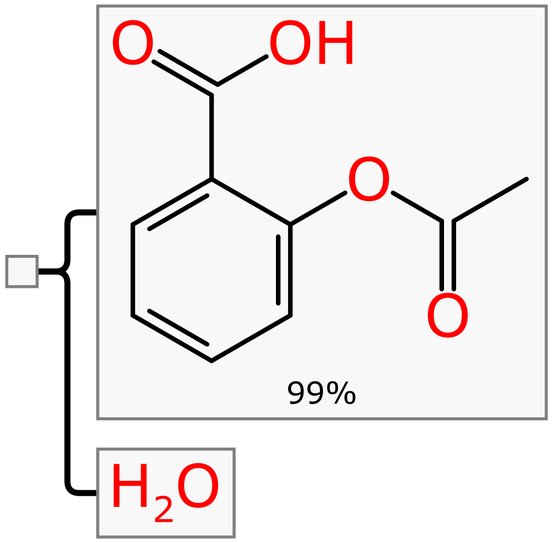
as first class citizens in the realm of informatics

Alex M. Clark

alex@collaborativedrug.com

Premise





Molfile

InChI

1980's

Mixfile

MInChI

2020's

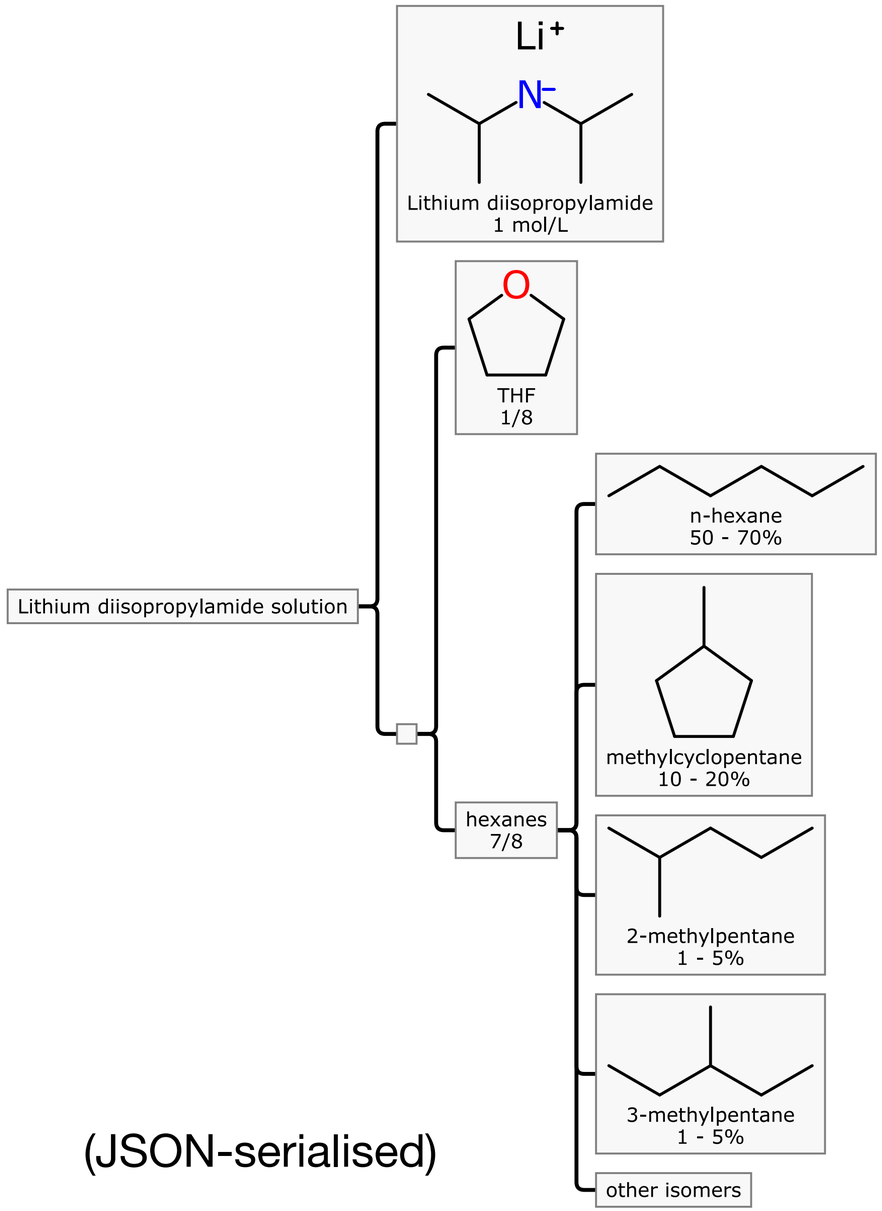
- Most mixtures stored as text or custom table layouts
- Value of upgrading to cheminformatics is well established...
- ... with the right datastructure, always just one script away from what you need
- If you can represent it, you can model it

CDD VAULT® Complexity Simplified

InChI TRUST

IUPAC INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

Mixfile/MInChI

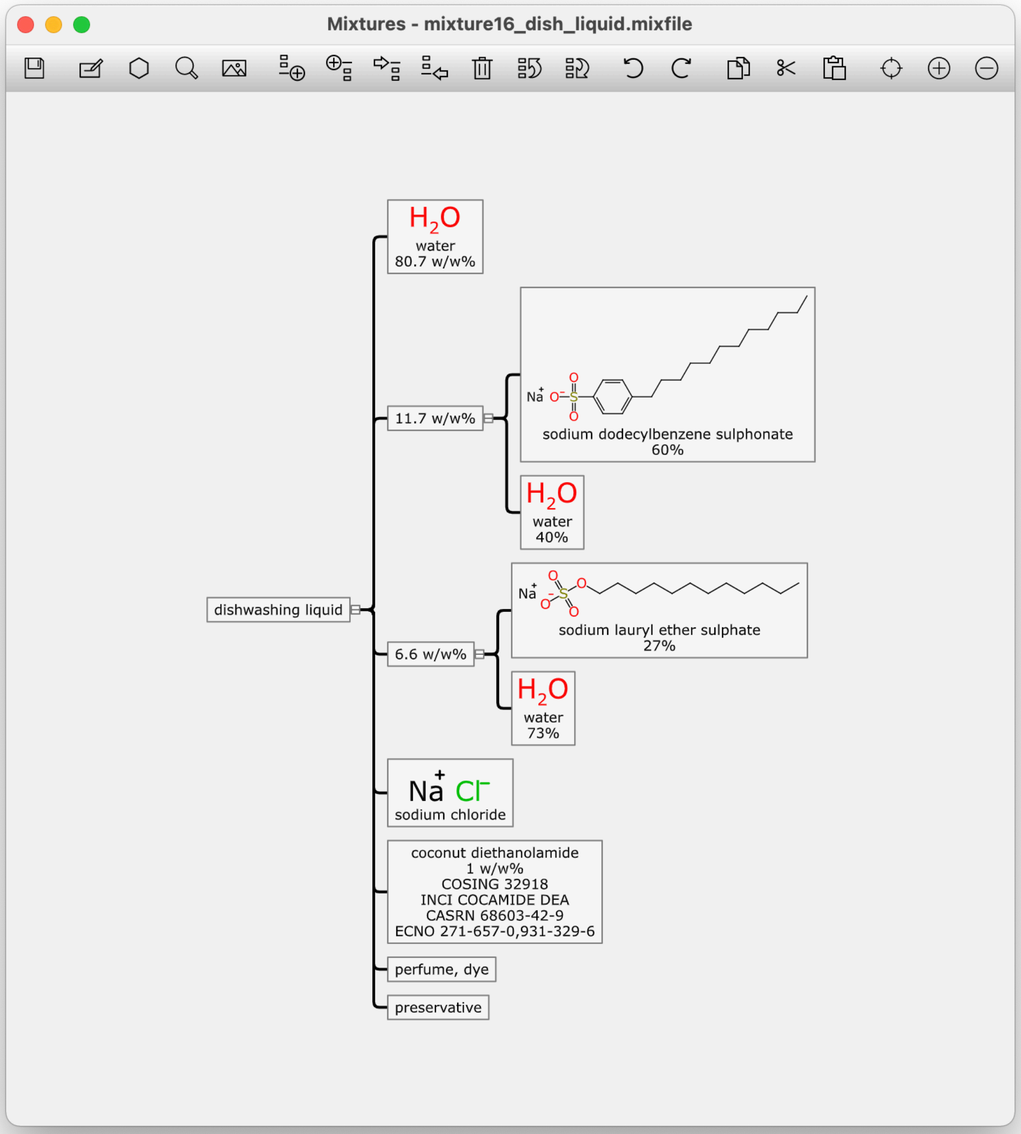


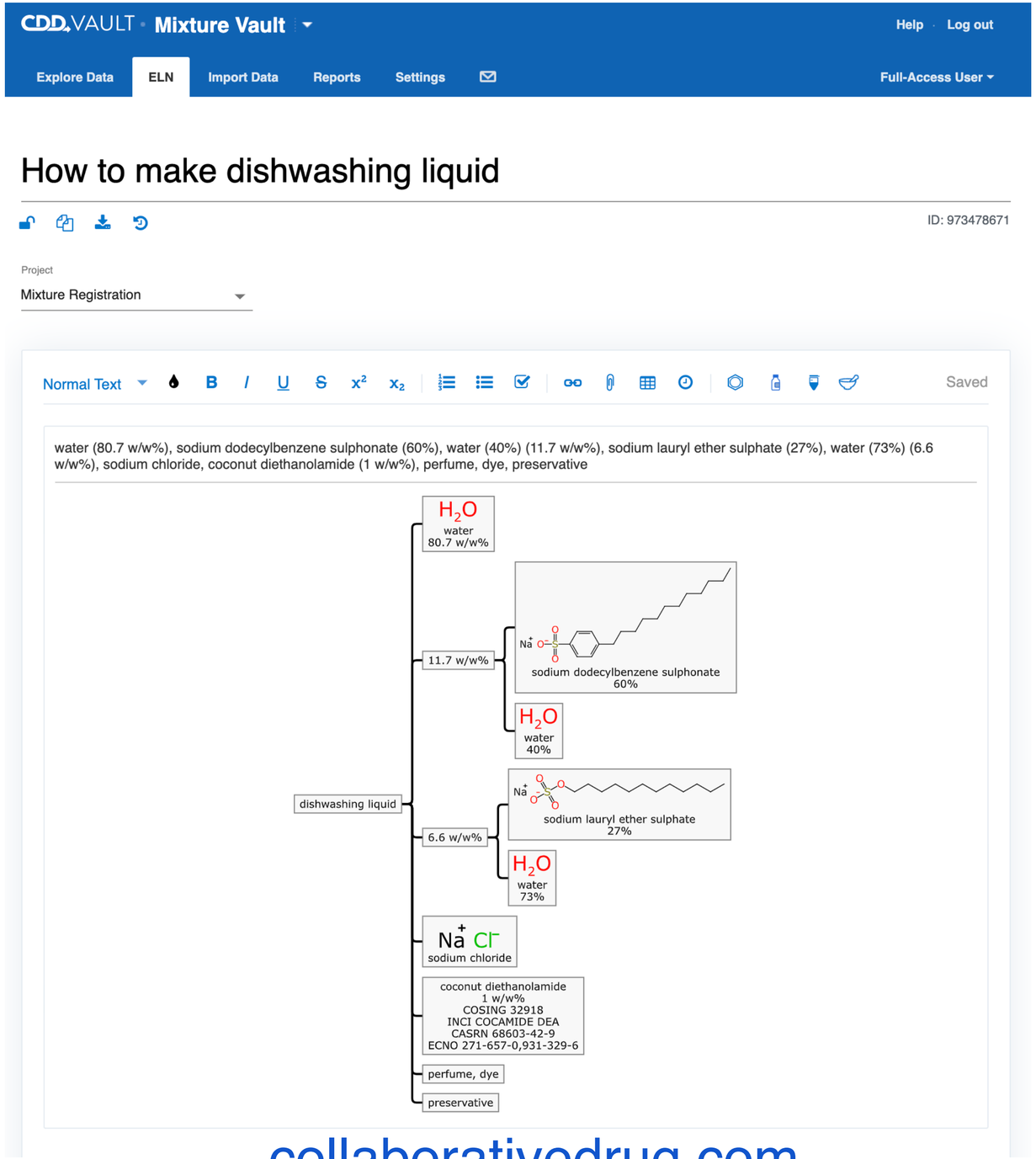
(JSON-serialised)

- Format needs to be:
  - hierarchical
  - embed structures when possible
  - include concentration information
  - tolerate uncertainty
- More verbose ELN-friendly form is **Mixfile**
- Concise form with canonical components is **MInChI** (*mixtures* InChI)

```
MInChI=0.00.1S/C4H8O/c1-2-4-5-3-1/h1-4H2&C6H12/
c1-6-4-2-3-5-6/h6H,2-5H2,1H3&C6H14/c1-3-5-6-4-2/
h3-6H2,1-2H3&C6H14/c1-4-5-6(2)3/h6H,4-5H2,1-3H3&C6H14/
c1-4-6(3)5-2/h6H,4-5H2,1-3H3&C6H14N.Li/c1-5(2)7-6(3)4;/
h5-6H,1-4H3;/q-1;+1/n{6&{1&{3&2&4&5}}}/
g{1mr0&{1vp0&{5:7pp1&1:2pp1&1:5pp0&1:5pp0}7vp0}}
```

Data Creation



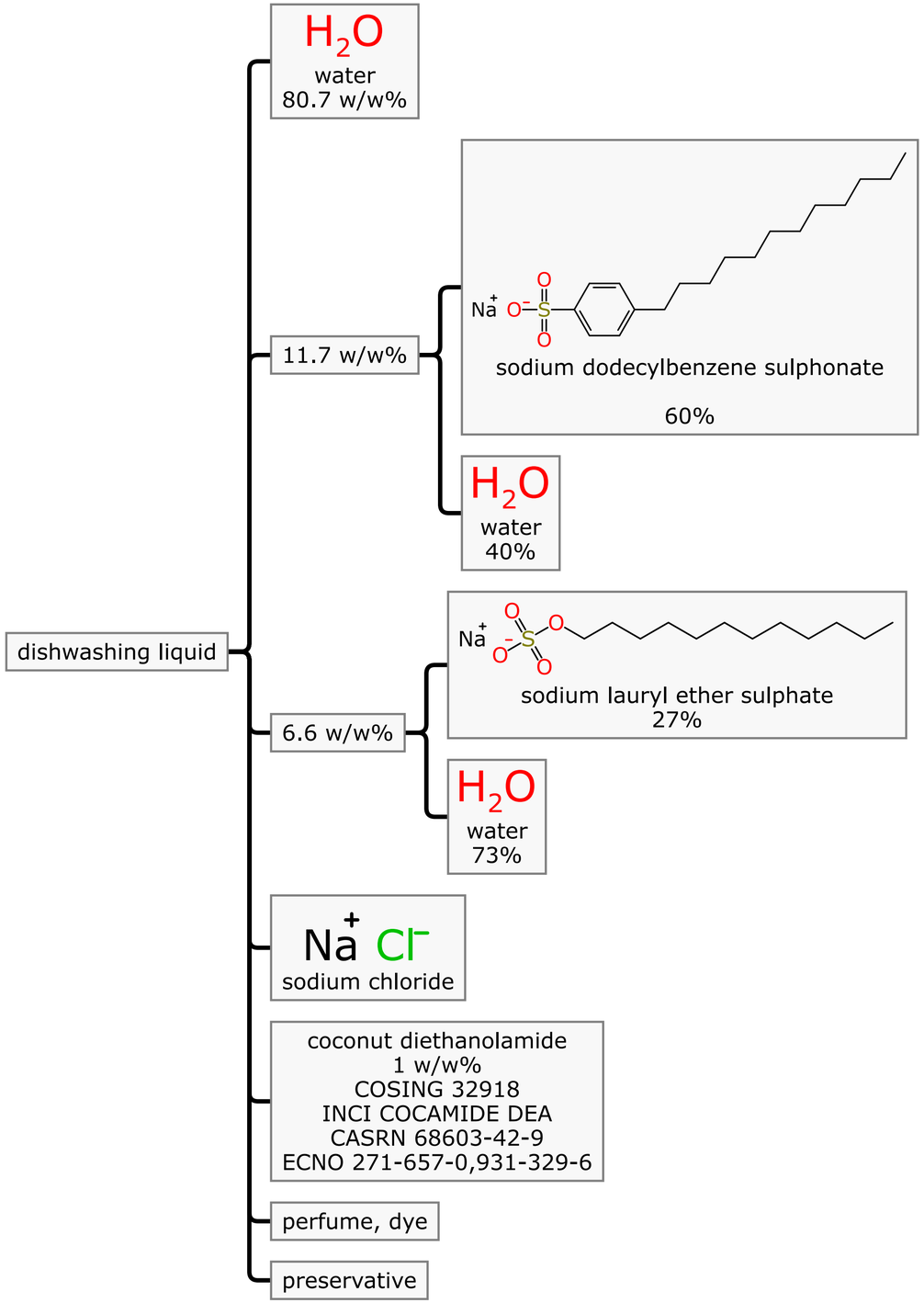


github.com/cdd/mixtures

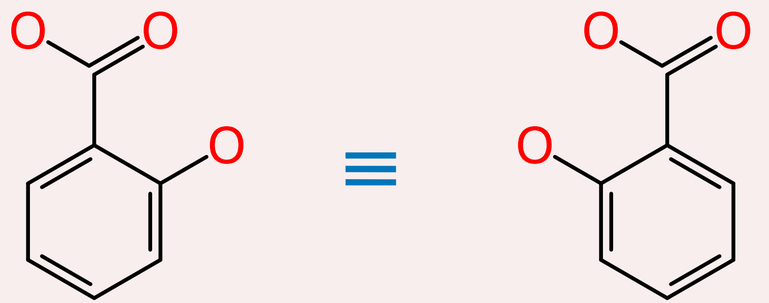
collaborativedrug.com

Formulation Example

- Many consumer products are well described from a chemical perspective
- Some components are more easily defined than others
- When structure is not available, can use external identifiers
- Hierarchy encodes information about the design of the product
- Concentrations can be expressed with uncertainties

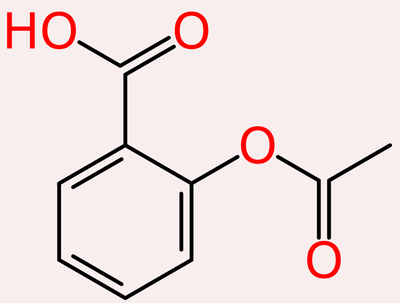


Comparisons with Structures

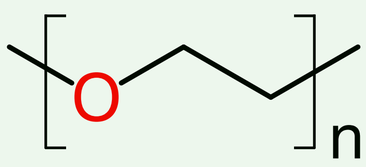


≡

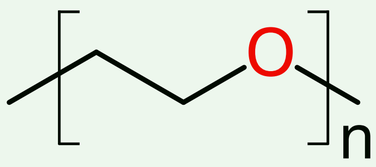
substructure of



InChI=1S/C7H6O3/c8-6-4-2-1-3-5(6)7(9)10/h1-4,8H,(H,9,10)



≅

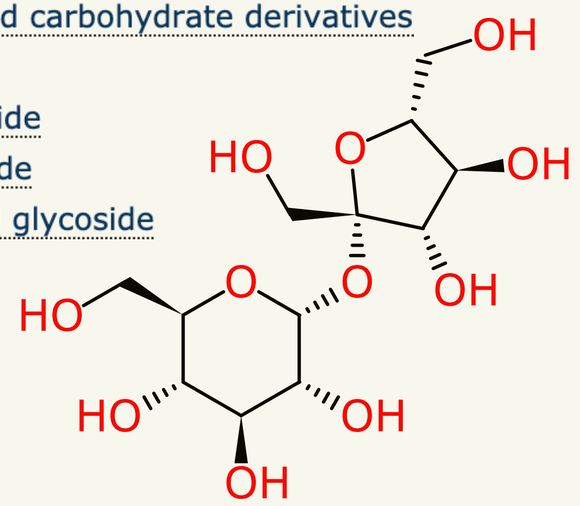


MW 300-500 MW 400-700

$Y_{1.2}Ba_{0.8}CuO_4$ ≅ $YBa_2Cu_3O_{7-\delta}$

- CHEBI:36962 organochalcogen compound
  - CHEBI:36963 organooxygen compound
    - CHEBI:78616 carbohydrates and carbohydrate derivatives
      - CHEBI:16646 carbohydrate
        - CHEBI:50699 oligosaccharide
          - CHEBI:36233 disaccharide
            - CHEBI:24407 glycosyl glycoside

≅



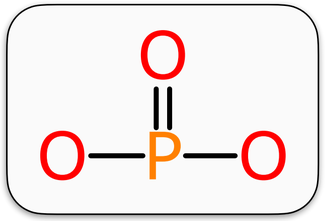
Search Queries

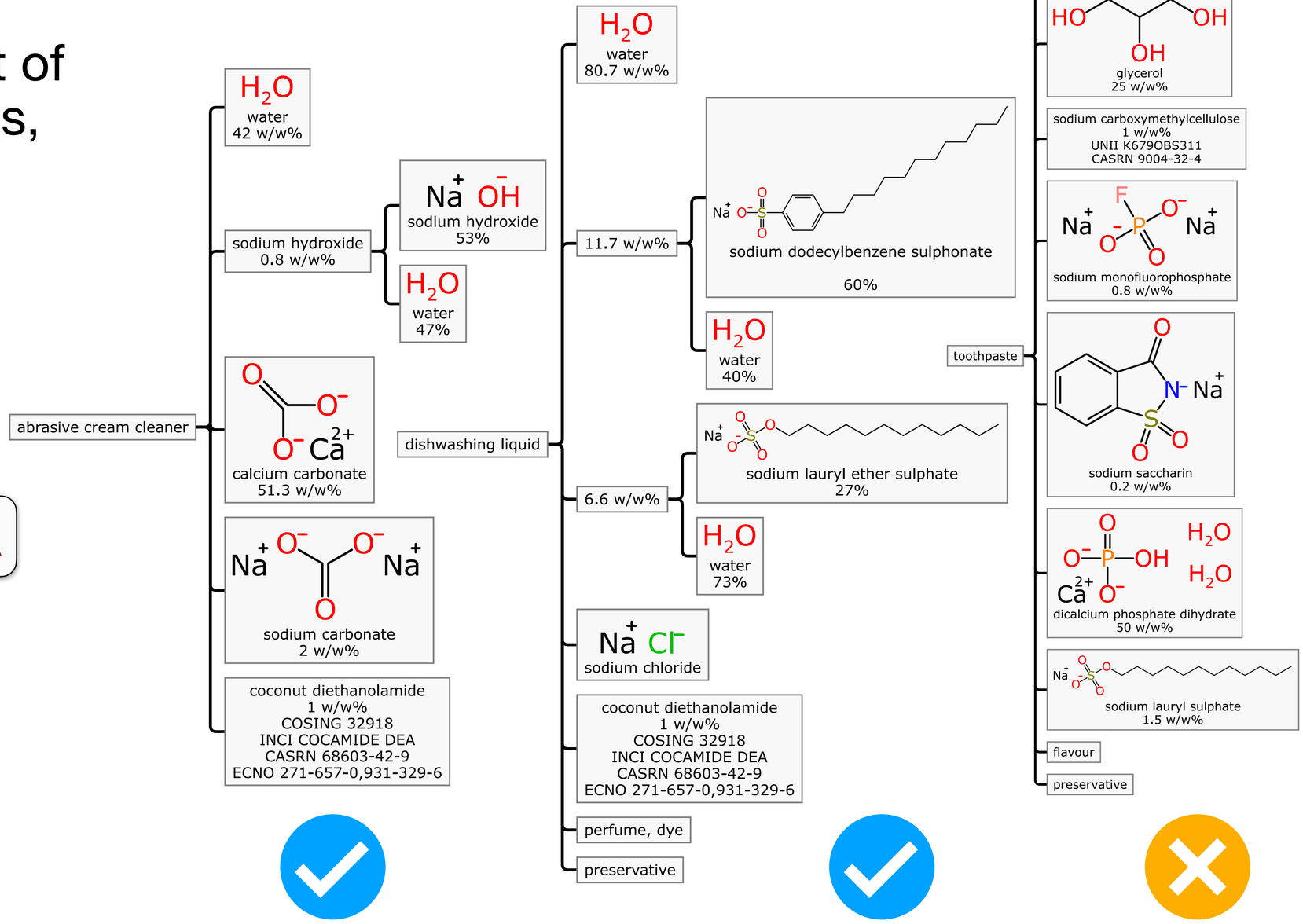
- Looking for a certain subset of external cleaning surfactants, phosphate-free

has H_2O >40%

has **INCI**: COCAMIDE DEA

has not substructure

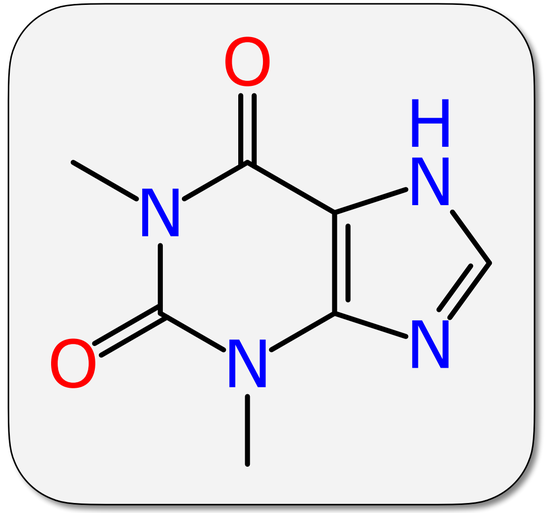




Informatics Example

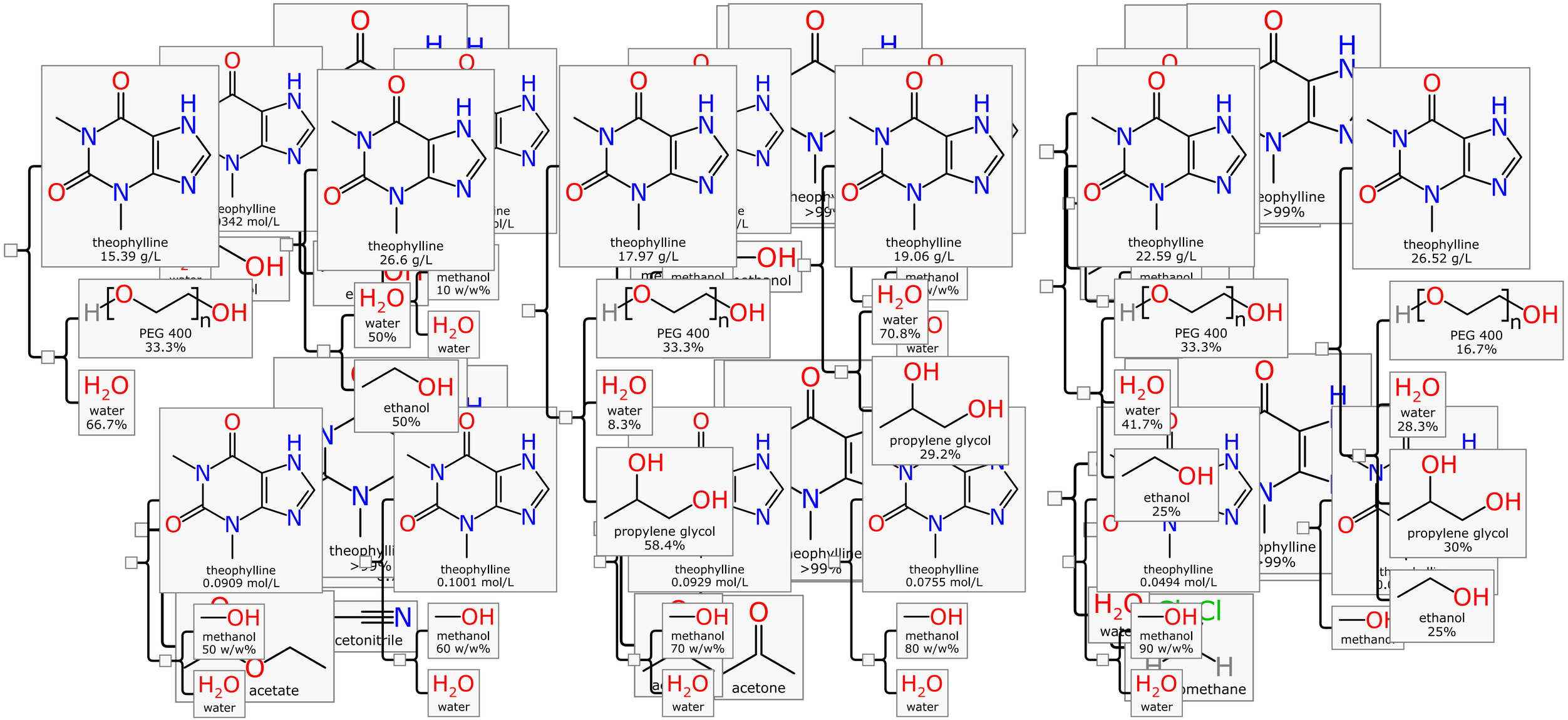
https://tinyurl.com/
y3svytfp
➪ 4:13:00

- Solubility of theophylline
- Often delivered in liquid form with mixed solvents: optimising proportion of drug is important
- Consider a scenario where:
  - all data was provided in *Mixtures InChI* form
  - these data exist in openly available repositories
- Query:
  - check that *theophylline* is present and has concentration
  - check that other ingredients are *solvents*
- Consider 4 papers with relevant solubility, published over 20 years...



theophylline
nasal anti-inflammatory

Papers (x4)


theophylline
15.39 g/L
PEG 400
33.3%
water
66.7%
theophylline
26.6 g/L
water
50%
methanol
10 w/w%
water
ethanol
50%
theophylline
0.0909 mol/L
theophylline
0.1001 mol/L
methanol
50 w/w%
acetonitrile
methanol
60 w/w%
water
acetate
water
theophylline
17.97 g/L
PEG 400
33.3%
water
8.3%
propylene glycol
58.4%
theophylline
0.0929 mol/L
theophylline
19.06 g/L
water
70.8%
propylene glycol
29.2%
theophylline
0.0755 mol/L
methanol
70 w/w%
acetone
water
methanol
80 w/w%
water
theophylline
22.59 g/L
PEG 400
33.3%
water
41.7%
ethanol
25%
theophylline
0.0494 mol/L
methanol
90 w/w%
water
theophylline
26.52 g/L
PEG 400
16.7%
water
28.3%
propylene glycol
30%
ethanol
25%

All Together for QSAR

Solubility	H_2O	—OH	OH	OH	OH	OH OH	H[O]$_n$OH	O	O O	N	Cl Cl H H
0.699		1									
15.19			1								
1.04					1						
3.142								1			
0.784										1	
0.91											1
6.3	1										
13.7		1									
11.6			1								
13.58				1							
6.73									1		
9.3								1			
8.20	0.8	0.2									
16.38	0.5	0.5									
13.60	0.2	0.8									
	(+8 more similar)										
15.39	0.333						0.667				
26.6	0.5		0.5								
17.97	0.083					0.584	0.333				
19.06	0.708					0.292					
22.59	0.417		0.25				0.333				
26.52	0.283		0.25			0.3	0.167				
	(+14 more similar)										

Deep Eutectics / Carbon Capture

- Mixtures of ionic & neutral solvents can absorb gases like CO_2

energy&fuels Article
pubs.acs.org/EF
DOI: 10.1021/ef5028873

Deep Eutectic Solvents: Physicochemical Properties and Gas Separation Applications

Gregorio García,[†] Santiago Aparicio,*[,†] Ruh Ullah,[‡] and Mert Atilhan*[,‡]

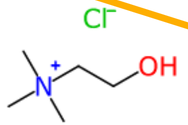
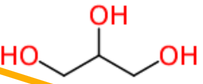
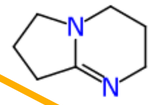
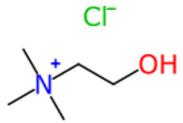
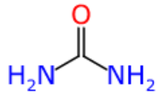
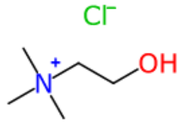
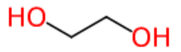
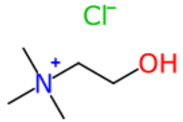
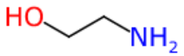
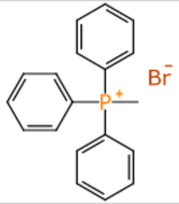
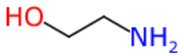
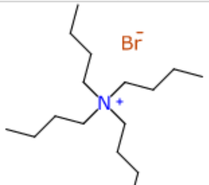
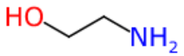
- Want to model?
  - finding data in literature is hard
  - curating is extremely laborious
- Each datapoint is a mixture...

Table 10. Summary of CO_2 and SO_2 Solubilities in DESs and in Some ILs for Comparison Purposes[a]

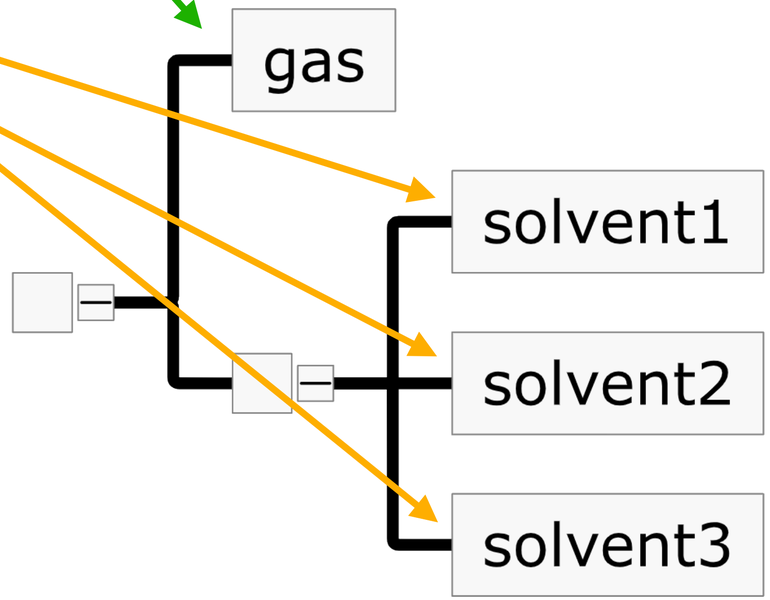
DES components						
IL	HBD	molar ratio	absorbate	solubility	T/P (K/bar)	ref
choline chloride	glycerol + DBN	1:2:6 CH_Cl:gly:DBN	CO_2	2.3−2.4 mmol/g	ambient	225
choline chloride	urea	1:2	CO_2	3.559 mmol/g	303.15/60	226
choline chloride	ethylene glycol	1:2	CO_2	3.1265 mmol/g	303.15/58.63	227
choline chloride	ethanolamine	1:6	CO_2	0.0749 mmol/g	298/10	228
MTPP_Br	ethanolamine		CO_2	0.0716 mmol/g	298/10	
TBA_Br	ethanolamine		CO_2	0.0591 mmol/g	298/10	
[BMIM][PF_6]	–	–	CO_2	0.200 mol/mol	298/12.99	204
[EMIM][BF_4]	–	–	CO_2	1.5999 mmol/g	298/41.55	229
choline chloride	triethylene glycol	4:1	CO_2	0.1941 mmol/g	293/5	230
choline chloride	phenol	4:1	CO_2	0.2108 mmol/g	293/5	
choline chloride	diethylene glycol	4:1	CO_2	0.1852 mol/mol	293/5	
choline chloride	glycerol	1:1	SO_2	0.678 g/g	290/1	231
[DMEA][glutarate] (2:1)	–	–	SO_2	0.623 mol/mol	313/0.004	232
CPL	KSCN	3:1	SO_2	1.38 mol/mol	313/1	233
acetamide	KSCN	3:1	SO_2	0.588 g/g	293/1	
acetamide	NH_4SCN	3:1		0.579 g/g	293/1	
CPL	NH_4SCN	3:1	SO_2	0.559 mol/mol	293/1	
urea	NH_4SCN	3:2	SO_2	0.372 mol/mol	303.1	
triethylene glycol (PEG150)	DBU	1:1	CO_2	1.04 mol/mol	298/1.0	234
[bmim][Tf_2N]	DBU	1:1	CO_2	1.0 mol/mol	298/1.0	235
[N_{2224}][CA]	H_2O	1:*n*	CO_2	0.66 mol/mol	298/4.4	236
choline chloride	glycerol	1:2	CO_2	3.692 mol/kg	303/58	237
[CPL][TBAB]		1:1	SO_2	0.680 mol/mol	298/1	238
[CPL][TBAB] + H_2O		1:1:4	SO_2	0.52 g/g	293/1.0	239
CPL−acetamide		1:1	SO_2	0.497 g/g	303/1	240
CPL−imidazole		1:1	SO_2	0.624 g/g	300/1	
[HMIM][Tf_2N]	–	–	SO_2	0.844 mol/mol	298/2.94	220
[hmpy][Tf_2N]	–	–	SO_2	0.844 mol/mol	298/2.96	
choline chloride	2,3-butanediol	1:4	CO_2	0.0188 mol/mol	298/5.08	241
choline chloride	1,4-butanediol	1:3	CO_2	0.0164 mol/mol	298/5.09	
choline chloride	1,2-propanediol	1:3	CO_2	0.0165 mol/mol	298/5.14	
choline chloride	lactic acid	1:2	CO_2	0.0248 mol/mol	348/19.27	242
choline chloride	urea	1:1.5	CO_2	0.201 mol/mol	313.15/118.4	243
choline chloride	urea	1:2	CO_2	0.309 mol/mol	313.15/1125	
choline chloride	urea	1:2.5	CO_2	0.203 mol/mol	313.15/1145	
choline chloride	urea + H_2O	50 wt % CH_Cl + urea (reline) + 50% H_2O	CO_2	0.111 mol/mol	313/7.8	244
choline chloride	urea + H_2O	60 wt % reline + 40% H_2O	CO_2	0.103 mol/mol	313/8.06	
choline chloride	urea + H_2O	70 wt % rel + 30% H_2O	CO_2	0.097 mol/mol	313/8.09	
choline chloride	urea + H_2O + MEA	50 wt % reline + 15 wt % MEA + H_2O	CO_2	0.229 mol/mol	313/8.18	
choline chloride	urea + H_2O + MEA	60 wt % reline + 10 wt % MEA + H_2O	CO_2	0.202 mol/mol	313/8.25	
choline chloride	urea + H_2O + MEA	70 wt % reline + 5 wt % MEA + H_2O	CO_2	0.189 mol/mol	313/8.13	

[a]Acronyms: CH_Cl, choline chloride; DBN, 1,5-diazabicyclo[4.3.0]non-5-ene; MTPP_Br, methyltriphenylphosphonium bromide; TBA_Br, tetrabutylammonium bromide; [BMIM][PF_6], 1-butyl-3-methylimidazolium hexafluorophosphate; [EMIM][BF_4], 1-ethyl-3-methylimidazolium tetrafluoroborate; [DMEA][glutarate], dimethylethanolammonium glutarate; CPL, caprolactam; DBU, diazabicyclo[5.4.0]undec-7-ene; [bmim][Tf_2N], 3-butyl-1-methylimidazolium bis(trifluoromethanesulfonyl)amide; [N_{2224}][CA], triethylbutylammonium carboxylate; [CPL][TBAB], caprolactam tetrabutylammonium bromide; [HMIM][Tf_2N], 1-*n*-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide; [hmpy][Tf_2N], 1-*n*-hexyl-3-methylpyridinium bis(trifluoromethylsulfonyl)imide.

Mixturfication

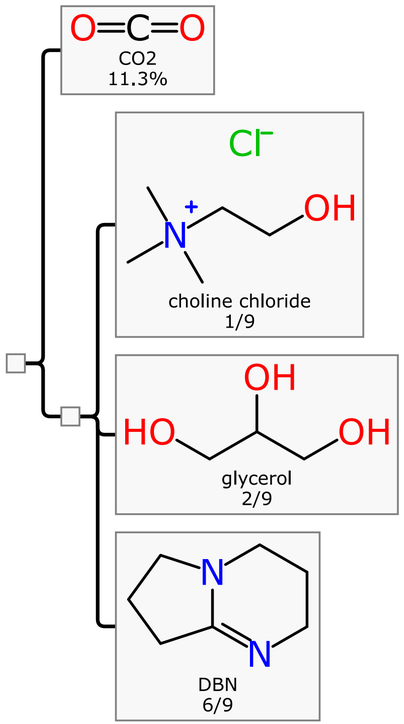
	Molecule1	Name1	Molecule2	Name2	Molecule3	Name3	Ratio1	Rati..	Ratio3	Gas	GasN..	Dissolve
1		choline chloride		glycerol		DBN	1	2	6	O=C=O	CO2	= 11.3 %
2		choline chloride		urea			1	2		O=C=O	CO2	= 17.1 %
3		choline chloride		ethylene glycol			1	2		O=C=O	CO2	= 15 %
4		choline chloride		ethanolamine			1	6		O=C=O	CO2	= 0.37 %
5		MTPP_Br		ethanolamine			1	6		O=C=O	CO2	= 0.34 %
6		TBA_Br		ethanolamine			1	6		O=C=O	CO2	= 0.28 %

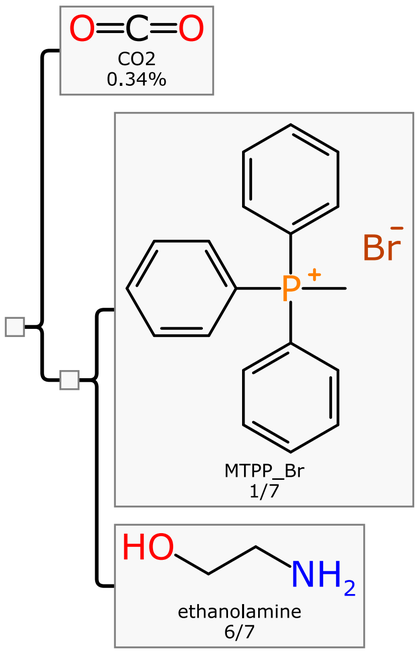
- Data entry easier as tabular content
- Mixture hierarchy is implicit

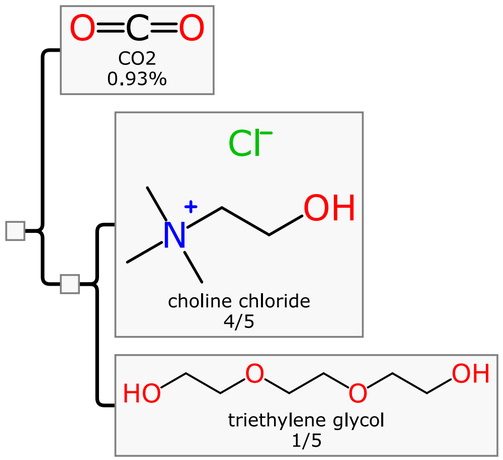


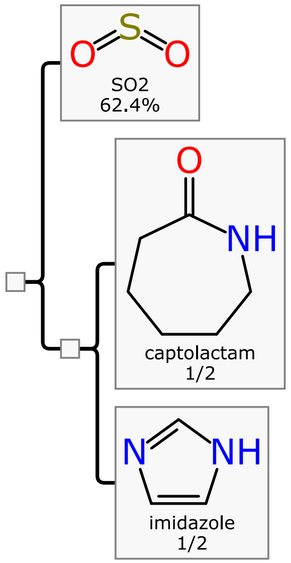
Mixtures

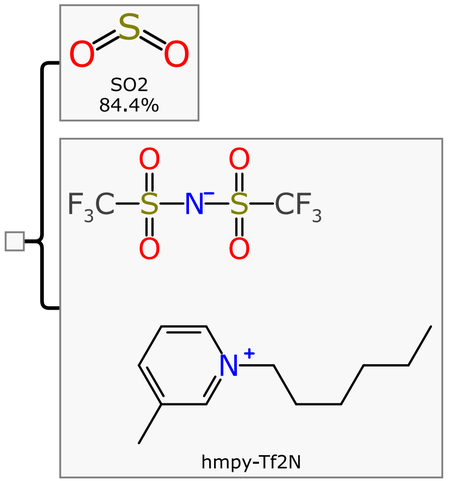
✤ Convert each row into a self-contained Mixfile...









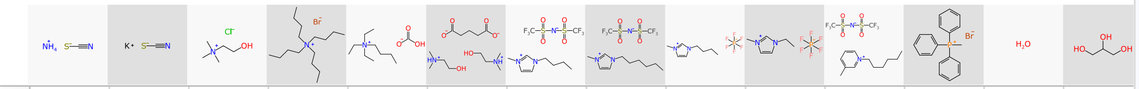


✤ ... have solubility of CO_2 and SO_2 in various solvent combinations

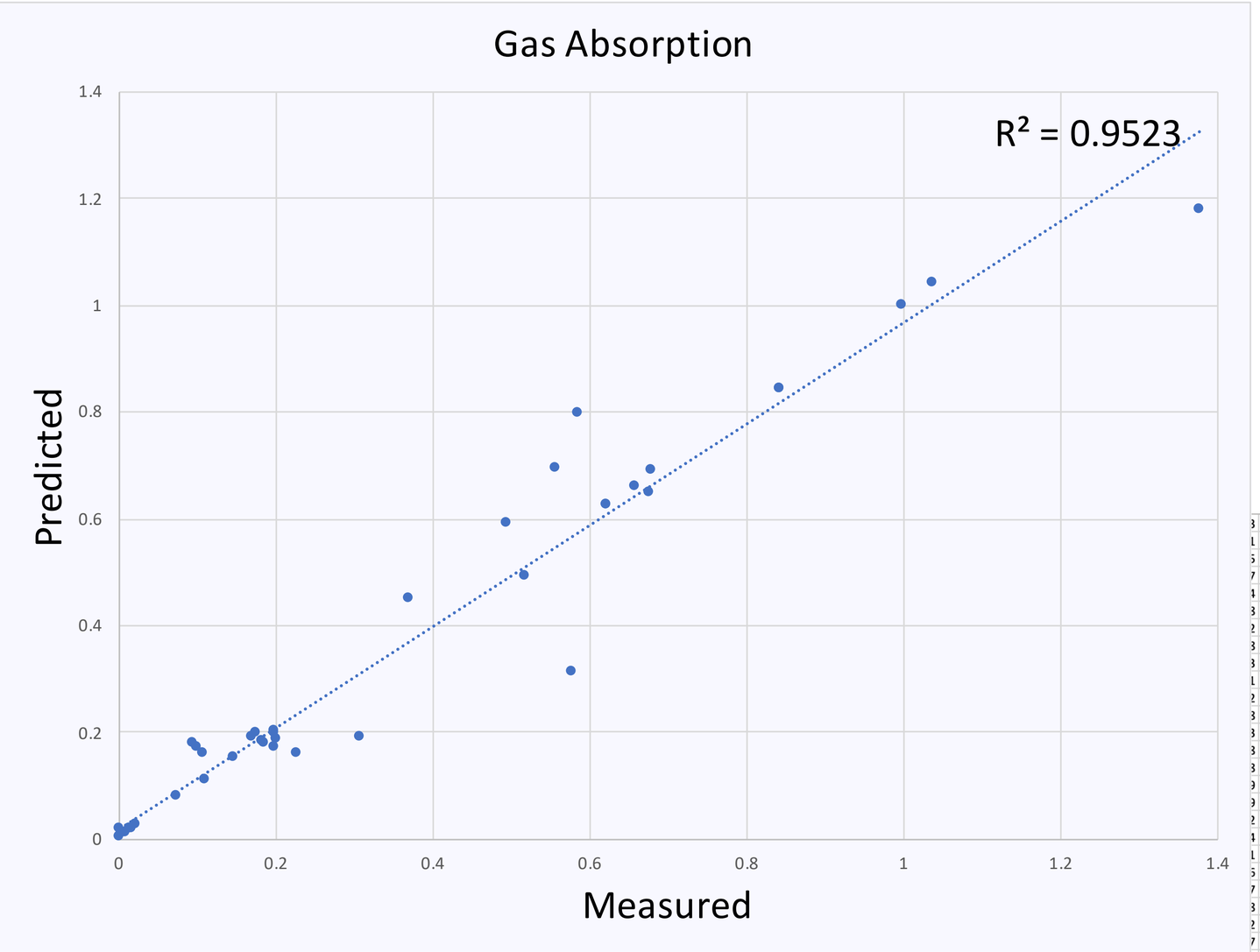
✤ Could be looked up in a database of mixtures from many different sources

Free Wilson-esque

- Tabulate components as column
- Build a basic regression model (F

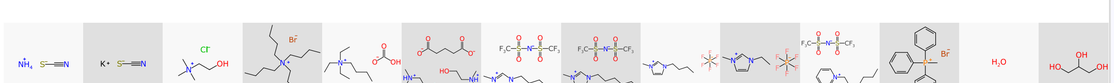


1	0	0	0.1111	0	0	0	0	0	0	0	0	0	0	0.222
2	0	0	0.3333	0	0	0	0	0	0	0	0	0	0	
3	0	0	0.3333	0	0	0	0	0	0	0	0	0	0	
4	0	0	0.1429	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	0	0	0.1429	0	
6	0	0	0	0.1429	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	1	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	1	0	0	0	
9	0	0	0.8	0	0	0	0	0	0	0	0	0	0	
10	0	0	0.8	0	0	0	0	0	0	0	0	0	0	
11	0	0	0.8	0	0	0	0	0	0	0	0	0	0	
12	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0.5
13	0	0	0	0	0	1	0	0	0	0	0	0	0	
14	0	0.25	0	0	0	0	0	0	0	0	0	0	0	
15	0	0.25	0	0	0	0	0	0	0	0	0	0	0	
16	0.25	0	0	0	0	0	0	0	0	0	0	0	0	
17	0.25	0	0	0	0	0	0	0	0	0	0	0	0	
18	0.4	0	0	0	0	0	0	0	0	0	0	0	0	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	0	0	0	0	0	0	0.5	0	0	0	0	0	0	
21	0	0	0	0	1	0	0	0	0	0	0	0	0	
22	0	0	0.3333	0	0	0	0	0	0	0	0	0	0	0.666
23	0	0	0	0.5	0	0	0	0	0	0	0	0	0	
24	0	0	0	0.1667	0	0	0	0	0	0	0	0.6667	0	
25	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	1	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	1	0	0	0	0
29	0	0	0.2	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0.75
32	0	0	0.3333	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0.3333	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0.2857	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0.5	0	0	0	0	0	0	0	0	0.25	0	0
37	0	0	0.5	0	0	0	0	0	0	0	0	0.2	0	0
38	0	0	0.5	0	0	0	0	0	0	0	0	0.15	0	0
39	0	0	0.5	0	0	0	0	0	0	0	0	0.175	0	0
40	0	0	0.5	0	0	0	0	0	0	0	0	0.15	0	0
41	0	0	0.5	0	0	0	0	0	0	0	0	0.125	0	0



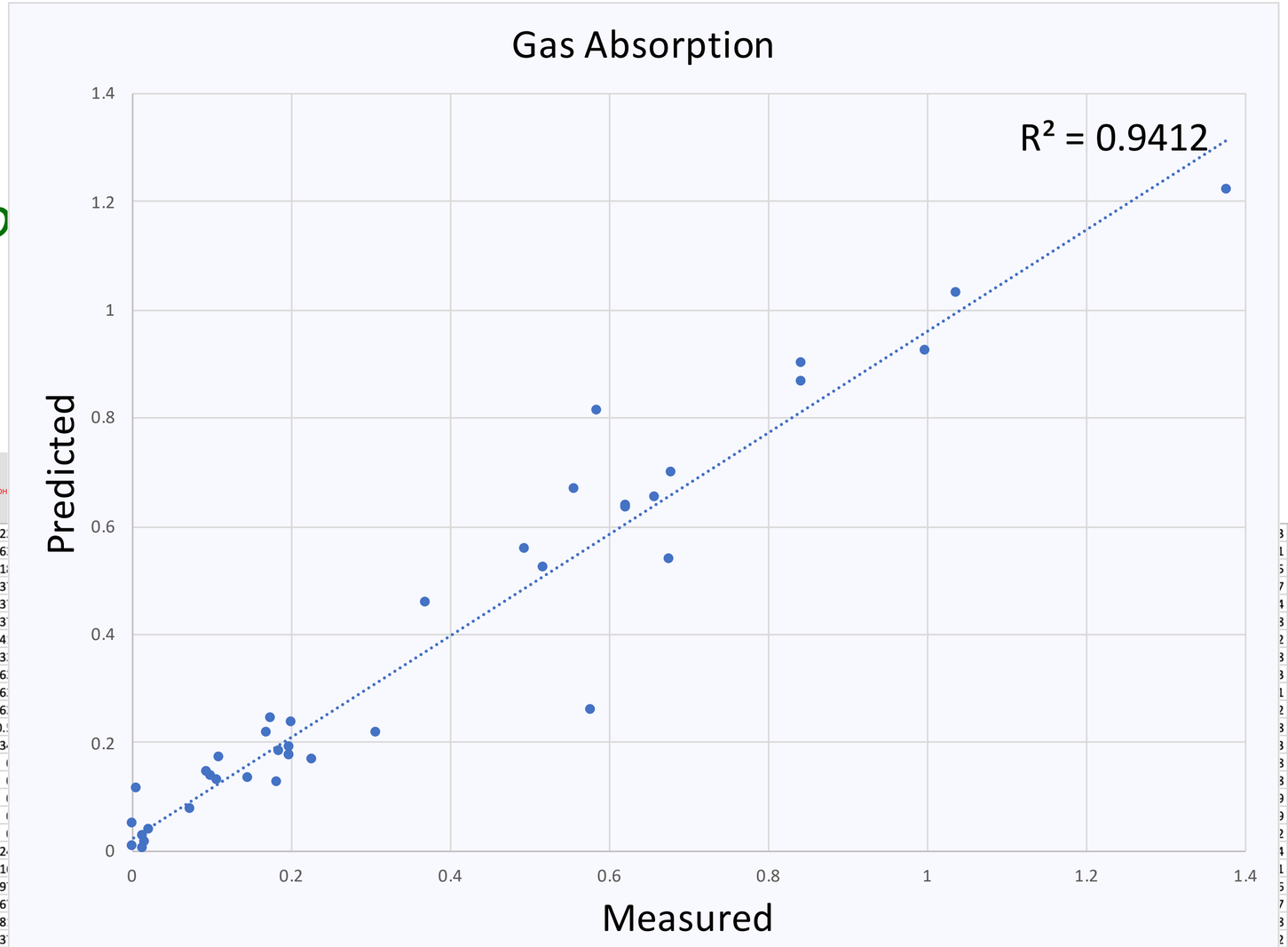
Cross-populate

✤ Empty cells set to **max(***Tanimoto*

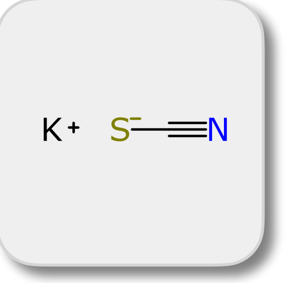
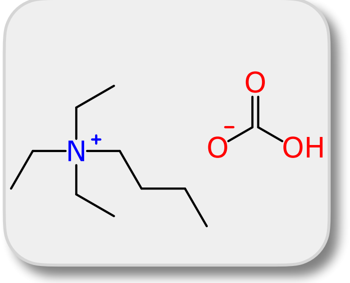
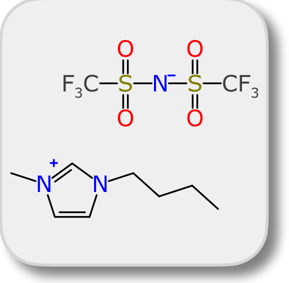
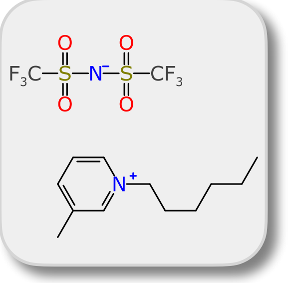
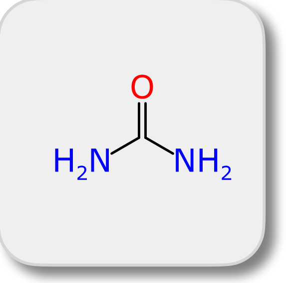
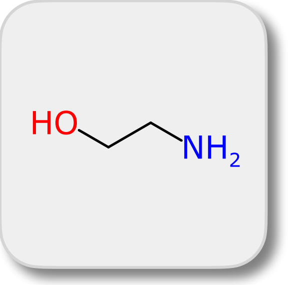
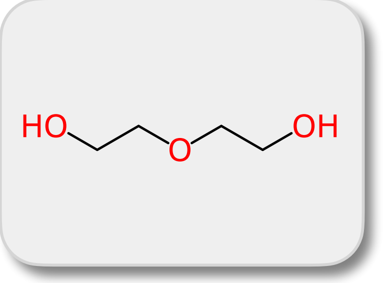
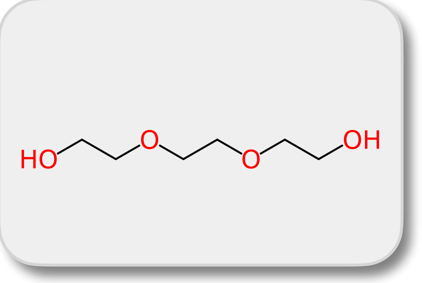
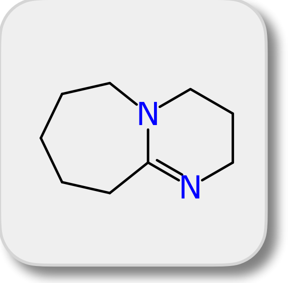
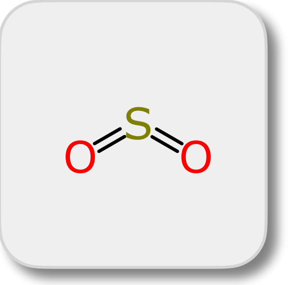


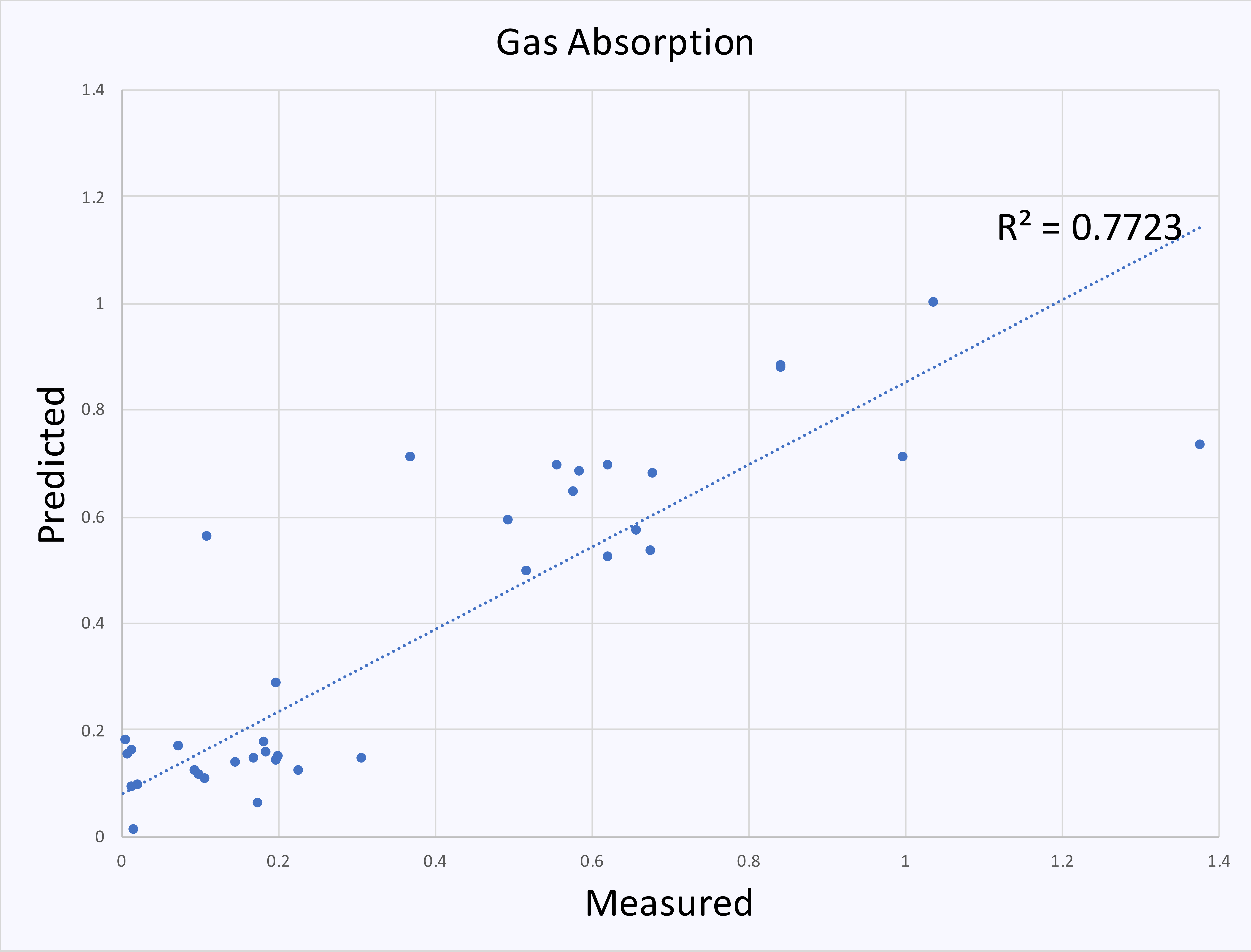
1	0	0	0.1111	0.0202	0.01852	0.02299	0.01389	0.01333	0.01515	0.01667	0.01258	0.0202	0	0.222
2	0	0	0.3333	0.06061	0.08	0.07692	0.01905	0.01802	0.01802	0.0202	0.01667	0.01235	0	0.0526
3	0	0	0.3333	0.06061	0.05556	0.1111	0.01961	0.01852	0.02222	0.02564	0.01709	0.01235	0	0.181
4	0	0	0.1429	0.04286	0.06122	0.127	0.02317	0.02198	0.02597	0.02956	0.02041	0.005291	0	0.183
5	0	0	0.2017	0.04286	0.06122	0.127	0.02317	0.02198	0.02597	0.02956	0.02041	0.1429	0	0.183
6	0	0	0.2017	0.1429	0.06548	0.127	0.02317	0.02198	0.02597	0.02956	0.02041	0.01099	0	0.183
7	0	0	0.05405	0.1818	0.1429	0.06522	0.6471	0.5676	1	0.6552	0.1509	0.05128	0	0.0294
8	0	0	0.06061	0.09375	0.07317	0.04651	0.4054	0.3846	0.6552	1	0.09615	0.05714	0	0.0333
9	0	0	0.8	0.1455	0.1333	0.129	0.03902	0.03721	0.04324	0.04848	0.03478	0.02963	0	0.126
10	0	0	0.8	0.1455	0.1333	0.129	0.03902	0.03721	0.04324	0.04848	0.03478	0.06	0	0.126
11	0	0	0.8	0.1455	0.1333	0.129	0.03902	0.03721	0.04324	0.04848	0.03478	0.02963	0	0.126
12	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0	0.
13	0	0	0.1613	0.09091	0.25	1	0.08163	0.07843	0.06522	0.04651	0.07407	0.02703	0	0.103
14	0.1875	0.25	0	0	0.01974	0.01923	0.0163	0.01563	0	0	0.03	0.02419	0	
15	0.1875	0.25	0.0375	0.0375	0.1154	0.1111	0.04167	0.03947	0.02273	0.02586	0.03659	0.03409	0	
16	0.25	0.1875	0.0375	0.0375	0.1154	0.1111	0.04167	0.03947	0.02273	0.02586	0.03659	0.03409	0	
17	0.25	0.1875	0	0	0.01974	0.01923	0.0163	0.01563	0	0	0.03	0.02419	0	
18	0.4	0.3	0	0	0.072	0.06923	0.01714	0.01622	0	0	0.015	0	0	
19	0	0	0.1	0.02174	0.03226	0.06667	0.0125	0.0119	0.01389	0.01563	0.01111	0.01429	0	0.0882
20	0	0	0.02439	0.08108	0.07778	0.04082	0.5	0.4394	0.3235	0.2027	0.1633	0.02326	0	0.0131
21	0	0	0.1667	0.4583	1	0.25	0.1556	0.1489	0.1429	0.07317	0.14	0.02778	0	0.0689
22	0	0	0.3333	0.06061	0.05556	0.06897	0.01754	0.01667	0.01961	0.02222	0.0155	0.01235	0	0.666
23	0	0	0.09091	0.5	0.2292	0.04545	0.08108	0.07692	0.09091	0.04688	0.07143	0.03846	0	0.0238
24	0	0	0.0303	0.1667	0.07639	0.01515	0.02703	0.02564	0.0303	0.01563	0.02381	0.01282	0.6667	0.00793
25	0	0	0.025	0.025	0.07692	0.07407	0.02778	0.02632	0.01515	0.01724	0.02439	0.02273	0	
26	0	0	0	0	0.01316	0.01282	0.01316	0.0125	0.01471	0.01667	0.02	0.02174	0	0
27	0	0	0.04651	0.1538	0.1489	0.07843	0.8788	1	0.5676	0.3846	0.3673	0.04444	0	0.025
28	0	0	0.04348	0.1429	0.14	0.07407	0.3265	0.3673	0.1509	0.09615	1	0.1111	0	0.02326
29	0	0	0.2	0.04211	0.05926	0.05714	0.02222	0.02105	0.025	0.02857	0.01951	0.0381	0	0.1846
30	0	0	0.25	0.08333	0.08654	0.15	0.04286	0.04054	0.04839	0.02679	0.0375	0.009259	0	0.1731
31	0	0	0.25	0.07143	0.07759	0.1034	0.03947	0.0375	0.04412	0.05	0.03488	0.03125	0	0.3462
32	0	0	0.3333	0.06061	0.1429	0.1111	0.03333	0.03175	0.01802	0.0202	0.02963	0.02564	0	0.1111
33	0	0	0.4	0.07273	0.072	0.06923	0.01951	0.0186	0.02162	0.02424	0.01739	0.01481	0	0.06316
34	0	0	0.3333	0.06061	0.08	0.07692	0.01905	0.01802	0.01802	0.0202	0.01667	0.01235	0	0.05263
35	0	0	0.2857	0.05195	0.08571	0.08242	0.02041	0.01931	0.01544	0.01732	0.01786	0.01058	0	0.04511
36	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0.25	0.07895
37	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0.2	0.07895
38	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0.15	0.07895
39	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0.175	0.07895
40	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0.15	0.07895
41	0	0	0.5	0.09091	0.08333	0.08065	0.02439	0.02326	0.02703	0.0303	0.02174	0.01852	0.125	0.07895

1																								
2																								
3																								
4																								
5																								
6																								
7																								
8																								
9																								
10																								
11																								
12																								
13																								
14																								
15																								
16																								
17																								
18																								
19																								
20																								
21																								
22																								
23																								
24																								
25																								
26	0	0	0.01786	0.02083	0.02273	0	0	0	0	0	0.02632	0.5	0.5	0.09375	0.0625	0	1	0.624
27	0.05	0.02632	0.04762	0.05263	0.02703	0.02564	0.02778	0.05405	0.02439	0.02381	0.02381	0.025	0.02083	0.01923	0.02	0	1	0.844
28	0.04651	0.02439	0.04444	0.04878	0.025	0.02381	0.02564	0.05	0.02273	0.02222	0.09524	0.02326	0.04	0.01818	0.01887	0	1	0.844
29	0.3333	0.8	0.2857	0.05714	0	0.05714	0.07273	0.06154	0.05	0.04706	0.04706	0	0	0	0	1	0	0.0188
30	0.1607	0.05769	0.04167	0.0125	0	0.2727	0.375	0.75	0.2308	0.2143	0.04412	0	0	0	0	1	0	0.0164
31	0.75	0.3125	0.2206	0.04412	0	0.15	0.1875	0.1607	0.1324	0.125	0.0375	0	0	0	0	1	0	0.0165
32	0.1961	0.2381	0.6667	0.1667	0.1333	0.07843	0.09524	0.08333	0.07018	0.06667	0.0303	0	0.02381	0	0	1	0	0.0248
33	0.08421	0.04444	0.12	0.3333	0.6	0.09412	0.1143	0.1	0.08421	0.08	0.01739	0	0.02727	0	0	1	0	0.201
34	0.07018	0.03704	0.1333	0.3704	0.6667	0.07843	0.09524	0.08333	0.07018	0.06667	0.01449	0	0.0303	0	0	1	0	0.309
35	0.06015	0.03175	0.1429	0.3968	0.7143	0.06723	0.08163	0.07143	0.06015	0.05714	0.01242	0	0.03247	0	0	1	0	0.203
36	0.1053	0.05556	0.05	0.1389	0.25	0.1176	0.1429	0.125	0.1053	0.1	0.02174	0	0.01136	0	0	1	0	0.111
37	0.1053	0.05556	0.06	0.1667	0.3	0.1176	0.1429	0.125	0.1053	0.1	0.02174	0	0.01364	0	0	1	0	0.103
38	0.1053	0.05556	0.07	0.1944	0.35	0.1176	0.1429	0.125	0.1053	0.1	0.02174	0	0.01591	0	0	1	0	0.097
39	0.1053	0.05556	0.05	0.1389	0.25	0.075	0.1429	0.125	0.1053	0.1	0.02174	0	0.01136	0	0	1	0	0.229
40	0.1053	0.05556	0.06	0.1667	0.3	0.05	0.1429	0.125	0.1053	0.1	0.02174	0	0.01364	0	0	1	0	0.202
41	0.1053	0.05556	0.07	0.1944	0.35	0.025	0.1429	0.125	0.1053	0.1	0.02174	0	0.01591	0	0	1	0	0.189



10-most Significant Descriptors

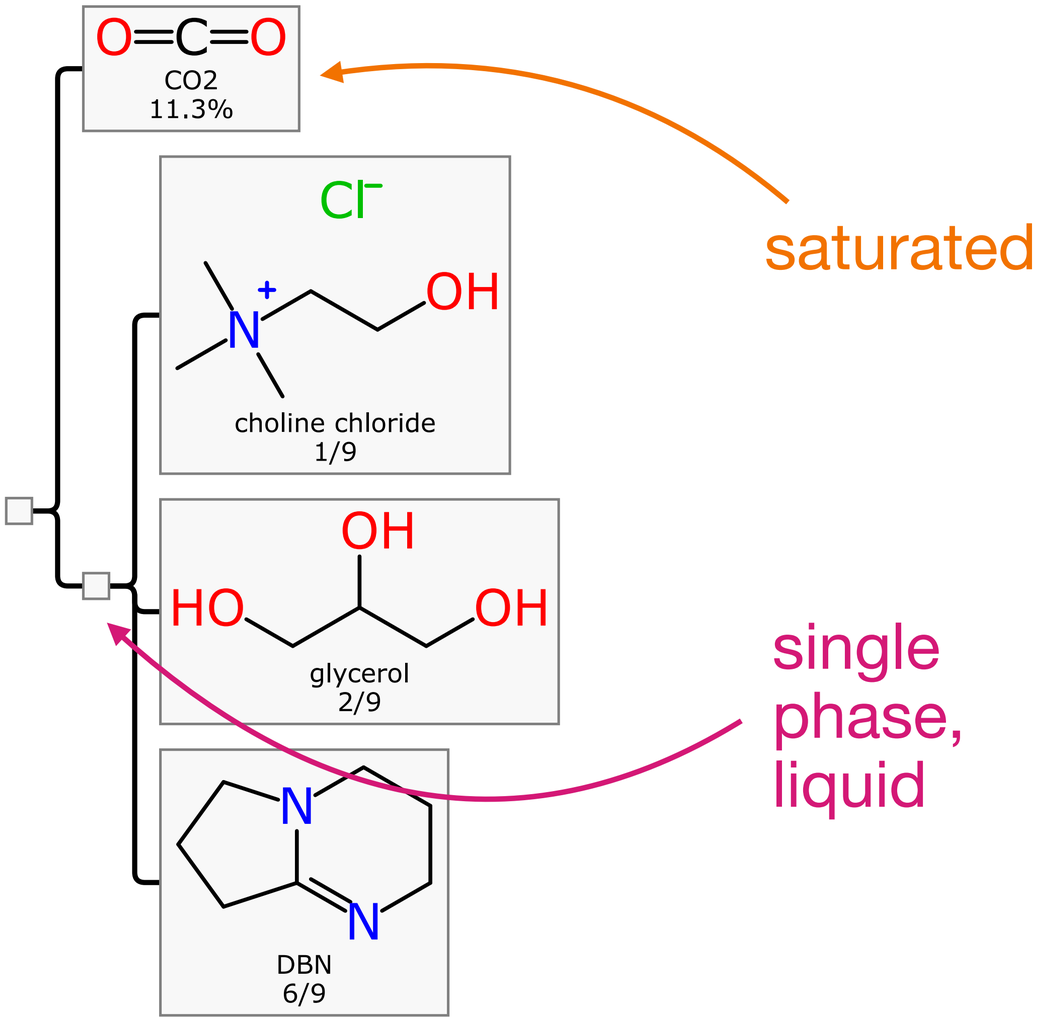













0	0.01852	0.01389	0.01258	0	0.04762	0.04167	0.03922	0.5455	0
0								0	0
0								0	0
0								0	0
0								0.004082	0
0								0	0
0								0.02174	0
0								0.02381	0
0								0	0
0								0.006452	0
0								0	0
0								0	1
0								0	1
0.25								0.1406	1
0.25								0	1
0.1875								0	1
0.1875								0.1406	1
0.3								0	1
0								0.5	0
0								0.5	0
0								0	0
0								0	0
0								0.09375	1
0								0.03125	1
0								0.09375	1
0								0.09375	1
0								0.01923	1
0								0.01818	1
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0								0	0
0	0.08333	0.02439	0.02174	0.35	0.025	0.1053	0.1	0	0

Labelling Caveat

✤ Mixture data not quite free of implicit assumptions...



✤ Need additional metadata, e.g.

- ontologies
- IUPAC terminology

✤ Work in progress

Summary

- **Open protocols**, tools & example data available for representing mixtures
- **New workflows** become practical with cheminformaticisation of mixtures
- **Much work** to be done to catch up with structure databases
- **Community** building is key
- **Many industries**: lab chemistry, drug formulations, consumer products, agriculture, analytical standards, safety...

Questions?

Journal of Cheminformatics (2019)
10.1186/s13321-019-0357-4

Open Source
github.com/cdd/mixtures

CDD Vault
collaborativedrug.com

✤ Contact:

Alex M. Clark **alex@collaborativedrug.com** (Collaborative Drug Discovery)

InChI TRUST

IUPAC INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

✤ *Also:* Leah McEwen (Cornell, InChI, IUPAC)