

# Towards Rational Molecular Design: Derivation of Property Guidelines for Reduced Acute Aquatic Toxicity

Adelina M. Voutchkova<sup>a</sup>, Jakub Kostal,<sup>a</sup> Justin B. Steinfeld,<sup>a</sup> John W. Emerson,<sup>b</sup> Bryan Brooks,<sup>c</sup> Paul Anastas,<sup>d</sup> Julie B. Zimmerman<sup>d,e\*</sup>

## Electronic Supplementary Information

### Table of Contents

- I. Supplementary Tables
- II. Supplementary Figures
- III. Recursive Modeling
- IV. Chemical structure tables

### Supplementary Tables

Supplementary Table 1. Univariate correlations between predicted properties and average acute toxicity values for three species: Fathead minnow, Japanese medaka and *Daphnia magna*.

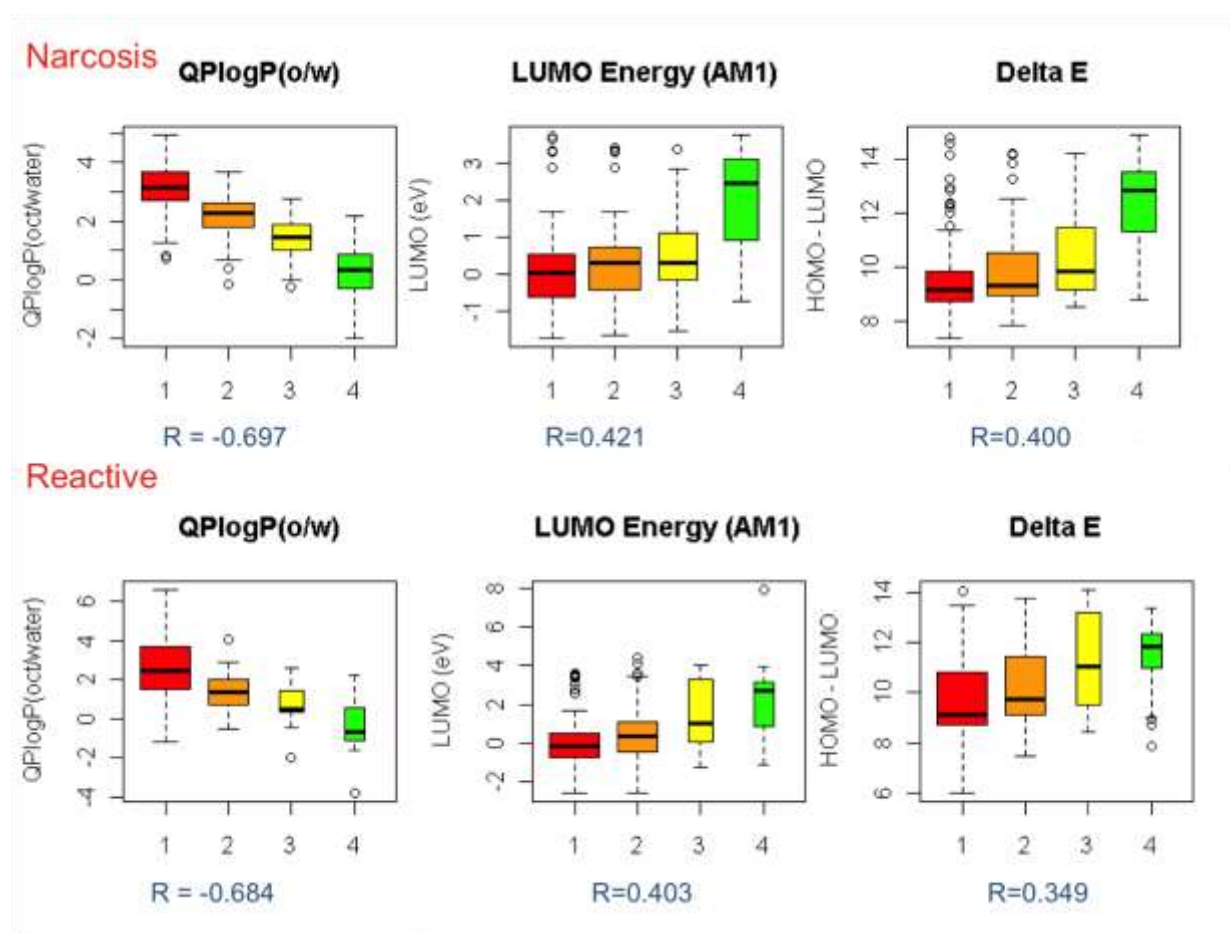
Property	Correlation coefficient (R)
# amine groups	0.020
# amidine groups	-0.008
# acid groups	-0.017
# amide groups	-0.004
# rotatable bonds	0.011
# reactive functional groups	-0.033
Molecular weight	-0.087
Dipole moment (Debye)	-0.04
Solvent Accessible Surface Area	-0.074
Hydrophobic Surface Area	0.013
Hydrophilic Surface Area	0.060
PI Surface Area	-0.106
Weakly Polar Surface Area	-0.057

volume	-0.078
# HB donors	0.08
# HB acceptors	0.156
dip <sup>2</sup> .Volume	-0.012
Globularity	0.054
Polarizability	-0.109
logP(oct/gas)	-0.034
logP(water/gas)	0.114
logP(octanol/warter)	-0.182
logSoubility (M)	0.132
Sluorine surface area	-0.017
Amide surface area	-0.002
Polar surface area	0.055
#N and O	0.049
Lipinski RuleOf5	-0.032
# ringatoms	-0.095
#non H-atoms	-0.09
HOMO (AM1) eV	-0.051
LUMO (AM1) eV	0.175
dE (AM1) eV	0.167

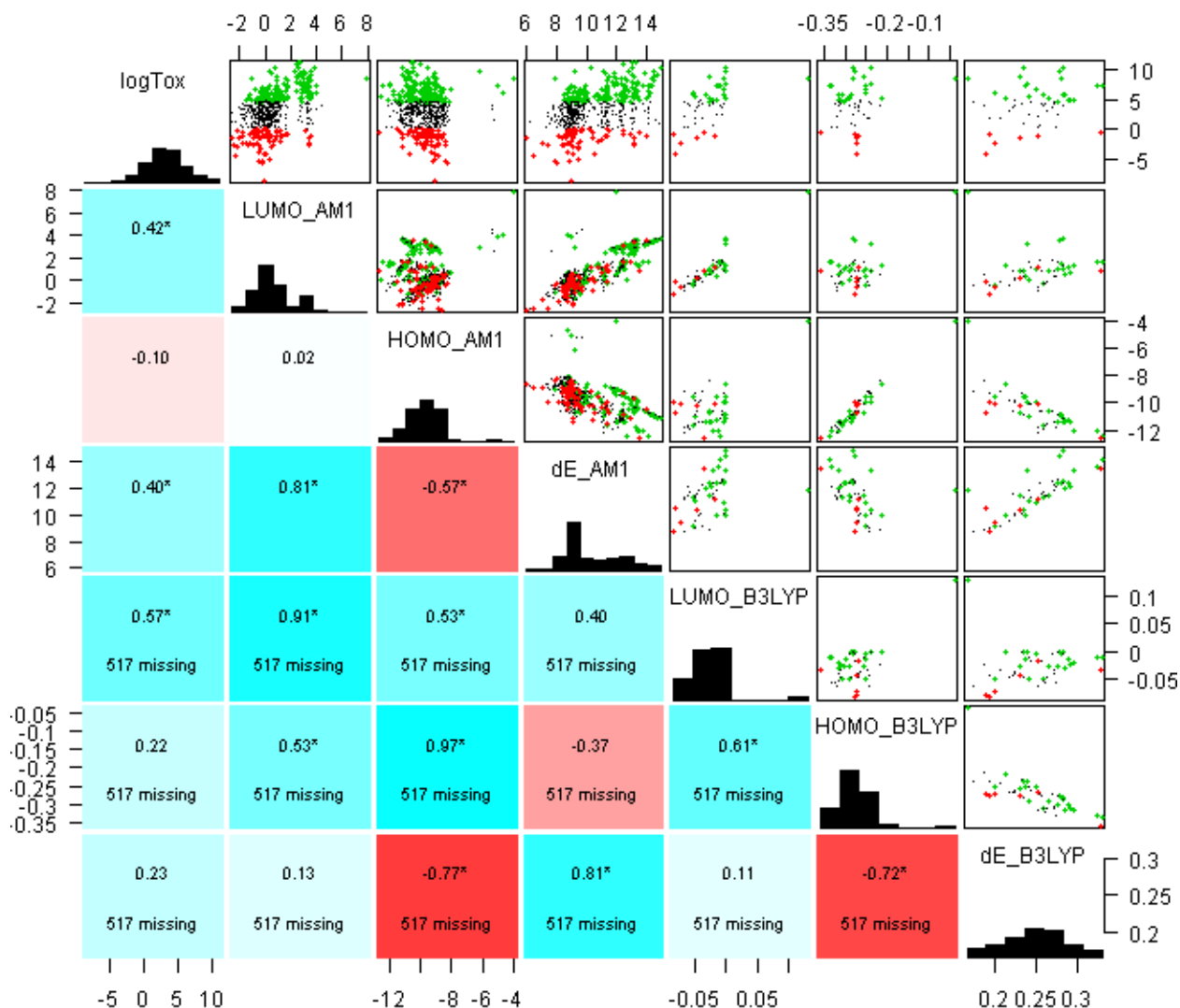
Supplementary Table 2. Summary of linear models for acute toxicity (LC50) to the fathead minnow (standard errors in parenthesis).

Property attributes	Linear Model			
	1	2	3	4
Intercept	6.11 (0.13)	5.61 (0.15)	-2.60(2.12)	0.69 (2.81)
logP <sub>o/w</sub>	-1.33 (0.05)	-1.21 (0.05)	-1.10(0.05)	-0.75 (0.11)
dE		0.42 (0.06)	0.42 (0.05)	0.36 (0.07)
LUMO			8.72 (2.25)	7.48 (2.89)
Globularity				-0.67 (0.17)
HOMO				-0.23 (0.04)
PSA				-0.04 (0.01)
logP(water/gas)				0.73 (0.12)
logP(oct/gas)				-0.60 (0.11)
NandO				0.47 (0.12)
QPlogS				-0.30 (0.11)
R <sup>2</sup>	0.56	0.60	0.61	0.65
RSE	1.91	1.82	1.80	1.72
BIC	234	2299	2290	2279

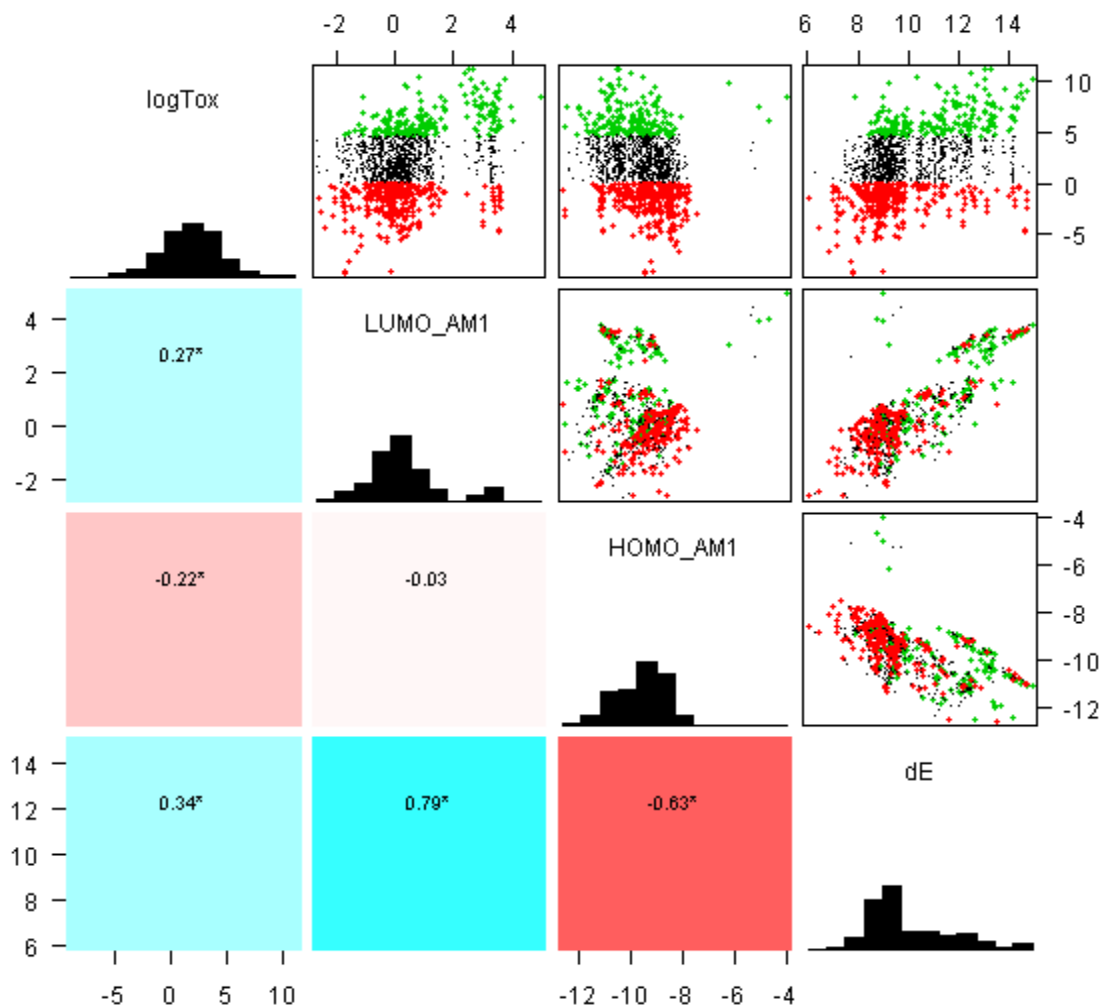
## Supplementary Figures



Supplementary Material, Figure 1. Box plots of  $\log P_{(o/w)}$ , LUMO energy and  $dE$  (HOMO-LUMO) of compounds in the EPA fathead minnow dataset that are toxic by narcosis and by reactive mechanisms. The compounds are divided by EPA level of concern for acute aquatic toxicity (high, medium, low, none); boxplot widths are proportional to the square root of the counts in each group.



**Supplementary Material, Figure 2.** Scatter plot matrix of HOMO and LUMO energies for 567 molecules in the fathead minnow dataset (unless otherwise noted) predicted by two methods: semi-empirical AM1 (Gaussian) and DFT B3LYP (Gaussian) (50 molecule random sample). Compounds in red have  $LC_{50} < 100$  mg/L, while those in green have  $LC_{50} > 100$  mg/L (low or no concern for acute aquatic toxicity). Because HOMO and LUMO energies at the DFT (B3LYP) level of theory were only calculated for a group of 50 randomly selected compounds, there were 517 “missing” observations in the corresponding tiles below the diagonal.



Supplementary Material, Figure 3. Scatter plot matrix of HOMO, LUMO and  $\Delta E$  (HOMO-LUMO) energies for all 865 compounds predicted by semi-empirical AM1 (Gaussian). Where more than one toxicity value was available or for more than one species, the geometric mean of toxicity values was used. Compounds in red have  $LC_{50} < 1$  mg/L (highest level of concern for acute aquatic toxicity), while those in green have  $LC_{50} > 100$  mg/L (no concern for acute aquatic toxicity).

### III. Recursive modeling

Recursive modeling was used to estimate limits of  $\log P_{(o/w)}$ , LUMO and  $\Delta E$  associated that capture the largest fraction of compounds with  $LC_{50}$  values  $>100$  mg/L (designated acutely “safe” compounds) while excluding the most compounds with  $LC_{50}$  values  $< 100$  mg/L (acutely “toxic”). This was achieved by defining a quantity termed *safe excess ratio*, as defined below:

$$\text{Safe excess ratio} = \frac{\# \text{safe}_{(\text{in subset})} - \# \text{toxic}_{(\text{in subset})}}{\# \text{safe}_{(\text{total})} + \# \text{toxic}_{(\text{total})}}$$

### IV. Chemical Structure Tables

Supplementary Table 3. Compounds from EPA Fathead Minnow database with  $LC_{50}$  (mg/L)  $> 500$  mg/L which have been (designated as acutely “safe”).

	CAS	FHM $LC_{50}$ (mg/L)	FHM $LC_{50}$ (mm ol/L)	$\log P_{o/w}$ (CLOGP) KOWWI N	LUMO (eV) AM1	dE (eV) AM1	Mode of Action
	111-46-6	75200	7.09E+02	-1.3	2.69	13.24	Baseline narcosis
	112-27-6	68900	4.59E+02	-1.24	2.54	13.04	Baseline narcosis
	100-97-0	49800	3.55E+02	<b>2.46</b> <b>-4.15</b>	2.4	12.44	MOA not determined due to insufficient evidence
	67-68-5	34000	4.35E+02	-1.35	0.81	10.34	Baseline narcosis
	67-56-1	29400	9.17E+02	-0.77	3.78	14.91	Baseline narcosis
	111-90-0	26500	1.98E+02	-0.54	2.59	13.04	Baseline narcosis

	82385 -42-0	18300	8.20E +01	0.91	3.04	9.22	Baseline narcosis
	100- 79-8	16700	1.26E +02	-0.07	2.25	12.7 3	Baseline narcosis
	77-71- 4	16460	1.29E +02	-0.65	0.36	11.1 3	MOA not determined due to insufficient evidence
	64-17- 5	14700	3.19E +02	-0.31	3.57	14.4 4	Baseline narcosis
	102- 71-6	11800	7.91E +01	-1.75	2.75	11.8 5	MOA not determined due to insufficient evidence
	126- 81-8	11500	8.20E +01	0.51	0.43	11.2	Baseline narcosis
	107- 41-5	10700	9.05E +01	-0.67	3.14	13.8 1	Baseline narcosis
	123- 91-1	10300	1.17E +02	-0.27	2.83	13.0 3	Baseline narcosis
	67-63- 0	8680	1.44E +02	0.05	3.58	14.7	Baseline narcosis
	67-64- 1	7160	1.23E +02	-0.24	0.84	11.5 1	Baseline narcosis
	75-65- 0	6410	8.65E +01	0.35	3.44	14.4 3	Baseline narcosis
	103- 76-4	6410	4.92E +01	-0.68	2.64	11.8 8	MOA not determined due to insufficient evidence
	110- 88-3	5950	6.61E +01	-0.43	2.39	13.1 5	Baseline narcosis
	51-79- 6	5240	5.88E +01	-0.15	1.54	12.3 6	Baseline narcosis

	137-40-6	4790	4.99E+01	0.33	7.93	11.92	MOA not determined due to insufficient evidence
	111-42-2	4710	4.48E+01	-1.43	2.97	12.53	Baseline narcosis
	71-23-8	4550	7.57E+01	0.25	3.49	14.34	Baseline narcosis
	5217-47-0	4510	2.25E+01	1.44	-1.09	<b>7.86</b>	Neurodepressant
	78-92-2	3670	4.95E+01	0.61	3.55	14.5	Baseline narcosis
	115-19-5	3290	3.91E+01	0.28	1.68	12.64	Baseline narcosis
	78-93-3	3220	4.47E+01	0.29	0.89	11.41	Baseline narcosis
	7209-38-3	3100	1.55E+01	-0.9	2.74	11.77	MOA not determined due to insufficient evidence
	622-40-2	2710	2.07E+01	-0.45	2.45	11.92	Baseline narcosis
	60-29-7	2560	3.45E+01	0.89	2.98	13.37	Baseline narcosis
	78-96-6	2520	3.36E+01	-0.96	3.32	13.07	MOA not determined due to insufficient evidence
	69723-94-0	2410	9.67E+00	<b>Undet</b> <b>-2.61</b>	-1.79	<b>8.82</b>	MOA not determined due to conflicting evidence
	109-01-3	2300	2.30E+01	-0.1	2.98	12.04	MOA not determined due to insufficient evidence
	109-07-9	2240	2.24E+01	-0.44	3.17	12.35	MOA not determined due to insufficient evidence
	140-	2190	1.70E	-0.68	2.73	11.8	MOA not determined due to



	31-8		+01			3	insufficient evidence
	109-99-9	2160	3.00E+01	0.46	3.11	13.29	Baseline narcosis
	141-43-5	2070	3.39E+01	-1.31	3.35	13.35	MOA not determined due to insufficient evidence
	111-69-3	1930	1.79E+01	-0.32	1.39	13.39	Baseline narcosis
	100-37-8	1780	1.52E+01	0.32	2.93	11.91	Baseline narcosis
	71-36-3	1730	2.33E+01	0.88	3.42	14.27	Baseline narcosis
	280-57-9	1730	1.54E+01	0.69	2.86	11.86	MOA not determined due to conflicting evidence
	54-21-7	1720	1.25E+01	<b>2.26</b> <b>-1.49</b>	3.93	<b>8.97</b>	MOA not determined due to insufficient evidence
	75-05-8	1644	4.01E+01	-0.34	1.66	14.13	Baseline narcosis
	96-22-0	1540	1.79E+01	0.79	0.91	11.33	Baseline narcosis
	107-12-0	1520	2.76E+01	0.16	1.71	13.7	Baseline narcosis
	110-73-6	1480	1.66E+01	-0.46	2.96	12.54	Baseline narcosis
	78-83-1	1430	1.93E+01	0.76	3.46	14.33	Baseline narcosis
	693-93-6	1390	1.67E+01	0.49	0.37	9.96	Baseline narcosis
	107-87-9	1240	1.44E+01	0.91	0.88	11.41	MOA not determined due to conflicting evidence

	77-75-8	1220	1.24E+01	0.86	1.7	12.53	Baseline narcosis
	109-76-2	1190	1.61E+01	-1.49	3.31	13.02	Polar narcosis
	4214-79-3	1140	8.80E+00	1.78	-0.21	9.19	MOA not determined due to insufficient evidence
	621-42-1	1130	7.48E+00	0.73	0.24	9.01	MOA not determined due to insufficient evidence
	78-90-0	1010	1.36E+01	-0.91	3.42	13.05	Polar narcosis
	17584-12-2	952	7.67E+00	-0.21	-0.22	<b>8.81</b>	Polar narcosis
	109-06-8	897	9.63E+00	1.11	0.16	9.79	Baseline narcosis
	563-80-4	864	1.00E+01	0.56	0.91	11.32	Baseline narcosis
	109-89-7	855	1.17E+01	0.58	3.23	12.52	MOA not determined due to insufficient evidence
	96-29-7	843	9.68E+00	0.65	1.08	11.01	MOA not determined due to conflicting evidence
	103-90-2	814	5.39E+00	0.32	0.25	<b>8.71</b>	MOA not determined due to insufficient evidence
	108-20-3	786	7.69E+00	1.52	2.87	13.25	Baseline narcosis
	120-07-0	735	4.06E+00	0.44	0.73	<b>8.8</b>	Baseline narcosis

	100-70-9	726	6.97E+00	0.5	-0.73	9.57	Polar narcosis
	108-93-0	704	7.03E+00	1.23	3.37	14.04	Baseline narcosis
	1634-04-4	672	7.62E+00	0.94	2.99	13.42	Baseline narcosis
	77-74-7	672	6.58E+00	1.53	3.42	14.14	Baseline narcosis
	55-21-0	661	5.46E+00	0.64	-0.13	9.82	Baseline narcosis
	6948-86-3	635	2.41E+00	1.15	2.49	11.21	Baseline narcosis
	6636-78-8	622	4.80E+00	1.5	-0.33	9.18	Polar narcosis
	108-94-1	621	6.33E+00	0.81	0.92	11.23	Baseline narcosis
	29553-26-2	582	3.64E+00	1.03	1.24	12.43	Baseline narcosis
	629-40-3	528	3.88E+00	0.59	1.55	13.31	Baseline narcosis

Supplementary Table 4. Compounds from EPA Fathead Minnow database with LC<sub>50</sub> (mg/L) < 10 mg/L (designated as acutely “most toxic”), which also meet the proposed design criteria for being “safe”

	CAS	logPo/w (CLOGP) KOWWIN Exptl value	LUMO (eV) AM1	dE (eV) AM1	LC <sub>50</sub> mg/L	Mode of Action
	57-14-7	-1.50 -1.19	2.76	12.22	7.85	electrophile

	90-02-8	1.81	-0.43	9.07	2.3	electrophile
	93-91-4	1.05 <b>0.61</b>	-0.55	9.55	1.1	electrophile
	96-05-9	1.57 <b>2.12</b>	0.05	10.23	0.99	electrophile
	96-17-3	1.14	0.90	11.35	9.97	electrophile
	100-25-4	1.46	-2.21	9.14	0.71	ND
	100-52-7	1.48	-0.43	9.57	9.87	electrophile
	104-88-1	2.1	-0.74	9.17	2.2	electrophile
	106-63-8	2.22	0.04	11.08	2.1	electrophile
	107-02-8	-0.01	-0.14	10.56	0.017	electrophile
	107-14-2	0.228	0.34	12.15	1.35	resp_blok
	107-18-6	0.17	1.22	11.26	0.32	electrophile
	107-19-7	0.139	1.70	12.33	1.48	electrophile
	109-77-3	-0.668	0.85	13.48	0.56	resp_blok
	111-86-4	1.613	3.43	13.12	5.19	ND
	112-20-9	1.975	3.42	13.11	2.16	ND
	114-26-1	1.567	0.29	9.17	8.8	Ache_inhib

	120-80-9	0.798	0.19	9.47	9.22	narcosis
	122-03-2	1.857	-0.41	9.34	6.62	narcosis
	387-45-1	1.59	-0.90	9.04	9.41	electrophile
	446-52-6	1.532	-0.70	9.21	1.35	electrophile
	590-86-3	0.827	0.89	11.45	3.25	electrophile
	693-16-3	1.672	3.49	13.15	5.19	ND
	818-61-1	0.079	0.02	11.08	4.8	electrophile
	999-61-1	0.58	0.00	11.07	3.34	electrophile
	1689-83-4	1.939	-1.07	9.13	6.8	narcosis
	3698-83-7	1.529	-2.14	9.04	0.0456	electrophile
	5292-45-5	0.718	-1.73	9.37	6.52	narcosis
	6361-21-3	1.168	-1.59	9.07	3.87	electrophile
	7383-19-9	1.879	1.65	12.49	1.76	electrophile