Supplementary Material

A)

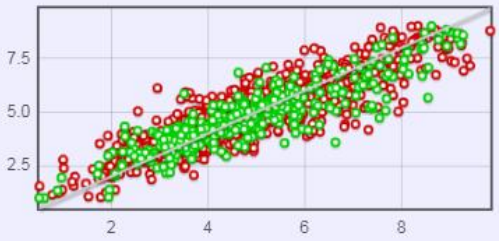
Overview of a multi-model

This is a multi-model that predicts several properties simultaneously

Model name: M1_EC50 aquatic+LC50 aquatic_TRANSNN (F) (3D) _ 25/25 - 504381[apply to new compounds]
Training method: TRANSNN

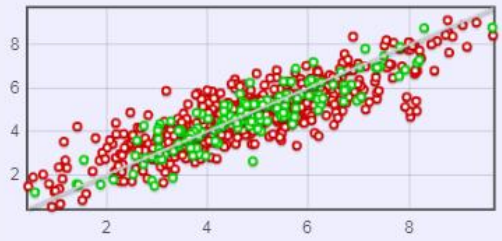
Property: EC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R_2	RMSE	MAE
Tox_Set_comb (training)(1502)	0.75	0.79	0.60
validation0	0.79	0.77	0.57



Property: LC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R_2	RMSE	MAE
Tox_Set_comb (training)(640)	0.72	0.85	0.65
validation0	0.82	0.69	0.53



B)

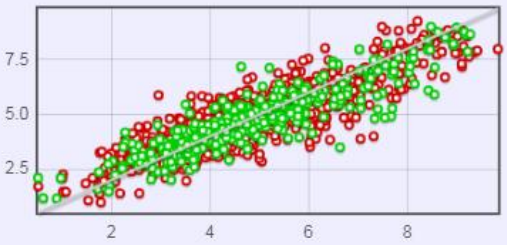
Overview of a multi-model

This is a multi-model that predicts several properties simultaneously

Model name: M2_EC50 aquatic+LC50 aquatic_CNF2 (F) (3D) _ 25/25 - 502660[apply to new compounds]
Training method: CNF2

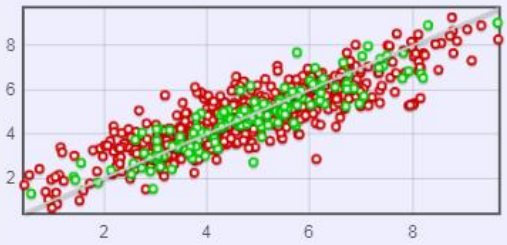
Property: EC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R_2	RMSE	MAE
Tox_Set_comb (training)(1502)	0.77	0.76	0.56
validation0	0.80	0.75	0.55



Property: LC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R_2	RMSE	MAE
Tox_Set_comb (training)(640)	0.73	0.83	0.62
validation0	0.83	0.67	0.50



C)

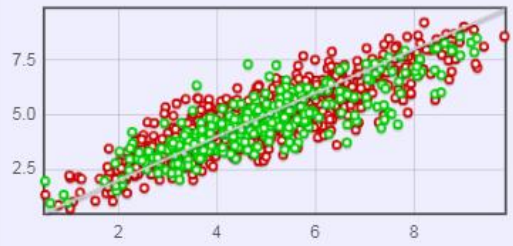
Overview of a multi-model

This is a multi-model that predicts several properties simultaneously

Model name: M3_EC50 aquatic+LC50 aquatic_DEEPCHEM (F) (3D) TEXTCNN_ 10/10 - 511001[apply to new compounds]
Training method: DEEPCHEM

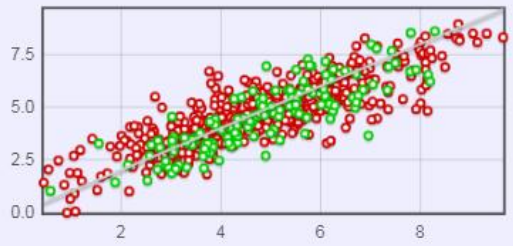
Property: EC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R2	RMSE	MAE
Tox_Set_comb (training)(1502)	0.74	0.82	0.62
validation0	0.74	0.87	0.64



Property: LC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R2	RMSE	MAE
Tox_Set_comb (training)(640)	0.71	0.86	0.65
validation0	0.72	0.90	0.70



D)

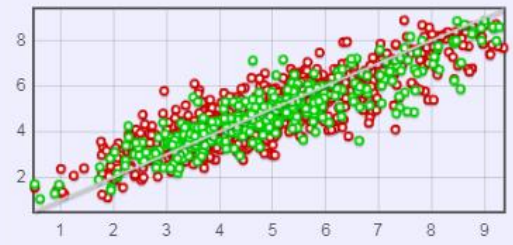
Overview of a multi-model

This is a multi-model that predicts several properties simultaneously

Model name: M4_Consensus EC50 aquatic+LC50 aquatic[apply to new compounds]
Training method: Consensus

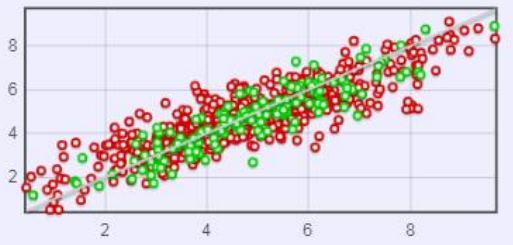
Property: EC50 aquatic measured in -log(mol/L) (Details..)

Dataset	R2	RMSE	MAE
Tox_Set_comb (training)(1502)	0.77	0.76	0.56
validation0	0.80	0.76	0.56



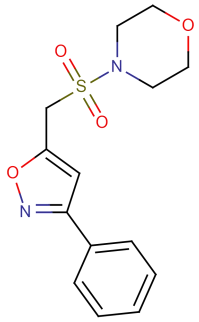
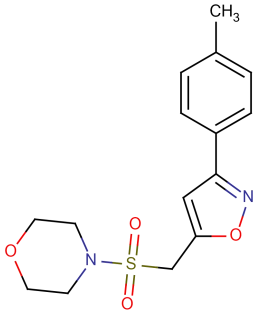
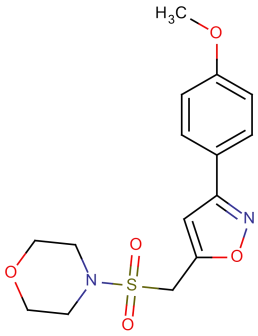
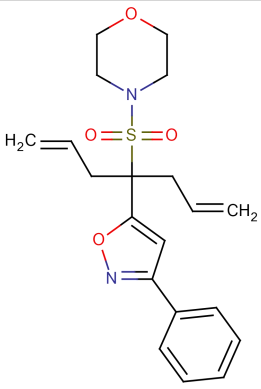
Property: LC50 aquatic measured in -log(mol/L) (Details..)

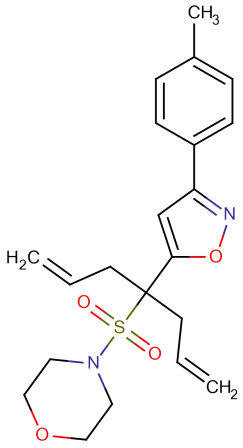
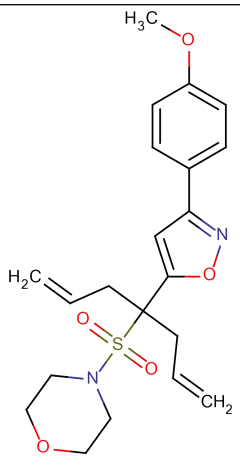
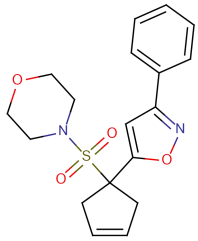
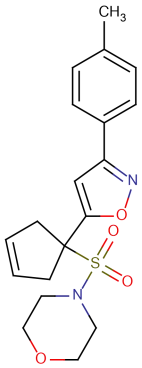
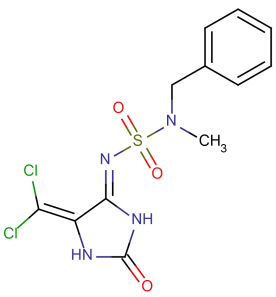
Dataset	R2	RMSE	MAE
Tox_Set_comb (training)(640)	0.74	0.81	0.61
validation0	0.82	0.69	0.53

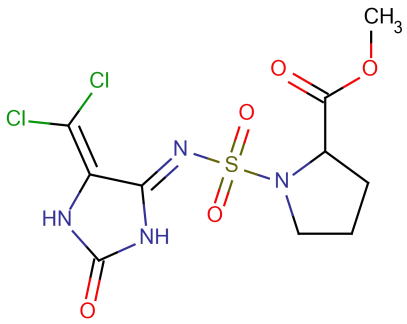
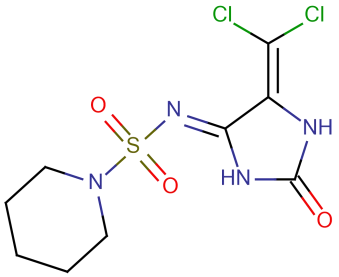
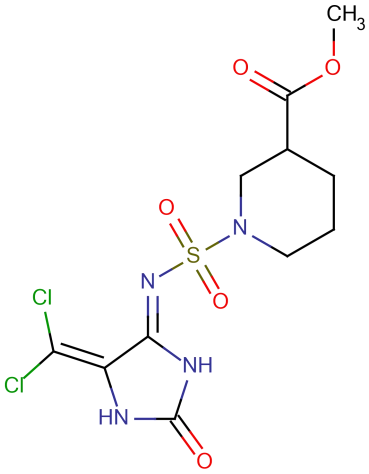
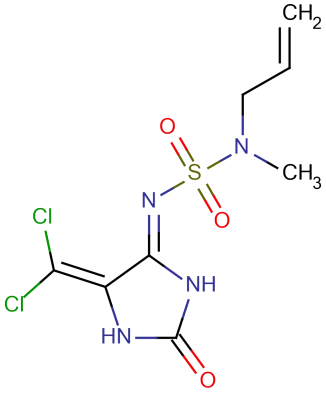


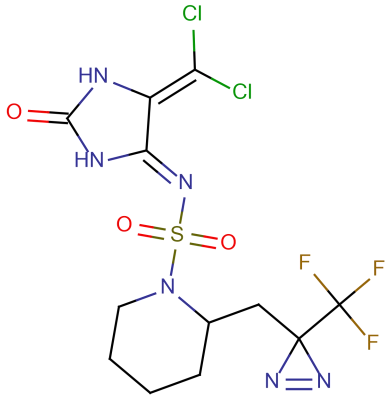
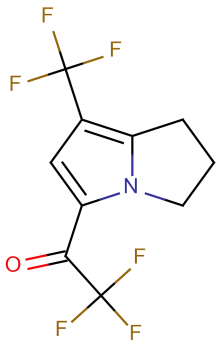
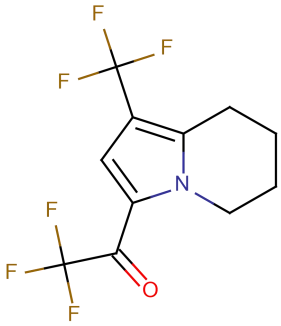
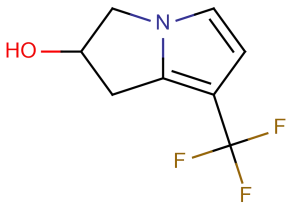
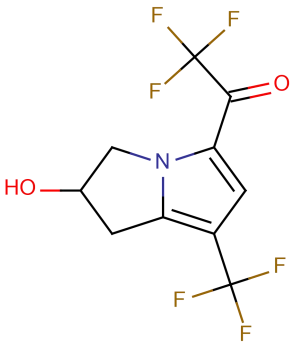
Supplementary Figure 1. QSAR models constructed using the OCHEM (http://ochem.eu). (A-C) Statistical coefficients calculated for machine learning models; (D) Consensus model calculated on the basis of three models.

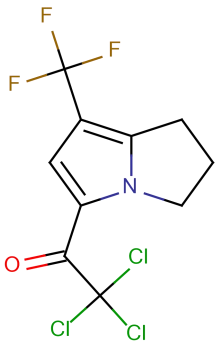
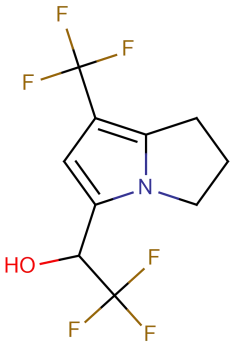
Supplementary Table 1. Structures of the 20 compounds analyzed in this work

Compound No	Chemical Structure	LC_{50} (mg/L)	Toxicity level classification (by D.R. Passino and S.B. Smith)[a]
1		>100.0	slightly toxic
2		45.33 ± 13.40	slightly toxic
3		37.32 ± 10.59	slightly toxic
4		25.69 ± 6.08	slightly toxic

5		21.84 ± 5.12	slightly toxic
6		33.23 ± 10.04	slightly toxic
7		44.45 ± 14.36	slightly toxic
8		41.08 ± 13.97	slightly toxic
9		16.01 ± 2.59	slightly toxic

10		22.35 ± 3.79	slightly toxic
11		17.62 ± 3.87	slightly toxic
12		44.34 ± 6.89	slightly toxic
13		18.62 ± 4.89	slightly toxic

14		44.35 ± 6.89	slightly toxic
15		4.89±1.32	moderately toxic
16		6.99±1.13	moderately toxic
17		33.39±4.57	slightly toxic
18		9.10±2.87	moderately toxic

19		7.76±1.75	moderately toxic
20		1.21±0.26	moderately toxic

Note: [a]Toxicity classification by LC_{50} range: practically harmless (100-1000 mg/L); slightly toxic (10-100 mg/L); moderately toxic (1-10 mg/L); highly toxic (0.1-1 mg/L) [1].

Supplementary Table 2. Statistical coefficients calculated by using consensus QSTR model for 20 compounds

No	Experimental values		Calculated values				
	LC_{50} (mg/L)	$\log(1/LC_{50})$	$\log(1/LC_{50})$	CONSENSUS-STD	$\log(1/EC_{50})$	CONSENSUS-STD	AD
1	>100.0	3.49	3.51	0.41	3.59	0.38	TRUE
2	45.33 ± 13.40	3.85	3.44	0.37	3.60	0.43	TRUE
3	37.32 ± 10.59	3.96	3.46	0.32	3.58	0.41	TRUE
4	25.69 ± 6.08	4.18	4.38	0.31	4.35	0.46	TRUE
5	21.84 ± 5.12	4.27	4.40	0.28	4.37	0.47	TRUE
6	33.23 ± 10.04	4.10	4.43	0.33	4.31	0.51	TRUE
7	44.45 ± 14.36	3.91	4.03	0.44	3.99	0.48	TRUE
8	41.08 ± 13.97	3.96	4.01	0.37	4.03	0.46	TRUE
9	16.01 ± 2.59	4.36	4.52	0.14	4.50	0.13	TRUE
10	22.35 ± 3.79	4.22	4.53	0.43	4.33	0.37	TRUE
11	17.62 ± 3.87	4.27	4.73	0.37	4.57	0.25	TRUE
12	44.34 ± 6.89	3.94	4.51	0.44	4.35	0.35	TRUE
13	18.62 ± 4.89	4.23	5.04	0.23	5.08	0.04	TRUE
14	44.35 ± 6.89	4.01	5.43	0.70	5.04	0.53	FALSE
15	4.89 ± 1.32	4.74	4.68	0.46	4.00	0.40	TRUE
16	6.99 ± 1.13	4.61	4.84	0.56	4.26	0.55	TRUE
17	33.39 ± 4.57	3.76	4.26	0.09	3.91	0.15	TRUE
18	9.10 ± 2.87	4.50	4.27	0.13	3.81	0.10	TRUE
19	7.76 ± 1.75	4.62	5.66	0.69	4.90	0.78	FALSE
20	1.21 ± 0.26	5.35	4.55	0.48	3.91	0.29	TRUE

Note: EC_{50} - half maximal effective concentration**;** LC_{50}- median lethal concentration; CONSENSUS-STD – the standard deviation of the predictions, obtained from an ensemble of models; AD –applicability domain

[1] Passino, D. R. M., & Smith, S. B. (1987). Acute bioassays and hazard evaluation of representative contaminants detected in Great Lakes fish, Environ. *Environmental Toxicology and Chemistry, 6*(11)*,* 901-907. https://doi.org/10.1002/etc.5620061111