Electronic Supplementary Material (ESI) for Toxicology Research.

Support Information

Table S1. Average Tanimoto similarity index in FHM、BS and RT data sets

	FHM	BS	RT
Total of compounds	995	900	903
Average Tanimoto similarity index	0.35	0.336	0.351

Table S2. Using 4 descriptors and FP4 modeling results in binary classification of local models

		Methods	Model ID	Q	SE	SP	FP	FN
BS	Test (162）	4D-RF	**1**	0.673	0.762	0.509	28	25
		4D-k-NN	**2**	0.716	0.810	0.544	26	20
		FP4-RF	**3**	0.747	0.905	0.456	31	10
		FP4-SVM	**4**	0.778	0.905	0.544	26	10
FHM	Test (181)	4D-RF	**5**	0.751	0.621	0.872	12	33
		4D-k-NN	**6**	0.768	0.759	0.777	21	21
		FP4-RF	**7**	0.691	0.575	0.798	19	37
		FP4-SVM	**8**	0.696	0.690	0.702	28	27
RT	Test (162)	4D-RF	**9**	0.790	0.931	0.435	26	8
		4D-k-NN	**10**	0.747	0.836	0.522	19	22
		FP4-RF	**11**	0.809	0.957	0.435	26	5
		FP4-SVM	**12**	0.796	0.897	0.544	21	12
Global	Test (320)	4D-RF	**13**	0.728	0.683	0.791	28	59
		4D-k-NN	**14**	0.719	0.785	0.627	50	40
		FP4-RF	**15**	0.756	0.790	0.709	39	39
		FP4-SVM	**16**	0.772	0.780	0.761	32	41

Table S3. The performance of ternary classification models in FHM, BS and RT

		Methods	Model ID	Q	0-0	0-1	0-2	1-0	1-1	1-2	2-0	2-1	2-2
BS	Test (162)	4D-RF	**a**	0.586	45	1	11	20	3	13	21	1	47
		4D-k-NN	**b**	0.549	31	12	14	11	16	9	12	15	42
		FP4-RF	**c**	0.562	32	6	19	13	5	18	9	6	54
		FP4-SVM	**d**	0.599	37	7	13	14	10	12	11	8	50
FHM	Test (181)	4D-RF	**e**	0.641	84	3	7	30	11	10	12	3	21
		4D-k-NN	**f**	0.641	74	14	6	16	23	12	7	10	19
		FP4-RF	**g**	0.514	85	2	7	38	3	10	30	1	5
		FP4-SVM	**h**	0.558	78	12	4	27	10	14	15	8	13
RT	Test (162)	4D-RF	**i**	0.562	29	1	16	9	3	34	10	1	59
		4D-k-NN	**j**	0.568	28	7	11	12	21	13	9	18	43
		FP4-RF	**k**	0.617	35	3	8	15	9	22	11	3	56
		FP4-SVM	**l**	0.648	39	4	3	8	20	18	12	12	47

0-0 means non-toxicity predicted to non-toxicity, 0-1 means non-toxicity predicted to moderate-toxicity, 0-2 means non-toxicity predicted to high-toxicity (just like false positive in binary classification), 1-0 means moderate-toxicity predicted to non-toxicity, 1-1 means moderate-toxicity predicted to moderate-toxicity, 1-2 means moderate-toxicity predicted to high-toxicity, 2-0 means high-toxicity predicted to non-toxicity (just like FN in binary classification), 2-1 means high-toxicity predicted to moderate-toxicity, 2-2 means high-toxicity predicted to high-toxicity.

Table S4. The sensitive substructure alerts in FHM

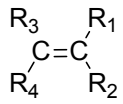
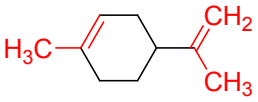
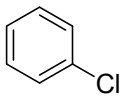
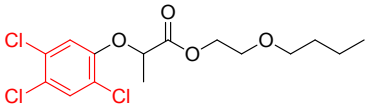
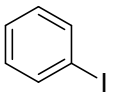
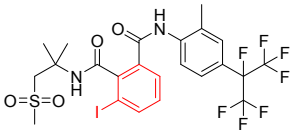
Fragment	Description	SMARTS	Structure	LC_{50}(PPM)		
				BS	FHM	RT
	Alkene	[CX3;$([H2]),$([H1][#6]),$(C([#6])[#6])]=[CX3;$([H2]),$([H1][#6]),$(C([#6])[#6])]			2	0
NO. 5			138-86-3		0.0385	80
	Arylchloride	[Cl][c]		0	2	
NO. 171			19398-13-1	44.5	1	
	Aryliodide	[I][c]			2	0
NO. 174			272451-65-7		0.0665	65.1

Figure Captions

Figure S1. Workflow of model building.

Figure S2. The tanimoto similarity index for the FHM, BS and RT's data set

Figure S3. The radar chart of five physicochemical descriptors: CrippenLogP, ATSm1, SwHBa and ETA_dEpsilon_D and molecular weight for the global model data set were presented. Each color line represents a compound.

Figure S4. Representation of receiver operating characteristics (ROC) curve for the validation set in model **2**, **5**, **10** and **14**.

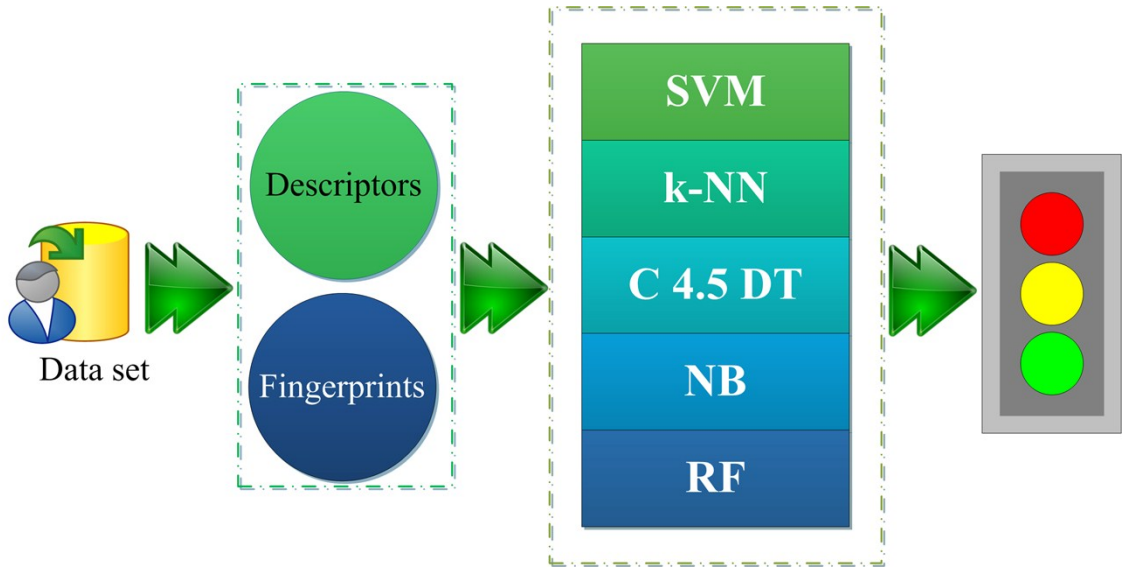


Figure S1. Workflow of model building.

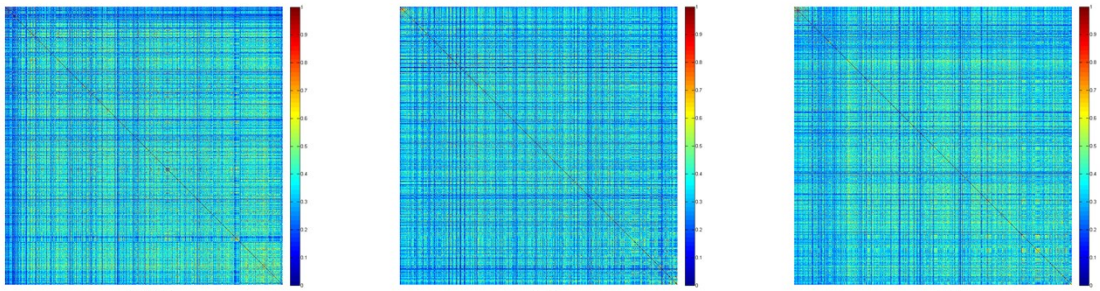


Figure S2. The Tanimoto similarity index for the FHM, BS and RT's data set.

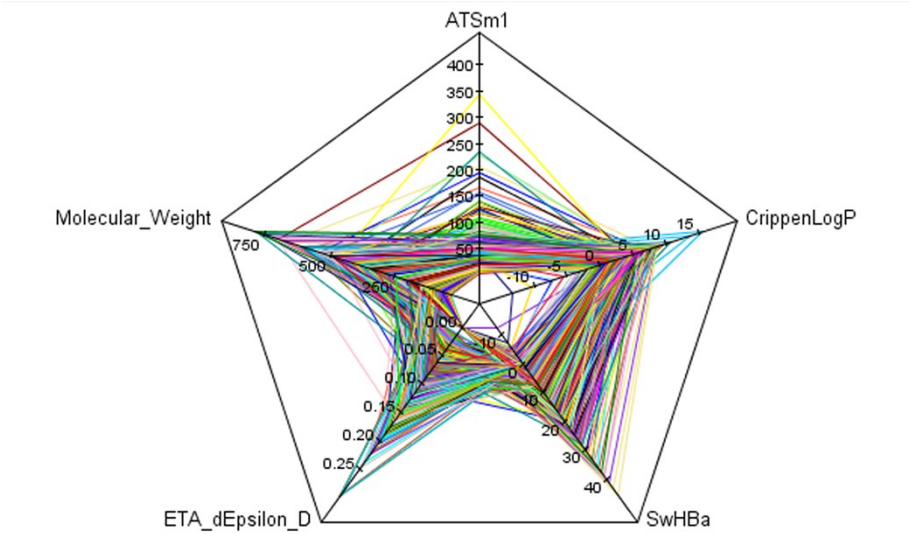


Figure S3. The radar chart of five physicochemical descriptors: CrippenLogP, ATSm1, SwHBa and ETA_dEpsilon_D and molecular weight for the global model data set were presented. Each color line represents a compound.

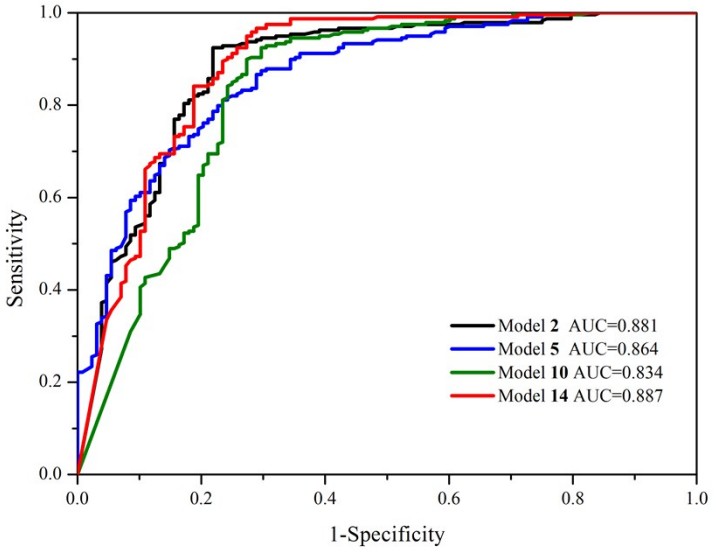


Figure S4. Representation of receiver operating characteristics (ROC) curve for the validation set in model **2**, **5**, **10** and **14**.