J. Serb. Chem. Soc. 86 (7–8) S136–S144 (2021)

Journal of the Serbian Chemical Society

JSCS-info@shd.org.rs • www.shd.org.rs/JSCS

Supplementary material

SUPPLEMENTARY MATERIAL TO

New structure-based models for the prediction of normal boiling point temperature of ternary azeotropes

ZOHREH FARAMARZI[1], FATEMEH ABBASITABAR[2]*, HOSSEIN JALALI JAHROMI[1] and MAZIAR NOEI[1]

[1]*Department of Chemistry, Mahshahr Branch, Islamic Azad University, Mahshahr, Iran and* [2]*Department of Chemistry, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran*

J. Serb. Chem. Soc. 86 (7–8) (2021) 685–698

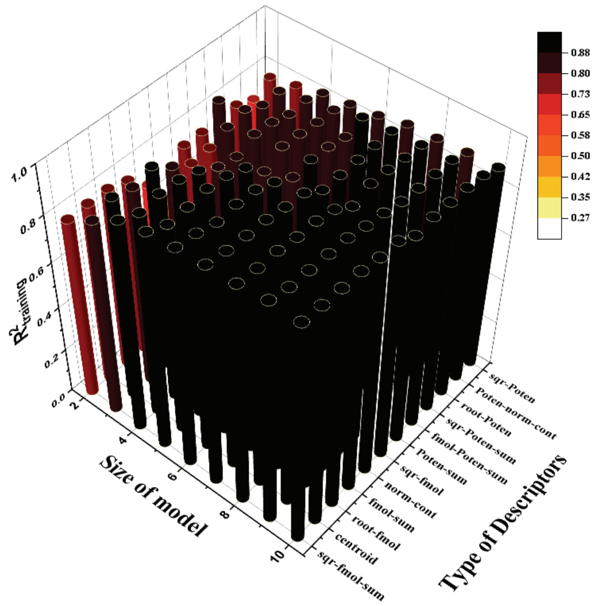

Fig. S-1. $R^2_{training}$ values for QSPR models with different sizes constructed on the basis of different types of mixture descriptors.

* Corresponding author. E-mail: fabbasitabar@gmail.com

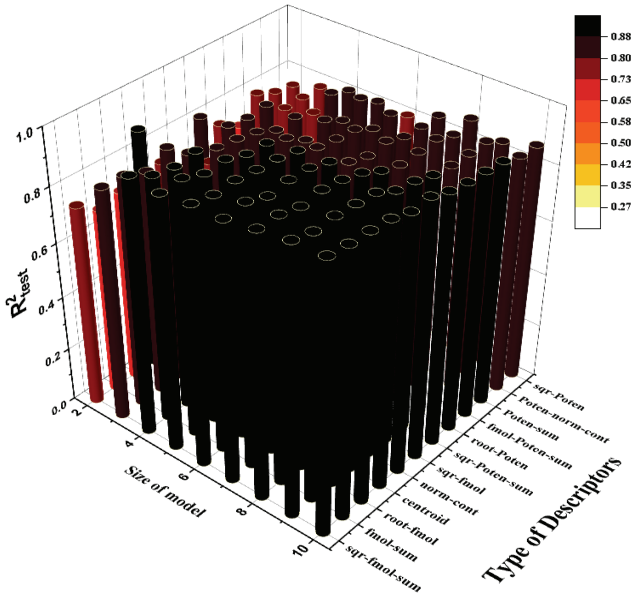

Fig. S-2. R^2_{test} values for QSPR models with different sizes constructed on the basis of different types of mixture descriptors.

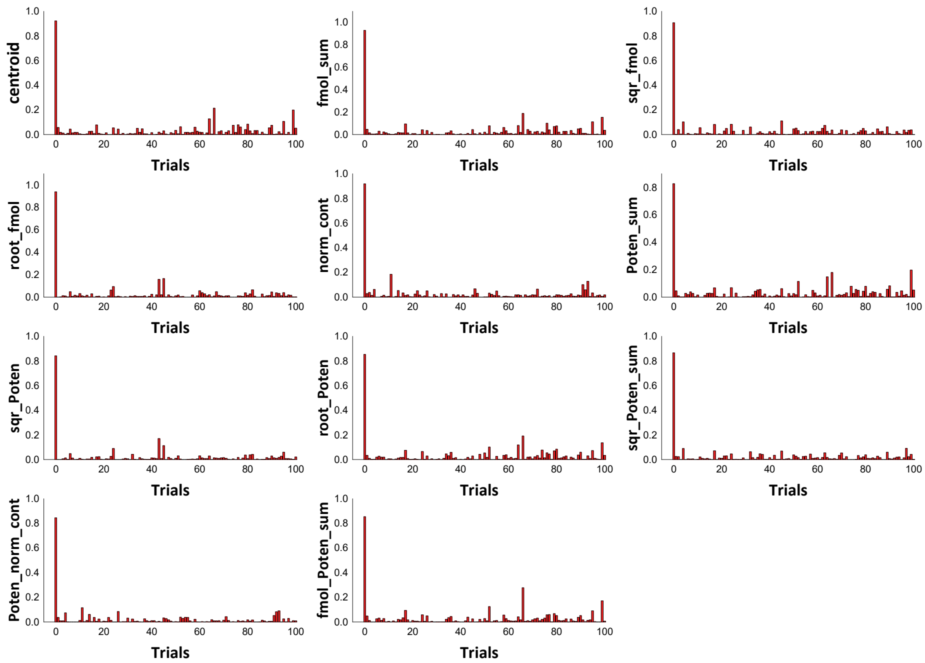

Fig. S-3. Y–randomization test applied to all QSAR model types; the first bar shows the $R^2_{training}$ value for corresponding original model.

Table S-I. Composition of mixtures appeared in the training set together with the observed and calculated boiling points by different QSPR models

Mixture composition			Observed	Calculated												
A	B	C		centroid	fmol-sum	sqr-fmol	root-fmol	sqr-fmol-sum	norm-cont	poten-sum	sqr-poten	root-poten	sqr-poten-sum	poten-norm-cont	fmol-poten-sum	Ensemble
Water	Carbon disulfide	Ethanol	314.4	313.7	309.8	316.9	313.6	316.8	315.1	336.8	337.6	334.7	339.3	314.3	332.9	323.5
Acetic acid	Pyridine	O-xylene	405.3	401.9	398.6	390.2	411.3	403.4	395.4	408.5	402.6	401.7	405.0	408.7	400.5	402.3
Water	Nitromethane	Dodecane	356.3	367.6	349.4	353.9	358.6	356.7	348.8	365.1	369.6	366.7	358.0	361.8	344.0	358.4
Methanol	Methyl acetate	Cyclohexane	323.9	331.4	334.3	341.8	330.4	334.0	324.6	326.2	331.3	332.5	331.9	333.0	339.1	332.6
Acetic acid	Pyridine	Octane	388.8	390.2	381.9	389.3	382.1	386.9	381.9	391.4	390.3	390.5	388.1	384.4	396.9	387.8
Water	Acetonitrile	Ethanol	346.0	350.0	339.5	343.0	340.0	341.2	342.0	345.8	335.9	330.3	342.1	334.9	334.1	339.9
Water	C-tetrachloride	Tert-butanol	337.8	342.3	337.5	335.5	340.8	341.6	337.3	337.9	337.8	336.1	336.9	340.9	336.7	338.4
Acetic acid	Benzene	Cyclohexane	350.3	360.5	347.1	337.6	353.1	346.9	347.0	349.8	352.6	353.7	349.2	355.8	337.2	349.2
Water	Butanol	Nonane	363.1	355.8	351.8	368.5	355.3	363.9	360.3	359.9	359.7	358.2	359.2	358.6	358.1	359.1
Acetic acid	Pyridine	Decane	407.2	402.3	409.4	408.8	403.6	402.3	412.6	386.3	387.9	394.2	404.2	396.4	387.8	399.7
Water	Chloroform	Methanol	325.4	324.2	320.5	327.6	326.4	325.2	328.9	331.5	330.6	333.1	334.8	336.3	331.2	329.2
Butanol	Pyridine	Toluene	381.8	376.4	377.5	392.3	379.1	385.9	385.5	378.7	378.1	387.9	379.8	382.7	375.2	381.6
Water	Butanol	Hexane	334.6	336.5	337.5	339.6	339.8	338.7	341.5	335.9	339.2	341.1	346.4	343.6	345.0	340.4
Water	Nitromethane	Octane	350.8	344.3	351.4	351.9	348.0	354.2	351.6	355.6	351.9	355.7	349.6	348.7	355.1	351.5
Water	2-butanone	Benzene	342.0	339.3	346.2	340.2	345.8	342.8	343.4	341.1	339.7	342.8	344.0	348.8	341.3	343.0
Water	Pyridine	Octane	359.8	356.8	361.7	357.7	358.4	352.8	361.2	361.1	358.7	360.3	354.4	350.2	363.4	358.0
Water	C-tetrachloride	Ethanol	334.9	338.4	337.7	336.0	337.6	339.0	337.6	342.0	337.1	338.9	339.9	338.2	335.4	338.1
Water	Pyridine	Decane	365.4	368.9	364.7	360.4	367.2	357.1	363.4	364.8	368.3	362.9	365.0	371.6	359.3	364.5
Water	Nitromethane	Undecane	355.9	362.1	349.9	353.7	357.2	356.9	349.4	360.3	364.3	359.8	361.4	364.4	346.3	357.1
Water	Isopropyl alcohol	Toluene	349.4	355.2	353.4	352.1	349.7	347.0	357.3	349.2	352.6	356.6	344.6	347.3	359.2	352.0
Chlorofor m	Methanol	Methyl acetate	329.5	329.4	325.0	333.9	326.3	325.4	324.8	319.4	326.5	331.6	326.0	325.6	334.3	327.4
Water	Trichloro ethylene	Acetonitrile	340.1	340.4	342.9	338.7	337.2	339.8	341.4	350.1	344.6	343.9	344.9	336.9	343.1	342.0
Chlorofor m	Ethanol	Acetone	341.3	336.3	331.7	336.0	337.7	332.4	339.7	343.5	329.9	332.5	340.7	333.7	330.8	335.4
Water	Acetonitrile	Benzene	339.1	341.2	341.5	338.8	336.6	337.7	341.8	340.7	350.9	340.6	339.2	348.2	340.2	341.5
Water	Nitromethane	Propyl alcohol	355.4	348.4	349.1	347.8	350.8	352.7	343.9	338.3	342.7	337.5	335.7	343.6	348.7	344.9
Water	Butanol	Octane	359.2	349.6	351.7	366.5	351.0	361.5	360.0	357.9	354.4	359.0	352.5	349.2	360.3	356.1
Water	Isobutanol	Toluene	354.4	355.8	357.0	363.3	351.8	349.4	361.6	343.1	342.9	355.7	342.1	349.0	360.3	352.7
Isopropyl alcohol	2-butanone	Cyclohexane	342.0	343.9	343.2	336.1	337.5	343.1	337.0	350.0	333.5	346.6	344.4	328.5	336.6	340.0
Water	Trichloro ethylene	Propyl alcohol	344.7	341.8	347.9	339.2	345.6	344.6	340.7	346.2	342.6	342.1	339.8	340.1	347.7	343.2
Water	Butanol	Heptane	351.2	343.4	349.9	360.2	349.1	354.6	355.7	342.7	343.7	348.3	347.5	343.3	356.7	349.6
Water	C-tetrachloride	Sec-butanol	338.1	342.0	336.6	335.6	338.3	340.0	335.7	337.9	337.8	336.2	336.8	340.9	338.4	338.0
Acetic acid	Pyridine	Ethylbenzene	402.2	404.5	400.1	393.4	404.5	405.0	397.3	402.9	406.7	400.4	401.5	397.4	404.3	401.5
Water	Nitromethane	1-decene	355.6	360.6	351.2	354.6	357.2	356.1	352.9	361.5	362.0	357.0	357.8	365.7	354.1	357.6
Water	Tert-butanol	Cyclohexane	338.1	339.8	342.9	338.1	337.6	333.4	335.9	340.1	338.3	344.1	337.5	344.4	342.0	339.5
Water	Ethanol	Cyclohexane	335.7	335.9	341.6	336.8	335.1	336.3	334.3	337.3	340.0	329.4	336.1	340.5	339.5	336.9
Water	Nitromethane	Nonane	353.8	350.4	352.1	352.7	354.5	356.1	352.5	358.1	357.2	356.8	356.9	357.2	355.7	355.0
Water	Nitromethane	Isopropyl alcohol	351.1	350.3	346.0	342.1	351.2	350.5	346.4	338.2	342.7	337.7	335.6	343.6	339.9	343.7
Water	Chloroform	Acetone	333.5	325.5	335.5	340.6	339.0	329.4	345.4	332.2	334.2	331.2	335.7	339.6	334.1	335.2
Water	Ethanol	Hexane	329.5	334.3	330.7	325.1	335.0	337.7	334.1	333.5	345.5	325.0	354.0	333.3	321.4	334.1
Water	Isopropyl alcohol	2-butanone	346.5	349.4	354.4	354.5	352.5	346.2	356.3	341.8	339.6	342.6	343.0	352.0	344.5	348.1
Water	Trichloro ethylene	Isopropyl alcohol	342.5	343.7	346.6	339.1	344.9	343.7	341.4	346.2	342.6	342.1	339.8	340.1	345.7	343.0
Water	Nitromethane	3-pentanone	355.5	349.8	356.6	346.3	347.3	353.0	346.4	342.9	340.2	347.1	344.5	341.4	350.3	347.1
Chlorofor m	Methanol	Acetone	330.6	326.6	326.6	334.3	326.1	324.2	328.0	323.4	327.9	326.7	330.2	330.1	332.0	328.0
Methanol	Acetone	Cyclohexane	324.2	328.6	335.0	338.8	329.2	331.1	332.4	323.2	330.4	324.9	327.3	323.1	333.8	329.8
Water	Ethanol	Heptane	341.9	341.2	345.6	343.5	341.8	346.8	346.1	351.9	358.7	342.8	353.8	346.9	339.3	346.5
Acetic	Pyridine	Nonane	401.1	396.4	401.9	398.6	399.2	399.7	402.5	387.3	387.6	391.0	400.5	392.9	407.4	397.1

acid																
Water	Butanol	Butylacetate	362.5	365.8	368.9	361.5	368.2	362.0	366.9	369.8	370.3	368.4	362.0	369.3	357.9	365.9
Water	Ethanol	Acrylonitrile	342.6	345.2	357.2	343.6	347.7	349.0	349.5	337.0	338.1	340.5	339.4	346.5	340.5	344.5
Water	Nitromethane	1-heptene	342.8	341.9	349.7	348.9	343.1	350.2	348.0	343.2	343.7	344.3	341.0	345.4	350.7	345.8
Water	Propyl alcohol	Cyclohexane	339.7	335.2	343.0	337.6	335.7	332.5	333.7	335.4	339.2	336.7	338.3	341.6	342.9	337.6
Isopropyl alcohol	Benzene	Cyclohexane	342.2	339.1	341.0	335.9	337.0	341.4	339.0	345.9	339.6	340.1	340.5	348.5	338.8	340.6
Water	Nitromethane	1-octene	350.5	348.9	353.7	352.7	354.1	356.1	352.6	358.0	352.1	356.9	347.1	349.7	358.0	353.3
Water	Chloroform	Ethanol	328.4	333.9	324.1	326.9	337.0	329.2	334.3	333.2	332.4	333.1	335.7	337.4	332.7	332.5
Water	Ethanol	Triethylamine	347.8	349.2	345.7	347.7	348.5	342.1	343.3	345.7	342.5	355.7	344.9	340.3	345.1	345.9

Table S-II. Composition of mixtures appeared in the test set together with the observed and calculated boiling points by different QSPR models

Mixture composition			Observed	Calculated												
A	B	C		centroid	fmol-sum	sqr-fmol	root-fmol	sqr-fmol- sum	norm-cont	poten-sum	sqr-poten	root-poten	sqr-poten-sum	poten-norm-cont	fmol-poten-sum	Ensemble
Water	Carbon tetrachloride	Propyl alcohol	338.5	337.8	340.0	336.8	337.3	338.5	337.1	339.6	337.4	338.5	337.9	339.1	337.7	338.1
Water	Carbon tetrachloride	2-butanone	338.8	334.4	342.7	338.5	339.5	344.6	341.2	338.2	337.7	335.2	337.0	340.5	335.7	338.8
Water	Carbon disulfide	Acetone	311.2	305.3	316.6	325.6	317.8	319.8	325.2	331.3	326.4	330.5	333.6	336.7	330.2	324.9
Water	Nitromethane	1-hexene	328.6	336.6	336.9	325.9	338.6	342.4	333.8	336.2	340.8	333.1	338.2	342.8	337.8	336.9
Water	Nitromethane	Hexane	330.0	331.2	332.8	327.5	339.4	338.3	332.9	335.2	340.6	333.6	339.5	342.1	336.7	335.8
Water	Nitromethane	Heptane	343.7	338.1	347.8	346.0	348.5	350.6	344.4	341.3	342.9	343.0	343.2	343.6	344.6	344.5
Water	Nitromethane	1-nonene	354.2	354.8	352.7	353.1	355.8	356.2	353.2	361.0	358.1	357.8	353.8	356.7	357.1	355.9
Water	Nitromethane	Decane	355.6	356.4	350.9	353.4	354.9	356.3	351.5	361.1	362.5	359.5	360.5	366.7	351.9	357.1
Water	Acetonitrile	Triethylamine	341.7	347.1	343.1	351.0	346.6	344.1	336.7	346.7	343.2	356.3	347.4	347.5	345.0	346.2
Water	1,2 dichloeroethane	Ethanol	340.9	342.6	342.0	332.8	339.8	346.2	337.8	339.7	328.8	332.6	346.5	336.2	340.5	338.8
Water	1,2 dichloeroethane	Isopropyl alcohol	342.8	343.9	342.4	333.0	340.1	343.7	336.9	336.9	334.0	337.5	343.7	342.1	341.7	339.7
Water	Ethanol	Ethyl acrylate	350.2	363.3	354.0	344.9	353.7	354.4	346.6	363.7	364.6	352.4	355.6	361.0	344.9	354.9
Water	Ethanol	Chlorobenzene	350.4	362.4	346.4	349.4	358.0	349.6	352.5	361.6	375.3	363.7	355.1	357.0	340.0	355.9
Water	Ethanol	Benzene	338.0	343.3	343.9	339.2	343.9	341.6	342.6	337.6	344.3	340.1	338.5	339.6	341.3	341.3
Water	Ethanol	Toluene	347.5	353.9	352.6	352.3	349.2	348.2	358.3	358.2	369.4	355.3	353.8	352.5	362.9	355.6
Water	Isopropyl alcohol	Benzene	339.6	344.6	344.1	338.8	344.6	341.6	341.5	335.2	339.8	345.4	338.0	343.2	341.1	341.5
Water	Isopropyl alcohol	Cyclohexane	337.4	337.1	342.4	337.2	335.1	333.4	334.2	335.2	339.2	337.0	337.2	341.2	341.6	337.6
Water	Propyl alcohol	Benzene	340.1	342.7	344.6	339.2	345.9	342.3	341.6	335.7	339.8	345.1	339.9	344.5	341.4	341.9
Water	2-butanone	Cyclohexane	336.7	331.9	346.5	339.4	336.9	338.8	339.0	338.7	339.2	338.0	342.0	346.3	338.8	339.6
Water	Butanol	Benzene	342.1	345.6	344.6	340.8	350.5	341.4	342.2	339.3	338.4	349.2	344.8	349.0	341.4	343.9
Water	Tert-butanol	Benzene	340.4	347.2	344.6	339.7	346.9	342.3	342.7	342.9	337.8	349.3	337.5	343.8	341.3	343.0
Water	Pyridine	Nonane	363.6	362.9	361.9	360.9	365.5	357.0	363.2	362.5	363.5	357.4	361.9	361.0	362.5	361.7
Chloroform	Ethanol	Hexane	330.4	326.8	324.2	328.5	327.1	333.5	332.3	349.7	331.5	329.4	340.8	329.2	331.4	332.0
Acetic acid	Pyridine	P-xylene	402.3	404.2	390.7	398.8	401.6	404.0	395.1	413.7	421.6	418.9	422.6	415.6	413.5	408.4

Table S-III. $R^2_{training}$, R^2_{test}, and q^2 of MLR models with different sizes developed based on different mixture descriptors

Model size	centroid	fmol–sum	sqr–fmol	root–fmol	sqr–fmol–sum	norm–cont	poten–sum	sqr–poten	root–poten	sqr–poten–sum	poten–norm–cont	fmol–poten–sum
$R^2_{training}$												
2	0.79	0.77	0.57	0.79	0.79	0.69	0.75	0.77	0.75	0.82	0.73	0.75
3	0.88	0.89	0.79	0.86	0.85	0.83	0.80	0.79	0.81	0.85	0.82	0.83
4	0.89	0.90	0.84	0.9	0.92	0.88	0.82	0.82	0.83	0.85	0.84	0.86
5	0.91	0.93	0.9	0.92	0.95	0.92	0.83	0.83	0.86	0.87	0.86	0.87
6	0.95	0.94	0.91	0.93	0.96	0.94	0.87	0.84	0.86	0.88	0.89	0.88
7	0.95	0.95	0.93	0.95	0.96	0.94	0.88	0.87	0.89	0.89	0.89	0.88
8	0.96	0.95	0.93	0.95	0.96	0.95	0.89	0.87	0.9	0.91	0.9	0.92
9	0.96	0.96	0.94	0.96	0.98	0.95	0.93	0.88	0.9	0.91	0.9	0.91
10	0.97	0.96	0.95	0.96	0.97	0.95	0.91	0.88	0.92	0.92	0.91	0.92
R^2_{test}												

Model size	centroid	fmol–sum	sqr–fmol	root–fmol	sqr–fmol–sum	norm–cont	poten–sum	sqr–poten	root–poten	sqr–poten–sum	poten–norm–cont	fmol–poten–sum
2	0.88	0.68	0.55	0.71	0.74	0.73	0.79	0.75	0.75	0.81	0.76	0.71
3	0.86	0.85	0.74	0.82	0.85	0.81	0.78	0.78	0.82	0.83	0.78	0.82
4	0.9	0.89	0.86	0.88	0.93	0.87	0.80	0.81	0.81	0.84	0.84	0.82
5	0.91	0.93	0.89	0.9	0.92	0.89	0.81	0.81	0.82	0.87	0.86	0.85
6	0.92	0.92	0.91	0.93	0.94	0.92	0.83	0.80	0.87	0.88	0.84	0.86
7	0.94	0.95	0.91	0.92	0.94	0.93	0.84	0.84	0.85	0.87	0.84	0.87
8	0.95	0.96	0.92	0.94	0.95	0.94	0.87	0.87	0.87	0.89	0.86	0.87
9	0.95	0.95	0.91	0.96	0.97	0.95	0.88	0.84	0.88	0.9	0.88	0.87
10	0.95	0.96	0.94	0.96	0.97	0.93	0.89	0.87	0.9	0.92	0.87	0.9
q^2												
2	0.77	0.73	0.46	0.77	0.76	0.63	0.71	0.74	0.71	0.81	0.69	0.7
3	0.86	0.87	0.74	0.84	0.82	0.81	0.77	0.76	0.77	0.83	0.79	0.79
4	0.87	0.88	0.78	0.87	0.91	0.86	0.79	0.77	0.8	0.83	0.81	0.81
5	0.89	0.91	0.86	0.9	0.93	0.9	0.79	0.78	0.82	0.85	0.83	0.83
6	0.93	0.91	0.87	0.91	0.94	0.91	0.83	0.8	0.82	0.85	0.86	0.84
7	0.93	0.92	0.89	0.93	0.94	0.91	0.84	0.82	0.85	0.87	0.86	0.84
8	0.94	0.93	0.9	0.94	0.95	0.92	0.84	0.82	0.86	0.89	0.86	0.88
9	0.94	0.93	0.91	0.94	0.96	0.93	0.89	0.82	0.86	0.89	0.87	0.87
10	0.95	0.94	0.93	0.95	0.96	0.93	0.86	0.83	0.88	0.89	0.87	0.87

Table S-IV. Statistical parameters of the QSPR models based on centroid descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.79	9.26	0.77	9.83	0.76	9.94	0.79	0.75	0.88	6.44	97.09
3	0.88	6.91	0.86	7.54	0.86	7.65	0.88	0.84	0.86	7.29	127.2
4	0.89	6.75	0.87	7.40	0.86	7.53	0.89	0.84	0.90	6.53	98.74
5	0.91	6.00	0.89	6.77	0.88	6.90	0.91	0.87	0.91	5.13	100.4
6	0.95	4.68	0.93	5.38	0.93	5.49	0.95	0.91	0.92	5.40	139.7
7	0.95	4.47	0.93	5.28	0.93	5.42	0.95	0.91	0.94	5.00	128.9
8	0.96	4.20	0.94	5.06	0.93	5.20	0.96	0.92	0.95	3.91	125.6
9	0.96	4.07	0.94	5.06	0.93	5.20	0.96	0.92	0.95	3.57	116.7
10	0.97	3.71	0.95	4.64	0.94	4.89	0.97	0.94	0.95	3.44	124.4

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-V. Statistical parameters of the QSPR models based on fmol-sum descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.77	9.74	0.73	10.56	0.73	10.59	0.77	0.73	0.68	9.18	85.46
3	0.89	6.75	0.87	7.31	0.87	7.46	0.89	0.83	0.85	6.27	134.32
4	0.90	6.38	0.88	7.02	0.88	7.20	0.90	0.85	0.89	5.21	112.04
5	0.93	5.31	0.91	5.99	0.91	6.10	0.93	0.89	0.93	5.09	130.96
6	0.94	5.06	0.91	6.04	0.91	6.22	0.94	0.89	0.92	4.66	118.40
7	0.95	4.68	0.92	5.72	0.91	6.12	0.95	0.91	0.95	4.04	117.12
8	0.95	4.57	0.93	5.47	0.92	5.62	0.95	0.90	0.96	3.57	123.10
9	0.96	4.29	0.93	5.27	0.93	5.46	0.96	0.91	0.95	4.45	120.31

10	0.96	3.93	0.94	4.90	0.94	5.11	0.96	0.93	0.96	3.96	110.68

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-VI. Statistical parameters of the QSPR models based on sqr-fmol descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.57	13.28	0.46	14.91	0.45	15.29	0.57	0.45	0.55	11.64	34.12
3	0.79	9.33	0.74	10.33	0.74	10.47	0.79	0.70	0.74	9.53	62.39
4	0.84	8.18	0.78	9.49	0.75	10.22	0.84	0.73	0.86	6.00	63.26
5	0.90	6.46	0.86	7.62	0.84	8.02	0.90	0.83	0.89	5.35	85.21
6	0.91	6.19	0.87	7.37	0.85	7.85	0.91	0.83	0.91	4.88	76.62
7	0.93	5.37	0.89	6.66	0.87	7.29	0.93	0.86	0.91	4.86	87.43
8	0.93	5.24	0.90	6.48	0.89	6.80	0.94	0.86	0.92	4.44	78.99
9	0.94	5.08	0.91	6.15	0.89	6.59	0.94	0.86	0.91	5.08	73.17
10	0.95	4.52	0.93	5.50	0.90	6.35	0.95	0.88	0.94	5.17	82.42

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-VII. Statistical parameters of the QSPR models based on root-fmol descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.79	9.24	0.77	9.69	0.77	9.75	0.79	0.74	0.71	8.85	97.83
3	0.86	7.47	0.84	8.08	0.84	8.17	0.86	0.81	0.82	7.32	106.71
4	0.90	6.27	0.87	7.23	0.87	7.35	0.90	0.84	0.88	5.71	116.50
5	0.92	5.77	0.90	6.47	0.89	6.62	0.92	0.86	0.90	5.13	109.24
6	0.93	5.42	0.91	6.07	0.91	6.25	0.93	0.88	0.93	4.45	102.21
7	0.95	4.67	0.93	5.48	0.92	5.77	0.95	0.90	0.92	4.70	117.74
8	0.96	4.09	0.94	4.94	0.94	5.01	0.96	0.93	0.94	4.96	133.34
9	0.96	4.16	0.94	5.13	0.93	5.32	0.96	0.90	0.96	3.69	111.86
10	0.96	3.84	0.95	4.74	0.94	4.88	0.97	0.93	0.96	3.92	116.17

[a]The reported values are average of 100 repetitions of 5-fold CV.
[b]The reported values are average of 100 repetitions of MCCV.

Table S-VIII. Statistical parameters of the QSPR models based on norm-count descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.69	11.24	0.63	12.36	0.62	12.61	0.69	0.60	0.73	8.60	57.77
3	0.83	8.29	0.81	8.76	0.81	8.85	0.83	0.79	0.81	7.26	83.46
4	0.88	7.03	0.86	7.62	0.85	7.80	0.88	0.82	0.87	5.82	89.89
5	0.92	5.59	0.90	6.36	0.90	6.61	0.92	0.88	0.89	5.50	117.02
6	0.94	5.17	0.91	6.12	0.90	6.47	0.94	0.89	0.92	4.84	113.09
7	0.94	5.16	0.91	6.05	0.90	6.39	0.94	0.90	0.93	4.18	95.11
8	0.95	4.60	0.92	5.66	0.92	5.82	0.95	0.91	0.94	4.83	104.15

9	0.95	4.39	0.93	5.33	0.92	5.64	0.95	0.91	0.95	4.28	100.01
10	0.95	4.51	0.93	5.30	0.92	5.67	0.95	0.89	0.93	4.20	82.74

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-IX. Sttistical parameters of the QSPR models based on Poten-sum descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.75	10.23	0.71	10.85	0.71	10.95	0.74	0.69	0.79	9.39	75.11
3	0.80	9.17	0.77	9.79	0.76	9.92	0.80	0.75	0.78	10.12	65.10
4	0.82	8.54	0.79	9.32	0.78	9.47	0.83	0.77	0.80	10.62	56.99
5	0.83	8.45	0.79	9.27	0.79	9.41	0.83	0.78	0.81	8.73	45.85
6	0.87	7.38	0.83	8.29	0.83	8.50	0.87	0.82	0.83	8.22	51.52
7	0.88	6.94	0.84	8.05	0.84	8.24	0.89	0.83	0.84	7.91	49.77
8	0.89	6.87	0.84	8.07	0.83	8.34	0.89	0.84	0.88	7.61	43.60
9	0.93	5.48	0.89	6.82	0.86	7.87	0.93	0.87	0.87	7.37	62.19
10	0.91	6.14	0.86	7.56	0.85	7.89	0.91	0.84	0.89	6.82	48.63

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-X. Statistical parameters of the QSPR models based on root-Poten descriptors.

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.75	10.08	0.71	10.87	0.71	11.00	0.76	0.72	0.75	8.64	78.10
3	0.81	8.84	0.77	9.65	0.77	9.78	0.81	0.77	0.82	7.45	71.42
4	0.83	8.32	0.80	9.14	0.79	9.30	0.83	0.78	0.81	8.53	60.82
5	0.86	7.63	0.82	8.56	0.82	8.71	0.86	0.79	0.82	7.85	58.48
6	0.86	7.57	0.82	8.55	0.82	8.74	0.87	0.78	0.87	7.47	48.51
7	0.89	6.75	0.85	7.95	0.84	8.09	0.89	0.82	0.85	7.93	52.91
8	0.90	6.32	0.86	7.66	0.85	7.85	0.91	0.83	0.87	6.35	52.56
9	0.90	6.31	0.86	7.58	0.85	7.86	0.90	0.84	0.88	7.67	52.59
10	0.92	5.91	0.88	7.18	0.87	7.39	0.92	0.86	0.90	7.50	52.84

[a] The reported values are average of 100 repetitions of 5-fold CV
[b] The reported values are average of 100 repetitions of MCCV

Table S-XI. Statistical parameters of the QSPR models based on sqr-Poten-sum descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a] q^2	5-CV[a] RMSE	MCCV[b] $R^2_{training}$	MCCV[b] R^2_{test}	R^2_{test}	RMSE-Test	F
2	0.82	8.54	0.81	8.86	0.81	8.94	0.82	0.75	0.81	8.98	118.91
3	0.85	8.00	0.83	8.45	0.82	8.55	0.85	0.77	0.83	9.22	90.92
4	0.85	7.80	0.83	8.31	0.83	8.40	0.85	0.76	0.84	9.32	70.92
5	0.87	7.43	0.85	7.88	0.85	7.96	0.86	0.78	0.87	7.73	62.09
6	0.88	7.15	0.85	7.92	0.84	8.07	0.88	0.79	0.88	6.95	55.35
7	0.89	6.82	0.87	7.43	0.86	7.70	0.89	0.82	0.87	8.47	51.70
8	0.91	6.03	0.89	6.62	0.89	6.89	0.91	0.85	0.89	7.23	58.16

9	0.91	6.24	0.89	6.86	0.88	7.19	0.91	0.84	0.90	7.25	53.91
10	0.92	5.86	0.89	6.75	0.89	6.90	0.92	0.85	0.92	6.76	61.88

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-XII. Statistical parameters of the QSPR models based on Poten-norm-cont descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a]		MCCV[b]		R^2_{test}	RMSE-Test	F
					q^2	RMSE	$R^2_{training}$	R^2_{test}			
2	0.73	10.64	0.69	11.32	0.68	11.46	0.72	0.66	0.76	11.32	67.40
3	0.82	8.70	0.79	9.35	0.79	9.42	0.82	0.78	0.78	8.63	74.18
4	0.84	8.11	0.81	8.84	0.80	8.98	0.84	0.77	0.84	8.20	64.58
5	0.86	7.70	0.83	8.47	0.82	8.63	0.86	0.78	0.86	7.68	57.24
6	0.89	6.81	0.86	7.69	0.85	8.01	0.89	0.82	0.84	8.56	61.78
7	0.89	6.65	0.86	7.58	0.85	7.95	0.89	0.82	0.84	8.49	54.77
8	0.90	6.44	0.86	7.57	0.85	8.06	0.90	0.81	0.86	8.39	50.26
9	0.90	6.34	0.87	7.38	0.83	8.53	0.90	0.80	0.88	7.64	45.30
10	0.91	6.22	0.87	7.46	0.78	10.10	0.91	0.80	0.87	8.01	41.52

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-XIII. Statistical parameters of the QSPR models based on fmol-Poten-sum descriptors

No. of descriptors	$R^2_{training}$	RMSE-Training	q^2	RMSE-CV	5-CV[a]		MCCV[b]		R^2_{test}	RMSE-Test	F
					q^2	RMSE	$R^2_{training}$	R^2_{test}			
2	0.75	10.24	0.70	11.18	0.69	11.49	0.75	0.66	0.71	9.51	74.84
3	0.83	8.45	0.79	9.33	0.78	9.76	0.83	0.79	0.82	7.44	79.60
4	0.86	7.64	0.81	8.82	0.81	9.07	0.86	0.83	0.82	7.14	74.36
5	0.87	7.29	0.83	8.34	0.81	8.91	0.87	0.80	0.85	6.70	65.01
6	0.88	6.98	0.84	8.21	0.82	8.75	0.88	0.79	0.86	6.78	58.60
7	0.88	6.92	0.84	8.17	0.83	8.47	0.89	0.84	0.87	6.98	50.08
8	0.92	5.91	0.88	6.92	0.88	7.16	0.92	0.88	0.87	7.10	84.72
9	0.91	6.03	0.87	7.43	0.85	7.87	0.91	0.82	0.87	7.28	58.11
10	0.92	5.74	0.87	7.30	0.86	7.66	0.92	0.82	0.90	6.54	56.24

[a]The reported values are average of 100 repetitions of 5-fold CV
[b]The reported values are average of 100 repetitions of MCCV

Table S-XIV. QSPR equation formulas established by MLR analysis for different types of mixture descriptors along with their statistical parameters

mixture descriptor	QSPR equation	$R^2_{training}$	q^2	R^2_{test}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC	r^2_m
centroid	y = 349.75 + 7.35 (±1.094) Ss + 11.04 (±1.060) piPC05 - - 4.70(±0.986) MATS2e + 8.75 (±0.869) a_donacc - - 8.65 (±0.978) GCUT_SMR_0 + 6.65 (±1.302) vsurf_D7	0.91	0.89	0.91	0.90	0.88	0.93	0.96	0.90
fmol–sum	y = 349.75 -6.33 (±0.966) BAC + 4.47 (±1.485) E2e + + 31.26 (±2.404) HATS1u + 5.44 (±1.262) vsurf_HL2 + + 36.20 (±2.637) vsurf_W1	0.9	0.88	0.89	0.91	0.90	0.94	0.95	0.87

sqr–fmol	*y* = 349.75 -9.11 (±0.976) MSD + 9.92 (±1.107) piPC06 + + 5.79 (±1.016) DP17 -5.80 (±0.984) GCUT_PEOE_1 - - 2.49 (±1.202) PEOE_VSA+1 + 10.89 (±1.059) vsurf_HL2	0.9	0.86	0.89	0.92	0.91	0.94	0.95	0.82
root–fmol	*y* = 349.75 + 9.60 (±1.499) ZM2V -5.86 (±1.143) MSD + + 1.38 (±1.025) piPC06 -5.36 (±1.110) E1e -3.60 (±0.854) G2s + + 4.49 (±0.973) R1m + 5.61 (±0.815) vsurf_DD23 + + 10.29 (±1.079) vsurf_HL2	0.95	0.93	0.92	0.92	0.90	0.94	0.98	0.92
sqr–fmol–sum	*y* = 349.75 + 8.59 (±1.010) ZM2V -7.22 (±1.445) X0A + + 5.25 (±1.238) E2e + 3.49 (±0.739) HATS3u + + 15.69 (±1.645) R1u+ + 23.16 (±1.776) vsurf_W1	0.95	0.93	0.92	0.92	0.90	0.94	0.97	0.91
norm–cont	*y* = 349.75 + 7.06 (±0.918) piPC06 + 3.87 (±0.975) DP17 + 23.40 (±2.465) R1u+ + 3.37 (±0.875) FCASA - - 17.87 (±2.529) Q_VSA_FPPOS + 11.55 (±1.109) vsurf_HL2	0.92	0.9	0.89	0.92	0.91	0.94	0.95	0.87
poten–sum	*y* = 349.75 + 18.52 (±2.646) SRW10 -6.76 (±2.317) X5A + + 8.35 (±1.411) BIC2 -2.87 (±1.859) RDF045v + +7.47 (±1.658) E2e + 10.65 (±1.790) R7v+	0.83	0.79	0.81	0.77	0.73	0.84	0.92	0.75
sqr–poten	*y* = 349.75 + 8.44 (±1.964) piPC06 + 5.88 (±2.093) Mor11m - 4.92 (±1.589) E1s + 7.22 (±1.599) R7v+ + + 4.07 (±1.858) DCASA - 6.06 (±1.994) vsurf_EDmin3	0.83	0.79	0.81	0.64	0.58	0.74	0.88	0.70
root–poten	*y* = 349.75 + 3.10 (±2.067) SPI + 19.34 (±2.881) SRW01 - - 8.49 (±2.903) X5A -6.48 (±1.663) X2Av - -8.87 (±1.534) Mor32v + 7.70 (±2.123) R7v+ + + 3.07 (±1.447) Q_PC+	0.86	0.82	0.87	0.78	0.75	0.85	0.93	0.79
sqr–poten–sum	*y* = 349.75 + 10.13 (±3.632) EEig01x + 15.22 (±2.486) EEig02r - - 6.02 (±1.365) EEig06r + 1.77 (±1.165) JGI2 + 8.83 (±2.494) E2e	0.85	0.83	0.84	0.79	0.76	0.86	0.94	0.74
poten–norm–cont	*y* = 349.75 -19.70 (±2.358) SIC5 -9.62 (±1.468) EEig05d + + 4.80 (±1.560) GGI4 -4.14 (±1.046) L/Bw - 7.65 (±1.424) VAdjEq - 3.13 (±1.706) vsurf_IW1	0.86	0.83	0.86	0.75	0.71	0.82	0.93	0.78
fmol–poten–sum	*y* = 349.75 + 12.48 (±1.634) piPC06 + 3.52 (±1.611) GGI2 + + 14.84 (±1.446) DP17 -7.37 (±1.939) Mor26e + + 4.67 (±1.272) Q_PC+	0.86	0.81	0.82	0.85	0.82	0.89	0.93	0.75
Ensemble model	---	0.97	0.96	0.95	0.92	0.91	0.95	0.98	0.87