Journal of Molecular Liquids 208 (2015) 269–279

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

Predicting the cytotoxicity of ionic liquids using QSAR model based on SMILES optimal descriptors

Abdolmohammad Ghaedi

Department of Chemistry, Gachsaran Branch, Islamic Azad University, P.O. Box 75818-63876, Gachsaran, Iran

ARTICLE INFO

Article history:
Received 24 January 2015
Received in revised form 21 April 2015
Accepted 22 April 2015
Available online xxxx

Keywords:
Cytotoxicity
Ionic liquids
QSAR
SMILES
CORAL software

ABSTRACT

In recent years, ionic liquids (ILs) have attracted great attention due to their remarkable physico-chemical properties. In spite of all advantages of ILs, these compounds probably cause persistent contaminations and constitute environmental risks. Cytotoxicity ($Log_{10}(CE_{50})$) data 225 ILs were used in the creation of quantitative structure–activity relationship (QSAR) models by CORrelation And Logic (CORAL) software (http://www.insilico.eu/coral) in which molecular structures are indicated by the simplified molecular input line entry system (SMILES) notation. These global SMILES descriptors are illustrating the presence of some chemical elements and various types of chemical bonds (double, triple, and stereochemical). The balance of correlations (BC) of constructing a quantitative structure–property/activity relationships and the classic scheme were compared. Results from three random splits displayed reliable models for predicting the external test set with the correlation coefficient (R^2) and cross validated correlation coefficient (Q^2) in ranges of 0.7315–0.8760 and 0.7062–0.8490, respectively. The best predictions obtained by a classic scheme along with the global SMILES descriptors are included in the modeling process. The average statistical characteristics of the external test set are the following: n = 44, R^2 = 0.8760, Q^2 = 0.8540, standard error of estimation (s) = 0.529, mean absolute error (MAE) = 0.400 and Fischer F-ratio (F) = 297.

1. Introduction

Ionic liquids are ionic and salt materials in the liquid state with a melting temperature below 100 °C and typically formed by large organic cations with delocalized charges and organic/inorganic anions, which are weakly coordinated to the organic cation [1]. They have recently attracted great attention due to their remarkable physico-chemical properties, such as thermal stability, low or negligible vapor pressure, non-inflammability, non-volatile, electrochemical stability, and favorable solvating properties for a wide range of polar and non-polar compounds [2]. Due to a negligible vapor pressure up to very high temperatures, which avoids the loss of solvent to the atmosphere and decreases the worker exposure risk, ILs have been considered as environmentally friendly and green solvents for the chemical, pharmaceutical and a wide range of industrial applications including separation, organic synthesis, sensors, catalysis, fuel cells, biocatalysis, batteries, electrochemistry, solar cells, capacitors, chromatography columns, biomass processing, extractants, drug delivery and in the reprocessing of nuclear waste [3–15].

In spite of all the advantages of ionic liquids it should be mentioned that increased industrial manufacture and application of these compounds probably cause persistent contaminations and constitute environmental risks [16]. Although their use is figured out not to induce any air pollution because of their low vapor pressure, some properties of ionic liquids such as high solubility in water, non-volatile, high chemical and thermal stability may consequently cause serious water pollution and show a potential toxicity risk for human beings and the environment [17,18]. In recent years, some studies have indicated that ILs have toxic impacts on human, bacteria, invertebrates, plants, and fish [19–23]. According to these results, ILs might not uncritically be considered as intrinsically "green", but it powerfully depends on their substructures. On the other hand, there is an enormous number of ILs which can be synthesized. Measuring the toxicity information of diverse ILs under a wide range of conditions using experimental methods is time-consuming, costly, resource-wasting and impractical. For this reason it is necessary and significant to develop and apply new and available models for the assessment of the toxicity of ILs without an experimental procedure [24–26].

Nowadays, quantitative structure–activity/property relationships (QSAR/QSPR) have been successfully used as a good alternative to provide information that is suitable for molecular design and medicinal chemistry [25,27]. The QSPR/QSAR model is a mathematical equation which has a main purpose of achieving an accurate and reliable mathematical equation that relates the chemical and physical properties or biological activity to a wide variety of descriptors. The QSPR/QSAR models have been applied widely to forecast the various properties of compounds, such as the toxicity of ILs compounds [25–32].

E-mail address: abm_ghaedi@yahoo.com.

http://dx.doi.org/10.1016/j.molliq.2015.04.049
0167-7322/

Table 1
Statistical characteristics of the models. The best models are displayed by bold.

T	Nact	Probe	Training set				Calibration set				Testing (validation) set				
			N_s	R^2_s	S_s	F_s	N_c	R^2_c	S_c	F_c	N_v	R^2_v	S_v	F_v	R^2_m
Version 1. Balance of correlations															
Split 1															
1	287	1	120	0.8766	0.316	838	60	0.8746	0.423	404	45	0.8675	0.527	282	0.7498
1	287	2	120	0.8754	0.318	829	60	0.874	0.424	402	45	0.8666	0.53	279	0.7471
1	287	3	120	0.8757	0.318	831	60	0.8751	0.421	406	45	**0.8678**	0.528	282	0.7451
Av	287	—	120	0.8759	0.317	833	60	0.8746	0.423	404	45	0.8673	0.528	281	0.7473
2	247	1	120	0.8761	0.317	834	60	0.8707	0.43	391	45	0.849	0.559	242	0.7368
2	247	2	120	0.8749	0.319	825	60	0.8712	0.429	392	45	0.85	0.556	244	0.7443
2	247	3	120	0.8746	0.319	823	60	0.8683	0.433	382	45	0.8491	0.556	242	0.7435
Av	247	—	120	0.8752	0.318	827	60	0.8701	0.431	388	45	0.8494	0.557	243	0.7415
3	209	1	120	0.8675	0.328	773	60	0.8671	0.435	379	45	0.852	0.559	248	0.7379
3	209	2	120	0.8709	0.324	796	60	0.868	0.433	381	45	0.8522	0.555	248	0.7371
3	209	3	120	0.8682	0.327	777	60	0.8674	0.435	379	45	0.8527	0.56	249	0.7343
Av	209	—	120	0.8689	0.326	782	60	0.8675	0.434	380	45	0.8523	0.558	248	0.7365
Split 2															
1	301	1	151	0.8756	0.391	1049	35	0.8035	0.57	135	39	0.7031	0.401	88	0.5989
1	301	2	151	0.8733	0.395	1027	35	0.804	0.572	135	39	0.7101	0.397	91	0.6039
1	301	3	151	0.8743	0.394	1036	35	0.8017	0.573	133	39	0.7214	0.39	96	0.6134
1	301	—	151	0.8744	0.393	1037	35	0.8031	0.572	135	39	0.7115	0.396	91	0.6054
2	249	1	151	0.8726	0.396	1021	35	0.8047	0.572	136	39	0.7658	0.36	121	0.6628
2	249	2	151	0.8715	0.398	1011	35	0.8024	0.576	134	39	**0.8063**	0.326	154	0.7186
2	249	3	151	0.8727	0.396	1022	35	0.8039	0.572	135	39	0.7715	0.355	125	0.6759
2	249	—	151	0.8723	0.397	1018	35	0.8037	0.573	135	39	0.7812	0.347	133	0.6858
3	218	1	151	0.8746	0.393	1039	35	0.801	0.573	133	39	0.8052	0.328	153	0.7019
3	218	2	151	0.8746	0.393	1039	35	0.798	0.577	130	39	0.8028	0.333	151	0.689
3	218	3	151	0.8736	0.395	1030	35	0.7978	0.578	130	39	0.8051	0.335	153	0.6867
3	218	—	151	0.8743	0.394	1036	35	0.7989	0.576	131	39	0.8044	0.332	152	0.6925
4	203	1	151	0.8667	0.405	969	35	0.8047	0.573	136	39	0.7886	0.34	138	0.6852
4	203	2	151	0.871	0.399	1006	35	0.8015	0.575	133	39	0.7706	0.356	124	0.6651
4	203	3	151	0.8694	0.401	992	35	0.8028	0.572	134	39	0.7907	0.34	140	0.6831
4	203	—	151	0.8691	0.402	989	35	0.803	0.573	135	39	0.7833	0.346	134	0.6778
Split 3															
1	305	1	168	0.8659	0.366	1072	27	0.7812	0.634	89	30	0.7379	0.54	79	0.6887
1	305	2	168	0.8646	0.368	1060	27	0.7809	0.633	89	30	0.7923	0.476	107	0.7909
1	305	3	168	0.8658	0.366	1071	27	0.7804	0.634	89	30	0.7531	0.523	85	0.7175
1	305	—	168	0.8654	0.367	1068	27	0.7808	0.634	89	30	0.7611	0.513	90	0.7324
2	249	1	168	0.8604	0.374	1023	27	0.7787	0.637	88	30	0.8383	0.426	145	0.8031
2	249	2	168	0.8612	0.373	1030	27	0.7785	0.637	88	30	0.8234	0.442	131	0.8008
2	249	3	168	0.8631	0.37	1046	27	0.7783	0.638	88	30	0.826	0.441	133	0.8027
2	249	—	168	0.8616	0.372	1033	27	0.7785	0.638	88	30	0.8292	0.436	136	0.8022
3	217	1	168	0.8594	0.375	1014	27	0.778	0.638	88	30	0.8409	0.423	148	0.804
3	217	2	168	0.8609	0.373	1027	27	0.778	0.639	88	30	0.8327	0.433	139	0.8037
3	217	3	168	0.8614	0.372	1032	27	0.778	0.638	88	30	**0.8414**	0.423	149	0.801
3	217	—	168	0.8606	0.373	1024	27	0.778	0.638	88	30	0.8383	0.426	145	0.8029
4	206	1	168	0.8562	0.379	988	27	0.7781	0.637	88	30	0.8385	0.425	145	0.8065
4	206	2	168	0.8573	0.378	997	27	0.7776	0.638	87	30	0.8254	0.443	132	0.7992
4	206	3	168	0.8564	0.379	990	27	0.778	0.637	88	30	0.8222	0.446	129	0.806
4	206	—	168	0.8566	0.379	992	27	0.7779	0.637	88	30	0.8287	0.438	136	0.8039
Version 2. Classic scheme															
Split 1															
1	307	1	181	0.8791	0.351	1301					44	0.8712	0.532	284	0.7321
1	307	2	181	0.8792	0.351	1303					44	0.8709	0.535	283	0.7284
1	307	3	181	0.8791	0.351	1302					44	0.869	0.533	279	0.7343
1	307	—	181	0.8791	0.351	1302					44	0.8704	0.533	282	0.7316
2	251	1	181	0.8792	0.351	1303					44	0.8718	0.526	286	0.7424
2	251	2	181	0.8791	0.351	1302					44	0.8686	0.533	278	0.7341
2	251	3	181	0.879	0.351	1301					44	0.8705	0.532	282	0.7357
2	251	—	181	0.8791	0.351	1302					44	0.8703	0.53	282	0.7374
3	232	1	181	0.8779	0.353	1287					44	0.8747	0.533	293	0.7239
3	232	2	181	0.8779	0.353	1287					44	0.8732	0.531	289	0.7297
3	232	3	181	0.8777	0.353	1284					44	0.8751	0.533	294	0.725
3	232	—	181	0.8778	0.353	1286					44	0.8743	0.533	292	0.7262
4	228	1	181	0.8779	0.353	1286					44	0.8734	0.531	290	0.7296
4	228	2	181	0.8772	0.354	1279					44	**0.876**	0.528	297	0.7296
4	228	3	181	0.8778	0.353	1286					44	0.8751	0.536	294	0.7212
4	228	—	181	0.8776	0.353	1284					44	0.8748	0.532	294	0.7268
5	201	1	181	0.8745	0.358	1247					44	0.8606	0.564	259	0.7067
5	201	2	181	0.8739	0.358	1240					44	0.8598	0.569	258	0.7005
5	201	3	181	0.8753	0.356	1257					44	0.8614	0.567	261	0.7007
5	201	—	181	0.8746	0.357	1248					44	0.8606	0.566	259	0.7026