Ebrahimpoor Gorji, Ali; Laakso, Juho-Pekka; Alopaeus, Ville; Uusi-Kyyny, Petri

Towards the prediction of CPA pure parameters for imidazolium, ammonium, and pyridinium based-ionic liquids using QSPR study: A comparative study

Published in:
Chemical Engineering Science

DOI:
10.1016/j.ces.2024.120825

Published: 05/02/2025

Document Version
Publisher's PDF, also known as Version of record

Chemical Engineering Science 302 (2025) 120825

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Towards the prediction of CPA pure parameters for imidazolium, ammonium, and pyridinium based-ionic liquids using QSPR study: A comparative study

Ali Ebrahimpoor Gorji, Juho-Pekka Laakso, Ville Alopaeus *, Petri Uusi-Kyyny

Aalto University, School of Chemical Technology, Department of Chemical and Metallurgical Engineering, Research Group of Chemical Engineering, P.O. Box 16100, FI-00076 Aalto, Finland

ARTICLE INFO

Keywords:
CPA pure parameters
QSPR
Ionic Liquids
Prediction
Thermodynamic model

ABSTRACT

The 'Quantitative Structure-Property Relationship' (QSPR) method for the prediction of CPA pure parameters has been applied on three datasets. New predictive QSPR models including new molecular descriptors have been developed and have been compared with former models. According to the obtained results of statistical parameters ($R^2 > 0.90$ and $Q^2_{LOO\text{-}CV} > 0.90$), the predictive capabilities of the QSPR models were better for both of training and test sets than former models. It was shown that CPA parameters for some new ILs can be predicted using QSPR models. The capability of the QSPR models were assessed by calculating density and vapor pressure of ILs. It was shown that the predicted parameters by QSPR models could predict density for none-studied ILs with *AARD of 5.0* % and have qualitatively low vapor pressures. A general QSPR model with cationic (ECCEN) and anionic descriptors (MW) for the prediction of 'b' parameter of huge number of ILs has been developed.

1. Introduction

In the past two decades, Ionic Liquids (ILs) have been considered as promising solvents (Maia et al., 2012). This kind of solvents are not only is an organic salt, but also is liquid at ambient temperature. ILs are constituted by two separate parts called 'organic cation' and 'organic or inorganic anion'. According to the sort of cations, types of ILs are often classified as Imidazolium, Ammonium, Pyridinium, Phosphonium etc. Due to some intrinsic and unique features of ILs, such as high degradation temperature, high selectivity, and low vapor pressure, ILs have received considerable attention in different applications. However, a molecular insight of their phase behavior and their interactions with other components is still required. For this regard, many thermodynamic models such as NRTL, UNIFAC, Equation of States (EoSs) like Cubic Plus Association (CPA), and ab initio method like COSMO-RS, have been frequently applied on different systems containing IL in the literature (Chen et al., 2020). Since CPA EoS showed an efficient capability for modeling of thermophysical properties of ILs as well as phase equilibria of binary and ternary systems containing IL, it has been investigated more, recently (Maia et al., 2012; Frias et al., 2023; Zhu et al., 2021; Baramaki et al., 2019).

In order to use thermodynamic models, some adjustable parameters are required as well as sigma-profile for COSMO-RS method. It should be noted that finding the CPA pure parameters (i.e., 'b', 'a_0'', 'c_1'', 'ε', and 'β') of either non-associating (i.e., alkanes, alkenes etc.) or associating compounds (i.e., ILs) is still a hard task (Velho et al., 2021). These parameters are optimized against experimental data of thermophysical properties such as vapor pressure or liquid density. If there is a lack of experimental data, it would become challenging task to simulate and control processes containing such compounds using CPA EoS (Velho et al., 2021). For this regard, many studies have already been focused on predicting the CPA pure parameters for non-associating compounds (Hansen, 2015; Palma et al., 2017; Coutinho et al., 2000). These studies have been conducted on small datasets (Velho et al., 2021). For this reason, Velho et al. (Velho et al., 2021) focused on a larger dataset to build some predictive correlations for 'b', 'a_0'', and 'c_1'' parameters of non-associating compounds. In their models, they used molecular variables such as van der Waals volume (V_w) and Molecular weight (Mw) along with critical properties (T_C and P_C). Regarding the associating compounds like ILs, Maia et al. (Maia et al., 2012) attempted to predict 'b' and 'a_0'' of ILs using V_w of the ionic liquids for further facilitation. They found that V_w can predict the 'b' parameter with acceptable deviation, however V_w had not enough accuracy for prediction of a_0 parameter. In most recent efforts, Frias et al. (Frias et al., 2023) suggested that two molecular descriptors which called V_w and 'Kier mass connectivity

* Corresponding author.
E-mail address: ville.alopaeus@aalto.fi (V. Alopaeus).

https://doi.org/10.1016/j.ces.2024.120825
Received 8 May 2024; Received in revised form 26 August 2024; Accepted 10 October 2024
Available online 15 October 2024

Nomenclature

Abbreviation Name of ionic liquid
[emim][PF6] 1-ethyl-3-methylimidazolium hexafluorophosphate
[emim] [TfO] 1-ethyl-3-methylimidazolium trifluoromethanesulfonate
[hmim] [TfO] 1-hexyl-3-methylimidazolium trifluoromethanesulfonate
[Hemim] [Tf2N] 1-(2-hydroxyethyl)-3-methylimidazolium bis (trifluoromethylsulfonyl)imide
[emim] [EtSO4] 1-ethyl-3-methylimidazolium ethyl sulfate
[emim] [SCN] 1-ethyl-3-methylimidazolium thiocyanate
[emim] [eFAP] 1-ethyl-3-methylimidazolium tris(pentafluoroethyl) trifluorophosphate
[emim] [Cl] 1-ethyl-3-methylimidazolium chloride
[Bmim] [NO3] 1-butyl-3-methylimidazolium nitrate
[NH2emim] [BF4] 1-aminoethyl-3-methylimidazolium fluoroborate
[emim] [BF4] 1-ethyl 3-methylimidazolium tetrafluoroborate
[bmim] [BF4] 1-butyl 3-methylimidazolium tetrafluoroborate
[hmim] [BF4] 1-hexyl-3-methylimidazolium tetrafluoroborate
[omim] [PF6] 1-octyl-3-methylimidazolium hexafluorophosphate
[emim] [Ac] 1-ethyl-3-methylimidazolium acetate
[hmim] [Ac] 1-hexyl-3-methylimidazolium acetate
[EA] [NO3] ethylammonium nitrate
[HEA] [L] 2-hydroxyethylammonium lactate
[BHEMA] [Ac] 2-Hydroxy-N-(2-hydroxyethyl)-N-methylethanaminium acetate
[BHEA] [L] Bis(2-hydroxyethyl)ammonium lactate
[BHEMA] [L] 2-Hydroxy-N-(2-hydroxyethyl)-N-methylethanaminium lactate
[THEMA] [MeSO4] tris(2-hydroxyethyl) methylammonium methylsulfate
[N1223] [FSI] N,N-diethyl-N-methyl-N-propylammonium bis (fluorosulfonyl)imide
[TBMA] [MeSO4] tributylmethylammonium methylsulfate
[N1125] [Tf2N] N,N-dimethyl-N-ethyl-N-pentylammonium bis (trifluoromethanesulfonyl)amide
[N112,3–C3OC] [Tf2N] N,N-dimethyl-N-ethyl-N-(3 methoxypropyl)ammonium bis(trifluoromethylsulfonyl) imide
[N122,2–2OC] [Tf2N] N,N-diethyl-N-methyl-N(2-methoxyethyl) ammonium bis(trifluoromethanesulfonyl)imide
[N112,2OCO1] [Tf2N] N-Acetoxyethyl-N,N-dimethyl-N-ethylammonium bis(trifluoromethanesulfonyl)amide
[N1127] [Tf2N] N,N-dimethyl-N-ethyl-N-heptylammonium bis (trifluoromethanesulfonyl)amide
[N112,2O2O1] [Tf2N] N,N-dimethyl-N-ethyl-N-methoxyethoxyethylammonium bis (trifluoromethanesulfonyl)amide
[Bpy] [SCN] 1-butylpyridinium thiocyanate
[emim] [NTF2] 1-ethyl-3-methylimidazolium bis (trifluoromethylsulfonyl)imide
[bmim] [NTF2] 1-butyl-3-methylimidazolium bis (trifluoromethylsulfonyl)imide
[hmim] [NTF2] 1-hexyl-3-methylimidazolium bis (trifluoromethylsulfonyl)imide
[omim] [NTF2] 1-octyl-3-methylimidazolium bis (trifluoromethylsulfonyl)imide
[Bzmim] [NTF2] 1-benzyl-3-methylimidazolium bis (trifluoromethylsulfonyl)imide
[Epy] [NTF2] 1-ethylpyridinium bis(trifluoromethylsulfonyl)imide
[Pmpip] [NTF2] 1-methyl-1-propylpiperidinium bis (trifluoromethylsulfonyl)imide
[Bmpyr] [NTF2] 1-butyl-1-methylpyrrolidinium bis (trifluoromethanesulfonyl)imide
[N1114] [NTF2] butyltrimethylammonium bis (trifluoromethylsulfonyl)imide
[omim] [BF4] 1-octyl-3-methylimidazolium tetrafluoroborate
[bmim] [PF6] 1-butyl-3-methylimidazolium hexafluorophosphate
[hmim] [PF6] 1-hexyl-3-methylimidazolium hexafluorophosphate
[C2OCpy] [Tf2N] 1–(2–methoxyethyl)pyridinium bis (trifluoromethylsulfonyl)amide
[hmpy] [Tf2N] 1-hexyl-3-methylpyridinium bis (trifluoromethylsulfonyl)amide
[C7py] [Tf2N] 1–heptylpyridinium bis(trifluoromethylsulfonyl) amide
[C5O2py] [Tf2N] 1–(2–(2–methoxyethoxy)ethyl)pyridinium bis (trifluoromethylsulfonyl)amide
[C10py] [Tf2N] 1–decylpyridinium bis(trifluoromethylsulfonyl) imide
[C7O3py] [Tf2N] 1–(2–(2–(2–methoxyethoxy)ethoxy)ethyl) pyridinium bis(trifluoromethylsulfonyl)amide
[HEpy] [Tf2N] 1–(2–hydroxyethyl)–pyridinium bis (trifluoromethylsulfonyl)imide
[3-MPpy] [Tf2N] 3–methyl–1–propylpyridinium bis (trifluoromethylsulfonyl)imide
[BPy] [Tf2N] 1–butylpyridinium bis[(trifluoromethyl)sulfonyl] imide
[BPy] [BF4] 1–butylpyridinium tetrafluoroborate
[b4mpy] [BF4] 1-butyl-4-methylpyridinium tetrafluoroborate
[Bpy] [NO3] 1-butylpyridinium nitrate

index' (λ) of ammonium and pyridinium based-ILs and they mentioned that these descriptors had a linear correlation with 'b', 'a_0', and 'c_1'. There are some sources in the literature that have reported the values of 'V_w' for different groups of ILs (Frias et al., 2023; Machida et al., 2009).

There are many research groups which have reported the CPA pure parameters for different types of ILs such as imidazolium, ammonium, pyridinium, phosphonium, and sulfonium (Maia et al., 2012; Frias et al., 2023; Zhu et al., 2021; Wang et al., 2023; Sousa et al., 2021; Panah, 2017; Hamedi et al., 2020; Ahmadian et al., 2018; Afsharpour, 2022; Afsharpour, 2021; Afsharpour, 2019; Sousa et al., 2014; Sousa et al., 2014; Manic et al., 2012). For this regard, a comprehensive and useful information has been tabulated in Table S1. Types, names of ILs and used experimental properties in the Objective Function (i.e., OF) have been listed in Table S1 along with literature references. The exact reported values of CPA pure parameters based on the two-site Scheme 2B for each IL can be found in Table S1.

As can be seen from Table S1, it was clear that different parametrization processes (i.e., 'OFs') affected the obtained values of CPA pure parameters for a constant IL, except 'b'. Regardless of used 'OF', different research groups reported similar values for 'b' parameter for the same IL. Unlike b, the values of other parameters (i.e., 'a_0', 'c_1', 'ε', and 'β') were strongly dependent on the used 'OF' (see Sheet 1 in the Supporting Information-Excel file). For example, three different research groups (i.e., (Zhu et al., 2021); (Panah, 2017), and (Hamedi et al., 2020) used three different 'OFs' (or parameterization process) to obtain CPA pure parameters of [hmim][NTf2] (see Table S1). Among the five pure parameters, they reported almost same value (approximately 0.3) for 'b' parameter, while for other parameters, the reported values had significant differences. Also, Sousa et al. (Sousa et al., 2014) and Hamedi et al. (Hamedi et al., 2020) reported similar values of 'b' parameter (approximately 0.128) for [m2HEA] [Pr], however, they used two different 'OFs'.

Table 1
The regressed values of CPA parameters in the literature and predicted values using QSPR models as well as values of ARD%.

Datasets	ILs[a]	b-Reg (L/mol)	b-Pre (L/mol) by Eq. (34)	ARD %	a_0-Reg (Pa m^6/mol^2)	a_0-Pre (Pa m^6/mol^2) by Eq. (35)	ARD %	c_1-Reg	c_1-Pre by Eq. (36)	ARD%	ε-Reg (J/mol)	ε-Pre (J/mol) by Eq. (37)	ARD %	β-Reg	β-Pre by Eq. (38)	ARD%
First	[emim] [TfO]	0.1767	0.1753	0.79	5.3922	3.8744	28.14	1.0283	1.3524	31.51	13,307	12641.2	5.00	0.0111	0.0063	43.24
	[hmim] [TfO]	0.2398	0.2393	0.20	3.5007	3.4595	1.17	2.5394	2.0876	17.79	10,486	11842.1	12.93	0.0093	0.0063	32.25
	[Hemim] [Tf2N]	0.2381	0.2413	1.34	1.123	1.3721	22.18	4.873	5.136	5.39	10,030	9321.6	7.06	0.0183	0.0241	31.69
	[emim] [EtSO4]	0.1767	0.1736	1.75	3.8128	3.6339	4.69	1.5207	1.3838	9.00	7643	8798.0	15.11	0.0066	0.0043	34.84
	[emim] [SCN]	0.1387	0.1293	6.77	2.2605	2.4412	7.99	1.7409	2.0385	17.09	14,476	14148.8	2.25	0.0457	0.0445	2.62
	[Bmim] [NO3]	0.1536	0.1668	8.59	2.953	4.1817	41.60	0.8385	1.2661	50.99	24,849	23162.2	6.78	0.0495	0.0511	3.23
	[NH2emim] [BF4]	0.1334	0.1389	4.12	2.9914	2.6538	11.28	0.9977	0.9962	0.15	20,691	20502.9	0.90	0.0492	0.0463	5.89
	[emim] [BF4]	0.1434	0.1389	3.13	6.147	6.8686	11.73	0.9579	0.6507	32.07	13,866	16278.4	17.39	0.0037	0.0029	21.62
	[bmim] [BF4]	**0.1751**	**0.1709**	**2.39**	**6.9942**	**7.6565**	**9.46**	**0.9955**	**0.511**	**48.66**	**16,664**	**20405.9**	**22.45**	**0.0013**	**0.0029**	**123.07**
	[hmim] [BF4]	0.2064	0.2029	1.69	6.9422	6.4537	7.03	1.2455	1.3859	11.27	17,784	15479.3	12.95	0.0008	0.0029	**262.5**
	[omim] [BF4]	0.238	0.2349	1.30	8.5927	7.4377	13.44	1.0655	1.4537	36.43	22,206	23176.1	4.36	0.0002	0.0029	***1350***
	[bmim] [PF6]	0.1965	0.198	0.76	7.0947	6.7513	4.84	1.1827	1.0375	12.27	14,230	11671.9	17.97	0.0035	0.0029	17.14
	[hmim] [PF6]	**0.2277**	**0.23**	**1.01**	**6.0745**	**5.5485**	**8.65**	**1.6775**	**1.9124**	**14.00**	**13,505**	**6745.3**	**50.05**	**0.0028**	**0.0029**	**3.57**
	[omim] [PF6]	0.26	0.262	0.76	5.5264	6.5325	18.20	2.1402	1.9801	7.48	14,489	14442.1	0.32	0.0018	0.0029	61.11
	[emim] [Ac]	**0.1424**	**0.1442**	**1.26**	**2.569**	**2.597**	**1.08**	**2.0961**	**1.5052**	**28.19**	**8625**	**12537.7**	**45.36**	**0.0163**	**0.0073**	**55.21**
	[hmim] [Ac]	0.2051	0.2082	1.51	2.5985	2.1821	16.02	2.9024	2.2404	22.80	11,095	11738.6	5.80	0.0061	0.0073	19.67
	[emim] [NTF2]	0.237	0.2413	1.81	5.6871	5.8579	3.00	0.9918	1.2801	29.06	11,861	12052.5	1.61	0.0034	0.0024	29.41
	[bmim] [NTF2]	0.2716	0.2733	0.62	7.1339	6.6457	6.84	0.9315	1.1404	22.42	13,822	16180.0	17.06	0.003	0.0024	20
	[hmim] [NTF2]	0.3075	0.3053	0.71	4.1393	5.443	31.49	2.4426	2.0153	17.49	11,596	11253.4	2.95	0.0011	0.0024	118.18
	[omim] [NTF2]	**0.3351**	**0.3373**	**0.65**	**3.6365**	**6.4269**	**76.73**	**2.6459**	**2.083**	**21.27**	**15,509**	**18950.2**	**22.18**	**0.0005**	**0.0024**	**380**
	[Bzmim] [NTF2]	0.2802	0.2733 By Eq. (39)	2.46	2.1293	2.2357 By Eq. (40)	4.99	3.4568	3.4105 By Eq. (41)	1.33	13,232	12973.0 By Eq. (42)	1.95	0.0011	0.0024 By Eq. (43)	**140**
Second	[EA] [NO3]	0.0826	0.0796	3.63	4.01551	4.6088	14.77	0.405	0.3003	25.85	42395.98	42295.2	0.23	0.00257	0.0005	80.76
	[HEA] [L]	0.118	0.1202	1.86	3.33558	3.1419	5.80	1.2	1.2495	4.125	31689.48	32933.4	3.92	0.045	0.0397	11.77
	[BHEMA] [Ac]	0.1488	0.1502	0.94	6.16294	5.6708	7.98	1.06	1.239	16.88	17,738	18424.5	3.87	0.06893	0.0656	4.78
	[BHEA] [L]	0.1521	0.1489	2.10	5.56244	5.4141	2.66	0.8892	0.9444	6.20	18459.22	19883.7	7.71	0.07171	0.0718	0.13
	[BHEMA] [L]	**0.1687**	**0.1677**	**0.59**	**6.45329**	**5.9818**	**7.30**	**1.0467**	**0.7189**	**31.31**	**18391.52**	**17536.5**	**4.64**	**0.07385**	**0.0776**	**5.00**
	[THEMA] [MeSO4]	0.1921	0.1952	1.61	6.03896	5.9319	1.77	1.1947	1.1856	0.76	16323.71	17335.5	6.19	0.08558	0.0815	4.78
	[N1223] [FSI]	0.2215	0.2256	1.85	7.14366	7.2965	2.13	1.2199	1.3191	8.13	16421.81	17574.6	7.02	0.07623	0.0714	6.29
	[TBMA] [MeSO4]	0.2848	0.2788	2.10	10.21653	10.3092	0.90	1.31	1.2575	4.00	14849.8	14603.0	1.66	0.08455	0.0882	4.25
	[N1125] [Tf2N]	0.3041	0.3056	0.49	10.45557	10.7937	3.23	1.12	1.0545	5.84	15100.55	15572.6	3.12	0.08693	0.0847	2.53

(*continued on next page*)

Table 1 (*continued*)

Datasets	ILs[a]	b-Reg (L/mol)	b-Pre (L/mol) by Eq. (34)	ARD %	a_0-Reg (Pa m^6/mol^2)	a_0-Pre (Pa m^6/mol^2) by Eq. (35)	ARD %	c_1-Reg	c_1-Pre by Eq. (36)	ARD%	ε-Reg (J/mol)	ε-Pre (J/mol) by Eq. (37)	ARD %	β-Reg	β-Pre by Eq. (38)	ARD%
	[N112,3–C3OC] [Tf2N]	**0.3021**	**0.2923**	**3.24**	**10.07925**	**8.8978**	**11.72**	**1.3103**	**1.3067**	**0.27**	**15208.72**	**15572.6**	**2.39**	**0.08688**	**0.0847**	**2.53**
	[N122,2–2OC] [Tf2N]	0.2919	0.2833	2.94	9.73665	8.8978	8.61	1.3464	1.3994	3.93	15294.9	15572.6	1.81	0.08227	0.0847	2.91
	[N112,2OCO1] [Tf2N]	0.283	0.2858	0.98	10.35197	9.3629	9.55	0.5236	0.542	3.51	15162.6	14328.9	5.49	0.08608	0.0878	1.97
	[N1127] [Tf2N]	0.3337	0.3448	3.32	11.48358	11.5937	0.95	1.1015	1.1141	1.14	14610.6	13260.4	9.24	0.0892	0.0904	1.34
	[N112,2O2O1] [Tf2N]	0.33	0.3222	2.36	6.97702	8.3584	19.79	1.3926	1.328	4.63	14852.1	13260.4	10.71	0.08789	0.0904	2.84
	[HEA] [Ac]	0.1003	0.1027	2.39	2.73	2.8309	3.69	1.94	1.7696	8.78	35,968	33821.3	5.96	0.01743	0.0278	59.77
	[BHEA] [Ac]	**0.1359**	**0.1314** By Eq. (44)	**3.31**	**4.09**	**5.1031** By Eq. (45)	**24.77**	**1.73**	**1.4645** By Eq. (46)	**15.34**	**14549.5**	**20771.7** By Eq. (47)	**42.76**	**0.0649**	**0.0598** By Eq. (48)	**7.85**
Third	[Bpy] [SCN]	0.16009	0.1585	1.00	4.47732	4.2868	4.25	1.122	1.0802	3.73	8741.08	8801.33	0.69	0.01527	0.0155	1.31
	[Bpy] [NO3]	0.15269	0.1555	1.83	4.2108	4.2345	0.56	0.9955	1.0802	8.51	8848.55	8801.33	0.53	0.01253	0.0152	21.60
	[BPy] [BF4]	0.178	0.1758	1.24	5.47558	5.8154	6.21	1.087	1.0802	0.63	14840.49	14332.9	3.42	0.028	0.0261	6.79
	[b4mpy] [BF4]	0.191	0.1916	0.31	6.3519	6.2786	1.15	1.803	1.8123	0.52	14017.4	14440.8	3.02	0.03435	0.0316	8.14
	[HEpy] [Tf2N]	0.24499	0.2458	0.33	7.46504	7.4336	0.42	1.25	1.1682	6.54	15820.05	15862.0	0.27	0.0395	0.0399	1.01
	[3-MPpy] [Tf2N]	**0.27012**	**0.2694**	**0.26**	**8.59229**	**8.3601**	**2.70**	**1.1898**	**1.2682**	**6.59**	**15646.94**	**15349.5**	**1.90**	**0.04588**	**0.0456**	**0.65**
	[BPy] [Tf2N]	0.27195	0.2694	0.96	8.58883	8.3601	2.66	0.8337	1.0802	29.57	15297.89	15349.5	0.34	0.04358	0.0436	0.00
	[C2OCpy] [Tf2N]	0.25945	0.2616	0.81	8.24731	8.3028	0.67	1.3	1.0965	15.65	15751.21	15924.5	1.10	0.04284	0.043	0.47
	[hmpy] [Tf2N]	0.31258	0.317	1.41	9.7322	9.7498	0.18	0.6116	0.6278	2.65	15697.96	15614.4	0.53	0.05446	0.0552	1.28
	[C7py] [Tf2N]	**0.32589**	**0.317**	**2.73**	**10.45819**	**9.7498**	**6.77**	**1.2315**	**0.9482**	**23.00**	**15684.18**	**15614.4**	**0.44**	**0.05212**	**0.0527**	**1.15**
	[C5O2py] [Tf2N]	0.30452	0.3014	1.02	9.46874	9.6352	1.76	0.9194	0.8626	6.18	16464.41	16523.4	0.36	0.0524	0.0526	0.38
	[C10py] [Tf2N]	0.36374	0.3646	0.25	11.27197	11.1396	1.17	1.0089	1.043	3.38	15810.43	15765.8	0.28	0.06218	0.0603	3.05
	[C7O3py] [Tf2N]	0.34343	0.3412	0.64	10.91432	10.9677	0.49	1.2714	1.2713	0.01	17016.46	16889.4	0.75	0.05869	0.0608	3.58

[a] Bold ILs mean test sets.

Table 2
Applied statistical parameters in this study.

Introduced parameters	Introduced parameters equations	Eqs. No
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i^{exp} - Y_i^{pre}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{exp} - \overline{Y}_i\right)^2}$	(8)
Adjustable coefficient of determination	$R^2_{Adj} = 1 - (1\text{-}R^2) \times (\frac{n-1}{n-p-1})$	(9)
Leave-one-out cross-validated coefficient of determination	$Q^2_{LOO\text{-}CV} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i^{exp} - Y_i^{pre-CV}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{exp} - \overline{Y}_i\right)^2}$	(10)
Fisher function	$F = \frac{\sum_{i=1}^{n}\left(Y_i^{pre} - \overline{Y}_i\right)^2/p}{\sum_{i=1}^{n}\left(Y_i^{exp} - Y_i^{pre}\right)^2/n-p-1}$	(11)
Standard residual	$S = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i^{exp} - Y_i^{pre}\right)^2}{n-p-1}}$	(12)
root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i^{exp} - Y_i^{pre}\right)^2}{n}}$	(13)
Average absolute deviation	$AAD = \frac{\sum_{i=1}^{n}(\lvert Y_i^{exp} - Y_i^{pre}\rvert)}{n}$	(14)
Average Absolute relative deviation %	$AARD\% = \frac{\sum_{i=1}^{n}(\lvert Y_i^{exp} - Y_i^{pre}\rvert)/Y_i^{exp}}{n} \times 100$	(15)
maximum Leverage	$h^* = 3(p+1)/n$	(16)

Y_i^{exp}, Y_i^{pre},$\overline{Y}_i$, n, and p demonstrate reported (or regressed) values, predicted values, average reported (or regressed) values, the number of the reported (or regressed) data set, and the number of employed descriptors, respectively.

In one hand, it may exist two variant sets of 5 parameters which were obtained with same used OF, could fit the vapor pressure and density of a constant IL with same Absolute Average Relative Deviations (AARD%) (Maia et al., 2012). This point is emphasizing that there is not necessarily a unique set of CPA pure parameters which can fit the thermophysical properties for a specific IL. On the other hand, it seems that there are some conflicts about the reported values. 'Order of Magnitude' point of view, the reported values of 'c_1' and 'ε' by Hamedi et al. (Hamedi et al., 2020) (see Table S1) had a significant difference with other related values which were reported by other groups. Another conflict which was observed in Table S1 and needs to be discussed more, is that 'why did Manic et al. (Manic et al., 2012) report the value of 'b' parameter of [bmim][NTf2], 0.322'. This value had noticeable deviation from other reported values by other researchers (Zhu et al., 2021; Hamedi et al., 2020). For these reasons, it is essential to figure out that what is the relationship between the CPA pure parameters of ILs which were obtained from a constant 'OF', and structures of cation and anion. Molecular insights might remove such inconsistency and help gaining a better understanding.

According to the literature survey, it was found that (V_w) of ILs (i.e., V_w-IL = V_w_Cation + V_w_Anion) as a molecular variable (i.e., descriptor) can predict 'b', 'a_0', and 'c_1' parameters of ILs, regardless of kind of ILs (Maia et al., 2012; Frias et al., 2023). Also, it was found that there were no models to predict the associating parameters such as 'ε' and 'β'. The vast number of potential ILs that can be synthesized makes it impossible to find the optimal IL for a particular task only using experiments. Therefore, the development of time-saving and cost-effective predictive models can be very useful.

The 'Quantitative Structure-Property Relationship' (QSPR) method as one of the robust Machine Learning (ML) methods is frequently applied to make qualitative and quantitative relations between IL: s and their specific properties (Gorji et al., 2021; Gorji and Sobati, 2019; Paduszynski, 2021; Billard et al., 2011). QSPR often isapplied on the experimental data, but there are some studies that proved its applicability on the regressed data (Brijmohan et al., 2024; Stavrou et al., 2016). Therefore, it is possible to design and/or develop new ILs using the QSPR method for different applications. The main aim of this study is to develop predictive 2D-QSPR models for prediction of CPA pure parameters of a vast number of ILs using QSARINS software (Gramatica, 2020; Gramatica and Sangion, 2016; Gramatica et al., 2013) which include different methods of internal and external validations as well as calculation of molecular descriptors. Another aim is to validate the available predictive models in the literature and to compare with the developed QSPR models in this study. The availability of QSPR models leads to the capability of predicting CPA pure component parameters for those ILs which have no experimental density and vapor pressure data in the literature.

Table 3
Developed predictive models using former descriptor (i.e., VABC-IL) by Frias et al. (Frias et al., 2023) for each dataset as well as statistical parameters.

Datasets	Models	Eq. No	Sets	Number of datapoints	R^2	AARD%
First	$b = 0.001VABC_{Cat} + 0.001VABC_{Ani} - 0.03$	(25)	Training	17	0.96	6.74
			Test	4	0.99	4.77
	$a0 = 0.004VABC_{Cat} + 0.004VABC_{Ani} + 3.56$	(26)	Training	17	0.01	59.03
			Test	4	0.006	41.77
	$c_1 = 0.0094VABC_{Cat} + 0.0094VABC_{Ani} - 0.571$	(27)	Training	17	0.21	46.32
			Test	4	0.35	31.39
Second	$b = 0.0009VABC_{Cat} + 0.0009VABC_{Ani} - 0.0195$	(28)	Training	13	0.99	5.45
			Test	3	0.99	4.69
	$a0 = 0.027VABC_{Cat} + 0.027VABC_{Ani} + 0.1744$	(29)	Training	13	0.83	15.81
			Test	3	0.98	10.27
	$c_1 = 0.0002VABC_{Cat} + 0.0002VABC_{Ani} + 1.0857$	(30)	Training	13	0.002	33.75
			Test	3	0.14	18.37
Third	$b = 0.001VABC_{Cat} + 0.001VABC_{Ani} - 0.0386$	(31)	Training	11	0.99	5.05
			Test	2	–[a]	2.88
	$a0 = 0.0316VABC_{Cat} + 0.0316VABC_{Ani} - 1.772$	(32)	Training	11	0.99	3.29
			Test	2	–[a]	4.09
	$c_1 = -0.0012VABC_{Cat} - 0.0012VABC_{Ani} + 1.4825$	(33)	Training	11	0.09	22.28
			Test	2	–[a]	11.66

[a] has not been calculated, due to limitation of datapoints.

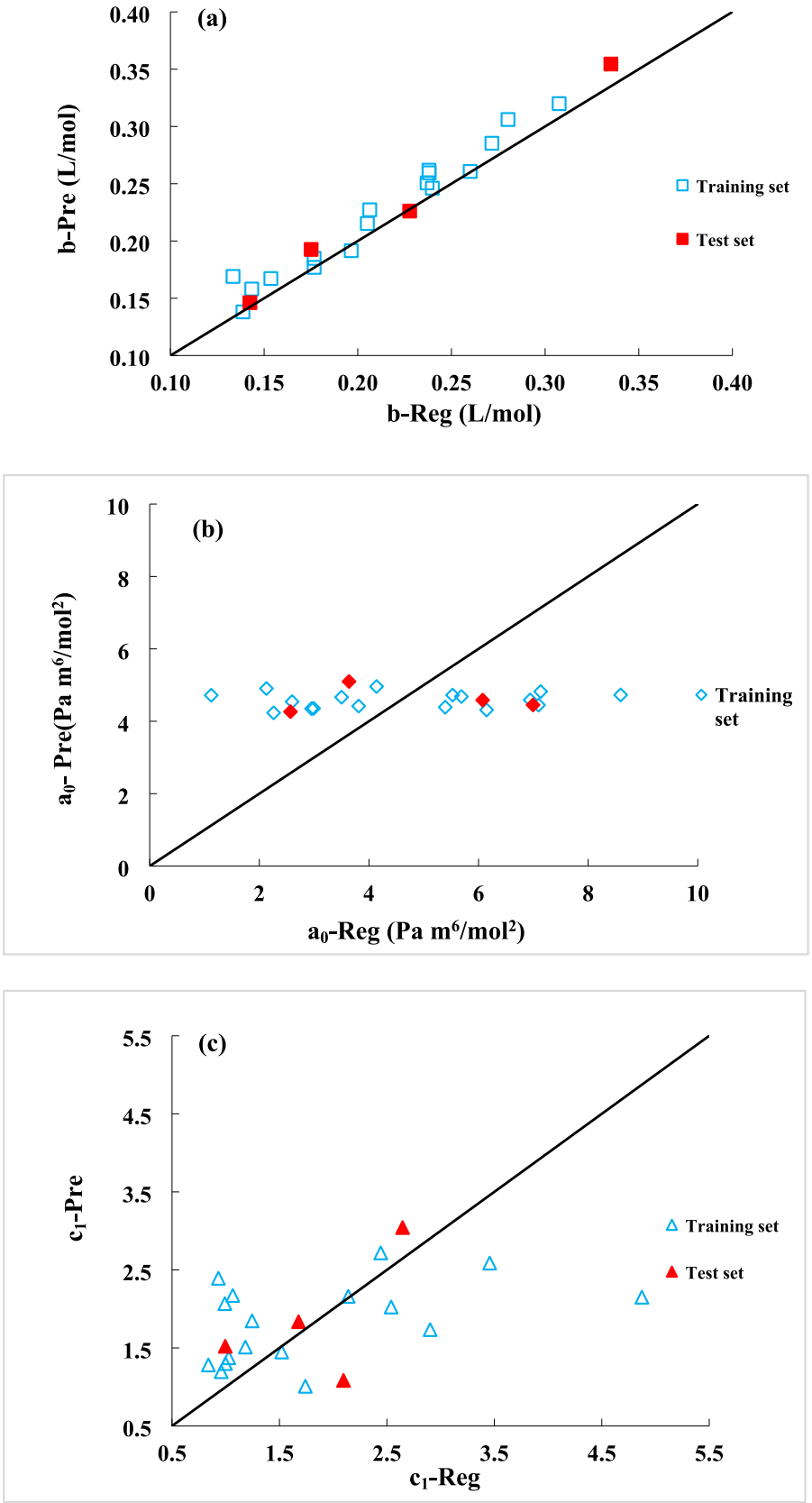


Fig. 1. The predicted versus regressed values of each parameter b (a), a_0 (b) and c_1 (c) for first dataset using VABC-IL descriptor.

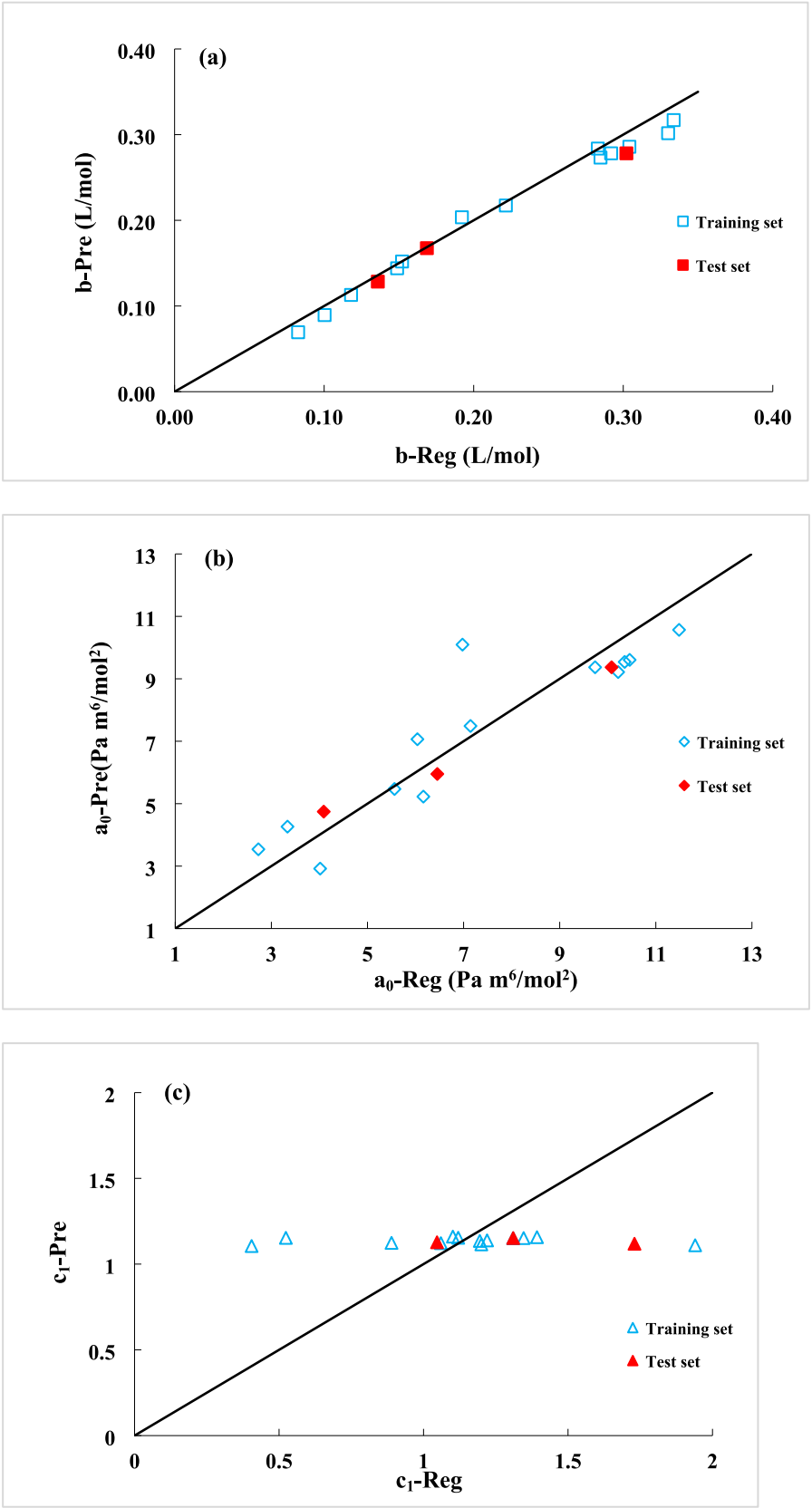


Fig. 2. The predicted versus regressed values of each parameter b (a), a_0 (b) and c_1 (c) for second dataset using VABC-IL descriptor.

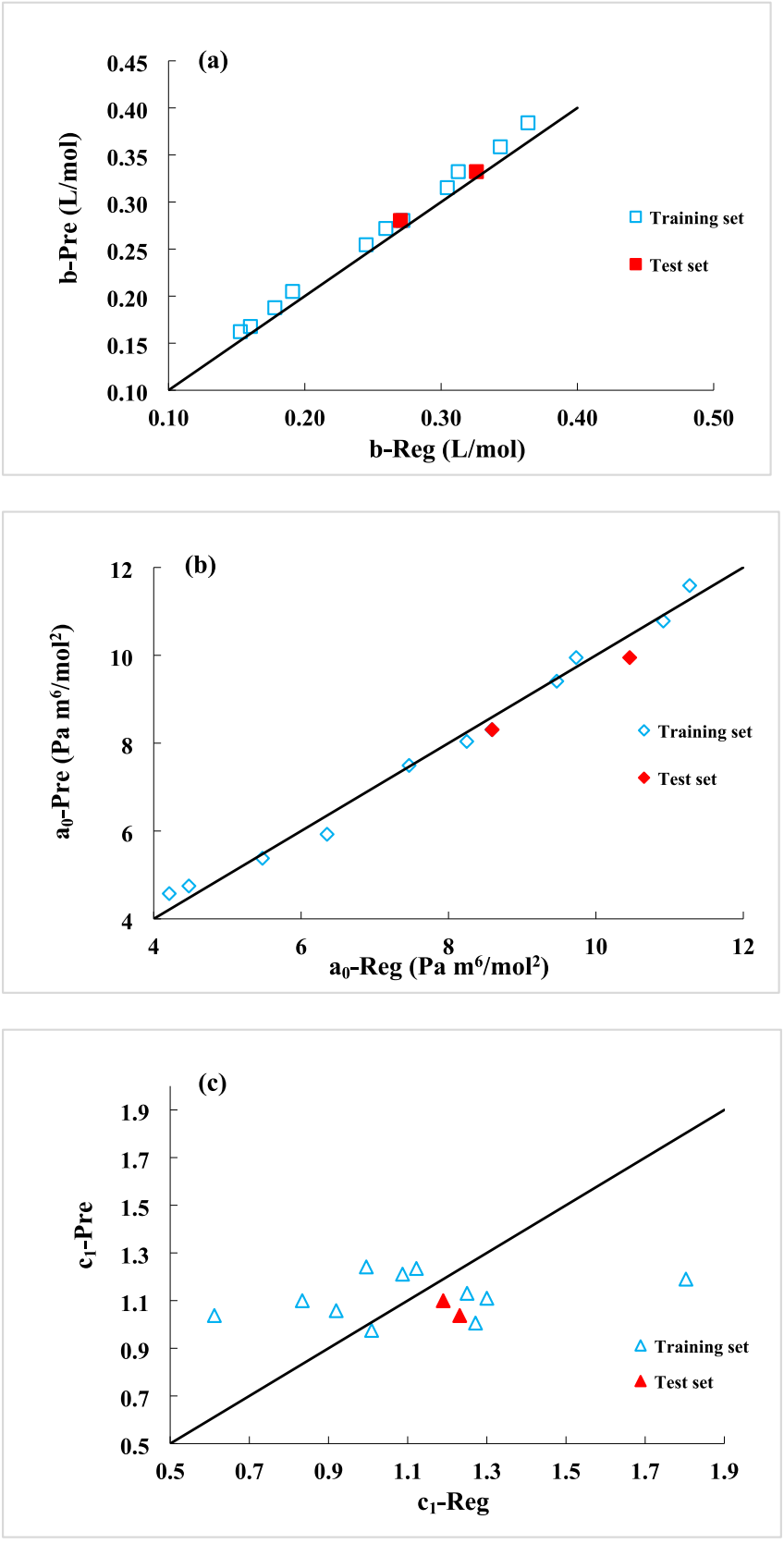


Fig. 3. The predicted versus regressed values of each parameter b (a), a_0 (b) and c_1 (c) for third dataset using VABC-IL descriptor.

Table 4
The values of statistical parameters of selected qspr models for the first dataset for both of training and test sets.

Selected QSPR Models	Eq. No	Sets	Number of datapoints	R2	R2-Adj	Q2-LOO	F	S	RMSE	AARD %	AAD
$b = 0.031994\ topoRadius_{Cat} + 0.000017\ ATS2v_{Ani} + 0.024853$	(34)	Training	17	0.99	0.98	0.98	693	0.0057	0.0051	2.25	0.0040
		Test	4	0.99	–	–	–	–	0.0028	1.33	0.0026
$a_0 = -155.5283\ ATSC8c_{Cat} + 0.0182\ ATSC3m_{Cat} + 0.1722\ ATSC2i_{Ani} + 9.2232$	(35)	Training	17	0.87	0.84	0.81	29	0.8517	0.7448	13.8	0.584
		Test	4	0.58	–	–	–	–	1.4580	23.98	1.000
$c_1 = 319.6869\ AATSC5e_{Cat} - 5.9341\ GATS1s_{Cat} - 0.2241\ AATSC2i_{Ani} + 8.8458$	(36)	Training	17	0.91	0.89	0.83	45	0.3648	0.3190	19.09	0.275
		Test	4	0.71	–	–	–	–	0.4888	28.03	0.468
$\varepsilon = -31827.2593\ AATS3e_{Cat} + 38.8214\ ATSC8m_{Cat} + 1378.1520\ AATS1s_{Ani} + 238869.19$	(37)	Training	17	0.90	0.88	0.83	40	1566.9	1370.2	7.79	1069
		Test	4	0.30	–	–	–	–	4659.5	35.01	4463
$\beta = 0.021704\ nHBDon\_Lipinski_{Cat} + 0.041629\ nHBd_{Ani} - 0.004138\ mindO_{Ani} + 0.002897$	(38)	Training	17	0.97	0.97	0.91	203	0.0028	0.0025	129	0.0020
		Test	4	0.99	–	–	–	–	0.0047	140	0.0031

2. Method

2.1. CPA EoS:

The proposed CPA EoS by Kontogeorgis (Kontogeorgis et al., 1996) in 1996, is consisted of two major parts: 1) physical and 2) association. In fact, it combined SRK EoS with association term from the Wertheim theory (see Eqs. ((1)-(2)). To model the phase behavior of chemical compounds, five pure parameters (i.e., b, a_0, c_1, ε, and β) are required for associating compounds (i.e., ILs) (Velho et al., 2021).

$$Z = Z^{phys} + Z^{assoc} \tag{1}$$

$$= \frac{1}{1-b\rho} - \frac{a\rho}{RT(1+b\rho)} - \frac{1}{2}\left(1+\rho\frac{\partial lng}{\partial\rho}\right)\sum_i x_i \sum_{A_i}(1-X_{A_i})$$

$$a(T) = a0\,[1 + c1\,(1-\sqrt{T_r})]2 \quad \text{energy term for SRK} \tag{2}$$

$$a0 = \Gamma bR \tag{3}$$

$$(R = 8.314\,(J/mol{\cdot}K) \text{ and } \quad T_r = \frac{T}{T_c}$$

$$X_{A_i} = \frac{1}{1+\rho\sum_j x_j \sum_{B_j} X_{B_j}\Delta^{A_iB_j}} \tag{4}$$

$$\Delta^{A_iB_j} = g(\rho)\,[\exp(\frac{\varepsilon^{A_iB_j}}{RT}) - 1]\,b_{ij}\,\beta^{A_iB_j} \tag{5}$$

$$g(\rho) = \frac{1}{1-1.9\eta} \; Where \eta = \frac{1}{4}b\rho \tag{6}$$

where 'a' (Eq. (1)) is the energy parameter of EoS SRK, 'b' the co–volume parameter, 'ρ' is the molar density, 'g' a simplified hard–-sphere radial distribution function, 'X_{Ai}' the mole fraction of pure component i not bonded at site A, and 'x_i' is the mole fraction of component i. The energy term of EoS SRK (i.e., Eq. (1)), 'a', described by Soave (Soave, 1972), which is related to 'a_0′' (energy parameter, Eqs. (2) and (3)), which in turn depends on the reduced energy parameter 'Γ', the gas constant 'R' and the co-volume parameter 'b'. 'a_0′' and 'c_1′' are often regressed (simultaneously with 'b') from pure component vapor pressure and liquid density data. 'X_{Ai}' (Eq. (4)) is related to the association strength 'Δ^{AiBj}' between sites belonging to two different molecules and is calculated by solving the following set of equations: 'ε^{AiBj}' and 'β^{AiBj}' (Eq. (5)) are the association energy and the association volume, respectively. The simplified radial distribution function, 'g(r)', is given by Eq. (6) (Kontogeorgis et al., 1999):

2.2. Datasets

According to the comprehensive literature survey in the proposed study, three datasets were found with enough variations of cation and anion structures for QSPR studies. The detail of these datasets can be found in Table 1 and Sheets 2–4 (Supporting Information-Excel File). Regarding the datasets, it is essential to mention some points. Firstly, the pure component parameters of regressed data of similar type of ILs (imidazolium, ammonium, or pyridinium) were compiled/collected into a dataset. Care was taken that the reported datasets were regressed using the same OF or parametrization process. Secondly, one dataset was created with variations of cation and anion for each type of IL. Regarding the reported CPA parameters for studied ILs by Panah (2017) (Panah, 2017), it should be added that those ILs were studied recently by Wang et al. (2023) (Wang et al., 2023) and Zhu et al. (2021) (Zhu et al., 2021). Therefore, these CPA parameters have been ignored in this study. All in all, three datasets including above-described features are tabulated in Table 1 for QSPR studies. As a brief description of datasets, it should be noted that first, second, and third datasets included 21 imidazolium, 16 ammonium, and 13 pyridinium-based ILs, respectively.

2.3. QSPR method

2.3.1. Calculation of molecular descriptors

In this study, it was tried to calculate 0D, 1D, and 2D descriptors which were independent of the optimization of molecular structures of cation and anion. Before calculation of such descriptors, each structure of cation and anion was drawn in ChemBioDraw-Ultra (ChemBioDraw, 2010). Then, the drawn structures in 'MDL mol' format fed to PaDEL-Descriptor software (Yap, 2011) for the calculation of descriptors. Descriptors with constant or almost constant values for each cation and anion group were eliminated. Finally, 748 anion descriptors, and 876 cation descriptors were calculated.

2.3.2. Model development

2.3.2.1. Basic theory. Multi-Linear Regression (MLR) is a most common method for QSPR models because of its reproducibility and interpretability (Lemaoui et al., 2020). In this study, we defined a MLR model for each CPA parameter as Eq. (7):

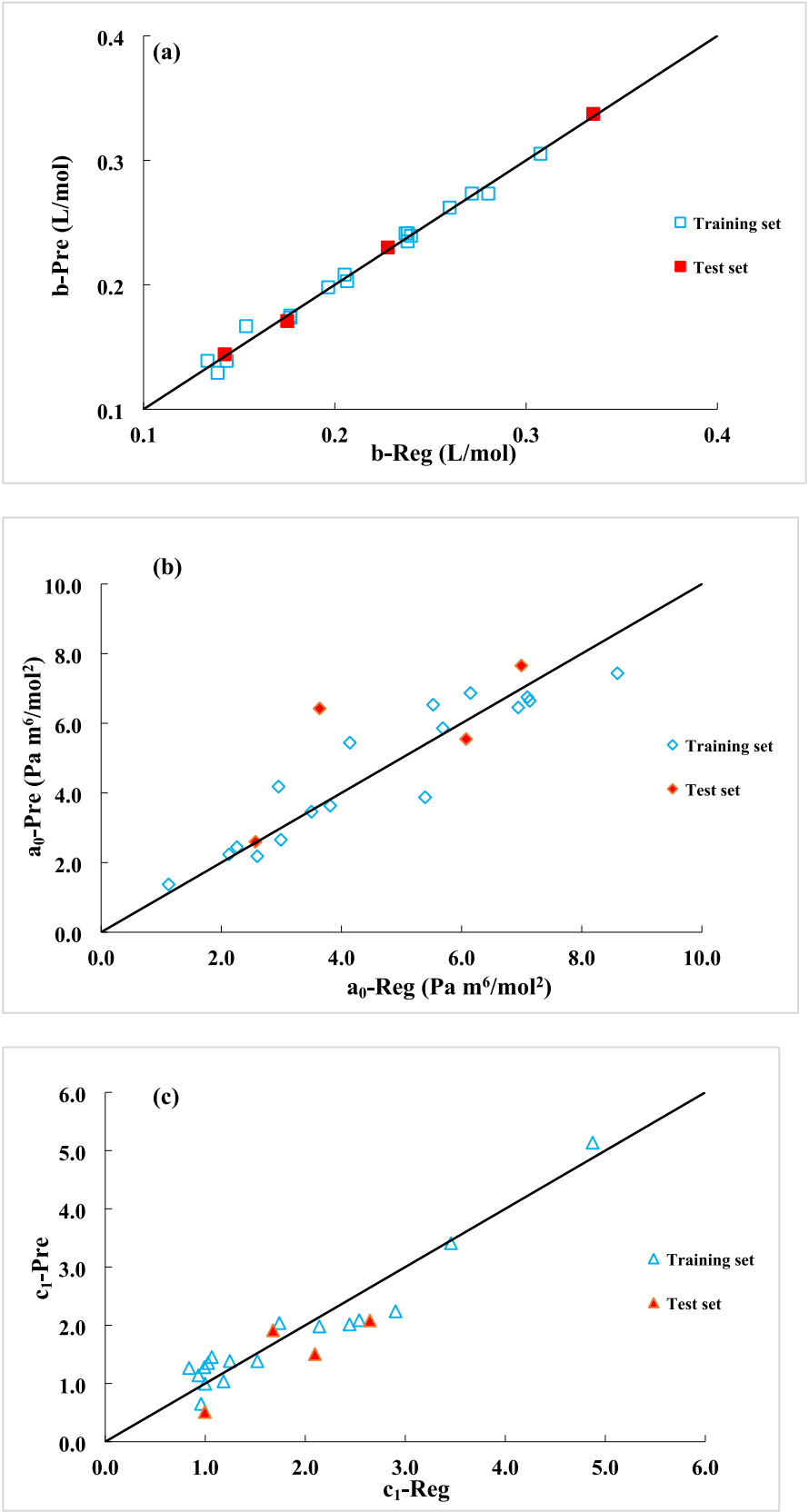
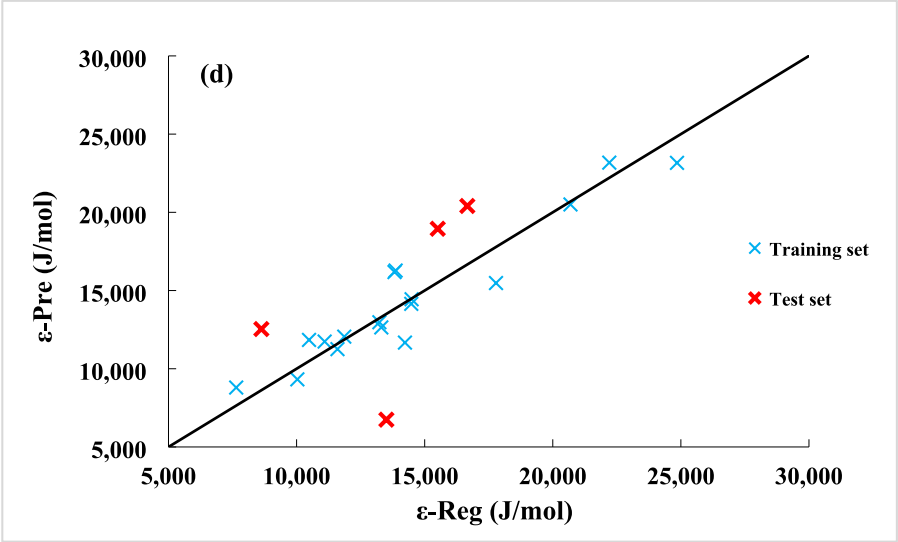


Fig. 4. The predicted versus regressed values of CPA pure parameters (a-e) for the first dataset.



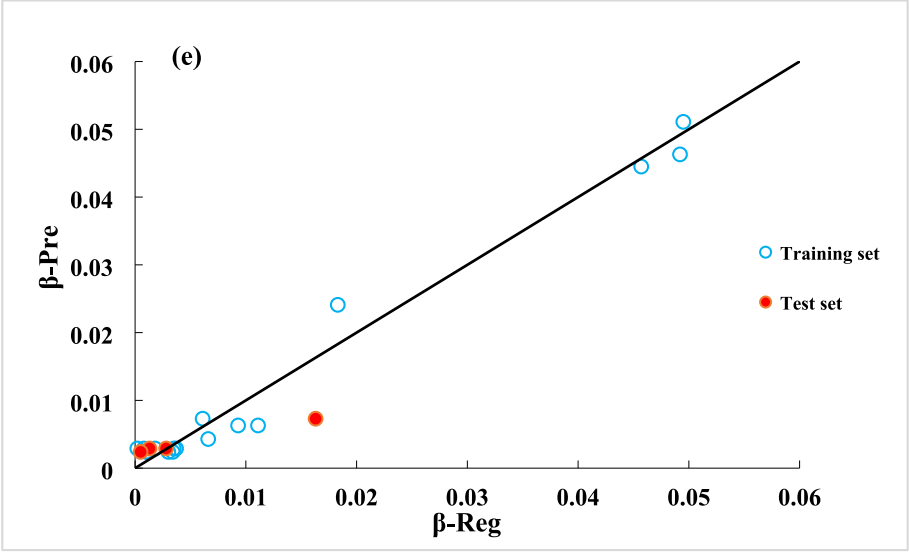


Fig. 4. (*continued*).

$$CPA\ pure\ parameter = F(Descriptors\ of\ cation) + F(Descriptors\ of\ anion) + C \quad (7)$$

In another words, each CPA parameter is simultaneously a function (i.e., F in Eq. (7)) of cation and anion descriptors. Since both cation and anion structures were changing in three datasets, a predictive QSPR model including cation and anion descriptors had to be developed to take into account the effect of cation and anion structures on each CPA parameter. For model construction, the suitable descriptors must be selected using variable selection method which is described upcoming sub-section.

2.3.2.2. Variable (or Descriptor) selection. There are some well-known variable selection methods such as artificial neural network (ANN) (Shahlaei, 2013), replacement method (RM) (Khooshechin et al., 2014), and genetic algorithm (GA) method (Modarresi et al., 2007). GA was applied in this study for the development of the MLR 2D-QSPR models. The details of the GA-MLR algorithm can be found elsewhere (Holland, 1975; Haupt et al., 2004). It should be added that the QSARINS software was utilized to develop the GA-MLR models.

2.3.3. Statistical Parameters

The goodness-of-fit of the developed QSPR models should be evaluated using the standard statistical parameters, including coefficient of determination (R^2), leave-one-out cross-validated coefficient of determination ($Q^2_{LOO\text{-}CV}$), adjustable coefficient of determination (R^2_{Adj}), average absolute relative deviation (%AARD), average absolute deviations (AAD), Fisher function (F), root mean square error (RMSE), standard residual (S), and maximum leverage (H*). More detailed information regarding the statistical parameters can be found in Table 2 (Eqs. (8)–(16)).

Applicability of Domain (AD) analysis as a vital concept of QSPR studies. AD analysis provides (Lemaoui et al., 2020): 1) the extent of extrapolation of developed QSPR models (Ojha and Roy, 2011; He et al., 2017) and 2) the uncertainty in prediction. In order to predict properties for new molecules, it is important that new molecules lie within the same AD space. In another words, it means that new molecules are physicochemically, biologically, or structurally similar with molecules used for model development (i.e., training set). The more space of AD, the more reliable predictions of new molecules. To perform the external validation using test set, it is essential to ensure that the test set of molecules is inside of QSPR model's AD (Amini et al., 2016).

The space of AD is specified using two main parameters: 1) the

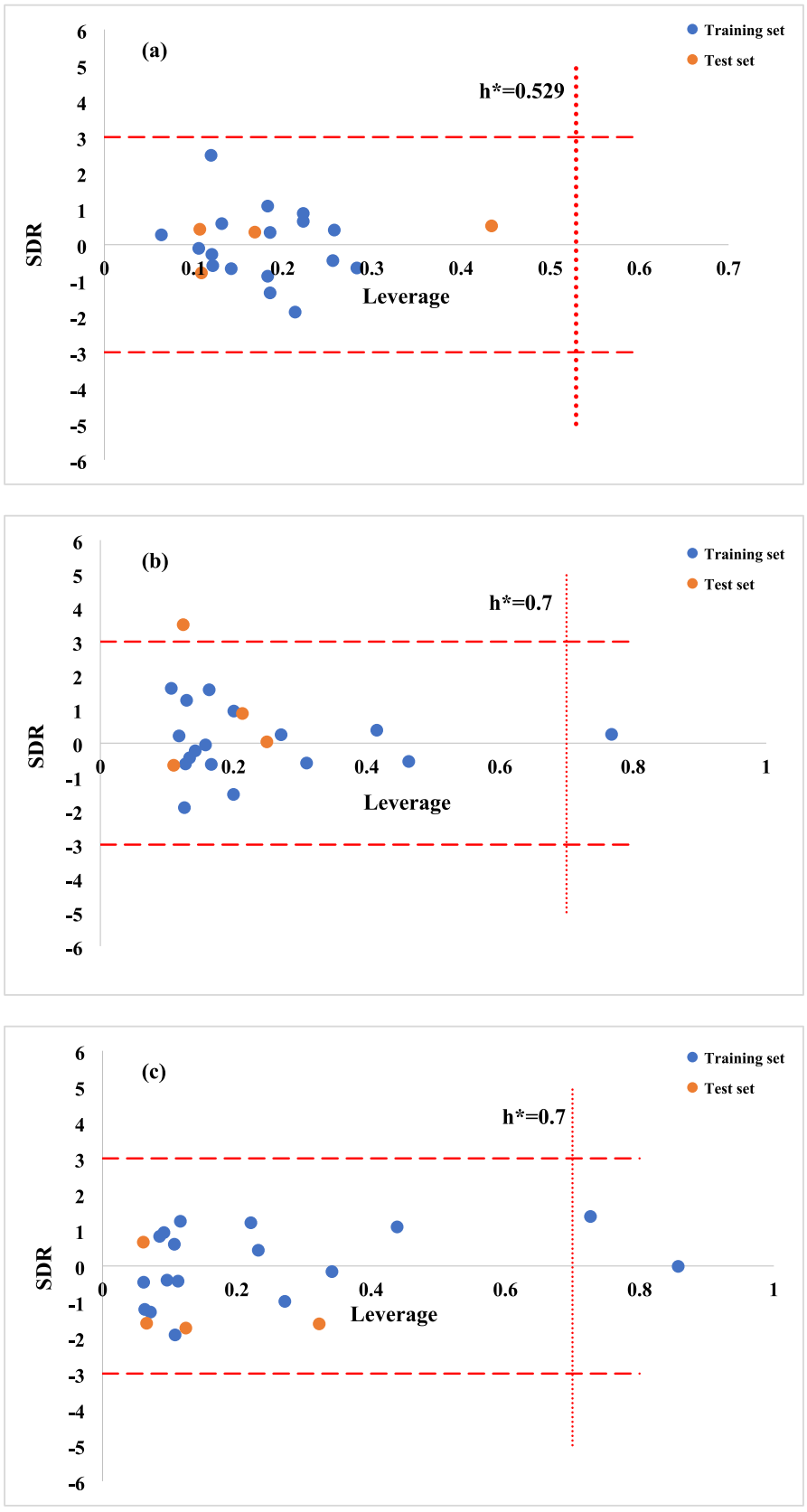


Fig. 5. Williams plots (i.e., SDR versus leverage) of CPA pure parameters b (a), a_0 (b), c_1 (c), ε (d) and β (e) for the first dataset.

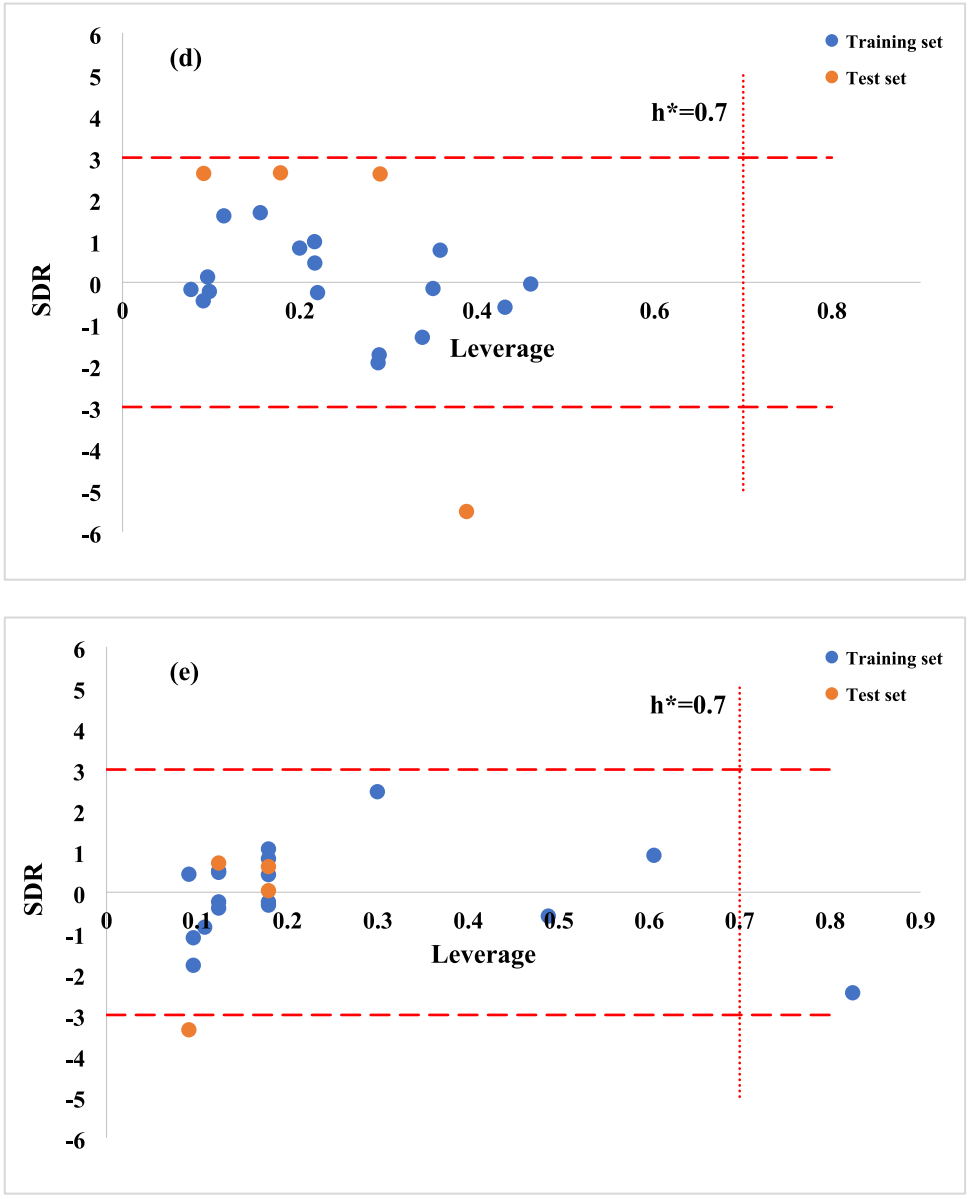


Fig. 5. (*continued*).

Table 5
The values of statistical parameters of selected qspr models for the second dataset for both of training and test sets.

Selected QSPR Models	Eq. No	Sets	Number of datapoints	R2	R2-Adj	Q2-LOO	F	S	RMSE	AARD %	AAD
$b = 0.001512\ TIC3_{Cat} + 0.000012\ ATS2v_{Ani} + 0.037778$	(39)	Training	13	0.99	0.99	0.99	1367	0.0060	0.0053	2.04	0.0040
		Test	3	0.99	–	–	–	–	0.0062	2.38	0.0051
$a_0 = 25.3141\ AATS1p_{Cat} - 6.8906\ AATSC4p_{Ani} - 28.0540$	(40)	Training	13	0.95	0.94	0.93	113	0.6651	0.5834	6.30	0.426
		Test	3	0.97	–	–	–	–	0.9389	14.60	0.888
$c_1 = 0.4776\ GATS4s_{Cat} - 2.0852\ MATS2s_{Cat} - 0.5201\ nHBDon_{Ani} + 1.2581$	(41)	Training	13	0.94	0.92	0.85	49	0.1064	0.0886	7.21	0.072
		Test	3	0.80	–	–	–	–	0.2436	15.64	0.199
$\varepsilon = -1751627.4525\ ETA\_Epsilon\_3_{Cat} + 2974.5314\ ATSC2e_{Ani} + 757103.2451$	(42)	Training	13	0.98	0.98	0.97	321	1281.8	1124.2	5.15	964.5
		Test	3	0.08	–	–	–	–	3632.2	16.6	2480
$\beta = 4.30383\ ETA\_Epsilon\_3_{Cat} + 0.00063\ TIC2_{Ani} - 1.76737$	(43)	Training	13	0.97	0.97	0.95	212	0.0047	0.0042	14.17	0.0033
		Test	3	0.88	–	–	–	–	0.0038	5.13	0.0036

standardized residual (SDR) and 2) the leverage values (h_i). SDR was defined as Eq. (17):

$$SDR = \frac{Y_i^{exp} - Y_i^{pre}}{\sqrt{\frac{\sum_{m=1}^{n}(Y_i^{exp} - Y_i^{pre})^2}{n}}} \quad (17)$$

h_i, represents a measure of a molecule's distance from the center of the training set. It is needed to determine whether new molecules are within the domain of applicability of the developed QSPR model or not. The parameter can be calculated with Eq. (18).

$$h_i\ (\text{or Leverage}\ (i)) = z_i.\ (Z_i^T z_i)^{-1}.Z_i^T \quad (18)$$

When z_i, Z are the descriptor row vector of point i and a n × p matrix of descriptors for compounds derived from the training set, respectively. It should be added that AD of developed QSPR models can be obtained in QSARINS software for each model. The space of AD can be observed in Williams plot. It should be added that maximum leverage (i.e., h*) can be calculated using Eq. (16).

2.3.4. Internal and external Validations

After the development of QSPR models, it is essential to conduct internal and external validations on the training (approx. 80 % of main dataset) and test (approx. 20 % of main dataset) sets. Regarding the internal validation, leave one out −cross validation (LOO-CV), leave multi out −cross validation (LMO-CV), and Y-Scrambling methods can be applied on the selected QSPR models. In fact, these methods were applied on the training set. Regarding the external validation, the prediction capability of selected QSPR models was assessed using a test set. It should be added that both internal and external validations of QSPR models can be done in QSARINS software one by one.

3. Results

3.1. Comparison with former models

The most quantitative and qualitative molecular descriptor which can predict the 'b', 'a_0', and 'c_1' parameters, was 'V_w-IL', (cm^3/mol) (Maia et al., 2012; Frias et al., 2023). This descriptor is a summation of 'V_w-Cation' and 'V_w-Anion' as Eq. (19). Zhao et al. (Zhao et al., 2003) proposed a faster method for calculation of 'V_w' for any structures using 'VABC' descriptor as Eqs. (20) and (21).

$$V_W - I_L = (V_W - Cat) + (V_W - Ani) \quad (19)$$

$$\text{VABC}\left(\frac{\text{Å3}}{\text{molecule}}\right) = \sum allatomcontributions - 5.92\,\text{NB} - 14.7\,\text{RA} - 3.8\,\text{RNR} \quad (20)$$

$N_B = N - 1 + R_g$, $R_g = R_A + R_{NA}$ (N_B = number of bonds, N = total number of atoms, Rg = total number of ring structures)

$$Vw - Cat\ \textbf{\textit{or}}\ Vw - Ani\ (cm3/mol) = 0.602\,VABC - (Cat\ \textbf{\textit{or}}\ Ani)(Å3/molecule) \quad (21)$$

Frias et al. (Frias et al., 2023) suggested a linear model for prediction of 'b', 'a_0', and 'c_1' parameters as Eqs. (22)-(24) (V_W converted to VABC in this study):

$$b = a_1\,(VABC - Cat) + a_1\,(VABC - Ani) + b_1\ Or\ b = a_1(VABC - IL) + b_1 \quad (22)$$

$$a_0 = a_2\,(VABC - Cat) + a_2\,(VABC - Ani) + a_2\ Or\ a_0 = a_2(VABC - IL) + b_2 \quad (23)$$

$$c_1 = c_3\,(VABC - Cat) + a_3\,(VABC - Ani) + b_3\ Or\ c_1 = a3(VABC - IL) + b_3 \quad (24)$$

When, a_1, a_2, a_3, b_1, b_2, and b_3 are adjustable parameters which are obtained using reported values of CPA parameters for each dataset.

In order to make comparison, new models including proposed molecular descriptor by Frias et al. (Frias et al., 2023) have been developed for 'b', 'a_0', and 'c_1' parameters using training sets of each dataset, separately. These models can be demonstrated as Eqs. (25)–(33) in Table 3, along with their statistical parameters (i.e., R^2 and AARD%) for both training and test sets.

The predicted versus regressed values of each parameter are shown in Figs. 1-3 for each dataset, respectively.

As can be seen from Figs. 1-3 and according to obtained statistical parameters in Table 3, 'VABC-IL' as a molecular descriptor can predict only 'b' parameter with high accuracy. The detail of former models can be found in Sheet 5 (Supporting Information-Excel File). According to the obtained results in the proposed study, it was found that 'VABC-IL' cannot be considered as a molecular descriptor for the prediction of 'c_1' parameter for any of the studied IL systems, and for 'a_0' parameter of imidazolium based-ILs. In contrast, the selected QSPR models in this study will show promising capability for the prediction of all five CPA

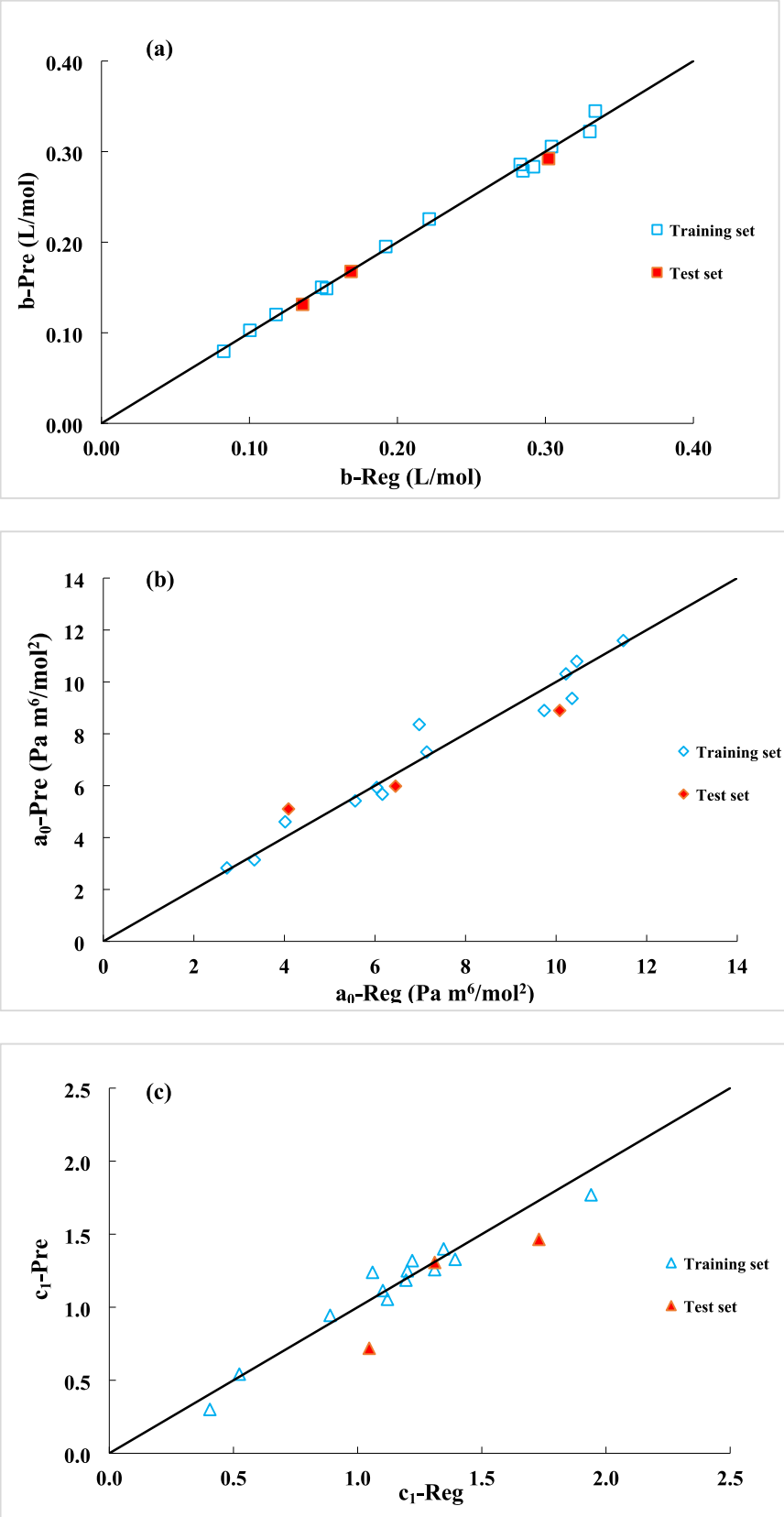
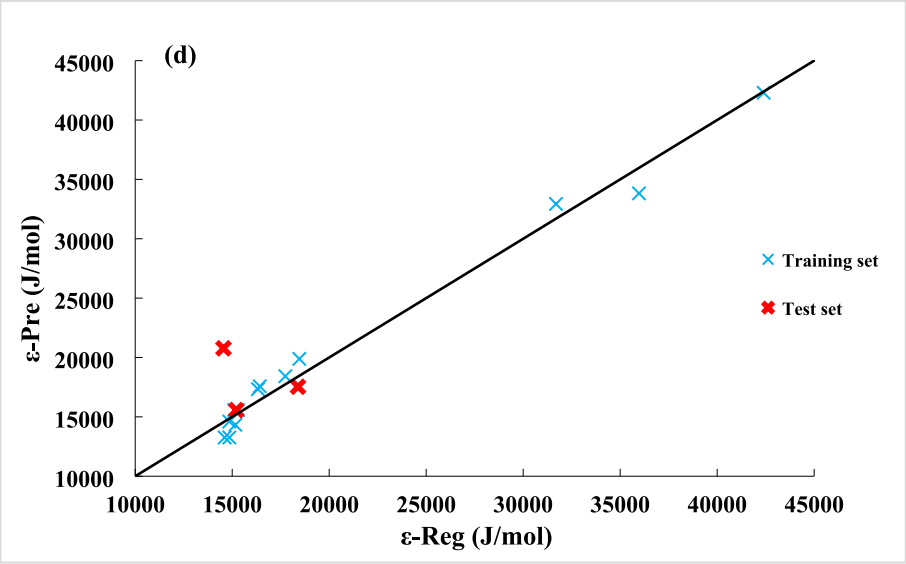


Fig. 6. The predicted versus regressed values of CPA pure parameters (a-e) for the second dataset.



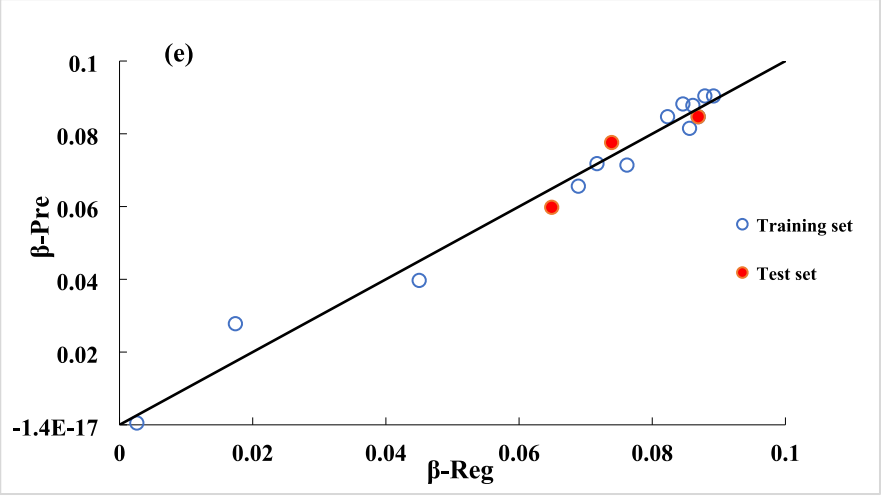


Fig. 6. (*continued*).

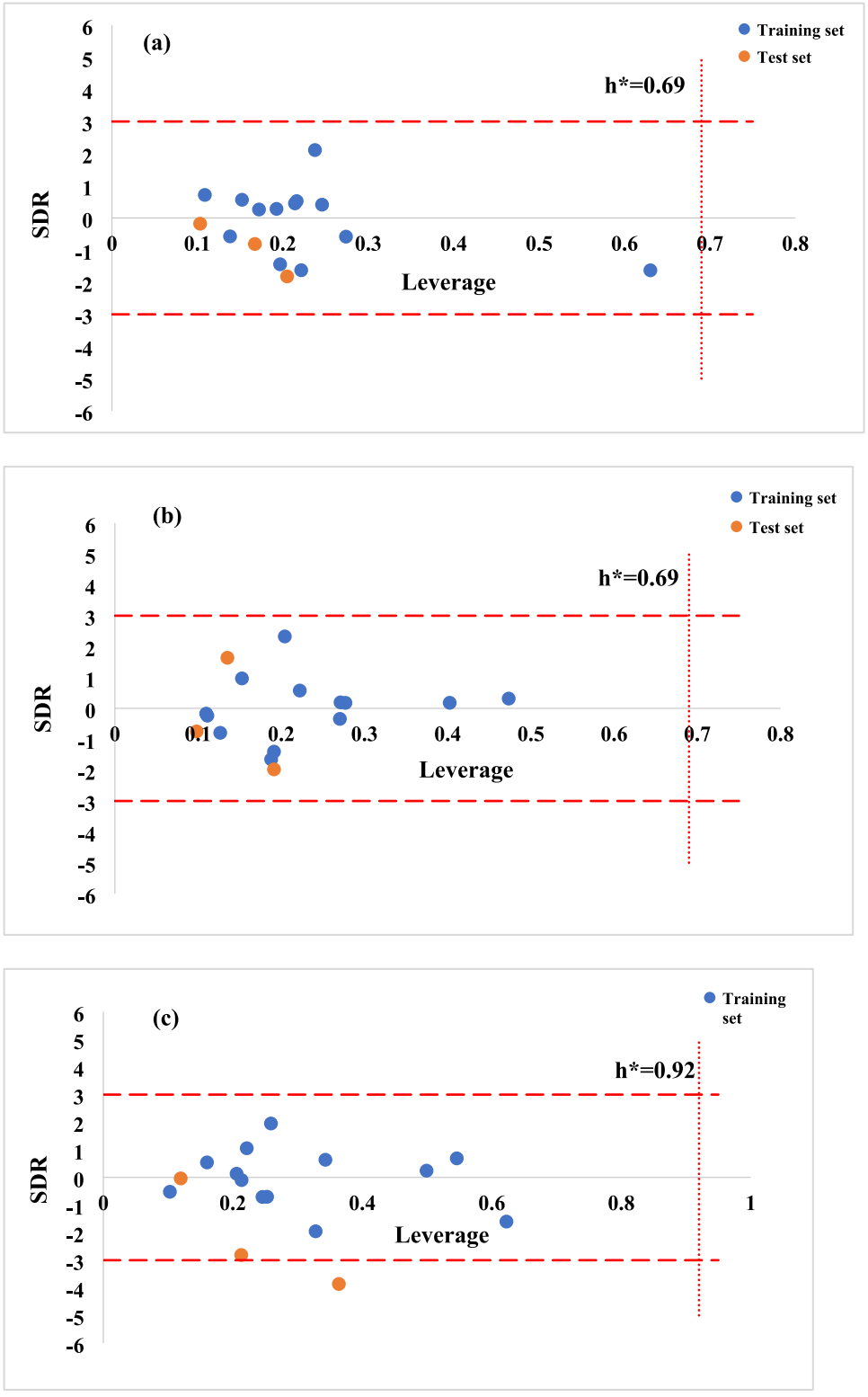
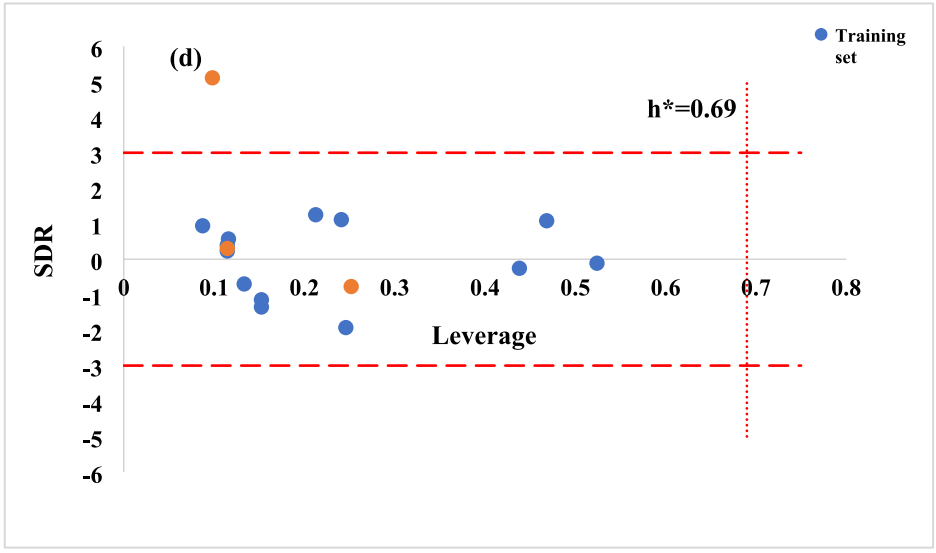


Fig. 7. Williams plots (i.e., SDR versus leverage) of CPA pure parameters b (a), a_0 (b), c_1 (c), ε (d) and β (e) for the second dataset.



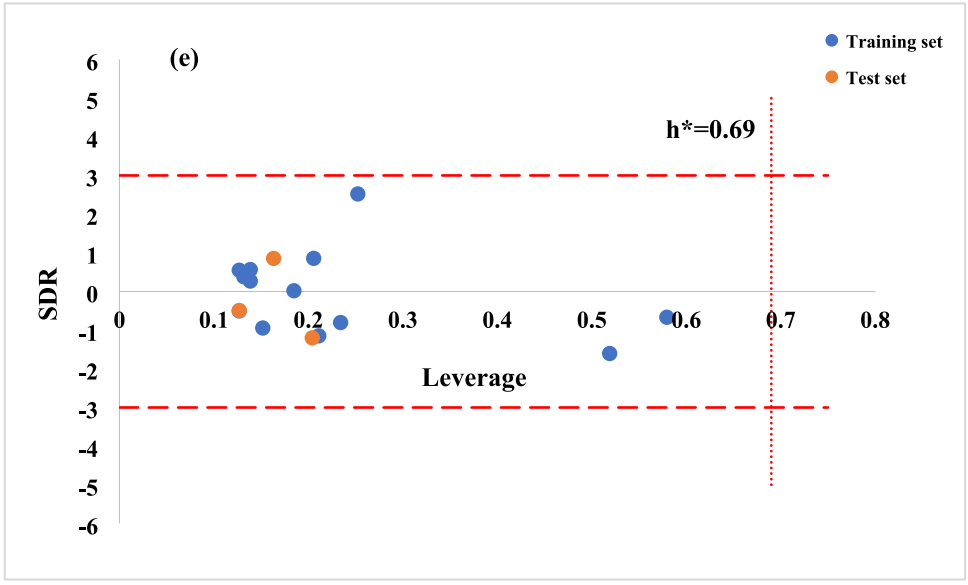


Fig. 7. (*continued*).

Table 6
The values of statistical parameters of selected qspr models for the third dataset for both of training and test sets.

Selected QSPR Models	Eq. No	Sets	Number of datapoints	R2	R2-Adj	Q2-LOO	F	S	RMSE	AARD %	AAD
$b = 0.000916\ VABC_{Cat} + 0.000028\ ATS1v_{Ani} - 0.011493$	(44)	Training	11	0.99	0.99	0.99	3527	0.0028	0.0024	–	–
		Test	2	–[a]	–	–	–	–	0.0063	–	–
$a_0 = 0.2119\ bpol_{Cat} - 19.1142\ ETA\_Shape\_P_{Ani} + 14.2250$	(45)	Training	11	0.99	0.99	0.99	919	0.1811	0.1545	–	–
		Test	2	–	–	–	–	–	0.5271	–	–
$c_1 = 0.050111\ VR2\_DzZ_{Cat} - 34.4245\ JGI8_{Cat} + 0.8126$	(46)	Training	11	0.87	0.83	0.81	26	0.1238	0.1056	–	–
		Test	2	–	–	–	–	–	0.2078	–	–
$\varepsilon = 733.0447\ AATS0i_{Cat} + 15014.0303\ ETA\_Shape\_X_{Ani} - 111577.9362$	(47)	Training	11	0.99	0.99	0.99	649	251.4	214.4	–	–
		Test	2	–	–	–	–	–	215.9	–	–
$\beta = 0.000545\ TIC1_{Cat} - 0.13165\ ETA\_Shape\_P_{Ani} + 0.07585$	(48)	Training	11	0.99	0.98	0.98	406	0.0018	0.0016	–	–
		Test	2	–	–	–	–	–	0.0005	–	–

[a] has not been calculated for test set, due to limitation of datapoints.

Table 7
A comparison of the calculation of IL's density using predicted CPA parameters and reported CPA parameters from the literature for the first group of ILs.

Ionic liquid	T(K)	P(Mpa)	AARD %[a]	AARD %[b]	No. of data	Ref
[emim] [TfO]	298 – 363	0.1 – 69	0.63	0.40	154	(Sequeira et al., 2020)
[Bmim] [NO3]	283 – 343	0.1	3.12	0.04	13	(Mokhtarani et al., 2009)
[bmim] [BF4]	283–323	0.1–60	1.88	0.24	117	(Sanmamed et al., 2010)
[Hemim] [NTF2]	283 – 353	0.1	0.88	0.10	15	(Liu et al., 2015)
[omim] [PF6]	313 – 472	0.1 – 200	1.21	1.12	180	(Taguchi et al., 2009)
[Bzmim] [NTF2]	293 – 373	0.1	3.31	0.06	11	(Mutelet et al., 2017)
[HEA] [L]	293 – 363	0.1	2.98	0.73	8	(Kurnia et al., 2009)
[hmpy] [NTF2]	278 – 353	0.1	2.25	0.87	16	(Liu et al., 2013)
[Bpy] [NO3]	293 – 338	0.1	1.30	0.19	10	(Wang et al., 2012)

a = calculated with predicted parameters by QSPR models.
b = calculated with reported parameters from the literature.

pure parameters for different kinds of ILs. These descriptors in each selected QSPR model will be selected from an extensive pool of molecular descriptors using GA method. Since, different validation techniques will be applied for each selected model, it is strongly suggested to predict the CPA pure parameters using QSPR models, instead of former models including 'VABC-IL'.

3.2. Developed QSPR models

In this study, each dataset was divided to training and test sets for performing of internal and external validations. Therefore, four ILs (i.e., ([bmim] [BF_4], [hmim] [PF_6], [emim] [Ac], and [omim] [NTF_2])), three ILs (i.e., ([BHEMA] [L], [N112,3–C3OC] [Tf_2N], and [BHEA] [Ac])), and two ILs (i.e., ([3-MPpy] [Tf_2N] and [C7py] [Tf_2N])) were set aside into test sets and remained ILs kept into training sets for first, second, and third datasets, respectively (see Table 1, Bold ILs).

3.2.1. Imidazolium based-ILs

Regarding the first dataset (i.e., Imidazolium), developed QSPR models with one, two, and three descriptors using training set (i.e., 17 datapoints), have been shown in Table S2.

According to the obtained values of R^2, the developed QSPR models with three descriptors have been selected for all parameters, except 'b'. As can be observed in Table S2, the developed QSPR model with two descriptors had acceptable accuracy for 'b' parameter. The predicted values of each parameter using QSPR model, are shown in Table 1. The statistical parameters of each selected model for each parameter are shown in Table 4.

As indicated in Table 4, the obtained value of Q^2_{LOO} (as internal validation) of each selected model for each parameter was high which is confirming that each model has acceptable capability for prediction of corresponded CPA parameter. Also, LMO-CV and Y-scrambling techniques have been done on the training set in the QSARINS software for each selected QSPR model and results confirmed the validity of each model. As external validation, there were also shown that the CPA pure parameters of each IL in test set predicted with enough accuracy based on the obtained values of AARD%. The plots of predicted versus regressed values for each parameter which were obtained using selected QSPR models (see Table 4), are shown in Fig. 4.

The Williams plots for each parameter which were obtained using selected QSPR models (see Table 4), can be seen in Fig. 5. According to these plots, it was found that there were no outlier datapoints in the first dataset.

The values of each descriptor in the selected models for the first dataset can be found in Sheet 6 (Supporting Information-Excel File). The intercorrelation between descriptors in the selected models, are presented in Sheet 7 (Supporting Information-Excel File). It can be seen that the descriptors in the selected models are independent from each other.

3.2.2. Ammonium based-ILs

Regarding the second dataset (i.e., ammonium), developed QSPR models with one and two descriptors using training set (i.e., 13 datapoints), are shown in Table S3.

According to the obtained values of R^2, the developed QSPR models with two descriptors have been selected for all parameters, except 'c_1''. For 'c_1'' parameter, a QSPR model with three descriptors was selected. Unlike first dataset and as can be observed in Table S3, the developed QSPR model with one descriptor (i.e., cationic descriptor) had

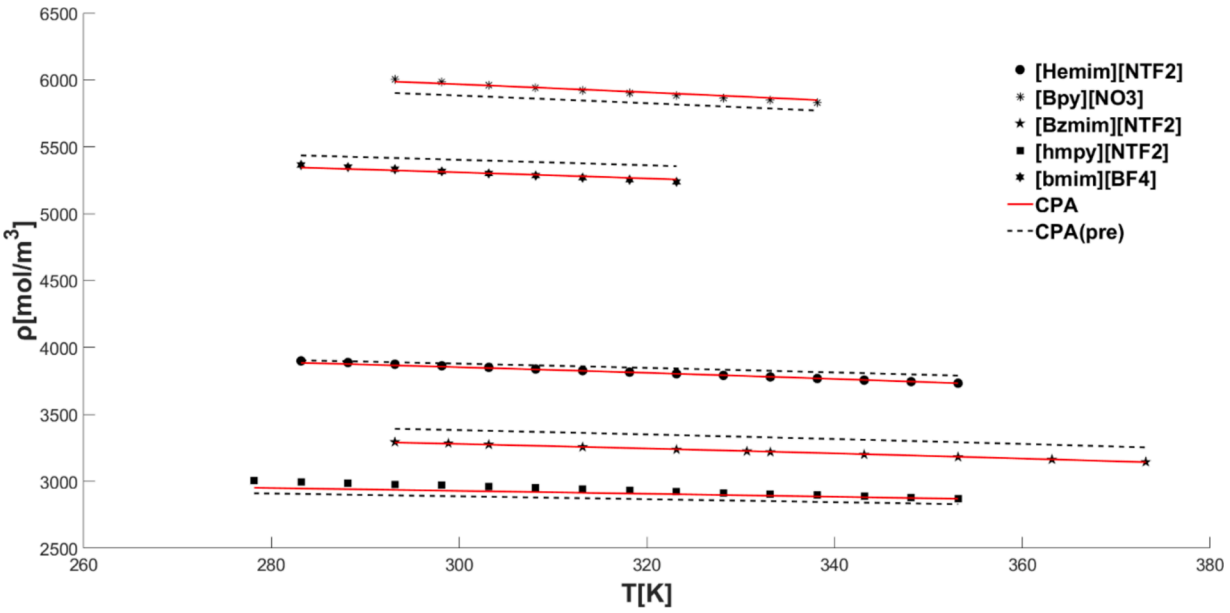


Fig. 8. Experimental (Sanmamed et al., 2010; Liu et al., 2015; Mutelet et al., 2017; Liu et al., 2013; Wang et al., 2012) and calculated densities with CPA for ILs included in training set using both fitted (solid line) and QSPR predicted (dash line) parameters.

Table 8
Density calculation using CPA and QSPR predicted parameters for second group of ILs.

Ionic liquid	T(K)	P(Mpa)	AARD %	No. of data	Ref
[hmim] [NO3]	283 – 323	0.1	3.27	8	(Khedri et al., 2019)
[bmim] [Ac]	273 – 363	0.1	3.61	13	(Harris, 2020)
[omim] [SCN]	288 – 343	0.1	5.17	12	(Heidari et al., 2012)
[bmim] [TfO]	283 – 413	0.1 – 140	1.16	146	(Safarov et al., 2019)
[TBMA] [NTF2]	283–353	0.1	8.30	15	(Bhattacharjee et al., 2014)
[HEA] [NO3]	303–333	0.1	11.12	7	(Alves Camêlo et al., 2020)
[b4mpy] [NTF2]	283–338	0.1	2.49	12	(Liu et al., 2012)

acceptable accuracy for all parameters (except 'c_1'). It means that these parameters can be predicted, regardless of anion structures. In another words, the cation structure had a significantly stronger effect on these parameters as the anion structures in the second dataset. The predicted values of each parameter using related QSPR model are shown in Table 1. The statistical parameters of each selected model for each parameter are shown in Table 5.

As can be indicated in Table 5 and according to the statistical parameters, the obtained value of Q^2_{LOO} (as internal validation) of each selected model for each parameter was high which is confirming that each model has acceptable capability for prediction of corresponded CPA parameter. Also, LMO-CV and Y-scrambling techniques have been done on the training set in the QSARINS software for each selected QSPR model and results confirmed the validity of each model. As external validation, there were also shown that the CPA pure parameters of each IL in test set predicted with enough accuracy based on the obtained values of AARD%. The plots of predicted versus regressed values for each parameter which were obtained using selected QSPR models (see Table 5), are shown in Fig. 6.

The Williams plots for each parameter which were obtained using selected QSPR models (see Table 5), can be seen in Fig. 7. According to these plots, it was found that there were no outlier datapoints in the second dataset.

The values of each descriptor in the selected models for the second dataset can be found in Sheet 8 (Supporting Information-Excel File). The intercorrelation between descriptors which were appeared in the selected models, has been presented in Sheet 9 (Supporting Information-Excel File). As can be resulted from Sheet 9 (Supporting Information-Excel File), the descriptors in the selected models were independent from each other.

3.2.3. *Pyridinium based-ILs*

Regarding the third dataset (i.e., pyridinium), developed QSPR models with one and two descriptors using training set (i.e., 11 datapoints),have been shown in Table S4.

According to the obtained values of R^2, the developed QSPR models with two descriptors have been selected for all parameters. For 'c_1' parameter, a QSPR model with two cationic descriptors was selected which showed that anion structures had not tangible effect on this parameter. The predicted values of each parameter using related QSPR model, are shown in Table 1. The statistical parameters of each selected model for each parameter are shown in Table 6.

As can be indicated in Table 6 and according to the statistical parameters, the obtained value of Q^2_{LOO} (as internal validation) of each selected model for each parameter was high which is confirming that each model has acceptable capability for prediction of corresponded CPA parameter. Also, LMO-CV and Y-scrambling techniques have been done on the training set in the QSARINS software for each selected QSPR model and results confirmed the validity of each model. As external validation, there were also found that the CPA pure parameters of each IL in test set predicted with enough accuracy based on the obtained values of RMSE.

The values of each descriptor in the selected models for the third dataset can be found in Sheet 10 (Supporting Information-Excel File). The intercorrelation between descriptors which were appeared in the selected models, are presented in Sheet 11 (Supporting Information-Excel File). As can be resulted from Sheet 11 (Supporting Information-Excel File), the appeared descriptors in the selected models were independent from each other.

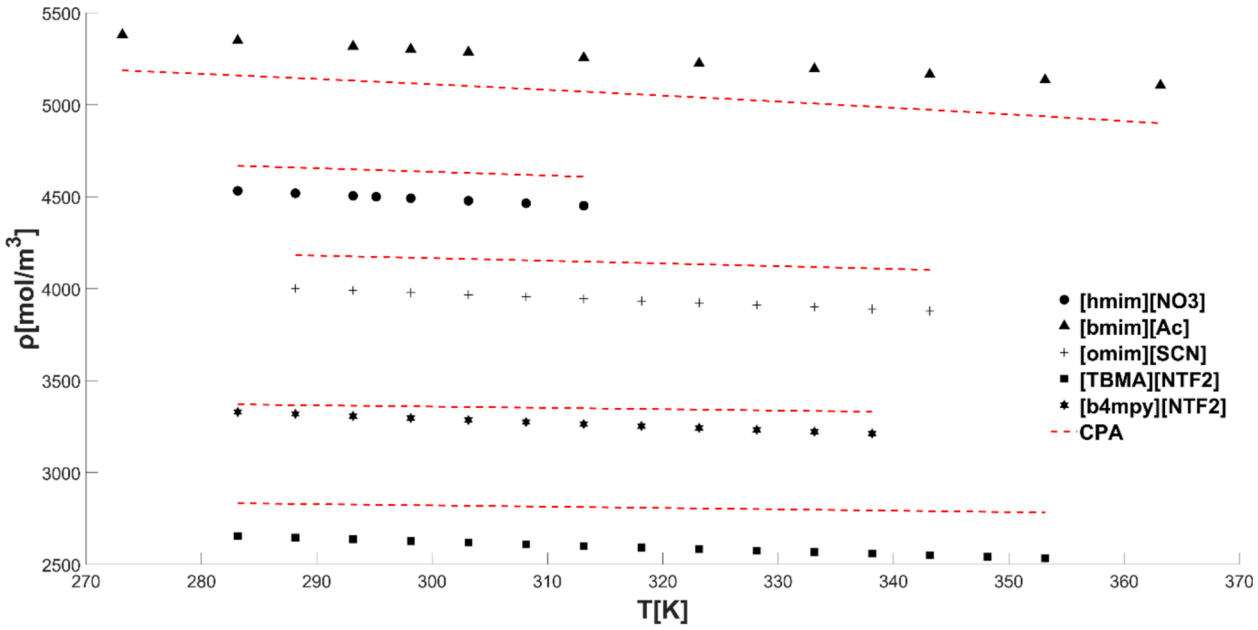


Fig. 9. Experimental (Khedri et al., 2019; Harris, 2020; Heidari et al., 2012; Bhattacharjee et al., 2014; Liu et al., 2012) and calculated densities of IL's using predicted CPA parameter by QSPR models.

Table 9
Vapor pressures for imidazolium based ionic liquids (included in training set) with QSPR predicted and fitted parameters.

T(K) IL	P(Pa) [emim] [TfO][a]	[emim] [TfO][b]	[Bmim] [NO3][a]	[Bmim] [NO3][b]	[omim] [PF6][a]	[omim] [PF6][b]
293.15	4.83	0.34	2.44	4.60	0.01	0.07
303.15	11.31	0.89	6.19	10.00	0.03	0.23
313.15	24.82	2.13	14.50	20.61	0.11	0.69
323.15	51.33	4.82	31.56	40.45	0.32	1.91
333.15	100.65	10.25	64.46	75.98	0.91	4.90
343.15	188.10	20.72	124.45	137.09	2.34	11.73
353.15	336.58	39.93	228.61	238.43	5.67	26.33
363.15	578.98	73.69	401.74	400.95	12.89	55.78
373.15	960.86	130.74	678.60	653.75	27.67	112.14

a = calculated with parameters obtained from QSPR models.
b = calculated with parameters from literature.

3.3. Calculation of IL's density using predicted CPA parameters

In this section, densities of two separate groups of ILs were calculated using predicted CPA parameters by QSPR models. For the first group, the ILs were present in the studied datasets (see Table 1), while, in the second group, ILs were absent in the studied datasets. Critical properties needed for all CPA calculation are estimated by Jock-Reid group contribution method (Joback and Reid, 1987) which is extended for ionic liquids (Valderrama et al., 2012).

Calculated densities for first group are shown in Table 7 and Fig. 8. In overall, QSPR-model could predict densities with reasonable accuracy (within 0.6–3.3 % AARD). One IL ([hmpy] [NTF2]) turn out to be interesting, since parameter deviation between QSPR and literature values were low for all parameters ($b = 1.4$, $a_0 = 0.2$, $c_1 = 2.7$, $\varepsilon = 0.5$ and $\beta = 1$ ARD%) still this is already producing over 1 % AARD difference for density, reflecting the challenge for predicting parameters for highly non-linear thermodynamical model with extremely high accuracy.

Calculated densities for second group of ILs using predicted CPA parameters from QSPR models, were shown in Table 8 and Fig. 9. Imidazolium and pyridinium based ILs predictions were more accurate than compared with ammonium based ionic liquids. This same trend is presented for CPA parameters that have been regressed in literature, thus it seems that density accuracy from literature is transferred to QSPR predictions.

3.4. Calculation of IL's vapor pressure using predicted CPA parameters

Vapor pressures were calculated for first group of studied ILs and the results for imidazolium, ammonium and pyridinium based IL are shown

Table 10
Vapor pressures for ammonium and pyridinium based ionic liquids (included in training set) with QSPR-predicted and fitted parameters.

T(K) IL	P(Pa) [HEA][L][a]	[HEA][L][b]	[THEMA][MeSO4][a]	[THEMA][MeSO4][b]	[hmpy][NTF2][a]	[hmpy][NTF2][b]	[Bpy][NO3][a]	[Bpy][NO3][b]
293.15	7.16E-04	4.62E-04	1.33E-03	8.27E-04	0.65	0.65	5.91	11.45
303.15	2.51E-03	1.65E-03	4.40E-03	2.81E-03	1.45	1.45	13.23	24.52
313.15	8.06E-03	5.40E-03	1.32E-02	8.67E-03	3.05	3.05	27.87	49.62
323.15	2.39E-02	1.63E-02	3.68E-02	2.46E-02	6.10	6.10	55.66	95.45
333.15	6.60E-02	4.56E-02	0.10	6.53E-02	11.65	11.64	105.88	175.38
343.15	0.17	0.12	0.23	0.16	21.32	21.28	192.73	309.16
353.15	0.42	0.30	0.54	0.38	37.56	37.47	337.11	524.86
363.15	0.96	0.69	1.17	0.84	63.89	63.70	568.58	861.05
373.15	2.10	1.53	2.44	1.78	105.24	104.86	927.71	1369.10

a = calculated with parameters obtained from QSPR models.
b = calculated with parameters from literature.

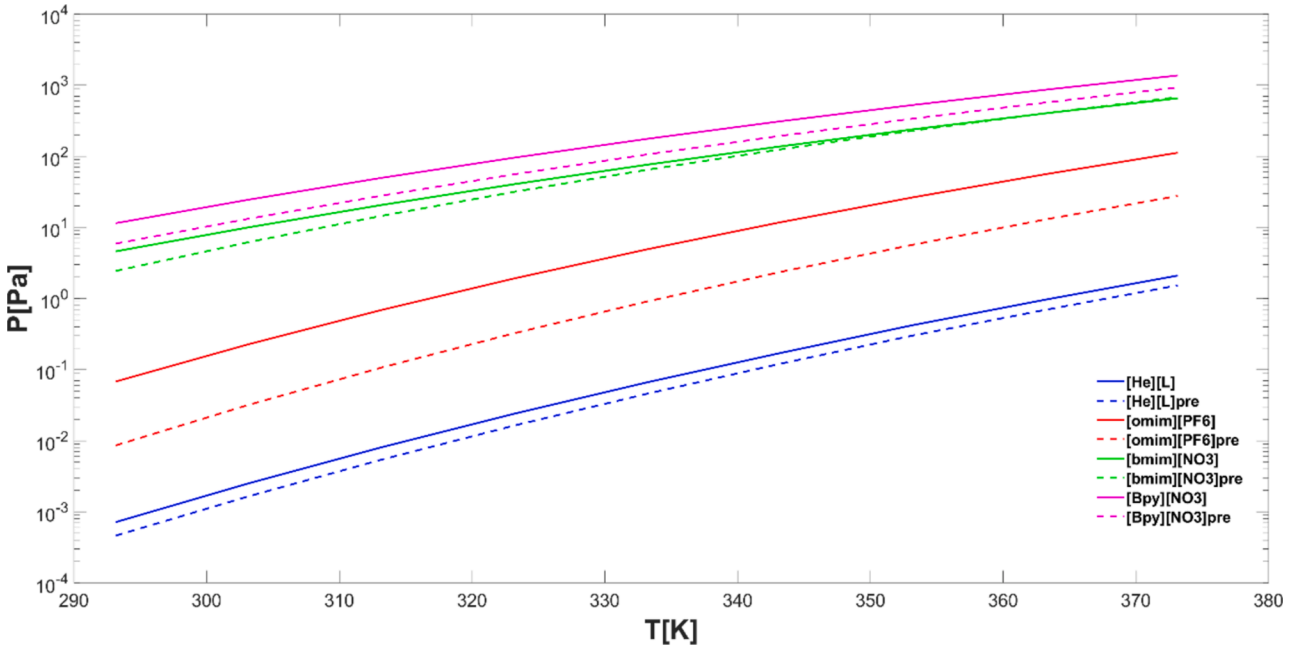


Fig. 10. CPA prediction for vapor pressure of ILs (included in training set or first group) using QSPR predicted parameters.

Table 11
Vapor pressures for second group of ILs (not included in training set) using QSPR predicted parameters.

T(K) IL	P(Pa) [hmim] [NO3]	[bmim] [Ac]	[bmim] [TfO]	[TBMA] [NTF2]	[HEA] [NO3]	[b4mpy] [NTF2]
293.15	0.13	28.66	3.05	1.43E-05	1.33E-11	6.20E-14
303.15	0.37	64.55	7.35	8.45E-05	1.48E-05	1.66E-13
313.15	1.01	136.39	16.53	3.23E-04	2.00E-04	4.19E-13
323.15	2.54	272.05	34.97	1.06E-03	9.24E-04	9.97E-13
333.15	6.01	515.09	70.04	3.17E-03	3.10E-03	9.83E-06
343.15	13.38	930.33	133.45	8.77E-03	9.60E-03	7.28E-05
353.15	28.22	1610.00	242.98	2.27E-02	2.77E-02	2.80E-04
363.15	56.67	2680.10	424.54	5.52E-02	7.48E-02	9.08E-04
373.15	108.75	4306.60	714.41	0.12733	0.19018	2.68E-03

in Tables 9 and 10 and Fig. 10. Low vapor pressures (in range of $1{\cdot}10^{-4}$ – $1{\cdot}10^{3}$ Pa) were obtained for all studied IL with minor differences between QSPR-predicted and fitted parameters. It should be noted that IL experimental vapor pressure values are not included in fitting process for majority of ionic liquids in training set. Interestingly, [hmpy][NTF2] produces almost precisely same vapor pressure (differences under 2 Pa) when it is calculated with QSPR-predicted and fitted parameters. It seemed that for this IL, small differences in parameters from QSPR-prediction did not affect of vapor pressure, but it was enough for producing difference in density predictions described in section 3-3. Vapor pressure calculation are showing that CPA is capable of describe ionic liquids low vapor pressure with both QSPR-predicted and fitted parameters.

Vapor pressure calculations for IL in second group (IL not in the training set) are shown in Table 11 and Fig. 11. QSPR predicted parameters for few IL's ([HEA][NO3] and [bmpy][NTF2]) produced extremely low vapor pressure at low temperature and this kind of vapor pressure were not presented in first group of IL's. However, it seemed QSPR predicted parameters followed same trend than in literature where imidazolium based ionic liquids produced higher vapor pressure than ammonium-based IL while all predicted vapor pressured are qualitatively low, which is typically expected for ILs.

It should be noted that IL experimental vapor pressure values were not included in fitting process of datasets 2 and 3 (see Table S1 (Frias et al. (2023)). this physical property has been predicted very well for [EA][NO3] (Emel'yanenko et al., 2014) (AAD = 0.43) using the predicted CPA parameters by QSPR models. But for another specified IL in Fig. 12, the predicted values of vapor pressure had deviation from experimental data (Rocha and Santos, 2013).

4. Discussion

To introduce the appeared descriptors in the selected QSPR models for each dataset, it should be mentioned that 'ATS2v', 'ATSC8c', 'ATSC3m', 'ATSC2i', 'AATSC5e', 'GATS1s', 'AATSC2i', 'AATS3e', 'ATSC8m', 'AATS1s', 'AATS1p', 'AATSC4p', 'GATS4s', 'MATS2s', 'ATSC2e', 'ATS1v', and 'AATS0i' descriptors are Autocorrelation descriptors (Todeschini and Consonni, 2009). These descriptors are topological and can be calculated using molecular graphs. Detail of these descriptors can be found in Moreau and Broto (Moreau and Broto, 1980; Moreau and Broto, 1980; Broto et al., 1984) as well as Moran (Moran, 1950) and Geary (Geary, 1954). 'TopoRadius' descriptor shows the minimum atom eccentricity (Lemaoui et al., 2020). 'nHBDon' and 'nHBDon_Lipinski' show the number of hydrogen bond donors and number of hydrogen bond donors using Lipinski's definition, respectively (Lemaoui et al., 2020). 'nHBd' and 'mindO' as the electrotopological state (E-state) atom type descriptors, express the counting of E-states for strong hydrogen bond donors and minimum atom-type E-state (O), respectively (Hall and Kier, 1995). 'TIC1′, 'TIC2′, and 'TIC3′ as information content descriptors, show the total information content index (Todeschini and Consonni, 2009). 'ETA_Epsilon_3′, 'ETA_Shape_P', and 'ETA_Shape_X' as extended topochemical atom descriptors (Lemaoui et al., 2020; Roy and Ghosh, 2004), show a measure of electronegative of atom count, shape index P, and shape index X, respectively. 'VR2_DzZ' as a Barysz matrix descriptor, shows normalized Randic-like eigenvector-based index and can be calculated from

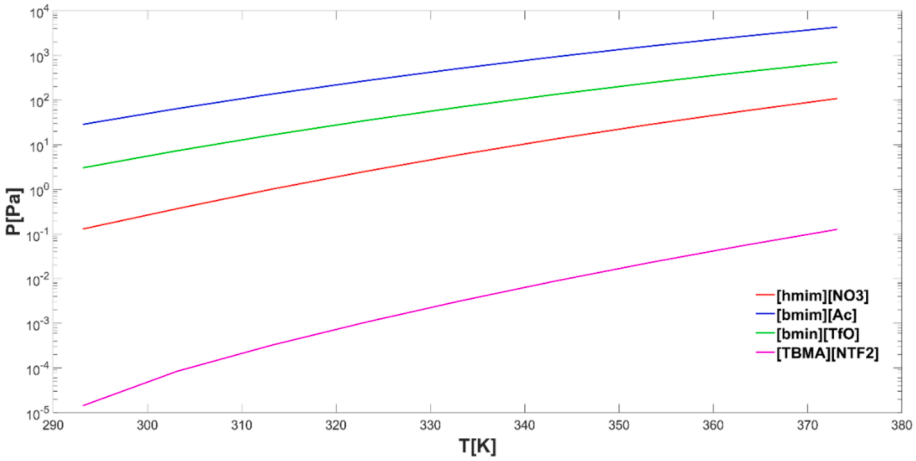


Fig. 11. Vapor pressure calculation of second group of IĹs (not included in training set) using QSPR predicted parameters.

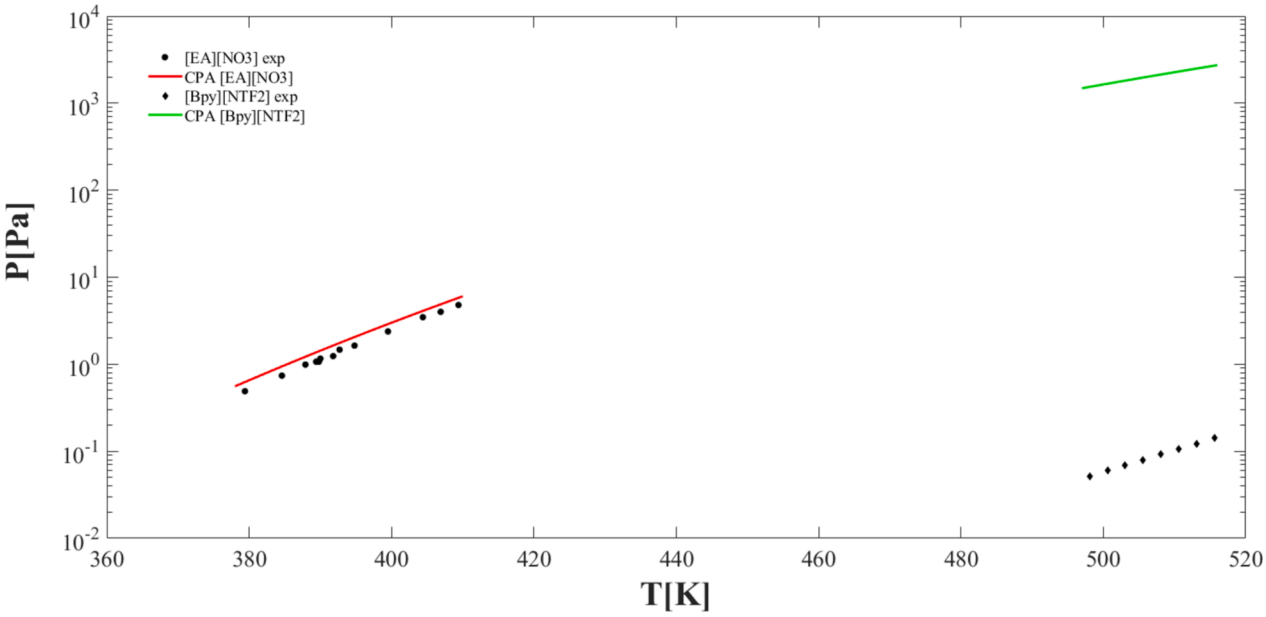


Fig. 12. Experimental (Emel'yanenko et al., 2014; Rocha and Santos, 2013) and predicted vapor pressures using predicted CPA parameters for [EA][NO3] and [Bpy][NTF2] ILs.

Barysz matrix (Todeschini and Consonni, 2009). 'bpol' shows the sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (Lemaoui et al., 2020). 'JGI8′ as topological charge descriptor,expresses mean topological charge index (Todeschini and Consonni, 2009). The definition of each above descriptor has been listed in Table 12.

According to the selected QSPR models for each dataset, the CPA pure parameters for some new imidazolium, ammonium, and pyridinium-based ILs which had never reported in the literature, can be predicted for the first time. For this regard, the leverage of each new IL should be calculated first. According to the Eq. (18), leverage (i.e., h) depends on the appeared descriptors of selected QSPR models for each parameter. As can be seen in Sheet 12 (Supporting Information-Excel File), the calculated leverage for some new ILs were within the space of AD of the developed QSPR models (lower than related h*), while, the h for some new ILs such as [Hemim][Ac], [NH_2emim][Ac], [Bzmim][Ac], [NH_2emim][NTF_2], [Hemim][BF_4], [Bzmim][BF_4], [Bzmim][PF_6], [Hemim][TfO], [NH_2emim][TfO], and [Bzmim][TfO] was higher than h* (see Sheet 12 (Supporting Information-Excel File)). This means that the selected QSPR models should not be applied for these ILs. All in all, it is possible to predict the CPA pure parameters for some new imidazolium based ILs which have been listed in Table 13. Same procedure can be applied for ammonium and pyridinium-based ILs, too.

Among the five CPA pure parameters, 'b' parameter was independent of used 'OF' or parametrization process. This point helps to create an extensive dataset of 'b' parameter for the development a general predictive model which can cover a large number of ILs. The detail of this dataset can be found in Sheet 13 (Supporting Information-Excel File).

General predictive QSPR model has two molecular descriptors (cationic and anionic descriptors) which can predict the values of 'b' parameter for 75 ILs. 'ECCEN' as cationic descriptor, shows the eccentric connectivity index of cations (Sharma et al., 1997) and 'MW' as anionic descriptor shows the molecular mass of anions. As can be seen in Fig. 13, Eq. (49) with these two descriptors has promising capability for the prediction of 'b' parameter for an extensive range (0.07 – 0.67). In comparison, the former suggested descriptor in the literature (VABC-IL) has not shown high accuracy for this range, particularly for large ILs. The availability of such molecular-based model helps to validate suspicious reported values of 'b' parameter. For example, one of observed conflicts in Table S1 was the reported values of 'b' for [bmim][NTf_2].

Table 12
Definition of each descriptor in PaDEL software (Yap, 2011).

Datasets	Descriptors	Definition
First	TopoRadius	Topological radius (minimum atom eccentricity)
	ATS2v	Broto-Moreau autocorrelation – lag 2 / weighted by van der Waals volumes
	ATSC8c	Centered Broto-Moreau autocorrelation – lag 8 / weighted by charges
	ATSC3m	Centered Broto-Moreau autocorrelation – lag 3 / weighted by mass
	ATSC2i	Centered Broto-Moreau autocorrelation – lag 2 / weighted by first ionization potential
	AATSC5e	Average centered Broto-Moreau autocorrelation – lag 5 / weighted by Sanderson electronegativities
	GATS1s	Geary autocorrelation – lag 1 / weighted by I-state
	AATSC2i	Average centered Broto-Moreau autocorrelation – lag 2 / weighted by first ionization potential
	AATS3e	Average Broto-Moreau autocorrelation – lag 3 / weighted by Sanderson electronegativities
	ATSC8m	Centered Broto-Moreau autocorrelation – lag 8 / weighted by mass
	AATS1s	Average Broto-Moreau autocorrelation – lag 1 / weighted by I-state
	nHBDon_Lipinski	Number of hydrogen bond donors (using Lipinski's definition: Any OH or NH. Each available hydrogen atom is counted as one hydrogen bond donor)
	nHBd	Count of E-States for (strong) Hydrogen Bond donors
	mindO	Minimum atom-type E-State: =O
Second	TIC3	Total information content index (neighborhood symmetry of 3-order)
	ATS2v	Broto-Moreau autocorrelation – lag 2 / weighted by van der Waals volumes
	AATS1p	Average Broto-Moreau autocorrelation – lag 1 / weighted by polarizabilities
	AATSC4p	Average centered Broto-Moreau autocorrelation – lag 4 / weighted by polarizabilities
	GATS4s	Geary autocorrelation – lag 4 / weighted by I-state
	MATS2s	Moran autocorrelation – lag 2 / weighted by I-state
	nHBDon	Number of hydrogen bond donors (using CDK HBondDonorCountDescriptor algorithm)
	ETA_Epsilon_3	A measure of electronegative atom count
	ATSC2e	Centered Broto-Moreau autocorrelation – lag 2 / weighted by Sanderson electronegativities
	TIC2	Total information content index (neighborhood symmetry of 2-order)
Third	VABC	Van der Waals volume
	ATS1v	Broto-Moreau autocorrelation – lag 1 / weighted by van der Waals volumes
	bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)
	ETA_Shape_P	Shape index P
	VR2_DzZ	Normalized Randic-like eigenvector-based index from Barysz matrix / weighted by atomic number
	JGI8	Mean topological charge index of order 8
	AATS0i	Average Broto-Moreau autocorrelation – lag 0 / weighted by first ionization potential
	ETA_Shape_X	Shape index X
	TIC1	Total information content index (neighborhood symmetry of 1-order)

Table 13
The predicted values of cpa pure parameters for some new imidazolium, ammonium and pyridinium based ILs using selected QSPR models.

	h (b-Para)	b-Pre (L/mol)	h (a_0-Para)	a_0-Pre (Pa m^6/mol^2)	h (c_1-Para)	c_1-Pre	h (ε-Para)	ε-Pre (J/mol)	h (β-Para)	β-Pre
[emim] [NO3]	0.1954	0.1348	0.1618	3.3939	0.0999	1.4058	0.3125	19034.7675	0.6052	0.0511
[hmim] [NO3]	0.1515	0.1988	0.2041	2.9789	0.0812	2.141	0.4531	18235.665	0.6052	0.0511
[omim] [NO3]	0.291	0.2308	0.1841	3.9629	0.0867	2.2088	0.5206	25932.4386	0.6052	0.0511
[bmim] [Ac]	0.0963	0.1762	0.249	3.3848	0.1354	1.3655	0.1104	16665.2522	0.091	0.0073
[omim] [Ac]	0.2745	0.2402	0.2775	3.166	0.1157	2.3082	0.2871	19435.4362	0.091	0.0073
[bmim] [SCN]	0.1361	0.1613	0.27	3.229	0.4253	1.8989	0.1185	18276.3371	0.4884	0.0445
[hmim] [SCN]	0.166	0.1933	0.3346	2.0262	0.4374	2.7738	0.227	13349.7475	0.4884	0.0445
[omim] [SCN]	0.3034	0.2253	0.3	3.0102	0.4569	2.8415	0.2966	21046.521	0.4884	0.0445
[bmim] [TfO]	0.0593	0.2073	0.1293	4.6623	0.11	1.2127	0.11	16768.7617	0.0959	0.0063
[omim] [TfO]	0.2599	0.2712	0.1453	4.4435	0.0756	2.1554	0.2867	19538.9457	0.0959	0.0063
[bmim] [EtSO4]	0.0597	0.2056	0.1447	4.4218	0.1131	1.2441	0.2148	12925.5413	0.1085	0.0043
[hmim] [EtSO4]	0.1056	0.2376	0.1794	3.219	0.0767	2.119	0.3161	7998.9518	0.1085	0.0043
[omim] [EtSO4]	0.2591	0.2696	0.1631	4.203	0.0818	2.1867	0.388	15695.7253	0.1085	0.0043
[TBMA] [NTF2]	0.2178	0.3365	0.3197	12.1704	0.1028	1.2575	0.1877	11596.4057	0.1540	0.0945
[HEA][NO3]	0.2306	0.0961	0.2721	3.2543	0.5459	1.2495	0.2321	33186.7719	0.4114	0.0229
[b4mpy] [NTF2]	0.1613	0.2853	0.1627	8.8233	0.8373	1.8123	0.2096	15457.4819	0.1445	0.0492

According to the result of molecular modelling which has been done for 75 ILs, it was found that the predicted value of 'b' for [bmim][NTf_2] is 0.27. It means that reported values (approx., 0.27) for this IL by (Maia et al., 2012; Zhu et al., 2021; Hamedi et al., 2020) are more reliable and acceptable than 0.32 which was reported by Manic et al. (Manic et al., 2012). Therefore, such molecular insights and structural-based model can remove such scientific conflict. The detail can be found in Sheet 13 (Supporting Information-Excel File).

$$b = 0.00043\ ECCEN_{Cat} + 0.00052\ MW_{Ani} + 0.07824 \quad (49).$$

5. Conclusion

In this study, CPA pure parameters of three kinds of ILs (imidazolium, ammonium, and pyridinium) have been predicted using QSPR models for the first time. The obtained values of statistical parameters (R^2, $Q^2_{LOO\text{-}CV}$, RMSE, and AARD%) of each model were satisfactory and acceptable for both of training and test sets. The obtained results of performing of internal and external validations verified that the prediction of CPA pure parameters of some new imidazolium, ammonium,

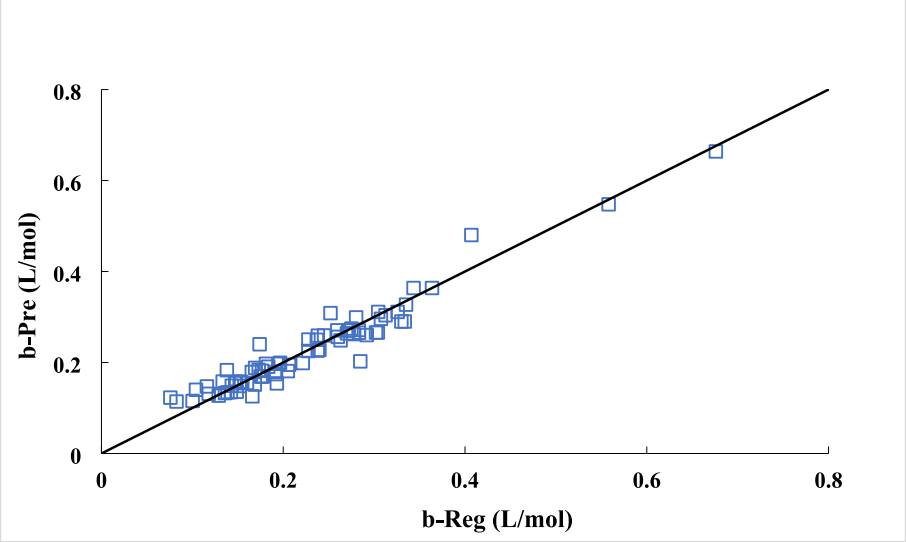


Fig. 13. A general predictive QSPR model for prediction of 'b' parameter for vast number of ILs using Eq. (49).

and pyridinium based ILs can be practical with respect to the leverage value of new IL and AD of each QSPR model. According to the results, it was found that 'VABC-IL' as former molecular descriptor in the literature had a significant deficiency for the prediction of 'c_1' parameter. And, it had not also acceptable accuracy for the prediction of 'a_0' parameter of imidazolium-based IL. The new suggested molecular descriptors in this study had promising capability for prediction of all 5 pure parameters. Among of 5 pure parameters, it was shown that only 'b' parameter of ILs was independent of used 'OFs' for the first time based on a comprehensive literature survey. For this reason, a general QSPR model was developed for this parameter. Also, it is concluded that the availability of such general QSPR model which was developed using a vast variety of ILs (75 ILs) can remove some observed conflicts in the literature. For example, it was confirmed that the value of 'b' parameter of [bmim][NTF_2] should be 0.27 using the general QSPR model. Density and vapor pressure calculations verified that QSPR models can predict CPA parameters for new ILs which have not reported before. Density calculations for studied ILs indicated that QSPR can predict parameters for ILs which CPA parameters have not been reported previously with 1.2–5.2 AARD% range for imidazolium based ILs and for ammonium and pyridium based ILs with 2.5–11.1 AARD%. QSPR-predicted parameters also produced qualitatively low vapor pressures for studied ILs showing promising result for QSPR-predicted CPA parameters.

CRediT authorship contribution statement

Ali Ebrahimpoor Gorji: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Juho-Pekka Laakso:** Writing – review & editing, Methodology, Data curation. **Ville Alopaeus:** Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition, Conceptualization. **Petri Uusi-Kyyny:** Writing – review & editing, Visualization, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was finically supported by the 'In-situ equilibrium shifting in CO2 utilization reactions by novel absorbents (CO2Shift)' Project (351113). The authors gratefully and thankfully appreciate to **Prof. Paola Gramatica** (University of Insubria) for providing the free license of QSARINS software for the development of Multilinear Regression models. Also, the authors gratefully and thankfully appreciate to **Prof. Georgios Kontogeorgis** (Technical university of Denmark) providing MATLAB code for CPA modelling.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ces.2024.120825.

Data availability

Data will be made available on request.

References

Afsharpour, A., 2019. Application of reaction equilibrium thermodynamic model for correlation of H2S solubility in ionic liquids [emim][Ace] and [hmim][Ace] using CPA equation of state. Pet. Sci. Technol. 37 (14), 1648–1654.

Afsharpour, A., 2021. An RETM approach to model CO2 and H2S solubility in four protic ionic liquids using mSRK and CPA EoSs. J. Mol. Liq. 324, 114684.

Afsharpour, A., 2022. A new approach for correlating of H2S solubility in [emim][Lac], [bmim][ac] and [emim][pro] ionic liquids using two-parts combined models. Chin. J. Chem. Eng. 44, 521–527.

Ahmadian, S., Mohammadi, M., Ehsani, M.R., 2018. Thermodynamic modeling of methane hydrate formation in the presence of imidazolium-based ionic liquids using two-step hydrate formation theory and CPA EoS. Fluid Phase Equilib. 474, 32–42.

Alves Camêlo, L.C., de Souza Dias, G., Santos, R.L., de Souza, C.M., Faria Soares, J.F., Pereira, B., Lima, Á.S., 2020. Protic ionic liquids as constituent of aqueous two-phase system based on acetonitrile: Synthesis, phase diagrams and genipin pre-purification. Fluid Phase Equilib. 507. https://doi.org/10.1016/j.fluid.2019.112425.

Amini, Z., Fatemi, M.H., Gharaghani, S., 2016. Hybrid docking-QSAR studies of DPP-IV inhibition activities of a series of aminomethyl-piperidones. Comput. Biol. Chem. 64, 335–345.

Baramaki, Z., Arab Aboosadi, Z., Esfandiari, N., 2019. Thermodynamic modeling of ternary systems containing imidazolium-based ionic liquids and acid gases using SRK, Peng-Robinson, CPA and PC-SAFT equations of state. Pet. Sci. Technol. 37 (24), 2420–2428.

Bhattacharjee, A., Luís, A., Santos, J.H., Lopes-da-Silva, J.A., Freire, M.G., Carvalho, P.J., Coutinho, J.A.P., 2014. Thermophysical properties of sulfonium- and ammonium-

based ionic liquids. Fluid Phase Equilib. 381, 36–45. https://doi.org/10.1016/j.fluid.2014.08.005.
Billard, I., Marcou, G., Ouadi, A., Varnek, A., 2011. In silico design of new ionic liquids based on quantitative structure− property relationship models of ionic liquid viscosity. J. Phys. Chem. B 115 (1), 93–98.
Brijmohan, N., Moodley, K., Narasigadu, C., 2024. The a priori screening of potential organic solvents using artificial neural networks. Fluid Phase Equilib. 577, 113960.
Broto, P., Moreau, G., Vandycke, C., 1984. Molecular structures: Perception, autocorrelation descriptor and SAR studies. Autocorrelation Descriptor. Eur J Med Chem 19, 66–70.
ChemBioDraw, Ultra level, 12.0.2.1076 version, 2010 CambridgeSoft.
Chen, Y., Liu, X., Woodley, J.M., Kontogeorgis, G.M., 2020. Gas solubility in ionic liquids: UNIFAC-IL model extension. Ind. Eng. Chem. Res. 59 (38), 16805–16821.
Coutinho, J.A., Vlamos, P.M., Kontogeorgis, G.M., 2000. General form of the cross-energy parameter of equations of state. Ind. Eng. Chem. Res. 39 (8), 3076–3082.
Emel'yanenko, V.N., Boeck, G., Verevkin, S.P., Ludwig, R., 2014. Volatile times for the very first ionic liquid: understanding the vapor pressures and enthalpies of vaporization of ethylammonium nitrate. *Chemistry–A*. European Journal 20 (37), 11640–11645.
Frias, R.M., Follegatti-Romero, L.M., Follegatti-Romero, L.A., 2023. Modelling mixtures of ammonium and pyridinium-based ionic liquids and carbon dioxide with the Cubic Plus Association Equation of state. Fluid Phase Equilib. 574, 113910.
Geary, R.C., 1954. The contiguity ratio and statistical mapping. Incorp Statist 5, 115–145.
Gorji, A.E., Sobati, M.A., 2019. Toward molecular modeling of thiophene distribution between the ionic liquid and hydrocarbon phases: Effect of hydrocarbon structure. J. Mol. Liq. 287, 110976.
Gorji, A.E., Sobati, M.A., Alopaeus, V., Uusi-Kyyny, P., 2021. Toward solvent screening in the extractive desulfurization using ionic liquids: QSPR modeling and experimental validations. Fuel 302, 121159.
Gramatica, P., 2020. Principles of QSAR modeling: comments and suggestions from personal experience. International Journal of Quantitative Structure-Property Relationships (IJQSPR) 5 (3), 61–97.
Gramatica, P., Chirico, N., Papa, E., Cassani, S., & Kovarich, S. (2013). QSARINS: A new software for the development, analysis, and validation of QSAR MLR models.
Gramatica, P., Sangion, A., 2016. A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. J. Chem. Inf. Model. 56 (6), 1127–1131.
Hall, L.H., Kier, L.B., 1995. Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. J Chem Inf Comput Sci 35, 1039–1045.
Hamedi, N., Rahimpour, M.R., Keshavarz, P., 2020. Methane solubility in ionic liquids: Comparison of cubic-plus-association and modified Sanchez-Lacombe equation of states. Chem. Phys. Lett. 738, 136903.
Hansen, T.A., 2015. Development of a predictive version of the CPA equation of state for applications in the chemical industry. Technical University of Denmark, Denmark. Master's thesis.
Harris, K.R., 2020. Temperature and Pressure Dependence of the Viscosity of the Ionic Liquid 1-Butyl-3-methylimidazolium Acetate. J. Chem. Eng. Data 65, 804–813. https://doi.org/10.1021/acs.jced.9b01018.
Haupt, R.L., Haupt, S.E., Algorithms, P.G., Edition, S., 2004. John Wiley & Sons Inc. Hoboken, New Jersey.
He, W., Yan, F., Jia, Q., Xia, S., Wang, Q., 2017. Description of the thermal conductivity λ (T, P) of ionic liquids using the structure–property relationship method. J. Chem. Eng. Data 62 (8), 2466–2472.
Heidari, M.R., Mokhtarani, B., Seghatoleslami, N., Sharifi, A., Mirzaei, M., 2012. Liquid-liquid extraction of aromatics from their mixtures with alkanes using 1-methyl 3-octylimidazolium thiocyanate ionic liquid. J. Chem. Thermodyn. 54, 310–315. https://doi.org/10.1016/j.jct.2012.05.006.
Holland, J.H., 1975. Adaption in natural and artificial systems. The University of MichiganPress, Ann Arbor MI.
Joback, K.G., Reid, R.C., 1987. Estimation of Pure-Component Properties from Group-Contributions. Chem Eng Commun 57, 233–243. https://doi.org/10.1080/00986448708960487.
Khedri, Z., Almasi, M., Maleki, A., 2019. Thermodynamic Properties of 1-Hexyl-3-methylimidazolium Nitrate and 1-Alkanols Mixtures: PC-SAFT Model. J. Chem. Eng. Data 64, 4465–4473. https://doi.org/10.1021/acs.jced.9b00507.
Khooshechin, S., Dashtbozorgi, Z., Golmohammadi, H., Acree Jr, W.E., 2014. QSPR prediction of gas-to-ionic liquid partition coefficient of organic solutes dissolved in 1-(2-hydroxyethyl)-1-methylimidazolium tris (pentafluoroethyl) trifluorophosphate using the replacement method and support vector regression. J. Mol. Liq. 196, 43–51.
Kontogeorgis, G.M., Voutsas, E.C., Yakoumis, I.V., Tassios, D.P., 1996. An equation of state for associating fluids. Ind. Eng. Chem. Res. 35 (11), 4310–4318.
Kontogeorgis, G.M., Yakoumis, I.V., Meijer, H., Hendriks, E., Moorwood, T., 1999. Multicomponent phase equilibrium calculations for water–methanol–alkane mixtures. Fluid Phase Equilib. 158, 201–209.
Kurnia, K.A., Wilfred, C.D., Murugesan, T., 2009. Thermophysical properties of hydroxyl ammonium ionic liquids. J. Chem. Thermodyn. 41, 517–521. https://doi.org/10.1016/j.jct.2008.11.003.
Lemaoui, T., Darwish, A.S., Hammoudi, N.E.H., Abu Hatab, F., Attoui, A., Alnashef, I.M., Benguerba, Y., 2020. Prediction of electrical conductivity of deep eutectic solvents using COSMO-RS sigma profiles as molecular descriptors: a quantitative structure–property relationship study. Ind. Eng. Chem. Res. 59 (29), 13343–13354.
Liu, Q.S., Li, P.P., Welz-Biermann, U., Liu, X.X., Chen, J., 2012. Density, electrical conductivity, and dynamic viscosity of N -alkyl-4-methylpyridinium bis (trifluoromethylsulfonyl)imide. J. Chem. Eng. Data 57, 2999–3004. https://doi.org/10.1021/je3004645.
Liu, Q.S., Li, P.P., Welz-Biermann, U., Chen, J., Liu, X.X., 2013. Density, dynamic viscosity, and electrical conductivity of pyridinium-based hydrophobic ionic liquids. J. Chem. Thermodyn. 66, 88–94. https://doi.org/10.1016/j.jct.2013.06.008.
Liu, Q.S., Liu, J., Liu, X.X., Zhang, S.T., 2015. Density, dynamic viscosity, and electrical conductivity of two hydrophobic functionalized ionic liquids. J. Chem. Thermodyn. 90, 39–45. https://doi.org/10.1016/j.jct.2015.06.010.
Machida, H., Taguchi, R., Sato, Y., Smith Jr, R.L., 2009. Analysis of ionic liquid PVT behavior with a Modified Cell Model. Fluid Phase Equilib. 281 (2), 127–132.
Maia, F.M., Tsivintzelis, I., Rodriguez, O., Macedo, E.A., Kontogeorgis, G.M., 2012. Equation of state modelling of systems with ionic liquids: Literature review and application with the Cubic Plus Association (CPA) model. Fluid Phase Equilib. 332, 128–143.
Manic, M.S., Queimada, A.J., Macedo, E.A., Najdanovic-Visak, V., 2012. High-pressure solubilities of carbon dioxide in ionic liquids based on bis (trifluoromethylsulfonyl) imide and chloride. J. Supercrit. Fluids 65, 1–10.
Modarresi, H., Modarress, H., Dearden, J.C., 2007. QSPR model of Henry's law constant for a diverse set of organic chemicals based on genetic algorithm-radial basis function network approach. Chemosphere 66 (11), 2067–2076.
Mokhtarani, B., Sharifi, A., Mortaheb, H.R., Mirzaei, M., Mafi, M., F., 2009. Sadeghian, Density and viscosity of 1-butyl-3-methylimidazolium nitrate with ethanol, 1-propanol, or 1-butanol at several temperatures. J. Chem. Thermodyn. 41, 1432–1438. https://doi.org/10.1016/j.jct.2009.06.023.
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.
Moreau, G., Broto, P., 1980. Autocorrelation of molecular structures, application to SAR studies. Nouv J Chim 4, 757–764.
Moreau, G., Broto, P., 1980. The autocorrelation of a topological structure: A new molecular descriptor. Nouv J Chim 4, 359–360.
Mutelet, F., Djebouri, H., Baker, G.A., Ravula, S., Jiang, B., Tong, X., Woods, D., Acree, W.E., 2017. Study of benzyl- or cyclohexyl-functionalized ionic liquids using inverse gas chromatography. J. Mol. Liq. 242, 550–559. https://doi.org/10.1016/j.molliq.2017.07.036.
Ojha, P.K., Roy, K., 2011. Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. Chemom. Intel. Lab. Syst. 109 (2), 146–161.
Paduszynski, K., 2021. Extensive databases and group contribution QSPRs of ionic liquid properties. 3: surface tension. Ind. Eng. Chem. Res. 60 (15), 5705–5720.
Palma, A.M., Oliveira, M.B., Queimada, A.J., Coutinho, J.A., 2017. Evaluating cubic plus association equation of state predictive capacities: A study on the transferability of the hydroxyl group associative parameters. Ind. Eng. Chem. Res. 56 (24), 7086–7099.
Panah, H.S., 2017. Modeling H2S and CO2 solubility in ionic liquids using the CPA equation of state through a new approach. Fluid Phase Equilib. 437, 155–165.
Rocha, M.A., Santos, L.M., 2013. First volatility study of the 1-alkylpyridinium based ionic liquids by Knudsen effusion. Chem. Phys. Lett. 585, 59–62.
Roy, K., Ghosh, G., 2004. QSTR with Extended Topochemical Atom Indices. 2. Fish Toxicity of Substituted Benzenes. J. Chem. Inf. Comput. Sci. 44, 559–567.
Safarov, J., Guluzade, A., Hassel, E., 2019. Thermophysical Properties of 1-Butyl-3-methylimidazolium Trifluoromethanesulfonate in a Wide Range of Temperatures and Pressures. J. Chem. Eng. Data 64, 2247–2258. https://doi.org/10.1021/acs.jced.8b00837.
Sanmamed, Y.A., González-Salgado, D., Troncoso, J., Romani, L., Baylaucq, A., Boned, C., 2010. Experimental methodology for precise determination of density of RTILs as a function of temperature and pressure using vibrating tube densimeters. J. Chem. Thermodyn. 42, 553–563. https://doi.org/10.1016/j.jct.2009.11.014.
Sequeira, M.C.M., Avelino, H.M.N.T., Caetano, F.J.P., Fareleira, J.M.N.A., 2020. Viscosity measurements of 1-ethyl-3-methylimidazolium trifluoromethanesulfonate (EMIM OTf) at high pressures using the vibrating wire technique. Fluid Phase Equilib. 505. https://doi.org/10.1016/j.fluid.2019.112354.
Shahlaei, M., 2013. Descriptor selection methods in quantitative structure–activity relationship studies: a review study. Chem. Rev. 113 (10), 8093–8103.
Sharma, V., Goswami, R., Madan, A.K., 1997. Eccentric connectivity index: a novel highly discriminating topological descriptor for structure− property and structure−activity studies. J. Chem. Inf. Comput. Sci. 37 (2), 273–282.
Soave, G., 1972. Equilibrium constants from a modified Redlich-Kwong equation of state. Chem. Eng. Sci. 27 (6), 1197–1203.
Sousa, J.M., Granjo, J.F., Queimada, A.J., Ferreira, A.G., Oliveira, N.M., Fonseca, I.M., 2014. Solubility of hydrofluorocarbons in phosphonium-based ionic liquids: Experimental and modelling study. J. Chem. Thermodyn. 79, 184–191.
Sousa, J.M., Granjo, J.F., Queimada, A.J., Ferreira, A.G., Oliveira, N.M., Fonseca, I.M., 2014. Solubilities of hydrofluorocarbons in ionic liquids: Experimental and modelling study. J. Chem. Thermodyn. 73, 36–43.
Sousa, J.M., Sintra, T.E., Ferreira, A.G., Carvalho, P.J., Fonseca, I.M., 2021. Solubility of H2S in ammonium-based ionic liquids. J. Chem. Thermodyn. 154, 106336.
Stavrou, M., Bardow, A., Gross, J., 2016. Estimation of the binary interaction parameter kij of the PC-SAFT Equation of State based on pure component parameters using a QSPR method. Fluid Phase Equilib. 416, 138–149.
Taguchi, R., Machida, H., Sato, Y., Smith, R.L., 2009. High-pressure densities of 1-alkyl-3-methylimidazolium hexafluorophosphates and 1-alkyl-3-methylimidazolium tetrafluoroborates at temperatures from (313 to 473) K and at pressures up to 200 MPa. J. Chem. Eng. Data 54, 22–27. https://doi.org/10.1021/je800224k.
R. Todeschini V. Consonni Molecular descriptors for chemoinformatics 2009 Wiley VCH) (Weinheim pg 27–37.

Valderrama, J.O., Forero, L.A., Rojas, R.E., 2012. Critical properties and normal boiling temperature of ionic liquids. Update and a new consistency test. Ind. Eng. Chem. Res. 51, 7838–7844. https://doi.org/10.1021/ie202934g.

Velho, P., Liang, X., Macedo, E.A., Gómez, E., Kontogeorgis, G.M., 2021. Towards a predictive Cubic Plus Association equation of state. Fluid Phase Equilib. 540, 113045.

Wang, Y., Huang, S., Liu, X., He, M., 2023. Thermodynamic model for CO2 absorption in imidazolium-based ionic liquids using cubic plus association equation of state. J. Mol. Liq. 378, 121587.

Wang, J.Y., Zhang, X.J., Hu, Y.Q., Di Qi, G., Liang, L.Y., 2012. Properties of n-butylpyridinium nitrate ionic liquid and its binary mixtures with water. J. Chem. Thermodyn. 45, 43–47. https://doi.org/10.1016/j.jct.2011.09.003.

Yap, C.W., 2011. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32 (7), 1466–1474.

Zhao, Y.H., Abraham, M.H., Zissimos, A.M., 2003. Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. J. Org. Chem. 68 (19), 7368–7373.

Zhu, C., He, M., Liu, X., Kontogeorgis, G.M., Liang, X., 2021. Quantification of dipolar contribution and modeling of green polar fluids with the polar cubic-plus-association equation of state. ACS Sustain. Chem. Eng. 9 (22), 7602–7619.