

Czech University of Life Sciences Prague

Faculty of Environmental Sciences



**Quantitative structure-property relationship
models for gas absorption and toxicity
prediction of ionic liquids**

Xuejing Kang

Ph.D. Thesis

Thesis supervisor: Associate Prof. Dr. Zhongbing Chen

Thesis reviewers: Doc. Dr. Ing. Petr Klusoň
(Charles University)

Associate Prof. Xiaodong Liang
(Technical University of Denmark)

Assistant Prof. Kamil Paduszyński
(Warsaw University of Technology)

Czech University of Life Sciences Prague

Faculty of Environmental Sciences

Department of Applied Ecology

Prague, 2022

**Quantitative structure-property relationship
models for gas absorption and toxicity
prediction of ionic liquids**

Xuejing Kang

Thesis

This thesis is submitted in fulfillment of the requirements for the Ph.D. degree at the
Czech University of Life Sciences Prague, Faculty of Environmental Sciences.

Prague (2022)

I hereby declare that the dissertation entitled “Quantitative structure-property relationship models for gas absorption and toxicity prediction of ionic liquids” submitted for the Ph.D. degree of Philosophy of Faculty of Environmental Sciences is my original work guided by my supervisor. All sources of information, text, illustration, tables and images have been specifically cited.

Acknowledgment

I would like to express my special gratitude and appreciate to my supervisor Associate Prof. Zhongbing Chen for constantly giving me support and guiding me both intellectually and professionally throughout my doctoral study. It has been an immense privilege to be your student and have worked with you. I would like to also express my gratitude to Prof. Jan Vymazal who gave me a lot of support and help for my research projects, including financial support, measurements of experimental samples and so on.

I sincerely thank my collaborator Dr. Yongsheng Zhao from University of California Santa Barbara, who gave me valuable guidance and enable my progress during my pursuit of a doctoral degree. I am also grateful to Prof. Maohong Fan and his team members for their experimental support, valuable advice, and consultations during my internship at the University of Wyoming.

To my family, my sisters and brothers, particularly my parents, thanks for your unwavering support. You are the foundation of my endeavors and I truly appreciate you for your continuous encouragement and companions throughout. To my group members, Shanshan Hu and Bo Hu, thank you for giving me a lot of help in both my study and life during my doctoral study.

Thanks to my dissertation committee, all my colleagues in our department, and everyone who have helped me navigate the process, supported my development and made my time in Prague enjoyable.

The experimental work included in this Ph.D. thesis was funded by the following research projects:

IGA 2020B0032 (Internal Grant Agency of the Faculty of Environmental Sciences, CULS Prague)

IGA 2022B0036 (Internal Grant Agency of the Faculty of Environmental Sciences, CULS Prague)

Abstract

Ionic liquids (ILs) have gained extensive attention in both industrial and academic areas in recent three decades. Due to their remarkable properties, ILs have been applied in diverse fields, such as chemical engineering, pharmaceuticals, biotechnology, etc. ILs have been regarded as novel potential solvents and alternative media for gas separation and adsorption in environmental and ecological protection fields due to their extraordinary non-volatility. Moreover, the toxicity assessment of ILs is required before they are widely applied and released into the environment. However, because of the possible combinations of cations and anions of ILs, it is impractical to rely entirely on experimental methods to acquire the property data for all ILs. Exploring efficient computer-aided methods for property prediction of ILs and screening suitable ILs for a specific target is desired and significant. Quantitative structure-property relationship (QSPR) models are commonly considered as a promising way to diminish and substitute experimental processes involving measurements for the properties of compounds.

This dissertation developed several novel methods based on the QSPR methodology to achieve better model performance to predict gas absorption and toxicity of ILs. The main research contents and results of this thesis include: (1) a dataset with 502 data points of ammonia solubility in 17 ILs and another dataset consisting of 122 Henry's law constant (HLC) data of hydrogen sulfide (H_2S) for 22 ILs were gathered; the relationships between the electrostatic potential surface areas (S_{EP}) and the gas absorption capability of ILs were studied; the extreme learning machine (ELM) method was applied to develop predictive models based on the S_{EP} descriptors of ILs. Results show that the proposed models were efficient and reliable, and can be potentially used for screening reasonable ILs to absorb NH_3/H_2S from chemical industry processes. (2) A QSPR model was developed to assess the ecotoxicity of ILs towards *Vibrio fischeri* using the ELM method; A comprehensive toxicity dataset for *Vibrio fischeri* involving 142 ILs were collected, while the parameters of cations and anions in ILs within various S_{EP} ranges were calculated. The developed model with the R^2 (0.9297) and the root-mean-square error (RMSE=0.3262) exhibited good reliability and predictive power for predicting ILs ecotoxicity on *Vibrio fischeri*. (3) Two new types of group/fragment contribution methods, atom surface fragment contribution (ASFC) method and non-integer group contribution (NGC) method, were first proposed to build predictive models for the ecotoxicity of ILs towards *leukemia rat cell*

line (IPC-81); A dataset of 140 ILs toxicity towards IPC-81 was built; the group parameters used for modeling were based on the sigma surface area ($S_{\sigma\text{-surface}}$) or the surface areas of electrostatic potential (S_{EP}) of atoms in ILs. The developed ASFC/NGC models showed better performance than the traditional group contribution method. Thus, the proposed techniques have novelty and superiority in assessing the properties of ionic liquids and other compounds. What's more, they can be applied in the screening and designing of green and functional ILs for special applications in various fields.

Abstrakt

Iontové kapaliny (IL) získaly v posledních třech desetiletích rozsáhlou pozornost v průmyslové i akademické oblasti. Díky svým pozoruhodným vlastnostem byly IL aplikovány v různých oblastech, jako je chemické inženýrství, farmacie, biotechnologie atd. IL byly považovány za nová potenciální rozpouštědla a alternativní média pro separaci a adsorpci plynů v oblastech ochrany životního prostředí a ekologie kvůli jejich mimořádné nevolatilitě. Kromě toho je nutné posouzení toxicity IL před tím, než jsou široce používány a uvolňovány do životního prostředí. Avšak kvůli možným kombinacím kationtů a aniontů IL je nepraktické spoléhat se výhradně na experimentální metody pro získání údajů o vlastnostech pro všechny IL. Prozkoumání účinných počítačově podporovaných metod pro predikci vlastností IL a screening vhodných IL pro specifický cíl je žádoucí a významné. Modely kvantitativního vztahu mezi strukturou a vlastnostmi (QSPR) jsou běžně považovány za slibný způsob, jak omezit a nahradit experimentální procesy zahrnující měření vlastností sloučenin.

Tato disertační práce vyvinula nové servalové metody založené na metodologii QSPR k dosažení lepší výkonnosti modelu pro předpovídání absorpce plynů a toxicity IL. Hlavní obsah výzkumu a výsledky této práce zahrnují: (1) byl shromážděn datový soubor s 502 datovými body rozpustnosti amoniaku v 17 IL a další datový soubor sestávající z 122 dat Henryho zákona (HLC) o sirovodíku (H_2S) pro 22 IL; byly studovány vztahy mezi povrchovými plochami elektrostatického potenciálu (S_{EP}) a schopností IL absorbovat plyn; metoda extrémního učícího se stroje (ELM) byla použita k vývoji prediktivních modelů založených na deskriptorech S_{EP} IL. Výsledky ukazují, že navržené modely byly účinné a spolehlivé a mohou být potenciálně použity pro screening rozumných IL pro absorpci NH_3/H_2S z procesů chemického průmyslu. (2) K posouzení ekotoxicity IL vůči *Vibrio fischeri* byl vyvinut model QSPR pomocí metody ELM; Byl shromážděn komplexní soubor údajů o toxicitě pro *Vibrio fischeri* zahrnující 142 IL, zatímco byly vypočteny parametry kationtů a aniontů v IL v různých rozmezech S_{EP} . Vyvinutý model s R^2 (0,9297) a střední hodnotou kvadratické chyby (RMSE=0,3262) vykazoval dobrou spolehlivost a prediktivní schopnost pro predikci ekotoxicity ILs na *Vibrio fischeri*. (3) Nejprve byly navrženy dva nové typy metod skupinového/fragmentového příspěvku, metoda příspěvku atomového povrchu (ASFC) a metoda neceločíselného příspěvku skupiny (NGC), aby se vytvořily prediktivní modely pro ekotoxicitu IL vůči leukemické krysí buněčné linii (IPC-81); Byl vytvořen soubor dat toxicity 140 IL vůči IPC-81; parametry skupiny použité pro modelování

byly založeny na sigma povrchové ploše ($S_{\sigma\text{-povrch}}$) nebo povrchové ploše elektrostatického potenciálu (S_{EP}) atomů v IL. Vyvinuté modely ASFC/NGC vykazovaly lepší výkon než tradiční metoda skupinového příspěvku. Navrhované techniky mají tedy novost a nadřazenost při posuzování vlastností iontových kapalin a dalších sloučenin. A co víc, mohou být použity při screeningu a navrhování zelených a funkčních IL pro speciální aplikace v různých oblastech.

Table of contents

Chapter I	Introduction	1
Chapter II	State of the art	5
Chapter III	Prediction of ammonia absorption in ionic liquids based on extreme learning machine modeling and a novel molecular descriptor S_{EP}	33
Chapter IV	A QSPR model for estimating Henry's law constant of H ₂ S in ionic liquids by ELM algorithm	55
Chapter V	Assessing the ecotoxicity of ionic liquids on <i>Vibrio fischeri</i> using electrostatic potential descriptors	77
Chapter VI	Atom surface fragment contribution method for predicting the toxicity of ionic liquids	99
Chapter VII	Application of atomic electrostatic potential descriptors for predicting the eco-toxicity of ionic liquids towards <i>leukemia rat cell line</i>	121
Chapter VIII	Discussion and summary	143
References		161
Curriculum vitae and List of Publications		203

Chapter I

Introduction

1. Introduction

Since ionic liquids (ILs) are categorized as a class of environmentally-friendly fluids that comply with regional and international environmental regulations, the interest in ILs has been booming in both academia and industry in recent two decades. Ionic liquids (ILs) consist of organic cations (such as pyrrolidinium, imidazolium, and pyridinium and ammonium) and organic/inorganic anions (e.g., Chloride, bromide, tetrafluoroborate, and hexafluorophosphate). Because of their extraordinary properties (e.g., low vapor pressure, high chemical stability, wide liquid temperature range, high ionic conductivity as well as adjustable structures (Blanchard et al., 1999; Isoaari et al., 2019; Ranke et al., 2007b; Tshibangu et al., 2011), multidisciplinary studies have emerged, including material sciences, chemical engineering, and environmental sciences.

ILs have been considered as appealing solvents for gas capture (Gomes and Husson, 2009). The solubility data of some gases in ILs have been experimentally measured and reported, such as CO₂ (Zeng et al., 2017), H₂S (Jalili et al., 2009), SO₂, NH₃, CO (Lei et al., 2014a) and so on. However, state-of-the-art research shows that the previously accepted advantage of low toxicity for ILs has proved to be overestimated, meaning that ILs pose a hazardous potential to humans and the environment to some extent. This is because they usually are soluble in water and can be released into soil or aquatic ecosystem (Ghanem et al., 2017; McFarlane et al., 2005; Yan et al., 2019). For example, some living organisms, such as *Green algae* (Latała et al., 2009; Pretti et al., 2009), *Caenorhabditis elegans* (Peng et al., 2018), *Vibrio fischeri* (Jafari et al., 2019; Stolte et al., 2007b; Ventura et al., 2012), *leukemia rat cell line* (Ranke et al., 2007b), can be affected by the ILs. To screen and design green and functional ILs, the toxicity assessment of ILs is required.

Due to the tremendous number of ILs, the time-consuming and hazardous experimental operations, some studies have investigated and reported a number of predictive models have been established for calculating the various properties of ILs by applying different algorithms (Belvèze, 2004; Jastorff et al., 2007; Katritzky et al., 2002; Palomar et al., 2007; Torrecilla et al., 2008a; Yan et al., 2008). Nevertheless, accurate evaluation for the features of ILs is complex and still desired for their further applications. Quantitative structure–property relationship (QSPR) models have been extensively employed for the estimation of the physicochemical characteristics of compounds.

This thesis focuses on the applications of ILs for gas absorption and the environmental risk assessment of ILs. Therefore, the objectives of this thesis include: (1) to collect NH_3 solubility data and henry's law constant data of H_2S in various ILs and develop efficient study QSPR models for evaluating the gas absorption ability of ILs based on different algorithms and quantum chemical descriptors; (2) to build datasets for toxicity of ILs towards *Vibrio fischeri* and *leukemia rat cell line* (IPC-81) and to develop predictive models with reliability and stability. (3) to study quantitative structure-property relationships of ILs, particularly the toxicity of ILs, and to propose new QSPR methods for property prediction of ILs based on the quantum chemical molecular parameters of the sigma surface areas ($S_{\sigma\text{-surface}}$) and electrostatic potential surface area (S_{EP}) of ILs.

Chapter II

State of the art

Contents

2.1 Ionic liquids	7
2.2 Applications of ionic liquids in gas absorption	11
2.2.1 <i>Ionic liquids for acid gases absorption</i>	<i>11</i>
2.2.2 <i>Ionic liquids for volatile organic compounds (VOCs).....</i>	<i>12</i>
2.3 Environmental fate and toxicity of ionic liquids	13
2.3.1 <i>Environmental fate of ionic liquids</i>	<i>13</i>
2.3.2 <i>Toxicological aspect of ionic liquids</i>	<i>15</i>
2.4 QSPR models for the property prediction of ionic liquids.....	25
2.4.1 <i>Description of QSPR method.....</i>	<i>25</i>
2.4.2 <i>QSPR models for predicting gas absorption in ionic liquids</i>	<i>26</i>
2.4.3 <i>QSPR models for predicting the toxicity of ionic liquids</i>	<i>27</i>

2.1 Ionic liquids

History

Molten salts comprised completely of cations and anions with the melting point below 100 °C are generally defined as ionic liquids (ILs) in fields of industry and academia. Paul Walden (1914) discovered the first IL, ethylammonium nitrate ($[\text{EtNH}_3][\text{NO}_3]$), which is a colorless and odorless liquid with low melting point (12 °C), negligible vapor pressure, and good conductivity (Nasirpour et al., 2020; Zandu et al., 2019). Later on, Hurley and Weir (1951) synthesized another IL by mixing 1-alkylpyridinium halides with metal halides, for instance, 1-ethypyridinium chloride and aluminum chloride (AlCl_3) in 1951. The chemical stabilities of these ILs, however, are not very good. Since Wilkes and Zaworotko (1992) produced a series of 1-ethyl-3-methylimidazolium based upon ILs that are air- and water-stable, the research studies related to ILs had a dramatic increase and began to become a significant scientific field (Dupont, 2004; Mallakpour and Dinari, 2012; Welton, 2018). The number of SCI publications about ILs has exponentially increased since the early 1990s, as shown by a search in Web of Science with the key word of “ionic liquids” (Figure 2.1), indicating that more and more scientists have been engaged in this burgeoning subject with plentiful achievements.

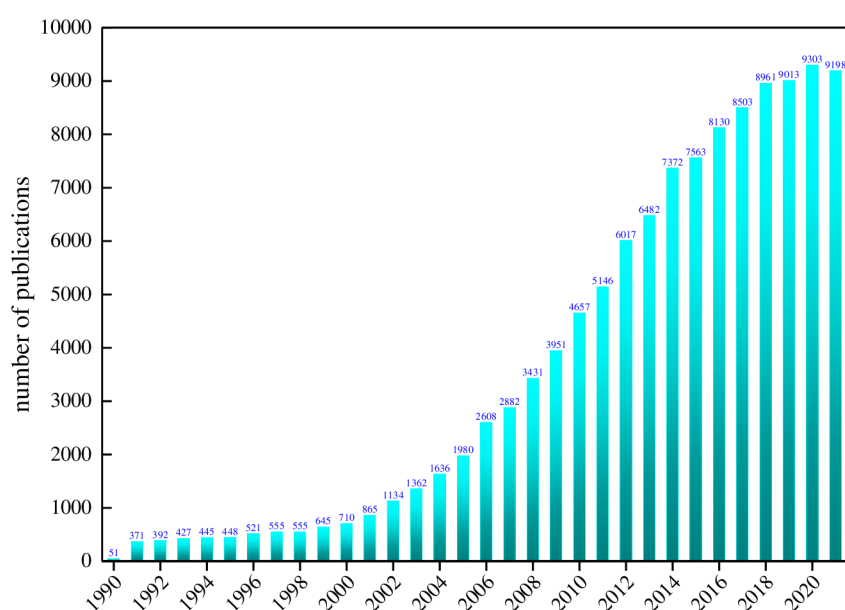
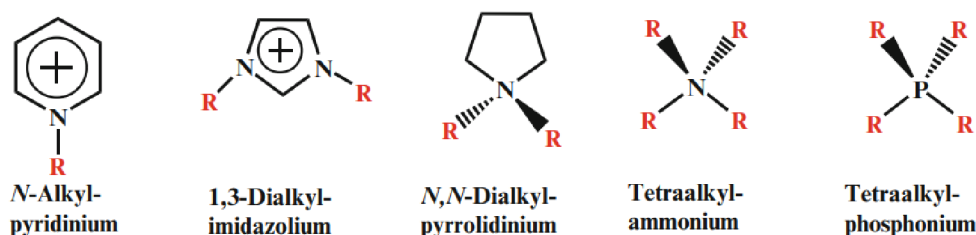


Figure 2.1 Publications about ionic liquids from Web of Science

Special properties

Typical cations and anions are presented in Figure 2.2. The unique properties of ILs include: (a) low melting point ($<100^{\circ}\text{C}$); (b) exceptionally low volatility; (c) nonflammability; (d) thermal and chemical stability; (e) high heat capacity; (f) high thermal conductivity; (g) strong ionic conductivity; (h) wide electrochemical potential window (Chiappe and Pieraccini, 2005; Laali, 2003; Lei et al., 2014b). ILs are normally known as “green solvent” and the alternatives of traditional organic solvents, particularly because of the crucial attribute of non-measurable vapor pressure, (Cho et al., 2019; Cull et al., 2000). Moreover, the combinations or the structural variability of ILs are enormous as the structures of cations and anions of ILs can differ independently. It has been investigated that possibly 10^6 pure ILs can be synthesized and approximately 10^{18} binary and ternary IL systems may exist (Castner and Wishart, 2010). Their physicochemical properties can be customized through tailoring cations and anions for desired applications, and hence ILs are also considered as “designer solvents” (Yang and Pan, 2005; Zhao, 2006).

Most Commonly Used Cations:



R: Methyl, Ethyl, Butyl, Hexyl, Octyl, Decyl

Most Commonly Used Anions:

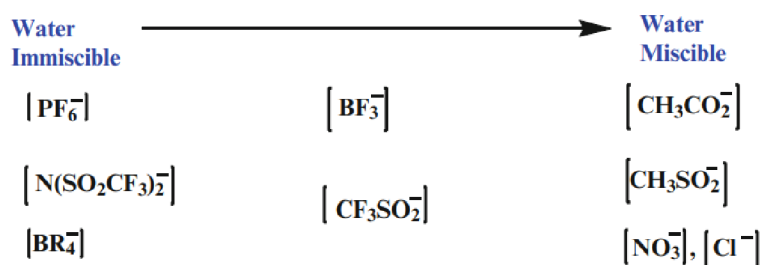


Figure 2.2 Typical structures of ionic liquids (Mallakpour and Dinari, 2012)

Applications of ionic liquids

Owing to their exceptional physicochemical characteristics described above, ILs have been extensively utilized in diverse industrial and research areas (Figure 2.3), involving chemistry (Greer et al., 2020; Plechkova and Seddon, 2008; Ratti, 2014), chemistry engineering (Gutowski, 2018; Nasirpour et al., 2020), electrochemistry (Doherty et al., 2012; Tiago et al., 2020), biotechnologies (Egorova et al., 2017; Koo and Ha, 2008; Roosen et al., 2008), pharmaceuticals (Bhupinder Singh, 2010; Egorova et al., 2017), coating (Saljooqi, 2021; Thuy Pham et al., 2010) and so on. According to Gordon (2001) that the improved process economics, reactivity, selectivity and yield make it distinct advantages to performing lots of reactions in ILs. The potential applications of ILs for the chemical industry have been gradually discovered and recognized, while the first industrial program related to ILs was proclaimed in 2003 (Tang et al., 2012). Conventional organic solvents applied in a variety of chemical processes can be substituted by ILs because ILs have superior efficiency and lower toxicity. Table 2.1 summarized the detailed comparison of ILs with conventional organic solvents.

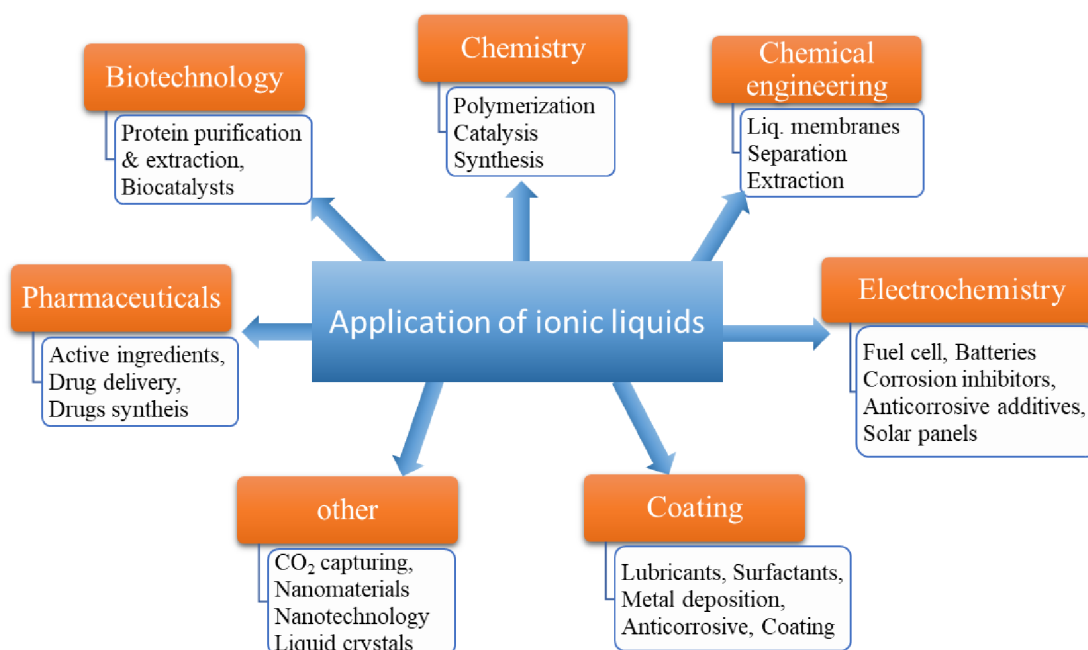


Figure 2.3 Applications of ionic liquids in various fields (Saljooqi, 2021)

Table 2.1 Comparison of properties between organic solvents and ionic liquids (Greer et al., 2020; Nasirpour et al., 2020)

Property	Organic Solvents	Ionic liquids
Number of solvents	> 1,000	> 1,000,000
Viscosity (Pa·s)	$2 \times 10^{-4} \sim 0.1$	$1.4 \times 10^{-2} \sim 40$
Density ($\text{Kg} \cdot \text{m}^{-3}$)	$6 \times 10^{-4} \sim 1.7 \times 10^{-3}$	$8 \times 10^{-4} \sim 3.3 \times 10^{-3}$
Refractive index	1.3 ~ 1.6	1.3 ~ 2.2
Electrical conductivity ($\text{mS} \cdot \text{cm}^{-1}$)	Usually insulator	Up to 120
Thermal conductivity ($\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$)	0.1 ~ 0.6	0.1 ~ 0.3
Applicability	Single function	Multifunction
Catalytic ability	Rare	Common and tunable
Chirality	Rare	Common and tunable
Vapor pressure	Obeys the Clausius-Clapeyron equation	Negligible vapor pressure under normal conditions
Flammability	Usually flammable	Usually nonflammable
Polarity	Conventional polarity concepts apply	Polarity concept questionable
Cost	Normally cheap	Expensive

As “designer solvents”, ILs have the capacity to carry out special tasks for a specific application. Due to their tunable hydrophobicity and solubility, ILs have been applied for extracting heavy metals, organic and biomolecules, as well as fuel desulfurization and CO₂ captures (Zhao, 2006). The nature of their interactions with biological systems and their antibacterial properties have contributed to advanced research into their utilization in different biocatalytic processes, drug delivery and drug synthesis. Their surface-active property along with other significant features of ILs led to their effective utilization instead of traditional surfactants in various areas, e.g., lubricants, paints, detergents, enhanced oil recovery (Rodríguez-Escontrela et al., 2016). Besides, ILs are easily recovered by means of distillation or extraction, which helps minimize their cost to increase their utilization in various industrial processes.

2.2 Applications of ionic liquids in gas absorption

2.2.1 Ionic liquids for acid gases absorption

ILs has emerged as a novel class of green materials for gas absorption and separations. Because of their desirable characteristics, ILs have the possibility to avoid the issues of secondary contaminates and high energy consumption caused by conventional (Lei et al., 2014b).

Khodadoust et al. (2006) utilized the ILs with 1-butylimidazole and 3-bromopropylamine hydrobromide as raw materials for reactions, executed the post-treatment and exchanges of anions to obtain room-temperature ILs with amine groups. The synthesized ILs were then used to absorb CO₂. This new type of ILs can be used repeatedly, with the advantages of non-volatility, no aquatic solvents needed, and so on. Wu et al. (2004) explored the adsorption of gaseous SO₂ in a simulated flue by using the ILs of 1,1,3,3-tetramethylguanidine lactate ([TMG][L]). It can be concluded that the removal of SO₂ mainly depended on chemical reactions under low pressure, whereas both chemical and physical reactions worked when the pressure was high. Compared with traditional methods, ILs have the advantages of fast absorption, high capture efficiency, good stability, and being recycled multiple times.

Under simulated atmospheric conditions with the temperature of 25°C, the humidity of 50%, as well as the concentration of gas pollutants of less than 10 mg/m³, the SO₂ and NO₂ removal ability of ILs, 1-ethyl-3-methylimidazole acetate ([C₂min][Ac]) was investigated by Li et al. (2018). The results show that the adsorption efficiency of SO₂ on the activated carbon supported by [C₂min][Ac] is better than that of pure activated carbon and KOH-carried activated carbon. ILs as adsorbents presented higher selectivity to SO₂ than to NO₂. Compared with the imidazolium cations, the acid gas pollutants preferentially and strongly interacted with the oxygen atom in the acetate anion. Zhou et al. (2016) reported the absorption capacity of seven ILs for H₂S, and studied their absorption mechanism based on the method of density functional theory (DFT). Consequently, ILs presented great efficiency of H₂S absorption. Besides, the H₂S absorption is a physical process, and the hydrogen bonding is the driving force for the absorption process.

Huang et al. (2017) synthesized a series of phenolic ILs with various cations and studied their adsorption capabilities for H₂S and CO₂. It proved that H₂S showed higher solubility than CO₂ as phenolic ILs had strong interactions with H₂S, while the

solubility of CO₂ decreased significantly as the increase of the number of cationic hydrogen bonds. Among them, tetramethylguanidine phenate ([TMGH][PhO]) showed the highest solubility for H₂S under the concentration of 0.85 mol/mol at 40°C and 0.1 MPa, which was higher than the solubility of other absorbents reported in the literature. The [TMGH][PhO] can be directly produced by the neutralization reaction of 1,1,3,3-tetramethylguanidine and phenol, causing an advantage in cost.

2.2.2 Ionic liquids for volatile organic compounds (VOCs)

In recent years, the applications of ILs based on imidazolium and ammonium for removing VOCs from the atmosphere has been explored.

Wang et al. (2017) studied a kind of imidazolium-based ILs, 1-butyl-3-methylimidazolium bis-(trifluoromethylsulfonyl)imide ([Bmim][NTf₂]) as absorbents to remove VOCs, with toluene as VOC pollutant samples to evaluate the absorption ability of [Bmim][NTf₂] to VOCs. With the conditions of 20 °C and 1 atm, the absorption efficiency of [Bmim][NTf₂] to toluene was reached to 61.5%. When concentration toluene was 300 ppm, the absorption rate (98.3%) of [Bmim][NTf₂] to toluene was the highest. Under the temperature of 20 °C and the flow rate of 50 mL/min, the absorption rate of the IL was over 94%. The results of the cycled absorption experiments show that the ILs could be recycled for at least 5 times, and the absorption efficiency was almost unchanged, which helps to reduce the cost of ILs in practice. Milota et al. (2007) developed a laboratory-scale system for absorption, using ammonium-based ILs (dodecyl(trihexyl)phosphonium dicyanamide) as absorbents to detect the methanol and α -pinene existing in VOCs at room temperature. Their results suggested that methanol, α -pinene and formaldehyde in VOCs were successfully removed with the totally removal rate of 73%-78%.

In summary, ILs for the absorption of gases in flues are mainly applied for the acid gases of CO₂, SO₂, NO₂, H₂S, and the organic substances in VOCs, e.g., toluene, alkenes, which have the advantages of high efficiency absorption, low cost and being easily recycled (Gomes and Husson, 2009). In terms of acid gas absorption, imidazolium- and guanidine-based ILs exhibit better absorption efficiency, which is possibly related to the interaction between ILs' structures and acid gases. The process of gas absorption in ILs is mostly manifested as a physical absorption process, hydrogen bonding as the driving force, and sometimes as a physical and chemical absorption process (Gomes and Husson, 2009).

2.3 Environmental fate and toxicity of ionic liquids

2.3.1 Environmental fate of ionic liquids

The increasing and extensive applications of ILs will negatively impact the environment and pose threat to the human health because of their inevitable discharge into the aquatic and terrestrial environments (Thamke et al., 2017). Learning the environmental fate of ILs, including their behaviors and fate in environment become more and more important (Figure 2.4).

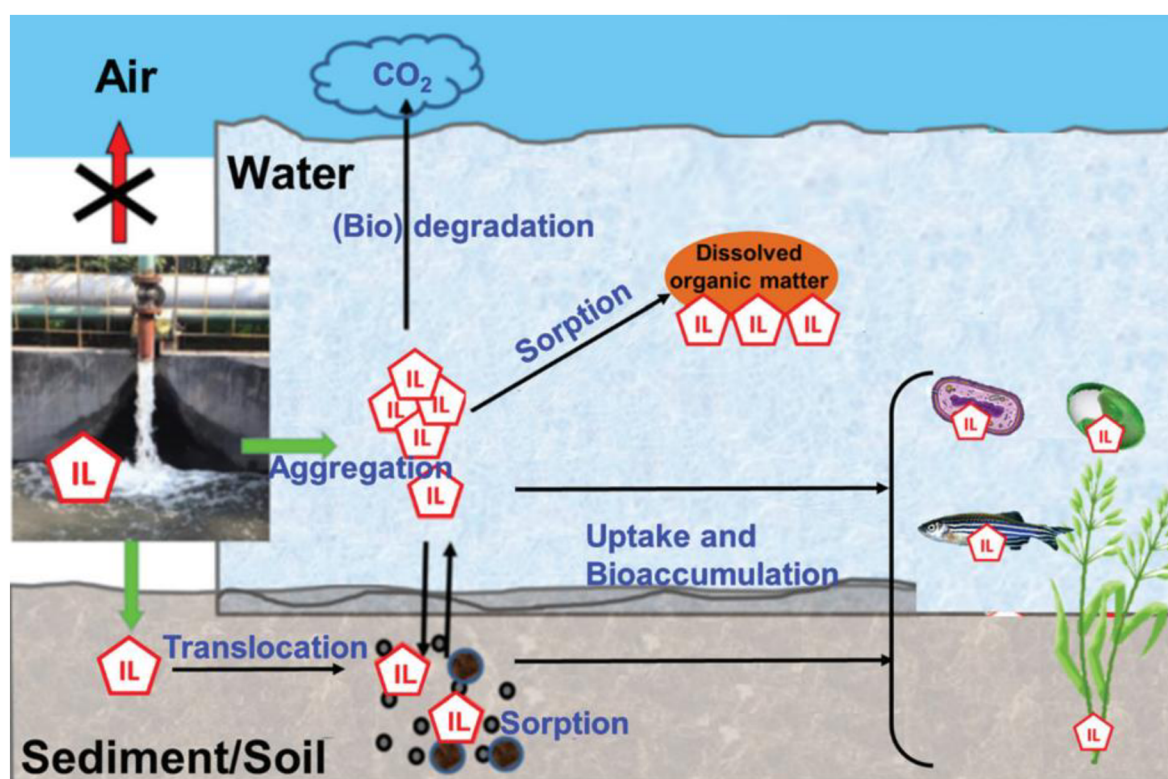


Figure 2.4 Transport, transformation, bio(degradation) and bioaccumulation of ionic liquids in the environment (Amde et al., 2015)

Chemical degradation of ionic liquids

The chemical methods for the degradation of ILs mainly include redox reactions and electrolysis. In particular, the oxidation processes, e.g., Fenton oxidation (101), UV/H₂O₂ oxidation (Czerwicka et al., 2009), ozonation (Pernak and Branicka, 2004)

and ultrasonic treatment method (Zhou et al., 2013), have been introduced for the degradation of ILs.

It has been demonstrated that the combinations of UV radiation and catalytic oxidants could improve the photodegradation efficiency of ILs (Stepnowski and Zaleska, 2005). The alkyl chain length has effects on the photodegradation of ILs. ILs with longer alkyl chains could be more easily degraded (Morawski et al., 2005). In terms of Fenton oxidation, studies showed that the degradation efficiency of several imidazolium-based and pyridinium-based ILs was in the range of 68% and 97% in a $\text{H}_2\text{O}_2/\text{Fe}^{3+}$ system (Siedlecka et al., 2008; Siedlecka and Stepnowski, 2009). It was also observed that ILs with imidazolium were more susceptible to be degraded than those with pyridinium. Ultrasonic irradiation can generate reaction radicals, which can result in the oxidation and decomposition of ILs. For example, almost 99% of imidazolium ILs were degraded within 72 hours in an ultrasonic/ H_2O_2 / CH_3COOH system (Li et al., 2007).

Biodegradation of ionic liquids

Compared to chemical degradation, biodegradation is a microbial process of chemicals breakdown, which is more economical and environmentally friendly (Jordan and Gathergood, 2015). The drawback is that biodegradation is less effective than chemical degradation currently. Therefore, the combinations of chemical and biological treatments have been introduced (Arellano et al., 2020; Gomez-Herrero et al., 2019). The chemical methods, such as Fenton-like oxidation may be used as pretreatment and then the generated intermediates could be degraded by microorganisms to improve the ability of IL biodegradation (Arellano et al., 2020; Gomez-Herrero et al., 2019).

The effects of IL structures on their biodegradability have been explored (Costa et al., 2017). Regarding cations, the extension of alkyl chains has a positive correlation with the toxicity and hydrophobicity of ILs. The function groups (e.g., ester, aldehyde, hydroxyl and carboxylic) with a hydrolytic site attacked by enzymes are beneficial for the improvements of biodegradability for imidazolium cations (Gathergood et al., 2004; Jordan and Gathergood, 2015). For the effects of anions, in contrast to inorganic anions, alkyl sulfates anions presented the best efficiency of biodegradation as they could provide extra carbon sources for microorganisms (Garcia et al., 2005). Additionally, the longer alkyl chains in sulfate anions exhibited higher biodegradability (Garcia et al., 2005). It has been found that the ILs were readily biodegradable if they come from

natural organic compounds, such as sugars, choline, amino acids, alcohols and so on (Gomes et al., 2019).

Adsorption of ionic liquids

The behavior of ILs in soil is significant because ILs may enter into soils and cause contaminations to the earth environment. Some publications have explored the adsorption mechanisms of ILs in soil and sediments, which strongly affected their retention and mobility (Beaulieu et al., 2008; Li et al., 2014; Matzke et al., 2009; Mrozik et al., 2012, 2008; Studzińska et al., 2009).

It has been suggested that the ionic interaction has influences on the adsorption of ILs in soil (Matzke et al., 2009). Stepnowski (2005) revealed that the adsorption of ILs with imidazolium was linked with their electrostatic interactions. In addition, this study also disclosed that the marine sediments with fine texture strongly and irreversibly absorbed ILs, whereas organic carbon-rich soil exhibited weak and reversible adsorption of ILs, demonstrating the significance of mineral elements in soil or sediments for the adsorption mechanism. Moreover, authors also proposed that longer alkyl chains in ILs led to irreversible bound between the soil and ILs, which was consistent with Matzke et al. (2009) and Stepnowski et al. (2007). It was found that ILs with hydrophobic chains had stronger ability of adsorption in soils and sediments than hydrophilic ILs (Mrozik et al., 2012, 2008). Conclusively, the adsorption ability of ILs is decided by their physical and chemical properties, which are highly related to their chemical structures.

2.3.2 Toxicological aspect of ionic liquids

As ILs can impede the emission of volatile organic compounds (VOCs) into the atmosphere to reduce the air pollution risks, ILs are considered as potential green substitutes which may take the place of conventional organic solvents in various chemical processes and technologies. However, they have non-negligible solubility in aqueous solutions and may reach surface water, groundwater, and soils by accidental or inevitable discharge. ILs are not inherently green, and some are quite toxic, leading to adverse effects on the living organisms and the environment (Thuy Pham et al., 2010). Because ILs' applications are expected to increase drastically, more research efforts are desired to recognize and mitigate the threats of ILs to human health and the environmental system (Pešić et al., 2020). Recently, comprehensive studies of ILs' environmental impact including toxicity and degradability of ILs have been studied

from 2010 (Bubalo et al., 2017). The current literature has reported many studies on the biological influences of ILs based on the evaluation systems of the toxicological test. For example, Thuy Pham et al. (2010) reviewed the ILs toxicities towards several systems with varying levels of biological complexity and to different environmental compartments (Figure 2.5).

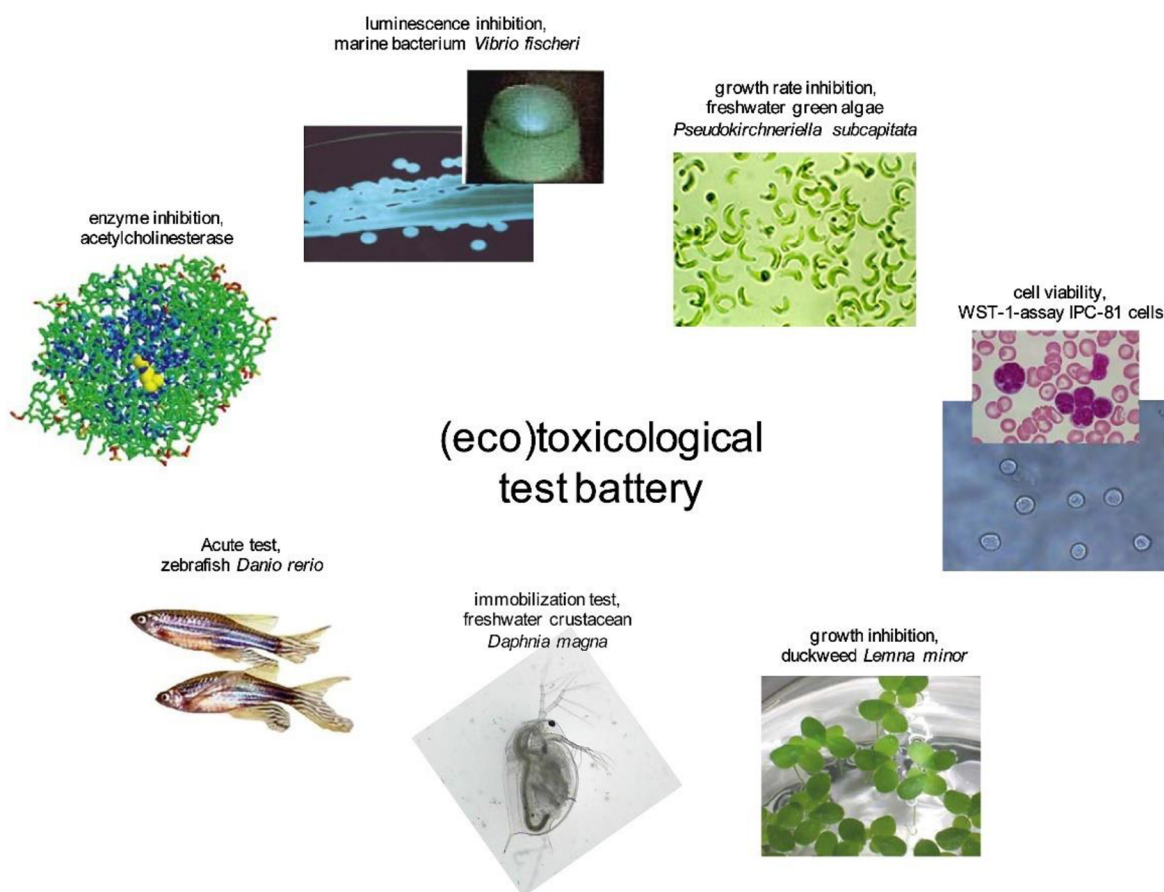


Figure 2.5 The flexible (eco)toxicological test battery considering aquatic and terrestrial compartments and different trophic levels (Matzke et al., 2007; Thuy Pham et al., 2010)

Toxic effects of ILs on enzyme activity

Due to multiple possibilities of contacts between enzymes and ILs in various industrial processes and the release of ILs may inhibit the activity of enzymes in organisms, the activity tests of enzymes have been applied for the toxicity assessment of ILs. Some researchers investigated the characteristics of the enzymes playing crucial roles for the balance of the environment (Cho et al., 2019; Fulton and Key, 2001).

A technique applying acetylcholinesterase (AChE) to investigate the effects from the alkyl side chain of ILs based imidazolium and pyridinium was first described by Stock et al. (2004). Their results demonstrated that the inhibitory effects on AChE strengthened with the length of alkyl side chains. Besides, ILs based on trihexyltetradecylphosphonium were determined to have the strongest toxicity (Stock et al., 2004). Studies performed by Arning et al. (2008a) proved that the extension of the alkyl chain negatively impacted the toxicity of ILs on the enzyme activity and they explained that the cationic moiety was the domain factor for the IL toxicity and no clear effects from anions were found.

A comprehensive dataset containing the inhibition values of various ILs and related salts towards AChE has been reported by Ranke and his co-workers (Ranke et al., 2007b). Among these data, it was discovered that all the detected inhibitory effects of ILs on the AChE could be attributed to the cations (Arning et al., 2008b). Particularly, pyridinium-based ILs exhibited stronger inhibition than imidazolium-based ILs, while the phosphonium-based ILs showed less inhibitory effects. ILs with the fluoride-based anions had inhibitory effects on AChE because these species are susceptible to hydrolysis. Nevertheless, all the other anion species imposed no effect towards the enzyme activity which could be accounted for the limited interactions between ILs and the active side of AChE (Matzke et al., 2007). [Table 2.2](#) summarized studies about ILs toxicity on AChE and other enzymes.

To design more benign and green structures of ILs, the introduction of less toxic ILs, ILs with biocompatible anions e.g., acesulfamate and saccharinates (Stasiewicz et al., 2008), and ILs with biocompatible cations, e.g., cholinium (Lai et al., 2011) and multicharged cations (Steudte et al., 2014) can be considered.

Table 2.2 Studied about ILs on enzyme inhibition (Cho et al., 2019)

No.	Enzymes	Cations	Side chains	Anions	Reference
1	AChE	Imidazolium, pyridinium, phosphonium	Alkyl, aryl	[BF ₄], Br, Cl, [CF ₃ SO ₃], [PF ₆], [8OSO ₃], [N(CN) ₂], [NO ₃], [(244Me ₃ Pen) ₂ PO ₂]	(Stock et al., 2004)
2	AChE	Imidazolium, pyridinium	Alkyl, ether	[BF ₄]	(Jastorff et al., 2005)
3	AChE	10 head groups	Alkyl, aryl, ether, cyano, hydroxyl, carboxyl, ester, sulfonyl	37 anions	(“The UFT/ Merck Ionic Liquids Biological Effects Database,” 2020)
4	AChE	Imidazolium	Alkyl	[(2-OPhO) ₂ B], [(CF ₃ SO ₂) ₂ N], [8OSO ₃], [(CF ₃) ₂ N], Cl	(Matzke et al., 2007)
5	AChE	Imidazolium, pyridinium, phosphonium, piperidinium, pyrrolidinium, quinolinium, morpholinium, ammonium, phosphonium	Alkyl, aryl, ether, cyano, hydroxyl, ketone	[BF ₄], Br, Cl, I	(Arning et al., 2008b)
6	AChE	Pyridinium	Alkyl, hydroxyl, alkoxy	Cl, saccharinate, acesulfamate	(Stasiewicz et al., 2008)
7		Imidazolium	Alkyl	Br	(Yu et al., 2008)
8	Chloroperoxidase	Imidazolium, pyridinium	Alkyl	[BF ₄], Br, [1OSO ₃], [PF ₆], [CF ₃ SO ₃]	(Boškin et al., 2009)
9	AChE cellulase	Imidazolium	Alkyl	, Br	(Luo et al., 2009)
10	AChE	Ammonium	Alkyl, aryl	N,N,N-trialkylammoniododecabotates	(Schaffran et al., 2009)

No.	Enzymes	Cations	Side chains	Anions	Reference
11	Penicillium expansum lipase, mushroom tyrosinase	Imidazolium, ammonium, cholinium	Alkyl	[1COO], [1OSO ₃], [NH ₃], [H ₂ PO ₄]	(Lai et al., 2011)
12	Catalase	Imidazolium, phosphonium, pyrrolidinium	Alkyl	[BF ₄], Cl, [CF ₃ SO ₃], [1OSO ₃]	(Pinto et al., 2011)
13	AChE	Imidazolium	Alkyl	[B(CN) ₄], [C(CN) ₃], [N(CN) ₂], [(CF ₃ SO ₂) ₂ N], [(C ₂ F ₅) ₃ PF ₃) ₂ N], Cl	(Steutde et al., 2012)
14	AChE	Imidazolium, cholinium	Alkyl, hydroxyalkyl	[BF ₄], Cl, 18 amino acids	(Hou et al., 2013)
15	AChE	Imidazolium, pyridinium, ammonium	Hydroxyalkyl	Alkanoate	(Peric et al., 2013)
16	AChE	Imidazolium	Alkyl, aryl, ester	I, Br	(Stolte et al., 2013)
17	Human carboxylase	Imidazolium, pyridinium, phosphonium, benzethonium	Alkyl	[(CF ₃ SO ₂) ₂ N], Cl, bistriflimide, salicylate, docusate	(Costa et al., 2014a)
18	AChE	Imidazolium-based dications	Alkyl, pentacyclo	I, Br, Cl, [1OSO ₃]	(Steutde et al., 2014)
19	Luciferase	Imidazolium, pyridinium	Alkyl	[BF ₄], Br, Cl, [(CF ₃ SO ₂) ₂ N], [CF ₃ SO ₃]	(Ge et al., 2014)
20	Glutathione reductase	Imidazolium, phosphonium, pyrrolidinium	Alkyl	[BF ₄], Cl, [CF ₃ SO ₃], [1OSO ₃]	(Cunha et al., 2015)
21	Trypsin	Imidazolium	Alkyl	[BF ₄], Br, Cl, [1COO], [CF ₃ SO ₃], [NH ₃]	(Fan et al., 2016)
22	Lactic dehydrogenase	Imidazolium	Alkyl	[BF ₄], Br, Cl, [1COO], [CF ₃ SO ₃]	(Dong et al., 2016)
23	Xylanase, endoglucanase	Imidazolium	Alkyl	[1COO]	(Wu et al., 2016)

Toxicity effects of ILs on bacteria

The toxicology bioassay using *Vibrio fischeri* (*V. fischeri*) is the most common assay method. Therefore, in studies on the IL toxicity to bacteria, the most frequently used bacterium is *V. fischeri* which is a species of marine bacterium with Gram-negative bioluminescent.

Costa et al. (2014b) investigated the toxicity of 7 ILs by developing a novel approach employing the *V. fischeri* assay. It was found that: 1) Regarding the cations, the toxicity of ILs incorporating aromatic groups in the cationic cores was higher than that of ILs without aromatic rings, while the toxicity of imidazolium-based ILs was lower than that of pyridinium-based ILs. 2) Regarding the anions, the toxicity of ILs with acetate and methanesulfonate anions to *V. fischeri* is less than that of ILs containing tetrafluoroborate anion. The toxicity of 29 pyridinium-, imidazolium- and ammonium-ILs to *V. fischeri* was investigated by Montalbán et al. (2016). In their project, the influences of the anions and the length of the alkyl chains on IL toxicity were investigated. The results show that: ILs based on longer alkyl chains had higher toxicity; ILs with fluorine-containing anions have higher toxicity and the growth in the amount of fluorine atoms caused their toxicity to increase.

Similar studies were carried out by Delgado-Mellado et al. (2019) and Kusumahastuti et al. (2019). The authors confirmed that the alkyl chain length was a main factor for the ILs toxicity again. All ILs having alkyl groups over C10 were poisonous, with almost no change in toxicity from C10 to C14, and a slight drop in toxicity when the alkyl chains were larger than C16. With the aim of obtaining qualitative structure-property relationships and providing guidance for the design of "greener" ILs, toxicity data for 16 ILs were reported by Hernandez-Fernandez et al. (2015), also by the means of *V. fischeri* bioassay to measure. The authors found that the incorporation of hydroxyl groups in the alkyl side chains of the imidazolium IL ([HOC₃mim]) remarkably reduced the toxicity of imidazolium-based IL ([C₄mim]). The toxicity sequence of ILs reported by the authors is shown in [Figure 2.6](#).

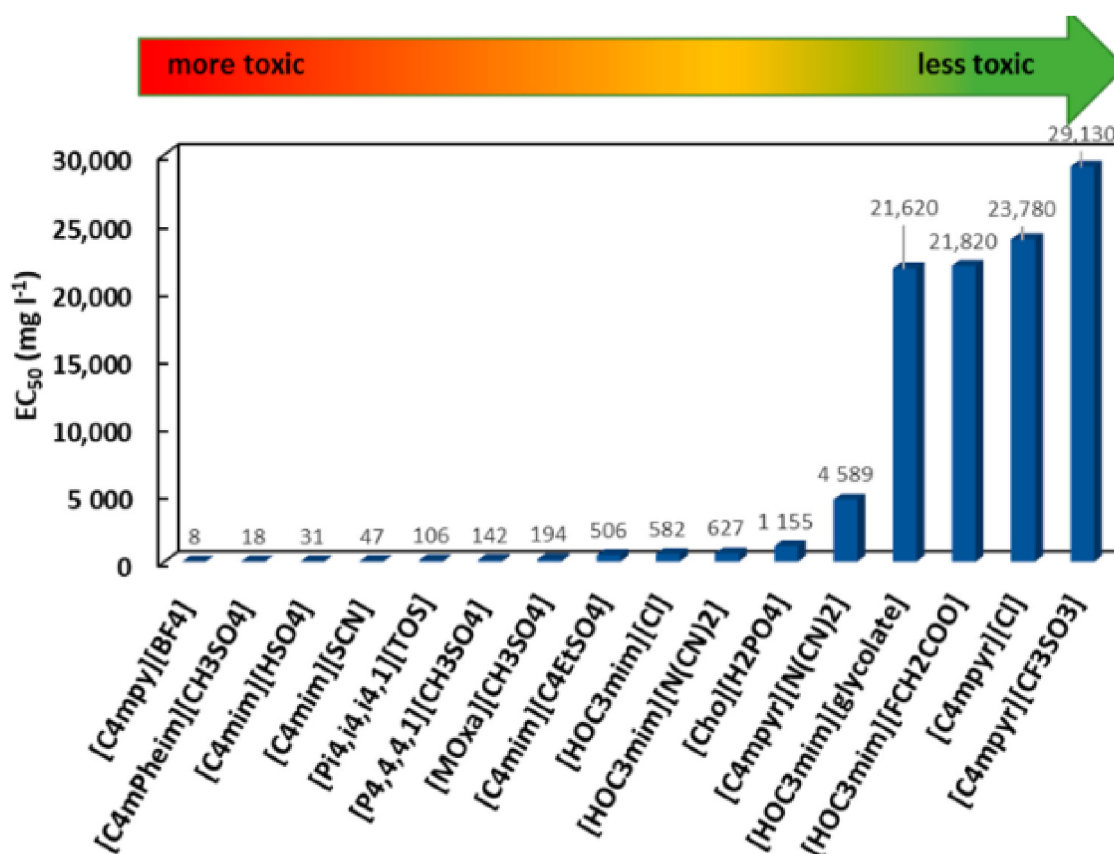


Figure 2.6 EC₅₀ data of ILs towards *V. fischeri* (Hernández-Fernández et al., 2015)

Toxicity effects of ILs on animal cells

Animal cell cultures, which can be used to investigate changes in fundamental functions of cells presence of chemicals, are often used to assess the toxicity of chemicals and have been applied to assess the toxicology of ILs. For the experimental estimation of ILs, a number of animal cells were used, for instance, *leukemia rat cell line* (IPC-81) (Ranke et al., 2004), *CaCo-2 cell line* (Frade et al., 2007; Vieira et al., 2019), C6 glioma (Ranke et al., 2004), *HeLa cells* (Pérez et al., 2020; Sehrawat et al., 2020; Voloshina et al., 2020), *MCF cells* (Pérez et al., 2020), *channel catfish ovary (CCO) cells*, *human immortal keratinocyte cells* (Thamke et al., 2019), *murine peritoneal macrophages* (Glanzmann et al., 2018), *human pancreatic cancerous cells* (Thamke et al., 2019), *green monkey kidney cells* (Thamke et al., 2019), *human renal cancer tumor cells* (Caparica et al., 2020), *Africanhuman breast cancerous cells* (Thamke et al., 2019), *human skin cell* (Arunkumar et al., 2020). These cell lines were utilized to test the IL toxicity by measuring various attributes of cells, such as viability, cytosolic enzyme release, vital staining and cell growth rate (Ekwall et al., 1990).

Ranke et al. (2004) initially investigated the effect of IL structures on cell viability, including the effect of ILs alkyl chain length and anion type on IPC-81 cells. They used imidazolium ILs with five anions ($[\text{BF}_4]$, Cl, Br, $[\text{PF}_6]$, $[\text{4MePhSO}_4]$) and various length of alkyl chains from C3 to C10 for performing the experiments. Results demonstrated that the prolongation of the alkyl chain improved the hydrophobicity of ILs, which was closely related to the inhibition of IPC-81 cell viability. However, the influence of anions on toxicology has not been manifested. Similar observations were executed by Stepnowski et al. (2004) in experiments of measuring HeLa cell. In another study, the toxicity of 1-Butyl-3-methylimidazolium tetrafluoroborate ($[\text{IM14}][\text{BF}_4]$) and 1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)amide ($[\text{IM14}][(\text{CF}_3\text{SO}_2)_2\text{N}]$) (Jastorff et al., 2005) exhibited significantly different, which proved that the anion species had inhibitory effects on cell viability, since $[\text{IM14}][(\text{CF}_3\text{SO}_2)_2\text{N}]$ was much more toxic than $[\text{IM14}][\text{BF}_4]$ (Jastorff et al., 2005). The inhibitory effects of ILs with $[(\text{CF}_3\text{SO}_2)_2\text{N}]$ anions on Caco-2 cell viability were also been studied by other researchers (Frade et al., 2007).

By investigated the viability of cells using 27 ILs based on different anions, Stolte et al. (2006) discovered that 10 of them presented distinct toxic effects towards IPC-81. The toxicology of anions in ILs may be attributed to the lipophilicity and/or the instability of them in an aqueous solution. Later, more detailed data for the toxicity of ILs with $[\text{IM12}]$ -containing cations and cyano-containing anions on animal cells, plants, algae, bacteria and crustaceans were tested by Steudte et al. (2012). Besides, Peric et al. (2013) investigated the (eco)toxicity of collected protic and aprotic ILs. Garcia-Lorenzo et al. (2008) studied the effects of different anions for the toxicity of imidazolium-derived ILs employing the CaCO-2 cell line of human. The detected anions of ILs presented strong toxicity in these reports, which could be caused by their hydrophobicity. Consequently, the toxic effects of anions could not be ignored when screening and designing green ILs. The relevant studies about the toxicity of ILs on animal cells are collected in [Table 2.3](#).

Table 2.3 Viability tests of ionic liquids, using several animal cell lines (Cho et al., 2019)

No.	Cell lines	Cations	Side chains	Anions	References
1	IPC-81	Imidazolium	Alkyl	Br, [(CF ₃ SO ₂) ₂ N]	(Jastorff et al., 2005)
2	IPC-81, C6 rat	Imidazolium	Alkyl	[BF ₄], [PF ₆], Cl	(Ranke et al., 2004)
3	IPC-81	Imidazolium, quinolinium, pyrrolidinium, pyridinium, ammonium, phosphonium	Alkyl, hydroxyl, ether, aryl, carboxyl	Cl, Br, [BF ₄], [PF ₆	(Ranke et al., 2007a)
4	IPC-81	Imidazolium	Alkyl	Cl, [BF ₄], [8OSO ₃], [(CF ₃ SO ₂) ₂ N], [(CF ₃) ₂ N], [(2-OPhO) ₂ B]	(“The UFT/ Merck Ionic Liquids Biological Effects Database,” 2020)
5	IPC-81	Imidazolium	Alkyl	27 anions	(Stolte et al., 2006)
6	IPC-81	4-(Dimethylamino) pyridinium, pyridinium, imidazolium, morpholinium, pyrrolidinium, piperidinium, ammonium	Alkyl, alkoxy, ether	Br, Cl	(Stolte et al., 2007a)
7	IPC-81	Cholinium, ammonium, pyrrolidinium	Alkyl, hydroxyl	[1OSO ₃], [1SO ₃], [(CF ₃ SO ₂) ₂ N]	(Stolte et al., 2012)
8	IPC-81	Morpholinium	Aryl, alkyl	17 anions	(Pernak et al., 2011)

No.	Cell lines	Cations	Side chains	Anions	References
9	IPC-81	Imidazolium	Alkyl	[CF ₃ SO ₂] ₂ N, [(C ₂ F ₅) ₃ PF ₃], [B(CN) ₄], [N(CN) ₂], [C(CN) ₃]	(Steudte et al., 2012)
10	IPC-81	Imidazolium	Aryl	Br, Cl, I	(Stolte et al., 2013)
11	IPC-81	Ammonium	Hydroxyalkyl	Alkanoate	(Peric et al., 2013)
12	HeLa	Imidazolium	Alkyl	[BF ₄], [PF ₆], Cl	(Stepnowski et al., 2004)
13	HT29, Caco-2	Imidazolium, cholinium, guanidinium, ammonium, phosphonium	Alkyl, ether, hydroxyl	[BF ₄], [PF ₆], [(CF ₃ SO ₂) ₂ N], [N(CN) ₂], acetatesulfame, saccharinate	(Frade et al., 2007)
14	Caco-2	Imidazolium	Alkyl, aryl, alkylaryl	[PF ₆], Cl, [1OSO ₃], [2OSO ₃]	(García-Lorenzo et al., 2008)
15	IPC-81	Hydroxypyridinium	Alkoxyethyl	Acetatesulfame, saccharinate, Cl	(Stasiewicz et al., 2008)
16	PC12	Imidazolium	Alkyl	Cl	(Li et al., 2012)
17	CCO	Imidazolium	Alkyl, aryl, dimethylaminoalkyl	[(CF ₃ SO ₂) ₂ N], [BF ₄], [PF ₆], Br	(Cvjetko Bubalo et al., 2015)
18	CCO of <i>Ictalurus punctatus</i>	Imidazolium	Alkyl	[(CF ₃ SO ₂) ₂ N], [BF ₄], [PF ₆]	(Cvjetko et al., 2012)
19	CCO of <i>Ictalurus punctatus</i>	Imidazolium	Alkyl	[(CF ₃ SO ₂) ₂ N], [BF ₄], [PF ₆], Br	(Radošević et al., 2013)
20	CCO of <i>Ictalurus punctatus</i>	Phosphonium	Alkyl	Alkanoate	(Ruokonen et al., 2016)

2.4 QSPR models for the property prediction of ionic liquids

2.4.1 Description of QSPR method

Definition of QSPR

During the 1960s, researchers had attempted to harness the computer to imitate and explain chemical phenomena in various chemistry fields. Moreover, methods for storing and retrieving the information of chemical structure were developed. (Dyson et al., 1968). Then, techniques were explored to establish quantitative structure-activity/property relationships (QSAR/QSPR) for predicting physical, chemical, biological or environmental data of chemicals were developed (Gasteiger, 2016; Hansch and T., 1964).

QSPR methods are computer-based mathematical models presenting the relationships between the physical-chemical property of compounds (experimental values) and their chemical structures (descriptors). QSPRs are important techniques which are used for the prediction of physicochemical and biological properties of compounds which have not been investigated through experimental methods. Generally, a QSPR model expressed by a mathematical equation (Eq.1).

$$F(y) = f(X_1, X_2, \dots, X_n) \quad (\text{Eq.1})$$

where $F(y)$ stands for the property of compounds; X_n represents the n -th molecular descriptor relating a property; the function (f) may be linear or non-linear.

This method assumes that strong correlations exist between the structures and the property of chemicals, and it has been proved that the structure and observed properties of organic compounds have strong relationship within the same group.

2.4.2 QSPR models for predicting gas absorption in ionic liquids

ILs for various gases absorption have been extensively investigated and applied because of their extraordinary nature recent years. However, measuring their absorptive ability for gases in laboratory are not only time- and labor-consuming but also costly. Therefore, the development of prediction models has been required to fill gaps in related data scarcity. The QSPR approach is one of the most important methods for developing computational model to predict properties of chemicals.

Safamirzaei and Modarress (2012) established a QSPR method to model solubility oxygen, nitrogen, argon, carbon monoxide, carbon dioxide, ethane and methane in an IL (1-butyl-3-methylimidazolium tetrafluoroborate). The molecular weight and sphericity of gas molecule were selected as descriptors and a neural network technique was applied. In their work, absolute pressure and temperature were the other two input variables. The superiority of the proposed model for gas solubility prediction were confirmed by low deviations of training and validation stages. Mehraein and Riahi (2017) built an efficient QSPR model to estimate CO₂ absorptivity applying a data set with 21 ILs. Three molecular descriptors including, number of single bonds, average bonding information content as well as image of the Onsager-Kirkwood were calculated based on the entire structure of ILs as inputs in their work. The network technique of genetic algorithm-multi linear regression (GA-MLR) coupled with least-squares support vector machines (LS-SVM) was employed for a nonlinear model development. Results showed that the developed mode was satisfactory with the square of correlation coefficient ($R^2=0.962$) and the root mean square error ($RMSE=0.015$). Two QSPR models for evaluating H₂S solubility based on a data set of 27 ILs were launched by Zhao et al. (2016b). The molecular descriptors based on the charge distribution area were obtained and an advanced algorithm, i.e., extreme learning machine were introduced to construct estimation models in their study. High R^2 values for two models were 0.998 and 0.997 respectively, indicating that their proposed QSPR models are reliable and efficient for the evaluation of H₂S absorptivity in ILs.

Some studies reported QSPR models for evaluating henry's law constant (HLC) of gases in ILs. Ghaslani et al. (2017) launched a QSPR study for the prediction of HLC of CO₂ in ILs with a 32 ILs-based dataset. Networks of LS-SVM (least squares support vector machine) and MLR (Multi linear regression) were harnessed, and three cation/anion structure descriptors selected by genetic algorithm (GA) for modeling. The non-linear LA-SVM-based models showed more accurate outcomes than the linear model using MLR. Kang et al. (2018a) developed three predictive QSPR models for

HLC of CO₂ in ILs using 7 molecular descriptors based the combination of electrostatic potential surface area (S_{EP}) and charge distribution area ($S_{\sigma\text{-profile}}$) descriptors of isolated cations and anions. 34 ILs involving 297 experiment data were collected for modeling. A best nonlinear model with high R^2 (0.995) and low average absolute relative deviation (AARD%=3.22) were successfully obtained to accurately estimate HLC values of CO₂ in ILs.

The related work for forecasting gas absorptivity in ILs based on the QSPR remains limited. However, using computer techniques to model the properties of chemicals replacing amount of laboratory work in various fields is significant to screen and design appropriate compounds. Therefore, more efforts to explore the relationships between properties and structures of compounds for efficiently modeling are desired.

2.4.3 QSPR models for predicting the toxicity of ionic liquids

It has been known that the toxicity of ILs towards a variety of living organisms and plants have been investigated and reported by many researchers. Subsequently, due to the countless number of ILs, the modeling methods for toxicity estimation of ILs have been made great progress based on various methods on request, such as equation of state (EoS) method, group contribution (GC) method, QSAR/QSPR method, empirical correlation (EC) methods, artificial intelligence (AI) algorithm as well as combinations of two or more of these methods and so on. The method of QSAR/QSPR recommended by the REACH (Registration, evaluation, authorization and Restriction of Chemicals) (Imaizumi, 2016; Lewis, 2018) and OECD (Organization for Economic Cooperation and Development) have been applied in a number of established models for estimating environmental risks of ILs on various biological species were also developed.

Peric et al (2015) using a QSPR method to predict the IL (eco)toxicity to five species, namely *Raphidocelis subcapitata*, *A. fischeri*, *Lemna minor*, IPC-81 cells as well as *acetylcholinesterase*. The established model was employed GC method with the number of groups/fragments (i.e., cations, anions, substitution) as descriptors. Conclusively, the QSPR model have the potential to screen low toxic ILs (Peric et al., 2015). Yan et al. (2019) proposed a QSPR model to evaluate the toxicity of ILs on IPC-81, using a set of norm indexes descriptors about space distribution of atoms in anions and cations.

Table 2.4 listed the studies involving predictive models for toxicity estimation of ILs based on QSPR/QSAR methods. It can be seen that, a majority of proposed QSPR

have quite good predictivity with the R^2 in the range of 0.8-0.99, demonstrating these models are satisfactory for the hazard assessment of IL and could be helpful to screen and design environmentally-friendly ILs.

Table 2.4 Summary of the experimental data used for the reported QSPR models and the state of the art of developed QSPR models for ILs

No.	End points	Number of ILs (anion and cation)	Descriptors	R^2	References
Tested species - <i>Vibrio fischeri</i> (<i>V. fischeri</i>)					
1	EC50	43 ILs	Group contribution methods	0.925	(Luis et al., 2007)
2	EC50	75 ILs with 17 anions and 9 cations	Group contribution methods	0.924	(Luis et al., 2010)
3	EC50	51 ILs	minimum net atomic charge for a C atom; WPSA-1; PPSA1; TMSA; maximum atomic orbital electronic population; LUMO + 1 energy	0.903 – 0.912	(Bruzzzone et al., 2011)
4	EC50	97 ILs	excess molar refraction; dipolarity/polarizability, hydrogen-bonding acidity, hydrogen-bonding basicity and McGowan volume	0.778 for IPC-81; 0.762 for <i>V. fischeri</i> ; 0.776 for <i>S. vacuolatus</i>	(Cho et al., 2013)
5	IC50, LC50	24 bromide based ILs	ELUMO; dipole moment; total energy, volume of ILs cation; molecular volume; the electron affinities	0.895 for <i>D. magna</i> 0.954 for <i>V. fischeri</i>	(C. Wang et al., 2015)
6	EC50	157 ILs composed of 74 cations and 22 anions	topological index, character vector CV of atoms, distance matrix for atom position	0.908	(Yan et al., 2015)

No.	End points	Number of ILs (anion and cation)	Descriptors	R ²	References
7	EC50	110 ILs with 29 anions and 49 cations	σ-Profile descriptors	0.906 - 0.910 for MLR; 0.961 - 0.979 for MLP	(Ghanem et al., 2017)
Tested species - <i>Staphylococcus aureus</i> (<i>S. aureus</i>)					
8	EC50	25 imidazolium based ILs	σ - Profile as molecular descriptors	0.963 - 0.972;	(Ghanem et al., 2015)
9	MIC, MBC	169 and 101 ILs with MICs and MBCs, respectively	Matrix norm index, atomic radius, atom weight, electronegativity, number of atoms, atom charge, molecular weight, branching degree	0.919 for pMIC; 0.913 for pMBC	(He et al., 2018)
10	MIC	131 ILs	E-State indices, ALogPS, ADRIANA.Code, Dragon 7.0, Chemaxon, Inductive descriptors, Fragmentor descriptors, GSFrag	0.83-0.88	(Hodyna et al., 2018)
11	MIC	242 ILs	E-State indices, ALogPS, Chemaxon, GSFrag, ToxAlerts (Structural Alerts)	0.85	(Trush et al., 2018)
Tested species - <i>Leukemia Rat Cell Line</i> (IPC-81)					
12	EC50	227 ILs with 25 anions and 227 cations	RARS, Kier symmetry index, heavy atom count, topological charge index of order 8	0.92	(Fatemi and Izadiyan, 2011)
13	EC50	97 ILs	excess molar refraction; dipolarity/polarizability, hydrogen-bonding acidity, hydrogen-bonding basicity and McGowan volume	0.778 for IPC-81; 0.762 for <i>V. fischeri</i> ; 0.776 for <i>S. vacuolatus</i>	(Cho et al., 2013)

No.	End points	Number of ILs (anion and cation)	Descriptors	R ²	References
14	EC50	100 ILs	Min partial charge for a N atom; relative number of O atoms; TMSA; number of C atoms	Training set: 0.918 - 0.959 Test set: 0.892 for MLR and 0.958 for SVM	(Zhao et al., 2014)
15	EC50	100 ILs	Ferreira–Kiralj hydrophobicity parameter; CrippenlogP and Mannhold log P descriptors	0.762 – 0.813	(de Melo, 2016)
16	EC50	253 ILs	XLogP, NAtoms, TPSA, Polariz, Dipole, InertiaZ, and Span	>0.9	(Gupta et al., 2015)
17	EC50	289 ILs	Two-dimensional structural and QTMS indices; Molar refraction, solute dipolarity, polarizability, hydrogen-bond acidity and basicity, McGowan volume; Lipophilicity, Randić's parameter, molecular connectivity, ETA indices, etc.	0.869	(Das et al., 2015)
18	EC50	17 ILs	Electrophilic indices (ω), the energy of highest occupied (EHOMO) and lowest unoccupied molecular orbital, (ELUMO) and energy gap (ΔE)	0.999	(Salam et al., 2016)
19	EC50	10 groups of 304 ILs	Weighted Holistic Invariant Molecular Descriptors (WHIM), ring descriptors, functional group counts, topological and constitutional indices	0.77~0.95 for local models, 0.772 for the global model	(Sosnowska et al., 2017)
20	EC50	119 ILs with 57 cations and 21 anions	S _{EP} and S _{σ-profile} , electrostatic potential V(r)	MLR – 0.92; SVM - 0.941; ELM – 0.969	(Cao et al., 2018)

No.	End points	Number of ILs (anion and cation)	Descriptors	R^2	References
Tested species - <i>Daphnia magna</i> (<i>D. magna</i>)					
21	EC50	64 ILs	Group contribution methods	0.974	(Ismail Hossain et al., 2011)
22	LC50	62 ILs	ETA index, topological non-ETA and thermodynamic parameters	0.948	(Roy and Das, 2013)
23	IC50, LC50	24 bromide based ILs	ELUMO; dipole moment; total energy, volume of ILs cation; molecular volume; the electron affinities	0.895 for <i>D. magna</i> , 0.954 for <i>V. fischeri</i>	(C. Wang et al., 2015)

Chapter III

Prediction of ammonia absorption in ionic liquids
based on extreme learning machine modeling and a
novel molecular descriptor S_{EP}

Xuejing Kang, Zuopeng Lv, Zhongbing Chen, Yongsheng Zhao

Adapted from Environmental Research, 189(2020): 109951

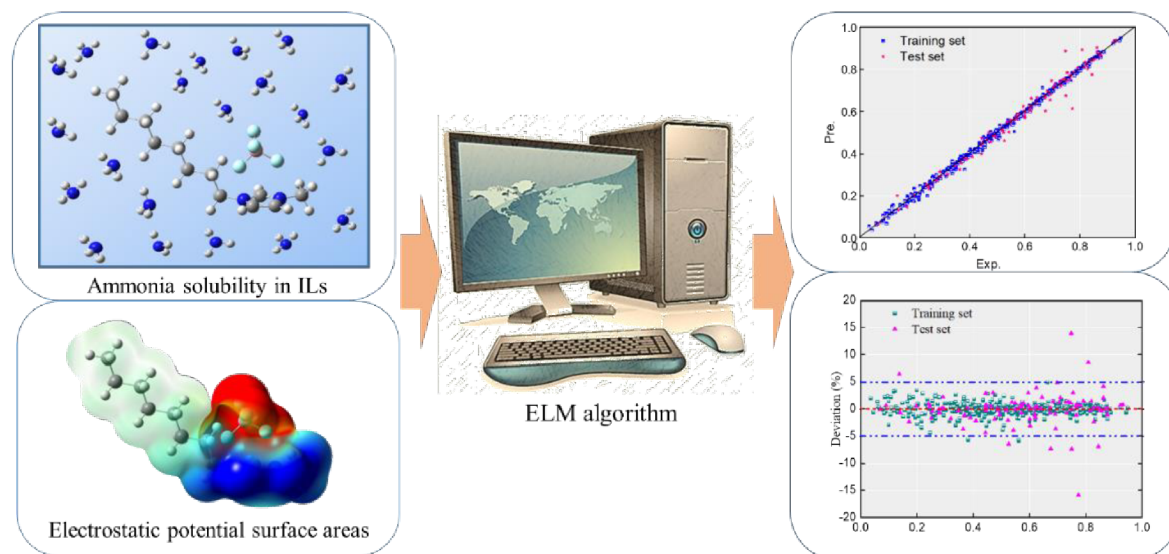
Contents

3.1 Abstract	35
3.2 Introduction	37
3.3 Theory	38
3.3.1 <i>Extreme learning machine (ELM)</i>	38
3.3.2 <i>Electrostatic potential</i>	41
3.4. Data and descriptors	42
3.4.1 <i>Data Acquisition</i>	42
3.4.2 <i>S_{EP} calculation</i>	45
3.4.3 <i>Executions of networks training</i>	47
3.5 Results and discussion	48
3.5.1 <i>Descriptor selection and analysis</i>	48
3.5.2 <i>ELM model development</i>	51
3.6 Conclusion	54
3.7 Supplementary data	54

3.1 Abstract

The large amounts of ammonia emissions generated from industrial production have caused serious environmental pollution problems, such as soil acidification, eutrophication, the formation of fine particles and changes in the global greenhouse balance, and also greatly endanger human health. At present, effectively reducing ammonia emissions or recovering ammonia is still a huge challenge. Ionic liquids (ILs) as a new class of green solvent have been introduced for ammonia absorption with great potential, but a huge number on combination systems of ILs lead to the difficulty of measuring the ammonia solubility in all ILs by experiments (e.g., danger and cost). Hereby, this study proposed a novel approach for estimating the ammonia solubility in different ILs. A predictive model was developed based on the novel algorithm - extreme learning machine (ELM) and the molecular descriptors of electrostatic potential surface areas (S_{EP}) as input parameters. Besides, 502 data points of ammonia solubility in 17 ILs were gathered with a wide range of pressure and temperature. For the total set, the determination coefficient (R^2) and the average absolute relative deviation (AARD) of the developed model were 0.9937 and 2.95%, respectively. The regression plots revealed good consistency between predictive and experimental data points. Results show the good performance and reliability of the developed model, indicating that the proposed approach can be potentially applied for screening reasonable ILs to absorb ammonia from chemical industry processes.

Graphical abstract



3.2 Introduction

Ammonia is a typical poisonous gas which is harmful to human health and its emission leads to a range of environmental issues, such as soil acidification, eutrophication, the fine particulate matter formation and the global greenhouse balance alteration (Erismann et al., 2007; Sutton et al., 2008; Sutton and Fowler, 2002). However, it belongs to a necessary gas for some biological and physiological processes and has been widely applied in various fields, for example, the production of nitric acid, nitrogenous fertilizers, as well as pharmaceuticals, petroleum refining, and so on (Li et al., 2010, 2015; J. Wang et al., 2019; Zeng et al., 2018). Therefore, absorption and reduction of ammonia from effluent gas in industries are greatly crucial. In traditional ammonia-water absorption systems, high water pressure leads to the need for expensive rectification equipment (Yokozeki and Shiflett, 2007a). Hence, environmentally-friendly solvents capturing ammonia are required to be developed.

Ionic liquids (ILs) are a special type of molten salts entirely formed by ions (Hosseini et al., 2019). They aroused considerable interest because of their distinctive physicochemical properties, especially the character of negligible vapor pressure (Xinyan Liu et al., 2018; Song et al., 2020b; G. Wang et al., 2015; Yokozeki and Shiflett, 2007b). Another important feature of ILs is that their properties can be designed for a particular application by tuning the cations and anions (Bedia et al., 2012; Song et al., 2018, 2017; Zhao et al., 2019). Therefore, ILs have been sanguinely viewed as not only green liquid solvents but also useful “designer solvents”. In recent two decades, ILs shows more promising perspectives in the field of absorption for industrial waste gases, such as CO₂ capture (Karadas et al., 2010; Palgunadi et al., 2009; Pinto et al., 2013; Song et al., 2020a), and the absorption of CO (Kumelan et al., 2005; Ohlin et al., 2004), H₂S (Jalili et al., 2013), SO₂ (Anderson et al., 2006), H₂ (Kumelan et al., 2006), NH₃ (Bedia et al., 2012) and so on.

Due to the possibility of designing ILs, extensive ILs can be synthesized by experiments. Additionally, taking into account the high operational costs and the difficulties of carrying out experiments, some estimation approaches have been employed to predict the solubility of gases in ionic liquids (Baghban et al., 2018; Song et al., 2020a). For example, Baghban et al. (2015) used multi-layer perceptron artificial neural network (MLP-ANN) and adaptive neuro-fuzzy interference system (ANFIS) to predict the absorption of carbon dioxide in ILs; the MLP-ANN, ANFIS and basis Function Artificial Neural Network (RBF-ANN) were utilized for predicting hydrogen sulfide solubility in various ILs (Amedi et al., 2016); Yokozeki and Shiflett (2010)

applied the generic van der Waals equation of state (EOS) to build models for several gases solubility (CO_2 , $\text{CF}_3\text{-CFH}_2$, SO_2 , and NH_3) in room-temperature ILs. To our best knowledge, however, the outcomes regarding the prediction of NH_3 solubility in ILs were barely published. Baghban et al. (2018) reported two predictive models based on Support Vector Machine (SVM), a Least Square Support Vector Machine (LS-SVM) approaches and 352 experimental data points. The results of these two models, higher regression coefficients and lower deviations for the training set and testing set, showed that both models had a good capability for estimating NH_3 solubility in ILs. In addition, it should be noted that the selection of descriptors is also crucial for developing models. the pressure (P), and temperature (T) and three physical properties of pure ILs (including molecular weight (MW), the critical temperature (T_c) and the critical pressure (P_c)) were used as input parameters for modeling by Baghban et al. (2018).

Different with the Baghban and his co-authors, in the current work, we intend to develop a new model for predicting ammonia solubility in various ILs through a completely different approach: (1) to gather an extensive dataset (502 data points) of ammonia solubility in various ILs; (2) to introduce a novel class of algorithm, Extreme learning machine (ELM) for modeling; (3) to propose a new type of input descriptions, electrostatic potential surface areas (S_{EP}) of ILs. ELM is a fast and efficient type of single-layer feedforward network (SLFN) learning algorithm, which is considered to be a specific universal learning machine. Compared with traditional neural networks (NN) algorithms, ELM possesses the advantages of fast training speed and sufficient generalization performance and has been extensively applied in the field of prediction (Li et al., 2019; Luo et al., 2019). The electrostatic potential is generated by the nuclei and electrons of a molecule and is a true physical property and observable, while the electrostatic potentials of a molecular surface can reflect the particular characters of the specific molecules (Murray and Politzer, 2017, 2002, 1998).

3.3 Theory

3.3.1 Extreme learning machine (ELM)

ELM has increasingly attracted more attention from academia and industry (Cui et al., 2018) and it was proposed by Huang et al. to solve the issues of conventional neural networks, such as the shortcomings of complex training parameters, slow learning rate and easy to capture the local minima (Deo and Şahin, 2015; Huang et al., 2006a, 2004). The network structure of ELM is presented in [Figure 3.1](#), consisting of

three layers, one input layer, one hidden layer, and one output layer. The input weight (α_j) is for connecting the input and hidden layers, while the output weight (β_k) is linked to the hidden and output layer. The main character of ELM is that the input weights (α_j) and the hidden layer biases (b_j) are randomly assigned (Sattar et al., 2019), in the training phase, but the output weights (β_k) are calculated by the generalized inverse of matrices. Assuming that the training sample set x is the input and the hidden layer output is $H(x)$, the calculation formula of the $H(x)$ is as follows:

$$H(x) = [h_1(x), \dots, h_j(x), \dots, h_q(x)] \quad (1)$$

where $h_j(x)$ is the output of the j -th hidden node, and its function can be expressed as: $h_j(x) = g(\alpha_j x + b_j)$ (2)

where $g(\alpha_j x + b_j)$ is a non-linear excitation function, and the common functions are Sigmoid function and Gaussian function. The sigmoid function is used in the current work, and its equation is:

$$g(\alpha_j x + b_j) = \frac{1}{1 + \exp(-\alpha_j x + b_j)} \quad (3)$$

If the weight matrix from the hidden layer to the output layer is:

$$\beta = [\beta_1, \dots, \beta_k, \dots, \beta_n] \quad (4)$$

where β_k stands for the weight linking the k -th hidden node and output nodes, then the output sample matrix (y) of the ELM network is

$$y = H(x) \beta \quad (5)$$

The output weights β can be obtained by calculating the least square solutions of the linear Eq. (6):

$$\|H\hat{\beta} - y\| = \|HH^+y - y\| = \min_{\beta} \|H\beta - y\| \quad (6)$$

Here the optimal solution of the above linear system is

$$\hat{\beta} = H^+y \quad (7)$$

where H^+ is the Moore–Penrose generalized inverse of output matrix of hidden layer H (Abdolkarimi et al., 2018; Huang et al., 2006a; Mayne et al., 1972; Sun and Sun, 2017).

In conclusion, the normal steps of ELM are summarized as follows (Abdolkarimi et al., 2018; Huang et al., 2006a):

Inputs: giving input parameters (x) with desired outputs (y) for training; specifying the activation function and the number of hidden nodes.

Algorithm: 1. Randomly initialize the input weight α_j and the bias b_j , $j = 1, \dots, q$; 2. The output matrix of hidden layer H is calculated; 3. The output weight β is calculated by Eq. (7).

In this work, the input variables (x) will be selected from pressure (P), temperature (T), and a series of S_{EP} descriptors, while the target outcomes (y) will be the predictive values of ammonia solubility in different ILs.

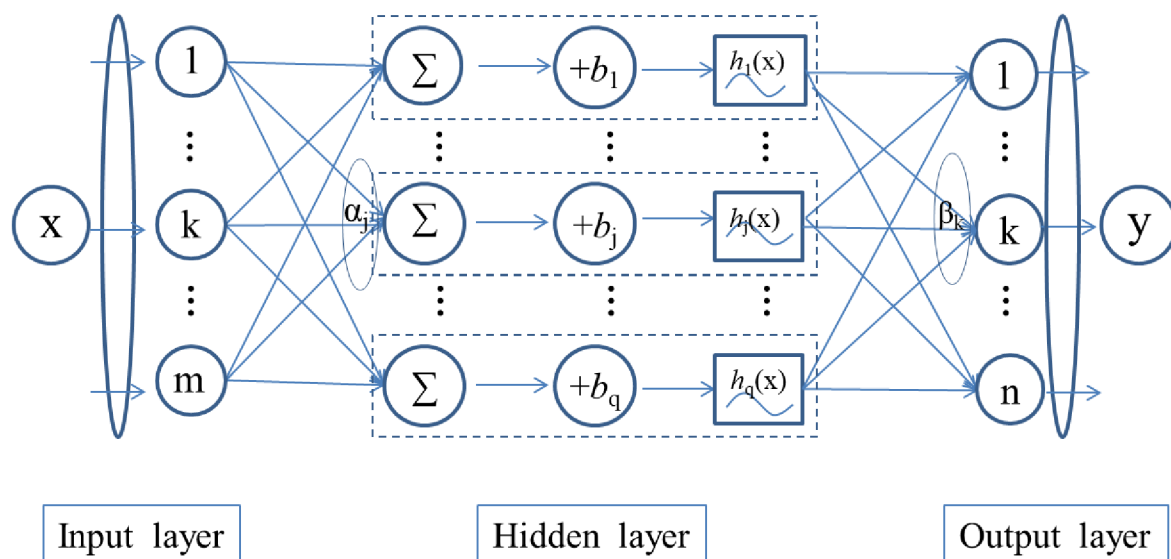


Figure 3.1 ELM network structure in this study for the prediction of NH_3 solubility in ILs

3.3.2 Electrostatic potential

The electrostatic potential, $V(r)$, is a three-dimensional local feature and a physical observable (Kolář and Hobza, 2016). The electrostatic potential can be calculated at all or any point in the space around the system (Murray and Politzer, 2011a). The Eq. (8) is the rigorous form of $V(r)$ at a point r (Murray and Politzer, 2017, 2011a).

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r') dr'}{|r' - r|} \quad (8)$$

Here, Z_A is the charge of nucleus A at position R_A , and $\rho(r)$ represents the electron density of the molecule at position r . $|R_A - r|$ denotes the distance between nucleus A and the point r and $|r' - r|$ means the distance of electron charge from r .

Various properties and reactive behaviour of atoms and molecules can be described based on the electrostatic potential which is proved to be beneficial for comprehending the characteristics and interaction preferences of molecules (Politzer and Murray, 2002). The electrostatic potential surface is to apply electrostatic potential values mapping with the isodensity surfaces and is a common approach to visualize the electrostatic properties of molecules (Rathi et al., 2019). Therefore, the electrostatic potential

surface area (S_{EP}), symbolizing the molecule surface area according to the electrostatic potential in different intervals, is possible to have the ability to evaluate the properties of compounds (Kang et al., 2018c).

3.4. Data and descriptors

3.4.1 Data Acquisition

In this work, a data set comprising 502 experimental data points of ammonia solubility in 17 ILs was gathered from the literature. The names of ILs, ammonia solubility ranges in different ILs and the corresponding data point numbers, temperature ranges, pressure ranges accompanied by the references have been listed in [Table 3.1](#). The specific information for the experimental data and corresponding ILs has been provided in Supplementary data. The entire data points were arbitrarily divided into two parts: 80% of them (402 data points) are used as a training set for developing model and 20% data (100 data points) is for a test set to validate the proposed model.

Table 3.1 Temperature, pressure, and NH₃ solubility range of ILs used in this study

No.	Ionic liquid	T range (K)	P range (bar)	NH ₃ solubility range (mole fraction)	No. of data points	AARD% (ELM)	Ref.
1	[BMIM][BF ₄]	282.2-355.1	0.91-25.7	0.068-0.844	36	4.52	(Yokozeki and Shiflett, 2007b)
		293.15-333.15	0.7-8.3	0.0608-0.7531	25	6.83	(Li et al., 2010)
2	[BMIM][PF ₆]	283.4-355.8	1.38-27	0.239-0.862	29	2.18	(Yokozeki and Shiflett, 2007b)
3	[EMIM][BF ₄]	293.15-333.15	1.1-6.3	0.0838-0.6921	25	3.56	(Li et al., 2010)
4	[EMIM][Tf ₂ N]	283.3-347.6	1.14-28.6	0.045-0.948	30	4.23	(Yokozeki and Shiflett, 2007b)
5	[HMIM][BF ₄]	293.15-333.15	1.4-7.1	0.128-0.7543	25	2.1	(Li et al., 2010)
6	[HMIM][Cl]	283.1-347.9	0.44-24.9	0.06-0.837	30	5.65	(Yokozeki and Shiflett, 2007b)
7	[OMIM][BF ₄]	293.15-333.15	1-6.1	0.1321-0.8081	25	2.85	(Li et al., 2010)
8	[EMIM][OAc]	282.5-348.5	3.21-28.91	0.473-0.877	30	0.85	(Yokozeki and Shiflett, 2007a)
9	[EMIM][EtSO ₄]	282.7-372.3	2.87-47.77	0.424-0.875	29	0.83	(Yokozeki and

							Shiflett, 2007a)
10	[EMIM][SCN]	283.2-372.8	2.44-50.07	0.34-0.876	36	1.17	(Yokozeki and Shiflett, 2007a)
11	[DMEA][OAc]	283.2-372.8	1.36-42.49	0.454-0.865	32	1.54	(Yokozeki and Shiflett, 2007a)
12	[EtOHMIM][Tf ₂ N]	298.15-343.15	1.055-6.4952	0.3299-0.7805	25	0.6	(Li et al., 2015)
13	[EtOHMIM][PF ₆]	298.15-343.15	0.4476-5.7544	0.187-0.7424	25	1.94	(Li et al., 2015)
14	[EtOHMIM][BF ₄]	298.15-343.15	0.9882-6.0442	0.2736-0.6983	25	1.17	(Li et al., 2015)
15	[EtOHMIM][SCN]	298.15-343.15	0.6798-5.9238	0.1095-0.672	25	2.08	(Li et al., 2015)
16	[EtOHMIM][DCA]	298.15-343.15	0.3996-5.7265	0.0354-0.6856	25	7.21	(Li et al., 2015)
17	[EtOHMIM][NO ₃]	298.15-343.15	0.6184-5.5965	0.0994-0.6665	25	4.25	(Li et al., 2015)

3.4.2 S_{EP} calculation

To obtain the parameters of S_{EP} for different ILs, there are two main steps to complete. The first one is to optimize the geometrical structures of the 17 ILs using the Gaussian 09 B.01 software at the level of B3LYP/6-31++G (d, p) based on the DFT theory. During this procedure, the total energy and vibration frequencies of ILs were calculated to confirm their local minimum of energy point without any virtual frequency. The next step is to calculate the electrostatic potential values of ILs by the software of Multiwfn 3.7 based on the optimal structure of ILs and the isodensity surface (electron density $\rho=0.001e/\text{bohr}^3$), the van der Waals surface defined by Bader (1987). The molecular surface is colored according to the electrostatic potential energy of each surface point (Roman-Vicharra et al., 2018). In this work, the calculated electrostatic potential on each IL surface ranges from -60 to 60 kcal/mol which evenly divided into 240 intervals to obtain the corresponding surface area. Hence, 240 S_{EP} values for each IL were obtained. The structures of ILs presenting electrostatic potential values by various colors on their surfaces were finally depicted in [Figure 3.2](#). It can be shown that the darker the red or blue color is, the stronger the polarity of the IL is. The electrostatic potentials versus the corresponding S_{EP} values of [HMIM][BF₄] were representatively expressed in [Figure 3.3](#).

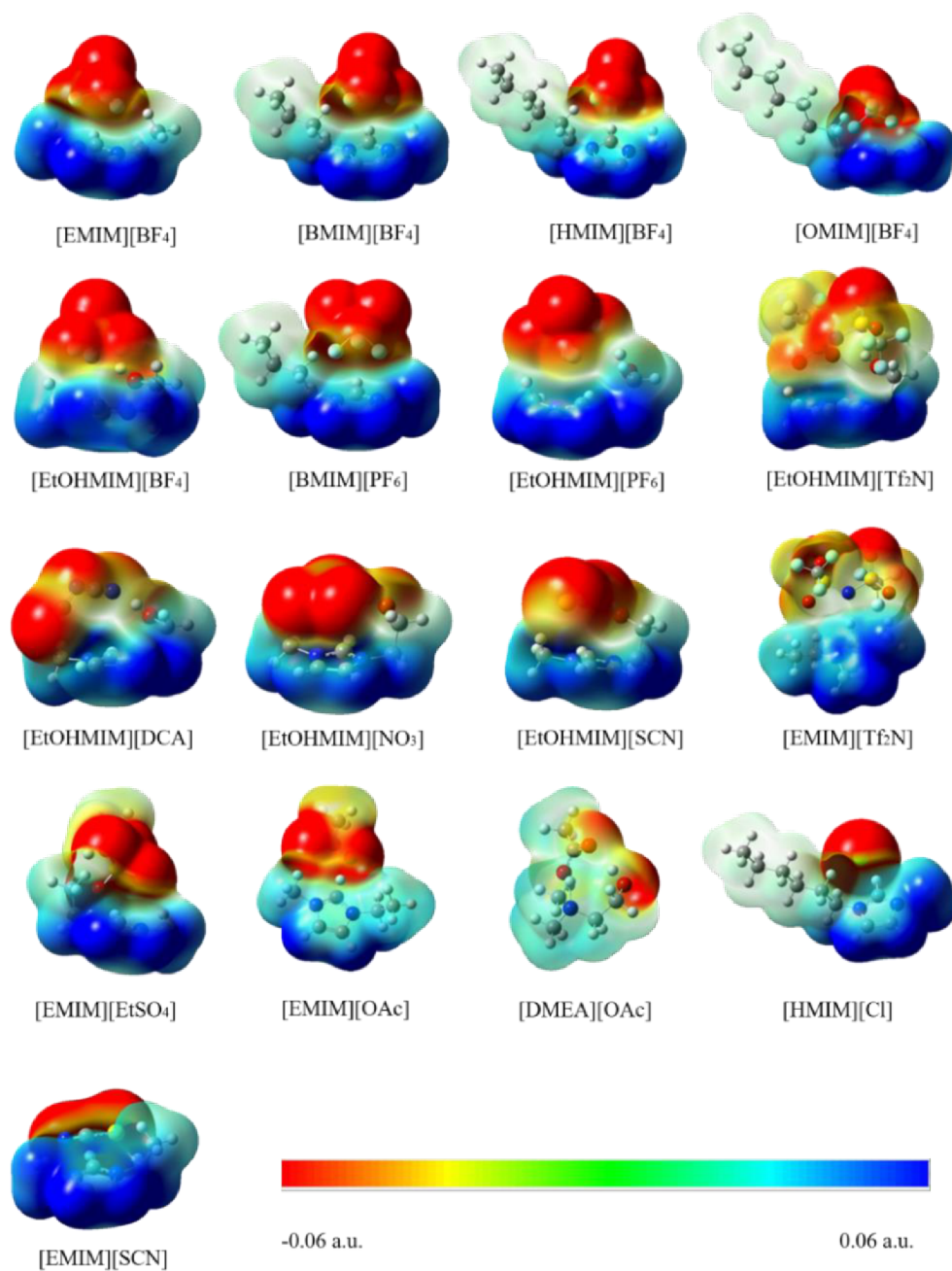


Figure 3.2 Electrostatic potential surface mapped electron total density with an isovalue 0.001 e/Bohr³ of the optimized ILs

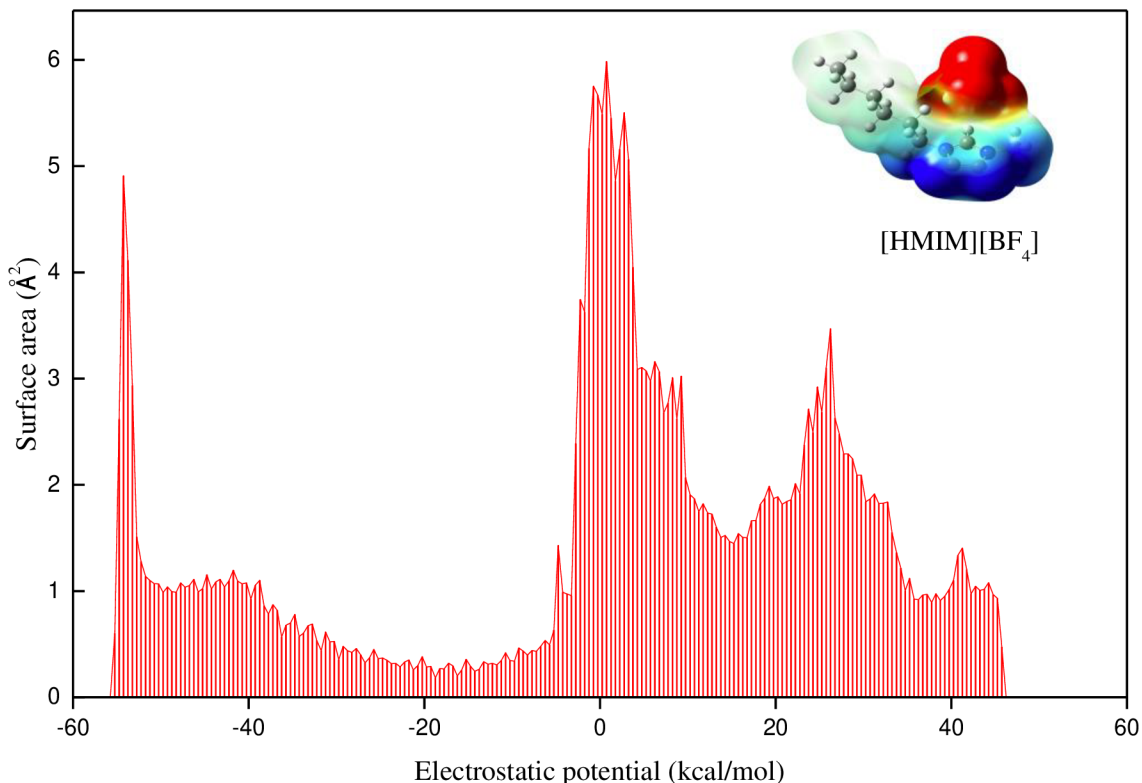


Figure 3.3 Electrostatic potentials versus corresponding surface areas of [HMIM][BF₄]

3.4.3 Executions of networks training

In the current study, an ELM model is built for estimating the ammonia solubility in ILs by employing the MATLAB package. To assess the performance of the predictive model by ELM, the regression plot as a graphical verification approach is applied. Meanwhile, some statistical methods are utilized, such as the determination coefficient (R^2), deviation (D), absolute deviation (AD), average absolute relative deviation (AARD), together with the mean-square error (MSE) and root mean square error (RMSE). Their expressions are shown as follows (Eq. (9) ~ (14)):

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2}{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2} \quad (9)$$

$$D (\%) = 100 \times (y_i^{cal} - y_i^{exp}) \quad (10)$$

$$\text{AAD (\%)} = 100 \times \frac{1}{N_p} \sum_{i=1}^{N_p} |y_i^{cal} - y_i^{exp}| \quad (11)$$

$$\text{AARD (\%)} = 100 \times \sum_{i=1}^{N_p} \left| \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \right| / N_p \quad (12)$$

$$\text{MSE} = \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p \quad (13)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p} \quad (14)$$

where N_p is the sample quantities of a data set; y_i^{cal} and y_i^{exp} signifies the predictive and experimental values of the i -th sample in a dataset, respectively; \bar{y}_m stands for the average data of the data points in a dataset.

3.5 Results and discussion

3.5.1 Descriptor selection and analysis

The multiple linear regression (MLR) method was utilized to pre-screen efficient descriptors from P , T , and 240 S_{EP} values of each IL in this part. To this end, the stepwise MLR procedure was executed by the SPSS software and then several reasonable MLR models with different parameters were acquired. As shown in [Figure 3.4](#), the R^2 of the model increases and the standard error (Std. Error) decreases gradually as the number of parameters adds. However, the variation tendencies become stable when the number of descriptors exceeds five. Therefore, the optimal MLR model with the R^2 (0.711) and AARD (35.29%) and the best number of parameters (5) was confirmed. [Figure 3.5](#) presents the Pearson correlation coefficients between the selected descriptors. All these correlation coefficients are less than 0.8, showing that no strong linear relationship exists between the descriptors. Therefore, each parameter is independent and carries particular physical meaning, demonstrating they can explain structure characteristics effecting the ammonia solubility in ILs (Zhao et al., 2015a).

The equation of the model was presented as Eq. (15) and the explicit information of the model and the five screened descriptors were listed in [Table 3.2](#). The S_{EP-j} symbolizes a value of surface area that within an electrostatic potential interval of 0.5

kcal/mol, for example, the $S_{EP24.75}$ stands for the surface area in the electrostatic potential range of 24.5~25 kcal/mol. The a , b , and C_i mean the coefficients of the corresponding parameters while K_0 is the intercept. The absolute value of t corresponds to the importance of the descriptor. It can be seen from Table 3.2, P and T , the first and second significant descriptors, have a critical influence on the ammonia solubility in ILs. $S_{EP24.75}$, $S_{EP7.25}$, and $S_{EP2.75}$ which are selected from 240 values show that they are more important than other S_{EP} values of ILs for predicting the ammonia solubility in ILs. Some conclusions can be deduced by qualitatively analysing the critical descriptors. For instance, these three screened S_{EP} descriptors are all positive, indicating that cations play a more vital role in the ammonia solubility in ILs than anions, which is consistent with the results of Yokozeki et al (Shi and Maginn, 2009; Yokozeki and Shiflett, 2007b, 2007a). The parameters, T , $S_{EP24.75}$, and $S_{EP7.25}$ with negative coefficients suggest the negative correlation between the parameters and ammonia solubility in ILs; conversely, the correlation between ammonia solubility in ILs and the parameters of P , $S_{EP2.75}$ is positive. Besides, it can be observed that the $S_{EP24.75}$ values of ILs with hydroxyl group are lower than those without hydroxyl group, such as the $S_{EP24.75}$ values of [EtOHMIM][BF₄] and [EMIM][BF₄] are 1.791 and 3.1035, respectively, meaning that ILs with hydroxyl group have larger ammonia absorption capacity. This situation can also be noticed in [EtOHMIM][Tf₂N] ($S_{EP24.75}=1.3292$) and [EMIM][Tf₂N] ($S_{EP24.75}=2.5842$) as well as [EtOHMIM][SCN] ($S_{EP24.75}=1.9728$) and [EMIM][SCN] ($S_{EP24.75}=2.9989$). This phenomenon can be attributed to that ammonia and hydroxyl-bearing ILs may more easily form strong hydrogen bonds, and thus increase the ammonia solubility. This is consistent with the conclusion calculated by quantum chemistry from the studies of Zhao et al (Zhao et al., 2017) and Li et al (Li et al., 2015).

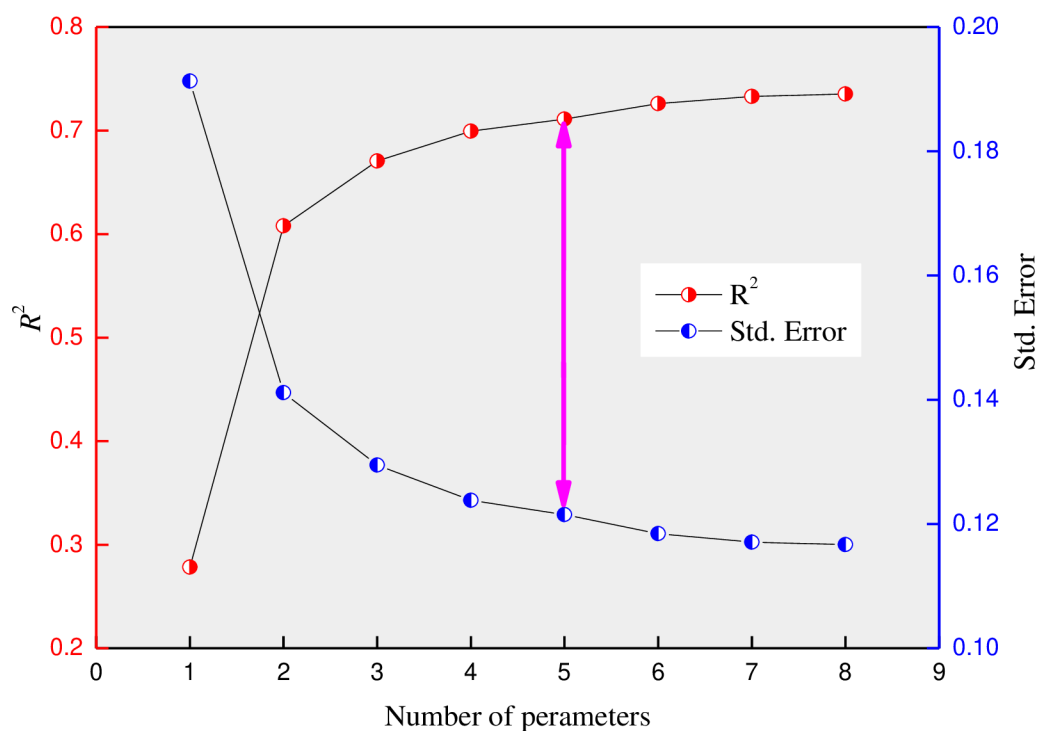


Figure 3.4 The R^2 of different MLR models for training set vs various number of parameters

$$y = P_0 + aP + bT + \sum_{i=1}^3 C_i S_{EP-j} \quad (15)$$

($n=502$, $R^2=0.711$, AARD %=35.29%)

Table 3.2 Coefficients and the t values for Eq. (15)

NO.	Parameter	Coefficient	t	
1	P	a	0.028	31.811
2	T	b	-0.007	-25.306
3	$S_{EP24.75}$	C_1	-0.117	-13.032
4	$S_{EP7.25}$	C_2	-0.063	-8.162
5	$S_{EP2.75}$	C_3	0.019	4.468
6		K_0	2.928	

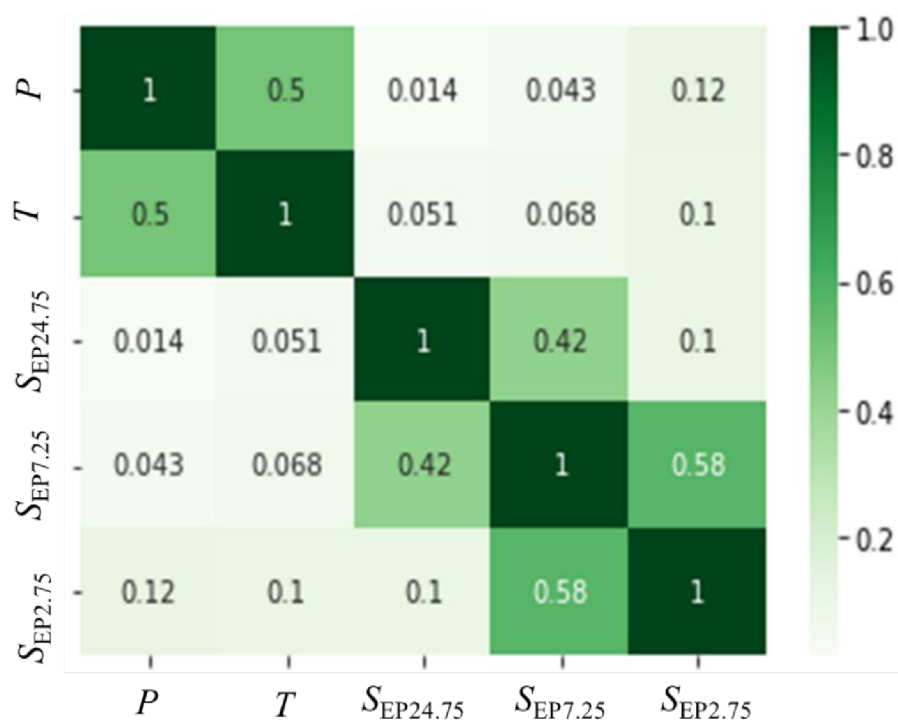


Figure 3.5 Correlation coefficients matrix of the five descriptors

3.5.2 ELM model development

From the results of the MLR model, it can be deduced that the linear relationship between input and output variables is not ideal. Hence, the nonlinear relationship of them is constructed by the ELM algorithm in the present work. The training set including 402 data points of ammonia solubility in ILs and the five input descriptors screened by MLR has been prepared for modeling. As mentioned above, the input weights and hidden bias were assigned randomly while the “sigmoid” was chosen as the activation function. To determine the number of hidden neurons, a total of 13 models were established evenly under the range of 35 to 215 neurons. Simultaneously, the R^2 and AARD of the models were gained and exhibited in Figure 3.6. It can be observed that the increasing R^2 and reducing AARD values become stabilized when the number of neurons was greater than 185. Therefore, the optimal neuron number (185) was decided, which implied that the predictive model was obtained. The most advantageous R^2 and AARD of the training set for the model are 0.9972 and 2.76 respectively. Subsequently, the 100 experimental values of the test set which were not employed in the training process were calculated by the proposed model for verifying the reliability of the model. The R^2 (0.9779) and AARD (3.69%) of the test set were calculated and AARD values for each IL were summarized in Table 3.1. All the ILs

possess rather low and acceptable AARD values.

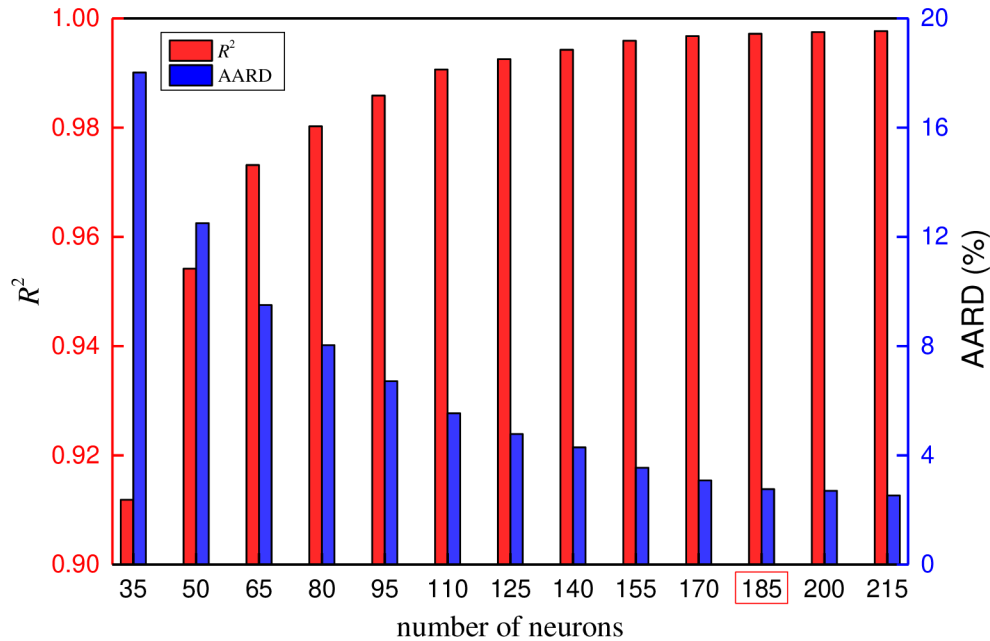


Figure 3.6 R^2 and AARD of the ELM model vs the number of neurons for the training sets

Regression plots were depicted in Figure 3.7 to compare the predictive data obtained by the ELM model for the training set and test set and the corresponding experimental values from literature. The high density of values for the training set and test set are located in the vicinity of the diagonal line, indicating that the ELM model has a good predictive ability. Linear regressions for the training and test set of the ELM model were presented as Eqs. (16) and (17), respectively.

$$y = 0.997x + 0.001, \quad R^2=0.9972 \quad (16)$$

$$y = 1.005x - 0.004, \quad R^2=0.9779 \quad (17)$$

The deviations between the predicted and experimental data were depicted in Figure 3.8. The estimated values of the total dataset with the deviations in the range of $\pm 5\%$ accounted for approximately 98%. Table 3.3 summarized various statistical parameters of different sets for the predictive model, including not only R^2 and AARD but also ARD, MSE, and RMSE. These results suggest the great goodness-of-fit, robustness, and reliability of the built model in the evaluation of ammonia solubility in

ILs.

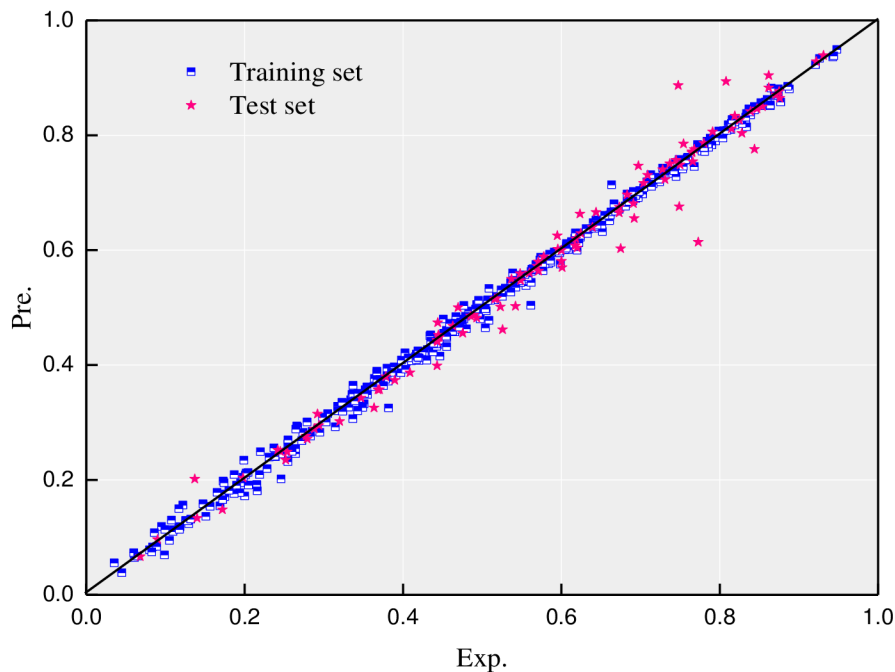


Figure 3.7 Predicted vs experimental NH₃ solubility in ILs of the ELM model in this study

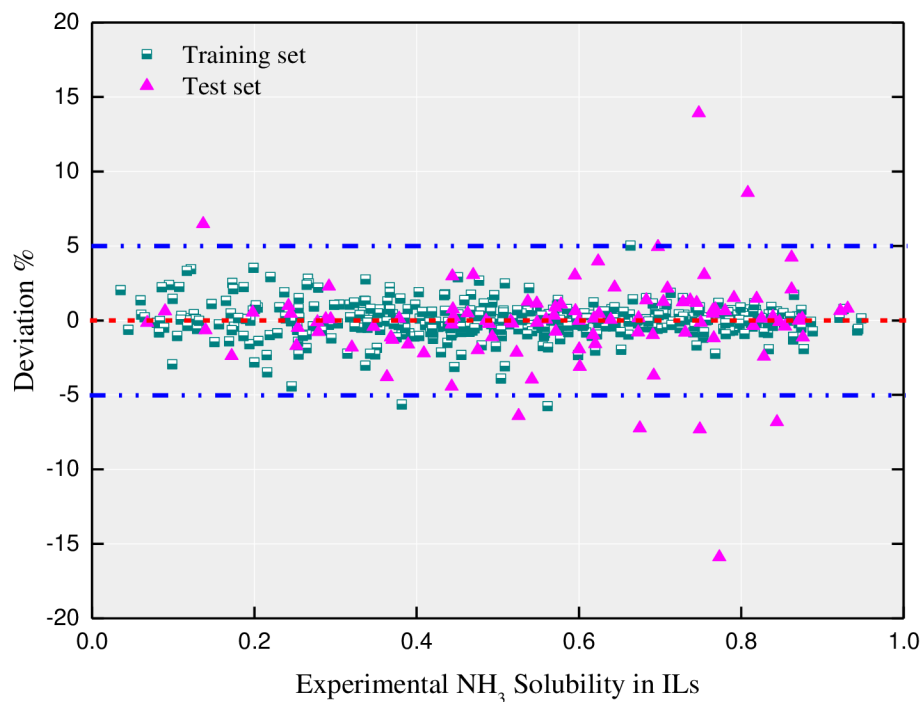


Figure 3.8 Deviations of the ELM model in this study

Table 3.3 The statistical parameters of the ELM model in this work

Dataset	No. of data points	R ²	AARD %	AAD%	MSE	RMSE
Training set	402	0.9972	2.76	0.83	0.0001	0.0120
Test set	100	0.9779	3.69	1.85	0.0010	0.0319
Total set	502	0.9937	2.95	1.03	0.0003	0.0178

3.6 Conclusion

A novel model was proposed to predict the solubility of ammonia in ILs by extreme learning machine algorithm based on P , T , and three screened S_{EP} values of ILs as input descriptors. The predictive ability of the built model was measured by the regression plots and five statistical parameters. The total dataset with the high R^2 (0.9937) and low AARD (2.95%) demonstrates a good performance of the prediction model. Statistical analyses prove that the proposed model owns the reliable accuracy to calculate ammonia solubility in ILs. Thus, the molecular descriptor, electrostatic potential surface area (S_{EP}), is proved to be of great potential in evaluating ammonia solubility in various ILs, and the ELM algorithm can be used to establish models with promising application in predicting gases absorption by ILs. This study can be significant for screening reasonable ILs to absorb ammonia and possibly make contributions for reducing ammonia emissions from industrial sources in the future.

3.7 Supplementary data

The following link is the Supplementary data to this chapter:

<https://ars.els-cdn.com/content/image/1-s2.0-S001393512030846X-mmc1.xlsx>

Chapter IV

A QSPR model for estimating Henry's law constant
of H₂S in ionic liquids by ELM algorithm

Xuejing Kang, Zuopeng Lv, Zhongbing Chen, Yongsheng Zhao

Adapted from Journal of Chemosphere, 269(2020): 128743

Contents

4.1 Abstract	57
4.2 Introduction	58
4.3 Methodologies	60
4.3.1 <i>Data collection and screening</i>	60
4.3.2 <i>Extreme learning machine (ELM) Algorithms</i>	62
4.3.3 <i>Descriptors calculation</i>	62
4.3.4 <i>Model evaluation</i>	66
4.4 Results and discussion	66
4.4.1 <i>Qualitative Analysis on HLC of H₂S in ILs</i>	67
4.4.2 <i>ELM Model development</i>	69
4.4.3 <i>ELM Model performance</i>	70
4.4.4 <i>Application domain (AD) analysis</i>	73
4.5 Conclusion	74
4.6 Supplementary data	75

4.1 Abstract

Ionic liquids (ILs) as green solvents have been studied in the application of gas sweetening. However, it is a huge challenge to obtain all the experimental values because of the high costs and generated chemical wastes. This study pioneered a quantitative structure–property relationship (QSPR) model for estimating Henry's law constant (HLC) of H₂S in ILs. A dataset consisting of the HLC data of H₂S for 22 ILs within a wide range of temperature (298.15–363.15 K) were collected from published reports. The electrostatic potential surface area (S_{EP}) and molecular volume of these ILs were calculated and used as input descriptors together with temperature. The extreme learning machine (ELM) algorithm was employed for nonlinear modeling. Results showed that the determination coefficient (R^2) of the training set, test set and total set were 0.9996, 0.9989, 0.9994, respectively, while the average absolute relative deviation (AARD%) of them were 1.3383, 2.4820 and 1.5820, respectively. The statistical parameters for the measurement of the model exhibited the great reliability, stability, and predictive power of the ELM model. The Applicability Domain (AD) of the ELM model is also investigated. It proves that the majority of ILs in the training and test sets are located in the model's AD and verifies the reliability of the model. The proposed model is potentially applicable to guide the application of ILs for gas sweetening.

4.2 Introduction

Hydrogen sulfide (H_2S) is one common acid gas generated in various industrial production processes (eg. oil-gas production, coal and biomass gasification, paper and pulp industry, wastewater treatment, etc.) (Nassar et al., 2016; Sakhaeinia et al., 2010b; Y. Wang et al., 2019). H_2S is a typical impurity for many industrial oil and gas products and is corrosive to pipelines and equipment (Li et al., 2016). Meanwhile, H_2S is an air pollutant and the main source of acid rain. It is also highly poisonous to living creatures and plants. Human health can be affected when the level of H_2S is more than 10 ppm (Potivichayanon et al., 2006; Y. Wang et al., 2019).

Traditionally, the primary techniques used to remove hydrogen sulfide are physical-chemical processes (for instance, cryogenics, adsorption, absorption, electrochemistry) (Berrouk and Ochieng, 2014; Fellah, 2016; Gholampour and Yeganegi, 2014; Liu et al., 2017; Zhang et al., 2008) and chemo-biological processes (Park et al., 2005; Son and Lee, 2005). Nevertheless, the drawbacks of these processes are obvious, such as high operating costs and chemical waste by-products need to be treated before discharge (Cao et al., 2019; Potivichayanon et al., 2006).

Ionic liquids (ILs) have emerged as a crucial class of compounds and attracted promising attention from many chemists since 1995 (Amir Hossein Jalili et al., 2012). These compounds are comprised of organic cations and organic/inorganic anions. They are also known as molten salts which are liquid within a wide-ranging temperature containing ambient conditions (Jalili et al., 2017; Rahmati-Rostami et al., 2009). Negligibly vapor pressure is one extraordinary and important property of ILs (Song et al., 2020a; J. Wang et al., 2020), meaning that they are non-volatile. They also show other remarkable characteristics like high ionic conductivities, high thermal stability, non-flammability, wide electrochemical window (Kang et al., 2018c; Kroon et al., 2006). For these reasons, ILs are deemed as environmentally benign solvents and have been investigated for acidic gas capture and separation of gaseous mixtures, such as carbon dioxides absorption, hydrogen sulfide removal of the natural and refinery gases (Bara et al., 2010; Galán Sánchez et al., 2007; Goodrich et al., 2011; Safavi et al., 2013; Song et al., 2020b).

The solubility of gases (e.g., CO_2 , H_2S) in ILs at diverse temperatures and pressures is an extremely important factor for gas absorption and separation process, and thus it is significant for the purification of natural gas (Jalili et al., 2009; Shafiei et al., 2014). The solubility and Henry's law constant (HLC) measurement of acidic gases

in various ILs in the laboratory have been carried and reported by some researchers (Ahmadi et al., 2014; K. Huang et al., 2014, 2013; Muldoon et al., 2007; Ying and Baltus, 2007; Yokozeki and Shiflett, 2010; Zhang et al., 2013). However, since the extensive number of ILs, experimental measuring methods are generally accompanied by high capital cost, high energy consumption, time-consuming problem, and so on (Eike et al., 2004). Thus, estimated approaches for predicting these properties in such systems over a wide range of conditions are of great interest. In the recent years, several QSPR (quantitative structure–property relationship) models have been explored for evaluating H₂S solubility in ILs using diverse techniques, such as gene-expression programming (GEP) (Ahmadi et al., 2014), equation of state (EoS) (Al-fnaish and Lue, 2017; Shojaeian, 2017), artificial neural network (ANN) (Amedi et al., 2016). Ahmadi et al. (Ahmadi et al., 2014) built a model by GEP methodology to evaluate the H₂S solubility in 11 ILs including 465 experimental data points. Their predictive model was obtained with the high coefficient of determination (R^2) of 0.9902 and the low mean absolute relative error (MARE) of 0.0438%. Al-fnaish and Lue (2017) described the H₂S solubility in several methylimidazolium bis(trifluoromethylsulfonyl)imide ILs by the means of the equation of state (EoS) combined Perturbed Chain Statistical Association Fluid Theory (PC-SAFT), and the AARD were successfully calculated and were between 2.76% and 6.62% for the systems of H₂S-IL. Three kinds of ANN models were proposed by Amedi et al. (2016) using three parameters, the critical temperature, critical pressure and molecular weight of ILs. With the good performance of R^2 (0.9951) and Mean Squared Error (MSE) (0.000117), the Multi-Layer Perceptron ANN (MLP-ANN) model exhibits the best predictive capability and accuracy.

Besides the algorithm, the descriptors for modeling are also a matter of extreme importance (Gao et al., 2020; Zhao et al., 2016b). The ANN approach based on the algorithm of particle swarm optimization (PSO) and backpropagation (BP) was used by Shafiei et al. (2014) to investigate the H₂S solubility in 11 ILs. In their study, five input descriptors containing pressure, temperature as well as acentric factor, critical pressure, critical temperature of ILs were employed and two models (PSO-ANN model and BP-ANN model) with good forecasting ability were acquired. Baghban et al. (2019) developed a LSSVM model for the estimation of H₂S solubility based on the descriptors of pressure, temperature and 15 types of groups of ILs. The R^2 and root mean square error (RMSE) are 0.9941 and 0.0122, respectively, indicating that the proposed model was greatly reliable and potentially useful. Zhao et al. (2016b) introduced a class of input descriptor, charge distribution area ($S_{\sigma\text{-profile}}$) of ILs, accompanied with extreme learning machine (ELM) algorithm to construct a QSPR model for correlating the solubility of H₂S in ILs.

There are some reports for evaluating the H₂S solubility in ILs but few studies to predict HLC of H₂S in ILs. Therefore, this work focused on developing reliable models for estimating the HLC of H₂S in different ILs. In our previous work, we exported two types of electrostatic potential surface area (S_{EP}) of ILs to calculate the H₂S solubility in ILs based on the ELM method (Kang et al., 2018d). The results implied the second model based on the S_{EP} of ion pairs was more accurate than that used the S_{EP} of isolated ions as input parameters. Hence, in this study, we gathered 122 data points on HLC of H₂S in 22 ILs and then utilize the S_{EP} of ion pairs (i.e., the S_{EP} of total ILs) as input descriptors linked with the intelligence algorithm of ELM to develop predictive models for estimating the HLC of H₂S in different ILs. It is worth noting that the S_{EP} of each total IL was finally divided into four different ranges as input variables in this work. Thus far, to the best of our knowledge, there is no publication aiming at the estimation on HLC of H₂S in various ILs. Meanwhile, we take the lead in attempting to use the neural network of the ELM method together with S_{EP} descriptors to establish the predictive models for HLC of H₂S in ILs.

4.3 Methodologies

4.3.1 Data collection and screening

The acquired HLC values of H₂S in 22 ILs involving 122 data points were selected from 14 references summarized in [Table 4.1](#). The ILs include 7 cations which are 1-butyl-3-methyl-imidazolium ([BMIM]⁺), 1-ethyl-3-methyl-imidazolium ([EMIM]⁺), 1-(2-hydroxyethyl)-3-methylimidazolium ([HEMIM]⁺), 1-hexyl-3-methyl-imidazolium ([HMIM]⁺), 1-octyl-3-methyl-imidazolium ([OMIM]⁺), methyldiethanolamin ([MDEAH]⁺), dimethylethanolammonium ([DMEAH]⁺), and 8 anions named Tetrafluoroborate ([BF₄]⁻), Hexafluorophosphate ([PF₆]⁻), Bis(trifluoromethylsulfonyl)imide ([Tf₂N]⁻), ethylsulfate ([EtSO₄]⁻), tris(pentafluoroethyl)trifluorophosphate ([eFAP]⁻), trifluoromethane-sulfonate ([TfO]⁻), acetate ([Ac]⁻), formate ([For]⁻). It is crucial to use reliable experimental data for building and validating the predictive models (Y. Huang et al., 2014). Therefore, we carefully analyzed and screened the data points from various literature during the acquisition process. The data analysis and selection procedure were available in Supplementary data A. The temperature and HLC ranges, volume of the ILs as well as the references were listed in [Table 4.1](#). The specific information of the experimental data and the full name of ILs were provided in Supplementary data B. The total 122 data points were randomly split into two subsets, which were the training set accounting

for 80% (98 data points) and the test set with 20% (24 data points), for modeling and validating respectively.

Table 4.1 Temperature, volume, HLC and AARD % values of the ILs used in this study

No.	IL	T range (K)	HLC (MPa)	AARD % (ELM)	No. of data points	Refs.
1	[BMIM][BF ₄]	303.15-343.15	1.55-3.34	0.6464	5	(Jalili et al., 2009)
		303.15-343.15	1.86-3.38	2.8905	5	(Jalili et al., 2009)
2	[BMIM][PF ₆]	298.15-343.15	1.43-3.26	4.5631	3	(Jou and Mather, 2007)
3	[BMIM][Tf ₂ N]	303.15-343.15	1.37-2.66	2.4289	5	(Jalili et al., 2009)
4	[BMIM][TfO]	303.15-343.15	1.06-2.60	1.5924	5	(Jalili et al., 2017)
5	[EMIM][EtSO ₄]	303.15-353.15	6.07-13.30	0.3964	6	(Jalili et al., 2010)
6	[EMIM][PF ₆]	333.15-363.15	3.74-5.29	0.4724	5	(Sakhaeinia et al., 2010a)
7	[EMIM][Tf ₂ N]	303.15-353.15	1.48-3.16	0.9932	6	(Sakhaeinia et al., 2010a)
8	[EMIM][TfO]	303.15-353.15	1.36-3.60	0.5877	6	(Nematpour et al., 2016)
9	[EMIM][eFAP]	303.15-353.15	1.53-3.50	0.572	6	(Jalili et al., 2013)
10	[EMIM][BF ₄]	298.15-353.15	1.78-4.86	1.551	7	(Jalili et al., 2019)
11	[HEMIM][BF ₄]	303.15-353.15	3.13-7.95	1.0925	6	(Shokouhi et al., 2010)
12	[HEMIM][PF ₆]	303.15-353.15	2.77-5.15	1.1437	6	(Sakhaeinia et al., 2010b)
13	[HEMIM][Tf ₂ N]	303.15-353.15	1.89-3.69	0.7258	6	(Sakhaeinia et al., 2010b)
14	[HEMIM][TfO]	303.15-353.15	1.93-4.24	1.1761	6	(Sakhaeinia et al., 2010b)
15	[HMIM][BF ₄]	303.15-343.15	1.25-2.57	1.8781	5	(Rahmati-Rostami et al., 2009)
16	[HMIM][Tf ₂ N]	303.15-353.15	1.25-2.86	0.8131	6	(Amir Hossein Jalili et al., 2012)
17	[OMIM][Tf ₂ N]	303.15-353.15	0.99-1.95	1.5056	6	(Amir Hossein Jalili et al., 2012)

No.	IL	T range (K)	HLC (MPa)	AARD % (ELM)	No. of data points	Refs.
18	[OMIM][PF ₆]	303.15-353.15	1.22-2.55	1.5621	6	(Safavi et al., 2013)
19	[MDEAH][Ac]	303.2-333.2	0.55-1.39	2.9714	4	(K. Huang et al., 2014)
20	[MDEAH][For]	303.2-333.2	1.15-2.32	2.0365	4	(K. Huang et al., 2014)
21	[DMEAH][Ac]	303.2-333.2	0.35-1.02	6.6141	4	(K. Huang et al., 2014)
22	[DMEAH][For]	303.2-333.2	0.59-1.39	2.2552	4	(K. Huang et al., 2014)

4.3.2 Extreme learning machine (ELM) Algorithm

ELM, a class of novel training computational method underlying the neural network of single-hidden-layer feedforward, was created by Huang (Huang et al., 2006b, 2004; Nabipour et al., 2020). It has been used with increasing interest from both industry and academia due to its least human intervention, remarkable generalization performance, and extraordinarily fast learning speed (Cui et al., 2018). In recent years, ELM has been widely applied in various fields, particularly for forecasting issues, such as electricity load (Ertugrul, 2016), wind speed (Liu et al., 2015), oil price (Yu et al., 2016). It has also been successfully introduced to establish models for predicting diverse ionic liquids properties in our previous studies (Kang et al., 2018a, 2018e, 2017; Zhu et al., 2019), where the detailed theory of ELM was introduced and described. In brief, the ELM network structure includes three units which are input variables, hidden neurons, and output variables. In this study, the input variables are volume, the S_{EP} values of total ILs and the temperature. The sole target variable (output variable) is the HLC of H₂S in ILs. The activation function between the input and hidden neurons is sigmoid function while the weights and bias between them were randomly generated, and a liner function connected the hidden and output units.

4.3.3 Descriptors calculation

It is potentially useful to assess the properties of compounds by S_{EP} as rich information from the electronic level is available by the electrostatic potentials, and the S_{EP} descriptor represents the electrostatic potential on the molecule surface area in various ranges. To obtain the S_{EP} values of ILs used in this study, the optimal structures of the 22 ILs were geometrically calculated by the software of Gaussian 09 (B.01

version) at the level of B3LYP/6-31 ++G (d, p) (Ortiz et al., 2009). The eventually optimized structures of ILs possess the local minimal energy and no imaginary frequency. Then the electrostatic potential surfaces of the optimized ILs were mapped to the Vander Waals surface (electron density at the isovalue of 0.001 e/Bohr³) (Bader et al., 1987) by Gaussian 09 B.01 (Figure 4.1). The different colors showed the varied electrostatic potential on the molecule surface of ILs in Figure 4.1. Meanwhile, the S_{EP} values ranging from -60 kcal/mol to 60 kcal/mol of each IL were obtained with the stepwise of 0.5 kcal/mol by the Multify 3.6 (dev) package (Lu and Chen, 2012) based on the Vander Waals surface, while the molecule volumes of ILs were simultaneously calculated as well.

Figure 4.2 plotted the electrostatic potentials versus corresponding surface areas of three ILs composed of cations with different alkyl chain length ([EMIM]⁺, [BMIM]⁺, and [HMIM]⁺) and consistent anion ([BF₄]⁻). It reveals that ILs with the same anions exist very close to S_{EP} values in the negative range. Figure 4.3 showed the S_{EP} of [BMIM][BF₄], [BMIM][PF₆] and [BMIM][Tf₂N], indicating that the S_{EP} values of ILs are pretty approximate in the positive region when including the same cation. It needs to mention that the various colors in Figure 4.2 and Figure 4.3 have the similar extend with the Figure 4.1. The color-coding of red-yellow-green-blue indicates the increase in the electrostatic potential value, meaning that the red and blue regions are symbols of negative and positive potential, while the green regions indicate neutral potential. All the S_{EP} values for each IL were finally divided into four ranges as showed in Figures 4.2 and 4.3, and the sum of S_{EP} values in each range became our ultimate descriptors are S_{EP1} , S_{EP2} , S_{EP3} , S_{EP4} , respectively.

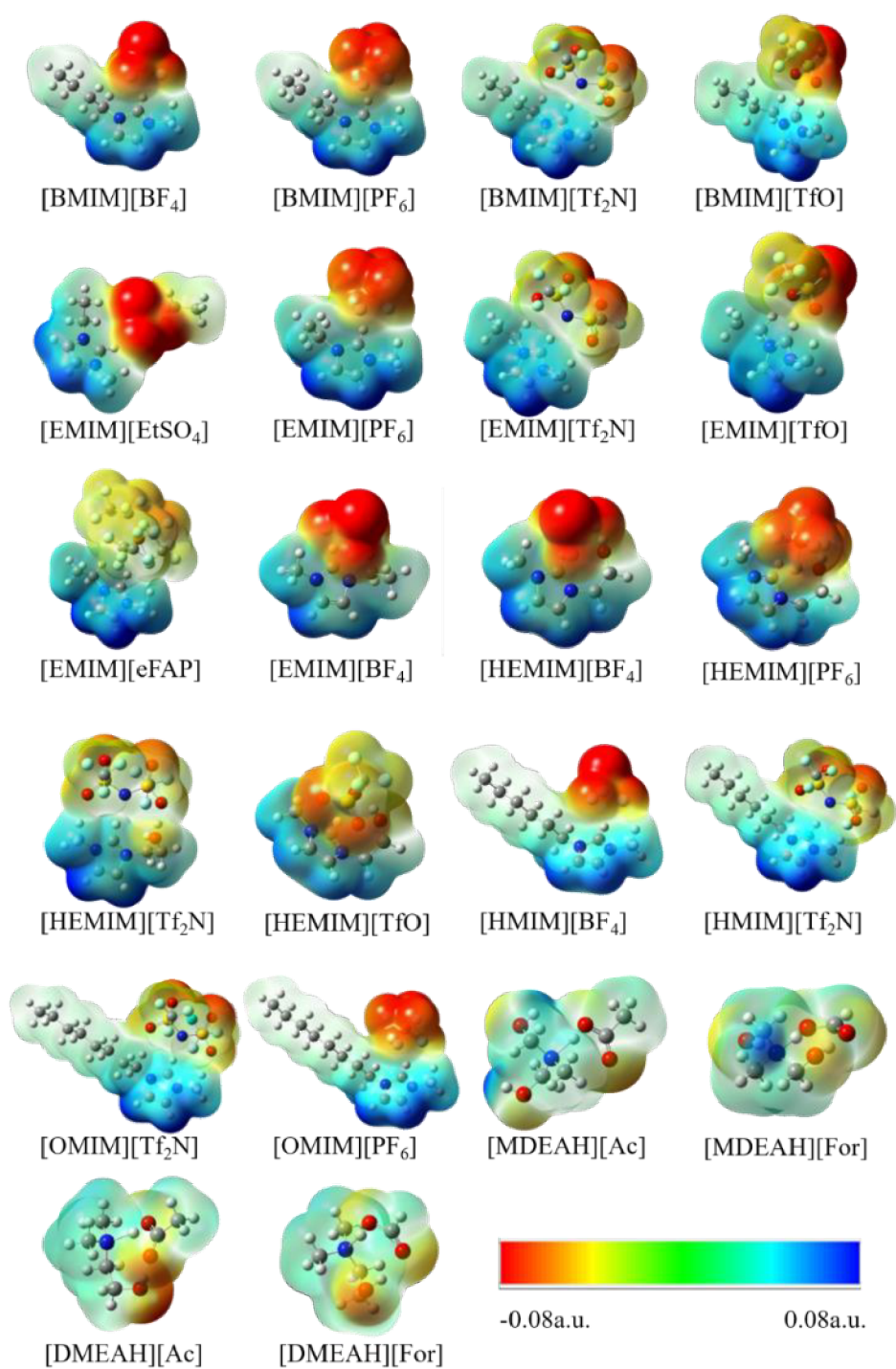


Figure 4.1 Electrostatic potential surface mapped electron density with an isovalue 0.001 e/Bohr^3 of the ILs used in this study

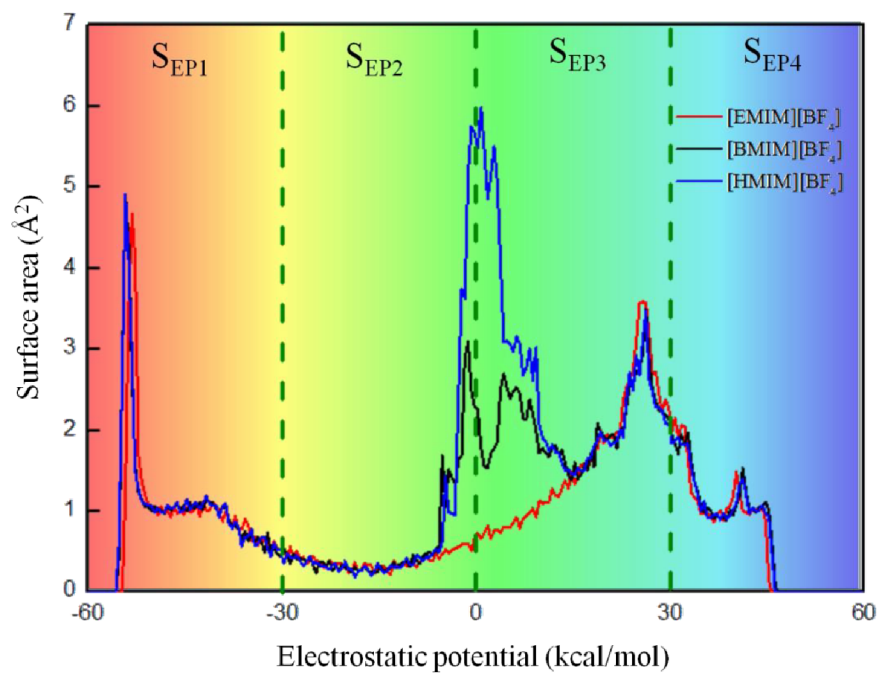


Figure 4.2 Electrostatic potentials versus corresponding surface areas of ILs with the same anion

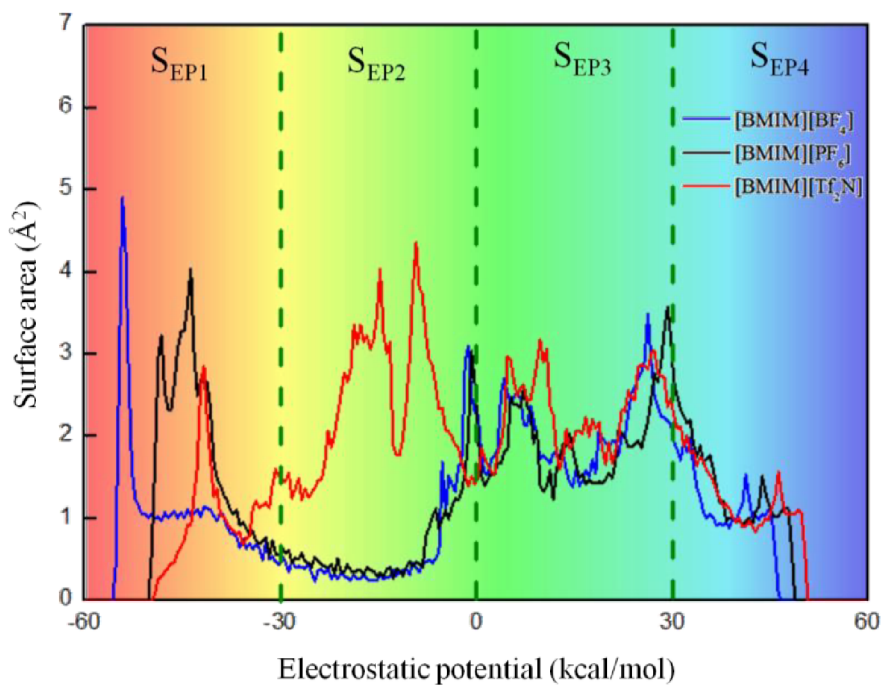


Figure 4.3 Electrostatic potentials versus corresponding surface areas of ILs with the same cation

4.3.4 Model evaluation

To measure the performance of the proposed models, a graphical approach of the regression plot is employed. Besides, seven metrics, namely determination coefficient (R^2), average absolute deviation (AAD), average absolute relative deviation (AARD), relative deviation (RD), absolute relative deviation (ARD), the mean-square error (MSE) accompanied with root mean square error (RMSE), are utilized. The detailed expressions of them are listed as follows:

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2}{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2} \quad (1)$$

$$AARD (\%) = 100 \times \sum_{i=1}^{N_p} \left| \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \right| / N_p \quad (2)$$

$$RD (\%) = 100 \times \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \quad (3)$$

$$ARD (\%) = 100 \times \left| \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \right| \quad (4)$$

$$AAD (\%) = 100 \times \sum_{i=1}^{N_p} |y_i^{cal} - y_i^{exp}| / N_p \quad (5)$$

$$MSE = \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p \quad (6)$$

$$RMSE = \sqrt{\sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p} \quad (7)$$

where y_i^{cal} is the predictive value of HLC of H₂S in the i -th IL, y_i^{exp} represents the corresponding experimental value; the average data and the sample number of a dataset were expressed by \bar{y}_m and N_p , respectively.

4.4 Results and discussion

4.4.1 Qualitative Analysis on HLC of H₂S in ILs

To build reliable models, the four S_{EP} parameters, the volume (V_m) of all the gathered ILs and the temperature (T) as input descriptors were qualitatively analyzed through multiple linear regression (MLR) model. The MLR procedure was carried out via the SPSS software, and the acquired MLR model was described as Eq. (8), where the a , b , C_i are the coefficients and P_0 is the intercept. The specific coefficients of the six input parameters and t values are given in Table 4.2. The absolute value of t symbolizes the importance of the parameter, meaning that the larger the absolute t value, the more vital the parameter. Therefore, the S_{EP1} , S_{EP2} are the first and second important descriptors, respectively, followed by the T and V_m . The HLC of H₂S in ILs was positively influenced by S_{EP1} , S_{EP2} , S_{EP3} , S_{EP4} and T , while negatively affected by V_m . This implies that the HLC of H₂S in ILs declines as the four S_{EP} descriptors increase, demonstrating the solubility of H₂S in ILs grows with the rising S_{EP1} , S_{EP2} , S_{EP3} , S_{EP4} . For instance, when the ILs possessing the same cation [BMIM]⁺, the S_{EP1} values of the ILs with [BF₄]⁻, [PF₆]⁻, [Tf₂N]⁻, [TfO] accord to the order of $S_{EP1-[BMIM][PF6]} > S_{EP1-[BMIM][BF4]} > S_{EP1-[BMIM][TfO]} > S_{EP1-[BMIM][Tf2N]}$ while the corresponding experimental HLC values of H₂S present the similar phenomenon, i.e., $HLC_{[BMIM][PF6]} > HLC_{[BMIM][BF4]} > HLC_{[BMIM][TfO]} > HLC_{[BMIM][Tf2N]}$, as shown in Figure 4.4. However, results do not always obey this rule. Although the S_{EP1} values of ILs with [HEMIM]⁺ cation have the same sequence, HLC of H₂S exhibited a different order, $HLC_{[HEMIM][BF4]} > HLC_{[HEMIM][PF6]} > HLC_{[HEMIM][TfO]} > HLC_{[HEMIM][Tf2N]}$, as presented in Figure 4.5. This circumstance was reported by Zhao et al. (Zhao et al., 2016a) as well. Moreover, growing T and decreasing V_m will raise the HLC of H₂S in ILs, which means that the solubility of H₂S in ILs when the T increases and V_m reduces. This indicates that the H₂S solubility in ILs has the typical features of physical solvents (Jalili et al., 2009; Sakhaeinia et al., 2010a).

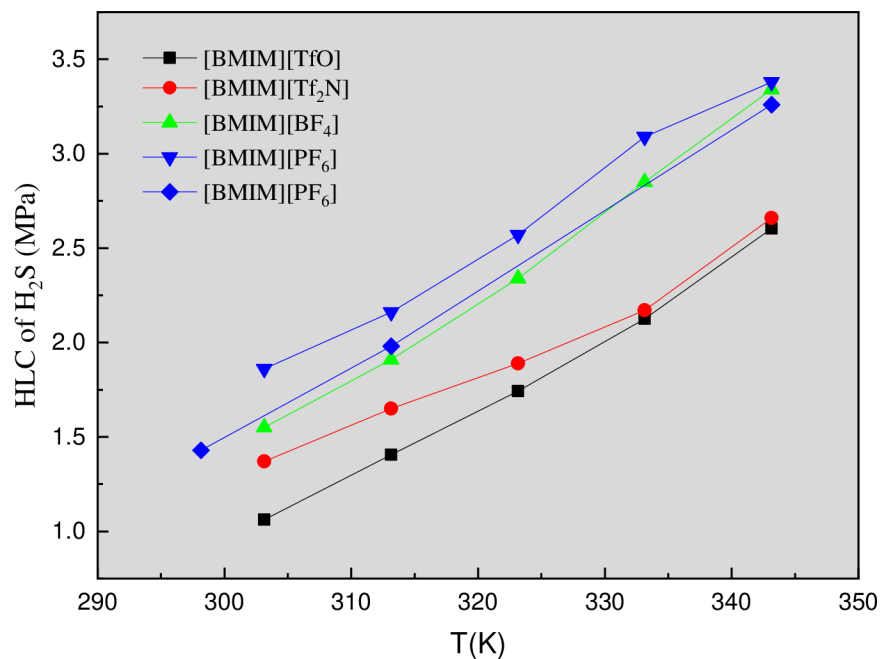


Figure 4.4 Experimental HLC of H₂S in different ILs including the same cation ([BMIM]⁺)

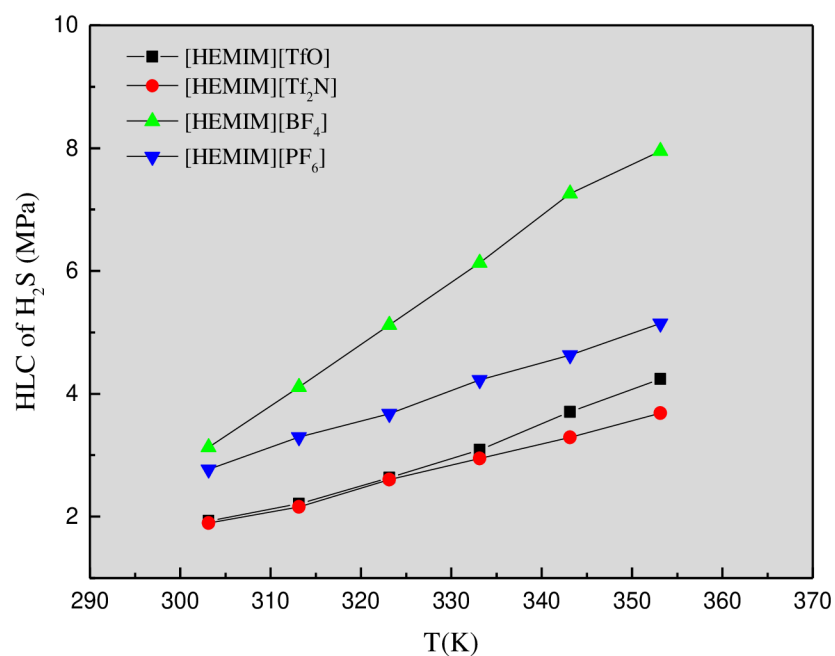


Figure 4.5 Experimental HLC of H₂S in different ILs including the same cation ([HEMIM]⁺)

Additionally, the negative potential regions of S_{EP1} and S_{EP2} have larger absolute values of t than the positive regions of S_{EP3} and S_{EP4} , showing that the anions of the ILs play a more important role than cations for the HLC of H₂S in ILs. This conclusion was also reported by Zhao et.al (2016a) who inferred that the cations have less influence on the solubility of H₂S than anions. The obtained MLR model with $R^2=0.749$ and AARD =37.05% were less accurate and inspired us to further develop a non-linear model.

$$y = P_0 + aV_m + bT + \sum_{i=1}^4 C_i S_{EPi} \quad (8)$$

$$(n=122, R^2=0.749, AARD =37.05\%)$$

Table 4.2 Coefficients and the t values for Eq. (8)

NO.	Parameter	Coefficient	t	
1	V_m	a	-0.066	-6.537
2	T	b	0.050	8.478
3	S_{EP1}	C_1	0.111	9.171
4	S_{EP2}	C_2	0.087	10.443
5	S_{EP3}	C_3	0.049	3.517
6	S_{EP4}	C_4	0.070	2.146
7		P_0	-14.178	-7.385

4.4.2 ELM Model development

As mentioned in section 2.2, the input and output variables as well as the activate function between different layers have been determined. To gain the ELM models for predicting the HLC of H₂S in ILs, the number of hidden neurons was the only factor that needs to be decided. Thus, the training set of the data points were utilized to calculate a series of ELM models under a different number of hidden neurons by Matlab software. The R^2 and AARD of different ELM models were shown in Figure 4.6. It can be seen that the R^2 values were grown dramatically as the number of hidden neurons increased while the AARD values tendency was in inverse. However, both of the R^2 and AARD values tended to be stable after the hidden neuron number exceeded 65. As a result, the optimal ELM model was established under the hidden neuron number of 65.

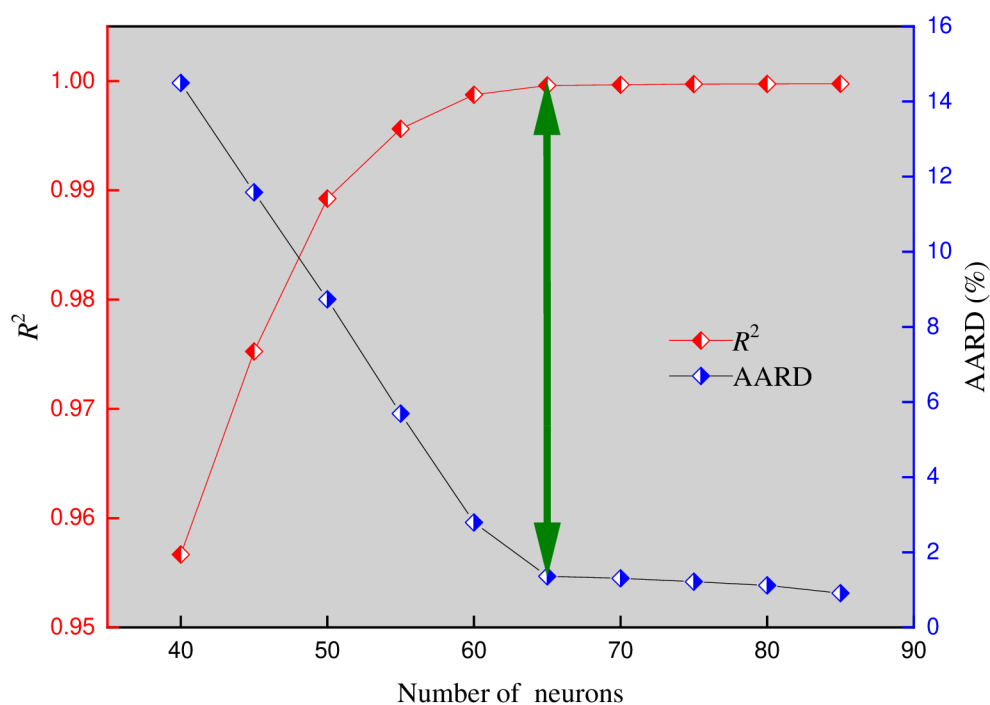


Figure 4.6 R^2 and AARD% of the ELM models versus the number of hidden neurons for the training set

4.4.3 ELM Model performance

When the best ELM model was confirmed, the R^2 and AARD of the model for the training set were 0.9996 and 1.3383%, respectively. Subsequently, the AAD, MSE and RMSE of the model were calculated and summarized in Table 4.3. Besides, the good performance of the test set with larger R^2 (0.9989) and smaller AARD (2.4820%) verified that the model was established without overfitting. The metrics of the test set, AAD (0.0534), MSE (0.0051), RMSE (0.0714) listed in Table 4.3 showed the fantastic precision and predictive power of the model.

Table 4.3 The statistical parameters of ELM model in this work

Dataset	No. of data points	R^2	AARD %	AAD%	MSE	RMSE
Training set	98	0.9996	1.3383	0.0275	0.0017	0.0407
Test set	24	0.9989	2.4820	0.0534	0.0051	0.0714
Total set	122	0.9994	1.5820	0.0330	0.0024	0.0489

Figure 4.7 compared the predictive HLC of H₂S in ILs by ELM and the experimental HLC values from the literature. The experimental values were close to the predictive ones for both training set and test set, demonstrating the great forecasting capability of the proposed ELM model. Based on the statistical theory, the linear regressions of the ELM model were created for the training and test dataset as Eq. (9) and (10), respectively.

$$y = 0.99999x - 0.00001, \quad R^2=0.9996 \quad (9)$$

$$y = 0.98966x + 0.01907, \quad R^2=0.9989 \quad (10)$$

The RD values and ARD in different regions of the data points for training and test sets were depicted in Figures 4.8 and 4.9, respectively. The majority of RD values for the whole dataset were within $\pm 5\%$ while the while only 6 RD values including 4 data from the training subset and 2 data from the test subset escaped this range (Figure 4.8). The different ranges of ARD are 51.64% (63 data points) in the scale of 0-1%, 25.41% (31 data points) in 1-2%, and 18.03% (22 data points) in 2-5% (Figure 4.9). These results of the entire dataset including the five values of the metrics in Table 4.3 proved the outstanding reliability, robustness and stability of the developed model.

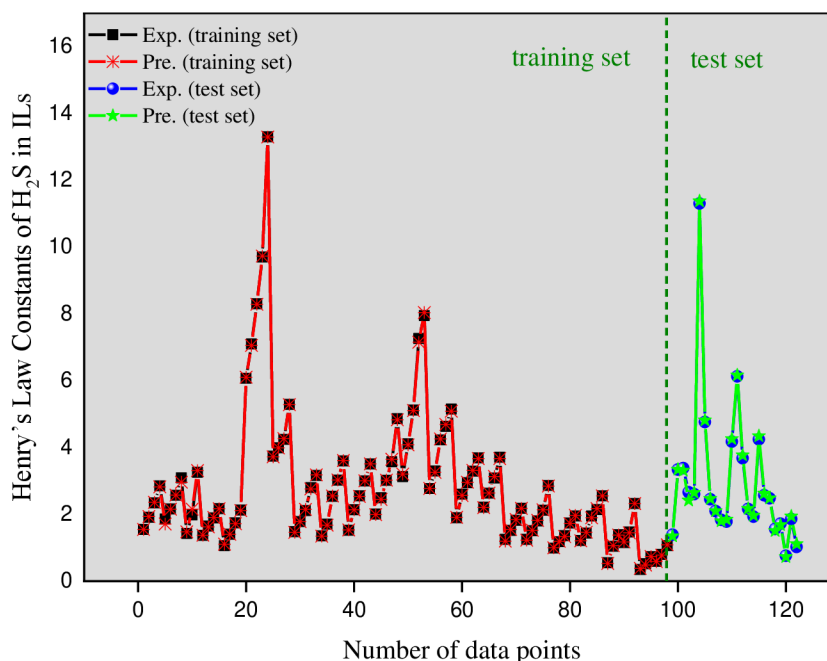


Figure 4.7 Predicted vs experimental HLC of H₂S in ILs for the ELM model

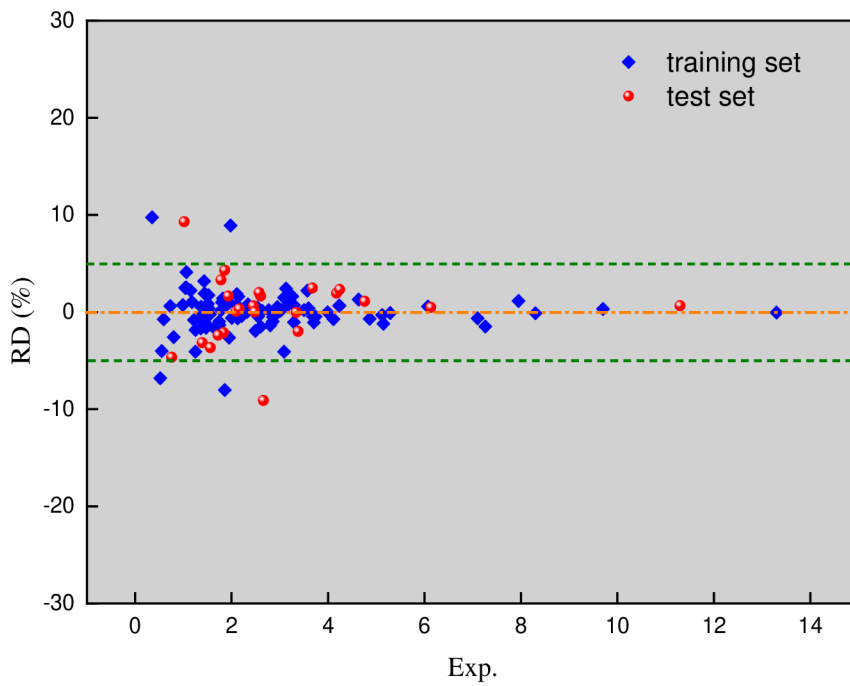


Figure 4.8. Percent of predicted values for the ELM model in different RD ranges

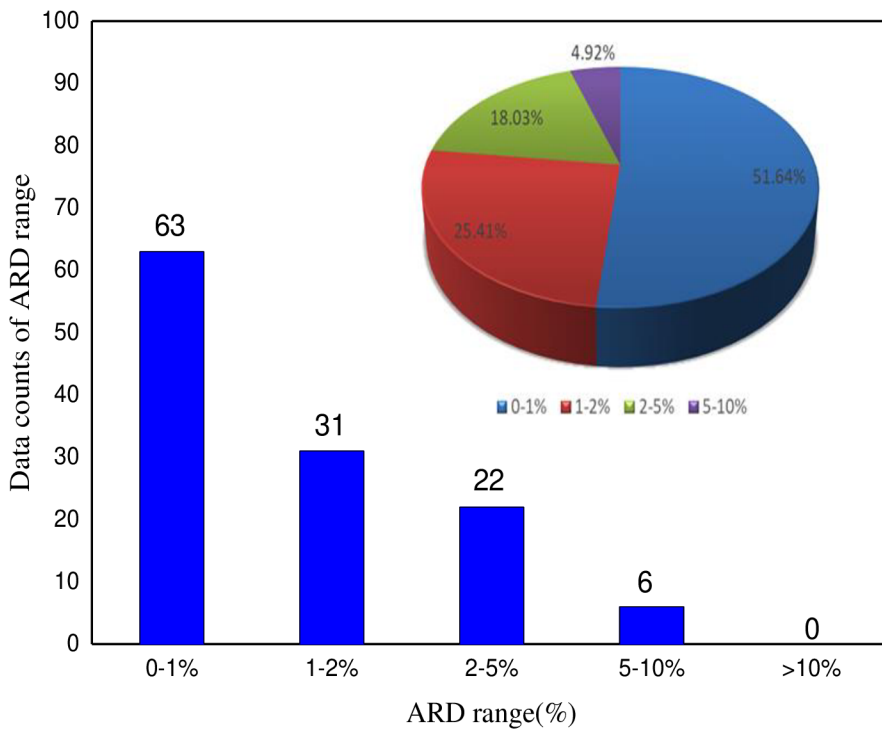


Figure 4.9 ARD% distribution in different ranges for the entire dataset

4.4.4 Application domain (AD) analysis

The AD of a model is essential to be defined (Fatemi and Izadiyan, 2011) since the AD allows for assessing the reliability of predictive data and thus to verify the usefulness of a model for new compounds (Rybińska-Fryca et al., 2020). The theoretical spatial domain of the AD is generated by the nature of the training set, the modeled responses and the molecular descriptors (Gharagheizi et al., 2012b). The leverage approach (Williams graph), i.e., the leverage (Hat values) vs standardized residuals (σ), is used to describe the AD in this work. The Hat values (h) are defined as follows:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i=1, 2, \dots, n) \quad (11)$$

where x_i is the i th descriptors in a row vector while X means the descriptor matrix of the training set. The critical value (h^*) is calculated by Eq. (12):

$$h^* = \frac{3(f + 1)}{n} \quad (12)$$

where f is the number of input descriptors, and the n is the number of samples in the training set.

The AD of a model is located in the region of $0 \leq h \leq h^*$ and $-3 \leq \sigma \leq 3$. ILs with $\sigma > 3$ are wrongly predicted and ILs with $h > h^*$ are structurally influential chemicals. It is worth noting that samples with $h > h^*$ and $-3 \leq \sigma \leq 3$ are out of AD but still successfully predicted by the model (Gharagheizi et al., 2012b).

Williams plot of the proposed ELM model is presented in [Figure 4.10](#) to define the application domain. The standardized residuals of the majority of ILs in training and test sets are less than 3 and all the ILs are with $h < h^*$ ($h^*=0.214$), showing that these ILs are in the AD of the model. However, three chemicals (6,9,12) from training set and one (18) from test set are outside of the AD, indicating that the predicted values are unreliable. As the h values of these chemicals (6,9,12 and 18) are less than the h^* , the four outliers are probably attributed to the incorrect experimental data instead of the molecular structure.

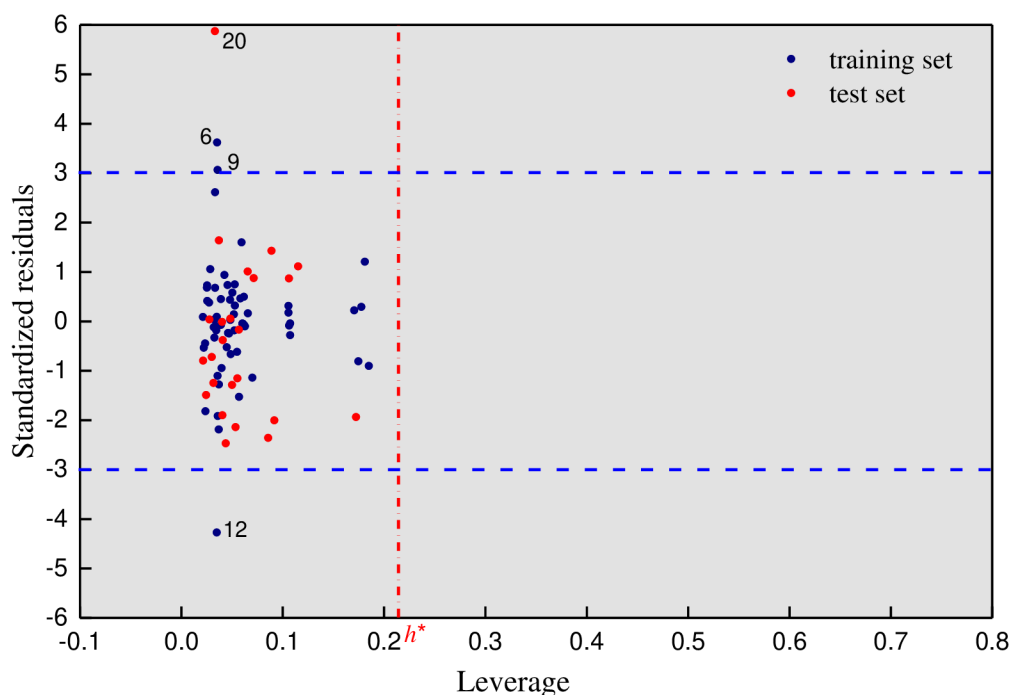


Figure 4.10 Williams plot of the ELM model for the training and test sets

4.5 Conclusion

In this work, 122 data points of HLC of H₂S in 22 ILs were gathered from fourteen references. Four descriptors regarding the electrostatic potential surface area of these ILs were calculated and allocated. Based on the parameters of temperature, molecule volume and S_{EP} of the training set ILs, the first ELM model for forecasting the HLC of H₂S in various ILs was developed with R^2 (0.9996) and AARD (1.3383%). Compared to the MLR method, the proposed ELM model was more accurate and reliable. The great predictive power of the model was verified by the performance of the test subset. The $R^2=0.9994$, AARD=1.5820% and RMSE=0.489 of the total set were acquired, revealing the excellent robustness and stability of the forecasting model. Most of ILs fallen down the AD of the ELM model turns out that the predictive model is relatively reliable. The method acquired in the current work is potentially applied to screen and design suitable ILs for H₂S absorption and gas sweetening.

4.6 Supplementary data

The following link is the Supplementary data to this chapter:

<https://ars.els-cdn.com/content/image/1-s2.0-S0045653520329416-mmc2.xlsx>

Chapter V

Assessing the ecotoxicity of ionic liquids on *Vibrio fischeri* using electrostatic potential descriptors

Xuejing Kang, Zuopeng Lv, Zhongbing Chen, Yongsheng Zhao

Adapted from Journal of Hazardous Materials 397 (2020): 122761

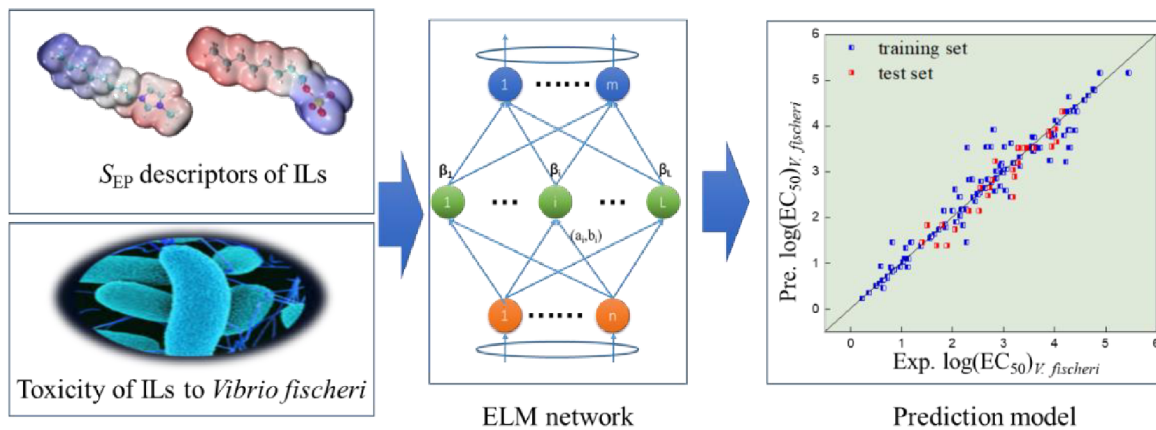
Contents

5.1 Abstract	79
5.2 Introduction	81
5.3 Dataset and method	83
5.3.1 <i>Ionic liquid data set and toxicity values</i>	83
5.3.2 <i>Extreme learning machine (ELM)</i>	83
5.3.3 <i>Electrostatic potential surface area (S_{EP}) descriptors</i>	84
5.3.4 <i>Model validation</i>	87
5.4 Results and discussion	89
5.4.1 <i>Screening and impact analysis of descriptors</i>	89
5.4.2 <i>Extreme learning machine (ELM) model establishment</i>	91
5.4.3 <i>The applicability domain (AD) of ELM</i>	95
5.4.4 <i>Comparison of different QSPR models for the ILs toxicity on <i>Vibrio fischeri</i></i>	96
5.5 Conclusions	97
5.6 Supplementary data	98

5.1 Abstract

Ionic liquids (ILs) have attracted increasing attention both in the scientific community and the industry in the past two decades. Their risk of being inevitable released to ecosystem lights up the urgent research on their toxicity to the environment. To reduce the time and capital consumption on testing tremendous ILs ecotoxicity experimentally, it is essential to construct predictive models for estimating their toxicity. The objective of this study is to provide a new approach for evaluating the ecotoxicity of ILs. A comprehensive ecotoxicity dataset for *Vibrio fischeri* involving 142 ILs, was collected and investigated. The electrostatic potential surface areas (S_{EP}) of separate cations and anions of ILs were firstly applied to develop predictive models for ecotoxicity on *Vibrio fischeri*. In addition, an intelligent algorithm named extreme learning machine (ELM) was employed to establish the predictive model. The squared correlation coefficients (R^2), the average absolute error (AAE%) and the root-mean-square error (RMSE) of the developed model are 0.9272, 0.2101 and 0.3262 for the entire set, respectively. The proposed approach based on the high R^2 and low deviation has remarkable potential for predicting ILs ecotoxicity on *Vibrio fischeri*.

Graphical abstract:



5.2 Introduction

Ionic liquids (ILs) have attracted extensive attention due to their remarkable properties in both industrial and academic fields for recent two decades. They generally consist of organic cations and inorganic or organic anions and are molten salts at near-ambient temperature (Delgado-Mellado et al., 2019). With the characteristics of negligible vapor pressure, non-volatility, non-flammability, high ionic conductivity, high electrochemical, thermal, and chemical stability (Blanchard et al., 1999; Isosaari et al., 2019; Ranke et al., 2007b; Tshibangu et al., 2011), they are considered to be powerful solvents, electrolytes and sealants. They have also been used in the fields of organic synthesis, biology, as well as extraction and separation of chemical substances (Brennecke and Maginn, 2001; Fredlake et al., 2004; Ghanem et al., 2018; Kang et al., 2017; Plechkova and Seddon, 2008). Due to the extraordinary non-volatility of ILs, they will not produce air pollutants, however they are toxic to the water environment and soil according to numerous reports. This is because that they usually are soluble in water and can be released into soils or aquatic ecosystem (Ghanem et al., 2017; McFarlane et al., 2005; Yan et al., 2019). For example, some living organisms can be affected by the ILs, such as green algae (Latała et al., 2009; Pretti et al., 2009), *Caenorhabditis elegans* (Peng et al., 2018), and *Vibrio fischeri* (Jafari et al., 2019; Stolte et al., 2007b; Ventura et al., 2012). Additionally, *leukemia rat cell line* (Ranke et al., 2007b) and acetylcholinesterase enzyme (Ranke et al., 2007b) can also be affected. Among them, *Vibrio fischeri* is a well-known type of marine luminescent bacteria with short reproductive cycle and the toxicity inference for *Vibrio fischeri* may be extrapolated for a wide variety of aquatic organisms (Jafari et al., 2019; Kaiser, 1998), and thus can be effectively applied for toxicological risk assessment.

In principal, it is estimated that more than 10^6 pure ILs can be experimentally synthesized and binary combinations of them can reach 10^{12} , while ternary systems are expected to be 10^{18} (Luis et al., 2010). Hence, assessing the toxicity of ILs through experimental methodology is time-consuming, costly and almost impossible to accomplish. Therefore, it is urgent to establish evaluation procedures to estimate the toxicity of ILs. The quantitative structure-activity relationship (QSAR) method has been employed to estimate different properties of ILs (Belvèze, 2004; Jastorff et al., 2007; Katritzky et al., 2002; Palomar et al., 2007; Torrecilla et al., 2008a; Yan et al., 2008), such as viscosity, heat capacity, melting point, Henry's law constant, water solubility (Kang et al., 2018c; Ranke et al., 2009). Several prediction models have been reported for estimating the ecotoxicity (EC_{50} *Vibrio fischeri*) of ILs using the technique of QSAR. A study on ILs toxicity towards *Vibrio fischeri* by Yan et al. (2015) proposed

a multiple linear regression (MLR) model employing QSAR method. In their work, 157 ILs were selected and the descriptors of topological indexes on the basis of character vector of atom and distance matrix of atom position were applied to construct the model, and good results with $R^2=0.908$ were obtained indicating the model can accurately evaluate the toxicity of ILs on *Vibrio fischeri*. Ghanem et al. (2017) built linear and non-linear models by QSAR method based on an ecotoxicity dataset of 110 ILs towards *Vibrio fischeri*. They selected five σ -profile descriptors for developing the MLR and non-linear (MLP) models. The well-performed correlation coefficients (R^2) were obtained, with 0.906 and 0.961, respectively. Jafari et al. (2019) utilized the elemental compositions of anions and cations as descriptors. They collected 122 ILs including 163 toxicity data points on *Vibrio fischeri* as the training set to develop a novel MLR model and 40 data points for 31 ILs as the test set. The values of R^2 were 0.877 and 0.850 for the training and test sets, respectively. Other 34 ILs containing 47 experimental data points were used to assess the newly acquired MLR model which turned out to be a highly reliable model with R^2 (0.895).

Our previous studies indicated that the selection of descriptors is crucial for building QSAR models to predict the toxicity of ILs (Cao et al., 2018; Zhao et al., 2015b, 2014). The electrostatic potential surface area (S_{EP}) for a molecule is defined as the surface areas of a molecular in different electrostatic potential intervals, showing a wealth of molecular information at electronic level (Kang et al., 2018d; Murray and Politzer, 2011b). Meanwhile, an advanced algorithm named extreme learning machine (ELM) has been successfully applied to evaluate a couple of properties of ILs in our previous work (Kang et al., 2018e, 2018b; Zhu et al., 2019). The best feature of ELM is that it is faster than traditional neural networks, especially single-hidden-layer feedforward neural networks. Beyond, ease of use, higher generalization performance, suitable for almost all nonlinear activation functions and fully complex activation functions are also the salient features of ELM (Huang et al., 2006a).

Therefore, the aims of this work are: (1) to calculate the electrostatic potential surface areas of collected ILs and screen reasonable S_{EP} descriptors for modeling; (2) to establish a novel and reliable model applying ELM method combined with S_{EP} descriptors; (3) to assess the ecotoxicity of ILs towards *Vibrio fischeri* using the established model and provide a new assessment approach for ILs toxicity.

5.3 Dataset and method

5.3.1 Ionic liquid data set and toxicity values

In this study, the experimental ecotoxicity data points of 142 ILs on *Vibrio fischeri* were collected from literature. These ILs consist of 54 cations and 28 anions. The toxicity data values ($\log(\text{EC}_{50})_{V. fischeri}$) were the logarithm form of half maximal effective concentration converted from EC_{50} values (μM) for *Vibrio fischeri*. All ecotoxicity data points were divided into two randomized groups. 80% of the data involved 113 ILs were utilized as a training set to build a predictive model, and the remaining set (29 ILs) was employed to measure the validity and stability of the model. The more detailed information of the selected ILs, such as the full name and abbreviation of ILs and ions, the $\log(\text{EC}_{50})_{V. fischeri}$ values, and the references for the data, was provided in the Supplementary data.

5.3.2 Extreme learning machine (ELM)

Extreme learning machine (ELM), a learning technique for the purpose of training single-hidden-layer feedforward neural networks, was firstly introduced by Huang (Huang et al., 2004). It guarantees faster learning accuracy than traditional learning algorithms as mentioned above. The network structure of ELM for the present work was shown in [Figure 5.1](#). It can be observed that input weights (a_i) (between the input neurons and the hidden neurons) and hidden biases (b_i) are selected at random, while the output weights (β_i) combining the hidden neurons and the output neurons can be determined by Moore–Penrose generalized inverse (Zhu et al., 2005). Mapping the data of the input layer from its original space to the feature space of the ELM is the activation function in this algorithm. ELM has been widely used in regression issues of various fields including computer vision, bioinformatics, earth science, and environmental science (Huang et al., 2015). It also has been applied for the prediction modeling of chemical properties.

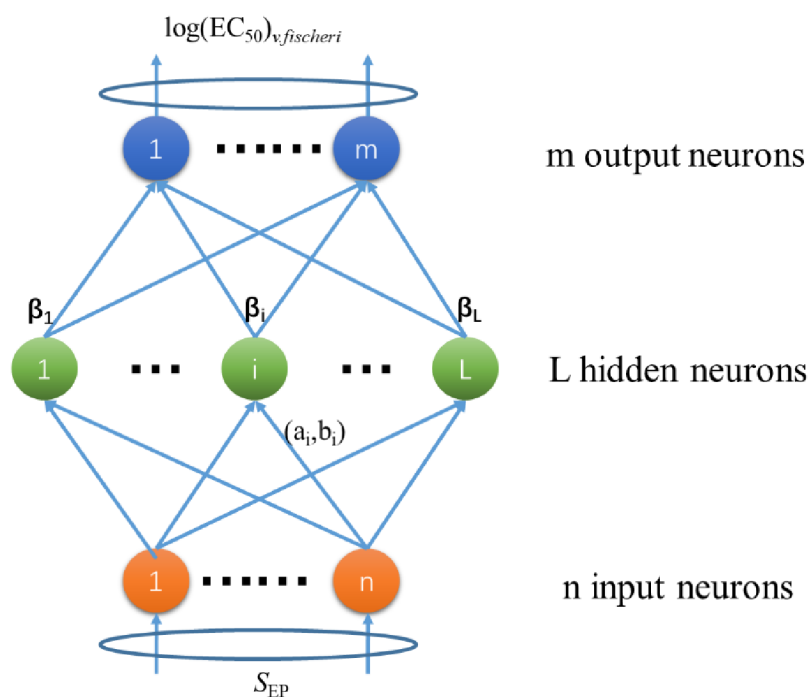


Figure. 5.1 ELM network structure in this study

5.3.3 Electrostatic potential surface area (S_{EP}) descriptors

For the sake of calculating S_{EP} descriptors, the structures of 54 cations and 28 anions of ILs were geometrically optimized at the theoretical level of B3LYP/6-31++G (d, p) by the software of Gaussian 09 B.01. Determining the local lowest energy point of cations and anions without a virtual frequency is a basic condition for the optimization process, thereby their energy as well as vibration frequencies are calculated. The optimal structures for cations and anions were eventually exhibited in [Figure. 5.2](#) and [Figure. 5.3](#), respectively. The coordinates and the total energy for the optimized cations and anions were included in the Supplementary data.

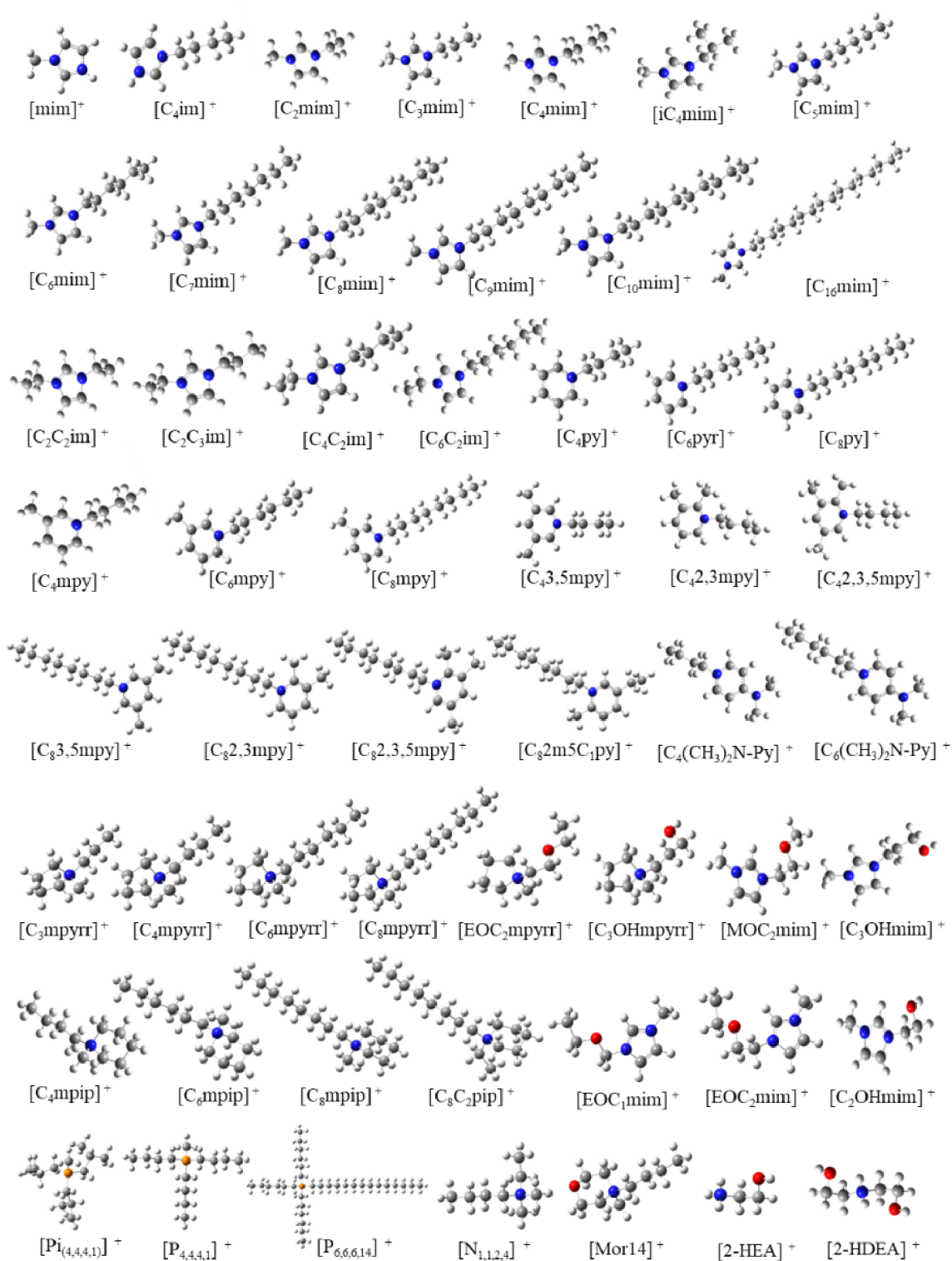


Figure 5.2 Structures of cations used in this work

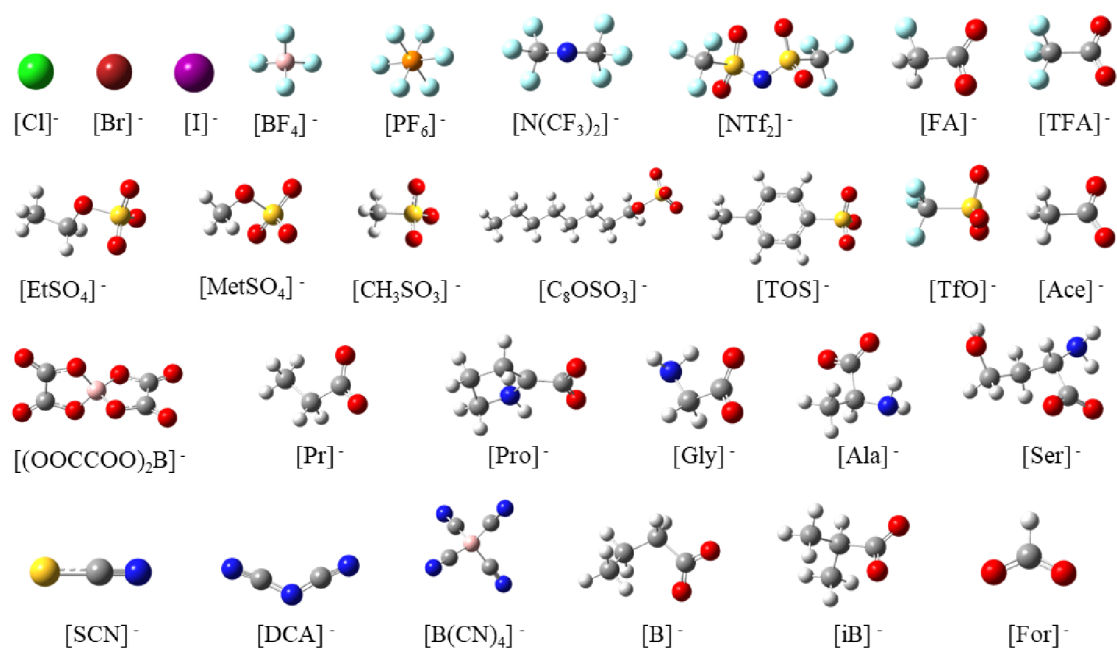


Figure 5.3 Structures of anions used in this work

Then the S_{EP} values were calculated by the Multiwfn 3.6(dev) software (Lu and Chen, 2012) under the condition that the electron density of the van der Waals surface was $0.001 \text{ e} / \text{Bohr}^3$ defined by Bader (Bader et al., 1987; Murray and Politzer, 2011b). The S_{EP} values were calculated in the step size of 0.5 kcal/mol when the anionic S_{EP} ranges from -160 kcal/mol to 0 kcal/mol and the range of cation is $0 \sim 160 \text{ kcal/mol}$. In order to screen reasonable descriptors, all the S_{EP} of separated anions and cations were treated in eight equal regions (S_{EP1} to S_{EP8} for cations and S_{EP9} to S_{EP16} for anions). For instance, S_{EP1} means the sum of S_{EP} values for cations in the range of $0 \text{ kcal/mol} \sim 20 \text{ kcal/mol}$, while the S_{EP9} adds up all the calculated S_{EP} values of anions between the range of -160 kcal/mol and -140 kcal/mol . The S_{EP} descriptors of the typical cation ($[\text{C}_{16}\text{mim}]^+$) and anion ($[\text{C}_8\text{OSO}_3]^-$) are depicted in Figure 5.4 and Figure 5.5, respectively.

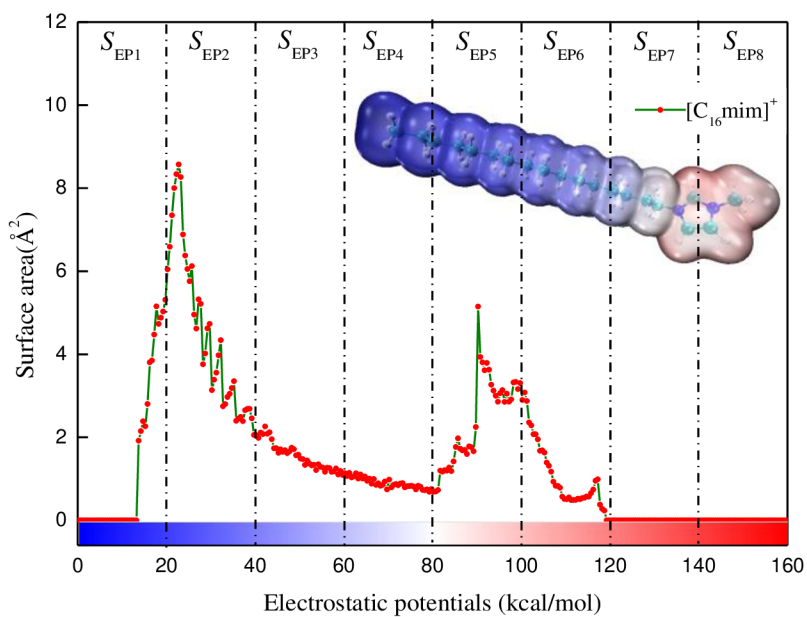


Figure 5.4 Electrostatic potentials versus corresponding surface areas of typical cation

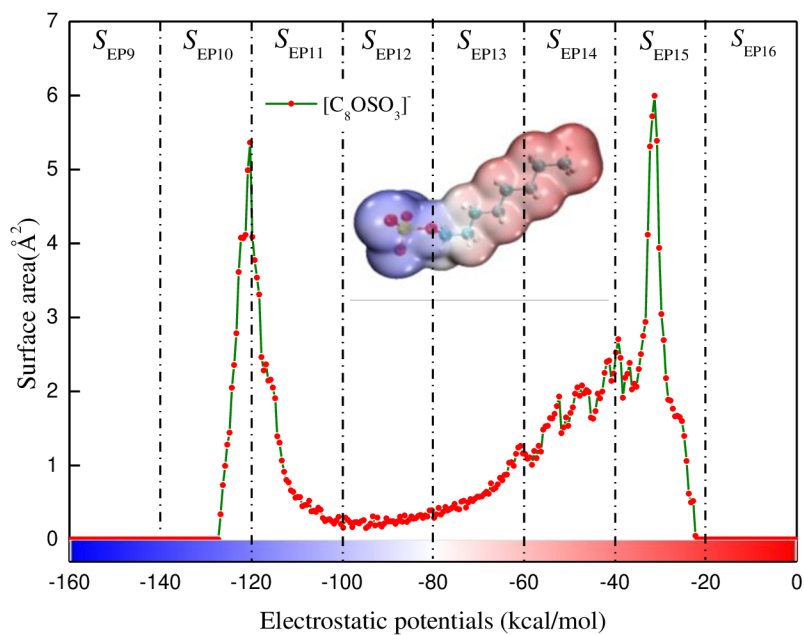


Figure 5.5 Electrostatic potentials versus corresponding surface areas of typical anion

5.3.4 Model validation

To valid the predictability of the linear and non-linear models in this work, six

metrics were introduced in the present work, namely the squared correlation coefficient (R^2), average absolute relative deviation (AARD%), the error (E), the average absolute error (AAE), mean-square error (MSE) and root-mean-square error (RMSE). The equations for these parameters are demonstrated in Eqs. (1)-(6) respectively.

$$R^2 = \frac{\sum_{i=1}^N (y_i^{Exp.} - \bar{y}_m)^2 - \sum_{i=1}^N (y_i^{Pre.} - y_i^{Exp.})^2}{\sum_{i=1}^N (y_i^{Exp.} - \bar{y}_m)^2} \quad (1)$$

$$AARD(\%) = 100 \times \sum_{i=1}^N \left| \frac{y_i^{Pre.} - y_i^{Exp.}}{y_i^{Exp.}} \right| / N \quad (2)$$

$$E = y_i^{Pre.} - y_i^{Exp.} \quad (3)$$

$$AAE = \sum_{i=1}^N |y_i^{Pre.} - y_i^{Exp.}| / N \quad (4)$$

$$MSE = \sum_{i=1}^N (y_i^{Pre.} - y_i^{Exp.})^2 / N \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^N (y_i^{Pre.} - y_i^{Exp.})^2 / N} \quad (6)$$

where N signifies the number of samples, i indicates the serial number of the data samples, $y_i^{Pre.}$ and $y_i^{Exp.}$ are the predictive and experimental $\log(EC_{50})_{V. fischeri}$, respectively. The average $\log(EC_{50})_{V. fischeri}$ for different datasets is expressed as \bar{y}_m .

In addition, the application domain (AD) is defined by the modeled response and the model descriptors, and can be used to estimate the predictive ability of QSAR model for the property of a new compound (Gramatica, 2007). In this work, the Williams plot, the standardized residuals (σ) versus the leverage values (h), is used to confirm the AD. The leverage approach is utilized to test whether a new compound is within the structural model domain and the leverage value h_i is calculated by the Eq. (7) as follows:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (7)$$

where x_i means a row vector of the i th IL descriptors and X represents the descriptor matrix for the training set. The threshold value (h^*) is defined by Eq. (8) as follows:

$$h^* = \frac{3(p+1)}{n} \quad (8)$$

where p denotes the number of model descriptors, and the n is the number of ILs in the

training set.

5.4 Results and discussion

5.4.1 Screening and impact analysis of descriptors

For the effective construction of promising non-linear models, the first step of this work is to screen reasonable descriptors by the approach of the multiple linear regression (MLR). Through the way of stepwise regression, the most accurate linear equation with the optimal number of parameters was determined (as shown in Eq. (9)). Five descriptors were ultimately selected from 16 descriptors of S_{EP1} - S_{EP16} , which are S_{EP1} , S_{EP2} , S_{EP6} , S_{EP7} , S_{EP13} , respectively. The rest descriptors were excluded due to their relative insignificance. The significance of the selected descriptors for the model is consistent with their order in Eq. (9).

$$\begin{aligned} \log(\text{EC}_{50})_{V.fischeri} = & 2.638 - 0.022 \times S_{EP2} - 0.09 \times S_{EP7} + 0.02 \times S_{EP6} \\ & - 0.01 \times S_{EP13} + 0.037 \times S_{EP1} \end{aligned} \quad (9)$$

(n=142, $R^2=0.492$)

It can be seen that four significant descriptors belong to cations while the S_{EP2} is the most important one. Nevertheless, only one descriptor (S_{EP13} , the fourth crucial descriptor), belongs to anions, manifesting the anions have a lower influence on the ILs toxicity than cations. This is in accordance with previous work which applied MLR algorithm to estimate the ecotoxicity of ionic liquids (EC_{50}) on *Vibrio fischeri* involving 9 kind of cations and 17 anions (Luis et al., 2010). Besides, it can be deduced from Eq. (9) that the S_{EP2} , S_{EP7} and S_{EP13} have negative effects on the predicted values which means the $\log(\text{EC}_{50})_{V.fischeri}$ value decreases with their rising, and the corresponding toxicity will be enhanced. For instance, the S_{EP2} values of $[\text{C}_6\text{mim}]^+$, $[\text{C}_7\text{mim}]^+$, $[\text{C}_8\text{mim}]^+$, $[\text{C}_9\text{mim}]^+$ and $[\text{C}_{10}\text{mim}]^+$ are 10.2982, 30.726, 52.5224, 76.7788, and 99.4291, respectively. The specific relationship for the surface areas of electrostatic potentials and the alkyl chain length was demonstrated in [Figure 5.6](#). It can be extrapolated that the cations with longer alkyl chain present higher S_{EP2} are responsible for the stronger ecotoxicity. This conclusion states a good consistence with the finding in previously published reports (Chen et al., 2014; Cho et al., 2007; Ghanem et al., 2017).

According to the observation, the anionic descriptor (S_{EP13}) negatively impacts the outcomes which implies that ILs with higher S_{EP13} possess lower $\log(\text{EC}_{50})_{V.fischeri}$ value

and stronger toxicity. For instance, compared with other anions in Fig. 7, the $[\text{NTf}_2]^-$ had the largest $S_{\text{EP}13}$ (70.6956), showing that ILs consisting of $[\text{NTf}_2]^-$ anion are the most toxic on the basis of possessing same cations. This may be because that the ILs with $[\text{NTf}_2]^-$ are recognized as hydrophobic compounds (Santos et al., 2014). In addition, the anion $[\text{BF}_4]^-$ is helpful for making ILs hydrophilic, but its S_{EP} distributed in the region of $S_{\text{EP}10}$ and $S_{\text{EP}11}$ (as shown in Figure. 3.7), indicating their insignificant for ILs toxicity. Above analysis demonstrates that ILs consisting of hydrophobic anions are more toxic compared to those with hydrophilic anions. This shows a good agreement with the conclusion done by Ghanem et al. (2017) and Santos et al. (2014).

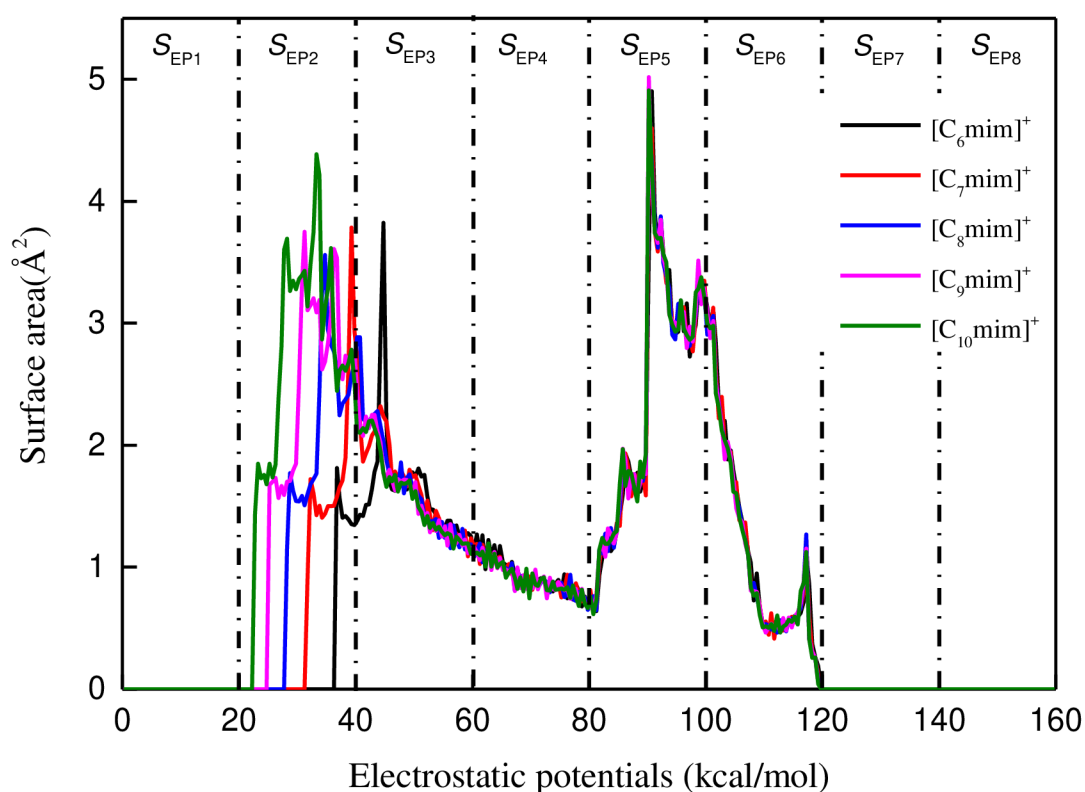


Figure 5.6 S_{EP} descriptors of various 1-alkyl-3-methylimidazolium cations

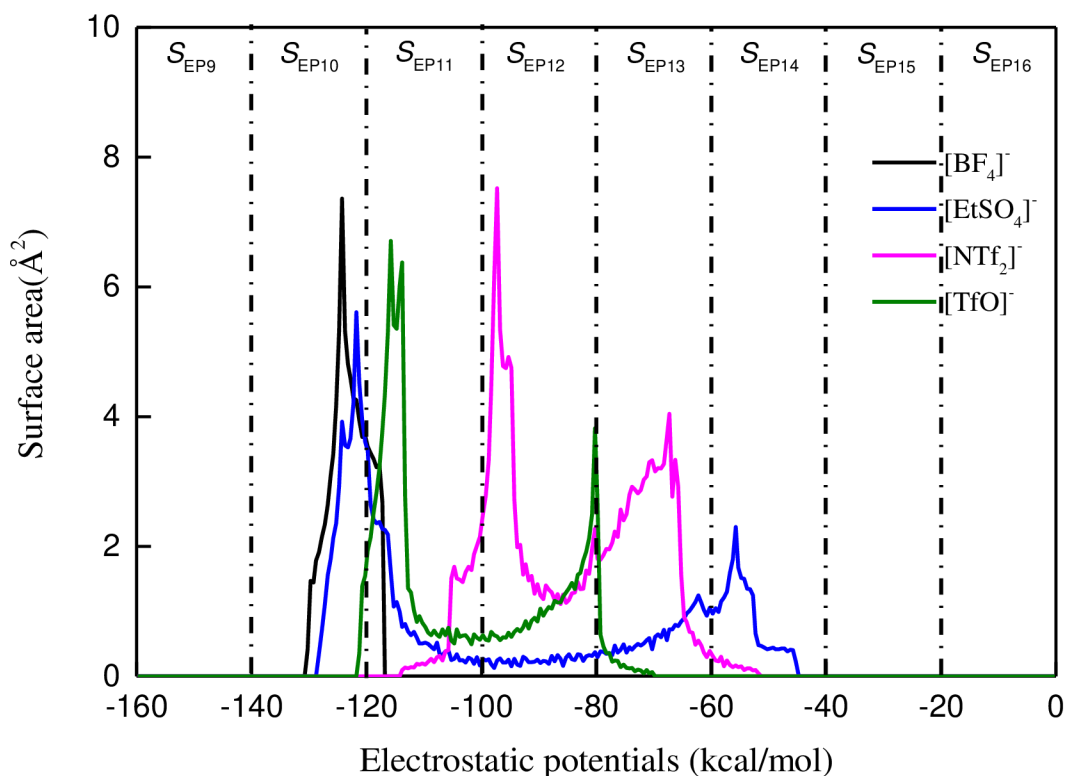


Figure. 3.7 S_{EP} descriptors of various hydrophobic and hydrophilic anions

5.4.2 Extreme learning machine (ELM) model establishment

Based on the above work, the screened five descriptors as input were employed to establish a more rational and accurate non-linear model by ELM algorithm. During the procedure of modeling, the “sig” function was selected as the activation function. Then, numerous models were developed by training set based on varied numbers of hidden layer neurons (from 40 to 80 under the step of 5) and the R^2 and AARD% values for each model were calculated simultaneously. After completing the establishment of the predictive models, the test set was applied for the validation of each model. The R^2 and AARD% of the ELM models based on the different number of neurons for different data sets were compared in Figure 5.8. It can be seen that the R^2 for training set gradually rose with the number of the hidden layer neurons growing while AARD% values conversely declined. Their trends were generally stabilized when the number of neurons exceeded 70. For the test set, the R^2 and AARD% had similar trends with the training set at the beginning. Nevertheless, the test set R^2 started to reduce sharply and its AARD% increased substantially when the neurons number was over 70. Hence, the optimal ELM model was obtained with the neurons number 70 in the hidden layer.

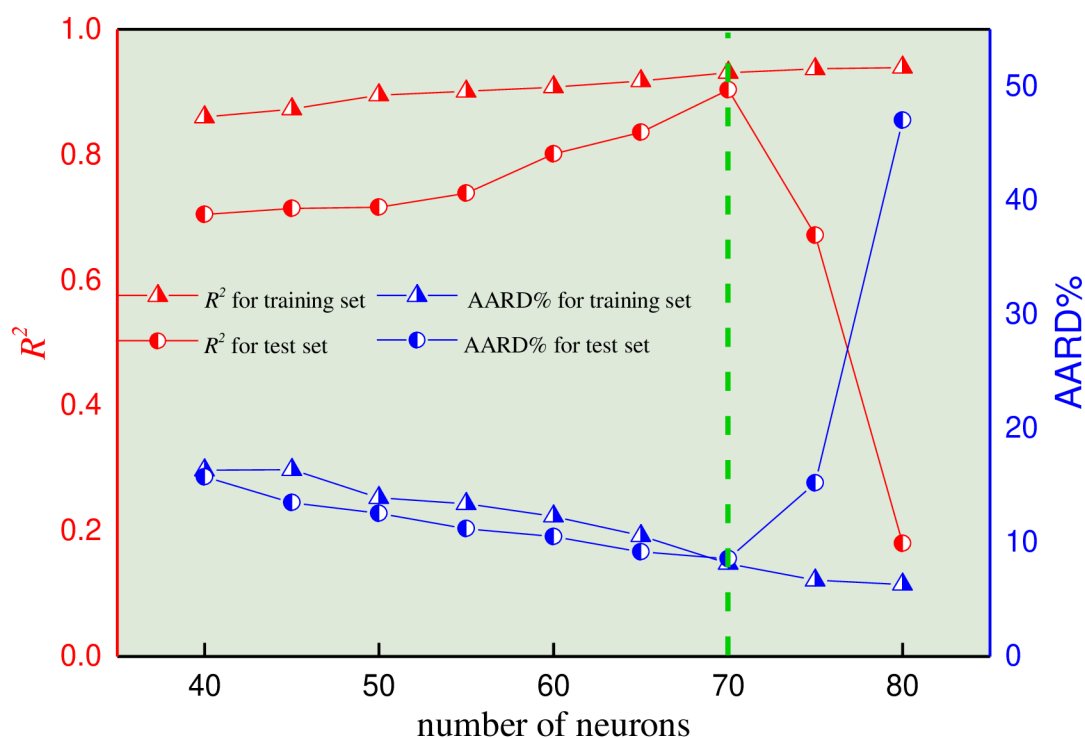


Figure. 5.8 R^2 and AARD% of the ELM model versus the number of neurons for different sets

As shown in the Figure 5.8, the R^2 (0.9308) and AARD% (8.1654) for the training set of the established ELM model were obtained. Simultaneously, those values for test set used as validation were 0.9035 and 8.5888, respectively. The $\log(\text{EC}_{50})_{V. fischeri}$ values measured by the ELM model were schematically compared with the corresponding experimental data in Figure 5.9. It can be observed that the predicted values and the data points from literature for the training and test sets are closely surrounded in the vicinity of diagonal. Detailed evaluated and experimental data were placed in the Supplementary data. The overall outcomes display a good performance for the toxicity evaluation of ILs on *Vibrio fischeri*.

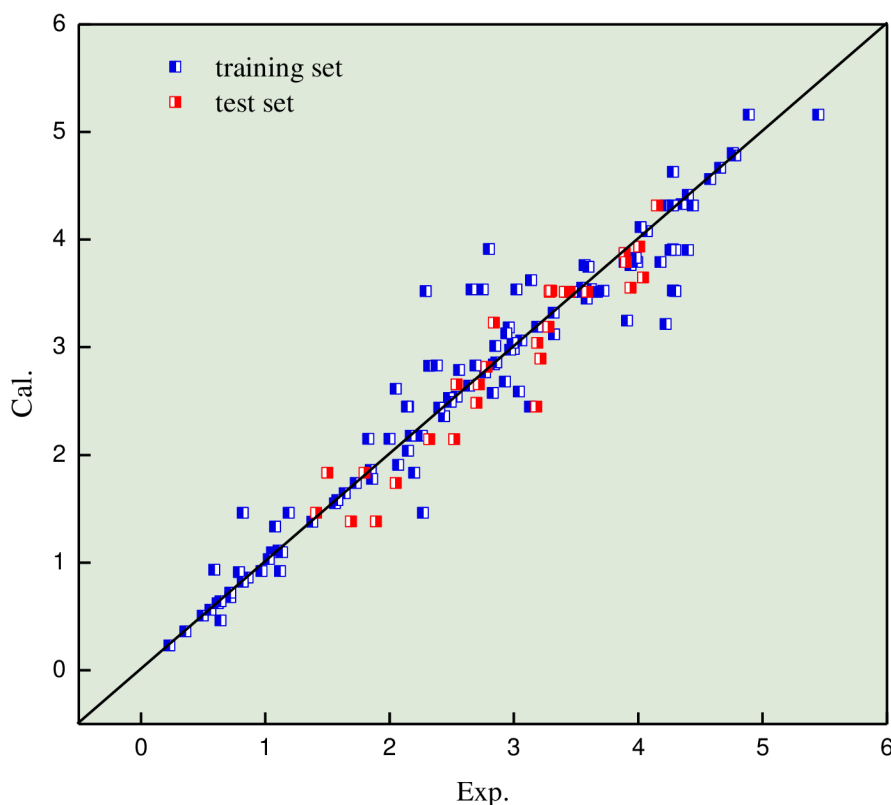


Figure 5.9 Predicted versus experimental $\log(\text{EC}_{50})_{V.fischeri}$ of ILs for the ELM model in this work

Error distributions for evaluated $\log(\text{EC}_{50})_{V.fischeri}$ values were exhibited in [Figure 5.10](#), which apparently displays that the majority of E% values for calculated toxicity data are between -1.0% and 1.0%. Beyond, the predictive abilities of the established model for different datasets were also detected by the metrics of R^2 , AARD%, AAE%, MSE, and RMSE which are presented in [Table 5.1](#). The listed statistical parameters in [Table 5.1](#) indicated the accurate estimation capability of the ELM model.

To further verify the stability of the developed ELM model, we employed three new ILs for the external validation. The results were presented in [Table 5.2](#) and the full names of the new ILs were provided in the Supplementary data [S2](#). As can be observed from [Table 5.2](#), the predicted values are close to the experimental values, showing that the ELM model is reliable and robust for the prediction of ILs toxicity towards *Vibrio fischeri*.

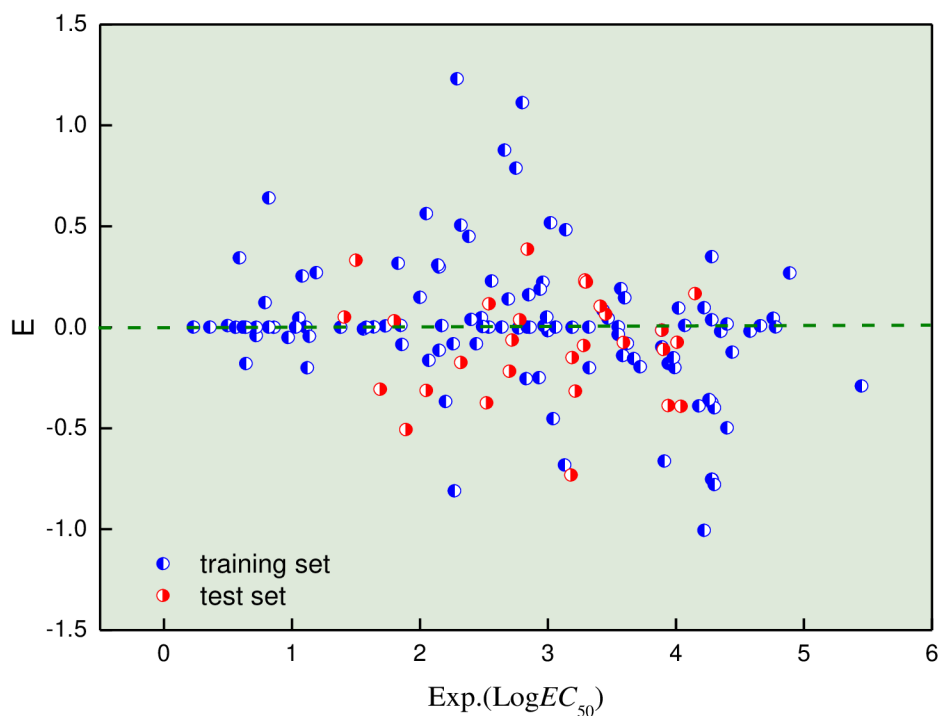


Figure 5.10 Error distributions of the ELM model developed in this work

Table 5.1 The statistical parameters of ELM model in this work

Data set	No. of data points	R^2	AARD %	AAE%	MSE	RMSE
Training set	113	0.9308	8.1654	0.2085	0.1147	0.3387
Test set	29	0.9035	8.5888	0.2163	0.0739	0.2718
Total set	142	0.9272	8.5023	0.2101	0.1080	0.3262

Table 5.2 Predictions of ILs toxicity to *Vibrio fischeri* for three external ILs

No.	ILs Abbreviation	Experimental $\log(\text{EC}50)_{v.fischeri}$ (mmol. L ⁻¹)	Predicted $\log(\text{EC}50)_{v.fischeri}$ (mmol. L ⁻¹)	Ref.
1	[C ₂ OHmim][I]	3.89	3.87	(Stolte et al., 2007b)
2	[C ₃ mim][Cl]	3.78	3.55	(Montalbán et al., 2016)
3	[[C ₄ 2,3mim][NTF ₂]	2.87	2.62	(Montalbán et al., 2016)

5.4.3 The applicability domain (AD) of ELM

The applicability domain analysis was performed to evaluate the robustness and predictability of the ELM model in this work. Generally, only those predicted values within the AD are reliable. As can be seen from Figure 5.11, the Williams plot shows the analyzing of AD to verify the reliability of ELM model. The results show that most of the ILs have lower leverage values than the critical value (h^*) and possess the standardized residuals within 3.0 standard deviation units (3σ), exhibiting the good predictive reliability of the model. As shown in Fig. 11, there are six outliers in the Williams plot including 4 X outliers and 2 Y outliers. The leverage values of 4 X outliers (involving ILs 57, 99, 100 and 101) are greater than h^* , indicating that these ILs are structurally influential chemicals (Ma et al., 2015). The standardized residuals of ILs 28 and 41 are Y-outliers ($>3\sigma$), but they belong to the AD of the ELM model as their leverage values are within the threshold value (h^*). These Y-outlier points are probably caused by the wrong experimental data rather than the molecular structure (Gramatica, 2007).

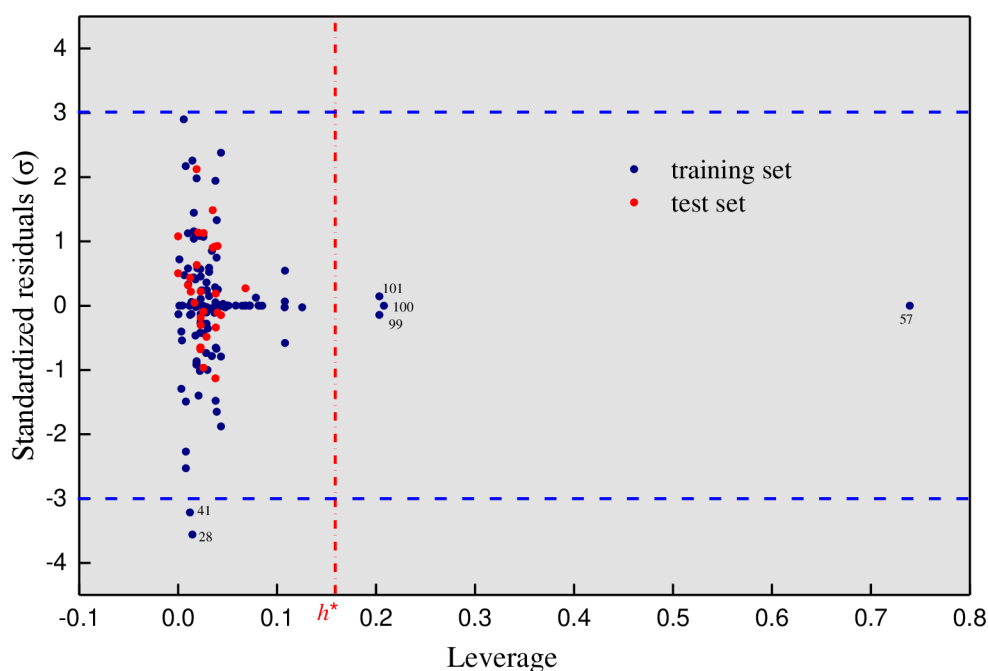


Figure 5.11 Williams plot of the ELM model for the train and test sets

5.4.4 Comparison of different QSPR models for the ILs toxicity on *Vibrio fischeri*

Table 5.3 summarised the published QSAR models for predicting the ecotoxicity of ILs on *Vibrio fischeri*. According to observation, the different descriptors were employed by relative researchers. The linear method (MLR) was widely applied in these studies and non-linear models were constructed by various approaches containing GFA, MLP and LSSVM. It obviously presents that the current study firstly combined S_{EP} descriptors and ELM technique for predicting the toxicity of ILs on *Vibrio fischeri*.

Additionally, this work assembled more extensive toxicity data points of the ILs (142) and introduced less descriptors (5) for the estimation model, whereas the majority of models in the bibliography were established based on the dataset around or below 100 endpoints and the number of applied molecular descriptors varied from 2 to 30.

Three models were carried out by Das and Roy (2012), Yan et al. (2015) and Jafari et al. (2019) utilizing rather comprehensive datasets which were 146 ILs, 157 ILs and 187 ILs respectively. In contrast, the authors (Yan et al., 2015) obtained superior R^2 indicating the good reliability of their models when they utilized 28 descriptors. Nevertheless, the greater R^2 values of the ELM model for the test set (0.9035) and the entire set (0.9272) in this work were achieved employing only five descriptors. Based on the above comparison, it can be obviously concluded that the ELM model with widespread data and rational descriptors in this work can be used reliably and accurately for the ILs toxicity prediction on *Vibrio fischeri*.

Table 5.3 Summary of published QSPR models for predicting the ecotoxicity of ILs for *Vibrio fischeri*

No. of IL	Method	No. of molecular descriptors	Descriptors type	R^2	Ref.
25	GFA ^b	4	Electronic, spatial, structural, thermodynamic, and topological descriptors	0.78	(Couling et al., 2006)
43	MLR	9	Group contribution	0.925	(Luis et al., 2007)
51	MLR	5	Minimum net atomic charge for a C atom; WPSA-1; PPSA1; TMSA); maximum atomic orbital electronic population; LUMO+1 energy	0.9017 (gas), 0.9209 (solvent)	(Bruzzone et al., 2011)
56	MLR	5	Lopping centric information index, the number of oxygen	Tr=0.78	(Grzonkowska et al., 2016)

No. of IL	Method	No. of molecular descriptors	Descriptors type	R^2	Ref.
96	MLR	15	atoms, Gutman molecular topological index Group contribution	0.924	(Luis et al., 2010)
103	MLR	30	Free energy relationship (LFER) descriptors	Tr=0.762 Te=0.812	(Cho et al., 2013)
24	MLR	2	The energy of the lowest unoccupied molecular orbital E LUMO & the molecular volume V	0.954	(C. Wang et al., 2015)
69	GFA & LSSVM ^c	5	The size; lipophilic ^a ; and 3D molecular structure of the cations core and molecular weight of anions	Tr=0.893 Te=0.903 &Tr=0.903 Te=0.933	(Ma et al., 2015)
110	MLR	5	Charge distribution density (σ -profile)	0.906	(Ghanem et al., 2017)
110	MLP ^d	5	Charge distribution density (σ -profile)	Tr=0.978 Tv=0.961 Te=0.979	(Ghanem et al., 2017)
146	MLR	5	DRAGON	Tr=0.694 Te=0.739	(Das and Roy, 2012)
157	MLR	28	Topological index	0.908	(Yan et al., 2015)
187	MLR	6	The number of carbon, hydrogen, nitrogen and oxygen atoms from cation as well as hydrogen and fluorine atoms from anion	Tr=0.8769 Te=0.8496	(Jafari et al., 2019)
142	ELM	5	Surface area of electrostatic potentials	Tr=0.9308 Te=0.9035 To=0.9272	This work

^a Affinity of the cation to penetrate to the lipid.

^b Genetic function approximation.

^c Least squares support vector machine, multifactorial analysis.

^d Multilayer perceptron neural network.

5.5 Conclusions

In present study, an extensive toxicity dataset of 142 ILs to *Vibrio fischeri* involving 54 cations and 28 anions was gathered. Then the S_{EP} descriptors of separate ions were calculated, and five crucial descriptors were selected by MLR method. S_{EP2} was proved to be the most significant descriptor and showed negative effect on the $\log(EC_{50})_{V. fischeri}$. The efficient non-linear model has been developed using the extreme learning machine technique combined with the S_{EP}

parameters. The predictive values of ILs toxicity are highly in line with the corresponding experimental data. The R^2 (0.9272) and RMSE (0.3262) of the model for the total set were gained, displaying a good performance and reliability for the toxicity evaluation of ILs towards *Vibrio fischeri*. The proposed approach is valuable for understanding the relationship between structure and toxicity of ILs and is helpful for screening and designing green and functional ILs for their future applications in various fields.

5.6 Supplementary data

The following link is the Supplementary data to this chapter:

Spreadsheet:

<https://ars.els-cdn.com/content/image/1-s2.0-S0304389420307500-mmc1.xlsx>

PDF file:

<https://ars.els-cdn.com/content/image/1-s2.0-S0304389420307500-mmc2.pdf>

Chapter VI

Atom surface fragment contribution method for predicting the toxicity of ionic liquids

Xuejing Kang, Yongsheng Zhao, Zhongbing Chen

Adapted from Journal of Hazardous Materials 421 (2022): 126705

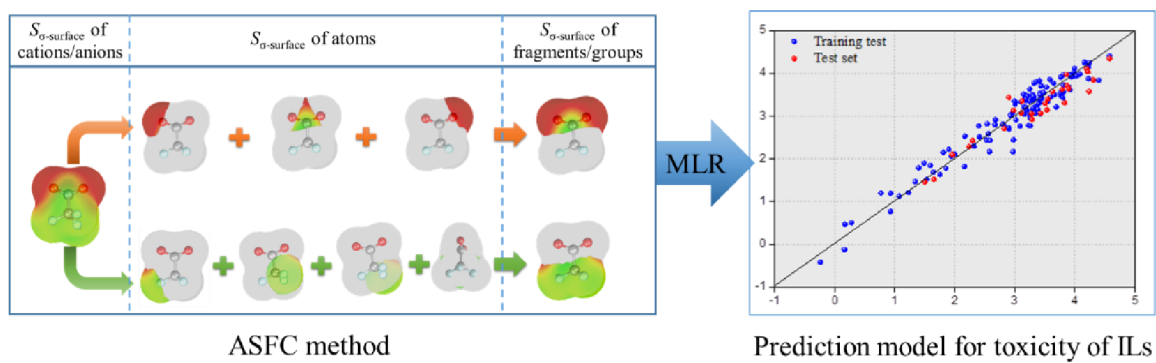
Contents

6.1 Abstract.....	101
6.2 Introduction.....	103
6.3 Theory and methodology.....	105
<i>6.3.1 Dataset.....</i>	<i>105</i>
<i>6.3.2 Traditional GC method.....</i>	<i>105</i>
<i>6.3.3 ASFC method.....</i>	<i>105</i>
<i>6.3.4 Model validation and performance</i>	<i>111</i>
6.4 Results and discussion	112
<i>6.4.1 Traditional GC model development</i>	<i>112</i>
<i>6.4.2 Novel ASFC model development</i>	<i>114</i>
<i>6.4.3 Comparisons.....</i>	<i>116</i>
6.5. Conclusion	118
6.6 Supplementary data.....	119

6.1 Abstract

In this study, a novel method—atom surface fragment contribution (ASFC)—was proposed for assessing the properties of compounds. We developed a predictive model using the ASFC method based on the sigma surface areas ($S_{\sigma\text{-surface}}$) of fragments/groups for estimating the toxicity of ILs. A toxicity dataset of 140 ILs towards *leukemia rat cell line* (IPC-81) was gathered and employed to train and validate models. The $S_{\sigma\text{-surface}}$ values of atoms in each group were firstly calculated from the COSMO profiles of cations and anions for ILs. Then the $S_{\sigma\text{-surface}}$ values of 26 groups were obtained and used as input descriptors for modeling. The R^2 and MSE of the built ASFC model were 0.924 and 0.071, respectively. Results indicate that the ASFC model developed by the new approach possesses great accuracy and reliability. In total, the ASFC method has extensive potential for the application of estimating diverse properties of ILs and other compounds due to its remarkable advantages.

Graphical abstract:



6.2 Introduction

Ionic liquids (ILs) have become a paramount research subject in modern chemistry and rapid progress has been achieved in the recent decades (Welton, 2018). ILs comprising entirely of ions are pure molten salts under ambient conditions (Gebbie et al., 2017) and have a couple of unique and benign features such as low volatility, low melting points, non-flammability as well as high electrochemical and thermal stabilities (Hosseini et al., 2020; Z. Wang et al., 2020). Because of these remarkable properties, ILs have been found applications in various areas of modern science, for example, pharmaceuticals (Lee and Lin, 2017), batteries, organic synthesis, catalysis, and gases (Ding et al., 2020; Egorova et al., 2017; Rosen et al., 2011). Tunable physicochemical properties are the key nature of ILs, which makes ILs have a wide range of applicability. Approximately 10^{18} ILs with different combinations of anions and cations can be synthesized (Clark et al., 2018). Hence, experimentally investigating the properties of ILs or screening desirable ILs for special targets one by one becomes unrealistic. Additionally, to avoid the resource-intensive, time-consuming, and even risky laboratory tests, many linear and nonlinear models for predicting the related properties of ILs were established by various methods, such as group contribution (GC), machine learning methods (Abramenko et al., 2020a; Gao et al., 2020; Song et al., 2020a; Torrecilla et al., 2008b; Zhao et al., 2015a).

GC is a class of semiempirical theory which is a successful and popular technique for estimating properties of compounds or mixtures (Bell et al., 2020; Y. Huang et al., 2013). During the GC method, one molecule is regarded as an integration of defined functional groups. It needs to fit the available data of compounds with the same groups to determine the group parameters, and then the properties of any compound can be evaluated (Lazzús, 2009; Mu et al., 2007). So far, GC models have been extensively used for estimating various properties (e.g., viscosity, density, electrical conductivity, melting point, surface tension, heat capacity) of ILs (Gardas and Coutinho, 2008; Gardas and P.Coutinho, 2009). The groups for ILs usually contain cations (e.g., imidazolium [Im]⁺, pyridinium [Py]⁺, piperidinium [Pip]⁺), anions (e.g., tetrafluoroborate [BF₄]⁻, chloride [Cl]⁻, hexafluorophosphate [PF₆]⁻), and some substituents (e.g., methylene (-CH₂-), methyl (-CH₃)) (Chen et al., 2019a; Y. Huang et al., 2013). Specifically, Gharagheizi et al (2012a) developed a GC model using 19 sub-structures plus temperature as model parameters to estimate the surface tension of different ILs. Several GC models were proposed by Chen et al. (2019b) to predict five types of properties for various ILs, namely heat capacity, viscosity, melting point, viscosity, density. The establishments of their models employed the GC parameters

which were calculated by fitting experimental data based on the number of different groups in each molecule. Ismail Hossain et.al. (2011) reported a GC model for evaluating the toxicity of ILs towards *D.magna*. Their work gathered the experimental toxicity data involving 64 ILs and the GC model was established via MLR (multiple linear regression) using the polynomial software and the model presented good performance with $R^2=0.974$. Their research results were in good agreement with the experimental data.

However, some deficiencies and disadvantages exist in the traditional GC models. The main drawbacks of GC include that only the types and frequencies of the groups are considered and the interactions of groups or neighbor groups are not described (Mu et al., 2007). To the end, the isomers of cations/anions including the isomeric fragments/groups cannot be distinguished. Nevertheless, strictly speaking, each group should have different (or slightly different) contributions in various molecules, since each group in different molecules should have different environments and various interactions with its surrounding atoms. Therefore, it is necessary to improve the traditional GC method for better accuracy and more extensive application.

Recent years, the toxicity of ILs has attracted much attention from researchers and some estimation models have been built (Cho and Yun, 2019; Jafari et al., 2019; Lan et al., 2020; Yan et al., 2019). This study planned to develop a new method—atom surface fragment contribution (ASFC)—utilizing some surface parameters of fragments, which were calculated based on quantum chemistry, to build predictive models for the toxicity of ILs against the *leukemia rat cell line* (IPC-81). Particularly, the surface area of screening charge density (i.e., the sigma surface area), or surface-area of the electrostatic potential of atoms, can reflect the interactions of atoms at the electric level and thus can distinguish the contributions of each group in different molecules to a certain extent. Consequently, they might have a high potential to be critical parameters that can make contributions for improving the reliability of GC models. For example, sigma profiles describe the distribution of the electronic polar charge related to a molecular surface (Motlagh et al., 2020). The sigma surface symbolized the charge distribution of a molecular surface. This study firstly introduced the sigma surface area ($S_{\sigma\text{-surface}}$) of atoms for the development of the proposed method. Whereas, in our future work, the predictive ASFC models will be developed based on the surface areas of electrostatic potential (S_{EP}) of atoms for the calculation of physicochemical characteristics of compounds.

In this work, prediction models for the toxicities of various ILs are constructed by the means of the ASFC method using the critical parameters of $S_{\sigma\text{-surface}}$ of fragments.

The objectives of our current work are (1) obtaining the $S_{\sigma\text{-surface}}$ values of every atom in each group; (2) estimating the contributions of groups in ILs; (3) developing an ASFC model for predicting the toxicities of various ILs towards IPC-81; (4) comparing the performances of predictive models using ASFC and general GC methods.

6.3 Theory and methodology

6.3.1 Dataset

A toxicity data set containing 140 ILs towards the *leukemia rat cell line* (EC50 values) was obtained from the widely acknowledged ILs database (The UFT/ Merck Ionic Liquids Biological Effects Database; Zhang et al., 2006). The full names and the toxicity data points of all the ILs were summarized in Supplementary data (Table S1). The entire set was arbitrarily split into two sub-datasets, a training data set with 113 ILs for modeling and a test set with 27 ILs for validating, respectively.

6.3.2 Traditional GC method

Traditional GC method is underlying the assumption that the property of a substance is the sum of the contributions of all the defined groups and one group's effect in the molecule is independent of the others (Lazzús, 2012; Mu et al., 2007). In this work, the toxicity values of ILs towards the IPC-81 were regarded as the outcome variable or dependent variable (Y) while the numbers of each group in one molecule were considered as predictors or independent variables (X). The Eq. (1) was utilized to express the relationship between X and Y , and thus the MLR model would be established for estimating the toxicity of ILs (Ismail Hossain et al., 2011).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (1)$$

where X_m is the number of the m th group in the molecule of an IL; β_0 , β_1 , β_2 and β_m , stands for the regression coefficients. The IBM SPSS Statistics 25 software was used to develop linear predictive models.

6.3.3 ASFC method

6.3.3.1 The acquisition of COSMO profiles and group segmentations

In this work, the COSMO files of 74 cations and 15 anions were available by the BIOVIA COSMObase 2020 provided with BIOVIA COSMOtherm 2020 software (Dassault Systèmes) at the quantum chemical level of BP-TZVPD-FINE. Figure 6.1 presents the sigma (σ) surface for a representative cation (Fig.1(a)) and anion (Fig.1(b)), respectively. The red part stands for positive COSMO charge density while the molecule surface charge is negative; meanwhile, the blue area represents the negative COSMO charge density while the molecule charge is positive. Then, all the cations and anions were decomposed into several groups according to their structures. For example, the procedure of group division (Figure 6.2). All the groups (26 groups) contained in the molecule structures of ILs are listed in Table 6.1.

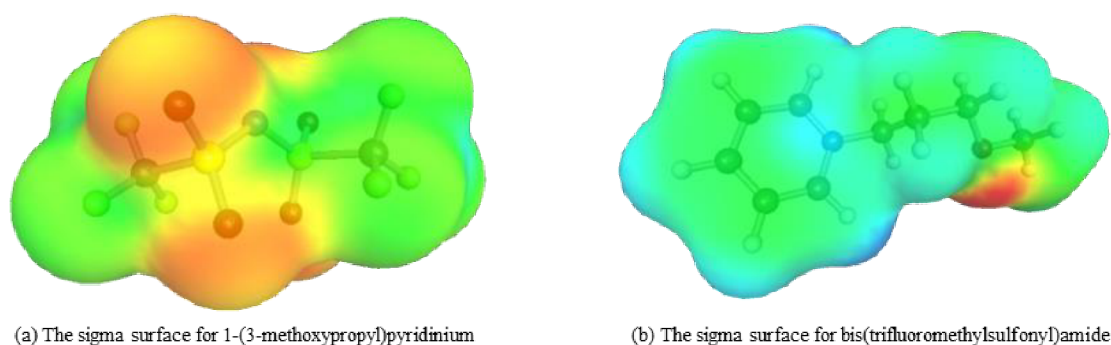


Figure 6.1 Sigma surface exemplified for (a) 1-(3-methoxypropyl)pyridinium;
(b) bis(trifluoromethylsulfonyl)amide

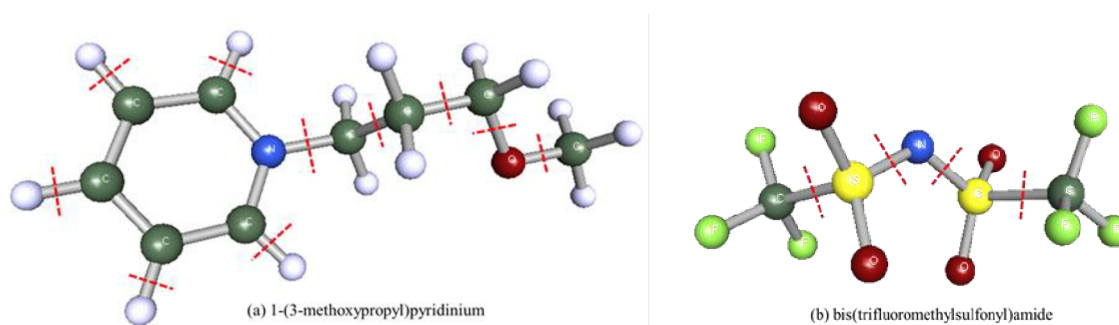
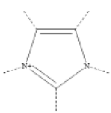
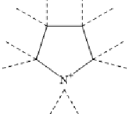
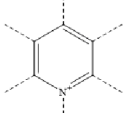
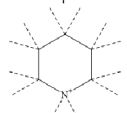
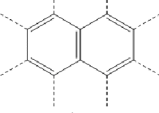
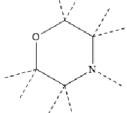
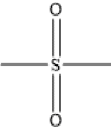
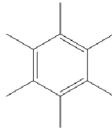


Figure 6.2 Group fragmentation exemplified for (a) 1-(3-methoxypropyl)pyridinium;
(b) bis(trifluoromethylsulfonyl)amide

Table 6.1 Structures and average surface values of groups for ILs used in this study

Group No.	Structures	Average S_{σ} -surface values (\AA^2)
Cation cores		
1*		50.139
2*		42.411
3*		59.964
4*		41.405
5	N-(CH ₂) ₂ (CH ₃) ₂	109.440
6*		106.505
7*		48.837
Anion cores		
8*	[TPTP] ⁻	262.489
9	[BF ₄] ⁻	94.145
10	[PF ₆] ⁻	114.339
11	[Cl] ⁻	52.682
12	[Br] ⁻	52.640
13	[I] ⁻	84.384
14*	[SCN] ⁻	86.624
Substituents		
15	-SO ₃	64.953
16		40.684
17	-COO-	49.997
18	-CF ₃	57.641
19	-CH ₃	36.291
20	-CH ₂ -	20.260
21	-H(ring)	5.547
22	-O- or [-O] ⁻	9.070

Group No.	Structures	Average $S_{\sigma\text{-surface}}$ values (\AA^2)
23	- OH	21.227
24		43.494
25	[-N-]- or > N -	8.769
26	-CN	42.821

Note: 1* imidazolium; 2* pyrrolidinium; 3* pyridinium; 4* piperidinium; 6* quinolinium; 7* morpholinium; 8* trifluorotris(pentafluoroethyl)phosphate; 14* thiocyanate.

6.3.3.2 The calculation of $S_{\sigma\text{-surface}}$ of atoms and groups

The $S_{\sigma\text{-surface}}$ values of atoms can be obtained from COSMO files which generally contain more COSMO charge information (Loschen and Klamt, 2012) and then employed to calculate the contributions of groups in the novel approach. To obtain the $S_{\sigma\text{-surface}}$ values of each atom of cations and anions, a set of codes were written and implemented via the MATLAB 2018 software to generate the $S_{\sigma\text{-surface}}$ values of each atom of ILs. Subsequently, the $S_{\sigma\text{-surface}}$ values of all the atoms for each group were added up and then the $S_{\sigma\text{-surface}}$ values of groups were used as predictor descriptors for modeling. For example, Table 6.2 lists the $S_{\sigma\text{-surface}}$ values of all atoms for a representative anion—trifluoroacetate ($[\text{CF}_3\text{COO}]^-$). Figure 6.3 described the procedure of calculating the $S_{\sigma\text{-surface}}$ values of groups for $[\text{CF}_3\text{COO}]^-$, which was divided into two groups (-COO- and -CF₃). Subsequently, the $S_{\sigma\text{-surface}}$ values of -COO- and -CF₃ were calculated, which are 49.997 and 61.804 respectively. If the group was contained in both the cation and anion of one IL, all the $S_{\sigma\text{-surface}}$ values of the group in the IL were added up as a descriptor. All the descriptors in different ILs were available in Supplementary data (Table S3 and Table S4), and the average $S_{\sigma\text{-surface}}$ values for groups (the whole dataset) used in this work were listed in Table 6.1. The ASFC model in this work was developed using the MLR method (Ismail Hossain et al., 2011) while the format is same with Eq (1). However, the X_m here was replaced by the parameter of $S_{\sigma\text{-surface}}$ of groups.

Table 6.2 The $S_{\sigma\text{-surface}}$ of atoms exemplified for trifluoroacetate ($[\text{CF}_3\text{COO}]^-$)

Groups	-COO-			-CF ₃			
	'C'	'O'	'O'	'C'	'F'	'F'	'F'
$S_{\sigma\text{-surface}}$ (Å ²)	11.145*	19.295	19.558	2.692	19.438	20.007	19.667

Note: *All the reported values in this paper were rounded to 3 effective digits while the specific data were available in Supplementary data.

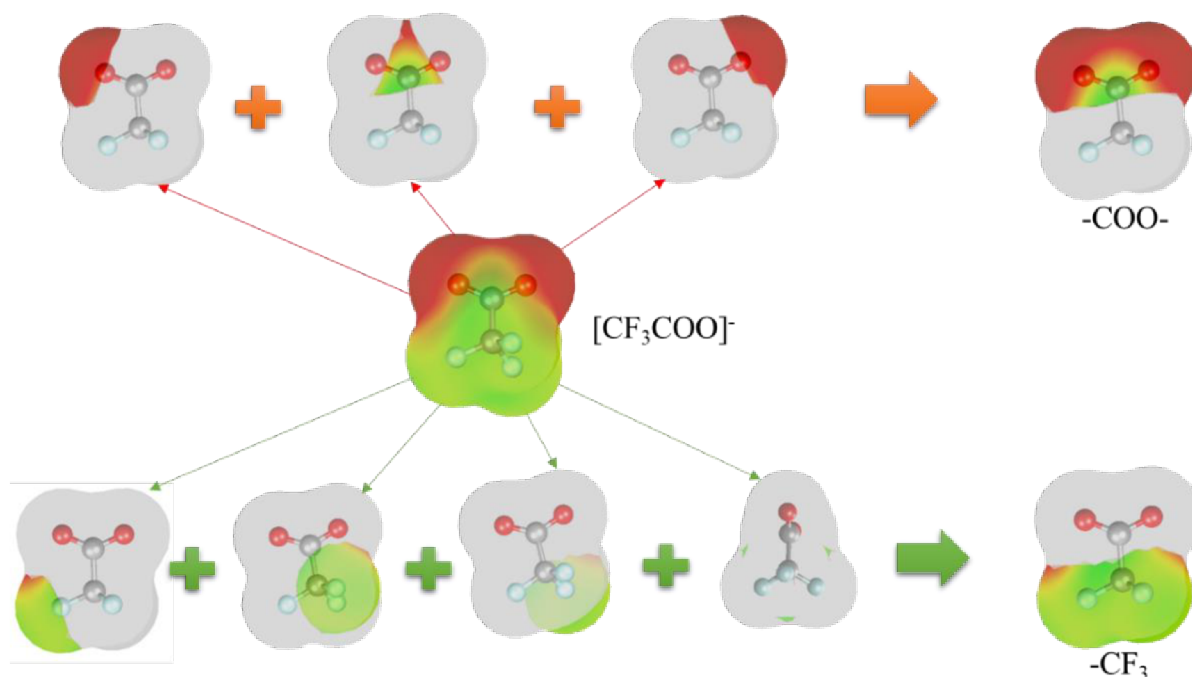


Figure 6.3 The calculation of sigma surface area of groups exemplified for trifluoroacetate ($[\text{CF}_3\text{COO}]^-$)

6.3.3.3 The superiority of ASFC method based on $S_{\sigma\text{-surface}}$

The primary advantage of this method is that it can distinguish not only the contributions of isomeric fragments/groups but also the contributions with slight differences of the same group in diverse molecules. For example, Figure 6.4 shows the isomeric cations ($\text{C}_8\text{H}_{12}\text{NO}$) of 1-(2-methoxyethyl) pyridinium and 1-(ethoxymethyl) pyridinium which have the same types and numbers of groups. As their difference is only the position of the group (-O-), they have the same number for each group and thus each group makes the same contribution in these two cations in the traditional GC

method. As a result, they always get the same predictive values for the property of ILs with these two cations and the same anions. However, in the ASFC method, each group in different molecules has its special $S_{\sigma\text{-surface}}$ value. We compared the $S_{\sigma\text{-surface}}$ values of groups between the above isomers of cations (Table 6.3), which implies their different contributions for the properties of ILs.

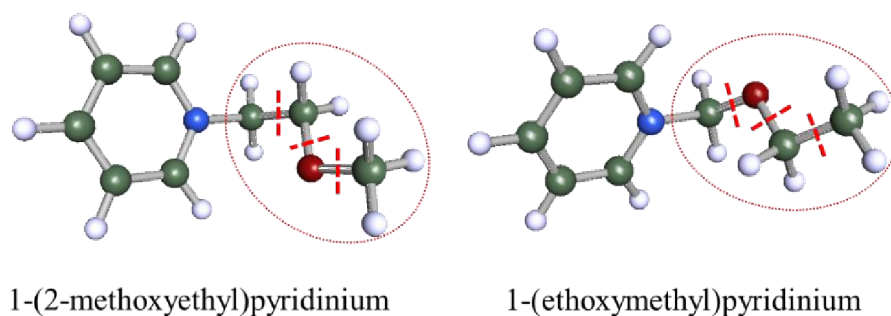
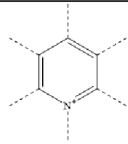


Figure 4.4 Two isomers of cations with the same molecular formula ($C_8H_{12}NO$)

Table 6.3 The comparison for $S_{\sigma\text{-surface}}$ of groups between two isomers of cations

Cations		The $S_{\sigma\text{-surface}}$ values of Groups (\AA^2)			
		—CH ₃	—CH ₂ —	—H(ring)	—O— /[—O] ⁻
1-(2-methoxyethyl)pyridinium	70.081	39.884	42.354	28.613	7.136
1-(ethoxymethyl)pyridinium	70.676	38.266	43.417	28.722	7.912

Additionally, the isomeric group, such as the group $(N-(CH_2)_2(CH_3)_2)$ in this study, has different structures in the cations, benzyltetradecyldimethylammonium and (ethoxymethyl) ethyldimethylammonium (Figure 6.5), but the $S_{\sigma\text{-surface}}$ values of them are 85.731 and 93.325 respectively. Consequently, this group makes different contributions in different cations for the toxicity values of ILs using our method, while it always shows the same contribution in the traditional GC method. Furthermore, one group owning a single structure can also present different contributions for the properties of various compounds. For instance, the $S_{\sigma\text{-surface}}$ values of the pyridinium group (excluding ring H) with the same structure in the cations 1-butyl-3,4-dimethylpyridinium, 1-butyl-3-methylpyridinium, and 1-butylpyridinium are 50.689, 59.693, 71.202, respectively. To sum up, this method can accurately depict the different contributions of each group in various molecules. Therefore, the obtained prediction

model using this method is highly potential to be more reliable and extensively applicable.

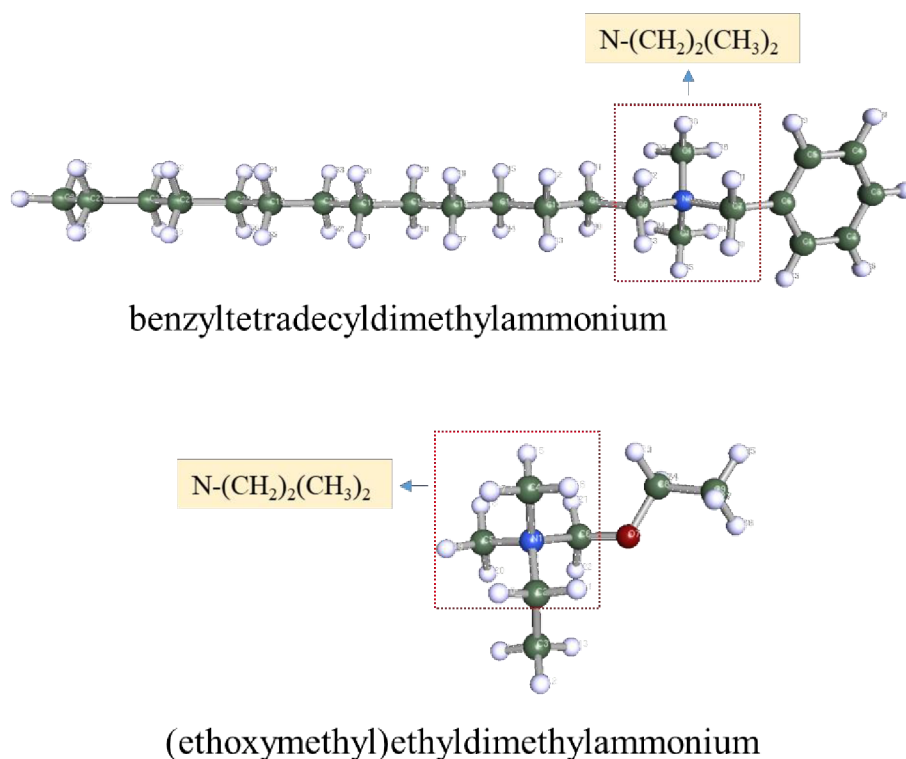


Figure 6.5 Two isomers of the group $(N-(CH_2)_2(CH_3)_3)$ indifferent cations

6.3.4 Model validation and performance

To investigate and compare the accuracy and reliability of the predictive models, the statistical parameters, namely average absolute relative deviation (AARD), relative deviation (RD), coefficient of determination (R^2), adjusted squared correlation coefficient (R_{adj}^2), mean-square error (MSE) and the root mean square error (RMSE), were employed. Besides, the regression plot was utilized as a graphical approach. The mathematical expressions of parameters are provided as follows:

$$AARD (\%) = 100 \times \sum_{i=1}^{N_p} \left| \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \right| / N_p \quad (2)$$

$$RD (\%) = 100 \times \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2}{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2} \quad (4)$$

$$R_{adj}^2 = 1 - \frac{(1 - R^2)(N_p - 1)}{N_p - k - 1} \quad (5)$$

$$\text{MSE} = \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p \quad (6)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p} \quad (7)$$

where y_i^{cal} is the predictive data while y_i^{exp} represents the experimental value from literature; the average data is expressed by \bar{y}_m ; N_p denotes the number of data points in a dataset and k means the number of independent variables.

6.4 Results and discussion

6.4.1 Traditional GC model development

The traditional GC model was built using the MLR based on the training set. The regression coefficients as well as *t* and *Sig.* values of groups were provided in [Table 6.2](#). It can find that the *Sig.* values of five groups (group 5, 8, 20, 24, 25) are below 0.01, dedicating that these groups have an extremely significant influence on the toxicity of ILs. Besides, the groups (group 9, 10, 11, 15, 17) with the *Sig.* values within the range of 0.01~0.05 also have a significant impact on the toxicity of ILs. The minus sign of coefficients represents the negative relationship between the calculated parameters and the toxicity of ILs, vice versa.

An example of calculating the predicted toxicity value of the IL (1-Benzyl-3-methylimidazolium tetrafluoroborate) was provided in Supplementary data ([Table S5](#)). The toxicity of ILs in the entire set was predicted by the developed GC model and then the predicted versus experimental data by this model were plotted in [Figure 6.6](#). All the spots are close to the diagonal. The R^2 for the training set and test set are 0.929 and 0.905, respectively. Results indicate a great correlation relationship exists between the calculated parameters and the toxicity of ILs towards IPC-81.

The RD values for different ILs were calculated and Figure 6.7 described the number of ILs located in different RD% ranges. The majority of ILs have the RD% within $\pm 10\%$. The two biggest RD% values belong to the ILs of 1-octylquinolinium tetrafluoroborate and benzyltetradecyldimethylammonium chloride which have lower experimental toxicity data. It probably could be ascribed to the experiment error.

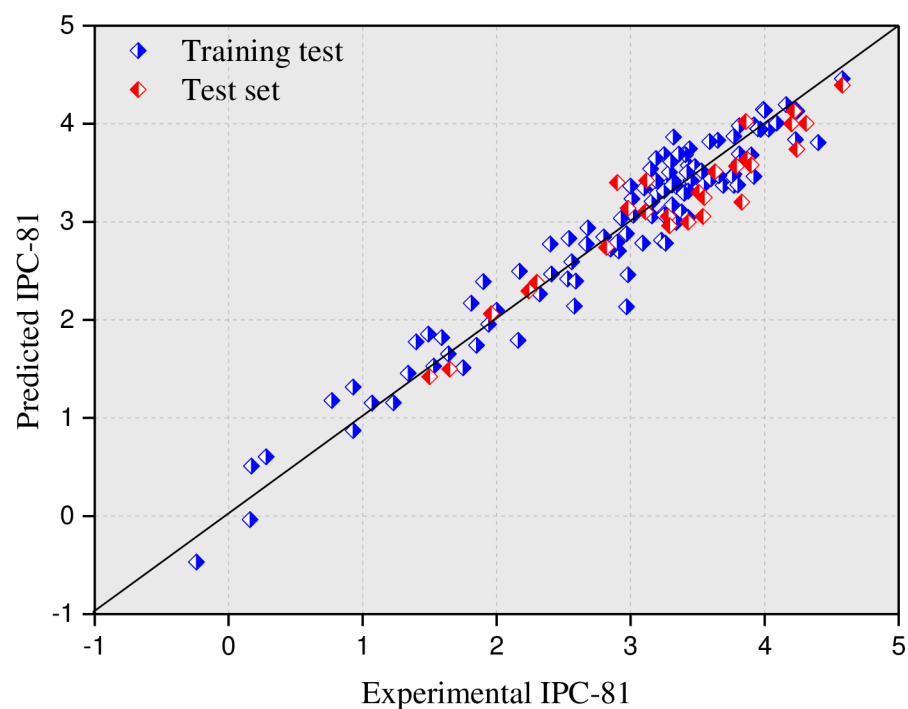


Figure 6.6 Predicted versus experimental toxicity values based on traditional GC method

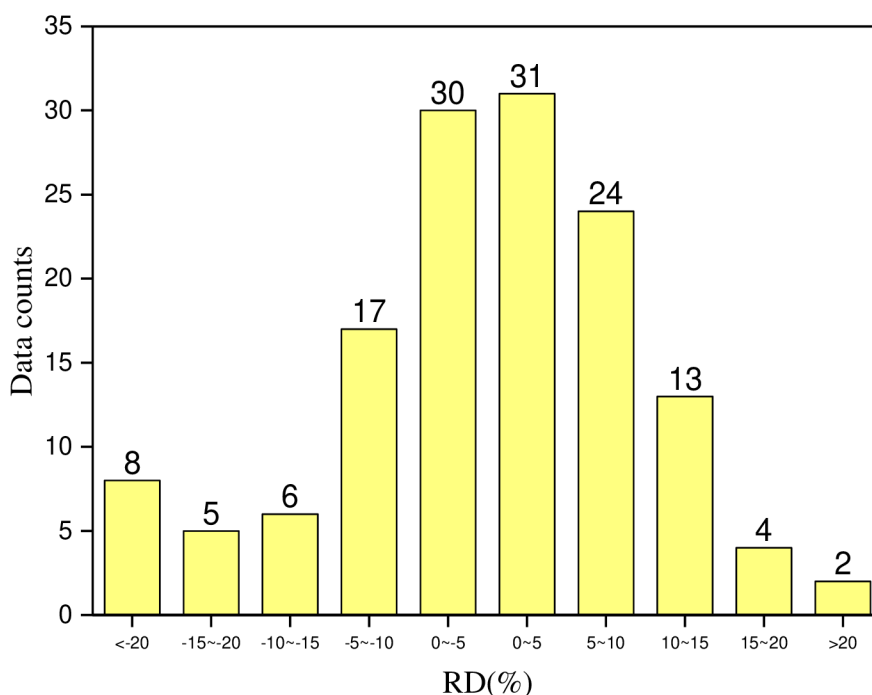


Figure 6.7 Histogram of relative deviation of the traditional GC method

6.4.2 Novel ASFC model development

The training set mentioned in section 2.1 was also used for building the ASFC model by MLR as the establishment of the general GC model. However, the difference is that the $S_{\sigma\text{-surface}}$ values of groups were considered as predictor variables. The regression coefficients, t values and $Sig.$ values of groups were provided in Table 6.4. It can be seen that all the groups made contributions for the dependent variables, which is different from the general GC model. Meanwhile, 14 groups possess the $Sig.$ values below 0.01 and four groups have the $Sig.$ values between 0.01 and 0.05, showing these groups have significant influences on the toxicity of ILs towards IPC-81.

The total data were predicted by the ASFC model. Figure 6.8 displayed the predicted data versus the original data. The scattered points of the training set approaching the diagonal and the R^2 of 0.936 reveal the good fitness of the model while the those of test set shows with $R^2 = 0.907$ the reliability of the model. Figure 6.9 exhibited the distribution of RD% for various ILs with 102 out of 140 ILs located in the range of $\pm 10\%$. Results demonstrate that the proposed ASFC model is relatively

reliable and accurate to forecast the toxicity of ILs towards IPC-81.

Additionally, 15 experimental data of ILs against IPC-81 were collected from the literature (Wu et al., 2020) as an external set (see [Table S2](#) in the Supplementary data) for the verification of the ASFC model. Then all the data including the training set, the test set and the external set were used as a training set to develop model. It can be seen from [Table 6.4](#), the R^2 of the two models were 0.886 and 0.911, respectively, which were satisfactory.

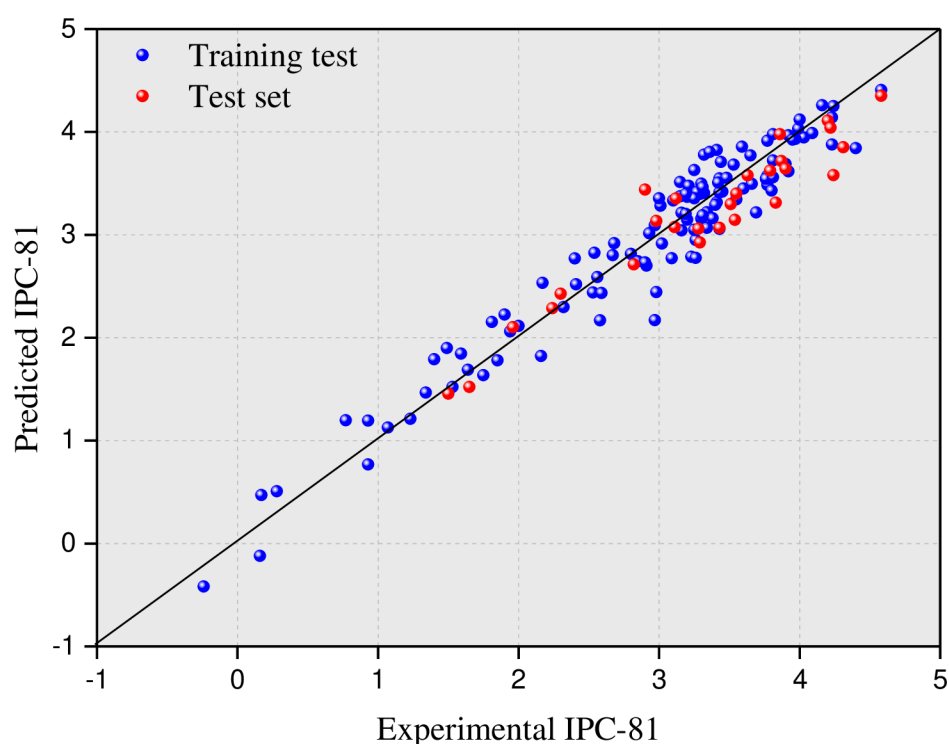


Figure 6.8 Predicted versus experimental toxicity values based on the proposed method

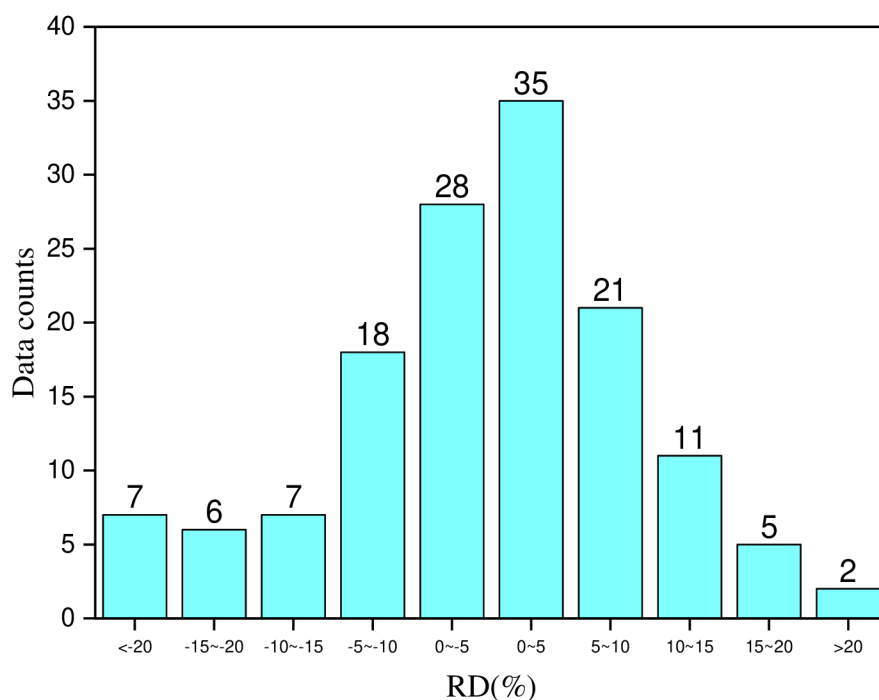


Figure 6.9 Histogram of relative deviation of the ASFC method

Table 6.4 The verification results of the model

Dataset	Number of data points	R^2	R_{adj}^2	ASFC model		
				MSE	RMSE	AARD%
The external dataset	15	0.886	0.877	0.078	0.280	6.114
The whole dataset as training set	155	0.911	0.893	0.086	0.294	10.613

6.4.3 Comparisons

From Table 6.5, it can be seen that group 3 (pyridinium), an important cation core for ILs, was removed because of the collinearity when the traditional GC model was established, but all the groups were considered in the novel ASFC model. As a result, the ASFC model has greater reliability. For the further analysis from Table 6.5, it can be found that the most critical group for the toxicity of ILs towards IPC-81 is group 20, i.e., methylene ($-\text{CH}_2-$). The increase of the number of methylene group highly generally means the longer chains in cations while the coefficient of this group is minus, which means it is negatively correlated with the toxicity values of ILs towards IPC-81.

In fact, our previous work (Kang et al., 2020c; Zhu et al., 2019) and most other studies (Abramenko et al., 2020b; Gao et al., 2020) also gave a conclusion that the cations play a remarkable role in the toxicity of ILs and a longer length of chain in cations of ILs usually shows more toxic behavior. This shows the consistency between the results of this study and the obtained deduction from the literature.

Table 6.6 compared the statistical parameters of the traditional GC model and ASFC model. It can be observed that the R^2 and R^2_{adj} values of and ASFC model are higher than those of the traditional GC model, while the AARD%, MSE and RMSE of the ASFC model are lower than those of the traditional GC model. Therefore, the proposed ASFC model shows better performance than the traditional GC model, demonstrating that the novel ASFC approach is more suitable and applicable for evaluating the toxicity of ILs than the traditional GC method.

Table 6.5 The coefficients, t values and $Sig.$ values of groups in different models

Group No.	Traditional GC model				ASFC model			
	coefficients	t value	$Sig.$ value	coefficients	t value	$Sig.$ value		
1	β_1 0.084	0.586	0.560	β_1 0.044	-1.433	0.156		
2	β_2 -0.215	-0.672	0.503	β_2 -0.179	-2.365	0.020		
3*	β_3 /	/	/	β_3 -0.053	-1.812	0.074		
4	β_4 -0.381	-0.791	0.431	β_4 -0.198	-2.450	0.016		
5	β_5 0.181	0.541	0.590	β_5 0.010	0.700	0.486		
6	β_6 -1.655	-6.567	3.60E-09	β_6 -0.067	-2.744	0.007		
7	β_7 -0.134	-0.397	0.693	β_7 -0.133	-2.281	0.025		
8	β_8 -2.779	-3.728	3.43E-04	β_8 -0.030	-6.956	6.47E-10		
9	β_9 -1.846	-2.510	0.014	β_9 -0.077	-6.101	2.93E-08		
10	β_{10} -1.604	-2.153	0.034	β_{10} -0.059	-5.906	6.81E-08		
11	β_{11} -1.569	-2.154	0.034	β_{11} -0.127	-5.979	4.98E-08		
12	β_{12} -1.373	-1.845	0.068	β_{12} -0.111	-5.789	1.13E-07		
13	β_{13} -1.385	-1.837	0.070	β_{13} -0.096	-5.653	2.01E-07		
14	β_{14} -1.446	-1.894	0.062	β_{14} -0.076	-5.695	1.68E-07		
15	β_{15} -1.697	-2.357	0.021	β_{15} -0.103	-6.510	4.81E-09		
16	β_{16} -0.365	-0.953	0.343	β_{16} -0.014	-1.476	0.144		
17	β_{17} -1.568	-2.055	0.043	β_{17} -0.136	-6.219	1.75E-08		
18	β_{18} 0.113	0.446	0.657	β_{18} 0.004	0.602	0.549		
19	β_{19} 0.048	0.456	0.650	β_{19} 0.014	1.834	0.070		
20	β_{20} -0.321	-20.719	2.13E-35	β_{20} -0.016	-22.113	3.06E-37		

Group No.	Traditional GC model				ASFC model			
	coefficients	<i>t</i> value	<i>Sig.</i> value	coefficients	<i>t</i> value	<i>Sig.</i> value		
21	β_{21} 0.119	1.524	0.131	β_{21} 0.187	2.852	0.005		
22	β_{22} 0.057	0.521	0.604	β_{22} 0.009	0.724	0.471		
23	β_{23} 0.253	1.324	0.189	β_{23} 0.036	2.401	0.019		
24	β_{24} -0.817	-2.888	0.005	β_{24} -0.079	-3.205	0.002		
25	β_{25} -1.508	-7.058	3.88E-10	β_{25} -0.696	-7.739	1.78E-11		
26	β_{26} 0.050	0.141	0.888	β_{26} 0.014	1.211	0.229		
	β_0 5.374	5.825	9.42E-08	β_0 8.955	3.907	1.86E-04		

Note: *The group 3 was removed because of its collinearity with the dependent variable during the development of traditional GC model.

Table 6.6 The statistical parameters of models in this work

Dataset	Number of data points	Traditional GC model					ASFC model				
		AARD%	R^2	R^2_{adj}	MSE	RMSE	AARD%	R^2	R^2_{adj}	MSE	RMSE
train	113	12.368	0.929	0.909	0.069	0.262	11.891	0.936	0.916	0.062	0.250
test	27	7.129	0.905	0.901	0.082	0.286	6.742	0.907	0.904	0.078	0.280
total	140	11.358	0.924	0.924	0.071	0.267	10.898	0.930	0.930	0.065	0.256

6.5. Conclusion

In this study, a dataset involving 140 ILs was collected from the widely acknowledged ILs database and the $S_{\sigma\text{-surface}}$ values of groups of various ILs were calculated and utilized as the predictor parameters for model training. A novel AFSC was proposed to evaluate the toxicity of ILs towards IPC-81. This approach can distinguish the contribution of isomers and even the same group in different molecules. It can be deduced that the group methylene has an extremely significant contribution to the toxicity of ILs towards IPC-81. Moreover, results imply that the ASFC model has better performance than the model constructed by the traditional GC method. The proposed ASFC model is reliable, precise as well as extensively applicable for forecasting the toxicity of ILs towards IPC-81. This work built a linear model by the ASFC method which presents the novelty, superiority and potential applications of this technique. Besides, the proposed method has the great potential to be widely applied in the evaluation of other properties of ILs and other compounds. Meanwhile, it can be possibly linked with advanced machine learning methods to establish the nonlinear models for assessing diverse properties of compounds in our future study.

6.6 Supplementary data

The following link is the Supplementary data to this chapter:

<https://ars.els-cdn.com/content/image/1-s2.0-S0304389421016708-mmc1.xlsx>

Chapter VII

Application of atomic electrostatic potential descriptors for predicting the eco-toxicity of ionic liquids towards *leukemia rat cell line*

Xuejing Kang, Yongsheng Zhao, Hongzhong Zhang, Zhongbing Chen

Adapted from Chemical Engineering Science (2022): 117941

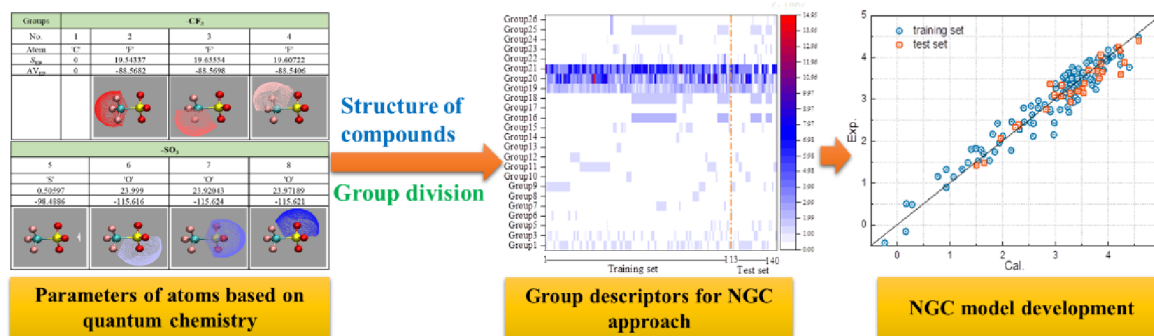
Contents

7.1 Abstract	123
7.2 Introduction	125
7.3 Materials and methods	127
7.3.1 <i>Data acquisition</i>	127
7.3.2 <i>Calculation of AV_{EP} and S_{EP} of atoms</i>	127
7.3.3 <i>Calculation of descriptors for NGC method</i>	132
7.3.4 <i>Model evaluation</i>	136
7.4 Results and discussion	136
7.5 Conclusions	142
7.6 Supplementary data	142

7.1 Abstract

The toxicity assessment of ionic liquids (ILs) towards the environment and living organisms has received great attention. Nevertheless, the huge number of ILs makes the toxicity data collection expensive and time-consuming, which motivates modeling development to fill data gaps of ILs toxicity. The group contribution (GC) method has been extensively applied for the estimation of various properties of ionic liquids. This study proposed a novel method, named the non-integer group contribution (NGC) method, which creatively utilizes the atomic electrostatic potential descriptors for modeling. Specifically, the average values of electrostatic potential (AV_{EP}) and the electrostatic potential surface area (S_{EP}) of atoms in the cations and anions of ILs were calculated and used to obtain the group descriptors in this work. Two NGC models were developed to predict the toxicity of ILs. Results show that both proposed models have satisfactory predictability. In contrast, the NGC-2 model based on S_{EP} descriptors exhibits better predictability due to its higher coefficient of determination ($R^2=0.927$), the lower average absolute relative deviation (AARD =11.257%) and root mean square error (RMSE=0.261) for the entire dataset. The NGC-2 model also shows better performance than the traditional GC method, demonstrating its advanced superiority. Therefore, the proposed approach has high potential in terms of generalization and applicability for predicting the property of compounds.

Graphical abstract:



7.2 Introduction

Product property data plays a significant role in modern chemical industries (Gani, 2019). Prediction models for molecular properties have been a strong demand as experimental data may be costly or time-consuming (Frutiger et al., 2016). Diverse methods have been implemented to develop estimation models for crucial issues in different fields, including chemical industry, chemical risk assessment, drug discovery, new material research, environmental impact evaluation, etc. (Kochev et al., 2019).

In recent decades, ionic liquids (ILs) have become a worldwide research topic and are ideal compounds for various applications, including catalysts, reaction solvents, electrolytes, drug delivery systems, lubricants, extraction media, and so on (Koutsoukos et al., 2021). ILs are considered benign solvents and can be alternatives to volatile organic solvents due to their advantageous physical and chemical properties, especially their wide liquid range and negligible vapor pressure (Lazidou et al., 2019; Xue Liu et al., 2018; Wang et al., 2016, 2021). Additionally, the fine-tuned feature of ILs is also a dominant property that makes them satisfy the requirements of target applications (Costa et al., 2017; Deng et al., 2010; Elfggen et al., 2020). However, the toxicity of ILs has been pointed out and has attracted much attention and concern because of their water-solubility and inevitable emissions in industrial applications (Cui et al., 2021; Thuy Pham et al., 2010). Due to the huge number of ILs (Cho and Yun, 2019; Zhao et al., 2016c), many researchers have reported prediction models for various properties of ILs using different techniques (Chen et al., 2020, 2019b; Gardas and P.Coutinho, 2009; Valderrama and Robles, 2007). The group contribution (GC) method is one of the most popular approaches for modeling (Eini et al., 2020; Gani, 2019). This method has developed over several decades (Simon, 1956; Van Krevelen and Chermin, 1951), and over 53,600 publications are available on the Web of Sciences in 2021.

The GC method can be widely used as it can provide fast and relatively accurate prediction (Kashinath et al., 2020; Mital et al., 2021). There is a basic assumption in the GC methods that the additive contribution of the molecule's functional groups can represent its physical property. Hence, the property of a compound is generally defined as a function that depends on the frequency of groups representing a molecule and their contribution (Gani et al., 2005; Klopman et al., 1994; Mital et al., 2021).

Various GC-based techniques for predicting various properties of compounds and mixtures have been widely applied and reported (Boethling et al., 1994; Eini et al.,

2020; Joback and Reid, 1987; Marrero and Gani, 2001; Serat et al., 2017; Simon, 1956; Van Krevelen and Chermin, 1951). Among these approaches, the one proposed by Marrero and Gani is well known (Eini et al., 2020; Marrero and Gani, 2001). The approach considered group contributions to perform the predictions at three levels: first-order groups describing a wide range of organic compounds; second-order groups providing the description of polyfunctional compounds and the differentiation of isomers; the third-order groups representing the systems of fused aromatic and/or nonaromatic rings as well as non-fused rings connected by chains (Marrero and Gani, 2001). Their method overcomes some drawbacks of traditional GC methods, such as unable to distinguish isomers, and limited applicability, exhibiting good performance. Thus, the method obtains wide attention and application, but it is quite complex (Hukkerikar et al., 2012; Marrero and Gani, 2001). The fragment contribution models (FCMs) are also employed GC-based methods using fragment counts as descriptors (Endo and Hammer, 2020). It was proposed by Hansch and Leo (1995) and is also widely adopted for assessing environmental properties. For example, Endo S. and Jort Hammer J. (2020) applied FCMs to predict partition coefficients of short-chain chlorinated paraffin congeners. Researchers also attempted to improve the accuracy of GC models by combining other algorithms (Y. Huang et al., 2013; Hukkerikar et al., 2012). However, since the group parameters are consistently acquired by counting the number of a group/fragment occurring in a molecule, and are only integer values, the information content of molecules from these parameters is limited.

Quantum chemistry (QC), also named molecular quantum mechanics, can provide more rich information about a molecular at the scale of atoms and has been increasingly applied in various areas of physics and chemistry for studying the properties and behaviors of molecules (Bocharnikova et al., 2020; Helgaker et al., 2008). To improve the accuracy of the classical GC method, this work proposed a new method named the non-integer group contribution method (NGC), which is based on QC calculations to calculate the group descriptors. In the new method, a single group is considered to have different contributions because of its special molecular environments, and the differences are reflected by a new approach to obtaining the frequency of groups existing in different molecules. For instance, molecular electrostatic potential (EP) may provide rich and accurate information about the active sites of several chemical entities (Bulat et al., 2010), and it is considered a significant concept for molecular modeling calculations and assessing chemical nature (Bayoumy et al., 2020). Hence, EP is used to determine the frequency of groups in a molecule for the new method. Besides, the electrostatic potential surface (S_{EP}) is a popular way of describing the electrostatic nature of molecules and has been applied to predict molecular properties (Rathi et al., 2019). In our previous study, the S_{EP} values of separate ions and entire molecules have

been successfully applied to build estimated models for various properties of ionic liquids (ILs) (Kang et al., 2020a, 2020b, 2018c, 2018d; Zhu et al., 2019). Distinguishing from our previous work, the EP and S_{EP} of atoms in this work were first obtained and introduced for calculating the frequency of groups in a molecule, and then creatively applied to establish novel GC models to evaluate the properties of compounds. However, the EP and S_{EP} of atoms and any character descriptor of atoms calculated by QC can be considered for modeling in the NGC method.

In this study, we targeted building more reliable GC models to evaluate the toxicities of ILs by proposing the novel NGC method. In total, the main objectives of our study are 1) to propose a new type of GC method (NGC) based on QC calculation; 2) to calculate the average EP and S_{EP} values to each atom of ILs; 3) to build new models to assess the toxicity of ILs towards the *leukemia rat cell line* (IPC-81) by the proposed NGC method; 4) to compare the performance of models established by the NGC and traditional GC methods.

7.3 Materials and methods

7.3.1 Data acquisition

To compare with our previous work for assessing the toxicity of ILs towards IPC-81 (Kang et al., 2021), the same datasets used were applied in this work. The toxicity of ILs was presented by $\log EC_{50}$ (μM), which means the logarithm of the half-maximal effective concentration. All of the data come from the widely acknowledged ILs database (The UFT/ Merck Ionic Liquids Biological Effects Database; Zhang et al., 2006), involving a training set (113 ILs) for developing models and a test set (27 ILs) for verifying the predictive ability of models.

7.3.2 Calculation of AV_{EP} and S_{EP} of atoms

The electrostatic potential of a molecule is generated by its nuclei and electrons around the surface of the molecule (Murray and Politzer, 2017; Rathi et al., 2019). Eq. (1) is the general formula of the electrostatic potential at any point \mathbf{r} .

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|R_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \quad (1)$$

where Z_A means the charge on nucleus A at the location of R_A ; $\rho(r')$ symbols the electronic density of the molecular at r' . $|R_A - r|$ and $|r' - r|$ represent the corresponding distances from r , respectively (Kolář and Hobza, 2016; Murray and Politzer, 2011a).

Here, the average values of electrostatic potential (AV_{EP}) and the electrostatic potential surface area (S_{EP}) of each atom in cations and anions of ILs were calculated and employed to obtain descriptors used in the new NGC method. The S_{EP} of atoms also means the local molecular surface area of electrostatic potential for each atom. It should be noted that the electrostatic potential surface employed in this work is the general definition of the van der Waals surface proposed by Bader (Bader et al., 1987; Bulat et al., 2010).

To obtain the AV_{EP} and S_{EP} values of atoms, all the geometrical structures of the 74 cations and 15 anions of ILs included in the whole dataset were first optimized by the Gaussian 16 C.01 software at the theoretical level B3LYP/6-311++g(d,p). Subsequently, the AV_{EP} and S_{EP} values of atoms in different cations and anions were calculated by the Multiwfn program (Lu and Chen, 2012) based on the optimized structures. [Figure 7.1](#) mapped the AV_{EP} and S_{EP} values of atoms 1-20 for a representative cation (1-(2-hydroxyethyl)-3-methylimidazolium) on the corresponding local surface, while [Figure 7.2](#) described the AV_{EP} and S_{EP} values of each atom for a representative anion (Trifluoromethanesulfonate). The surface regions of each atom were also mapped in [Figure 7.1](#) and [Figure 7.2](#), where different colors represent various S_{EP} and AV_{EP} . The method for group division is the same as in our previous work (Kang et al., 2021). Accordingly, the structures of collected ILs contained 26 groups, which were listed in [Table 7.1](#). The frequencies of groups in cations and anions acquired by the traditional GC method were depicted in [Figure 7.3](#), and their specific values were summarized in Supplementary data ([Table S7](#)), including the full names and abbreviations of cations and anions in these ILs.

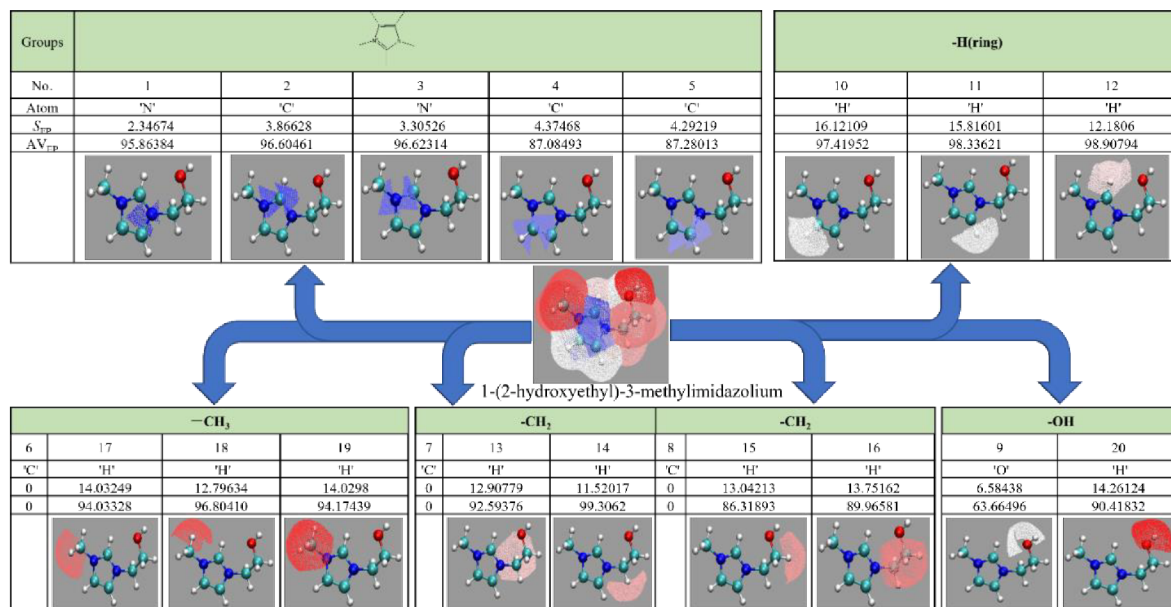


Figure 7.1. The S_{EP} values, AV_{EP} values and the surface regions of atoms for a representative cation (1-(2-hydroxyethyl)-3-methylimidazolium)

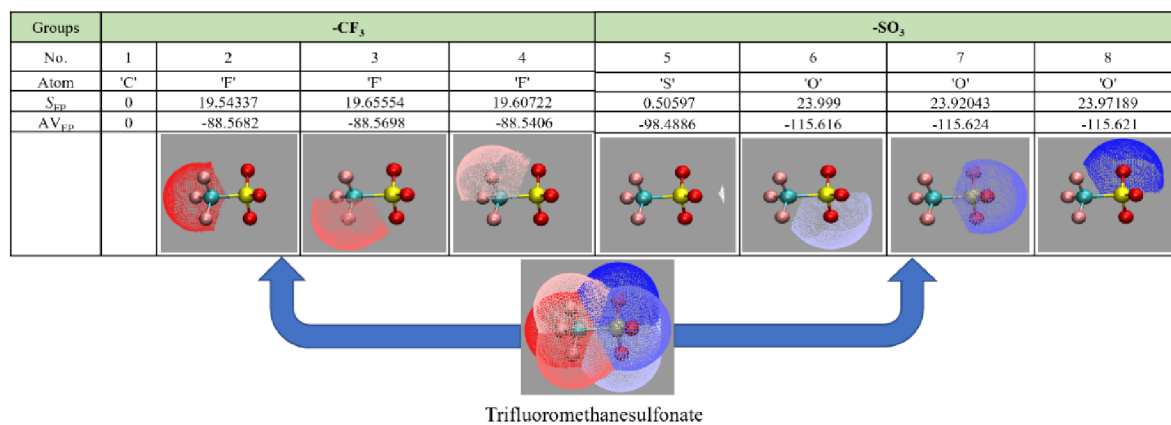
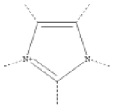
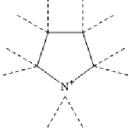
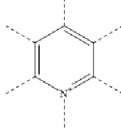
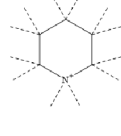
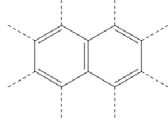
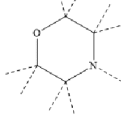
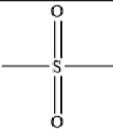
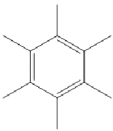


Figure 7.2. The S_{EP} values, AV_{EP} values and the surface regions of atoms for a representative anion (Trifluoromethanesulfonate)

Table 7.1. Structures and average AV_{EP} and S_{EP} values of groups in ILs

Group No.	Structures	AV_{SEP-m} of groups (\AA^2)	AV_{AVEP-m} of groups (KJ/mol)
Cation cores			
1*		17.9186	470.8723
2*		0	0
3*		23.4064	549.0793
4*		0	0
5	N-(CH ₂) ₂ (CH ₃) ₂	95.5149	982.8595
6*		42.7083	807.5683
7*		10.8042	51.1633
Anion cores			
8*	[TPTP] ⁻	257.4821	-1406.8239
9	[BF ₄] ⁻	91.6475	-614.7037
10	[PF ₆] ⁻	112.0738	-673.4562
11	[Cl] ⁻	68.8858	-137.8724
12	[Br] ⁻	74.6353	-130.6107
13	[I] ⁻	81.5302	-128.0210
14*	[SCN] ⁻	96.3014	-351.1232
Substituents			
15	-SO ₃	72.4665	-466.9298

Application of atomic electrostatic potential descriptors for predicting the eco-toxicity of ionic liquids towards leukemia rat cell line

Group No.	Structures	AV_{SEP-m} of groups (\AA^2)	AV_{AVEP-m} of groups (KJ/mol)
16		42.4020	-269.4060
17	-COO-	54.4661	-355.1193
18	-CF ₃	56.7387	-226.9724
19	-CH ₃	38.2560	211.4609
20	-CH ₂ -	21.7449	149.9960
21	-H(ring)	12.8905	92.3870
22	-O-or [-O] ⁻	5.8577	47.9049
23	-OH	22.1680	78.2402
24		23.0875	-73.5647
25	[-N-]- or > N-	8.6645	-81.9983
26	-CN	39.8786	-38.4675

Note:1* imidazolium; 2* pyrrolidinium; 3* pyridinium; 4* piperidinium; 6* quinolinium; 7* morpholinium; 8* trifluorotris(pentafluoroethyl)phosphate;14* thiocyanate.

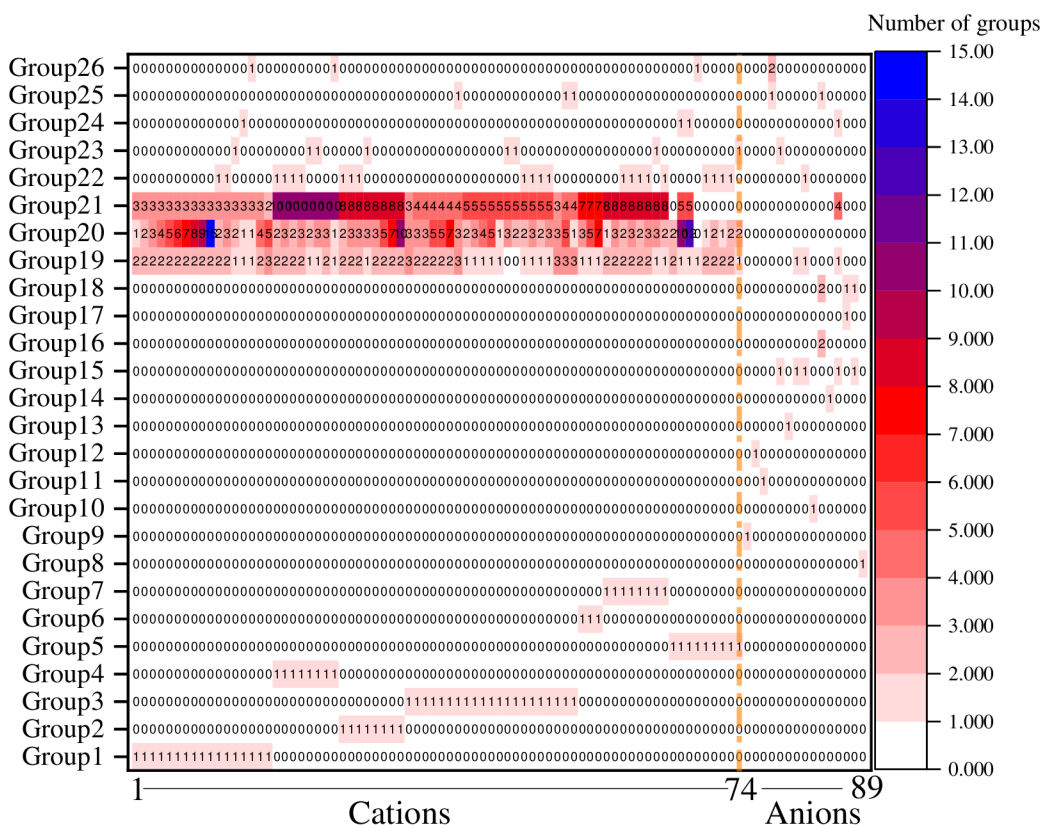


Figure 7.3. The number of groups in different cations and anions

7.3.3 Calculation of descriptors for NGC method

As we mentioned above, in the NGC method, the frequency of a group is calculated based on QC calculations instead of an integer value by counting. In this work, we used two descriptors to create NGC models for prediction based on the calculation of AV_{EP} and S_{EP} values of atoms.

Generally, GC models are developed using multiple linear regression (MLR) algorithms. For instance, our previous work (Kang et al., 2021) employed the formula of the classical GC method (Eq.(2)) :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \quad (2)$$

where m represents the m -th group in an IL; X_m is the number of the corresponding group occurring in an IL, including its cation and anion; β_m is the regressed contribution of the corresponding group; Y is the dependent variable, which usually is

the predictive data (Lazzús, 2012; Mu et al., 2007). Figure 7.4 presents the X_m values of 26 groups in the 140 ILs for the traditional GC model, while the specific values were presented in Supplementary data (Table S4).

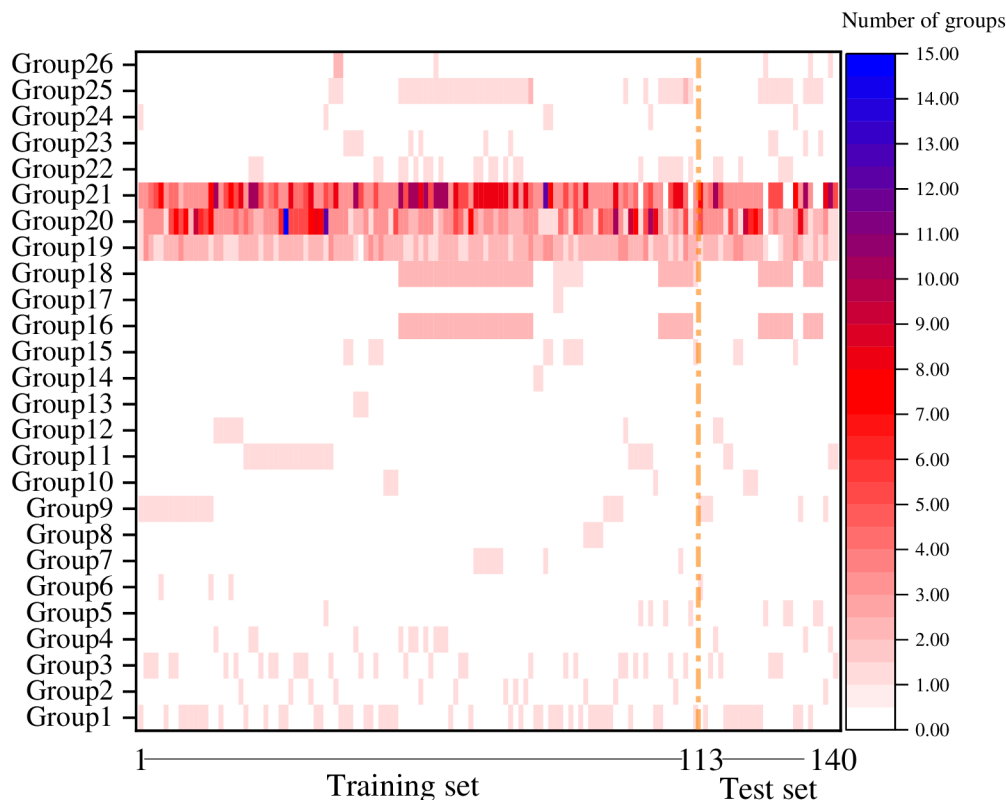


Figure 7.4. The number of groups of the training set and test set used in the traditional GC model

However, in the proposed NGC method, the descriptor (X_m) was revised and replaced by the ratio (X'_m) of a specific property value to the average property value of the group in a specified dataset, which is a novel method for calculating the frequency of groups. Additionally, the property value of a group is generally calculated by the QC methods. The new NGC method was represented by Eq. (3).

$$Y = \beta'_0 + \beta'_1 X'_1 + \beta'_2 X'_2 + \dots + \beta'_m X'_m \quad (3)$$

Two methods for calculating X'_m are provided in this work, the ratios of the AV_{EP} (and S_{EP}) of a group in one IL to the corresponding average AV_{EP} (and S_{EP}) value of the group in the collected dataset, while the values of X'_m are usually non-integer or even minus. The two techniques for X'_m in this work are expressed as Eq. (4) and Eq. (5),

respectively.

$$X'_m = \frac{AV_{EP-i,m}}{AV_{AVEP-m}} \quad (4)$$

$$X'_m = \frac{S_{EP-i,m}}{AV_{SEP-m}} \quad (5)$$

$$AV_{AVEP-m} = \frac{\sum_{i=1}^{N_p} AV_{EP-i,m}}{N} \quad (6)$$

$$AV_{SEP-m} = \frac{\sum_{i=1}^{N_p} S_{EP-i,m}}{N} \quad (7)$$

where i notes the i -th IL in the dataset, while m means the m -th group in the IL; N is the number of each group appearing in the whole dataset, while N_p is the number of ILs in the corresponding dataset. Eq. (6) and Eq. (7) show the calculation process, which stands for the average AV_{EP} value of a group containing all the ILs of the corresponding dataset. Table 7.1 exhibits the structures, the AV_{AVEP-m} and AV_{SEP-m} values of groups in ILs used in this study. The non-integer number of groups in 74 cations and 15 anions were summarized in Supplementary data (Tables S5 and S6), respectively. The final descriptors (X'_m) of a IL can be obtained by summing of the non-integer number of groups of the cation ($X'_{m-cation}$) and anion ($X'_{m-anion}$). Two classes of the calculated descriptors (X'_m) of 140 ILs were plotted in Figure 7.5 (a) and (b), respectively, while the detailed values were provided in the Supplementary data (Table S2 and Table S3). It is necessary to note that some atoms may have no contribution to the van der Waals surface when they are sandwiched by the atoms of the outer layers, leading to the zero AV_{AVEP-m} and AV_{SEP-m} values. Because the AV_{AVEP-m} and AV_{SEP-m} values of atoms in groups 2 and 4 were zero in all ILs, the corresponding descriptors (X'_m) of them were also zero, so groups 2 and 4 were excluded in the input descriptors, as shown in Figures 7.5 (a) and (b). Then, all the descriptors were employed to build NGC models for toxicity estimation of ILs by the MLR method.

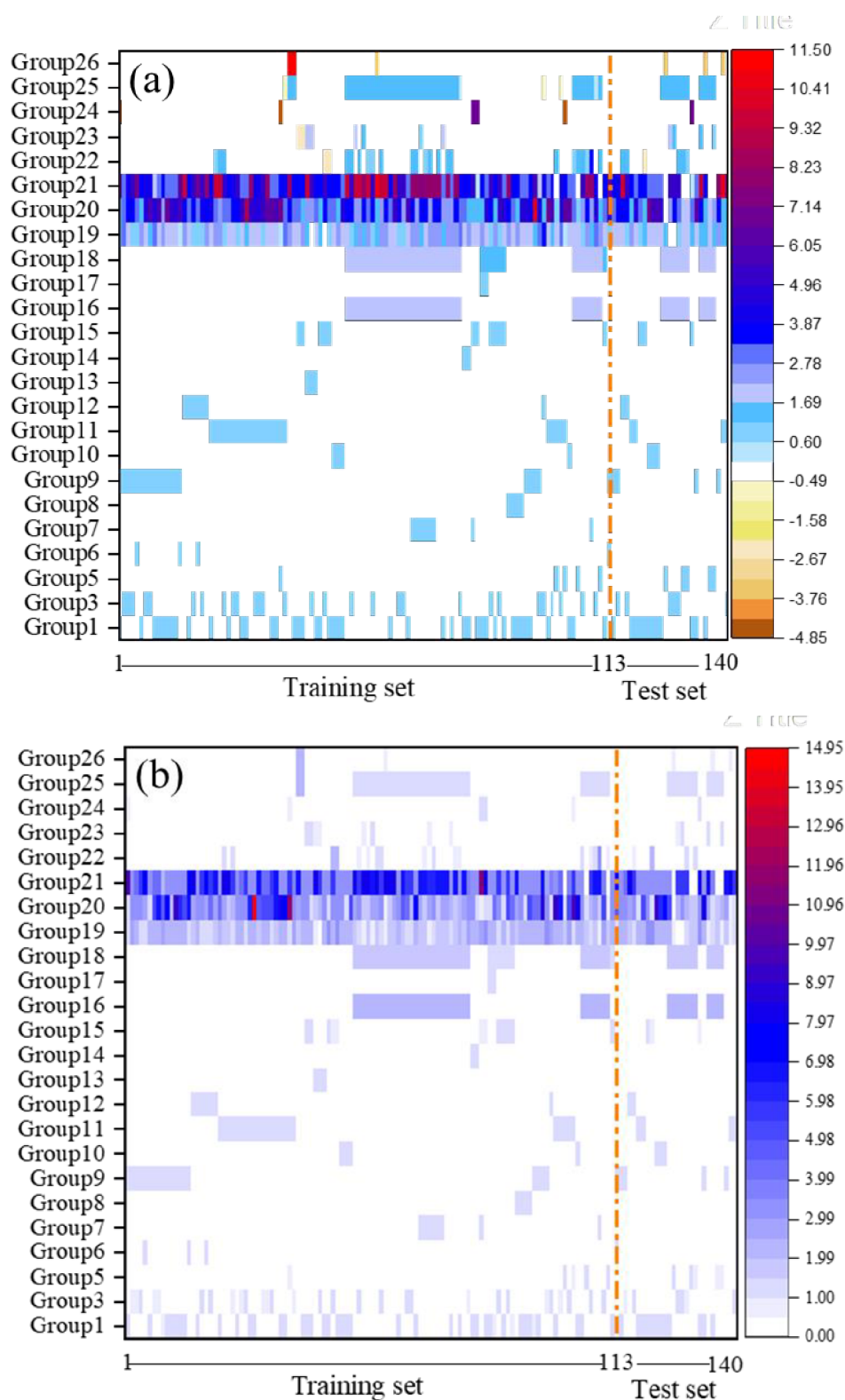


Figure 7.5. The descriptors used in (a) the NGC-1 model based on AV_{EP} and (b) the NGC-2 model based on SEP

7.3.4 Model evaluation

Five statistical indexes, including coefficient of determination (R^2), absolute relative deviation (ARD) and average ARD (AARD), mean square error (MSE), and root MSE (RMSE), are applied to investigate the goodness-of-fit and performance of developed estimation models. The expressions (Eq.8-12) of them are listed as follows:

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i^{cal} - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2}{\sum_{i=1}^N (y_i^{cal} - \bar{y}_m)^2} \quad (8)$$

$$AARD (\%) = 100 \times \sum_{i=1}^{N_p} \left| \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \right| / N_p \quad (9)$$

$$ARD (\%) = 100 \times \left| \frac{y_i^{cal} - y_i^{exp}}{y_i^{exp}} \right| \quad (10)$$

$$MSE = \sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p \quad (11)$$

$$RMSE = \sqrt{\sum_{i=1}^{N_p} (y_i^{cal} - y_i^{exp})^2 / N_p} \quad (12)$$

where y_i^{cal} represents the estimated toxicity data of ILs; y_i^{exp} stands for the experimental toxicity data of ILs from literature; \bar{y}_m notes the corresponding average data in different datasets; N_p is the number of total data points in a specified dataset.

7.4 Results and discussion

Because the AV_{EP} and S_{EP} values of group 2 (pyrrolidinium) and group 4 (piperidinium) in the ILs are zero, their descriptors were removed when models were developed. Then the rest of the 24 descriptors were employed to create NGC models by the MLR algorithm. NGC-1 and NGC-2 models based on AV_{EP} and S_{EP} values of groups were established, respectively. The contributions of different groups were obtained and presented in Table 7.2, including the t and $Sig.$ values of the descriptors. The minus in front of a contribution value indicates the corresponding parameter has a negative impact on the toxicity of ILs. In contrast, the contribution with a positive value stands for a positive correlation between the toxicity of ILs and the parameter. Besides, the larger the absolute t value and the lower the $Sig.$ value, the more important the corresponding descriptor. As can be seen from Table 7.2, the group 20 of methylene (-

CH₂-) with the most significant absolute t value and smallest Sig. value is the most crucial parameter for the toxicity of ILs. The negative sign in front of the contribution of β_{20} reveals that the higher the AV_{EP} or S_{EP} of the -CH₂- group, the lower the toxicity value of the IL, and the more toxic the IL. Meanwhile, most -CH₂- groups are from cations, which can be seen from Figure 7.3. Thus, an IL with a greater number of -CH₂- groups or longer alkyl chains in its cation is always more poisonous. This can be confirmed by experimental toxicity values of ILs from the literature. For example, Figure 7.6 compared the experimental toxicity results of ILs acquired in this work with the same anion ([BF₄]⁻) but different cations, implying that the ILs with a shorter alkyl chain in the cation have higher toxicity data and indeed weaker toxicity. This conclusion is also consistent with some reported studies (Abramenko et al., 2020b; Kang et al., 2020c; Peng and Picchioni, 2020; Zhu et al., 2019).

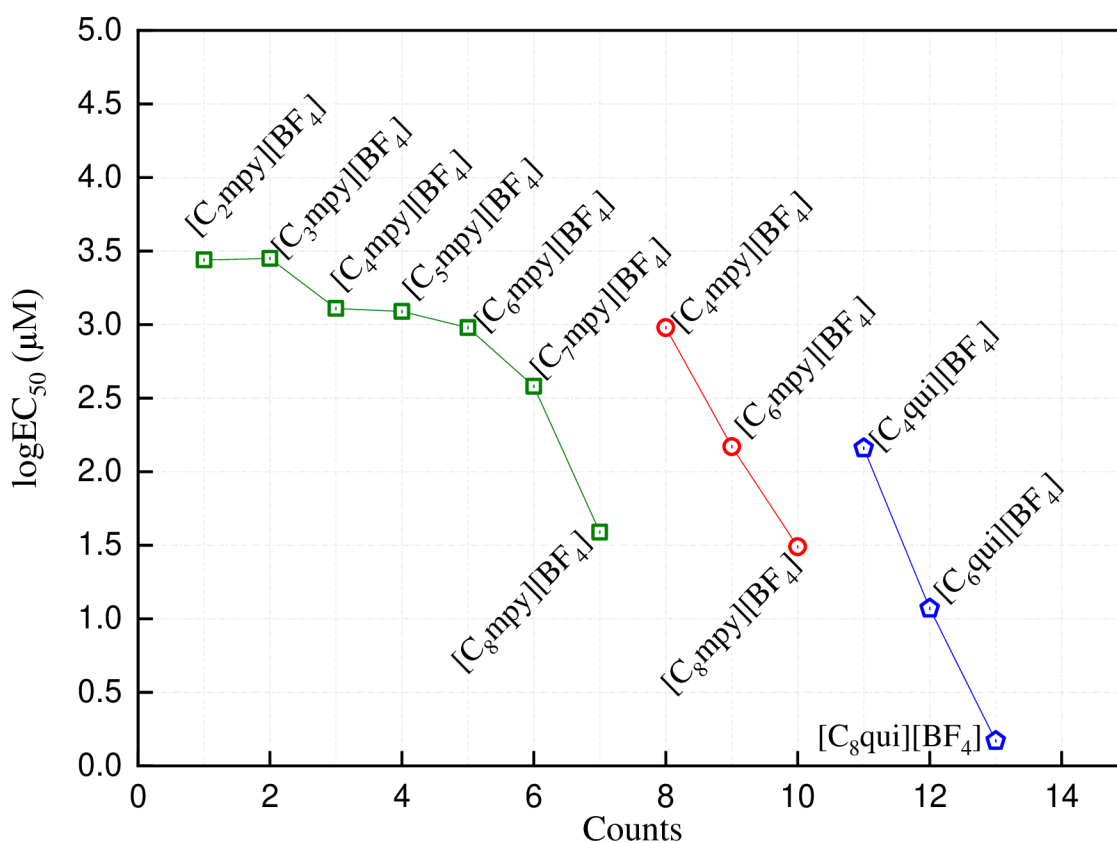


Figure 7.6. The toxicity comparison between the ILs with the same anion and various cations

Table 7.2. The coefficients, t values and Sig. values of groups in different models

Group No.	NGC-1 model (AV_{EP})			NGC-2 model (S_{EP})				
	contributions	t value	Sig. value	contributions	t value	Sig. value		
1	β_1	-2.000	-3.817	2.51E-04	β_1	0.118	0.360	0.720
2	β_2	/	/	/	β_2	/	/	/
3	β_3	-1.590	-3.289	0.001	β_3	0.055	0.206	0.837
4	β_4	/	/	/	β_4	/	/	/
5	β_5	-3.579	-4.096	9.32E-05	β_5	0.212	0.319	0.751
6	β_6	-2.841	-6.138	2.35E-08	β_6	-1.554	-7.399	7.68E-11
7	β_7	-0.252	-1.363	0.176	β_7	0.183	1.056	0.294
8	β_8	2.006	1.417	0.160	β_8	-14.158	-7.947	5.95E-12
9	β_9	2.975	2.087	0.040	β_9	-13.249	-7.384	8.26E-11
10	β_{10}	3.232	2.265	0.026	β_{10}	-12.986	-7.265	1.43E-10
11	β_{11}	3.251	2.278	0.025	β_{11}	-12.923	-7.272	1.38E-10
12	β_{12}	3.455	2.448	0.016	β_{12}	-12.753	-7.213	1.81E-10
13	β_{13}	3.611	2.495	0.014	β_{13}	-12.809	-7.100	3.06E-10
14	β_{14}	3.326	2.326	0.022	β_{14}	-12.827	-7.154	2.38E-10
15	β_{15}	2.340	1.741	0.085	β_{15}	-12.925	-6.886	8.17E-10
16	β_{16}	-0.402	-0.639	0.525	β_{16}	-3.301	-5.150	1.58E-06
17	β_{17}	1.881	1.440	0.153	β_{17}	-12.767	-6.727	1.68E-09
18	β_{18}	0.915	2.929	0.004	β_{18}	-0.036	-0.122	0.903
19	β_{19}	-0.363	-2.311	0.023	β_{19}	0.040	0.250	0.803
20	β_{20}	-0.558	-12.070	2.32E-20	β_{20}	-0.325	-21.960	1.80E-37
21	β_{21}	-0.263	-2.798	0.006	β_{21}	0.081	0.879	0.382
22	β_{22}	-0.176	-1.923	0.058	β_{22}	-0.017	-0.236	0.814
23	β_{23}	-0.233	-1.723	0.088	β_{23}	0.406	1.696	0.093
24	β_{24}	-0.023	-0.406	0.686	β_{24}	-1.140	-2.463	0.016
25	β_{25}	1.633	6.093	2.85E-08	β_{25}	-6.422	-7.221	1.75E-10
26	β_{26}	0.114	0.910	0.365	β_{26}	0.098	0.234	0.816
	β_0	5.557	3.004	0.003	β_0	16.861	9.654	1.84E-15

The R^2 values of training sets for the NGC-1 and NGC-2 models are 0.880 and

0.928, respectively, showing the satisfied fitness of the models. The predictive and experimental values using the NGC-1 and NGC-2 models for all ILs were plotted in Figures 7.7 (a) and (b), respectively. Most data points are close to the diagonal, demonstrating that the predictive data is quite consistent with the experimental values.

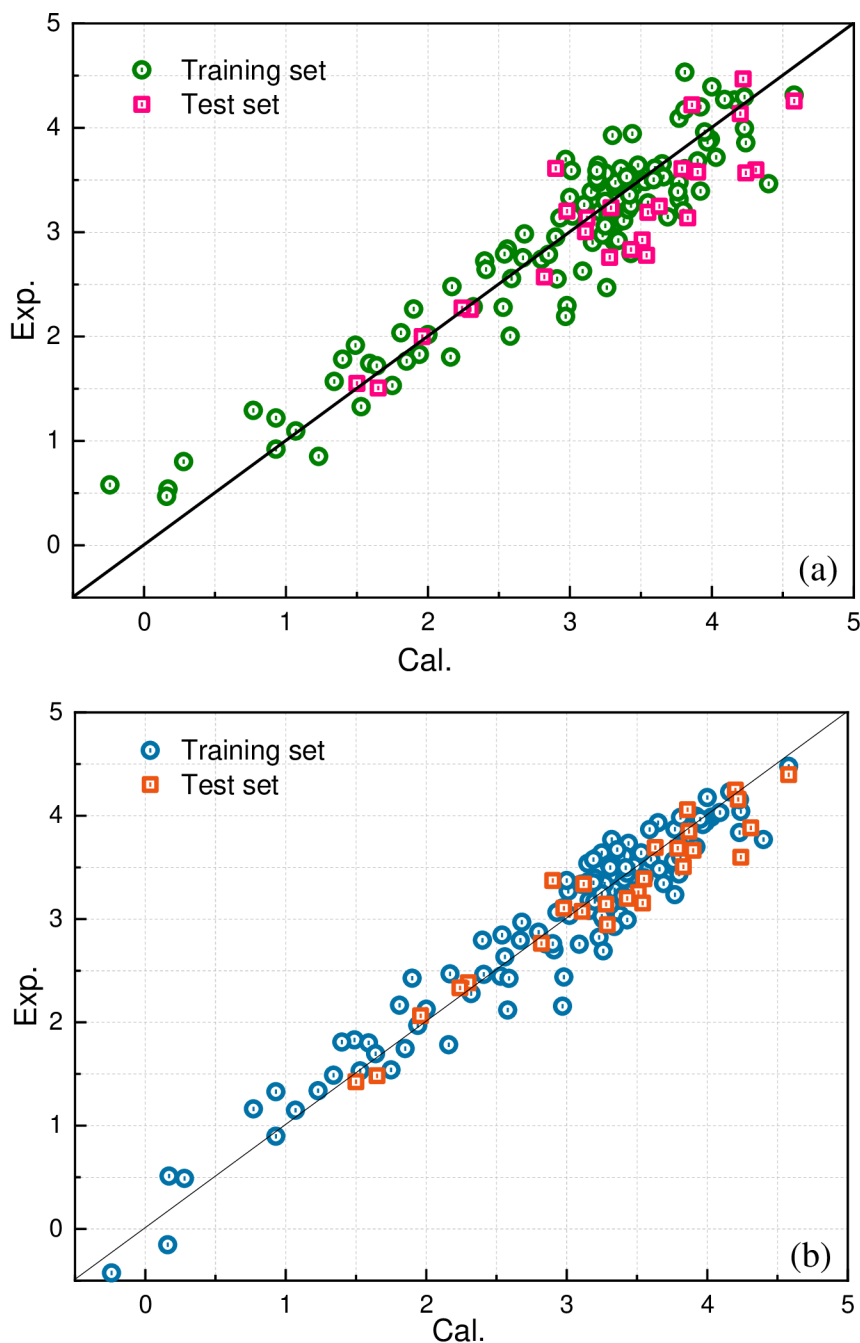


Figure 7.7. Calculated vs experimental toxicity values of ILs by (a) the NGC-1 model based on AV_{EP} and (b) the NGC-2 model based on S_{EP}

For the NGC-1 model, the R^2 for the training, the test, and the total sets are 0.880, 0.808, and 0.866, respectively. The distribution of AAD (%) for the NGC-1 model in different ranges was depicted in Figure 7.8 (a). It is obvious that the majority of AAD (%) values are less than 20%, accounting for 87.86% (123 data points), while the proportion of AAD (%) over 20% is only 12.14% (17 data points). The R^2 of the training, test and total sets for the NGC-2 model are 0.928, 0.920, and 0.927, respectively. Figure 7.8 (b) shows that the percentage of AAD (%) below 20% for the NGC-2 model is 92.86% including 130 data points. Moreover, the other statistical parameters of the NGC-2 model, containing AARD, MSE, and RMSE are also lower than those of NGC-1 model. In total, the NGC-2 model shows better goodness-of-fit and accuracy.

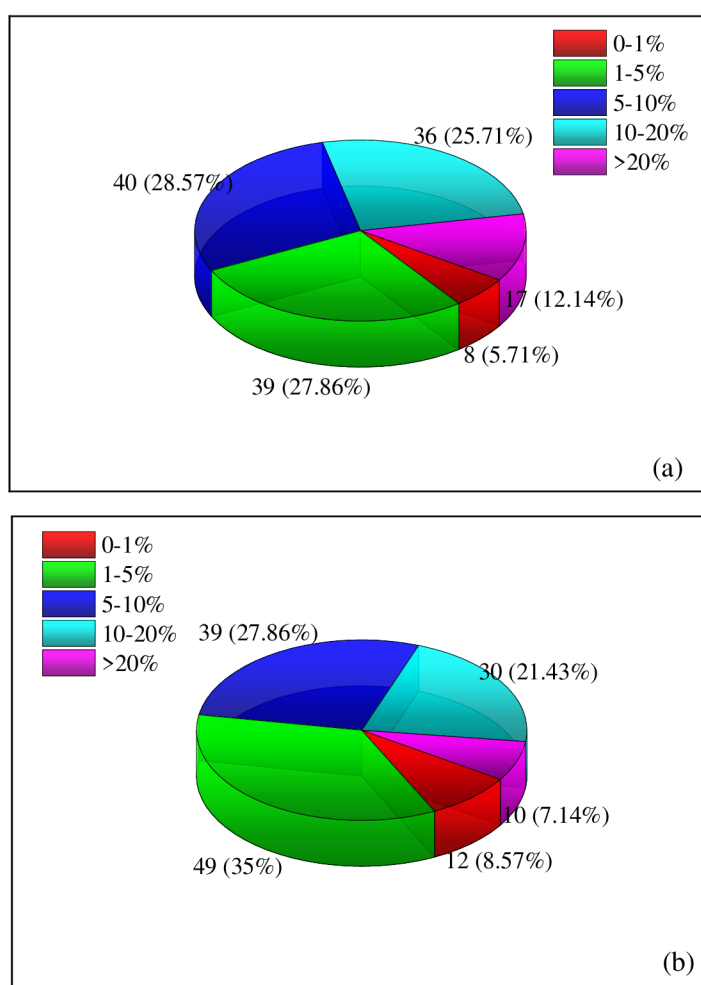


Figure 7.8 ARD% distribution of (a) the NGC-1 model based on AV_{EP} and (b) the NGC-2 model based on S_{EP}

The AARD% values of ILs in various classes were presented in Table 7.3. It can be found that two types of ILs with ammonium and quinolinium have larger AARD%

values. After analyzing, the three largest AARD% values belong to the ILs, [C₈qui][BF₄], [N_{1,1,12}Bn][Cl] and [N_{1,1,14} Bn][Cl], respectively, which should be attributed to the experimental error.

The comparison between the proposed NGC models and the traditional GC model created in our previous study is summarized in Table 7.4. It can be observed that the NGC-2 model depending on S_{EP} values, has higher R^2 and lower AARD, MSE, and RMSE, revealing better performance than the traditional GC model. We also provided an example of predicting a new IL by the models GC, NGC-1, and NGC-2 for better understanding and ease of use by readers.

Table 7.3. The summary of the number of points and AARD% for different classes of ILs

NO.	Class	Number of data points	AARD%	
			NGC model-1 (AV_{EP})	NGC model-2 (S_{EP})
1	imidazolium	56	16.100	9.483
2	pyridinium	35	8.439	8.406
3	piperidinium	14	7.769	4.895
4	pyrrolidinium	13	11.479	5.779
5	ammonium	10	48.101	32.389
6	morpholinium	8	9.654	6.985
7	quinolinium	4	59.417	56.822

Table 7.4. The comparison of the statistical indexes for different models

Datasets		Train set	Test set	Total set
Number of Data points		113	27	140
NGC-1 model (this work)	AARD%	17.701	9.286	16.078
	R^2	0.880	0.808	0.866
	MSE	0.117	0.165	0.126
	RMSE	0.342	0.406	0.355
NGC-2 model (this work)	AARD%	12.538	5.897	11.257
	R^2	0.928	0.920	0.927
	MSE	0.070	0.060	0.068
	RMSE	0.265	0.246	0.261
General GC	AARD%	12.368	7.129	11.358

Datasets		Train set	Test set	Total set
Number of Data points		113	27	140
model (Kang et al., 2021)	R^2	0.929	0.905	0.924
	MSE	0.069	0.082	0.071
	RMSE	0.262	0.286	0.267

7.5 Conclusions

This work creatively proposed the NGC approach, a novel type of GC method for predicting the toxicity of ILs towards IPC-81. In the NGC method, quantum chemistry calculations are applied to obtain the group descriptors for developing NGC models, i.e., the frequency of a group occurring in a molecule, and to improve the accuracy and reliability of predictive models of the typical GC method. The AV_{EP} and S_{EP} of groups were employed to calculate the number of groups, which reflects richer molecule information than those utilized in the traditional GC method. Two new models, NGC-1 and NGC-2 were constructed with satisfactory statistical indexes. In comparing the predictive performance of the NGC and typical GC models, the model NGC-2 with the R^2 value of 0.927, AARD value of 11.257%, and the RMSE value of 0.261 for the total dataset exhibits better predictability and accuracy than the traditional GC method. It can be inferred that the proposed NGC technique using the non-integer number of groups as descriptors is suitable to estimate the toxicity of ILs. More atomic parameters from quantum chemistry calculations for the descriptors of NGC models will be attempted and implemented in our future work. The NGC method will also be explored to evaluate other properties of ILs and other compounds.

7.6 Supplementary data

The following link is the Supplementary data to this chapter:

<https://ars.els-cdn.com/content/image/1-s2.0-S0009250922005255-mmc1.xlsx>

Chapter VIII

Discussion and Summary

Contents

8.1 The workflow of the thesis.....	145
8.2 QSPR model development for gas absorption of ionic liquids	148
8.3 QSPR model development for toxicity assessment of ionic liquids ...	152
8.4 Conclusions.....	158
8.5 Future work.....	159

8.1 The workflow of the thesis

The application of ionic liquids (ILs) for gases are reviewed in this thesis. Mainly due to the feature of negligible vapor pressure of ILs, ILs are extensively defined as a novel class of green solvent and have the potential to take place of conventional organic solvents in lots of industrial areas. In the fields of environmental sciences, ILs have been widely used for acid gases and VOCs absorption and the solubility data and/or HLC values of gases in ILs have been measured and reported. Therefore, to make contribution to further application of ILs for contaminated gases absorption, predictive models for solubility (i.e. ammonia (NH₃)) and HLC (i.e. H₂S) of gases in various ILs are successfully established based QSPR approaches.

The (eco)toxicity of ILs and their environmental fate were introduced in this thesis. Because of their nonnegligible solubility in liquid solutions, their hazard evaluation and risk to the environment and human health are urgent to explore. A great deal of laboratory work for measuring the (eco)toxicity of ILs have been done, the computer-aided toxicity assessment of ILs have also been implemented by researchers. However, duo to the countless for ILs. The current achievements on this topic are still limited. In this thesis, the further study for toxicity evaluation of ILs based on the QSPR methodologies, and novel molecular descriptors are proposed to build predictive models. The studies in the followed three Chapters focused on the toxicity assessment of ILs.

This dissertation developed two new QSPR models for predicting the gas absorption capability and one novel QSPR model for toxicity of ILs to *Vibrio fischeri* (*V. fischeri*), using extreme learning machine (ELM) algorithm coupled with novel different quantum chemical descriptors. Besides, two new methods based on the quantum chemical descriptors of atoms in ILs were proposed to estimate the *IPC-81 cells*. All the models with satisfactory statistical parameters show good reliability and predictivity, and thus could provide valuable guidance for the application of ILs.

The workflow of this thesis can be observed in Figure 8.1 and Figure 8.2. The thesis mainly contains two parts. One is the estimated models' development for gases absorption in ionic liquids, and the other is the toxicity assessment of ionic liquids by new QSPR models.

- Chapter III studied the relationships between the electrostatic potential surface area (S_{EP}) values of each entirely optimized IL and ammonia

solubility in ILs. A new predictive model for the ammonia solubility in ILs were constructed by using the selected molecular descriptors and ELM algorithm.

- Chapter IV investigated the relationships between the S_{EP} descriptors in four ranges of the entire ILs and Henry's law constant of H_2S of ILs. A highly efficient estimation model was successfully developed for the Henry's law constant of H_2S of ILs.
- The following three Chapters focused on the environmental risk evaluation of ILs. Chapter V explored the relationships between the S_{EP} descriptors in sixteen ranges of the cations and anions of ILs and the toxicity of ILs to *V. fischeri*. The five most important descriptors were employed as inputs for the ELM method and then the QSPR model for forecasting the toxicity of ILs to *V. fischeri* were established.
- Chapter VI first proposed a new method which applied the atom surface fragment properties of ILs to calculate groups' contribution and then developed predictive models for the toxicity evaluation of ILs to IPC-81.
- Chapter VII initially introduced the non-integer group contribution method for the property prediction of chemicals. In this study the method was first used for the toxicity estimation of ILs to IPC-81.

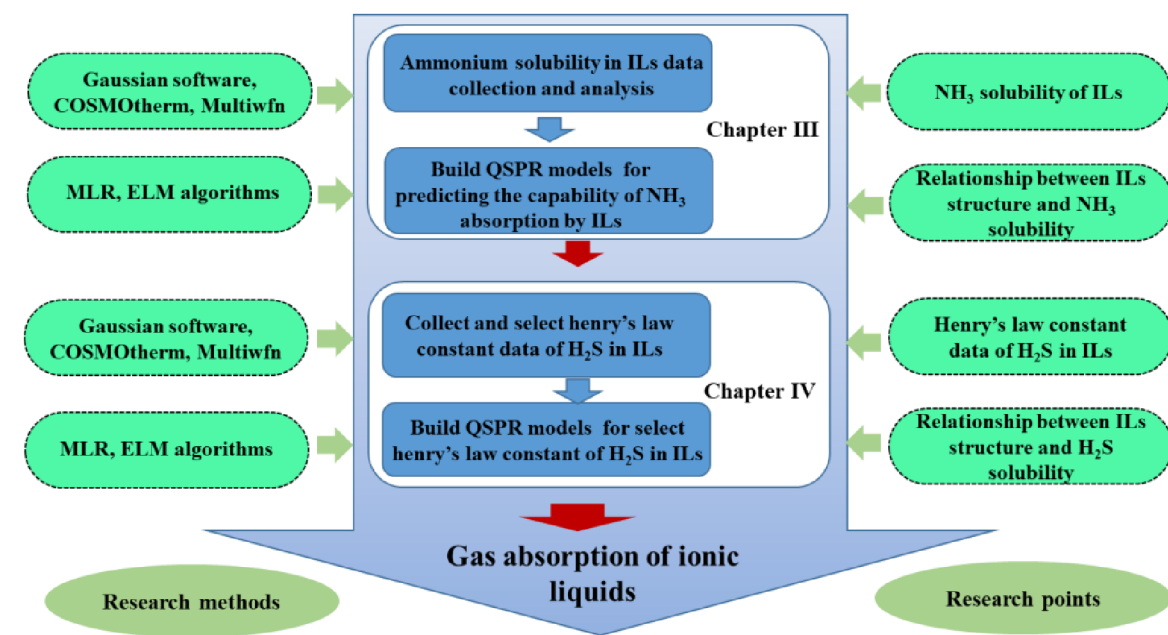


Figure 8.1 Workflow of Chapter III and Chapter IV

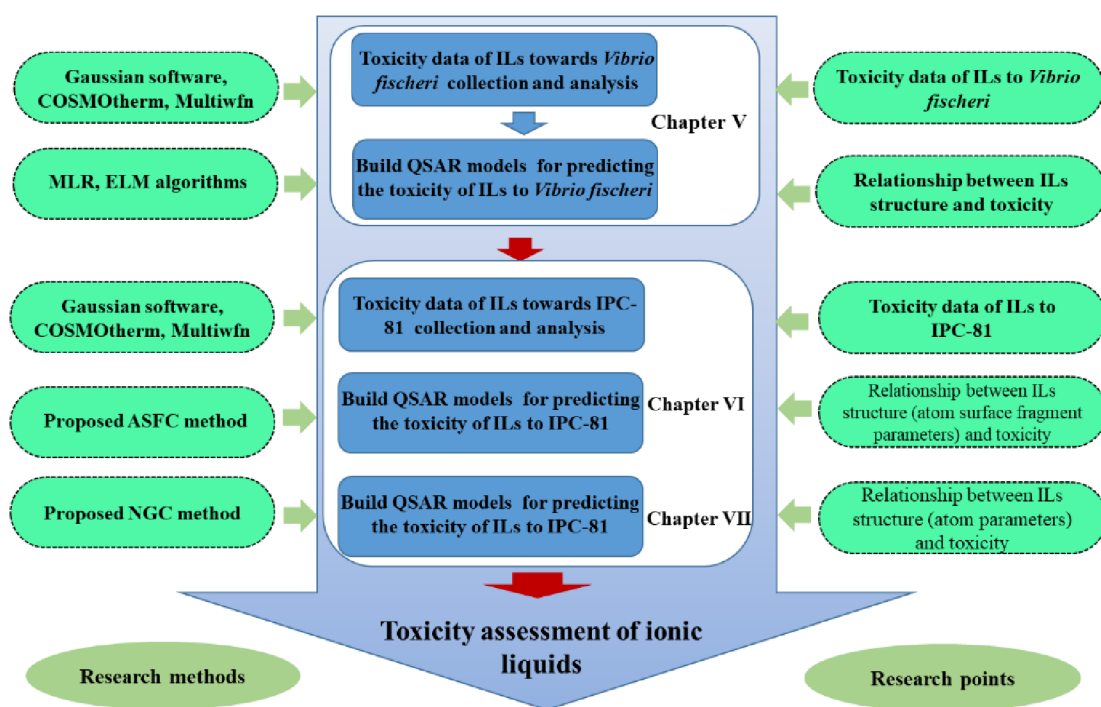


Figure 8.2 Workflow of Chapters V, VI and VII

8.2 QSPR model development for gas absorption of ionic liquids

QSPR model development and Molecular descriptors

Ionic liquids (ILs) are composed of organic cations and organic/inorganic anions, which have extraordinary properties including high chemical stability, low vapor pressure, high ionic conductivity, wide liquid temperature range, and also fine-tunable nature. However, since the characteristics investigation for tremendous number of ILs by experimental approaches are time-consuming and inevitably accompanied by hazardous experimental operations, a number of predictive models have been explored and reported for calculating the ILs' properties applying different computational algorithms (Belvèze, 2004; Jastorff et al., 2007; Katritzky et al., 2002; Palomar et al., 2007; Torrecilla et al., 2008a; Yan et al., 2008). So far, quantitative structure–property relationship (QSPR) models have been well established and various descriptors are used to correlate the physicochemical properties of a compound with its molecular structure and to estimate the characteristics of compounds.

A QSPR study commonly contains three main steps. The first step is collection of experimental data and data preparation; the next contains data processing for descriptor computation and model building; the last step involves the validation of the obtained model and the mechanism analysis of the results. Each step consists of multiple operations to develop efficient and reliable models (Roy et al., 2015). The typical process of QSPR modeling is shown in [Figure 8.3](#).

Typically, the molecular descriptors of ions can be classified into four types which based on the various dimensionality representing the chemical structures of compounds/ILs: 1) 0D, number of atoms/bonds/heavy atoms/specific atoms, molecular weight, atomic properties; 2) 1D, number of functional groups/fragments; 3) 2D, topological descriptors calculated from molecular graphs, dipolarity/polarizability; 4) 3D, geometrical molecular descriptors gained from optimal 3D coordinates of atoms; for example, molecular volume/surface, energy parameters, electrostatic potentials, electron density descriptors; 5) 4D, 3D parameters plus extra properties calculated through quantum chemical calculations, for example, the atomic charges (Nantasenamat et al., 2009). The stepwise approaches in multiple linear regression (MLR) are commonly used to further reduce descriptors which are redundant or show collinearity (Eichenlaub et al., 2022).

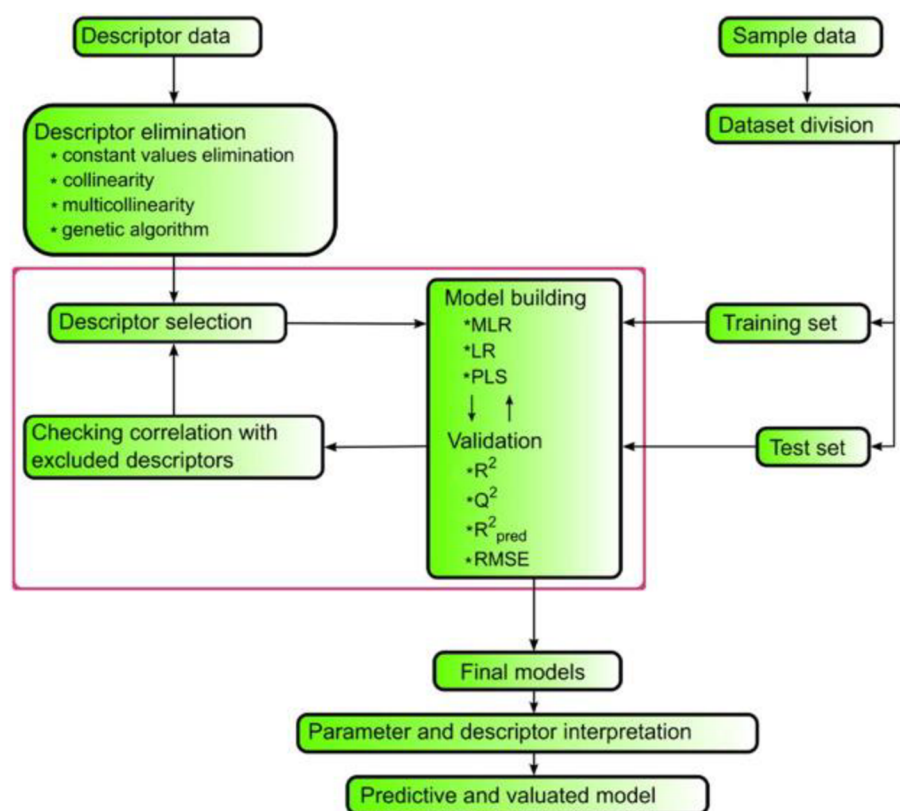


Figure 8.3. Schematic overview of the QSPR process. (Eichenlaub et al., 2022)

In general, the innovation of the model development commonly comes from three aspects: (1) Innovations in input parameters/descriptors; (2) Innovations in the algorithms used to build the model; (3) Innovations of new theoretical methods. In this thesis, the novel classes of input descriptors were proposed in chapters III, IV and V, while the new theoretical methods were first introduced in chapters VI and VII. The detailed innovations and research work of this thesis are summarized as follows.

QSPR model development for gas absorption in ILs

In recent years, ILs have been classified as environmentally-friendly chemicals in compliance with environmental regulations. Hence, interest in research and applications of ILs has been thriving and studies on multidisciplinary fields (e.g., environmental sciences, chemical engineering and material sciences) have been explored. Due to the special features of them, ILs have been commonly regarded as promising solvents for gas absorption (Gomes and Husson, 2009). Some studies for measuring the gases solubility in various ILs have been performed and reported, for

example, H₂S (Jalili et al., 2009), CO₂ (Zeng et al., 2017), NH₃, CO, SO₂ (Lei et al., 2014a) and so on.

Due to the fine-tuning nature of ILs, a great number of ILs are probably synthesized in laboratory. To reduce the high operational costs and the difficulties of carrying out experiments, some neural network approaches are used to predict the gas solubility in ILs. Baghban et al. (2015) predicted the carbon dioxide absorption in ILs by two computational methods which are multi-layer perceptron artificial neural network (MLP-ANN) and adaptive neuro-fuzzy interference system (ANFIS). Later, three methods, namely MLP-ANN, ANFIS and RBF-ANN (basis Function Artificial Neural Network) were applied for H₂S solubility prediction in ILs by Amedi et al. (2016). In 2018, Baghban et al. (2018) investigated the predictivity of Support Vector Machine (SVM), a Least Square Support Vector Machine (LS- SVM) approaches for the NH₃ solubility estimation in ILs.

Extreme learning machine (ELM) is an efficient and universal learning algorithm. It has been discovered that ELM has faster training speed and stronger generalization than those traditional neural networks (NN) algorithms. Currently, ELM has been widely utilized in the areas for property prediction of chemicals (Li et al., 2019; Luo et al., 2019). Therefore, ELM is considered to develop the QSPR models for property estimation in my studies.

In Chapter III, a novel QSPR model to evaluate the NH₃ solubility in ILs were proposed.

In the process of the model establishment, the 240 electrostatic potential surface area (S_{EP}) values of each optimized IL within the range of ± 60 kcal/mol were firstly obtained. Then three S_{EP} molecular descriptors, of the entire ILs used were selected as inputs for modeling by the stepwise procedure of MLR. The estimated model was developed by applying the extreme learning machine (ELM) algorithms, in which three S_{EP} descriptors of the entire ILs, $S_{EP24.75}$, $S_{EP7.25}$, and $S_{EP2.75}$, along with the temperature (T) and pressure (P) of the IL-gas systems were employed. All the selected S_{EP} parameters are positive potentials which mainly depend on cations' surface, suggesting that cations are more significant than anions the ammonia solubility in ILs, which agrees with the conclusions reported by Yokozeki et al (Shi and Maginn, 2009; Yokozeki and Shiflett, 2007b, 2007a). The negative coefficients of T , $S_{EP24.75}$, and $S_{EP7.25}$ show that they have the negative influences on the ammonia solubility in ILs. To the contrast, the P and $S_{EP2.75}$ parameters have positive effects on the ammonia solubility in ILs. Moreover, it was found that the $S_{EP24.75}$ values of ILs without hydroxyl group are

larger than those with hydroxyl group. For example, the $S_{EP24.75}$ values of [EMIM][BF₄] is 3.1035 and that of [EtOHMIM][BF₄] is 1.791. Consequently, the ammonia solubility of [EMIM][BF₄] is lower than that of [EtOHMIM][BF₄]. The reason is probably that the ammonia may more readily generate hydrogen bonds with the hydroxyl group in ILs, which result in the increasing ammonia solubility, which is consistent with the findings by Zhao et al (2017) and Li et al (2015).

To develop reliable models using ELM algorithm, the number of neurons in its hidden layer was investigated. Compared to the models with different number of neurons, the model with 185 neurons showed the best performance and thus the optimal model was confirmed. The obtained statistical parameters of the developed QSPR model are R^2 (0.9937) and AARD (2.95%). Besides, almost 98% of all the predicted data have quite low absolute deviations (less than 5%). Showing the good reliability and predictivity of the predictive model for NH₃ solubility in ILs. To our best knowledge, this work is the first QSPR model for predicting the NH₃ solubility in ILs, which will help to reduce the laboratory work for the measurements of NH₃ absorption ability of ILs. This work will also be significant selecting and designing efficient ILs for NH₃ absorption.

Chapter IV developed a novel model for estimating the Henry's law constant (HLC) of hydrogen sulfide (H₂S) in ILs.

The model development was also based on the ELM method in Chapter. The molecular descriptors were firstly calculated and screened. In detail, the S_{EP} values between -60 kcal/mol and 60 kcal/mol of each entire IL were calculated and divided into four ranges. Four descriptors of S_{EP} , namely S_{EP1} , S_{EP2} , S_{EP3} , S_{EP4} , were all employed for modeling, indicating that the influences of all S_{EP} values of ILs were considered. Results demonstrated that the four S_{EP} descriptors have positive influences with the HLC values of H₂S in ILs, whereas the S_{EP1} was proved to be the most important descriptors for the HLC of H₂S in ILs. Particularly, the ILs have the same cation [BMIM]⁺ and different anions of [BF₄]⁻, [PF₆]⁻, [Tf₂N]⁻, [TfO], the S_{EP1} values of them have the order of $S_{EP1-[BMIM][PF6]} > S_{EP1-[BMIM][BF4]} > S_{EP1-[BMIM][TfO]} > S_{EP1-[BMIM][Tf2N]}$, while their HLC values of H₂S follow the sequence of $HLC_{[BMIM][PF6]} > HLC_{[BMIM][BF4]} > HLC_{[BMIM][TfO]} > HLC_{[BMIM][Tf2N]}$. However, this rule is not always correct. For instance, ILs with the same cation of [HEMIM]⁺ and four anions have the same order of S_{EP1} values, HLC of H₂S exhibited a different sequence, $HLC_{[HEMIM][BF4]} > HLC_{[HEMIM][PF6]} > HLC_{[HEMIM][TfO]} > HLC_{[HEMIM][Tf2N]}$, which was previously concluded by Zhao et al.(2016a) as well. Additionally, the HLC values of H₂S in ILs increase with larger T and smaller V_m , which proves that the H₂S solubility in ILs when the T increases and

V_m reduces. This agrees with the typical rules of the H₂S solubility in solvents (Jalili et al., 2009; Sakhaeinia et al., 2010a).

The R^2 and AARD of the developed model are 0.9994 and 1.582%, respectively. Around 95% of relative deviations (RD) values of the predicted data for the whole dataset were in the range of $\pm 5\%$, showing the excellent performance of the proposed model. Moreover, the application domain (AD) of the developed model was analyzed by Williams plot, i.e., the leverage (h) vs standardized residuals (σ) (Gharagheizi et al., 2012b). Results show that the majority of σ values of ILs in the entire dataset are lower than 3 and all the h values of ILs are less than the critical value h^* (0.214), revealing that most predicted data of the ILs are in the AD of the model. However, there are four ILs from the dataset outside of the AD, which are unreliable predicted values. This may be probably attributed to the experimental errors when measured (Gharagheizi et al., 2012b). Overall, most of ILs located in the AD of the ELM model and the satisfactory statistical parameters (e.g. high R^2 and low AARD) turns out the great predictive ability and accuracy of the model. This work is potentially helpful to screen and design optimal ILs for H₂S absorption.

8.3 QSPR model development for toxicity assessment of ionic liquids

The state-of-the-art research shows some living organisms can be impacted by the ILs., such as *Vibrio fischeri*, *Caenorhabditis elegans*, *Green algae*, *leukemia rat cell line* (Jafari et al., 2019; Latała et al., 2009; Peng et al., 2018; Pretti et al., 2009; Ranke et al., 2007b; Stolte et al., 2007b; Ventura et al., 2012). Therefore, the toxicity of ILs towards living creatures and environments cannot be underestimated, as they can be dissolved in water and released into ecosystem (Ghanem et al., 2017; McFarlane et al., 2005; Yan et al., 2019).

Some researchers have aimed to find general correlations between the structures and the toxicity of ILs. According to the current research results, a part of ILs have relatively low toxicity while some ILs exhibit apparent inhibitory effects on a variety of biological systems. Till now, no uniform rules have been discovered, but it can be mainly concluded about the influence factors of the IL structures on the toxicity as follows (Egorova et al., 2017; Gouveia et al., 2014; Ranke et al., 2004):

a) The length of the alkyl chain of the cationic side chain: most related studies show that the length of the alkyl side chain and the toxicity of ILs have a positive relationship. Besides, the lipophilicity of the IL gradually become strong with the increase of its alkyl chain. Meanwhile, the interaction with the phospholipids on the cell membrane and the permeability increase, and thus the damage to the cell membrane will be more severe, resulting in an increase of the toxicity.

b) Functional groups on cations: functional groups on cations of ionic liquids have no general rules on toxicity, and their functional groups have different effects for different cations.

c) The nature and structure of the anions and cations: for ILs with several typical types of cations that have been frequently studied, it is generally regarded that their toxicity has followed order: ammonium cations < pyridinium cations < imidazolium cations, while for anions of ILs, the IL toxicity is generally derived from the new ions generated by the hydrolysis of anions.

d) The interactions between anions and cations of ILs: the interactions between ILs and the third substance are affected by the hydrogen bonds between anions and cations, and the strength of ionic bonds affect. Therefore, their toxicity is also varied with different physiological environments.

Table 8.1 listed the structural effects on the IL toxicity. Overall, the anions and cations of ionic liquids have different effects on distinct biological individuals. The current studies still exist certain limitations, and the mechanism of action, especially the molecular mechanism, is far from enough. In order to effectively and purposefully predict the impacts of ILs on the entire ecological environment and human health, it is necessary to gradually improve the research data of ILs on diverse cells and organisms, and then provide guidance for designing and synthesizing non-toxic, green, and easy-to-degrade ILs.

The relationships between the structure and properties of ILs are also investigated in our works. However, different with previous studies, the relationships between the toxicity of ILs and the electrostatic potential as well as the charge distribution of the corresponding molecular surfaces were explored and used for predictive model establishment.

Table 8.1 Structural effects of ILs on toxicity (Cho et al., 2021)

Structure	Low toxicity	High toxicity
Cations	Cholinium, Guanidium, Sulfonium, Morpholinium Dicationic moiety Protic moiety Short alkyl chain length Less number of substitutes Hydrophobic functionalization (e.g., ester, hydroxyl)	Imidazolium, Phosphonium, Ammonium, Pyridinium, Quinolinium Hydrophobic functionalization (e.g., aryl) Long alkyl chain More number of substitutes (e.g., peralkylated) Bioactive molecules (e.g., herbicide-, menthol-, ampicillin-, theophylline-based)
Anions	Halides Hydrophilic anions Small molecular volume Anion with high Bronsted acidity Amino acid-based anions	Hydrophobic structures Metal containing molecules (e.g., GdCl ₄ , MnCl ₄ , CoCl ₄) Unstable in aquatic environment (e.g., BF ₄ , PF ₆ , SbF ₆) Bioactive molecules

The toxicity of ionic liquids towards *Vibrio fischeri* (*V. fischeri*)

V. fischeri can be effectively used in toxicological risk assessments because it is a species of the marine luminescent bacterium with a short reproductive cycle, and the inference of toxicity against *V. fischeri* can be extrapolated to various aquatic organisms (Jafari et al., 2019; Kaiser, 1998). Therefore, the toxicological risk assessment of ILs can be effectively performed through *V. fischeri*. Nevertheless, because of the extensive number of ILs, some estimation models have been published for the ecotoxicity (EC₅₀ *Vibrio fischeri*) of ILs to fill in the data gap and reduce the laboratory work. Furthermore, to screen and design green and functional ILs, the toxicity assessment of ILs is required.

Yan et al. (2015) proposed a QSPR model based on the multiple linear regression (MLR) method for ILs toxicity evaluation towards *V. fischeri*. During their study, some topological index descriptors about the character vector of atom as well as the distance matrix for atom position were employed for the models of evaluating the ILs toxicity against *V. fischeri*. Ghanem et al. (2017) built QSAR models for toxicity evaluation of ILs towards *Vibrio fischeri* using the linear (MLR) and non-linear (MLP) algorithms, in which five sigma profile descriptors were used for the model development, while the elemental compositions from anions and cations of ILs were considered as variables for modelling by Jafari et al. (2019) utilized. Overall, the confirmation of descriptors for QSAR models is quite significant in predicting the toxicity of ILs (Cao et al., 2018; Zhao et al., 2015b, 2014). In my studies, the electrostatic potential surface areas (S_{EP})

of molecules was considered as descriptors for QSPR models as it can display abundant molecular information at electronic level (Kang et al., 2018d; Murray and Politzer, 2011b).

In Chapter V of this thesis, a toxicity dataset of 142 ILs towards *V. fischeri* were gathered and a QSPR model for predicting this kind of toxicity of ILs were built. In this study, the cations and anions of ILs were separately optimized and used for calculating molecular descriptors. All the obtained S_{EP} values of cations and anions were then split to 16 descriptors. Finally, five most important S_{EP} descriptors were selected as input variables for the model development, which are S_{EP1} , S_{EP2} , S_{EP6} , S_{EP7} , S_{EP13} , respectively. Results shows that four significant descriptors (i.e., S_{EP1} , S_{EP2} , S_{EP6} , S_{EP7}) come from cations while only one descriptor (S_{EP13}) belongs to anions. Meanwhile, the most significant descriptor is the S_{EP2} and S_{EP13} are the fourth key descriptor, implying that cations have more influences than anions on the ILs toxicity (Luis et al., 2010). Additionally, the descriptors of S_{EP2} , S_{EP7} and S_{EP13} negatively impacted on the toxicity values of ILs. Results show that longer alkyl chain of ILs which always has higher S_{EP2} led to the stronger toxicity of ILs, which is consistent with the conclusion in previous literature (Chen et al., 2014; Cho et al., 2007; Ghanem et al., 2017). As for the anionic descriptor, higher S_{EP13} of ILs led to lower $\log(EC_{50})_{V. fischeri}$ value of ILs and thus the corresponding ILs have stronger toxicity. For instance, the anion of $[NTf_2]^-$ had the bigger S_{EP13} value than the others and ILs with $[NTf_2]^-$ anion are the most toxic when ILs have same cations. This may be attributed to the hydrophobic nature of $[NTf_2]^-$ (Santos et al., 2014). However, the anion $[BF_4]^-$ are regarded as hydrophilic compounds, but its corresponding S_{EP} values located in the descriptors of S_{EP10} and S_{EP11} which are insignificant factors for ILs toxicity. Therefore, ILs with hydrophobic anions are more toxic than those with hydrophilic anions, which also previously pointed out by Ghanem et al. (2017) and Santos et al. (2014). It needs to be noted that the current study firstly employed S_{EP} descriptors along with ELM technique for forecasting the toxicity of ILs on *V. fischeri*. Meanwhile, the model was developed based on 142 data points of ILs and only 5 input descriptors. Compared to the models in previous publications, most of them utilized around or below 100 data and the number of molecular descriptors was from 5 to 30 (Table 5.3).

Based on the ELM algorithm, the predictive model for the total set with the R^2 of 0.9272 and AARD of 0.3262% were constructed, and the majority of absolute errors of the evaluated toxicity data are less than 1.0%. Besides, The AD analysis was also carried out to assess the robustness and reliability of the proposed model. Except for 6 outliers, the rest 136 ILs have h values lower than h^* as well as the standardized residuals $\sigma < 3$. Four outliers with $h > h^*$ are structurally influential chemicals (Ma et

al., 2015) while tow outliers with $\sigma > 3$ are due to the incorrect experimental data (Gramatica, 2007). These outcomes demonstrate that the proposed approach is satisfactory for the prediction of IL toxicity towards *V. fischeri*.

The toxicity of ionic liquids against IPC-81

A variety of group contribution (GC) and machine learning methods have been applied to develop efficient linear and nonlinear models for the property evaluation of ILs to the problems from experiments, including resource consumption, time consumption and various risks in laboratory tests (Abramenko et al., 2020a; Gao et al., 2020; Song et al., 2020a; Torrecilla et al., 2008b; Zhao et al., 2015a).

GC method has been successfully and widely applied for characteristics evaluation of chemicals or mixtures. It is a type of semiempirical theory and also have been extensively utilized to assess various properties of ILs, such as electrical conductivity, density, heat capacity, melting point, surface tension, viscosity and so on (Gardas and Coutinho, 2008; Gardas and P.Coutinho, 2009).

The common groups of ILs can be typically categorized as three species, cation cores anions and some substituents. For example, imidazolium [Im]⁺, piperidinium [Pip]⁺ and pyridinium [Py]⁺ are representative cation cores; tetrafluoroborate [BF₄]⁻, hexafluorophosphate [PF₆]⁻ and chloride [Cl]⁻ are common anion cores; methylene (-CH₂-) and methyl (-CH₃) are typical substituents (Chen et al., 2019a; Y. Huang et al., 2013). A GC model was established for the surface tension estimation of ILs by Gharagheizi et.al (2012a) using the input parameters of nineteen sub-structures and the temperature of the systems. In 2019, Chen et al. (2019b) proposed several GC models to predict the melting point, density, viscosity, heat capacity, density of ILs, utilizing the GC parameters obtained based on the number of groups in the structure of each IL. Another GC model was reported by Ismail Hossain et.al. (2011) to predict the ILs toxicity against *Daphnia magna*. In their work, the number of groups (cation cores, anions cores and substitutions) by counting in ILs were used as descriptors to calculate the contribution coefficients and built predictive models using the MLR method.

The traditional GC model obviously has some shortcomings. The theoretical basis of the GC method assumes that the property of a molecule can be described by adding all the contributions of defined groups in the substance, and each group independently contributes and affects the properties of the molecule (Lazzús, 2012; Mu et al. People, 2007).

Since traditional GC only considers the type and frequency of groups, and does not describe the interactions of groups or adjacent groups (Mu et al., 2007), the inability to distinguish cationic/anionic isomers, including isomer fragments/groups. In fact, since each group in different molecules should have a different environment and various interactions with its surrounding atoms, the contribution of each group should be different (or slightly different) in different molecules. Therefore, it is necessary to improve the traditional GC method for better accuracy and wider application.

However, only the species and number of the groups by counting are generally considered while the interactions of groups or adjacent groups are ignored in the conventional GC methods (Mu et al., 2007), resulting in the inability of distinguish the cationic and anionic isomers as well as the isomeric fragments/groups in molecules. Because each group in different molecules has various surroundings and interactions with its neighboring atoms/fragments, the contribution of each group should strictly be calculated to be distinct (or slightly different) in various molecules. Therefore, improving the traditional GC method to be more accurate and extensively applicable is realistic and significant.

In Chapter IV, based on the theory of group contribution (GC) method, a novel comprehensive approach, named atom surface fragment contribution (ASFC), was first created and used to predict the toxicity of ILs towards IPC-81. The inspiration of the ASFC method was from the basic theory of the conventional group contribution method as the innovation of this method lies on the calculation scheme of the group parameters. In this new method, the sigma surface area ($S_{\sigma\text{-surface}}$) of atoms in cations and anions of ILs were obtained from COSMO sigma files and utilized to obtain group/fragment descriptors to build predictive models. The detailed calculation approach has been described in the section 6.3.3. For comparison, the traditional GC method were also harnessed to develop a predictive mode for comparison. Compared with conventional GC model, ASFC model shows better performance and thus has high potential to be used for predictive model development for the toxicity of ILs.

Furthermore, a more general method, non-integer group contribution method (NGC) is initially proposed to estimate the toxicity of ILs towards IPC-81 cells in Chapter VII. In this method, a totally different and advanced method for getting the number of groups in compounds were described, which was motivated by the methods of ASFC proposed in my previous work. In the current work, the average values of electrostatic potential (AV_{EP}) and the electrostatic potential surface area (S_{EP}) of each atom in cations and anions of ILs were obtained by quantum calculations and exemplified for calculating the number of groups in ILs, and separately applied for the

NGC model development. Two NGC models with satisfactory statistical parameters show their reliability and accuracy for evaluating the toxicity of ILs to IPC-81. The R^2 and RMES of NGC model using the average S_{EP} values of groups in ILs is 0.927 and 0.261 respectively, which proves that its performance is better than the model developing by traditional GC method.

8.4 Conclusions

Two estimation models for NH_3 solubility and HLC of H_2S in ILs were developed respectively by combination of the electrostatic potential surface area (S_{EP}) descriptors and the advanced ELM method. The S_{EP} descriptors used in both models were calculated based on each entire optimized IL. However, the difference between them was that three crucial S_{EP} descriptors as inputs used in the model for NH_3 solubility were screened from 240 S_{EP} values of each IL, while four input descriptors S_{EP} values were obtained by dividing the whole S_{EP} values for each IL into four ranges. The built models have potential in selecting and designing efficient ILs for NH_3 absorption and gas sweetening. What's more, the proposed approaches can be applied for developing estimation models for not only the gas absorption ability of ILs but also the physico-chemical properties (e.g., viscosity, melting points, heat capacity) of ILs, and even other classes of compounds (e.g., organic compounds).

A new QSPR model for predicting the toxicity of ILs towards *V. fischeri* was built. Different from the calculation methods of the above S_{EP} descriptors, the selected S_{EP} descriptors were based on the separately optimized cations and anions of ILs. With the use of ELM algorithm, the proposed model with satisfactory statistical parameters were acquired. The developed models may be helpful to provide guidance to select and design low toxic ILs for the future applications of ILs. Moreover, the proposed method has potential applications in evaluating the toxicity of ILs to various living organisms and plants, and also other properties of compounds.

Two novel methods in accordance with the basic theory of group contribution (GC) method were initially proposed, atom surface fragment contribution (ASFC) method and non-integer group contribution (NGC) method. The sigma surface area ($S_{\sigma\text{-surface}}$) of atoms in ILs were utilized to calculate group/fragment parameters as input variables for ASFC models, whereas the average values of electrostatic potential (AV_{EP}) and the electrostatic potential surface area (S_{EP}) descriptors of atoms in ILs were separately employed to calculate the non-integer number of groups in ILs for modeling. The high R^2 and low RMES of the ASFC model indicating the reliability and accuracy of the

developed model. For NGC models, the NGC-2 model using the average S_{EP} values of atoms in ILs exhibits greater performance than NGC-1 model and also the traditional GC model. Therefore, both of the proposed approaches have high potential for predictive model development for not only the toxicity of ILs but also the other properties of chemicals.

It is worth mentioning that the NGC method can be considered a more comprehensive method to be used than the ASFC method since the theory of NGC method which is an advanced variant of the conventional GC method is easier to be understood. Additionally, except for the atom surface fragment parameters, any other types of atomic descriptors (e.g., atomic volumes, atomic charges) based on the quantum chemistry calculations can be applied for NGC model establishment. Therefore, NGC method may have the potential to be alternatives of the traditional GC method in the field of property prediction for compounds.

8.5 Future work

The topic of this this thesis can be further extended from different directions to make contribution to the application of ILs in environmental sciences and assessment of the (eco)toxicity and the environment of ILs in the future. The details are as follows.

- Firstly, the developed models can be further applied to screen and design optimal and specific-tasked ILs for practical applications. And the proposed method can be further utilized by other gases (CO_2 , SO_2).
- The ASFC method and NGC methods can be further applied to develop more predictive models for other properties of compounds. Meanwhile, the ASFC and NGC methods can be further updated based on different quantum chemical descriptors of atoms (e.g., atomic volumes and atomic charges) in molecules.
- Moreover, as the rapid development of computer technologies, more advanced method, e.g., deep learning, can be used for the development of QSPR models to acquire more efficient and accurate predictive models.
- Furthermore, ILs can be applied not only for gas absorption but also for heavy metals and organic pollutants removal. The related experiments about the removal efficacy of heavy metals and organic pollutants are

deserved to investigate. Some of our preliminary explorations show that ionic liquids have good effects on the extraction of heavy metal Cd. Therefore, further in-depth research is meaningful. Suitable ILs for different pollutants removal can be selected by computer-aided methods, such as QSPR methods, COSMO-RS/SAC methods to further experimentally verify the effects, while the quantum chemical method can be used to study their mechanism in depth.

References

- Abdolkarimi, E.S., Abaei, G., Mosavi, M.R., 2018. A wavelet-extreme learning machine for low-cost INS/GPS navigation system in high-speed applications. *GPS Solut.* 22, 1–13. <https://doi.org/10.1007/s10291-017-0682-x>
- Abramenko, N., Kustov, L., Metelytsia, L., Kovalishyn, V., Tetko, I., Peijnenburg, W., 2020a. A review of recent advances towards the development of QSAR models for toxicity assessment of ionic liquids. *J. Hazard. Mater.* 384. <https://doi.org/10.1016/j.jhazmat.2019.121429>
- Abramenko, N., Kustov, L., Metelytsia, L., Kovalishyn, V., Tetko, I., Peijnenburg, W., 2020b. A review of recent advances towards the development of QSAR models for toxicity assessment of ionic liquids. *J. Hazard. Mater.* 384. <https://doi.org/10.1016/j.jhazmat.2019.121429>
- Ahmadi, M.A., Haghbakhsh, R., Soleimani, R., Bajestani, M.B., 2014. Estimation of H₂S solubility in ionic liquids using a rigorous method. *J. Supercrit. Fluids* 92, 60–69. <https://doi.org/10.1016/j.supflu.2014.05.003>
- Al-fnaish, H., Lue, L., 2017. Modelling the solubility of H₂S and CO₂ in ionic liquids using PC-SAFT equation of state. *Fluid Phase Equilib.* 450, 30–41. <https://doi.org/10.1016/j.fluid.2017.07.008>
- Amde, M., Liu, J.F., Pang, L., 2015. Environmental Application, Fate, Effects, and Concerns of Ionic Liquids: A Review. *Environ. Sci. Technol.* <https://doi.org/10.1021/acs.est.5b03123>
- Amedi, H.R., Baghban, A., Ahmadi, M.A., 2016. Evolving machine learning models to predict hydrogen sulfide solubility in the presence of various ionic liquids. *J. Mol. Liq.* 216, 411–422. <https://doi.org/10.1016/j.molliq.2016.01.060>
- Amir Hossein Jalili, M.S., Cyrus, G., Mehdizadeh, A., Hosseini-Jenab, M., Taghikhani, V., 2012. solubility of CO₂, H₂S, and their mixture in the ionic liquid 1-Octyl-3-methylimidazolium Bis(trifluoromethyl) sulfonylimide. *J. Phys. Chem. B* 116, 2758–2774. <https://doi.org/10.1021/jp305719w>
- Anderson, J.L., Dixon, J.N.K., Maginn, E.J., Brennecke, J.F., 2006. Measurement of SO₂ solubility in ionic liquids. *J. Phys. Chem. B.* <https://doi.org/10.1021/jp063547u>

- Arellano, M., Oturan, N., Pazos, M., Ángeles Sanromán, M., Oturan, M.A., 2020. Coupling electro-Fenton process to a biological treatment, a new methodology for the removal of ionic liquids? *Sep. Purif. Technol.* 233. <https://doi.org/10.1016/j.seppur.2019.115990>
- Arning, J., Dringen, R., Schmidt, M., Thiessen, A., Stolte, S., Matzke, M., Bottin-Weber, U., Caesar-Geertz, B., Jastorff, B., Ranke, J., 2008a. Structure-activity relationships for the impact of selected isothiazol-3-one biocides on glutathione metabolism and glutathione reductase of the human liver cell line Hep G2. *Toxicology* 246. <https://doi.org/10.1016/j.tox.2008.01.011>
- Arning, J., Stolte, S., Bösch, A., Stock, F., Pitner, W.R., Welz-Biermann, U., Jastorff, B., Ranke, J., 2008b. Qualitative and quantitative structure activity relationships for the inhibitory effects of cationic head groups, functionalised side chains and anions of ionic liquids on acetylcholinesterase. *Green Chem.* 10. <https://doi.org/10.1039/b712109a>
- Arunkumar, R., Abraham, A.N., Shukla, R., Drummond, C.J., Greaves, T.L., 2020. Cytotoxicity of protic ionic liquids towards the HaCat cell line derived from human skin. *J. Mol. Liq.* 314. <https://doi.org/10.1016/j.molliq.2020.113602>
- Bader, R.F.W., Carroll, M.T., Cheeseman, J.R., Chang, C., 1987. Properties of Atoms in Molecules: Atomic Volumes. *J. Am. Chem. Soc.* 109(26), 7968–7979. <https://doi.org/10.1021/ja00260a006>
- Baghban, A., Ahmadi, M.A., Shahraki, B.H., 2015. Prediction carbon dioxide solubility in presence of various ionic liquids using computational intelligence approaches. *J. Supercrit. Fluids.* <https://doi.org/10.1016/j.supflu.2015.01.002>
- Baghban, A., Bahadori, M., Lemraski, A.S., Bahadori, A., 2018. Prediction of solubility of ammonia in liquid electrolytes using Least Square Support Vector Machines. *Ain Shams Eng. J.* 9, 1303–1312. <https://doi.org/10.1016/j.asej.2016.08.006>
- Baghban, A., Sasanipour, J., Habibzadeh, S., Zhang, Z., 2019. Estimating solubility of supercritical H₂S in ionic liquids through a hybrid LSSVM chemical structure model. *Chinese J. Chem. Eng.* 27, 620–627. <https://doi.org/10.1016/j.cjche.2018.08.026>

- Bara, J.E., Camper, D.E., Gin, D.L., Noble, R.D., 2010. Room-Temperature ionic liquids and composite materials: Platform technologies for CO₂ capture. *Acc. Chem. Res.* 43(1), 152–159. <https://doi.org/10.1021/ar9001747>
- Bayoumy, A.M., Ibrahim, M., Omar, A., 2020. Mapping molecular electrostatic potential (MESP) for fulleropyrrolidine and its derivatives. *Opt. Quantum Electron.* 52, 1–13. <https://doi.org/10.1007/s11082-020-02467-6>
- Beaulieu, J.J., Tank, J.L., Kopacz, M., 2008. Sorption of imidazolium-based ionic liquids to aquatic sediments. *Chemosphere* 70. <https://doi.org/10.1016/j.chemosphere.2007.07.046>
- Bedia, J., Palomar, J., Gonzalez-Miquel, M., Rodriguez, F., Rodriguez, J.J., 2012. Screening ionic liquids as suitable ammonia absorbents on the basis of thermodynamic and kinetic analysis. *Sep. Purif. Technol.* 95, 188–195. <https://doi.org/10.1016/j.seppur.2012.05.006>
- Bell, I.H., Mickoleit, E., Hsieh, C.M., Lin, S.T., Vrabec, J., Breitkopf, C., Jäger, A., 2020. A Benchmark Open-Source Implementation of COSMO-SAC. *J. Chem. Theory Comput.* <https://doi.org/10.1021/acs.jctc.9b01016>
- Belvèze, L.S., 2004. Modeling and Measurement of Thermodynamic Properties of Ionic Liquids. Thesis.
- Berrouk, A.S., Ochieng, R., 2014. Improved performance of the natural-gas-sweetening Benfield-HiPure process using process simulation. *Fuel Process. Technol.* 127, 20–25. <https://doi.org/10.1016/j.fuproc.2014.06.012>
- Bhupinder Singh, S., 2010. Ionic Liquid Applications: Pharmaceuticals, Therapeutics, and Biotechnology. *J. Am. Chem. Soc.* 132. <https://doi.org/10.1021/ja1098947>
- Blanchard, L.A., Hancu, D., Beckman, E.J., Brennecke, J.F., 1999. Green processing using ionic liquids and CO₂. *Nature* 399, 28–29. <https://doi.org/10.1038/19887>
- Bocharnikova, E.N., Tchaikovskaya, O.N., Bazyl, O.K., Artyukhov, V.Y., Mayer, G. V., 2020. Theoretical study of bisphenol A photolysis, in: *Advances in Quantum Chemistry*. pp. 191–217. <https://doi.org/10.1016/bs.aiq.2019.12.001>
- Boethling, R.S., Howard, P.H., Meylan, W., Stiteler, W., Beauman, J., Tirado, N., 1994. Group Contribution Method for Predicting Probability and Rate of Aerobic

- Biodegradation. Environ. Sci. Technol. 28, 459–465.
<https://doi.org/10.1021/es00052a018>
- Boškin, A., Tran, C.D., Franko, M., 2009. Oxidation of organophosphorus pesticides with chloroperoxidase enzyme in the presence of an ionic liquid as co-solvent. Environ. Chem. Lett. 7. <https://doi.org/10.1007/s10311-008-0161-2>
- Brennecke, J.F., Maginn, E.J., 2001. Ionic liquids: Innovative fluids for chemical processing. AIChE J. 47, 2384–2389. <https://doi.org/10.1002/aic.690471102>
- Bruzzone, S., Chiappe, C., Focardi, S.E., Pretti, C., Renzi, M., 2011. Theoretical descriptor for the correlation of aquatic toxicity of ionic liquids by quantitative structure-toxicity relationships. Chem. Eng. J. <https://doi.org/10.1016/j.cej.2011.08.073>
- Bubalo, M.C., Radošević, K., Redovniković, I.R., Slivac, I., Srček, V.G., 2017. Toxicity mechanisms of ionic liquids. Arh. Hig. Rada Toksikol. <https://doi.org/10.1515/aiht-2017-68-2979>
- Bulat, F.A., Toro-Labbé, A., Brinck, T., Murray, J.S., Politzer, P., 2010. Quantitative analysis of molecular surfaces: Areas, volumes, electrostatic potentials and average local ionization energies. J. Mol. Model. 16, 1679–1691. <https://doi.org/10.1007/s00894-010-0692-x>
- Cao, L., Zhu, P., Zhao, Y., Zhao, J., 2018. Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids. J. Hazard. Mater. <https://doi.org/10.1016/j.jhazmat.2018.03.025>
- Cao, X., Song, X., Chu, Q., Mu, L., Li, Y., Bian, J., 2019. An efficient method for removing hydrogen sulfide from natural gas using supersonic Laval nozzle. Process Saf. Environ. Prot. 129, 220–229. <https://doi.org/10.1016/j.psep.2019.07.008>
- Castner, E.W., Wishart, J.F., 2010. Spotlight on ionic liquids. J. Chem. Phys. 132. <https://doi.org/10.1063/1.3373178>
- Chen, H., Zou, Y., Zhang, L., Wen, Y., Liu, W., 2014. Enantioselective toxicities of chiral ionic liquids 1-alkyl-3-methylimidazolium lactate to aquatic algae. Aquat. Toxicol. <https://doi.org/10.1016/j.aquatox.2014.05.010>

- Chen, Y., Cai, Y., Thomsen, K., Kontogeorgis, G.M., Woodley, J.M., 2020. A group contribution-based prediction method for the electrical conductivity of ionic liquids. *Fluid Phase Equilib.* 509. <https://doi.org/10.1016/j.fluid.2020.112462>
- Chen, Y., Kontogeorgis, G.M., Woodley, J.M., 2019a. Group Contribution Based Estimation Method for Properties of Ionic Liquids. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/acs.iecr.8b05040>
- Chen, Y., Kontogeorgis, G.M., Woodley, J.M., 2019b. Group Contribution Based Estimation Method for Properties of Ionic Liquids. *Ind. Eng. Chem. Res.* 58, 4277–4292. <https://doi.org/10.1021/acs.iecr.8b05040>
- Chiappe, C., Pieraccini, D., 2005. Ionic liquids: Solvent properties and organic reactivity. *J. Phys. Org. Chem.* <https://doi.org/10.1002/poc.863>
- Cho, C.W., Pham, T.P.T., Jeon, Y.C., Vijayaraghavan, K., Choe, W.S., Yun, Y.S., 2007. Toxicity of imidazolium salt with anion bromide to a phytoplankton *Selenastrum capricornutum*: Effect of alkyl-chain length. *Chemosphere.* <https://doi.org/10.1016/j.chemosphere.2007.06.023>
- Cho, C.W., Pham, T.P.T., Zhao, Y., Stolte, S., Yun, Y.S., 2021. Review of the toxic effects of ionic liquids. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2021.147309>
- Cho, C.W., Ranke, J., Arning, J., Thöming, J., Preiss, U., Jungnickel, C., Diedenhofen, M., Krossing, I., Stolte, S., 2013. In silico modelling for predicting the cationic hydrophobicity and cytotoxicity of ionic liquids towards the Leukemia rat cell line, *Vibrio fischeri* and *Scenedesmus vacuolatus* based on molecular interaction potentials of ions. *SAR QSAR Environ. Res.* <https://doi.org/10.1080/1062936X.2013.821092>
- Cho, C.W., Song, M.H., Pham, T.P.T., Yun, Y.S., 2019. Environmental Concerns Regarding Ionic Liquids in Biotechnological Applications, in: *Advances in Biochemical Engineering/Biotechnology.* https://doi.org/10.1007/10_2018_79
- Cho, C.W., Yun, Y.S., 2019. Application of general toxic effects of ionic liquids to predict toxicities of ionic liquids to *Spodoptera frugiperda*, *Eisenia fetida*, *Caenorhabditis elegans*, and *Danio rerio*. *Environ. Pollut.*

<https://doi.org/10.1016/j.envpol.2019.113185>

- Clark, K.D., Emaus, M.N., Varona, M., Bowers, A.N., Anderson, J.L., 2018. Ionic liquids: solvents and sorbents in sample preparation. *J. Sep. Sci.* 41, 209–235. <https://doi.org/10.1002/jssc.201700864>
- Costa, S.P.F., Azevedo, A.M.O., Pinto, P.C.A.G., Saraiva, M.L.M.F.S., 2017. Environmental Impact of Ionic Liquids: Recent Advances in (Eco)toxicology and (Bio)degradability. *ChemSusChem*. <https://doi.org/10.1002/cssc.201700261>
- Costa, S.P.F., Justina, V.D., Bica, K., Vasiloiu, M., Pinto, P.C.A.G., Saraiva, M.L.M.F.S., 2014a. Automated evaluation of pharmaceutically active ionic liquids' (eco)toxicity through the inhibition of human carboxylesterase and *Vibrio fischeri*. *J. Hazard. Mater.* 265. <https://doi.org/10.1016/j.jhazmat.2013.11.052>
- Costa, S.P.F., Pinto, P.C.A.G., Lapa, R.A.S., Saraiva, M.L.M.F.S., 2014b. Toxicity assessment of ionic liquids with *Vibrio fischeri*: An alternative fully automated methodology. *J. Hazard. Mater.* 284. <https://doi.org/10.1016/j.jhazmat.2014.10.049>
- Couling, D.J., Bernot, R.J., Docherty, K.M., Dixon, J.N.K., Maginn, E.J., 2006. Assessing the factors responsible for ionic liquid toxicity to aquatic organisms via quantitative structure-property relationship modeling. *Green Chem.* <https://doi.org/10.1039/b511333d>
- Cui, D., Huang, G. Bin, Liu, T., 2018. ELM based smile detection using Distance Vector. *Pattern Recognit.* 79, 356–369. <https://doi.org/10.1016/j.patcog.2018.02.019>
- Cui, Y.H., Shi, Q.S., Zhang, D.D., Wang, L.L., Feng, J., Chen, Y.W., Xie, X.B., 2021. Detoxification of ionic liquids using glutathione, cysteine, and NADH: Toxicity evaluation by *Tetrahymena pyriformis*. *Environ. Pollut.* 276. <https://doi.org/10.1016/j.envpol.2021.116725>
- Cull, S.G., Holbrey, J.D., Vargas-Mora, V., Seddon, K.R., Lye, G.J., 2000. Room-temperature ionic liquids as replacements for organic solvents in multiphase bioprocess operations. *Biotechnol. Bioeng.* 69. [https://doi.org/10.1002/\(SICI\)1097-0290\(20000720\)69:2<227::AID-BIT12>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0290(20000720)69:2<227::AID-BIT12>3.0.CO;2-0)

- Cunha, E., Passos, M.L.C., Pinto, P.C.A.G., Saraiva, M.L.M.F.S., 2015. Automated evaluation of the inhibition of glutathione reductase activity: application to the prediction of ionic liquids' toxicity. *RSC Adv.* 5. <https://doi.org/10.1039/c5ra04029a>
- Cvjetko Bubalo, M., Radošević, K., Gaurina Srček, V., Das, R.N., Popelier, P., Roy, K., 2015. Cytotoxicity towards CCO cells of imidazolium ionic liquids with functionalized side chains: Preliminary QSTR modeling using regression and classification based approaches. *Ecotoxicol. Environ. Saf.* 112. <https://doi.org/10.1016/j.ecoenv.2014.10.029>
- Cvjetko, M., Radošević, K., Tomica, A., Slivac, I., Vorkapić-Furač, J., Gaurina Srček, V., 2012. Cytotoxic effects of imidazolium ionic liquids on fish and human cell lines. *Arh. Hig. Rada Toksikol.* 63. <https://doi.org/10.2478/10004-1254-63-2012-2132>
- Czerwicka, M., Stolte, S., Müller, A., Siedlecka, E.M., Gołebiowski, M., Kumirska, J., Stepnowski, P., 2009. Identification of ionic liquid breakdown products in an advanced oxidation system. *J. Hazard. Mater.* 171. <https://doi.org/10.1016/j.jhazmat.2009.06.027>
- Das, R.N., Roy, K., 2012. Development of classification and regression models for *Vibrio fischeri* toxicity of ionic liquids: Green solvents for the future. *Toxicol. Res. (Camb)*. <https://doi.org/10.1039/c2tx20020a>
- Das, R.N., Roy, K., Popelier, P.L.A., 2015. Exploring simple, transparent, interpretable and predictive QSAR models for classification and quantitative prediction of rat toxicity of ionic liquids using OECD recommended guidelines. *Chemosphere* 139. <https://doi.org/10.1016/j.chemosphere.2015.06.022>
- de Melo, E.B., 2016. Correction: A structure–activity relationship study of the toxicity of ionic liquids using an adapted Ferreira–Kiralj hydrophobicity parameter. *Phys. Chem. Chem. Phys.* 18. <https://doi.org/10.1039/c5cp90218e>
- Delgado-Mellado, N., Ayuso, M., Villar-Chavero, M.M., García, J., Rodríguez, F., 2019. Ecotoxicity evaluation towards *Vibrio fischeri* of imidazolium- and pyridinium-based ionic liquids for their use in separation processes. *SN Appl. Sci.*

<https://doi.org/10.1007/s42452-019-0916-3>

- Deng, Y., Morrissey, S., Gathergood, N., Delort, A.M., Husson, P., Gomes, M.F.C., 2010. The presence of functional groups key for biodegradation in ionic liquids: Effect on gas solubility. *ChemSusChem* 3. <https://doi.org/10.1002/cssc.200900241>
- Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmos. Res.* <https://doi.org/10.1016/j.atmosres.2014.10.016>
- Ding, Y., Chen, M., Guo, C., Zhang, P., Wang, J., 2020. Molecular fingerprint-based machine learning assisted QSAR model development for prediction of ionic liquid properties. *J. Mol. Liq.* <https://doi.org/10.1016/j.molliq.2020.115212>
- Doherty, A.P., Diaconu, L., Marley, E., Spedding, P.L., Barhdadi, R., Troupel, M., 2012. Application of clean technologies using electrochemistry in ionic liquids. *Asia-Pacific J. Chem. Eng.* <https://doi.org/10.1002/apj.529>
- Dong, X., Fan, Y., Zhang, H., Zhong, Y., Yang, Y., Miao, J., Hua, S., 2016. Inhibitory effects of ionic liquids on the lactic dehydrogenase activity. *Int. J. Biol. Macromol.* 86. <https://doi.org/10.1016/j.ijbiomac.2016.01.059>
- Dupont, J., 2004. On the solid, liquid and solution structural organization of imidazolium ionic liquids. *J. Braz. Chem. Soc.* <https://doi.org/10.1590/S0103-50532004000300002>
- Dyson, G.M., Lynch, M.F., Morgan, H.L., 1968. A modified IUPAC-Dyson notation system for chemical structures. *Inf. Storage Retr.* 4. [https://doi.org/10.1016/0020-0271\(68\)90004-1](https://doi.org/10.1016/0020-0271(68)90004-1)
- Egorova, K.S., Gordeev, E.G., Ananikov, V.P., 2017. Biological Activity of Ionic Liquids and Their Application in Pharmaceutics and Medicine. *Chem. Rev.* <https://doi.org/10.1021/acs.chemrev.6b00562>
- Eichenlaub, J., Rakowska, P.W., Kloskowski, A., 2022. User-assisted methodology targeted for building structure interpretable QSPR models for boosting CO₂ capture with ionic liquids. *J. Mol. Liq.* 350. <https://doi.org/10.1016/j.molliq.2022.118511>

- Eike, D.M., Brennecke, J.F., Maginn, E.J., 2004. Predicting Infinite-Dilution Activity Coefficients of Organic Solutes in Ionic Liquids. *Ind. Eng. Chem. Res.* 43(4), 1039–1048. <https://doi.org/10.1021/ie034152p>
- Eini, S., Jhamb, S., Sharifzadeh, M., Rashtchian, D., Kontogeorgis, G.M., 2020. Developing group contribution models for the estimation of Atmospheric Lifetime and Minimum Ignition Energy. *Chem. Eng. Sci.* 226. <https://doi.org/10.1016/j.ces.2020.115866>
- Ekwall, B., Silano, V., Zucco, F., 1990. Chapter 7 - Toxicity Tests with Mammalian Cell Cultures. *Short-term Toxic. Tests Non-genotoxic Eff.*
- Elfgen, R., Gehrke, S., Hollóczki, O., 2020. Ionic Liquids as Extractants for Nanoplastics. *ChemSusChem* 13. <https://doi.org/10.1002/cssc.202001749>
- Endo, S., Hammer, J., 2020. Predicting Partition Coefficients of Short-Chain Chlorinated Paraffin Congeners by COSMO-RS-Trained Fragment Contribution Models. *Environ. Sci. Technol.* 54, 15162–15169. <https://doi.org/10.1021/acs.est.0c06506>
- Erisman, J.W., Bleeker, A., Galloway, J., Sutton, M.S., 2007. Reduced nitrogen in ecology and the environment. *Environ. Pollut.* <https://doi.org/10.1016/j.envpol.2007.06.033>
- Ertugrul, Ö.F., 2016. Forecasting electricity load by a novel recurrent extreme learning machines approach. *Int. J. Electr. Power Energy Syst.* 78, 429–435. <https://doi.org/10.1016/j.ijepes.2015.12.006>
- Fan, Y., Dong, X., Yan, L., Li, D., Hua, S., Hu, C., Pan, C., 2016. Evaluation of the toxicity of ionic liquids on trypsin: A mechanism study. *Chemosphere* 148. <https://doi.org/10.1016/j.chemosphere.2016.01.033>
- Fatemi, M.H., Izadiyan, P., 2011. Cytotoxicity estimation of ionic liquids based on their effective structural features. *Chemosphere.* <https://doi.org/10.1016/j.chemosphere.2011.04.021>
- Fellah, M.F., 2016. Adsorption of hydrogen sulfide as initial step of H₂S removal: A DFT study on metal exchanged ZSM-12 clusters. *Fuel Process. Technol.* 144, 191–196. <https://doi.org/10.1016/j.fuproc.2016.01.003>

- Frade, R.F.M., Matias, A., Branco, L.C., Afonso, C.A.M., Duarte, C.M.M., 2007. Effect of ionic liquids on human colon carcinoma HT-29 and CaCo-2 cell lines. *Green Chem.* 9. <https://doi.org/10.1039/b617526k>
- Fredlake, C.P., Crosthwaite, J.M., Hert, D.G., Aki, S.N.V.K., Brennecke, J.F., 2004. Thermophysical properties of imidazolium-based ionic liquids. *J. Chem. Eng. Data* 49, 954–964. <https://doi.org/10.1021/je034261a>
- Frutiger, J., Marcarie, C., Abildskov, J., Sin, G., 2016. A Comprehensive Methodology for Development, Parameter Estimation, and Uncertainty Analysis of Group Contribution Based Property Models-An Application to the Heat of Combustion. *J. Chem. Eng. Data* 61, 602–613. <https://doi.org/10.1021/acs.jced.5b00750>
- Fulton, M.H., Key, P.B., 2001. Acetylcholinesterase inhibition in estuarine fish and invertebrates as an indicator of organophosphorus insecticide exposure and effects. *Environ. Toxicol. Chem.* <https://doi.org/10.1002/etc.5620200104>
- Galán Sánchez, L.M., Meindersma, G.W., de Haan, A.B., 2007. Solvent Properties of Functionalized Ionic Liquids for CO₂ Absorption. *Chem. Eng. Res. Des.* 85(1A), 31–39. <https://doi.org/10.1205/cherd06124>
- Gani, R., 2019. Group contribution-based property estimation methods: advances and perspectives. *Curr. Opin. Chem. Eng.* 23, 184–196. <https://doi.org/10.1016/j.coche.2019.04.007>
- Gani, R., Harper, P.M., Hostrup, M., 2005. Automatic creation of missing groups through connectivity index for pure-component property prediction. *Ind. Eng. Chem. Res.* 44, 7262–7269. <https://doi.org/10.1021/ie0501881>
- Gao, X., Qu, H., Shan, S., Song, C., Baranenko, D., 2020. Prediction of toxicity of Ionic Liquids based on GC-COSMO method. *J. Hazard. Mater.* 122964. <https://doi.org/10.1016/j.carbpol.2020.115849>
- García-Lorenzo, A., Tojo, E., Tojo, J., Teijeira, M., Rodríguez-Berrocal, F.J., González, M.P., Martínez-Zorzano, V.S., 2008. Cytotoxicity of selected imidazolium-derived ionic liquids in the human Caco-2 cell line. Sub-structural toxicological interpretation through a QSAR study. *Green Chem.* 10. <https://doi.org/10.1039/b718860a>

- Garcia, M.T., Gathergood, N., Scammells, P.J., 2005. Biodegradable ionic liquids Part II. Effect of the anion and toxicology. *Green Chem.* <https://doi.org/10.1039/b411922c>
- Gardas, R.L., Coutinho, J.A.P., 2008. A group contribution method for viscosity estimation of ionic liquids. *Fluid Phase Equilib.* <https://doi.org/10.1016/j.fluid.2008.01.021>
- Gardas, R.L., P.Coutinho, J.A., 2009. Group Contribution Methods for the Prediction of Thermophysical and Transport Properties of Ionic Liquids. *AIChE J.* 55, 1274–1290. <https://doi.org/10.1002/aic>
- Gasteiger, J., 2016. Chemoinformatics: Achievements and challenges, a personal view. *Molecules.* <https://doi.org/10.3390/molecules21020151>
- Gathergood, N., Garcia, M.T., Scammells, P.J., 2004. Biodegradable ionic liquids: Part I. Concept, preliminary targets and evaluation. *Green Chem.* 6. <https://doi.org/10.1039/b315270g>
- Ge, H.L., Liu, S.S., Su, B.X., Zhu, X.W., 2014. Two-stage prediction of the effects of imidazolium and pyridinium ionic liquid mixtures on luciferase. *Molecules* 19. <https://doi.org/10.3390/molecules19056877>
- Gebbie, M.A., Smith, A.M., Dobbs, H.A., Lee, A.A., Warr, G.G., Banquy, X., Valtiner, M., Rutland, M.W., Israelachvili, J.N., Perkin, S., Atkin, R., 2017. Long range electrostatic forces in ionic liquids. *Chem. Commun.* <https://doi.org/10.1039/c6cc08820a>
- Ghanem, O. Ben, Mutalib, M.I.A., El-Harbawi, M., Gonfa, G., Kait, C.F., Alitheen, N.B.M., Lévêque, J.M., 2015. Effect of imidazolium-based ionic liquids on bacterial growth inhibition investigated via experimental and QSAR modelling studies. *J. Hazard. Mater.* 297. <https://doi.org/10.1016/j.jhazmat.2015.04.082>
- Ghanem, O. Ben, Mutalib, M.I.A., Lévêque, J.M., El-Harbawi, M., 2017. Development of QSAR model to predict the ecotoxicity of *Vibrio fischeri* using COSMO-RS descriptors. *Chemosphere* 170, 242–250. <https://doi.org/10.1016/j.chemosphere.2016.12.003>
- Ghanem, O. Ben, Shah, S.N., Lévêque, J.M., Mutalib, M.I.A., El-Harbawi, M., Khan,

- A.S., Alnarabiji, M.S., Al-Absi, H.R.H., Ullah, Z., 2018. Study of the antimicrobial activity of cyclic cation-based ionic liquids via experimental and group contribution QSAR model. *Chemosphere* 195, 21–28. <https://doi.org/10.1016/j.chemosphere.2017.12.018>
- Gharagheizi, F., Ilani-Kashkouli, P., Mohammadi, A.H., 2012a. Group contribution model for estimation of surface tension of ionic liquids. *Chem. Eng. Sci.* 78, 204–208. <https://doi.org/10.1016/j.ces.2012.05.008>
- Gharagheizi, F., Sattari, M., Ilani-Kashkouli, P., Mohammadi, A.H., Ramjugernath, D., Richon, D., 2012b. Quantitative structure-property relationship for thermal decomposition temperature of ionic liquids. *Chem. Eng. Sci.* <https://doi.org/10.1016/j.ces.2012.08.036>
- Ghaslani, D., Eshaghi Gorji, Z., Ebrahimipour Gorji, A., Riahi, S., 2017. Descriptive and predictive models for Henry's law constant of CO₂ in ionic liquids: A QSPR study. *Chem. Eng. Res. Des.* 120. <https://doi.org/10.1016/j.cherd.2016.12.020>
- Gholampour, F., Yeganegi, S., 2014. Molecular simulation study on the adsorption and separation of acidic gases in a model nanoporous carbon. *Chem. Eng. Sci.* 117, 426–435. <https://doi.org/10.1016/j.ces.2014.07.003>
- Glanzmann, N., Carmo, A.M.L., Antinarelli, L.M.R., Coimbra, E.S., Costa, L.A.S., da Silva, A.D., 2018. Synthesis, characterization, and NMR studies of 1,2,3-triazolium ionic liquids: a good perspective regarding cytotoxicity. *J. Mol. Model.* 24. <https://doi.org/10.1007/s00894-018-3682-z>
- Gomes, J.M., Silva, S.S., Reis, R.L., 2019. Biocompatible ionic liquids: Fundamental behaviours and applications. *Chem. Soc. Rev.* <https://doi.org/10.1039/c9cs00016j>
- Gomes, M.F.C., Husson, P., 2009. Ionic liquids: Promising media for gas separations, in: ACS Symposium Series. <https://doi.org/10.1021/bk-2009-1030.ch016>
- Gomez-Herrero, E., Tobajas, M., Polo, A., Rodriguez, J.J., Mohedano, A.F., 2019. Removal of imidazolium-based ionic liquid by coupling Fenton and biological oxidation. *J. Hazard. Mater.* 365. <https://doi.org/10.1016/j.jhazmat.2018.10.097>
- Goodrich, B.F., De La Fuente, J.C., Gurkan, B.E., Zadigian, D.J., Price, E.A., Huang, Y., Brennecke, J.F., 2011. Experimental measurements of amine-functionalized

- anion-tethered ionic liquids with carbon dioxide. *Ind. Eng. Chem. Res.* 50(1), 111–118. <https://doi.org/10.1021/ie101688a>
- Gordon, C.M., 2001. New developments in catalysis using ionic liquids. *Appl. Catal. A Gen.* 222. [https://doi.org/10.1016/S0926-860X\(01\)00834-1](https://doi.org/10.1016/S0926-860X(01)00834-1)
- Gouveia, W., Jorge, T.F., Martins, S., Meireles, M., Carolino, M., Cruz, C., Almeida, T. V., Araújo, M.E.M., 2014. Toxicity of ionic liquids prepared from biomaterials. *Chemosphere* 104. <https://doi.org/10.1016/j.chemosphere.2013.10.055>
- Gramatica, P., 2007. Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* <https://doi.org/10.1002/qsar.200610151>
- Greer, A.J., Jacquemin, J., Hardacre, C., 2020. Industrial Applications of Ionic Liquids. *Molecules.* <https://doi.org/10.3390/molecules25215207>
- Grzonkowska, M., Sosnowska, A., Barycki, M., Rybinska, A., Puzyn, T., 2016. How the structure of ionic liquid affects its toxicity to *Vibrio fischeri*? *Chemosphere.* <https://doi.org/10.1016/j.chemosphere.2016.06.004>
- Gupta, S., Basant, N., Singh, K.P., 2015. Nonlinear QSAR modeling for predicting cytotoxicity of ionic liquids in leukemia rat cell line: an aid to green chemicals designing. *Environ. Sci. Pollut. Res.* 22. <https://doi.org/10.1007/s11356-015-4526-3>
- Gutowski, K.E., 2018. Industrial uses and applications of ionic liquids. *Phys. Sci. Rev.* 3. <https://doi.org/10.1515/psr-2017-0191>
- Hansch, C., T., F., 1964. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* 86.
- He, W., Yan, F., Jia, Q., Xia, S., Wang, Q., 2018. QSAR models for describing the toxicological effects of ILs against *Staphylococcus aureus* based on norm indexes. *Chemosphere* 195. <https://doi.org/10.1016/j.chemosphere.2017.12.091>
- Helgaker, T., Klopper, W., Tew, D.P., 2008. Quantitative quantum chemistry. *Mol. Phys.* 106, 2107–2143. <https://doi.org/10.1080/00268970802258591>
- Hernández-Fernández, F.J., Bayo, J., Pérez de los Ríos, A., Vicente, M.A., Bernal, F.J., Quesada-Medina, J., 2015. Discovering less toxic ionic liquids by using the Microtox® toxicity test. *Ecotoxicol. Environ. Saf.* 116.

<https://doi.org/10.1016/j.ecoenv.2015.02.034>

- Hodyna, D., Kovalishyn, V., Semenyuta, I., Blagodatnyi, V., Rogalsky, S., Metelytsia, L., 2018. Imidazolium ionic liquids as effective antiseptics and disinfectants against drug resistant *S. aureus*: In silico and in vitro studies. *Comput. Biol. Chem.* 73. <https://doi.org/10.1016/j.compbiolchem.2018.01.012>
- Hosseini, M., Rahimi, R., Ghaedi, M., 2020. Hydrogen sulfide solubility in different ionic liquids: an updated database and intelligent modeling. *J. Mol. Liq.* <https://doi.org/10.1016/j.molliq.2020.113984>
- Hosseini, S.M., Mulero, A., Alavianmehr, M.M., 2019. Predictive methods and semi-classical Equations of State for pure ionic liquids: A review. *J. Chem. Thermodyn.* <https://doi.org/10.1016/j.jct.2018.09.022>
- Hou, X.D., Liu, Q.P., Smith, T.J., Li, N., Zong, M.H., 2013. Evaluation of Toxicity and Biodegradability of Cholinium Amino Acids Ionic Liquids. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0059145>
- Huang, G. Bin, Zhu, Q.Y., Siew, C.K., 2006a. Extreme learning machine: Theory and applications. *Neurocomputing.* <https://doi.org/10.1016/j.neucom.2005.12.126>
- Huang, G. Bin, Zhu, Q.Y., Siew, C.K., 2006b. Extreme learning machine: Theory and applications. *Neurocomputing* 70, 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- Huang, G. Bin, Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: A new learning scheme of feedforward neural networks, in: *IEEE International Conference on Neural Networks - Conference Proceedings.* <https://doi.org/10.1109/IJCNN.2004.1380068>
- Huang, Gao, Huang, Guang Bin, Song, S., You, K., 2015. Trends in extreme learning machines: A review. *Neural Networks* 61, 32–48. <https://doi.org/10.1016/j.neunet.2014.10.001>
- Huang, K., Cai, D.N., Chen, Y. Le, Wu, Y.T., Hu, X.B., Zhang, Z.B., 2013. Thermodynamic validation of 1-alkyl-3-methylimidazolium carboxylates as task-specific ionic liquids for H₂S absorption. *AIChE J.* 59(6), 2227–2235. <https://doi.org/10.1002/aic.13976>

- Huang, K., Zhang, X.M., Xu, Y., Wu, Y.T., Hu, X.B., Xu, Y., 2014. Protic ionic liquids for the selective absorption of H₂S from CO₂: Thermodynamic analysis. *AIChE J.* 60(12), 4232–4240. <https://doi.org/10.1002/aic.14634>
- Huang, K., Zhang, X.M., Zhou, L. Sen, Tao, D.J., Fan, J.P., 2017. Highly efficient and selective absorption of H₂S in phenolic ionic liquids: A cooperative result of anionic strong basicity and cationic hydrogen-bond donation. *Chem. Eng. Sci.* 173. <https://doi.org/10.1016/j.ces.2017.07.048>
- Huang, Y., Dong, H., Zhang, X., Li, C., Zhang, S., 2013. A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids. *AIChE J.* 59, 1348–1359. <https://doi.org/10.1002/aic.13910>
- Huang, Y., Zhao, Y., Zeng, S., Zhang, X., Zhang, S., 2014. Density prediction of mixtures of ionic liquids and molecular solvents using two new generalized models. *Ind. Eng. Chem. Res.* 53(39), 15270–15277. <https://doi.org/10.1021/ie502571b>
- Hukkerikar, A.S., Sarup, B., Ten Kate, A., Abildskov, J., Sin, G., Gani, R., 2012. Group-contribution + (GC +) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilib.* 321, 25–43. <https://doi.org/10.1016/j.fluid.2012.02.010>
- Hurley, F.H., Wier, T.P., 1951. Electrodeposition of Metals from Fused Quaternary Ammonium Salts. *J. Electrochem. Soc.* 98. <https://doi.org/10.1149/1.2778132>
- Imaizumi, Y., 2016. Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Japanese J. Pestic. Sci.* 41. <https://doi.org/10.1584/jpestics.w15-48>
- Ismail Hossain, M., Samir, B.B., El-Harbawi, M., Masri, A.N., Abdul Mutalib, M.I., Hefter, G., Yin, C.Y., 2011. Development of a novel mathematical model using a group contribution method for prediction of ionic liquid toxicities. *Chemosphere.* <https://doi.org/10.1016/j.chemosphere.2011.06.088>
- Isosaari, P., Srivastava, V., Sillanpää, M., 2019. Ionic liquid-based water treatment technologies for organic pollutants: Current status and future prospects of ionic liquid mediated technologies. *Sci. Total Environ.* 690, 604–619.

<https://doi.org/10.1016/j.scitotenv.2019.06.421>

- Jafari, M., Keshavarz, M.H., Salek, H., 2019. A simple method for assessing chemical toxicity of ionic liquids on *Vibrio fischeri* through the structure of cations with specific anions. *Ecotoxicol. Environ. Saf.* 182, 109429. <https://doi.org/10.1016/j.ecoenv.2019.109429>
- Jalili, A.H., Mehdizadeh, A., Shokouhi, M., Ahmadi, A.N., Hosseini-Jenab, M., Fateminassab, F., 2010. Solubility and diffusion of CO₂ and H₂S in the ionic liquid 1-ethyl-3-methylimidazolium ethylsulfate. *J. Chem. Thermodyn.* 42, 1298–1303. <https://doi.org/10.1016/j.jct.2010.05.008>
- Jalili, A.H., Mehrabi, M., Zoghi, A.T., Shokouhi, M., Taheri, S.A., 2017. Solubility of carbon dioxide and hydrogen sulfide in the ionic liquid 1-butyl-3-methylimidazolium trifluoromethanesulfonate. *Fluid Phase Equilib.* 453, 1–12. <https://doi.org/10.1016/j.fluid.2017.09.003>
- Jalili, A.H., Rahmati-Rostami, M., Ghotbi, C., Hosseini-Jenab, M., Ahmadi, A.N., 2009. Solubility of H₂S in ionic liquids [bmim][PF₆], [bmim][BF₄], and [bmim][Tf₂N]. *J. Chem. Eng. Data* 54, 1844–1849. <https://doi.org/10.1021/je8009495>
- Jalili, A.H., Shokouhi, M., Maurer, G., Hosseini-Jenab, M., 2013. Solubility of CO₂ and H₂S in the ionic liquid 1-ethyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate. *J. Chem. Thermodyn.* 55(4), 1663–1668. <https://doi.org/10.1016/j.jct.2013.07.022>
- Jalili, A.H., Shokouhi, M., Maurer, G., Zoghi, A.T., Sadeghzah Ahari, J., Forsat, K., 2019. Measuring and modelling the absorption and volumetric properties of CO₂ and H₂S in the ionic liquid 1-ethyl-3-methylimidazolium tetrafluoroborate. *J. Chem. Thermodyn.* 131, 544–556. <https://doi.org/10.1016/j.jct.2018.12.005>
- Jastorff, B., Mölter, K., Behrend, P., Bottin-Weber, U., Filser, J., Heimers, A., Ondruschka, B., Ranke, J., Schaefer, M., Schröder, H., Stark, A., Stepnowski, P., Stock, F., Störmann, R., Stolte, S., Welz-Biermann, U., Ziegert, S., Thöming, J., 2005. Progress in evaluation of risk potential of ionic liquids - Basis for an eco-design of sustainable products, in: *Green Chemistry*. <https://doi.org/10.1039/b418518h>

- Jastorff, B., Störmann, R., Ranke, J., 2007. Thinking in structure-activity relationships - A way forward towards sustainable chemistry. *Clean - Soil, Air, Water* 35, 399–405. <https://doi.org/10.1002/clen.200720018>
- 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>
- Jordan, A., Gathergood, N., 2015. Biodegradation of ionic liquids-a critical review. *Chem. Soc. Rev.* <https://doi.org/10.1039/c5cs00444f>
- Jou, F.Y., Mather, A.E., 2007. Solubility of hydrogen sulfide in [bmim][PF₆]. *Int. J. Thermophys.* 28, 490–495. <https://doi.org/10.1007/s10765-007-0185-z>
- Kaiser, K.L.E., 1998. Correlations of *Vibrio fischeri* bacteria test data with bioassay data for other organisms, in: *Environmental Health Perspectives*. <https://doi.org/10.2307/3433809>
- Kang, X., Chen, Z., Zhao, Y., 2020a. Assessing the ecotoxicity of ionic liquids on *Vibrio fischeri* using electrostatic potential descriptors. *J. Hazard. Mater.* 397. <https://doi.org/10.1016/j.jhazmat.2020.122761>
- Kang, X., Liu, C., Zeng, S., Zhao, Z., Qian, J., Zhao, Y., 2018a. Prediction of Henry's law constant of CO₂ in ionic liquids based on SEP and S σ -profile molecular descriptors. *J. Mol. Liq.* 262, 139–147. <https://doi.org/10.1016/j.molliq.2018.04.026>
- Kang, X., Liu, C., Zeng, S., Zhao, Z., Qian, J., Zhao, Y., 2018b. Prediction of Henry's law constant of CO₂ in ionic liquids based on SEP and S σ -profile molecular descriptors. *J. Mol. Liq.* <https://doi.org/10.1016/j.molliq.2018.04.026>
- Kang, X., Liu, X., Li, J., Zhao, Y., Zhang, H., 2018c. Heat Capacity Prediction of Ionic Liquids Based on Quantum Chemistry Descriptors. *Ind. Eng. Chem. Res.* 57(49), 16989–16994. <https://doi.org/10.1021/acs.iecr.8b03668>
- Kang, X., Lv, Z., Chen, Z., Zhao, Y., 2020b. Prediction of ammonia absorption in ionic liquids based on extreme learning machine modelling and a novel molecular descriptor SEP. *Environ. Res.* 189. <https://doi.org/10.1016/j.envres.2020.109951>
- Kang, X., Lv, Z., Zhao, Y., Chen, Z., 2020c. A QSPR model for estimating Henry's law

- constant of H₂S in ionic liquids by ELM algorithm. *Chemosphere* 269, 128743. <https://doi.org/10.1016/j.chemosphere.2020.128743>
- Kang, X., Qian, J., Deng, J., Latif, U., Zhao, Y., 2018d. Novel molecular descriptors for prediction of H₂S solubility in ionic liquids. *J. Mol. Liq.* 265, 756–764. <https://doi.org/10.1016/j.molliq.2018.06.113>
- Kang, X., Zhao, Y., Chen, Z., 2021. Atom surface fragment contribution method for predicting the toxicity of ionic liquids. *J. Hazard. Mater.* 421. <https://doi.org/10.1016/j.jhazmat.2021.126705>
- Kang, X., Zhao, Y., Li, J., 2018e. Predicting refractive index of ionic liquids based on the extreme learning machine (ELM) intelligence algorithm. *J. Mol. Liq.* 250, 44–49. <https://doi.org/10.1016/j.molliq.2017.11.166>
- Kang, X., Zhao, Z., Qian, J., Afzal, R.M., 2017. Predicting the viscosity of ionic liquids by the ELM intelligence algorithm. *Ind. Eng. Chem. Res.* 56(39), 11344–11351. <https://doi.org/10.1021/acs.iecr.7b02722>
- Karadas, F., Atilhan, M., Aparicio, S., 2010. Review on the use of ionic liquids (ILs) as alternative fluids for CO₂ capture and natural gas sweetening. *Energy and Fuels.* <https://doi.org/10.1021/ef1011337>
- Kashinath, S.A.A., Hashim, H., Mustaffa, A.A., Yunus, N.A., 2020. Cetane number estimation of pure compound using group contribution method. *Chem. Eng. Trans.* 78. <https://doi.org/10.3303/CET2078056>
- Katritzky, A.R., Jain, R., Lomaka, A., Petrukhin, R., Karelson, M., Visser, A.E., Rogers, R.D., 2002. Correlation of the melting points of potential ionic liquids (imidazolium bromides and benzimidazolium bromides) using the CODESSA program. *J. Chem. Inf. Comput. Sci.* 42, 225–231. <https://doi.org/10.1021/ci0100494>
- Khodadoust, A.P., Chandrasekaran, S., Dionysiou, D.D., 2006. Preliminary assessment of imidazolium-based room-temperature ionic liquids for extraction of organic contaminants from soils. *Environ. Sci. Technol.* 40. <https://doi.org/10.1021/es051563j>
- Klopman, G., Li, J.Y., Wang, S., Dimayuga, M., Wang, S., Dimayuga, M., 1994.

- Computer Automated log P Calculations Based on an Extended Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* 34, 752–781. <https://doi.org/10.1021/ci00020a009>
- Kochev, N., Paskaleva, V., Pukalov, O., Jeliaskova, N., 2019. Ambit-GCM: An Open-source Software Tool for Group Contribution Modelling. *Mol. Inform.* 38. <https://doi.org/10.1002/minf.201800138>
- Kolář, M.H., Hobza, P., 2016. Computer Modeling of Halogen Bonds and Other σ -Hole Interactions. *Chem. Rev.* 116, 5155–5187. <https://doi.org/10.1021/acs.chemrev.5b00560>
- Koo, Y.-M., Ha, S.H., 2008. Application of ionic liquid in biotechnology. *J. Biotechnol.* 136. <https://doi.org/10.1016/j.jbiotec.2008.07.1792>
- Koutsoukos, S., Philippi, F., Malaret, F., Welton, T., 2021. A review on machine learning algorithms for the ionic liquid chemical space. *Chem. Sci.* <https://doi.org/10.1039/d1sc01000j>
- Kroon, M.C., Buijs, W., Peters, C.J., Witkamp, G.J., 2006. Decomposition of ionic liquids in electrochemical processing. *Green Chem.* 8, 241–245. <https://doi.org/10.1039/b512724f>
- Kumelan, J., Kamps, Á.P.S., Tuma, D., Maurer, G., 2005. Solubility of CO in the ionic liquid [bmim][PF₆]. *Fluid Phase Equilib.* 228–229, 207–211. <https://doi.org/10.1016/j.fluid.2004.07.015>
- Kumelan, J., Pérez-Salado Kamps, Á., Tuma, D., Maurer, G., 2006. Solubility of H₂ in the Ionic Liquid [bmim][PF₆]. *J. Chem. Eng. Data.* <https://doi.org/10.1021/je050362s>
- Kusumahastuti, D.K.A., Sihtmäe, M., Kapitanov, I. V., Karpichev, Y., Gathergood, N., Kahru, A., 2019. Toxicity profiling of 24 L-phenylalanine derived ionic liquids based on pyridinium, imidazolium and cholinium cations and varying alkyl chains using rapid screening *Vibrio fischeri* bioassay. *Ecotoxicol. Environ. Saf.* 172. <https://doi.org/10.1016/j.ecoenv.2018.12.076>
- Laali, K.K., 2003. *Ionic Liquids in Synthesis.* Synthesis (Stuttg). <https://doi.org/10.1055/s-2003-40869>

- Lai, J.Q., Li, Z., Lü, Y.H., Yang, Z., 2011. Specific ion effects of ionic liquids on enzyme activity and stability. *Green Chem.* 13. <https://doi.org/10.1039/c1gc15140a>
- Lan, T., Yan, X., Yan, F., Xia, S., Jia, Q., Wang, Q., 2020. Norm index in QSTR work for predicting toxicity of ionic liquids on *Vibrio fischeri*. *Ecotoxicol. Environ. Saf.* <https://doi.org/10.1016/j.ecoenv.2020.111187>
- Latała, A., Ndzi, M., Stepnowski, P., 2009. Toxicity of imidazolium and pyridinium based ionic liquids towards algae. *Bacillaria paxillifer* (a microphytobenthic diatom) and *Geitlerinema amphibium* (a microphytobenthic blue green alga). *Green Chem.* <https://doi.org/10.1039/b901887e>
- Lazidou, D., Mastrogeorgopoulos, S., Panayiotou, C., 2019. Thermodynamic characterization of ionic liquids. *J. Mol. Liq.* 277, 10–21. <https://doi.org/10.1016/j.molliq.2018.12.023>
- Lazzús, J.A., 2012. A group contribution method to predict the melting point of ionic liquids. *Fluid Phase Equilib.* 313, 1–6. <https://doi.org/10.1016/j.fluid.2011.09.018>
- Lazzús, J.A., 2009. Neural network based on quantum chemistry for predicting melting point of organic compounds. *Chinese J. Chem. Phys.* <https://doi.org/10.1088/1674-0068/22/01/19-26>
- Lee, B.S., Lin, S.T., 2017. Prediction and screening of solubility of pharmaceuticals in single- and mixed-ionic liquids using COSMO-SAC model. *AIChE J.* <https://doi.org/10.1002/aic.15595>
- Lei, Z., Dai, C., Chen, B., 2014a. Gas solubility in ionic liquids. *Chem. Rev.* 114, 1289–1326. <https://doi.org/10.1021/cr300497a>
- Lei, Z., Dai, C., Chen, B., 2014b. Gas solubility in ionic liquids. *Chem. Rev.* <https://doi.org/10.1021/cr300497a>
- Lewis, E., 2018. REGULATION (EC) No 1907/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. *Off. J. Eur. Union.*
- Li, F., Yang, Z., Weng, H., Chen, G., Lin, M., Zhao, C., 2018. High efficient separation of U(VI) and Th(IV) from rare earth elements in strong acidic solution by selective sorption on phenanthroline diamide functionalized graphene oxide.

- Chem. Eng. J. 332. <https://doi.org/10.1016/j.cej.2017.09.038>
- Li, G., Zhou, Q., Zhang, X., LeiWang, Zhang, S., Li, J., 2010. Solubilities of ammonia in basic imidazolium ionic liquids. *Fluid Phase Equilib.* 297, 34–39. <https://doi.org/10.1016/j.fluid.2010.06.005>
- Li, L.L., Sun, J., Tseng, M.L., Li, Z.G., 2019. Extreme learning machine optimized by whale optimization algorithm using insulated gate bipolar transistor module aging degree evaluation. *Expert Syst. Appl.* <https://doi.org/10.1016/j.eswa.2019.03.002>
- Li, X., Zhao, J., Li, Q., Wang, L., Tsang, S.C., 2007. Ultrasonic chemical oxidative degradations of 1,3-dialkylimidazolium ionic liquids and their mechanistic elucidations. *Dalt. Trans.* <https://doi.org/10.1039/b618384k>
- Li, X.Y., Jing, C.Q., Zang, X.Y., Yang, S., Wang, J.J., 2012. Toxic cytological alteration and mitochondrial dysfunction in PC12 cells induced by 1-octyl-3-methylimidazolium chloride. *Toxicol. Vitro.* 26. <https://doi.org/10.1016/j.tiv.2012.07.006>
- Li, Z., Jiang, W.T., Chang, P.H., Lv, G., Xu, S., 2014. Modification of a Ca-montmorillonite with ionic liquids and its application for chromate removal. *J. Hazard. Mater.* 270. <https://doi.org/10.1016/j.jhazmat.2014.01.054>
- Li, Z., Liao, F., Jiang, F., Liu, B., Ban, S., Chen, G., Sun, C., Xiao, P., Sun, Y., 2016. Capture of H₂S and SO₂ from trace sulfur containing gas mixture by functionalized UiO-66(Zr) materials: A molecular simulation study. *Fluid Phase Equilib.* 427, 259–267. <https://doi.org/10.1016/j.fluid.2016.07.020>
- Li, Z., Zhang, Xiangping, Dong, H., Zhang, Xiaochun, Gao, H., Zhang, S., Li, J., Wang, C., 2015. Efficient absorption of ammonia with hydroxyl-functionalized ionic liquids. *RSC Adv.* 5, 81362–81370. <https://doi.org/10.1039/c5ra13730f>
- Liu, H., Tian, H.Q., Li, Y.F., 2015. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. *Energy Convers. Manag.* <https://doi.org/10.1016/j.enconman.2015.04.057>
- Liu, X., Li, J., Wang, R., 2017. Study on the desulfurization performance of hydramine/ionic liquid solutions at room temperature and atmospheric pressure. *Fuel Process. Technol.* 167, 382–387.

<https://doi.org/10.1016/j.fuproc.2017.07.023>

- Liu, Xue, Nie, Y., Liu, Y., Zhang, S., Skov, A.L., 2018. Screening of Ionic Liquids for Keratin Dissolution by Means of COSMO-RS and Experimental Verification. *ACS Sustain. Chem. Eng.* 6. <https://doi.org/10.1021/acssuschemeng.8b04830>
- Liu, Xinyan, Zhou, T., Zhang, X., Zhang, S., Liang, X., Gani, R., Kontogeorgis, G.M., 2018. Application of COSMO-RS and UNIFAC for ionic liquids based gas separation. *Chem. Eng. Sci.* 192, 816–828. <https://doi.org/10.1016/j.ces.2018.08.002>
- Loschen, C., Klamt, A., 2012. COSMO quick: A novel interface for fast σ -profile composition and its application to COSMO-RS solvent screening using multiple reference solvents. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/ie3023675>
- Lu, T., Chen, F., 2012. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* 33(5), 580–592. <https://doi.org/10.1002/jcc.22885>
- Luis, P., Garea, A., Irabien, A., 2010. Quantitative structure-activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC50 (*Vibrio fischeri*). *J. Mol. Liq.* <https://doi.org/10.1016/j.molliq.2009.12.008>
- Luis, P., Ortiz, I., Aldaco, R., Irabien, A., 2007. A novel group contribution method in the development of a QSAR for predicting the toxicity (*Vibrio fischeri* EC50) of ionic liquids. *Ecotoxicol. Environ. Saf.* <https://doi.org/10.1016/j.ecoenv.2006.06.010>
- Luo, X., Jiang, C., Wang, W., Xu, Y., Wang, J.H., Zhao, W., 2019. User behavior prediction in social networks using weighted extreme learning machine with distribution optimization. *Futur. Gener. Comput. Syst.* <https://doi.org/10.1016/j.future.2018.04.085>
- Luo, Y.R., Wang, S.H., Yun, M.X., Li, X.Y., Wang, J.J., Sun, Z.J., 2009. The toxic effects of ionic liquids on the activities of acetylcholinesterase and cellulase in earthworms. *Chemosphere* 77. <https://doi.org/10.1016/j.chemosphere.2009.07.026>
- M., M.W., H., H.P., 1995. Atom Fragment Contribution Method for Estimating Octanol-Water Partition-Coefficients. *J. Pharm. Sci.* 84, 83–92.

- Ma, S., Lv, M., Deng, F., Zhang, X., Zhai, H., Lv, W., 2015. Predicting the ecotoxicity of ionic liquids towards *Vibrio fischeri* using genetic function approximation and least squares support vector machine. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2014.10.011>
- Mallakpour, S., Dinari, M., 2012. Ionic liquids as green solvents: Progress and prospects, in: *Green Solvents II: Properties and Applications of Ionic Liquids*. https://doi.org/10.1007/978-94-007-2891-2_1
- Marrero, J., Gani, R., 2001. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* 183, 183–208. [https://doi.org/10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9)
- Matzke, M., Stolte, S., Thiele, K., Juffernholz, T., Arning, J., Ranke, J., Welz-Biermann, U., Jastorff, B., 2007. The influence of anion species on the toxicity of 1-alkyl-3-methylimidazolium ionic liquids observed in an (eco)toxicological test battery. *Green Chem.* <https://doi.org/10.1039/b705795d>
- Matzke, M., Thiele, K., Müller, A., Filser, J., 2009. Sorption and desorption of imidazolium based ionic liquids in different soil types. *Chemosphere* 74. <https://doi.org/10.1016/j.chemosphere.2008.09.049>
- Mayne, A.J., Rao, C.R., Mitra, S.K., 1972. Generalized Inverse of Matrices and Its Applications. *Oper. Res. Q.* <https://doi.org/10.2307/3007981>
- McFarlane, J., Ridenour, W.B., Luo, H., Hunt, R.D., DePaoli, D.W., Ren, R.X., 2005. Room temperature ionic liquids for separating organics from produced water. *Sep. Sci. Technol.* 40, 1245–1265. <https://doi.org/10.1081/SS-200052807>
- Mehraein, I., Riahi, S., 2017. The QSPR models to predict the solubility of CO₂ in ionic liquids based on least-squares support vector machines and genetic algorithm-multi linear regression. *J. Mol. Liq.* 225. <https://doi.org/10.1016/j.molliq.2016.10.133>
- Milota, M., Mosher, P., Li, K., 2007. VOC and HAP removal from dryer exhaust gas by absorption into ionic liquids. *For. Prod. J.* 57.
- Mital, D.K., Nancarrow, P., Zeinab, S., Jabbar, N.A., Ibrahim, T.H., Khamis, M.I., Taha, A., 2021. Group contribution estimation of ionic liquid melting points: Critical

- evaluation and refinement of existing models. *Molecules* 26, 66–80. <https://doi.org/10.3390/molecules26092454>
- Montalbán, M.G., Hidalgo, J.M., Collado-González, M., Díaz Baños, F.G., Villora, G., 2016. Assessing chemical toxicity of ionic liquids on *Vibrio fischeri*: Correlation with structure and composition. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2016.04.042>
- Morawski, A.W., Janus, M., Goc-Maciejewska, I., Syguda, A., Pernak, J., 2005. Decomposition of ionic liquids by photocatalysis. *Pol. J. Chem.* 79.
- Motlagh, S.R., Harun, R., Awang Biak, D.R., Hussain, S.A., Omar, R., Elgharbawy, A.A., 2020. COSMO-RS based prediction for alpha-linolenic acid (ALA) extraction from microalgae biomass using room temperature ionic liquids (RTILS). *Mar. Drugs*. <https://doi.org/10.3390/md18020108>
- Mrozik, W., Jungnickel, C., Skup, M., Urbaszek, P., Stepnowski, P., 2008. Determination of the adsorption mechanism of imidazolium-type ionic liquids onto kaolinite: Implications for their fate and transport in the soil environment. *Environ. Chem.* 5. <https://doi.org/10.1071/EN08015>
- Mrozik, W., Kotłowska, A., Kamysz, W., Stepnowski, P., 2012. Sorption of ionic liquids onto soils: Experimental and chemometric studies. *Chemosphere* 88. <https://doi.org/10.1016/j.chemosphere.2012.03.070>
- Mu, T., Rarey, J., Gmehling, J., 2007. Group contribution prediction of surface charge density profiles for COSMO-RS(01). *AIChE J.* 53, 3231–3240. <https://doi.org/10.1002/aic.11338>
- Muldoon, M.J., Aki, S.N.V.K., Anderson, J.L., Dixon, J.K., Brennecke, J.F., 2007. Improving carbon dioxide solubility in ionic liquids. *J. Phys. Chem. B* 111(30), 9001–9009. <https://doi.org/10.1021/jp071897q>
- Murray, J.S., Politzer, P., 2017. Molecular electrostatic potentials and noncovalent interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7, 1–10. <https://doi.org/10.1002/wcms.1326>
- Murray, J.S., Politzer, P., 2011a. The electrostatic potential: An overview. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 154–163. <https://doi.org/10.1002/wcms.19>

- Murray, J.S., Politzer, P., 2011b. The electrostatic potential: An overview. Wiley Interdiscip. Rev. Comput. Mol. Sci. <https://doi.org/10.1002/wcms.19>
- Murray, J.S., Politzer, P., 2002. Electrostatic Potentials: Chemical Applications, in: Encyclopedia of Computational Chemistry. <https://doi.org/10.1002/0470845015.cca014>
- Murray, J.S., Politzer, P., 1998. Statistical analysis of the molecular surface electrostatic potential: an approach to describing noncovalent interactions in condensed phases. *J. Mol. Struct. TheoChem* 425, 107–114.
- Nabipour, N., Qasem, S.N., Salwana, E., Baghban, A., 2020. Evolving LSSVM and ELM models to predict solubility of non-hydrocarbon gases in aqueous electrolyte systems. *Meas. J. Int. Meas. Confed.* 164, 107999. <https://doi.org/10.1016/j.measurement.2020.107999>
- Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., Prachayasittikul, V., 2009. A practical overview of quantitative structure-activity relationship. *EXCLI J.*
- Nasirpour, N., Mohammadpourfard, M., Zeinali Heris, S., 2020. Ionic liquids: Promising compounds for sustainable chemical processes and applications. *Chem. Eng. Res. Des.* <https://doi.org/10.1016/j.cherd.2020.06.006>
- Nassar, I.M., Noor El-Din, M.R., Morsi, R.E., El-Azeim, A.A., Hashem, A.I., 2016. Eco Friendly nanocomposite materials to scavenge hazard gas H₂S through fixed-bed reactor in petroleum application. *Renew. Sustain. Energy Rev.* 65, 101–112. <https://doi.org/10.1016/j.rser.2016.06.019>
- Nematpour, M., Jalili, A.H., Ghotbi, C., Rashtchian, D., 2016. Solubility of CO₂ and H₂S in the ionic liquid 1-ethyl-3-methylimidazolium trifluoromethanesulfonate. *J. Nat. Gas Sci. Eng.* 30, 583–591. <https://doi.org/10.1016/j.jngse.2016.02.006>
- Ohlin, C.A., Dyson, P.J., Laurency, G., 2004. Carbon monoxide solubility in ionic liquids: Determination, prediction and relevance to hydroformylation. *Chem. Commun.* <https://doi.org/10.1002/chin.200432032>
- Ortiz, J. V, Cioslowski, J., Fox, D.J., 2009. Gaussian 09, revision B. 01. Wallingford CT.
- P. Walden, 1914. Molecular weights and electrical conductivity of several fused salts.

Bull. Acad. Imper. Sci. 8.

- Palgunadi, J., Kang, J.E., Nguyen, D.Q., Kim, J.H., Min, B.K., Lee, S.D., Kim, H., Kim, H.S., 2009. Solubility of CO₂ in dialkylimidazolium dialkylphosphate ionic liquids. *Thermochim. Acta*. <https://doi.org/10.1016/j.tca.2009.04.022>
- Palomar, J., Ferro, V.R., Torrecilla, J.S., Rodríguez, F., 2007. Density and molar volume predictions using COSMO-RS for ionic liquids. An approach to solvent design. *Ind. Eng. Chem. Res.* 46, 6041–6048. <https://doi.org/10.1021/ie070445x>
- Park, D., Lee, D.S., Joung, J.Y., Park, J.M., 2005. Comparison of different bioreactor systems for indirect H₂S removal using iron-oxidizing bacteria. *Process Biochem.* 40(3–4), 1461–1467. <https://doi.org/10.1016/j.procbio.2004.06.034>
- Peng, D., Picchioni, F., 2020. Prediction of toxicity of Ionic Liquids based on GC-COSMO method. *J. Hazard. Mater.* 398. <https://doi.org/10.1016/j.jhazmat.2020.122964>
- Peng, Y., Tong, Z.-H., Chong, H.-J., Shao, X.-Y., 2018. Toxic effects of prolonged exposure to [C14mim]Br on *Caenorhabditis elegans*. *Chemosphere* 208, 226–232. <https://doi.org/10.1016/J.CHEMOSPHERE.2018.05.176>
- Pérez, S.A., Montalbán, M.G., Carissimi, G., Licence, P., Villora, G., 2020. In vitro cytotoxicity assessment of monocationic and dicationic pyridinium-based ionic liquids on HeLa, MCF-7, BGM and EA.hy926 cell lines. *J. Hazard. Mater.* 385. <https://doi.org/10.1016/j.jhazmat.2019.121513>
- Peric, B., Sierra, J., Martí, E., Cruañas, R., Garau, M.A., 2015. Quantitative structure-activity relationship (QSAR) prediction of (eco)toxicity of short aliphatic protic ionic liquids. *Ecotoxicol. Environ. Saf.* 115. <https://doi.org/10.1016/j.ecoenv.2015.02.027>
- Peric, B., Sierra, J., Martí, E., Cruañas, R., Garau, M.A., Arning, J., Bottin-Weber, U., Stolte, S., 2013. (Eco)toxicity and biodegradability of selected protic and aprotic ionic liquids. *J. Hazard. Mater.* 261. <https://doi.org/10.1016/j.jhazmat.2013.06.070>
- Pernak, J., Borucka, N., Walkiewicz, F., Markiewicz, B., Fochtman, P., Stolte, S., Steudte, S., Stepnowski, P., 2011. Synthesis, toxicity, biodegradability and

- physicochemical properties of 4-benzyl-4-methylmorpholinium-based ionic liquids. *Green Chem.* 13. <https://doi.org/10.1039/c1gc15468k>
- Pernak, J., Branicka, M., 2004. Synthesis and Aqueous Ozonation of Some Pyridinium Salts with Alkoxyethyl and Alkylthiomethyl Hydrophobic Groups. *Ind. Eng. Chem. Res.* 43. <https://doi.org/10.1021/ie030118z>
- Pešić, J., Watson, M., Papović, S., Vraneš, M., 2020. Ionic Liquids: Review of their Current and Future Industrial Applications and their Potential Environmental Impact. *Recent Pat. Nanotechnol.* 15. <https://doi.org/10.2174/1872210513999190923121448>
- Pinto, A.M., Rodríguez, H., Colón, Y.J., Arce, A., Soto, A., 2013. Absorption of carbon dioxide in two binary mixtures of ionic liquids. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/ie303238h>
- Pinto, P.C.A.G., Costa, A.D.F., Lima, J.L.F.C., Saraiva, M.L.M.F.S., 2011. Automated evaluation of the effect of ionic liquids on catalase activity. *Chemosphere* 82. <https://doi.org/10.1016/j.chemosphere.2010.11.046>
- Plechkova, N. V., Seddon, K.R., 2008. Applications of ionic liquids in the chemical industry. *Chem. Soc. Rev.* 37, 123–150. <https://doi.org/10.1039/b006677j>
- Politzer, P., Murray, J.S., 2002. The fundamental nature and role of the electrostatic potential in atoms and molecules. *Theor. Chem. Acc.* 108, 134–142. <https://doi.org/10.1007/s00214-002-0363-9>
- Potivichayanon, S., Pokethitiyook, P., Kruatrachue, M., 2006. Hydrogen sulfide removal by a novel fixed-film bioscrubber system. *Process Biochem.* 41, 708–715. <https://doi.org/10.1016/j.procbio.2005.09.006>
- Pretti, C., Chiappe, C., Baldetti, I., Brunini, S., Monni, G., Intorre, L., 2009. Acute toxicity of ionic liquids for three freshwater organisms: *Pseudokirchneriella subcapitata*, *Daphnia magna* and *Danio rerio*. *Ecotoxicol. Environ. Saf.* 72, 1170–1176. <https://doi.org/10.1016/j.ecoenv.2008.09.010>
- Radošević, K., Cvjetko, M., Kopjar, N., Novak, R., Dumić, J., Srček, V.G., 2013. In vitro cytotoxicity assessment of imidazolium ionic liquids: Biological effects in fish Channel Catfish Ovary (CCO) cell line. *Ecotoxicol. Environ. Saf.* 92.

<https://doi.org/10.1016/j.ecoenv.2013.03.002>

- Rahmati-Rostami, M., Ghotbi, C., Hosseini-Jenab, M., Ahmadi, A.N., Jalili, A.H., 2009. Solubility of H₂S in ionic liquids [hmim][PF₆], [hmim][BF₄], and [hmim][Tf₂N]. *J. Chem. Thermodyn.* 41, 1052–1055. <https://doi.org/10.1016/j.jct.2009.04.014>
- Ranke, J., Mölter, K., Stock, F., Bottin-Weber, U., Poczobutt, J., Hoffmann, J., Ondruschka, B., Filser, J., Jastorff, B., 2004. Biological effects of imidazolium ionic liquids with varying chain lengths in acute *Vibrio fischeri* and WST-1 cell viability assays. *Ecotoxicol. Environ. Saf.* [https://doi.org/10.1016/S0147-6513\(03\)00105-2](https://doi.org/10.1016/S0147-6513(03)00105-2)
- Ranke, J., Müller, A., Bottin-Weber, U., Stock, F., Stolte, S., Arning, J., Störmann, R., Jastorff, B., 2007a. Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity. *Ecotoxicol. Environ. Saf.* 67. <https://doi.org/10.1016/j.ecoenv.2006.08.008>
- Ranke, J., Othman, A., Fan, P., Müller, A., 2009. Explaining ionic liquid water solubility in terms of cation and anion hydrophobicity. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms10031271>
- Ranke, J., Stolte, S., Störmann, R., Arning, J., Jastorff, B., 2007b. Design of sustainable chemical products - The example of ionic liquids. *Chem. Rev.* 107, 2183–2206. <https://doi.org/10.1021/cr050942s>
- Rathi, P.C., Ludlow, R.F., Verdonk, M.L., 2019. Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network. *J. Med. Chem.* 63, 8778–8790. <https://doi.org/10.1021/acs.jmedchem.9b01129>
- Ratti, R., 2014. Ionic Liquids: Synthesis and Applications in Catalysis. *Adv. Chem.* 2014. <https://doi.org/10.1155/2014/729842>
- Rodríguez-Escontrela, I., Rodríguez-Palmeiro, I., Rodríguez, O., Arce, A., Soto, A., 2016. Characterization and phase behavior of the surfactant ionic liquid tributylmethylphosphonium dodecylsulfate for enhanced oil recovery. *Fluid Phase Equilib.* 417. <https://doi.org/10.1016/j.fluid.2016.02.021>
- Roman-Vicharra, C., Franco-Gallo, F., Alaminsky, R.J., Galvez-Aranda, D.E.,

- Balbuena, P.B., Seminario, J.M., 2018. Sigma-holes in battery materials using iso-electrostatic potential surfaces. *Crystals* 8, 1–13. <https://doi.org/10.3390/cryst8010033>
- Roosen, C., Müller, P., Greiner, L., 2008. Ionic liquids in biotechnology: Applications and perspectives for biotransformations. *Appl. Microbiol. Biotechnol.* <https://doi.org/10.1007/s00253-008-1730-9>
- Rosen, B.A., Salehi-Khojin, A., Thorson, M.R., Zhu, W., Whipple, D.T., Kenis, P.J.A., Masel, R.I., 2011. Ionic liquid-mediated selective conversion of CO₂ to CO at low overpotentials. *Science* (80-.). <https://doi.org/10.1126/science.1209786>
- Roy, K., Das, R.N., 2013. QSTR with extended topochemical atom (ETA) indices. 16. Development of predictive classification and regression models for toxicity of ionic liquids towards *Daphnia magna*. *J. Hazard. Mater.* 254–255. <https://doi.org/10.1016/j.jhazmat.2013.03.023>
- Roy, K., Kar, S., Das, R.N., 2015. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, SpringerBriefs in Molecular Science,.
- Ruokonen, S.K., Sanwald, C., Sundvik, M., Polnick, S., Vyavaharkar, K., Duša, F., Holding, A.J., King, A.W.T., Kilpeläinen, I., Lämmerhofer, M., Panula, P., Wiedmer, S.K., 2016. Effect of Ionic Liquids on Zebrafish (*Danio rerio*) Viability, Behavior, and Histology; Correlation between Toxicity and Ionic Liquid Aggregation. *Environ. Sci. Technol.* 50. <https://doi.org/10.1021/acs.est.5b06107>
- Rybińska-Fryca, A., Sosnowska, A., Puzyn, T., 2020. Representation of the structure-A key point of building QSAR/QSPR models for ionic liquids. *Materials* (Basel). <https://doi.org/10.3390/ma13112500>
- Safamirzaei, M., Modarress, H., 2012. Correlating and predicting low pressure solubility of gases in [bmim][BF₄] by neural network molecular modeling. *Thermochim. Acta* 545. <https://doi.org/10.1016/j.tca.2012.07.005>
- Safavi, M., Ghotbi, C., Taghikhani, V., Jalili, A.H., Mehdizadeh, A., 2013. Study of the solubility of CO₂, H₂S and their mixture in the ionic liquid 1-octyl-3-methylimidazolium hexafluorophosphate: Experimental and modelling. *J. Chem. Thermodyn.* 65, 220–232. <https://doi.org/10.1016/j.jct.2013.05.038>

- Sakhaeina, H., Jalili, A.H., Taghikhani, V., Safekordi, A.A., 2010a. Solubility of H₂S in ionic liquids 1-Ethyl-3-methylimidazolium Hexafluorophosphate ([emim][PF₆]) and 1-Ethyl-3-methylimidazolium Bis(trifluoromethyl)sulfonylimide ([emim][Tf₂N]). *J. Chem. Eng. Data* 55(12), 5839–5845. <https://doi.org/10.1021/je100794k>
- Sakhaeina, H., Taghikhani, V., Jalili, A.H., Mehdizadeh, A., Safekordi, A.A., 2010b. Solubility of H₂S in 1-(2-hydroxyethyl)-3-methylimidazolium ionic liquids with different anions. *Fluid Phase Equilib.* 298, 303–309. <https://doi.org/10.1016/j.fluid.2010.08.027>
- Salam, M.A., Abdullah, B., Ramli, A., Mujtaba, I.M., 2016. Structural feature based computational approach of toxicity prediction of ionic liquids: Cationic and anionic effects on ionic liquids toxicity. *J. Mol. Liq.* 224. <https://doi.org/10.1016/j.molliq.2016.09.120>
- Saljooqi, A., 2021. Introduction to Ionic Liquids and Their Environment-Friendly Applications, in: *Ionic Liquid-Based Technologies for Environmental Sustainability*. <https://doi.org/10.1016/B978-0-12-824545-3.00001-5>
- Santos, A.G., Ribeiro, B.D., Alviano, D.S., Coelho, M.A.Z., 2014. Toxicity of ionic liquids toward microorganisms interesting to the food industry. *RSC Adv.* <https://doi.org/10.1039/c4ra05295a>
- Sattar, A.M.A., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. *Neural Comput. Appl.* 31, 157–169. <https://doi.org/10.1007/s00521-017-2987-7>
- Schaffran, T., Justus, E., Elfert, M., Chen, T., Gabel, D., 2009. Toxicity of N, N, N-trialkylammoniododecaborates as new anions of ionic liquids in cellular, liposomal and enzymatic test systems. *Green Chem.* 11. <https://doi.org/10.1039/b906165g>
- Sehrawat, H., Kumar, N., Tomar, R., Kumar, L., Tomar, V., Madan, J., Dass, S.K., Chandra, R., 2020. Synthesis and characterization of novel 1,3-benzodioxole tagged noscapine based ionic liquids with in silico and in vitro cytotoxicity analysis on HeLa cells. *J. Mol. Liq.* 302.

<https://doi.org/10.1016/j.molliq.2020.112525>

- Serat, F.Z., Benkouider, A.M., Yahiaoui, A., Bagui, F., 2017. Nonlinear group contribution model for the prediction of flash points using normal boiling points. *Fluid Phase Equilib.* 449, 52–59. <https://doi.org/10.1016/j.fluid.2017.06.008>
- Shafiei, A., Ahmadi, M.A., Zaheri, S.H., Baghban, A., Amirfakhrian, A., Soleimani, R., 2014. Estimating hydrogen sulfide solubility in ionic liquids using a machine learning approach. *J. Supercrit. Fluids* 95, 525–534. <https://doi.org/10.1016/j.supflu.2014.08.011>
- Shi, W., Maginn, E.J., 2009. Molecular Simulation of Ammonia Absorption in the Ionic Liquid 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ([emim][Tf2N]). *AIChE J.* 55, 2414–2421. <https://doi.org/10.1002/aic.11910>
- Shojaeian, A., 2017. Thermodynamic modeling of solubility of hydrogen sulfide in ionic liquids using Peng Robinson-Two State equation of state. *J. Mol. Liq.* 229, 591–598. <https://doi.org/10.1016/j.molliq.2016.12.001>
- Shokouhi, M., Adibi, M., Jalili, A.H., Hosseini-Jenab, M., Mehdizadeh, A., 2010. Solubility and diffusion of H₂S and CO₂ in the ionic liquid 1-(2-Hydroxyethyl)-3-methylimidazolium tetrafluoroborate. *J. Chem. Eng. Data* 55, 1663–1668. <https://doi.org/10.1021/jc900716q>
- Siedlecka, E.M., Mroziak, W., Kaczyński, Z., Stepnowski, P., 2008. Degradation of 1-butyl-3-methylimidazolium chloride ionic liquid in a Fenton-like system. *J. Hazard. Mater.* 154. <https://doi.org/10.1016/j.jhazmat.2007.10.104>
- Siedlecka, E.M., Stepnowski, P., 2009. The effect of alkyl chain length on the degradation of alkylimidazolium- and pyridinium-type ionic liquids in a Fenton-like system. *Environ. Sci. Pollut. Res.* 16. <https://doi.org/10.1007/s11356-008-0058-4>
- Simon, R.H.M., 1956. Estimation of critical properties of organic compounds by the method of group contributions. A. L. Lyderren. Engineering Experiment Station Report 3. College of Engineering, University of Wisconsin, Madison, Wisconsin(1955). 22 pages, *AIChE Journal*. <https://doi.org/10.1002/aic.690020328>

- Son, H.J., Lee, J.H., 2005. H₂S removal with an immobilized cell hybrid reactor. *Process Biochem.* 40(6), 2197–2203. <https://doi.org/10.1016/j.procbio.2004.08.013>
- Song, Z., Shi, H., Zhang, X., Zhou, T., 2020a. Prediction of CO₂ solubility in ionic liquids using machine learning methods. *Chem. Eng. Sci.* 223, 115752. <https://doi.org/10.1016/j.ces.2020.115752>
- Song, Z., Zhang, C., Qi, Z., Zhou, T., Sundmacher, K., 2018. Computer-aided design of ionic liquids as solvents for extractive desulfurization. *AIChE J.* <https://doi.org/10.1002/aic.15994>
- Song, Z., Zhou, T., Qi, Z., Sundmacher, K., 2020b. Extending the UNIFAC model for ionic liquid–solute systems by combining experimental and computational databases. *AIChE J.* 66(2), e16821. <https://doi.org/10.1002/aic.16821>
- Song, Z., Zhou, T., Qi, Z., Sundmacher, K., 2017. Systematic Method for Screening Ionic Liquids as Extraction Solvents Exemplified by an Extractive Desulfurization Process. *ACS Sustain. Chem. Eng.* <https://doi.org/10.1021/acssuschemeng.7b00024>
- Sosnowska, A., Grzonkowska, M., Puzyn, T., 2017. Global versus local QSAR models for predicting ionic liquids toxicity against IPC-81 leukemia rat cell line: The predictive ability. *J. Mol. Liq.* 231. <https://doi.org/10.1016/j.molliq.2017.02.025>
- Stasiewicz, M., Mulkiwicz, E., Tomczak-Wandzel, R., Kumirska, J., Siedlecka, E.M., Golebiowski, M., Gajdus, J., Czerwicka, M., Stepnowski, P., 2008. Assessing toxicity and biodegradation of novel, environmentally benign ionic liquids (1-alkoxymethyl-3-hydroxypyridinium chloride, saccharinate and acesulfamates) on cellular and molecular level. *Ecotoxicol. Environ. Saf.* 71. <https://doi.org/10.1016/j.ecoenv.2007.08.011>
- Stepnowski, P., 2005. Preliminary assessment of the sorption of some alkyl imidazolium cations as used in ionic liquids to soils and sediments. *Aust. J. Chem.* 58. <https://doi.org/10.1071/CH05018>
- Stepnowski, P., Mrozik, W., Nichthauser, J., 2007. Adsorption of alkylimidazolium and alkyloxy-pyridinium ionic liquids onto natural soils. *Environ. Sci. Technol.* 41.

<https://doi.org/10.1021/es062014w>

- Stepnowski, P., Składanowski, A.C., Ludwiczak, A., Łaczyńska, E., 2004. Evaluating the cytotoxicity of ionic liquids using human cell line HeLa. *Hum. Exp. Toxicol.* 23. <https://doi.org/10.1191/0960327104ht480oa>
- Stepnowski, P., Zaleska, A., 2005. Comparison of different advanced oxidation processes for the degradation of room temperature ionic liquids. *J. Photochem. Photobiol. A Chem.* 170. <https://doi.org/10.1016/j.jphotochem.2004.07.019>
- Steutde, S., Bemowsky, S., Mahrova, M., Bottin-Weber, U., Tojo-Suarez, E., Stepnowski, P., Stolte, S., 2014. Toxicity and biodegradability of dicationic ionic liquids. *RSC Adv.* 4. <https://doi.org/10.1039/c3ra45675g>
- Steutde, S., Stepnowski, P., Cho, C.W., Thöming, J., Stolte, S., 2012. (Eco)toxicity of fluoro-organic and cyano-based ionic liquid anions. *Chem. Commun.* 48. <https://doi.org/10.1039/c2cc34955h>
- Stock, F., Hoffmann, J., Ranke, J., Störmann, R., Ondruschka, B., Jastorff, B., 2004. Effects of ionic liquids on the acetylcholinesterase - A structure-activity relationship consideration. *Green Chem.* 6. <https://doi.org/10.1039/b402348j>
- Stolte, S., Arning, J., Bottin-Weber, U., Matzke, M., Stock, F., Thiele, K., Uerdingen, M., Welz-Biermann, U., Jastorff, B., Ranke, J., 2006. Anion effects on the cytotoxicity of ionic liquids. *Green Chem.* 8. <https://doi.org/10.1039/b602161a>
- Stolte, S., Arning, J., Bottin-Weber, U., Müller, A., Pitner, W.R., Welz-Biermann, U., Jastorff, B., Ranke, J., 2007a. Effects of different head groups and functionalised side chains on the cytotoxicity of ionic liquids. *Green Chem.* 9. <https://doi.org/10.1039/b615326g>
- Stolte, S., Matzke, M., Arning, J., Bösch, A., Pitner, W.R., Welz-Biermann, U., Jastorff, B., Ranke, J., 2007b. Effects of different head groups and functionalised side chains on the aquatic toxicity of ionic liquids. *Green Chem.* 9, 1170–1179. <https://doi.org/10.1039/b711119c>
- Stolte, S., Schulz, T., Cho, C.W., Arning, J., Strassner, T., 2013. Synthesis, toxicity, and biodegradation of tunable aryl alkyl ionic liquids (TAAILs). *ACS Sustain. Chem. Eng.* 1. <https://doi.org/10.1021/sc300146t>

- Stolte, S., Steudte, S., Areitioaurtena, O., Pagano, F., Thöming, J., Stepnowski, P., Igartua, A., 2012. Ionic liquids as lubricants or lubrication additives: An ecotoxicity and biodegradability assessment. *Chemosphere* 89. <https://doi.org/10.1016/j.chemosphere.2012.05.102>
- Studzińska, S., Kowalkowski, T., Buszewski, B., 2009. Study of ionic liquid cations transport in soil. *J. Hazard. Mater.* 168. <https://doi.org/10.1016/j.jhazmat.2009.03.029>
- Sun, W., Sun, J., 2017. Prediction of carbon dioxide emissions based on principal component analysis with regularized extreme learning machine: The case of China. *Environ. Eng. Res.* 22, 302–311. <https://doi.org/10.4491/eer.2016.153>
- Sutton, M.A., Erisman, J.W., Dentener, F., Möller, D., 2008. Ammonia in the environment: From ancient times to the present. *Environ. Pollut.* <https://doi.org/10.1016/j.envpol.2008.03.013>
- Sutton, M.A., Fowler, D., 2002. Introduction: Fluxes and impacts of atmospheric ammonia on national, landscape and farm scales. *Environ. Pollut.* [https://doi.org/10.1016/S0269-7491\(01\)00145-2](https://doi.org/10.1016/S0269-7491(01)00145-2)
- Tang, S., Baker, G.A., Zhao, H., 2012. Ether- and alcohol-functionalized task-specific ionic liquids: Attractive properties and applications. *Chem. Soc. Rev.* 41. <https://doi.org/10.1039/c2cs15362a>
- Thamke, V.R., Chaudhari, A.U., Tapase, S.R., Paul, D., Kodam, K.M., 2019. In vitro toxicological evaluation of ionic liquids and development of effective bioremediation process for their removal. *Environ. Pollut.* 250. <https://doi.org/10.1016/j.envpol.2019.04.043>
- Thamke, V.R., Tapase, S.R., Kodam, K.M., 2017. Evaluation of risk assessment of new industrial pollutant, ionic liquids on environmental living systems. *Water Res.* 125. <https://doi.org/10.1016/j.watres.2017.08.046>
- The UFT/ Merck Ionic Liquids Biological Effects Database [WWW Document], 2020. URL <http://www.il-eco.uft.uni-bremen.de>
- Thuy Pham, T.P., Cho, C.W., Yun, Y.S., 2010. Environmental fate and toxicity of ionic liquids: A review. *Water Res.* <https://doi.org/10.1016/j.watres.2009.09.030>

- Tiago, G.A.O., Matias, I.A.S., Ribeiro, A.P.C., Martins, L.M.D.R.S., 2020. Application of Ionic Liquids in Electrochemistry—Recent Advances. *Molecules*. <https://doi.org/10.3390/MOLECULES25245812>
- Torrecilla, J.S., Palomar, J., García, J., Rojo, E., Rodríguez, F., 2008a. Modelling of carbon dioxide solubility in ionic liquids at sub and supercritical conditions by neural networks and mathematical regressions. *Chemom. Intell. Lab. Syst.* 93, 149–159. <https://doi.org/10.1016/j.chemolab.2008.05.004>
- Torrecilla, J.S., Rodríguez, F., Bravo, J.L., Rothenberg, G., Seddon, K.R., López-Martin, I., 2008b. Optimising an artificial neural network for predicting the melting point of ionic liquids. *Phys. Chem. Chem. Phys.* <https://doi.org/10.1039/b806367b>
- Trush, M.M., Kovalishyn, V., Ocheretniuk, A.D., Kalashnikova, L.E., Prokopenko, V.M., Holovchenko, O. V., Kobzar, O.L., Brovarets, V.S., Metelytsia, L.O., 2018. New 1,3-oxazolylphosphonium Salts as Potential Biocides: QSAR Study, Synthesis, Antibacterial Activity and Toxicity Evaluation. *Lett. Drug Des. Discov.* 15. <https://doi.org/10.2174/1570180815666180219164334>
- Tshibangu, P.N., Ndwandwe, S.N., Dikio, E.D., 2011. Density, viscosity and conductivity study of 1-Butyl-3-methylimidazolium bromide. *Int. J. Electrochem. Sci.* 6, 2.
- Valderrama, J.O., Robles, P.A., 2007. Critical properties, normal boiling temperatures, and acentric factors of fifty ionic liquids. *Ind. Eng. Chem. Res.* 46. <https://doi.org/10.1021/ie0603058>
- Van Krevelen, D.W., Chermín, H.A.G., 1951. Estimation of the free enthalpy (Gibbs free energy) of formation of organic compounds from group contributions. *Chem. Eng. Sci.* 1, 66–80. [https://doi.org/10.1016/0009-2509\(51\)85002-4](https://doi.org/10.1016/0009-2509(51)85002-4)
- Ventura, S.P.M., Marques, C.S., Rosatella, A.A., Afonso, C.A.M., Gonçalves, F., Coutinho, J.A.P., 2012. Toxicity assessment of various ionic liquid families towards *Vibrio fischeri* marine bacteria. *Ecotoxicol. Environ. Saf.* 76, 162–168. <https://doi.org/10.1016/j.ecoenv.2011.10.006>
- Vieira, N.S.M., Bastos, J.C., Rebelo, L.P.N., Matias, A., Araújo, J.M.M., Pereira, A.B.,

2019. Human cytotoxicity and octanol/water partition coefficients of fluorinated ionic liquids. *Chemosphere* 216, 576–586. <https://doi.org/10.1016/J.CHEMOSPHERE.2018.10.159>
- Voloshina, A.D., Gumerova, S.K., Sapunova, A.S., Kulik, N. V., Mirgorodskaya, A.B., Kotenko, A.A., Prokopyeva, T.M., Mikhailov, V.A., Zakharova, L.Y., Sinyashin, O.G., 2020. The structure – Activity correlation in the family of dicationic imidazolium surfactants: Antimicrobial properties and cytotoxic effect. *Biochim. Biophys. Acta - Gen. Subj.* 1864. <https://doi.org/10.1016/j.bbagen.2020.129728>
- Wang, C., Wei, Z., Wang, L., Sun, P., Wang, Z., 2015. Assessment of bromide-based ionic liquid toxicity toward aquatic organisms and QSAR analysis. *Ecotoxicol. Environ. Saf.* <https://doi.org/10.1016/j.ecoenv.2015.02.012>
- Wang, G., Zhang, S., Xu, W., Qi, W., Yan, Y., Xu, Q., 2015. Efficient saccharification by pretreatment of bagasse pith with ionic liquid and acid solutions simultaneously. *Energy Convers. Manag.* <https://doi.org/10.1016/j.enconman.2014.09.029>
- Wang, J., Luo, J., Feng, S., Li, H., Wan, Y., Zhang, X., 2016. Recent development of ionic liquid membranes. *Green Energy Environ.* 1, 43–61. <https://doi.org/10.1016/j.gee.2016.05.002>
- Wang, J., Song, Z., Cheng, H., Chen, L., Deng, L., Qi, Z., 2020. Multilevel screening of ionic liquid absorbents for simultaneous removal of CO₂ and H₂S from natural gas. *Sep. Purif. Technol.* 248, 117053. <https://doi.org/10.1016/j.seppur.2020.117053>
- Wang, J., Zeng, S., Huo, F., Shang, D., He, H., Bai, L., Zhang, X., Li, J., 2019. Metal chloride anion-based ionic liquids for efficient separation of NH₃. *J. Clean. Prod.* 206, 661–669. <https://doi.org/10.1016/j.jclepro.2018.09.192>
- Wang, K., Xu, H., Yang, C., Qiu, T., 2021. Machine learning-based ionic liquids design and process simulation for CO₂ separation from flue gas. *Green Energy Environ.* 6, 432–443. <https://doi.org/10.1016/j.gee.2020.12.019>
- Wang, W., Ma, X., Grimes, S., Cai, H., Zhang, M., 2017. Study on the absorbability, regeneration characteristics and thermal stability of ionic liquids for VOCs

- removal. *Chem. Eng. J.* 328. <https://doi.org/10.1016/j.cej.2017.06.178>
- Wang, Y., Wang, Z., Pan, J., Liu, Y., 2019. Removal of gaseous hydrogen sulfide using Fenton reagent in a spraying reactor. *Fuel* 239, 70–75. <https://doi.org/10.1016/j.fuel.2018.10.143>
- Wang, Z., Song, Z., Zhou, T., 2020. Machine Learning for Ionic Liquid Toxicity Prediction. *Processes*. <https://doi.org/10.3390/pr9010065>
- Welton, T., 2018. Ionic liquids: a brief history. *Biophys. Rev.* 10, 691–706. <https://doi.org/10.1007/s12551-018-0419-2>
- Wilkes, J.S., Zaworotko, M.J., 1992. Air and water stable 1-ethyl-3-methylimidazolium based ionic liquids. *J. Chem. Soc. Chem. Commun.* <https://doi.org/10.1039/C39920000965>
- Wu, T., Li, W., Chen, M., Zhou, Y., Zhang, Q., 2020. Estimation of Ionic Liquids Toxicity against Leukemia Rat Cell Line IPC-81 based on the Empirical-like Models using Intuitive and Explainable Fingerprint Descriptors. *Mol. Inform.* <https://doi.org/10.1002/minf.202000102>
- Wu, W., Han, B., Gao, H., Liu, Z., Jiang, T., Huang, J., 2004. Desulfurization of flue gas: SO₂ absorption by an ionic liquid. *Angew. Chemie - Int. Ed.* 43. <https://doi.org/10.1002/anie.200353437>
- Wu, Y.-W., Higgins, B., Yu, C., Reddy, A.P., Ceballos, S., Joh, L.D., Simmons, B.A., Singer, S.W., VanderGheynst, J.S., 2016. Ionic Liquids Impact the Bioenergy Feedstock-Degrading Microbiome and Transcription of Enzymes Relevant to Polysaccharide Hydrolysis. *mSystems* 1. <https://doi.org/10.1128/msystems.00120-16>
- Yan, C.Q., Wan, H., Guan, G.F., 2008. Prediction of melting points for 1,3-disubstituent imidazolium ionic liquids. *Acta Phys. - Chim. Sin.* 24, 2198–2202. <https://doi.org/10.3866/PKU.WHXB20081209>
- Yan, F., Lan, T., Yan, X., Jia, Q., Wang, Q., 2019. Norm index-based QSTR model to predict the eco-toxicity of ionic liquids towards Leukemia rat cell line. *Chemosphere* 234. <https://doi.org/10.1016/j.chemosphere.2019.06.064>
- Yan, F., Shang, Q., Xia, S., Wang, Q., Ma, P., 2015. Topological study on the toxicity

- of ionic liquids on *Vibrio fischeri* by the quantitative structure-activity relationship method. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2015.01.016>
- Yang, Z., Pan, W., 2005. Ionic liquids: Green solvents for nonaqueous biocatalysis. *Enzyme Microb. Technol.* <https://doi.org/10.1016/j.enzmictec.2005.02.014>
- Ying, H., Baltus, R.E., 2007. Experimental measurement of the solubility and diffusivity of CO₂ in room-temperature ionic liquids using a transient thin-liquid-film method. *Ind. Eng. Chem. Res.* 46(24), 8166–8175. <https://doi.org/10.1021/ie070501u>
- Yokozeki, A., Shiflett, M.B., 2010. Gas solubilities in ionic liquids using a generic van der Waals equation of state. *J. Supercrit. Fluids* 55(2), 846–851. <https://doi.org/10.1016/j.supflu.2010.09.015>
- Yokozeki, A., Shiflett, M.B., 2007a. Vapor-liquid equilibria of ammonia + ionic liquid mixtures. *Appl. Energy.* <https://doi.org/10.1016/j.apenergy.2007.02.005>
- Yokozeki, A., Shiflett, M.B., 2007b. Ammonia solubilities in room-temperature ionic liquids. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/ie061260d>
- Yu, L., Dai, W., Tang, L., 2016. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Eng. Appl. Artif. Intell.* 47, 110–121. <https://doi.org/10.1016/j.engappai.2015.04.016>
- Yu, M., Li, S.M., Li, X.Y., Zhang, B.J., Wang, J.J., 2008. Acute effects of 1-octyl-3-methylimidazolium bromide ionic liquid on the antioxidant enzyme system of mouse liver. *Ecotoxicol. Environ. Saf.* 71. <https://doi.org/10.1016/j.ecoenv.2008.02.022>
- Zandu, S.K., Chopra, H., Singh, I., 2019. Ionic Liquids for Therapeutic and Drug Delivery Applications. *Curr. Drug Res. Rev.* 12. <https://doi.org/10.2174/2589977511666191125103338>
- Zeng, S., Liu, L., Shang, D., Feng, J., Dong, H., Xu, Q., Zhang, X., Zhang, S., 2018. Efficient and reversible absorption of ammonia by cobalt ionic liquids through Lewis acid-base and cooperative hydrogen bond interactions. *Green Chem.* 20, 2075–2083. <https://doi.org/10.1039/c8gc00215k>

- Zeng, S., Zhang, Xiangping, Bai, L., Zhang, Xiaochun, Wang, H., Wang, J., Bao, D., Li, M., Liu, X., Zhang, S., 2017. Ionic-Liquid-Based CO₂ Capture Systems: Structure, Interaction and Process. *Chem. Rev.* 117, 9625–9673. <https://doi.org/10.1021/acs.chemrev.7b00072>
- Zhang, L., De Schryver, P., De Gusseme, B., De Muynck, W., Boon, N., Verstraete, W., 2008. Chemical and biological technologies for hydrogen sulfide emission control in sewer systems: A review. *Water Res.* 42, 1–12. <https://doi.org/10.1016/j.watres.2007.07.013>
- Zhang, S., Sun, N., He, X., Lu, X., Zhang, X., 2006. Physical properties of ionic liquids: Database and evaluation. *J. Phys. Chem. Ref. Data* 35, 1475–1517. <https://doi.org/10.1063/1.2204959>
- Zhang, Y., Yu, P., Luo, Y., 2013. Absorption of CO₂ by amino acid-functionalized and traditional dicationic ionic liquids: Properties, Henry's law constants and mechanisms. *Chem. Eng. J.* 214, 355–363. <https://doi.org/10.1016/j.cej.2012.10.080>
- Zhao, H., 2006. Innovative applications of ionic liquids as “green” engineering liquids. *Chem. Eng. Commun.* <https://doi.org/10.1080/00986440600586537>
- Zhao, L., Wang, Q., Ma, K., 2019. Solubility Parameter of Ionic Liquids: A Comparative Study of Inverse Gas Chromatography and Hansen Solubility Sphere. *ACS Sustain. Chem. Eng.* 7, 10544–10551. <https://doi.org/10.1021/acssuschemeng.9b01093>
- Zhao, Y., Gani, R., Afzal, R.M., Zhang, X., Zhang, S., 2017. Ionic liquids for absorption and separation of gases: An extensive database and a systematic screening method. *AIChE J.* <https://doi.org/10.1002/aic.15618>
- Zhao, Y., Gao, H., Zhang, X., Huang, Y., Bao, D., Zhang, S., 2016a. Hydrogen Sulfide Solubility in Ionic Liquids (ILs): An Extensive Database and a New ELM Model Mainly Established by Imidazolium-Based ILs. *J. Chem. Eng. Data* 61, 3970–3978. <https://doi.org/10.1021/acs.jced.6b00449>
- Zhao, Y., Gao, J., Huang, Y., Afzal, R.M., Zhang, X., Zhang, S., 2016b. Predicting H₂S solubility in ionic liquids by the quantitative structure-property relationship

- method using: S σ -profile molecular descriptors. *RSC Adv.* 6, 70405–70413. <https://doi.org/10.1039/c6ra15429h>
- Zhao, Y., Huang, Y., Zhang, X., Zhang, S., 2015a. A quantitative prediction of the viscosity of ionic liquids using S σ -profile molecular descriptors. *Phys. Chem. Chem. Phys.* <https://doi.org/10.1039/c4cp04712e>
- Zhao, Y., Zeng, S., Huang, Y., Afzal, R.M., Zhang, X., 2015b. Estimation of Heat Capacity of Ionic Liquids Using S \sim -profile Molecular Descriptors. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/acs.iecr.5b03576>
- Zhao, Y., Zhang, X., Deng, L., Zhang, S., 2016c. Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms. *Comput. Chem. Eng.* 92, 37–42. <https://doi.org/10.1016/j.compchemeng.2016.04.035>
- Zhao, Y., Zhao, J., Huang, Y., Zhou, Q., Zhang, X., Zhang, S., 2014. Toxicity of ionic liquids: Database and prediction via quantitative structure-activity relationship method. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2014.06.018>
- Zhou, H., Lv, P., Shen, Y., Wang, J., Fan, J., 2013. Identification of degradation products of ionic liquids in an ultrasound assisted zero-valent iron activated carbon micro-electrolysis system and their degradation mechanism. *Water Res.* 47. <https://doi.org/10.1016/j.watres.2013.03.057>
- Zhou, X., Cao, B., Liu, S., Sun, X., Zhu, X., Fu, H., 2016. Theoretical and experimental investigation on the capture of H₂S in a series of ionic liquids. *J. Mol. Graph. Model.* 68. <https://doi.org/10.1016/j.jmglm.2016.06.013>
- Zhu, P., Kang, X., Zhao, Y., Latif, U., Zhang, H., 2019. Predicting the toxicity of ionic liquids toward acetylcholinesterase enzymes using novel QSAR models. *Int. J. Mol. Sci.* 20(9), 2186. <https://doi.org/10.3390/ijms20092186>
- Zhu, Q.Y., Qin, A.K., Suganthan, P.N., Huang, G. Bin, 2005. Evolutionary extreme learning machine. *Pattern Recognit.* 38, 1759–1763. <https://doi.org/10.1016/j.patcog.2005.03.028>

Curriculum vitae

&

List of publications

Xuejing Kang

Date of birth: 14.07.1986

Place of birth: Neihuang, Henan Province, China

Address: Kamycka 1280, Suchdol, 16521, Prague 6, Czech Republic

E-mail: kang@fzp.czu.cz, Xuejing_kang@hotmail.com

Education

2019 – present: Ph.D. candidate
Department of Applied Ecology, Faculty of Environmental
Sciences
Czech University of Life Sciences Prague, Czech Republic

2009 – 2012: Master programme
Department of Material Sciences, School of Material and
Chemical Engineering
Zhengzhou University of Light Industry, Zhengzhou, China

2005 – 2009: Bachelor programme
Department of Environmental Engineering, School of Material
and Chemical Engineering
Zhengzhou University of Light Industry, Zhengzhou, China

Publications

Xuejing Kang, Bo Hu, Mayang Christy Perdana, Yongsheng Zhao, Zhongbing Chen. Extreme leaning machine models for predicting the *n*-octanol/water partition coefficient (K_{ow}) data of organic compounds [J]. Journal of Environmental Chemical Engineering, 2022, 108552. <https://doi.org/10.1016/j.jece.2022.108552>

Xuejing Kang, Yongsheng Zhao, Hongzhong Zhang, Zhongbing Chen. Application of atomic electrostatic potential descriptors for predicting the eco-toxicity of ionic liquids towards leukemia rat cell line [J]. Chemical Engineering Science, 2022, 260: 117941. <https://doi.org/10.1016/j.ces.2022.117941>

Xuejing Kang, Yongsheng Zhao, and Zhongbing Chen. Atom surface fragment contribution method for predicting the toxicity of ionic liquids [J]. Journal of Hazardous Materials 421 (2022): 126705. <https://doi.org/10.1016/j.jhazmat.2021.126705>

Xuejing Kang, Zuopeng Lv, Zhongbing Chen, Yongsheng Zhao. Assessing the ecotoxicity of ionic liquids on *Vibrio fischeri* using electrostatic potential descriptors[J]. Journal of Hazardous Materials, 2020, 397: 122761. <https://doi.org/10.1016/j.jhazmat.2020.122761>

Xuejing Kang, Zuopeng Lv, Zhongbing Chen*, Yongsheng Zhao*. Prediction of ammonia absorption in ionic liquids based on extreme learning machine modelling and a novel molecular descriptor SEP [J]. Environmental Research, 2020, 189: 109951. <https://doi.org/10.1016/j.envres.2020.109951>

Xuejing Kang, Zuopeng Lv, Zhongbing Chen*, Yongsheng Zhao*. A QSPR model for estimating Henry's law constant of H₂S in ionic liquids by ELM algorithm[J]. Chemosphere, 2020, 269: 128743. <https://doi.org/10.1016/j.chemosphere.2020.128743>

Peng Zhu, **Xuejing Kang**, Ullah Latif, Maoming Gong, Yongsheng Zhao. A Reliable Database for Ionic Volume and Surface: Its Application to Predict Molar Volume and Density of Ionic Liquid[J]. *Industrial & Engineering Chemistry Research*, 2019, 58(23): 10073-10083. <https://doi.org/10.1021/acs.iecr.9b00720>

Peng Zhu, **Xuejing Kang**, Yongsheng Zhao, Ullah Latif, Hongzhong Zhang et al. Predicting the toxicity of ionic liquids toward acetylcholinesterase enzymes using novel QSAR models[J]. *International journal of molecular sciences*, 2019, 20(9): 2186. <https://doi.org/10.3390/ijms20092186>

Xuejing Kang, Xinyan Liu, Jianqing Li, Yongsheng Zhao*, Hongzhong Zhang* Heat capacity prediction of ionic liquids based on quantum chemistry descriptors. [J]. *Industrial & Engineering Chemistry Research*, 2018, 57(49): 16989-16994. <https://doi.org/10.1021/acs.iecr.8b03668>

Xuejing Kang, Yongsheng Zhao*, Jinjin Li*. Predicting refractive index of ionic liquids based on the extreme learning machine (ELM) intelligence algorithm. *Journal of Molecular Liquids*, 2018, 250: 44-49. <https://doi.org/10.1016/j.molliq.2017.11.166>

Xuejing Kang, Chunjiang Liu, Shaojuan Zeng, et al. Prediction of Henry's law constant of CO₂ in ionic liquids based on S_{EP} and S_{σ-profile} molecular descriptors [J]. *Journal of Molecular Liquids*, 2018, 262: 139-147. <https://doi.org/10.1016/j.molliq.2018.04.026>

Xuejing Kang, Jing Deng, Yongsheng Zhao. Novel molecular descriptors for the prediction of H₂S solubility in ionic liquids[J], *Journal of Molecular Liquids*, 2018, 265:756-764. <https://doi.org/10.1016/j.molliq.2018.06.113>

Xuejing Kang, Zhijun Zhao, Jianguo Qian, Raja Muhammad Afzal. Prediction the viscosity of ionic liquids by the ELM intelligence algorithm. *Industrial & Engineering Chemistry Research*, 2017, 56(39): 11344-11351. <https://doi.org/10.1021/acs.iecr.7b02722>

Yongsheng Zhao, Mingguang Pan, **Xuejing Kang**, et al. Gas separation by ionic liquids: A theoretical study [J]. *Chemical Engineering Science*, 2018, 189: 43-55. <https://doi.org/10.1016/j.ces.2018.05.044>

Grants and projects

IGA 2020B0032 (Internal Grant Agency of the Faculty of Environmental Sciences, CULS Prague).

IGA 2022B0036 (Internal Grant Agency of the Faculty of Environmental Sciences, CULS Prague).

UGC 52_2021 (University Grant Competition, CULS Prague).

Participation in Conferences

- | | |
|---------|--|
| 2019.11 | Kosteleckého inspirování 2019, Prague, Czech Republic |
| 2020.11 | Kosteleckého inspirování 2020 (virtual), Prague, Czech Republic |
| 2021.3 | 12 th International Conference on Environmental Science and Development (virtual), Prague, Czech Republic |
| 2021.11 | 2021 AIChE Annual meeting (virtual), Boston, USA |

