Investigation the effects of cation/anion structure on distribution of Thiophene between IL-rich and Hydrocarbon-rich phases by QSPR approach

Zahra Moheb-Aleaba, Mohammad Reza Khosravi-Nikou*

Gas Engineering Department, Petroleum University of Technology, Ahvaz, Iran

*: Corresponding author: Dr. Mohammad Reza Khosravi-Nikou, mr.khosravi@put.ac.ir, Tel.Fax: 00986135550868

Abstract

Selecting an appropriate ionic liquid as extraction solvent for extracting sulfur compounds from fuel, is a promising way to reduce SOx emissions and to prevent environmental pollution. In this regard, the best structural descriptors of different cation and anion, which play an important role on the thiophene distribution between ionic liquid (IL) and hydrocarbon phases in the ternary systems, that is helpful for selecting a proper IL, have been investigated using quantitative structure–property relationship (QSPR) approach. Five different datasets were collected from the literature containing {((Cation/[NTf₂])–thiophene–n-heptane), (Cation/[NTf₂])–thiophene–n-hexane), (Cation/[FAP])–thiophene–n-heptane), (Cation/[TCM])–thiophene–n-heptane), ([C₂MIM]/Anion–thiophene–hydrocarbon), ([C₈MIM]/Anion–thiophene–hydrocarbon)} with 664 data points. By investigating the various kinds of molecular descriptors of 1D, 2D and 3D, it was found that the increment of Spanning Tree Number (STN) descriptor as cation structure descriptor, causes a decrease in mole fraction of thiophene in IL-rich phase due to the increase in steric hindrance. STN descriptor reflects the complexity of structure of cation and the information about steric hindrance. Also, it was found that an increase in Elp value as anion structure descriptor leads to the decrement of mole fraction of thiophene in IL-rich phase, in order
to increase the interaction between cation and anion in ionic liquid, that reduced the available space of cation to interact with thiophene. The $E_{1p}$ descriptor represents the density of atoms projected along the length or unfilled space of the anion, thus it is informative of the accessibility of them by thiophene.

**Keywords:** QSPR; Ionic liquids; Extractive desulfurization

**Introduction**

The presence of sulfur and its derivatives in transportation fuels becomes an environmental hazard because of $SO_x$ emissions. Therefore, in the last few decades, environmental concern has increased and severe standards and regulations are implemented all over the world to minimize the negative health and environmental effects from automobile exhausts. The maximum sulfur content in the transportation fuels has been limited to less than 15 ppm in many countries. Consequently, the deep desulfurization of fuels has attracted increased attention in the research community worldwide. Hydrodesulfurization (HDS) is the commercialized technology to remove sulfur compounds from fuels, but due to the lack of high performance catalyst and highly cost and energy consumptive due to high operating temperatures (300–400 $^\circ$C) as well as elevated pressures (20–100 atm for hydrogen stream) is difficult to use [1, 2]. Therefore, some alternative technologies have to be explored, which extractive desulfurization (EDS) is of practical interest considering the fact that the extraction process is a well established technology and can be operated under ambient conditions.

Recent investigations have shown that ionic liquids have a great potential for removing refractive sulfur compounds from the hydrocarbon liquid fuel. Accordingly, EDS by ionic liquid has become an area of great interest to researchers in recent years, as shown in Fig.1.
Figure 1. Number of publications about extractive desulfurization by ionic liquids published in journals (ISI web of science)

However, successful use of this technique depends on the selection of efficient solvents. Many ionic liquids can be synthesized through combination of different cations and anions, but all the studied ionic liquids for this purpose have been selected principally for their ready availability at relatively low cost and environmental impact. Compared with the very huge number of potential ionic liquids, the number of ionic liquids that were studied experimentally in desulfurization is very limited. There is still large room to more design and select the ionic liquids with higher desulfurization ability and better physiochemical properties. Since, there are numerous possible ionic liquids due to the large number of possible cation-anion combinations, selecting an appropriate one is important. Most extractive desulfurization studies conduct experimentally for selecting ionic liquids. The structural features of ionic liquid such as cation family, cation alkyl chain length, anion nature, etc., play an important role in removing sulfur compounds from fuels [1, 3-18]. Literature reviews show that increasing the size of cation/anion of ionic liquid increases the ionic liquid’s ability for sulfur removal [1, 3-6]. Also, the larger alkyl chain of cation/anion and cation symmetry cause more capacity of ionic liquid to sulfur removal [8-18]. It should be noted that an experimental testing approach is very time-consuming, and detecting of appropriate ionic liquid is even unrealistic. Consequently, reliable theoretical methods should be applied for selection of proper ionic liquids. Prediction of thermodynamic properties of ionic liquid is essential...
for theoretical screening of ionic liquids, which provide useful insights into certain thermodynamic properties of ionic liquids, but they demand computation, even for studying a single ionic liquid molecule [19, 20]. This makes them impractical for screening ionic liquids within a large set of potential candidates. On the other hand, methods using the conductor-like screening model for real solution (COSMO-RS), have been demonstrated to predict the activity coefficients and phase equilibria of systems involving ionic liquid [14, 17, 18, 21-26]. Despite many uses of this method in literature, but there is still large room for improvement the quantitative prediction accuracy of the COSMO models and the development of suitable COSMO databases [25]. For instance, Gao et al. [17], investigated the influence of different parameters on capacity of several ionic liquids for thiophene and dibenzothiophene to remove them from fuels. They concluded that the desulfurization efficiencies rigidly increase with the capacity of ionic liquids for thiophene and dibenzothiophene, which is consistent with the expected fact that higher capacity brings better desulfurization efficiency. However, by comparison with other works, it can be observed that some ionic liquids, which are reported in the Gao’s results, do not follow the same trend as reported in Gao’s paper [15, 27]. Due to the presence of the large number of experimental data on extractive desulfurization by ionic liquid systems, it is better to find a systematic screening method. Quantitative structure–property relationship (QSPR) is a robust method to relate macroscopic properties to molecular level descriptors [28]. Holbrey et al. [16], investigated extraction of sulfur compound from dodecane using ionic liquids and used QSPR analysis of solvent group contributions to extraction. Gorji et al. [29], employed QSPR methodology to propose models for the prediction of thiophene distribution between the IL- and hydrocarbon-rich phases in ternary systems containing IL, thiophene, and hydrocarbon solvent. Moreover, they investigated the effects of different anion and cation structures on the thiophene distribution between ionic liquid and hydrocarbon phases in the ternary systems using QSPR approach [30, 31].

In this work, QSPR has been used to investigate the effect of cation/anion structures on the sulfur-compound distribution coefficient between IL- and hydrocarbon-rich phases in ternary systems of extractive desulfurization.

**Methods**
**Experimental database**

In order to obtain generalized models that are able to provide predictions, the experimental domain must be large and evenly covered. Therefore, by investigating literature, the available experimental data of mole fractions of components in both IL- and hydrocarbon-rich phases in ternary systems of extractive desulfurization, were collected. To eliminate the other effective parameters on sulfur removal contents in extractive desulfurization process, the data of mixtures with the same hydrocarbon and operating condition, were selected between all collected experimental data. Consequently, five datasets were used in the present study, containing \{(Cation/[NTf$_2$]–thiophene–n-heptane), (Cation/[NTf$_2$]–thiophene–n-hexane), (Cation/[FAP]–thiophene–n-heptane), ([C$_2$MIM]/Anion–thiophene–hydrocarbon), ([C$_8$MIM]/Anion–thiophene–hydrocarbon)\} with 664 data points which introduced in Table 1.

| Dataset | Hydrocarbon | Ionic Liquid | data points | Ref. |
|---------|-------------|--------------|-------------|------|
| 1       | n-Heptane   | [PMPIP][NTf$_2$] | 12          | [32] |
|         |             | [BMPPIP][NTf$_2$] | 14          | [32] |
|         |             | [hmmpy][NTf$_2$] | 11          | [33] |
|         |             | [bzmim][NTf$_2$] | 10          | [34] |
|         |             | [C$_8$MIM][NTf$_2$] | 9           | [35] |
|         |             | [BCN$_3$Py][NTf$_2$] | 10          | [36] |
|         |             | [BCN$_4$Py][NTf$_2$] | 10          | [36] |
|         |             | [HCN$_4$Py][NTf$_2$] | 12          | [36] |
|         |             | [COC$_2$mMOR][NTf$_2$] | 14          | [37] |
|         |             | [COC$_2$mPIP][NTf$_2$] | 14          | [37] |
|         |             | [COC$_2$mPYR][NTf$_2$] | 14          | [37] |
|         |             | [C$_2$MIM][NTf$_2$] | 12          | [38] |
|   |    |                           |    |    |    |
|---|----|--------------------------|----|----|----|
| 2 | n-Hexane | [C₅MIM][NTf₂] | 17 | [39] |
|   |          | [C₂MIM][NTf₂] | 12 | [40] |
|   |          | [h₂mmpy][NTf₂] | 12 | [41] |
|   |          | [hmmpy][NTf₂] | 13 | [42] |
|   |          | [bzmim][NTf₂] | 10 | [34] |
|   |          | [C₈MIM][NTf₂] | 10 | [35] |
| 3 | n-Heptane | [BMPYR][FAP] | 17 | [43] |
|   |          | [C₂MIM][FAP] | 21 | [44] |
|   |          | [C₂OHMIM][FAP] | 23 | [44] |
|   |          | [COC₂N₁,₂][FAP] | 22 | [44] |
|   |          | [COC₂mPIP][FAP] | 14 | [45] |
|   |          | [COC₂mPYR][FAP] | 16 | [45] |
|   |          | [COC₂mMOR][FAP] | 15 | [45] |
| 4 | n-hexane | [C₂MIM][NTf₂] | 12 | [40] |
|   |          | [C₂MIM][DEP] | 11 | [46] |
|   |          | [C₂MIM][EtSO₄] | 10 | [47] |
|   |          | [C₂MIM][OAc] | 11 | [46] |
| n-heptane | [C₂MIM][FAP] | 21 | [44] |
|   |          | [C₂MIM][SCN] | 16 | [48] |
|   |          | [C₂MIM][TCM] | 29 | [49] |
|   |          | [C₂MIM][EtSO₄] | 9 | [47] |
| n-dodecane | [C₂MIM][NTf₂] | 12 | [38] |
| n-hexadecane | [C₂MIM][EtSO₄] | 14 | [47] |
| 5 | n-hexane | [C₈MIM][BF₄] | 13 | [50] |
|   |          | [C₈MIM][SCN] | 10 | [51] |
|   |          | [C₈MIM][NTf₂] | 10 | [35] |
| n-heptane | [C₈MIM][NO₃] | 10 | [52] |
|   |          | [C₈MIM][BF₄] | 10 | [53] |
| n-octane | [C₈MIM][SCN] | 13 | [51] |
|   |          | [C₈MIM][NO₃] | 10 | [52] |
| n-decane | [C₈MIM][BF₄] | 11 | [54] |
| n-dodecane | [C₈MIM][SCN] | 11 | [51] |
|   |          | [C₈MIM][NTf₂] | 14 | [55] |
|   |          | [C₈MIM][BF₄] | 12 | [53] |
Methodology

At first, before any computation of molecular descriptors, the molecular structures should be at their least energy level. Therefore, by using Chembio3D Ultra software [56], the structures of both cation and anion for each ionic liquid were found at optimum level by minimizing the energy level. The method of optimization of the Molecular Mechanics 2 (MM2) software feature of Chem3D was used. Molecular mechanics is a simulation operation which employs equations of “classical physics” enabling the computation of various bonded attributes such as bond stretching, angle bending, and torsional energy along with other non-bonded features [57]. Molecular mechanics considers the attractive and repulsive forces to control the relative positions of the nuclei of the atoms constituting a structure. The potential energy of a given molecule can be represented by the following simplified equation:

\[
E_{Total} = \sum E_{Stretching} + \sum E_{Bend} + \sum E_{vdW} + \sum E_{Coulombic} + \sum E_{Torsion}
\]  

(1)

Here, a mechanical model is hypothesized considering that spheres representing atoms are joined by mechanical springs representing covalent bonds. The energy terminologies shown in equation (1) have been formally defined in Table 2. The interaction and energy functionalities explained by classical physics are also termed as “force fields”. The steric energy for a molecule is first determined employing force fields followed by the adjustment of conformational stability leading to the minimization of the steric energy.

| Force/treatment | Equation | Brief details |
|-----------------|----------|---------------|
| Torsion         | \[ E_{Torsion} = \frac{1}{2} k_\phi [1 + \cos m(\phi + \phi_{offset})] \] | Torsional energy presents the energy required for free rotation of a sigma bond. The dihedral angle depicting the relative orientation of the... |
where $\phi_{offset}$ is the ideal torsion angle relative to a staggered conformation of two atoms and $k_\phi$ represents the energy barrier for rotation about the torsion angle $\phi$. The periodicity of rotation is denoted by $'m'$.

**Bond stretching**

$$E_{\text{Stretching}} = \frac{1}{2} k_{\text{stretch}} \times (r - r_0)^2$$

where the ideal and stretched bond lengths are, respectively, denoted by $r_0$ and $r$, and $k_{\text{stretch}}$ is a force constant giving a measure of the strength of the spring, i.e., bond.

Hooke’s law can be employed for the computation of bond stretching energy considering a covalent bond to be made up of a spring. However, Morse function containing complex mathematical terms also allows computation of bond stretching.

**Angle bending**

$$E_{\text{Bend}} = \frac{1}{2} k_\theta \times (\theta - \theta_0)^2$$

where the ideal bond angle is denoted by $\theta_0$ and $\theta$ is the bond angle in the bend position.

The ideal bending angle is the angle formed by three consecutive atoms at their minimum energy position.

Bending angle $\theta$ can be represented as follows:

**van der Waals force**

$$E_{\text{vdW}} = \varepsilon \times \left[ \left( \frac{r_{\text{min}}}{r} \right)^{12} - 2 \times \left( \frac{r_{\text{min}}}{r} \right)^6 \right]$$

Here, at minimum energy value $\varepsilon$, $r_{\text{min}}$ presents the distance between atoms $i$ and $j$, while the actual distance between the atoms is $r$

The van der Waals force of interaction can be represented by Lennard–Jones potential equation where the first term bearing power 6 $\varepsilon r_{\text{min}}^{12}$ represents forces of attraction and the term with 12th power $\varepsilon r_{\text{min}}^{12}$ denotes short-range repulsive forces involved.

**Coulombic force**

$$E_{\text{Coulombic}} = \frac{q_i \times q_j}{D \times r_{ij}}$$

where $q_i$ and $q_j$ represent the point charges on atoms $i$ and $j$, respectively, with $r_{ij}$ being the distance $r_{ij}$.

It measures the effect of charges between two points. The attractive or repulsive interaction between two atoms $i$ and $j$ separated by distance $r_{ij}$.
Afterward, the various kinds of molecular descriptors of 1D, 2D and 3D were computed using Dragon software [58]. It should be noted that some of the calculated descriptors are non-informative, since they are not relevant descriptor for the present study, or have the same values for all studied cations/anions in this study. So, by excluding inapplicable descriptors, around 1200 descriptors for cations and 750 descriptors for anions remained to applied for modeling. Distribution coefficients of thiophene, which calculated by dividing the values of mole fraction in the IL-rich phase ($Y_2$) to the values of mole fraction in the hydrocarbon-rich phase ($X_2$), provides a quantitative description of the distributing behavior of the sulfur-compound between the coexisting phases, and can be used as an appropriate parameter for investigating the ability of IL as extractant in EDS:

$$\beta = \frac{Y_2}{X_2}$$

(2)

On the ternary systems, $Y_2$ can be related to $X_2$ as the following [29, 59]:

$$Y_2 = aX_2 + b$$

(3)

where $a$, and $b$ are adjustable parameters which express slope or average distribution coefficient and intercept for each ternary systems, respectively. In this study, QSPR approach has been used to relate $Y_2$ to $X_2$ based on a linear equation, obtained by substituting the molecular descriptors for cations with $b$ at equation (3). For anions, since the collected datasets with the same hydrocarbon do not have enough variant data for calculating models, the datasets with different hydrocarbon were considered. Therefore, the effect of each hydrocarbon on the mole fraction of thiophene in the IL-rich phase ($Y_2$) should be investigated. Gorji et al. [29], used the Wiener polarity number (Pol) as the molecular characteristic of hydrocarbon for this purpose, as following:
Among the pool of molecular descriptors calculated by Dragon software, the best ones should be selected as variables along with $X_2$ to find the QSPR models for cations and anions. For this purpose, the genetic algorithm (GA) are applied. It should be mentioned that GA-MLR models by employing QSARINS software, were used in the current study [60].

After developing the QSPR models, it is essential to validate them to assess its reliability and predictivity by several statistical parameters including coefficient of determination ($R^2$), adjustable coefficient of determination ($R_{adj}^2$), root-mean-square error (RMSE), average absolute deviations (\%AAD), average absolute relative deviation (\%AARD), F-value ($F$) and standard residual ($s$). $R^2$ describes the proportion of variability in a data set that is accounted for by the statistical model and provides a measure of how well the model can predict new outcomes. Good QSPR models have $R^2$ and $R_{adj}^2$ values close to 1.0, and RMSE, \%AAD and \%AARD values close to 0 [28]. F-value indicates whether the selected model considering descriptors, has a better fitting than the model without descriptors.

Table 3. List of statistical parameters and their formulas

| Definition                        | Equations |
|----------------------------------|-----------|
| Coefficient of determination     | $R^2 = 1 - \frac{\sum(Y_{\text{exp},i} - Y_{\text{pred},i})^2}{\sum(Y_{\text{exp},i} - \bar{Y}_{\text{exp}})^2}$ |
| Adjusted coefficient of determination | $R_{adj}^2 = 1 - (1 - R^2) \left(\frac{n - 1}{n - p}\right)$ |

\(n=\) number of objects  
\(p=\) number of predictor variables
Results and Discussion

The target property, which has been examined in this work, was the correlation between $Y_2$ and $X_2$ for each ternary system involved different cations, anions and hydrocarbons for each dataset, by applying the QSPR approach.

At first, the correlation between $Y_2$ and $X_2$ in each dataset, without consideration the effects of cation/anion and hydrocarbon, has been examined. Then, the models have been developed by considering the influence of hydrocarbon on $Y_2$, by adding Pol descriptor for anion’s datasets. Afterwards, the effects of cation/anion structure have been evaluated, by adding their molecular descriptors. The developed models for each dataset are shown in Table 4.
### Table 4. QSPR models for each dataset and selected variables

| Dataset | Model                                                                 | R²  |
|---------|-----------------------------------------------------------------------|-----|
| 1       | \( Y_2 = 0.7127(\pm0.0697)X_2 + 0.1345(\pm0.0411) \) (5)              | 0.811 |
|         | \( Y_2 = 0.725(\pm0.045)X_2 - 0.1756(\pm0.03)STN + 0.4646(\pm0.0622) \) (6) | 0.9221 |
| 2       | \( Y_2 = 0.68(\pm0.1167)X_2 + 0.1552(\pm0.0783) \) (7)               | 0.7669 |
|         | \( Y_2 = 0.7808(\pm0.11)X_2 + 0.1303(\pm0.0624)Mor04v + 0.1023(\pm0.0709) \) (8) | 0.8374 |
| 3       | \( Y_2 = 0.7954(\pm0.0552)X_2 + 0.119(\pm0.0283) \) (9)               | 0.9043 |
|         | \( Y_2 = 0.7764(\pm0.0503)X_2 - 0.2918(\pm0.1249)Mor2hv + 0.0467(\pm0.0401) \) (10) | 0.9235 |
| 4       | \( Y_2 = 0.6953(\pm0.0674)X_2 + 0.095(\pm0.0376) \) (11)              | 0.7901 |
|         | \( Y_2 = 0.7409(\pm0.0497)X_2 - 0.0287(\pm0.0057)Pol + 0.2109(\pm0.0355) \) (12) | 0.8906 |
|         | \( Y_2 = 0.7408(\pm0.0411)X_2 - 0.0249(\pm0.0048)Pol - 0.2994(\pm0.0823)Elp + 0.3279(\pm0.0436) \) (13) | 0.926 |
| 5       | \( Y_2 = 0.7258(\pm0.0515)X_2 + 0.1354(\pm0.0298) \) (14)             | 0.8595 |
|         | \( Y_2 = 0.7789(\pm0.0387)X_2 - 0.0197(\pm0.0036)Pol + 0.2344(\pm0.0284) \) (15) | 0.9264 |
|         | \( Y_2 = 0.8198(\pm0.0376)X_2 - 0.0176(\pm0.0033)Pol + 1.2032(\pm0.4253)Mor28v + 0.2134(\pm0.0265) \) (16) | 0.9412 |

### Table 5. Obtained statistical parameters applying developed QSPR models for different datasets

| Dataset | No. of data | Eq. | R²  | R²adj | F    | s    | RMSE | %AAD | %AARD |
|---------|-------------|-----|-----|-------|------|------|------|------|-------|
| 1       | 142         | 5   | 0.811 | 0.809 | 411.92 | 0.117 | 0.102 | 7.794 | 23.273 |
| 2       | 6           | 0.922 | 0.92 | 562.11 | 0.075 | 0.078 | 6.516 | 29.107 |
As can be seen in Tables 4 and 5, for dataset 1, the addition of a cation descriptor (STN descriptor) to the $X_2$ variable, improved the coefficient of determination ($R^2$) and F-value (F), and also a decrement of Root Mean Square Error (RSME) and Average Absolute Deviations (AAD). Accordingly, by comparing the statistical parameters of equations (5) and (6) in Table 5, adding a cation descriptor to $X_2$ variable, improved the prediction ability of QSPR model. Therefore, equation (6) is an appropriate model to predict $Y_2$ for dataset 1. For dataset 2, the added cation descriptor (Mor04v descriptor) causes the improvement in statistical parameters except F-value, which indicates the added descriptor are not a proper one. For dataset 3, the added cation descriptor
(Mor04v descriptor) leads to no significant enhancement in statistical parameters. Thus, the $Y_2$ can be predicted without adding a cation descriptor in this dataset.

Fig. 2 shows the predicted $Y_2$ using QSPR models vs. experimental data of $Y_2$ for dataset 1. As can be seen in this figure, equation (6) can predict $Y_2$ with acceptable accuracy, while the predicted data by equation (5), are not in desirable agreement with experimental data. In Figs. 3 and 4 plots of predicted $Y_2$ vs. experimental data for datasets 2 and 3 are depicted. As can be observed in these figures, the predicted data with equations (8) and (10), are in acceptable agreement with experimental data compare to equations (7) and (9). Due to the decrease of $F$-value and no noticeable improvement of statistical parameters in equation (10) for dataset 3, the added cation descriptor in this model (Mor21v descriptor) cannot be a proper descriptor to improve the capability of model for predicting the mole fraction of thiophene in the IL-rich phase. As demonstrated in Table 5, the coefficient of determination value in equation (6) ($R^2=0.92$) is noticeably higher than its value in equation (8) ($R^2=0.84$). Consequently, the STN descriptor is a more desirable cation’s descriptor than Mor04v descriptor to enhance the predictive models. STN descriptor is a structural descriptor of topological descriptors category, used as a measure of molecular complexity; which increased with the complexity of the molecular structure [61]. Some specific algorithms have been proposed to calculate the number of spanning trees in molecular graphs. The more connectivities and branches of the structure of a molecule, the higher value of STN, so the steric hindrance increases. Accordingly, there is less available space around cation to interact with thiophene regarding the increment of steric hindrance. As a consequence, the STN descriptor was added with the negative sign in the QSPR model for dataset 1. It should be mentioned that Joule et al. [62] and Gupa et
al. [63] reported the same view regarding the effect of steric hindrance on cation-thiophene interaction.

**Figure 2.** Predicted value of $Y_2$ using a) equation (4), b) equation (5) for trainset and c) equation (4), d) equation (5) for testset versus experimental $Y_2$ for dataset 1.
Figure 3. Predicted value of $Y_2$ using a) equation (6), b) equation (7) for trainset and c) equation (6), d) equation (7) for testset versus experimental $Y_2$ for dataset 2.

Figure 4. Predicted value of $Y_2$ using a) equation (8), b) equation (9) for trainset and c) equation (8), d) equation (9) for testset versus experimental $Y_2$ for dataset 3.
For datasets 4 and 5, as observed in Tables 4 and 5, the addition of Pol descriptor which reflects the effect of variation of hydrocarbon, improved the ability of models to predict $Y_2$ as expected, and the reason why the addition Pol descriptor can boost the prediction ability of QSPR models reported in Gorji et al. study [29]. By adding hydrocarbon descriptor in equations (12) and (15), there is an increase in the $R^2$ value and a reduction of F-value for both datasets. Furthermore, the statistical parameters for these two equations, represented in Table 5, demonstrate the addition of Pol descriptor leads to improve the accuracy of models for predicting $Y_2$. In datasets 4 and 5, the third variable in equations (13) and (16), is an anion descriptor to take into account the effects of structural features of anion on the capability of models for predicting $Y_2$, which are E1p and Mor28v descriptors for equations (13) and (16), respectively, as shown in Figs. 5 and 6. As can be concluded from Table 5, adding anion descriptor leads to enhance the predictive ability of model for dataset 4, while leads to no noticeable enhancement for dataset 5. Accordingly, equation (15) is a proper predictive model for fifth dataset with $X_2$ and Pol descriptors without taking into account anion effects. However, it should be noted that this equation, is not an appropriate model, since it cannot consider the variation of anions and structural feature of them in dataset 5.

Fig. 5 depicts predicted $Y_2$ using models vs. experimental $Y_2$ for dataset 4. As can be seen, equation (13) predicts $Y_2$ values with a good agreement with experimental data, and the addition of an anion descriptor (E1p descriptor) leads to an improvement in predictive model’s ability. However, as illustrated in Fig. 6, adding an anion descriptor (Mor28v descriptor) along with $X_2$ and Pol descriptors, as mentioned previously, cannot improve the ability of model to predict $Y_2$ significantly.
Comparing equations (13) and (16), it can be concluded that equation (13) is a proper predictive model with a desirable ability to predict $Y_2$, due to the consideration of the effects of an anion descriptor (E1p descriptor). E1p is a structural descriptor of weighted holistic invariant molecular (WHIM) descriptor category, which introduced by Todeschini et al. [64]. WHIM descriptors are three-dimensional descriptors, containing information about the entire molecular structure in terms of size, shape, symmetry and atom distribution. These descriptors are calculated for a molecule within six different weighting schemes. E1p is related to the atom distribution and density around the origin and along the first principal component of the molecule (the length of the molecule), where the atoms are weighted by the atomic polarizability, and can be related to the quantity of unfilled space per projected atom and has been called density (or emptiness) [65]. Thus, it called first component accessibility directional WHIM index / weighted by atomic polarizabilities.

It is noteworthy that the ionic liquid’s potential for the extraction of thiophene from hydrocarbon depends on the intermolecular interactions between thiophene and ionic liquid [23]. Accordingly, the increment of unfilled space of anion leads to stronger cation-anion interaction in ionic liquid, which causes the reduction of the availability of cation to interact with the thiophene. Therefore, it is acceptable that E1p descriptor has a negative sign in equation (13). Briefly, the higher value of E1p descriptor reflects more unfilled space of anion and more affinity of it to interact with cation, leading the decrease of $Y_2$. Similar results were reported in previously studies [14, 18, 23].
Figure 5. Predicted value of $Y_2$ using a) equation (10), b) equation (11), c) equation (12) for trainset and d) equation (10), e) equation (11), f) equation (12) for testset versus experimental $Y_2$ for dataset 4.
Figure 6. Predicted value of $Y_2$ using a) equation (13), b) equation (14), c) equation (15) for trainset and d) equation (13), e) equation (14), f) equation (15) for testset versus experimental $Y_2$ for dataset 5.
Conclusion

In the present study, the effect of cation/anion structure of ionic liquids on the distribution coefficient of thiophene in ternary systems was investigated by applying QSPR approach. Five datasets were collected from literature, three of them including different cation with the fixed anion and hydrocarbon for each datasets, two others including different anion and hydrocarbon with the fixed cation. The results depicted significant role of cation structure on thiophene distribution coefficient. Therefore, the STN descriptor was added to the $X_2$ as cation descriptor in QSPR model, led to improve the predictive capability of model (from $R^2=0.81$ to $R^2=0.92$). The increment of STN causes a decrease in mole fraction of thiophene in IL-rich phase due to the increase in steric hindrance. Also, it was found that anion structure plays an important role for thiophene distribution coefficient along with hydrocarbon structure features. According to this, E1p descriptor as anion structure descriptor was added to the model of $X_2$ and Pol variables, and improved the QSPR model predictivity (i.e. $R^2=0.93$). In summary, E1p descriptor represents the density of atoms projected along the length or unfilled space of the anion, thus it is informative of the accessibility of them by thiophene. As a result, an increase in E1p value causes the decrement of $Y_2$, in order to stronger cation–anion interaction in ionic liquid. This work could provide a basis for studying the selection of a proper ionic liquid based on cation/anion structure descriptor in future research.

Availability of data and materials

Not applicable
Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

Zahra Moheb-Aleaba used the software, collected the data, implemented the methods, analyzed the data, and wrote this manuscript.

Mohammad Reza Khosravi-Nikou read and approved the final manuscript.

Acknowledgements

The authors would like to thank Prof. Paola Gramatica and Dr. Zhen Song for useful discussions and suggestions.

References

[1] X. Jiang, Y. Nie, C. Li, and Z. Wang, "Imidazolium-based alkylphosphate ionic liquids—a potential solvent for extractive desulfurization of fuel," Fuel, vol. 87, no. 1, pp. 79-84, 2008.
[2] K. Kędra-Krolik, F. Mutelet, J.-C. Moise, and J.-N. Jaubert, "Deep Fuels Desulfurization and Denitrogenation Using 1-Butyl-3-methylimidazolium Trifluoromethanesulfonate," Energy & Fuels, vol. 25, no. 4, pp. 1559-1565, 2011/04/21 2011.
[3] S. Zhang, Q. Zhang, and Z. C. Zhang, "Extractive Desulfurization and Denitrogenation of Fuels Using Ionic Liquids," Industrial & Engineering Chemistry Research, vol. 43, no. 2, pp. 614-622, 2004/01/01 2004.
[4] G. Liu, J. Sun, Y. Liu, W. Shen, and X. Kong, "Extraction of Sulfur-containing Compounds from Diesel by 1-Ethyl-3-methylimidazolium Diethyl Phosphate in a Lab-scale Extractive Column," Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, vol. 37, no. 1, pp. 61-67, 2014/12/13 2014.
[5] A. Bösmann, L. Datsevich, A. Jess, A. Lauter, C. Schmitz, and P. Wasserscheid, "Deep desulfurization of diesel fuel by extraction with ionic liquids," Chemical Communications, no. 23, pp. 2494-2495, 2001.
[6] S. Zhang and Z. C. Zhang, "Novel properties of ionic liquids in selective sulfur removal from fuels at room temperature," Green Chemistry, vol. 4, no. 4, pp. 376-379, 2002.
X. Wang, M. Han, H. Wan, C. Yang, and G. Guan, "Study on extraction of thiophene from model gasoline with Brønsted acidic ionic liquids," *Frontiers of Chemical Science and Engineering*, vol. 5, no. 1, pp. 107-112, 2011.

W. Jian-long, D.-S. Zhao, E.-P. Zhou, and D. Zhi, "Desulfurization of gasoline by extraction with N-alkyl-pyridinium-based ionic liquids," *Journal of fuel chemistry and technology*, vol. 35, no. 3, pp. 293-296, 2007.

Y. Mochizuki and K. Sugawara, "Removal of organic sulfur from hydrocarbon resources using ionic liquids," *Energy & Fuels*, vol. 22, no. 5, pp. 3303-3307, 2008.

O. U. Ahmed, F. S. Mjalli, T. Al-Wahaibi, Y. Al-Wahaibi, and I. M. AlNashef, "Optimum Performance of Extractive Desulfurization of Liquid Fuels Using Phosphonium and Pyrrolidinium-Based Ionic Liquids," *Industrial & Engineering Chemistry Research*, vol. 54, no. 25, pp. 6540-6550, 2015/06/16 2015.

F. Yu, C. Liu, B. Yuan, P. Xie, C. Xie, and S. Yu, "Energy-efficient extractive desulfurization of gasoline by polyether-based ionic liquids," *Fuel*, vol. 177, pp. 39-45, 2016.

P. Matheswaran, C. D. Wilfred, K. A. Kurnia, and A. Ramli, "Overview of activity coefficient of thiophene at infinite dilution in ionic liquids and their modeling using COSMO-RS," *Industrial & Engineering Chemistry Research*, vol. 55, no. 3, pp. 788-797, 2016.

T. T. Le Bui, S. Van Ho, B. T. Nguyen, and H. T. N. Uong, "Synthesis, characterization and application of some non-halogen ionic liquids as green solvents for deep desulfurization of diesel oil," *Fuel*, vol. 191, pp. 54-61, 2017.

J. J. Raj, S. Magaret, M. Pranesh, K. C. Lethesh, W. C. Devi, and M. I. A. Mutalib, "Extractive desulfurization of model fuel oil using ester functionalized imidazolium ionic liquids," *Separation and Purification Technology*, vol. 196, pp. 115-123, 2018/05/08/ 2018.

J. J. Raj, S. Magaret, M. Pranesh, K. C. Lethesh, W. C. Devi, and M. I. A. Mutalib, "Dual functionalized imidazolium ionic liquids as a green solvent for extractive desulfurization of fuel oil: Toxicology and mechanistic studies," *Journal of Cleaner Production*, vol. 213, pp. 989-998, 2019/03/10/ 2019.

J. D. Holbrey, I. López-Martin, G. Rothenberg, K. R. Seddon, G. Silvero, and X. Zheng, "Desulfurisation of oils using ionic liquids: selection of cationic and anionic components to enhance extraction efficiency," *Green Chemistry*, vol. 10, no. 1, pp. 87-92, 2008.

S. Gao, X. Chen, R. Abro, A. A. Abdelawab, S. S. Al-Deyab, and G. Yu, "Desulfurization of Fuel Oil: Conductor-like Screening Model for Real Solvents Study on Capacity of Ionic Liquids for Thiophene and Dibenzothiophene," *Industrial & Engineering Chemistry Research*, vol. 54, no. 38, pp. 9421-9430, 2015/09/30 2015.

R. Anantharaj and T. Banerjee, "COSMO-RS based predictions for the desulphurization of diesel oil using ionic liquids: Effect of cation and anion combination," *Fuel processing technology*, vol. 92, no. 1, pp. 39-52, 2011.

X. Liu, G. Zhou, X. Zhang, and S. Zhang, "Molecular dynamics simulation of desulfurization by ionic liquids," *AIChE Journal*, vol. 56, no. 11, pp. 2983-2996, 2010/01/20 2010.

R. Anantharaj and T. Banerjee, "Quantum chemical studies on the simultaneous interaction of thiophene and pyridine with ionic liquid," *AIChE Journal*, vol. 57, no. 3, pp. 749-764, 2011.

M. Diedenhofen and A. Klamt, "COSMO-RS as a tool for property prediction of IL mixtures—A review," *Fluid Phase Equilibria*, vol. 294, no. 1, pp. 31-38, 2010/07/15/ 2010.

C. D. Wilfred, Z. Man, and Z. P. Chan, "Predicting methods for sulfur removal from model oils using COSMO-RS and partition coefficient," *Chemical Engineering Science*, vol. 102, pp. 373-377, 2013.

A. R. Ferreira, M. G. Freire, J. C. Ribeiro, F. M. Lopes, J. G. Crespo, and J. A. Coutinho, "Ionic liquids for thiols desulfurization: Experimental liquid–liquid equilibrium and COSMO-RS description," *Fuel*, vol. 128, pp. 314-329, 2014.
[24] Z. Song, J. Zhang, Q. Zeng, H. Cheng, L. Chen, and Z. Qi, "Effect of cation alkyl chain length on liquid-liquid equilibria of {ionic liquids + thiophene + heptane}: COSMO-RS prediction and experimental verification," *Fluid Phase Equilibria*, vol. 425, pp. 244-251, 2016/10/15/ 2016.

[25] Z. Song, C. Zhang, Z. Qi, T. Zhou, and K. Sundmacher, "Computer-aided design of ionic liquids as solvents for extractive desulfurization," *AIChE Journal*, vol. 64, no. 3, pp. 1013-1025, 2018/03/01 2018.

[26] Z. Song *et al.*, "Effect of water on extractive desulfurization of fuel oils using ionic liquids: A COSMO-RS and experimental study," *Chinese Journal of Chemical Engineering*, vol. 25, no. 2, pp. 159-165, 2017/02/01/ 2017.

[27] H. Zhao, G. A. Baker, D. V. Wagle, S. Ravula, and Q. Zhang, "Tuning Task-Specific Ionic Liquids for the Extractive Desulfurization of Liquid Fuel," *ACS Sustainable Chemistry & Engineering*, vol. 4, no. 9, pp. 4771-4780, 2016/09/06 2016.

[28] T. Le, V. C. Epa, F. R. Burden, and D. A. Winkler, "Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties," *Chemical Reviews*, vol. 112, no. 5, pp. 2889-2919, 2012/05/09 2012.

[29] A. E. Gorji and M. A. Sobati, "Toward molecular modeling of thiophene distribution between the ionic liquid and hydrocarbon phases: Effect of hydrocarbon structure," *Journal of Molecular Liquids*, p. 110976, 2019.

[30] A. E. Gorji and M. A. Sobati, "How Anion Structures Can Affect the Thiophene Distribution between Imidazolium-Based Ionic Liquid and Hydrocarbon Phases? A Theoretical QSPR Study," *Energy & Fuels*, vol. 33, no. 9, pp. 8576-8587, 2019/09/19 2019.

[31] A. E. Gorji and M. A. Sobati, "Effect of the cation structure on the thiophene distribution between the ionic liquid with NTf2 anion and the hydrocarbon rich phases: A QSPR study," *Journal of Molecular Liquids*, vol. 313, p. 113551, 2020/09/01/ 2020.

[32] M. Wlazło, D. Ramjugernath, P. Naidoo, and U. Domańska, "Effect of the alkyl side chain of the 1-alkylpiperidinium-based ionic liquids on desulfurization of fuels," *The Journal of Chemical Thermodynamics*, vol. 72, pp. 31-36, 2014/05/01/ 2014.

[33] A. Arce, M. Francisco, and A. Soto, "Evaluation of the polysubstituted pyridinium ionic liquid [hmmpy][Ntf2] as a suitable solvent for desulfurization: Phase equilibria," *The Journal of Chemical Thermodynamics*, vol. 42, no. 6, pp. 712-718, 2010/06/01/ 2010.

[34] M. Rogošić, A. Sander, V. Kojić, and J. P. Vuković, "Liquid–liquid equilibria in the ternary and multicomponent systems involving hydrocarbons, thiophene or pyridine and ionic liquid (1-benzyl-3-metylimidazolium bis(trifluoromethylsulfonyl)imide)," *Fluid Phase Equilibria*, vol. 412, pp. 39-50, 2016/03/25/ 2016.

[35] L. Alonso, A. Arce, M. Francisco, and A. Soto, "Phase behaviour of 1-methyl-3-octylimidazolium bis[trifluoromethylsulfonyl]imide with thiophene and aliphatic hydrocarbons: The influence of n-alkane chain length," *Fluid Phase Equilibria*, vol. 263, no. 2, pp. 176-181, 2008/02/15/ 2008.

[36] U. Domańska, K. Walczak, and M. Zawadzki, "Separation of sulfur compounds from alkanes with 1-alkylcyanopyridinium-based ionic liquids," *The Journal of Chemical Thermodynamics*, vol. 69, pp. 27-35, 2014/02/01/ 2014.

[37] A. Marciniak, M. Wlazło, and J. Gawkowska, "Ternary (liquid+liquid) equilibria of {bis(trifluoromethylsulfonyl)-amide based ionic liquids+butan-1-ol+water}," *The Journal of Chemical Thermodynamics*, vol. 94, pp. 96-100, 2016/03/01/ 2016.

[38] H. Rodriguez, M. Francisco, A. Soto, and A. Arce, "Liquid–liquid equilibrium and interfacial tension of the ternary system heptane+thiophene+1-ethyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide," *Fluid Phase Equilibria*, vol. 298, no. 2, pp. 240-245, 2010/11/25/ 2010.
M. Rogošić, A. Sander, and M. Pantaler, "Application of 1-pentyl-3-methylimidazolium bis(trifluoromethylsulfonyl) imide for desulfurization, denitrification and dearomatization of FCC gasoline," *The Journal of Chemical Thermodynamics*, vol. 76, pp. 1-15, 2014/09/01/ 2014.

B. Rodríguez-Cabo, A. Soto, and A. Arce, "Desulfurization of fuel-oils with [C2mim][NTf2]: A comparative study," *The Journal of Chemical Thermodynamics*, vol. 57, pp. 248-255, 2013/02/01/.

B. Rodríguez-Cabo, M. Francisco, A. Soto, and A. Arce, "Hexyl dimethylpyridinium ionic liquids for desulfurization of fuels. Effect of the position of the alkyl side chains," *Fluid Phase Equilibria*, vol. 314, pp. 107-112, 2012/01/25/ 2012.

M. Francisco, A. Arce, and A. Soto, "Ionic liquids on desulfurization of fuel oils," *Fluid Phase Equilib.,* vol. 294, no. 1, p. 39, 2010.

U. Domańska, E. V. Lukoshko, and M. Królikowski, "Separation of thiophene from heptane with ionic liquids," *The Journal of Chemical Thermodynamics*, vol. 61, pp. 126-131, 2013/06/01/ 2013.

K. Kędra-Królik, M. Fabrice, and J.-N. Jaubert, "Extraction of thiophene or pyridine from n-heptane using ionic liquids. Gasoline and diesel desulfurization," *Industrial & Engineering Chemistry Research*, vol. 50, no. 4, pp. 2296-2306, 2011.

A. Marciniak and M. Włazło, "Ternary (liquid+liquid) equilibria of trifluorotris(perfluoroethyl)phosphate based ionic liquids+thiophene+heptane): Part 2," *The Journal of Chemical Thermodynamics*, vol. 86, pp. 196-201, 2015/07/01/ 2015.

B. Rodríguez-Cabo, A. Arce, and A. Soto, "Desulfurization of fuels by liquid–liquid extraction with 1-ethyl-3-methylimidazolium ionic liquids," *Fluid Phase Equilibria*, vol. 356, pp. 126-135, 2013/10/25/ 2013.

L. Alonso, A. Arce, M. Francisco, O. Rodríguez, and A. Soto, "Liquid–Liquid Equilibria for Systems Composed by 1-Methyl-3-octylimidazolium Tetrafluoroborate Ionic Liquid, Thiophene, and n-Hexane or Cyclohexane," *Journal of Chemical & Engineering Data*, vol. 52, no. 5, pp. 1729-1732, 2007/09/01 2007.

L. Alonso, A. Arce, M. Francisco, and A. Soto, "Solvent extraction of thiophene from n-alkanes (C7, C12, and C16) using the ionic liquid [C8mim][BF4]," *The Journal of Chemical Thermodynamics*, vol. 40, no. 6, pp. 966-972, 2008/06/01/ 2008.

L. Alonso, A. Arce, M. Francisco, O. Rodríguez, and A. Soto, "Gasoline desulfurization using extraction with [C8mim][BF4] ionic liquid," *AIChE Journal*, vol. 53, no. 12, pp. 3108-3115, 2007.
[55] L. Alonso, A. Arce, M. Francisco, and A. Soto, "(Liquid+liquid) equilibria of [C8mim][NTf2] ionic liquid with a sulfur-component and hydrocarbons," *The Journal of Chemical Thermodynamics*, vol. 40, no. 2, pp. 265-270, 2008/02/01/ 2008.

[56] U. ChemBio3D, "CambridgeSoft Corporation," 14.0.0.117 ed: a subsidiary of PerkinElmer, 1998-2014.

[57] K. Roy, S. Kar, and R. N. Das, *A primer on QSAR/QSPR modeling: fundamental concepts*. Springer, 2015.

[58] DRAGON, "Software for Molecular Descriptor Calculations," 5.5 ed. http://www.talete.mi.it: Talete srl, 2007.

[59] O. U. Ahmed, F. S. Mjalli, M. K. Hadj-Kali, T. Al-Wahaibi, and Y. Al-Wahaibi, "Measurements and prediction of ternary liquid–liquid equilibria for mixtures of IL + sulfur compound + hexadecane," *Fluid Phase Equilibria*, vol. 421, pp. 16-23, 2016/08/15/ 2016.

[60] P. Gramatica, N. Chirico, E. Papa, S. Kovarich, and S. Cassani, "QSARINS: A New Software for the Development, Analysis, and Validation of QSAR MLR Models," *Journal of Computational Chemistry, Software news and updates*, vol. 34, pp. 2121-2132, 2013.

[61] N. Trinajstic, D. Babic, S. Nikolic, D. Plavsic, D. Amic, and Z. Mihalic, "The Laplacian matrix in chemistry," *Journal of Chemical Information and Computer Sciences*, vol. 34, no. 2, pp. 368-376, 1994/03/01 1994.

[62] J. A. Joule and K. Mills, *Heterocyclic chemistry*. John Wiley & Sons, 2008.

[63] R. R. Gupta, M. Kumar, and V. Gupta, *Heterocyclic Chemistry: Volume II: Five-Membered Heterocycles*. Springer Science & Business Media, 2013.

[64] R. Todeschini, P. Gramatica, R. Provenzani, and E. Marengo, "Weighted holistic invariant molecular descriptors. Part 2. Theory development and applications on modeling physicochemical properties of polyaromatic hydrocarbons," *Chemometrics and Intelligent Laboratory Systems*, vol. 27, no. 2, pp. 221-229, 1995/02/01/ 1995.

[65] R. Todeschini and P. Gramatica, "SD-modelling and Prediction by WHIM Descriptors. Part 5. Theory Development and Chemical Meaning of WHIM Descriptors," *Quantitative Structure-Activity Relationships*, vol. 16, no. 2, pp. 113-119, 1997.