Korean J. Chem. Eng., 37(1), 72-79 (2020)
DOI: 10.1007/s11814-019-0408-6

pISSN: 0256-1115
eISSN: 1975-7220

Molecular investigation of amine performance in the carbon capture process: Least squares support vector machine approach

Bijan Rezaei, Siavash Riahi†, and Ali Ebrahimpoor Gorji

Institute of Petroleum Engineering, Faculty of Chemical Engineering, College of Engineering, University of Tehran, Tehran, Iran

(*Received 17 June 2019 • accepted 15 October 2019*)

Abstract−The growing threat of global warming has raised more attention towards carbon capture. Current amine plants used for carbon removal suffer from great costs inflicted by high energy demand of the solvent regeneration step. Recently, looking for amines with proper performance in reduced temperatures has been the subject of many researches. Clearly, conducting these researches without any criterion and based only on trial and error wastes large amounts of money and time; thus, it is highly needed that the effect of different amine structural parameters be studied on the amine's cyclic capacity. Quantitative structure property relationship (QSPR) provides an effective method for predicting amines capacity for CO_2 absorption. In this work, density functional theory (DFT) was employed for optimization of the molecular geometries, and linear and nonlinear models based on parameters related to the molecular structure are presented. The value of the square of the correlation coefficient (R^2) for the MLR and SVM models are 0.894 and 0.973, respectively. Developed models can be used as a criterion for amine selection. Reliability and high predictability of the models are confirmed based on statistical tests. Moreover, mechanistic interpretation of models for better understanding of the reaction mechanism of carbon capture was discussed.

Keywords: Carbon Capture, Amine, Cyclic Capacity, QSPR, LS-SVM

INTRODUCTION

Despite recent efforts to mitigate global warming, atmospheric carbon dioxide is still increasing [1]. Considering the 42 percent share of heat and electricity sector in global carbon dioxide emissions [2], carbon capture from flue gas of power plants is the most viable option to address this problem. Cyclic chemical absorption/desorption using amine based solutions is a conventional process for carbon capture [3].

Despite the long history of the conventional process, it has some drawbacks of which high energy consumption is the main. The high energy demand of the amine regeneration step in the conventional process inflicts large costs on industry. The regeneration step consumes up to 15 percent of the power plant output [4], which can be 70 percent of the total operational costs of the process. Corrosion and solvent loss due to volatility and degradation are other problems of the conventional carbon capture process [5]. Reducing the temperature of the regeneration step not only decreases the required energy for regeneration, but also reduces corrosion and solvent loss. For this purpose, a number of approaches including process design/modification [6-10] and using novel solvents [11-13] have been suggested. For example, Jiang et al. [6] performed a sensitivity study to optimize the parameters of the process. They proved that the Piperazine-promoted NH_3 process enhanced the CO_2 capture significantly and suggested that it can be used in commercial applications [6]. Oh et al. [7] focused on the minimizing the total efficiency penalty and used commercial simulator Aspen Plus in developing the process models. In this work, the total energy has been reduced by 20-40%. Regarding the new amine solvents, absorption performance of the aromatic amines molecules was used by Puxty et al. [11], which had not received remarkable consideration in comparison to cyclic and noncyclic aliphatic amine molecules [11].

There are three methods to CO_2 capture: pre-combustion, oxy-combustion, and post-combustion [14]. Post-combustion capture (PCC) is the most conventional method for removing the CO_2 after the burning of the fossil fuels [14]. This method can be carried out using different amines. So, the search for new amines by means of solvent screening or solvent synthesis is costly and time consuming. A theoretical model for the prediction of amine performance can give us a criterion for primary evaluation of amines before synthesis or performing any other experiments. Since amine performance in carbon capture changes with an alteration in the molecular structure [15-18], modeling of amine cyclic capacity based on structural characteristics sounds rational. To describe the mathematical relationship between structural characteristics and cyclic capacity, quantitative structure-property relationship (QSPR) is a proper method for this purpose. In QSPR studies the most relevant characteristics of molecular structure can be identified among the huge number of characteristics. QSPR is based on the basic assumption that difference in compound properties arises from difference in molecular structures [19,20]. In this modeling method, a mathematical expression of structural features of molecules called descriptors is used as variables [21]. Recently, Momeni and Riahi developed QSPR mod-

†To whom correspondence should be addressed.
E-mail: riahi@ut.ac.ir

els for rich and lean CO_2 loadings of linear amines. The focus of their study was on linear amines [22,23]. Rezaei and Riahi used the QSPR method to model rich and lean CO_2 loadings of cyclic and linear amines in carbon capture using membranes [24].

In spite of the importance of the cyclic capacity in the evaluation of the solvent performance, a proper QSPR model for cyclic capacity has not yet been developed. In this paper, the QSPR method was employed to develop valid models for prediction and description of the cyclic capacity of linear and cyclic amines in the carbon capture process with reduced temperature in the regeneration step. Both linear and nonlinear models were developed and compared. Least squares support vector machine [25,26] was applied for nonlinear model development.

To the best of the authors' knowledge, the largest and widest varieties of amine structures are employed in this study. In addition to the prediction capability of the developed model, a mechanistic interpretation of the model variables provides insight into desirable structural characteristics of amines for carbon capture.

MATERIAL AND METHODS

1. Data Set

The cyclic capacity of amines in the carbon capture process was calculated using the experimental data reported in the paper published by Singh and Versteeg, performing the regeneration step in 80 °C (40 °C lower than conventional temperature) in hope of finding high performance amines in low temperatures [18]. The reported data set consists of 35 different structures, including cyclic and linear amines.

2. Descriptor Calculation and Variable Selection

First, molecular geometries were optimized by employing the semi-empirical optimization method RM1. A final optimization was performed using the density functional theory (DFT) at the level of B3LYP and 6-311++G (d,p) basis set [27]. These calculations were conducted by Gaussian software [28]. Then, geometrically optimized 3D structures of amines were fed to the Dragon software, which is a commercial software that calculates a large set of molecular descriptors [29]. As a result, 4885 descriptors for each molecule were calculated. Since this large number of data could make further calculations slow and imprecise, a portion of descriptors were excluded. The elimination procedure consists of three steps: (i) elimination of descriptors with constant values or low information (Shannon entropy [30] lower than 0.5), (ii) exclusion of one of the highly correlated descriptors (correlation coefficient more than 0.95), which has lower variance, and (iii) elimination of descriptors which have low normalized mutual information with cyclic capacity (lower than 0.1); mutual information is a measure of dependence between two variables [26]. After this procedure, a subset of 352 descriptors was selected for each molecule.

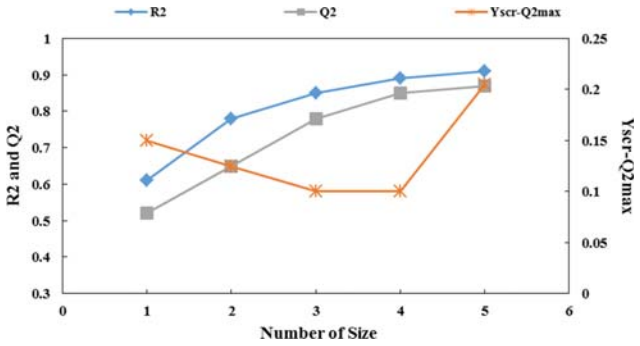

Fig. 1. Model performance vs. number of employed descriptors.

Genetic algorithm (GA) and reshaped sequential selection (RSS) [24] were used for variable selection based on Q^2 (the square correlation coefficients of cross validation). RSS consumed less time to converge and gave approximately the same and slightly better results in comparison to GA. The performance of the best linear models in respect to the size of the model is given in Fig. 1, where it is clearly observed that employing more than four variables did not make any significant improvement in the statistical quality of the linear models. Therefore, using more than four variables only adds to the complexity of the model. Moreover, y-scrambling test was used to evaluate the risk of chance correlation [31,32]. As shown in Fig. 1, the model with four variables has the lowest value of maximum Q^2 for scrambled Y vectors. This indicates that using more than four variables in the model increases the risk of chance correlation.

VE2_Dz(Z), RDF050m, E1s and ALOGP2 are variables of the best linear model based on two criteria of low number of variables and high statistical performance. VE2_Dz(Z) represents the average coefficient of the last eigenvector from Barysz distance matrix [33,34] which is a 2D matrix-based descriptor. RDF050m is one of the radial distribution function (RDF) descriptors and it is weighted by atomic mass. E1s belongs to WHIM descriptors and it describes first-component accessibility directional WHIM index weighted by intrinsic state [35]. ALOGP2 represents the squared octanol-water partition coefficient, which is calculated based on the Ghose-Crippen method [36]. As presented in Table 1, the selected descriptors are not highly correlated; therefore, the correlation between descriptors cannot negatively affect the regression coefficients and the stability of model.

RESULTS AND DISCUSSION

In spite of the high reliability of Q^2, one cannot thoroughly assess

Table 1. Correlation between selected descriptors

	VE2_Dz(Z)	RDF050m	E1s	ALOGP2
VE2_Dz(Z)	1	−0.66487	−0.31697	−0.37394
RDF050m	−0.66487	1	0.64821	0.66139
E1s	−0.31697	0.64821	1	0.30069
ALOGP2	−0.37394	0.66139	0.30069	1

the predictability of the model only based on this statistic [37]. To assure the model's predictability performance, using the molecules which were not employed in the model development is necessary. Therefore, the data set should be split into two groups of training and test sets. The training set is used for model development and the test set is employed for the model validation. Correct categorization of the data set is crucial and random splitting can negatively affect the quality of validation and regression model. Both training and test sets should be composed of structurally various molecules and should represent the whole descriptor space occupied by the data set. Moreover, the training set should contain at least one compound that is structurally similar to every compound in the test set [37]. Principal component analysis (PCA) and K-means clustering were used for this purpose. It is suggested that test set should contain at least five molecules [37]. In this study, seven compounds (20 percent of the whole data set) were chosen as test set molecules. The best linear model is given by the following equation:

$$\text{Cyclic Capacity} = 2.269 - 2.551 \times \text{VE2\_Dz(Z)} + 0.095 \times \text{RDF050m} - 1.704 \times \text{E1s} + 0.161 \times \text{ALOGP2} \quad (1)$$

As mentioned, the training set was used for regression. Statistical parameters of the linear model for the training set, test set and the whole data set (overall) are presented in Table 2. These results indicate the reliability and high predictability power of the linear model. These statistical parameters include the coefficient of determination (R^2), adjusted R^2, root mean square error (RMSE), Fisher function (F), the slopes of regression lines forced through zero (k, k′) and standard error of the estimate (s).

LS-SVM technique was applied to the variables of the best linear model in order to develop a nonlinear model. Regularization parameter (γ) and RBF kernel parameter ($\sigma 2$) were tuned by means of a grid search in the range of 0-500 for both of the parameters. Parameters of LS-SVM model (γ and $\sigma 2$) are equal to 22.4163 and 1.6558. Considering Table 2 and Table 3, the nonlinear model has a higher R^2, which shows better agreement with the experimental values. k and k′ of both linear and nonlinear models have values in the proper range suggested by golbarikh [37]. Lower RMSE and s of the nonlinear model indicate lower error in comparison with the linear model. Therefore, one can conclude that the nonlinear model is more precise.

In view of the proximity of points to the bisector of the first quadrant in Fig. 2, one can conclude that the predicted values are in good agreement with the experimental values for both models. However, lower distance of points from the bisector for the non-

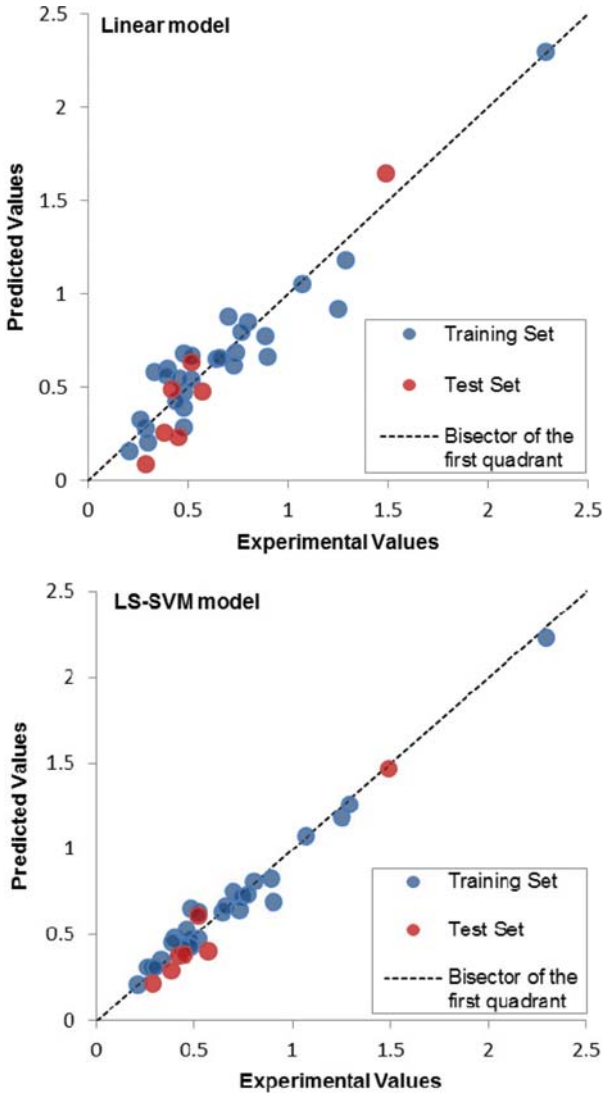

Fig. 2. Experimental vs. predicted cyclic capacity values (mol CO_2/mol amine) for linear and LS-SVM models.

Table 2. Statistical parameters of the linear model

Data set	Number of compounds	R^2	R^2adj	F	k	k′	RMSE	s
Training	28	0.895	0.877	49	1	0.97	0.135	0.149
Test	7	0.847	0.820	61.3	0.94	1.02	0.148	0.163
Overall	35	0.894	0.880	63.6	1	0.97	0.133	0.144

Table 3. Statistical parameters of the nonlinear model

Data set	Number of compounds	R^2	R^2adj	F	K	k′	RMSE	s
Training	28	0.972	0.967	180.2	1.01	0.98	0.07	0.077
Test	7	0.945	0.936	117	1.04	0.95	0.088	0.097
Overall	35	0.973	0.97	253.9	1.01	0.98	0.067	0.072

linear model indicates a higher accuracy than the linear model.

Applicability domain of models was investigated using Williams plot. Leverage value and standardized residual compose the axes of Williams plot. Leverage provides a distance measure of the compounds from the center of the chemical space formed by the data set compounds. Critical leverage is calculated as follows:

$$h^{*}=3(p+1)/n \quad (2)$$

where p and n represent the number of variables and the number of training compounds, respectively. Training compounds with leverages more than the critical value and standardized residuals outside the range of [−3,3] are considered outliers and should be excluded [38,39]. Training molecules with high leverage values and low standardized residuals are not outlier and have a positive influence on the model [38,39]. Williams plots for linear and nonlinear models are plotted in Fig. 3. It can be seen that only the leverage of one training point in each model is more than the critical value. Since standardized residuals for both points are lower than the cut-off value, none of them can be considered as an outlier. Both points belong to Tetraethylenepentamine and the high leverage value is caused by a large number of amine functional groups in its structure compared to some other amines which are drawn in Fig. 4. The number of amine functional groups (nN) has a significant impact on the value of important descriptors (including variables of the model). The relationship between the number of amine groups

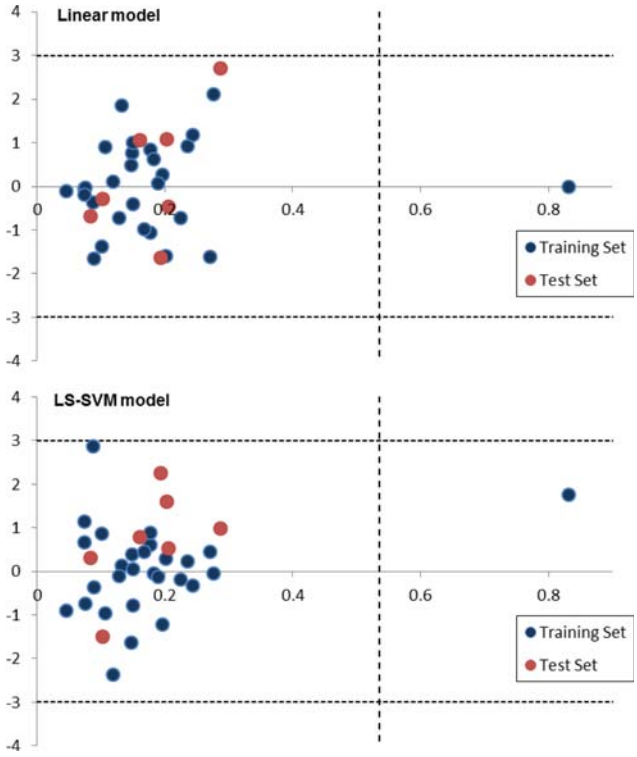

Fig. 3. William's plots of models.

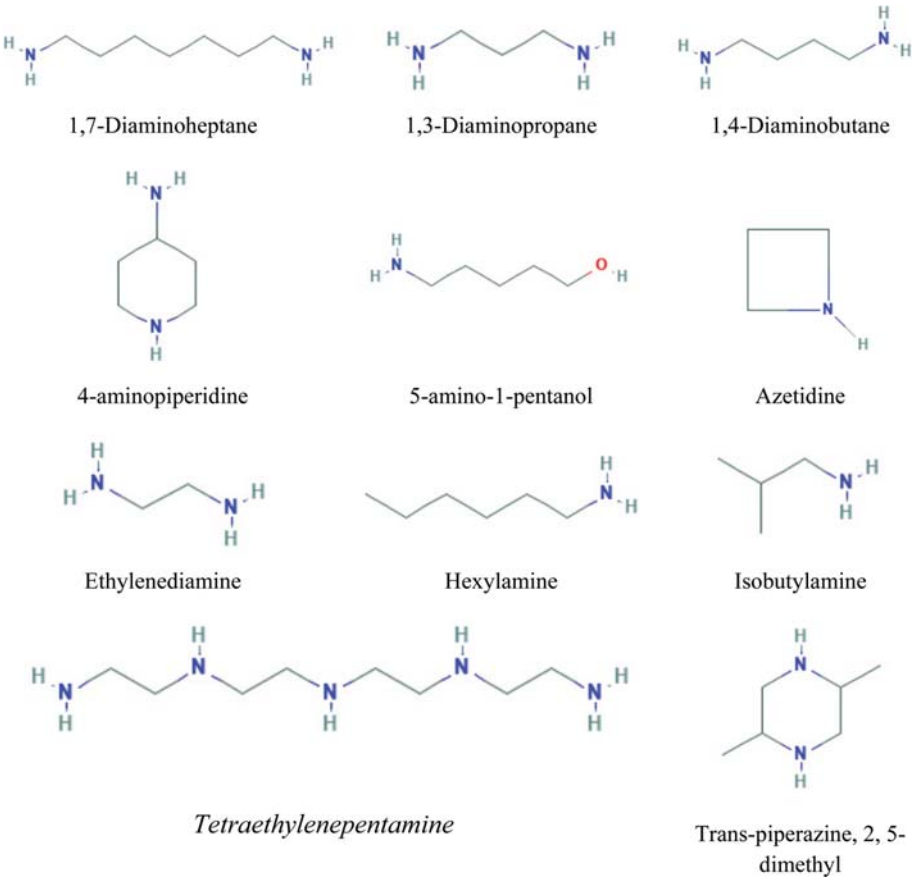

Fig. 4. Molecular structures of different amines.

Table 4. Experimental and calculated values of amines' cyclic capacities

No	Name	Experimental	Linear model	LS-SVM model	AE1	AE2
1	1,7-Diaminoheptane	0.4	0.599	0.484	0.199	0.084
2	1,3-Diamino propane	0.33	0.58	0.357	0.25	0.027
3	1,4-Diamino butane	0.48	0.471	0.452	0.009	0.028
4	1,4-Diazabicyclo [2.2.2] octane	0.48	0.685	0.652	0.205	0.172
5	1-Amino-2-propanol	0.52	0.67	0.637	0.15	0.117
6	2-(1-Piperazinyl)ethanol	0.52	0.545	0.478	0.025	0.042
7	2-(1-Piperazinyl)ethylamine	1.25	0.922	1.187	0.328	0.063
8	2-Amino-1-butanol	0.48	0.287	0.471	0.193	0.009
9	2-Methylpiperidine	0.64	0.65	0.636	0.01	0.004
10	4-Aminopiperidine	0.7	0.878	0.756	0.178	0.056
11	5-Amino-1-pentanol	0.29	0.281	0.312	0.009	0.022
12	Azetidine	0.89	0.773	0.828	0.117	0.062
13	Butylamine	0.48	0.389	0.431	0.091	0.049
14	Diethylenetriamine	1.07	1.053	1.078	0.017	0.008
15	Ethylamine	0.21	0.156	0.213	0.054	0.003
16	Ethylenediamine	0.44	0.428	0.424	0.012	0.016
17	Hexylamine	1.29	1.18	1.261	0.11	0.029
18	Isobutylamine	0.26	0.329	0.315	0.069	0.055
19	Monoethanolamine	0.3	0.203	0.313	0.097	0.013
20	N-(2-Hydroxyethyl)ethylenediamine	0.66	0.661	0.663	0.001	0.003
21	N-ethylpiperazine	0.74	0.688	0.72	0.052	0.02
22	N,N′-Bis(2-hydroxyethyl)ethylenediamine	0.8	0.849	0.809	0.049	0.009
23	N-Pentylamine	0.39	0.567	0.457	0.177	0.067
24	Piperazine	0.9	0.662	0.688	0.238	0.212
25	Piperidine	0.73	0.618	0.645	0.112	0.085
26	sec-Butylamine	0.46	0.549	0.53	0.089	0.07
27	Tetraethylenepentamine	2.29	2.297	2.234	0.007	0.056
28	trans-Piperazine, 2,5-dimethyl	0.77	0.799	0.738	0.029	0.032
29	**1,2-Diamino propane**	0.42	0.488	0.38	0.068	0.04
30	**1-Methyl piperazine**	0.52	0.634	0.616	0.114	0.096
31	**3-Amino-1-propanol**	0.38	0.257	0.291	0.123	0.089
32	**4-Amino-1-butanol**	0.45	0.231	0.39	0.219	0.06
33	**Hexamethylenediamine**	0.57	0.479	0.406	0.091	0.164
34	**Propylamine**	0.29	0.091	0.22	0.199	0.07
35	**Triethylenetetramine**	1.49	1.645	1.466	0.155	0.024

All of the numerical values are on the basis of (mol CO_2/mol amine). Bold molecules at end of the table are test set molecules

and the descriptors of the model is discussed in the model interpretation section.

MODEL INTERPRETATION

Mechanistic interpretation of the selected descriptors gives a better insight into the mechanism of reaction between carbon dioxide and amines.VE2_Dz(Z) is calculated by the following equation:

$$\mathrm{VE2\_Dz(Z)} = \frac{\sum_{i=1}^{A} l_{iA}}{\mathrm{A}} \quad (3)$$

where A is the number of non-hydrogen atoms and values of l_{iA} are equal to the coefficients of the eigenvector associated with the last (largest negative) eigenvalue of the Barysz distance matrix [35]. Higher values of l_{iA} correspond to atoms of lower degree (connected to lower non-hydrogen atoms). Atoms of molecules with lower number of branches have lower degrees and therefore higher l_{iA} and VE2_Dz(Z) values [35]. However, the smallest values of VE2_Dz(Z) in the data set correspond to the molecules without branch and with large chain length (largest path in the molecular structure). This is due to the large value of A in the denominator of VE2_Dz(Z) formula. VE2_Dz(Z) can be simply thought of as a measure of branching and chain length. It is noteworthy that the chain length has more influence on VE2_Dz(Z) value than branching.

Steric hindrance, due to the negative influence on carbamate stability, has a positive effect on the regeneration yield [40]. Considering the direct relationship between the level of branching and

steric hindrance, VE2_Dz(Z) accounts for steric hindrance. Moreover, Singh et al. reported a positive effect of increasing chain length on the desorption yield due to the reduction in carbamate stability [18]. Thus, both aspects of VE2_Dz(Z) are in agreement with the reaction mechanism.

Directional WHIM descriptors are obtained based on statistical indices which are calculated on the projections of the atoms along the first principal axis. The projections of the atoms are performed by calculation of the eigenvalues and eigenvectors of a weighted covariance matrix of the centered Cartesian coordinates of a molecule [35]. In case of E1s, the intrinsic state of atoms is employed as a weighting parameter. The intrinsic state of an atom can be considered as the ratio of π and lone-pair electrons over the number of σ bonds in the molecular structure of the atom [35]. Therefore, the intrinsic state of amine's atoms reflects the ability of the considered atom for interaction with CO_2. Amine and alcohol functional groups have high intrinsic state values.

E1s is one of the directional WHIM density descriptors. These descriptors are related to the quantity of unfilled space per projected atom. The greater the values of these descriptors, the more the projected unfilled space [35]; thus, these descriptors consider accessibility to atoms. Since E1s use the intrinsic state as a weighting parameter, it can be related to the accessibility to atoms with higher intrinsic state values. Therefore, E1s represent accessibility to amine and alcohol functional groups [41]. This descriptor has a negative effect on the cyclic capacity because of the reverse relationship between the accessibility and steric hindrance.

RDF descriptors can be defined as the probability distribution of finding an atom in a spherical volume of radius R. The radial distribution function is calculated by:

$$RDFRw = f \cdot \sum_{i=1}^{nAT-1} \sum_{j=i+1}^{nAT} w_i \cdot w_j \cdot e^{-\beta(R-r_{ij})^2} \quad (4)$$

where f is a scaling factor (assumed equal to one in the calculations), w indicates the value of the weighting parameter for the atoms i and j, rij is the interatomic distance and nAT represents the number of atoms in the molecule. β in the exponential term is the smoothing parameter (A^{o-2}) which can be thought as a temperature factor that defines the movement of atoms [35]. In the Dragon software, a β value equal to 100 is used. RDF050m considers the effect of steric hindrance around nitrogen by determination of the probability distribution of finding an atom in a spherical volume of radius five A^o around the amine functional group. In RDF050m, atomic mass is employed as a weighting parameter. Doing so would increase the influence of heavier atoms like oxygen on the descriptor value. Since heavier or bulkier atoms like oxygen increase the steric hindrance around amine functional group, enhanced influence of these atoms in RDF050m is in good agreement with the reaction mechanism. Considering present molecules in the data set, five A^o is approximately equal to the distance between the amine functional group and substituents at γ carbons[1]. Therefore, along with previously confirmed effect of substituents at α and β carbons on carbamate stability, the influence of substituents at γ carbons should also be taken into account.

[1]This estimation is based on geometries optimized by Gaussian software

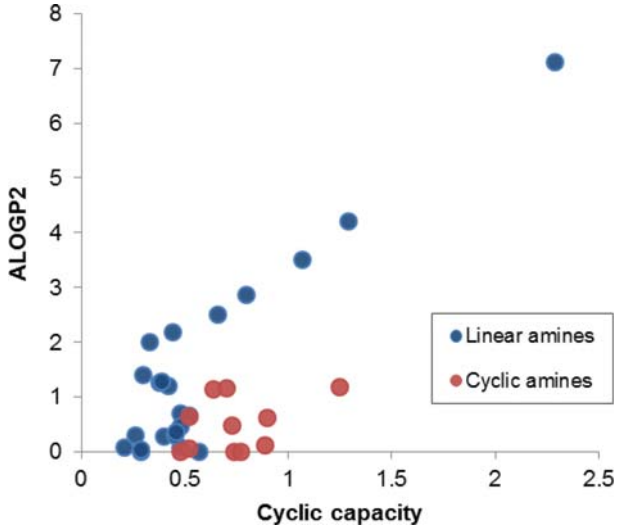

Fig. 5. ALOGP2 vs. cyclic capacity of amines.

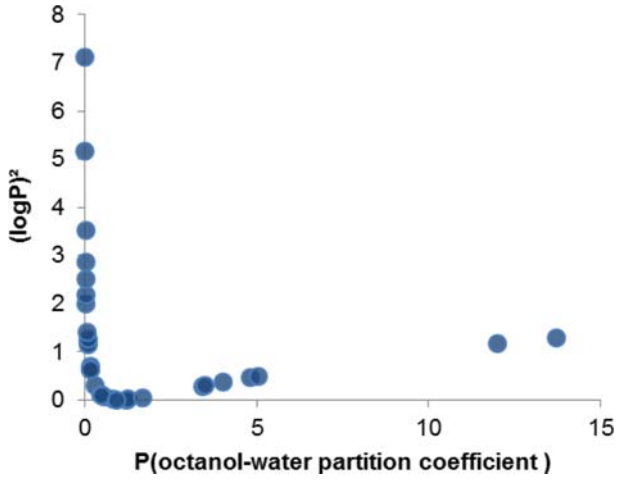

Fig. 6. Squared logP vs. P.

Octanol-water partition coefficient is a measure of molecular hydrophobicity which equals the ratio of the concentrations of the un-ionized solute in octanol to its concentration in water [35]. The value of this coefficient is usually expressed as the logarithm of the mentioned ratio. As can be seen in Fig. 5, linear amines with ALOGP2 (squared octanol-water partition coefficient) values more than one have a high correlation with cyclic capacity. As shown in Fig. 6, highly polar amines (points with P value lower than 0.1) have ALOGP2 more than one. CO_2, despite its symmetric structure, is slightly polar (P=0.713, ALOGP=−0.1468) and has a small negative charge on its oxygen ends. This is the reason that explains the physical solubility of CO_2 in polar solvents like water or amines. The polarity of amines has a direct relationship with the negative charge of certain atoms of amine (such as nitrogen). In addition, the negative charge of the atom has a direct relationship with the tendency of the atom for a reaction with the positively charged atom of carbon dioxide (carbon). Therefore, the absorption power of amine increases with an increase in polarity. Note that P and ALOGP

of carbon dioxide were also calculated by the Dragon software for the purpose of providing a consistent frame for comparison.

All of the highly polar amines in the data set have multiple amine functional groups and polarity has a direct relationship with the number of amine functional groups. Considering the reported values for rich and lean loadings of the data set amines by Singh and Versteeg, an increase in the number of amine functional groups has a greater influence on rich loading (absorption) than lean loading (desorption) [18]. This indicates that the positive effect of polarity on cyclic capacity is due to its positive influence on absorption power.

After understanding the relationship between descriptors and mechanism of reaction, the reason of high leverage value of Tetraethylenepentamine can be comprehended. The larger the nN in a linear molecule, the larger the chain length in the molecule. Therefore, due to the reverse relationship between VE2_Dz(Z) and chain length, Tetraethylenepentamine has smallest VE2_Dz(Z) in the data set. In addition, based on the value of ALOGP, Tetraethylenepentamine is the most polar molecule in the data set. For that reason, the largest ALOGP2 in the data set belongs to this molecule, which is at least more than 35 percent larger than ALOGP2 of other compounds. As a result, Tetraethylenepentamine is the furthest molecule from the center of the chemical space.

CONCLUSION

A linear model based on the QSPR method was developed. The LS-SVM technique was applied on the descriptors of the linear model to obtain a nonlinear model with a higher predictability. Both linear and nonlinear models were evaluated by various statistical tests and validated with a test set composed of seven molecules. Results showed proper statistical performance and high predictability. It is noteworthy that the results of nonlinear approach were more promising.

Employed descriptors were interpreted based on the reaction mechanism to extract important structural characteristics. It was revealed that chain length, branching, and steric hindrance around the amine functional group are key factors with positive influence on the cyclic capacity. Furthermore, the steric hindrance effect caused by substituents at γ carbons is significant, and influence of farther substituents from the amine functional group can be neglected. However, in case of linear amines with large chain length, the possibility of hydrogen bond formation between hydrogen bond donors and acceptors in two opposed parts of the molecule should be investigated. This phenomenon results in creating a ring shaped structure which reduces the distance between atoms in two opposed parts of the linear structure [17]. In this case, applying the γ carbons criterion based on the linear conformation is incorrect and would cause considerable error.

Results of this work can help researchers to find better amines for carbon capture through solvent synthesis or screening. However, other important properties of new amines like corrosivity should be taken into account too.

REFERENCES

1. Weekly average atmospheric CO_2 by the Mauna Loa Observatory. Available: http://www.esrl.noaa.gov/gmd/ccgg/trends/weekly.html.
2. M. Van der Hoeven, CO_2 emissions from fuel combustion-highlights, IEA Statistics (2014).
3. G. Puxty, R. Rowland, A. Allport, Q. Yang, M. Bown, R. Burns, M. Maeder and M. Attalla, *Environ. Sci. Technol.*, **43**, 6427 (2009).
4. Z. Wang, M. Fang, Y. Pan, S. Yan and Z. Luo, *Chem. Eng. Sci.*, **93**, 238 (2013).
5. F. Karadas, M. Atilhan and S. Aparicio, *Energy Fuels*, **24**, 5817 (2010).
6. K. Jiang, K. Li, H. Yu and P. H. M. Feron, *Chem. Eng. J.*, **347**, 334 (2018).
7. S. Y. Oh, S. Yun and J. K. Kim, *Appl. Energy*, **216**, 311 (2018).
8. B. Zhao, F. Liu, Z. Cui, C. Liu, H. Yue, S. Tang, Y. Liu, H. Lu and B. Liang, *Appl. Energy*, **185**, 362 (2017).
9. A. Cousins, L. T. Wardhaugh and P. H. M. Feron, *Int. J. Greenhouse Gas Control*, **5**, 605 (2011).
10. A. García-Abuín, D. Gómez-Díaz and J. M. Navaza, *Fuel*, **135**, 191 (2014).
11. G. Puxty, W. Conway, Q. Yang, R. Bennett, D. Fernandes, P. Pearson, D. Maher and P. Feron, *Int. J. Greenhouse Gas Control*, **83**, 11 (2019).
12. S. Murai, M. Daigo, Y. Kato, Y. Maesawa, T. Muramatsu and S. Saito, *Energy Procedia*, **63**, 1933 (2014).
13. J. Zhang, R. Misch, Y. Tan and D. W. Agar, *Chem. Eng. Technol.*, **34**, 1481 (2011).
14. Z. Zhang, Y. Li, W. Zhang, J. Wang, M. R. Soltanian and A. G. Olabi, *Renew. Sust. Energy Rev.*, **98**, 179 (2018).
15. A. K. Chakraborty, G. Astarita and K. B. Bischoff, *Chem. Eng. Sci.*, **41**, 997 (1986).
16. G. Sartori and D. W. Savage, *Ind. Eng. Chem. Fundam.*, **22**, 239 (1983).
17. P. Singh, J. P. Niederer and G. F. Versteeg, *Int. J. Greenhouse Gas Control*, **1**, 5 (2007).
18. P. Singh and G. F. Versteeg, *Process Saf. Environ. Prot.*, **86**, 347 (2008).
19. D. Ghaslani, Z. E. Gorji, A. E Gorji and S. Riahi, *Chem. Eng. Res. Des.*, **120**, 15 (2017).
20. A. E. Gorji, Z. E. Gorji and S. Riahi, *Korean J. Chem. Eng.*, **34**, 1405 (2017).
21. W. M. Berhanu, G. G. Pillai, A. A. Oliferenko and A. R. Katritzky, *ChemPlusChem*, **77**, 507 (2012).
22. M. Momeni and S. Riahi, *J. Nat. Gas Sci. Eng.*, **21**, 442 (2014).
23. M. Momeni and S. Riahi, *Int. J. Greenhouse Gas Control*, **42**, 157 (2015).
24. B. Rezaei and S. Riahi, *J. Nat. Gas Sci. Eng.*, **33**, 388 (2016).
25. J. A. K. Suykens, Least squares support vector machines, World Scientific (2002).
26. I. Mehraein and S. Riahi, *J. Mol. Liq.*, **225**, 521 (2017).
27. C. J. Cramer and F. M. Bickelhaupt, *Angew. Chem. Int. Ed.*, **42**, 381 (2003).
28. M. J. Frisch, A. B. Nielsen and A. Frisch, Gaussian 98: Gaussian Incorporated (1998).
29. R. Todeschini, V. Consonni, A. Mauri and M. Pavan, DRAGON version 6, Talete srl, Milan, Italy (2011).
30. R. M. Gray, Entropy and information theory, Springer Science & Business Media (2011).
31. J. G. Topliss and R. J. Costello, *J. Med. Chem.*, **15**, 1066 (1972).

32. J. G. Topliss and R. P. Edwards, *J. Med. Chem.*, **22**, 1238 (1979).
33. M. Barysz, G. Jashari, R. S. Lall, V. K. Srivastava and N. Trinajstic, *Stud. Phys. Theor. Chem.*, **28**, 222 (1983).
34. A. T. Balaban, D. Ciubotariu and M. Medeleanu, *J. Chem. Inf. Comput. Sci.*, **31**, 517 (1991).
35. R. Todeschini and V. Consonni, Molecular Descriptors for Chemoinformatics, John Wiley & Sons, 41 (2009).
36. A. K. Ghose and G. M. Crippen, *J. Comput. Chem.*, **7**, 565 (1986).
37. A. Golbraikh and A. Tropsha, *Mol. Diver.*, **5**, 231 (2000).
38. A. Tropsha, P. Gramatica and V. Gombar, *QSAR Comb. Sci.*, **22**, 69 (2003).
39. J. Jaworska, N. Nikolova-Jeliazkova and T. Aldenberg, *ATLA-NOTTINGHAM*, **33**, 445 (2005).
40. S. Gangarapu, A. T. Marcelis and H. Zuilhof, *ChemPhysChem*, **14**, 3936 (2013).
41. E. F. Da Silva and H. F. Svendsen, *Int. J. Greenhouse Gas Control*, **1**, 151 (2007).