J. DRUG DEL. SCI. TECH., 20 (6) 467-476 2010

# A simple QSPR model to predict aqueous solubility of drugs

**A. Shayanfar[1], M.A.A. Fakhree[2], A. Jouyban[3]***

[1]Pharmaceutical Analysis Laboratory, Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz 51664, Iran
[2]Liver and Gastrointestinal Research Center, Tabriz University of Medical Sciences, Tabriz 51664, Iran
[3]Pharmaceutical and Food Control Dept., Faculty of Pharmacy, Tabriz University of Medical of Sciences, Tabriz 51664, Iran
*Correspondence: ajouyban@hotmail.com

*Aqueous solubility of a drug/drug candidate is essential data in drug discovery, and an in silico method for predicting the aqueous solubility of drug candidates provides a valuable tool to speed up the process of drug discovery and development. This paper describes a simple quantitative structure property relationship (QSPR) model for predicting the aqueous solubility of drugs which is validated by cross-validation methods. A data set of 220 drug or drug like molecules as a train set was employed and the accuracy of the proposed QSPR model was compared with those of the general solubility equation (GSE) and the linear solvation energy relationship (LSER). Also, a test set containing the aqueous solubility of 75 official drugs which are structurally and physico-chemically diverse, was proposed to compare the accuracy of the aqueous solubility prediction models as a reference data set. The developed model is: $logS_w$ = -1.120E - 0.599ClogP, in which $S_w$ is the molar aqueous solubility of a drug, E is the excess molar refraction and ClogP is the computed logarithm of partition coefficient of drug. The E and ClogP values for a drug candidate could be computed using Pharma-Algorithms software. Average absolute error (AAE) and mean percentage deviation (MPD) were used as comparison criteria. The proposed QSPR provided better AAE and MPD for solubility prediction in comparison with GSE and LSER models.*

*Key words: Aqueous Solubility – Prediction – QSPR – Drug.*

Aqueous solubility of a drug/drug candidate is essential data in drug discovery because poor soluble compounds have low absorption and failed to proceed in drug discovery processes. In addition, aqueous solubility investigations are important for oral or parenteral drug liquid formulations. Also the solubility in water shows its importance where it is necessary to make a solution to test pharmacological or toxicological activities of the compounds. Solubility data is required for determination of absorption, distribution, metabolism and excretion (ADME) and is an important topic in drug extraction and analysis [1, 2].

Determination of solubility values is time-consuming, costly and affected by various factors, e.g. the aqueous solubility of ionizable compounds measured at pH 7 or in unbuffered water, however total solubility is equal to intrinsic solubility when pH < pKa for acids and when pH > pKa for bases [2, 3]. The other factors which affect the experimental solubility measurement are solute purity, equilibration time, temperature, and laboratory technique [2]. As an alternative, a number of quantitative structure property relationships (QSPRs) have been reported for solubility prediction of chemical compounds in water. Due to economic and humanitarian pressures, these models can improve the efficiency of drug discovery but solubility is a difficult physicochemical parameter to be predicted [4]. One reason for this problem is the lack of high quality dataset for designing and comparing the solubility prediction models [2]. Another problem is the applicability of the model for solubility prediction by using the accessible calculated parameters which make it easier to use for the researchers of the pharmaceutical and chemical sciences.

The solubility prediction models could be classified in two categories; models using 1D or 2D descriptors and models using 3D descriptors. The models using 1D or 2D descriptors are simple and their calculations are straightforward. The most accurate and famous models are: i) general solubility equation (GSE) of Yalkowsky and ii) the linear solvation energy relationship (LSER) of Abraham. The GSE consists of two parameters; melting point (*mp*) and logarithm of partition coefficient (*logP*) and is expressed as:

$$\log S_w = 0.5 - 0.01(mp - 25) - \log P \qquad \textbf{Eq. 1}$$

where $S_w$ is the molar aqueous solubility of a drug at 25 ºC. If the solute has a melting point of less than 25 ºC, the (*mp* - 25) term is set to zero [5]. The two parameters, *logP* and *mp* are good representatives of the effects of hydrophobicity and crystal packing on the solubility of a certain solute [5]. A possible disadvantage is the melting point as an experimental parameter which may not be available for some of the compounds in the early stages of drug discovery. Also, drugs with high melting points which decompose before melting are not suitable to be predicted by the GSE model [6, 7]. The *logP* can be measured using experimental methods such as HPLC [6], and/or calculated by some computational methods [8, 9], then applied to solubility prediction. The *logP* is the important parameter for solubility prediction where some models proposed are based on *logP*. For the first time, Hansch *et al*. proposed a linear relationship between aqueous solubility and *logP*. But for complex molecules addition of other parameters are required. A number of models based on *logP* could be found in the literature [2, 3, 6, 10].

The LSER is another model developed by Abraham that composed of five properties of the solute [11]:

$$\log S_w = 0.318 - 1.004E + 0.77S + 2.168A + 4.238B - 3.362A{\cdot}B - 3.987V \qquad \textbf{Eq. 2}$$

and was updated as [12]:

$$\log S_w = 0.395 - 0.955E + 0.320S + 1.155A + 3.255B - 0.785A{\cdot}B - 3.330V \qquad \textbf{Eq. 3}$$

in which *E* is the excess molar refraction of the compound, *S* is the dipolarity/polarizability, *A* and *B* are hydrogen bond acidity and basicity, respectively [13]. These last three parameters (*S, A* and *B*) are determined from solubility data of a compound in water and different organic solvents, the *A·B* term is representative of hydrogen-bond interactions between acidic and basic functional groups of the drug in its pure solid or liquid [11] and *V* is one percent of the McGowan volume and can be calculated simply using a group contribution method [13]. In this work, an updated version of LSER (*Equation 3*) was used for comparing with other models.

The *E* descriptor is defined as:

$$E = MRx - aV + b \quad \textbf{Eq. 4}$$

in which *MRx* is molar refraction and the units of *E* and *MRx* are ($cm^3$ $mol^{-1}$)/10, *a* and *b* are the model constant. *MRx* is calculated by [13]:

$$MRx = 10[(\eta^2 - 1)/(\eta^2 + 2)]V \quad \textbf{Eq. 5}$$

where η is the refractive index of the compound as a pure liquid at 20 °C.

Unfortunately, experimental Abraham's solute parameters are not available for some drugs but the parameters can be calculated by Pharma-Algorithms software [14], thus with calculated parameters, the LSER model can be considered as a computational model.

The GSE and LSER approaches are the golden and famous models for predicting aqueous solubility of chemical compounds, but prediction error of these models are relatively high for solubility prediction of drug and drug like molecules [1].

The models based on 3D descriptors were reviewed by Balakin *et al*. [15]. These descriptors have some problems such as difficulties in calculation and optimization of molecular geometry which can influence the capability of the prediction methods [15]. Also there are few examples of 3D methods which show good prediction in comparison with other descriptors [6]. Recently, Duchowicz *et al*. proposed a new QSPR model for the prediction of aqueous solubility using three DRAGON descriptors. One of the descriptors is a radial distribution function-6.0/unweighted (a 3D descriptor) [16].

The number of articles dealing with the aqueous solubility prediction methods has increased in recent years, revealing the importance of the subject in the pharmaceutical area [5]. Most of the presented methods were reviewed by Taskinen and Norinder [3]. In these articles, different data sets have been reported as test sets and as a consequence the accuracy of the models could not be directly compared. To cover this, structurally and physico-chemically diverse drugs as a test set are required. For the first time, a data set is developed by Yalkowsky consisting of 21 chemical and pharmaceutical compounds and is used by other researchers to compare the accuracies of aqueous solubility prediction models [6]. These 21 compounds are not more pharmaceutically interested compounds and also are not structurally and physico-chemically diverse.

The aims of this work are to propose a simple QSPR for predicting the aqueous solubility of drugs and also to provide a diverse test set for comparing the accuracies of the aqueous solubility prediction methods.

## I. EXPERIMENTAL

The experimental aqueous solubilities of 220 drugs and/or drug like molecules at 25 °C were collected from the literature as a train data set (data points dealing with the intrinsic solubility were excluded) [2, 16-20]. The selected solutes for the train set including drug and/or drug like molecules were also extracted from recently published works. Melting points and experimental *logP* (*ElogP*) values of the studied solutes were taken from ChemIDplus (National Library of Medicine) [21], and calculated Abraham solvation parameters and calculated *logP* (*ClogP*) from Pharma-Algorithms software [14]. We selected Abraham solvation parameters, *ClogP* and melting points for developing a new QSPR model by stepwise multiple linear regression (MLR). To validate the proposed model and in order to assess its prediction capability, cross validation methods were used. The data is sorted by aqueous solubility and divided into two, four and ten groups. Each group was excluded from the training process and the excluded data was considered as a validation set. The predictive squared correlation coefficient ($Q^2$) calculated by *Equations 6 to 8* were considered to deal with the validity of the model [22]:

$$Q^2_{F1} = 1 - \frac{\sum_{i=1}^{n\ Validation\ set} (y_i' - y_i)^2}{\sum_{i=1}^{n\ Validation\ set} (y_i - \bar{y}_{Train\ set})^2} \quad \textbf{Eq. 6}$$

$$Q^2_{F2} = 1 - \frac{\sum_{i=1}^{n\ Validation\ set} (y_i' - y_i)^2}{\sum_{i=1}^{n\ Validation\ set} (y_i - \bar{y}_{Validation\ set})^2} \quad \textbf{Eq. 7}$$

$$Q^2_{F3} = 1 - \frac{\left[\sum_{i=1}^{n\ Validation\ set} (y_i' - y_i)^2\right] / n_{Validation\ set}}{\left[\sum_{i=1}^{n\ Train\ set} (y_i - \bar{y}_{Train\ set})\right] / n_{Train\ set}} \quad \textbf{Eq. 8}$$

In these equations, $y_i$ and $y_i'$ are experimental and predicted values respectively, *n* is number of the compounds in the validation set and $\bar{y}$ indicate the means of the training and validation sets.

To compare the accuracy of aqueous solubility prediction using GSE, LSER and the proposed models, a data set composed of 75 structurally and physico-chemically diverse drugs as a test set is proposed. Normal distribution of aqueous solubility data and solute's parameters were checked by Kolmogorov-Smirov and Shapiro-Wilk tests. The range of the experimental solubility and physicochemical properties (e.g. *logP, mp*) are as wide as possible and approximately equal numbers of acidic, basic and neutral drugs were employed.

An external validation method employing the proposed test set was used to check the validity of QSPR model considering the following criteria taken from the literature [23, 24]:

1) $R^2 > 0.6$ and $Q^2 > 0.5$, where $R^2$ is the squared correlation coefficient between the predicted ($y_i''$) and experimental ($y_i$) solubilities of compounds in the test set and $Q^2$ is defined as:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n\ Test\ set} (y_i - y_i'')^2}{\sum_{i=1}^{n\ Test\ set} (y_i - \bar{y}_{Train\ set})^2} \quad \textbf{Eq. 9}$$

in which $n_{Test\ set}$ is the number of compounds in the test set and $\bar{y}_{Train\ set}$ indicates the mean of the train set;

2) $[(R^2 - R_0^2)/R^2]$ or $[(R^2 - R_0'^2)/R^2] < 0.1$, where $R^2$ from a test set should be close to $R_0^2$ or $R_0'^2$ ($R_0^2$ is the squared correlation coefficient of the predicted versus experimental values using with intercept analysis and $R_0'^2$ is that of a no intercept regression analysis);

3) k (slope of the regression line of the predicted versus experimental solubilities using intercept regression) or k' (slope of the regression line of the predicted versus experimental solubilities using with no intercept regression) value should be between 0.85 to 1.15 [23, 24].

Accuracy of the solubility prediction by the proposed model in train and test sets was compared with those of GSE and LSER by performing a paired t-test. Abraham *et al*. used the experimental Abraham's solvation parameters but these parameters have not been reported for some compounds therefore, we used the Abraham's solvation parameters computed by Pharma-Algorithms software. Different *logP* values are available for a given solute and the impact of the experimental *logP* and calculated *logP* by Pharma-Algorithms [14], KowWin [25], ACD/Labs [26], and *alogPs* [27] values on the prediction capability of the proposed model and GSE for test set was also studied.

The accuracy of the predicted solubilities is calculated by average absolute error (AAE) and mean percentage deviation (MPD) criteria, defined as:

$$AAE = \frac{\sum \left| \log S_w^{Calculated} - \log S_w^{Observed} \right|}{N} \quad \textbf{Eq. 10}$$

$$MPD = \frac{100}{N} \sum \frac{\left| S_w^{Calculated} - S_w^{Observed} \right|}{S_w^{Observed}} \quad \textbf{Eq. 11}$$

in which N is the number of data points. All analyses were performed by Excel 2003 software.

## II. RESULTS AND DISCUSSION

The log$S_w$ of 220 drugs (details listed in *Table I*) are regressed against seven investigated descriptors (Abraham solvation parameters, *MP* and *ClogP*), *E* and *ClogP* were selected by stepwise MLR method to develop a two variable model for aqueous solubility prediction. The two parameters made a significant ($p < 0.0005$) contribution to the model as:

$$\log S_w = -1.120E - 0.599 \text{Clog}P \quad \textbf{Eq. 12}$$
$$N = 220,\ R^2 = 0.934,\ s = 0.893,\ F = 1538$$

N is the number of drugs in the training set, $R^2$ is the squared correlation coefficient, s is the standard error of estimate and F is F-value (Fisher variance ratio). The inverse relation between *ClogP* and aqueous solubility has been reported earlier and shows the relation between solubility and lipophilicity of the solute [2]. The *E* parameter or excess molar refraction which could be calculated by Pharma-Algorithms software or according to *Equation 4* is composed of two parameters: the *V* and *MRx* that are indicators of the aqueous solubility of a molecule because these parameters can be calculated by atomic fragmental and the number of bonds in the molecule [13, 28]. A mechanistic interpretation is required for a valid QSPR model [29]. In this work, the employed descriptors have physicochemical interpretations as explained above, that are in agreement with the solubility mechanism and shows the validity of model. The *ClogP* and *E* values are scattered and there is no cross correlation between the employed independent variables as shown in *Figure 1*.

*Tables I and II* listed the details of train and test sets. The overall errors in correlation and prediction analyses for the investigated models for train and test sets are shown in *Tables III and IV*, respectively. Graphs of the estimates from three methods for train and test sets are shown in *Figures 2 and 3*. The proposed model shows better correlation between experimental and predicted aqueous solubility and AAE and MPD values of the proposed model are less than those of the GSE and LSER models. The overall deviations of these models have high standard deviations, because of some outliers of the predicted aqueous solubility data. Possible reasons for these outliers in predictions could be any impurity of drugs, polymorphism, any systematic errors in solubility experiments, inadequate equilibration time and temperature variations [1]. As an example, different polymorphs possess different solubilities [1], however, this is not considered in the independent variables of the GSE, LSER and the proposed models. Computational weakness of the model is another reason for the predicted outliers. To further investigate on the models' accuracies, the 10 % highest and lowest MPDs or AAEs from train and test sets for each model are excluded and the overall deviations are shown in *Tables III and IV*. In this case, the proposed model provided better results when compared with the AAE and MPD of GSE and LSER models.

The *logP* is an experimental parameter but it can be calculated by some software. In this work, the impact of the various *logP* values on the prediction capability of the GSE and the proposed model was studied. The correlations between calculated *ClogPs* and *ElogP* for test set are investigated and shown in *Figure 4*. The aqueous solubilities are predicted by GSE and the proposed model using *ElogP* and *Clogp*

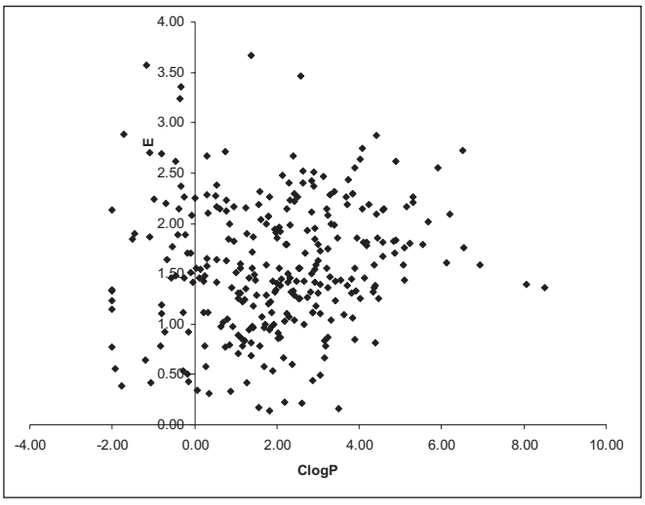


**Figure 1 -** Scatter plot of *ClogP* versus *E* values for data points of train and test sets (N = 295).

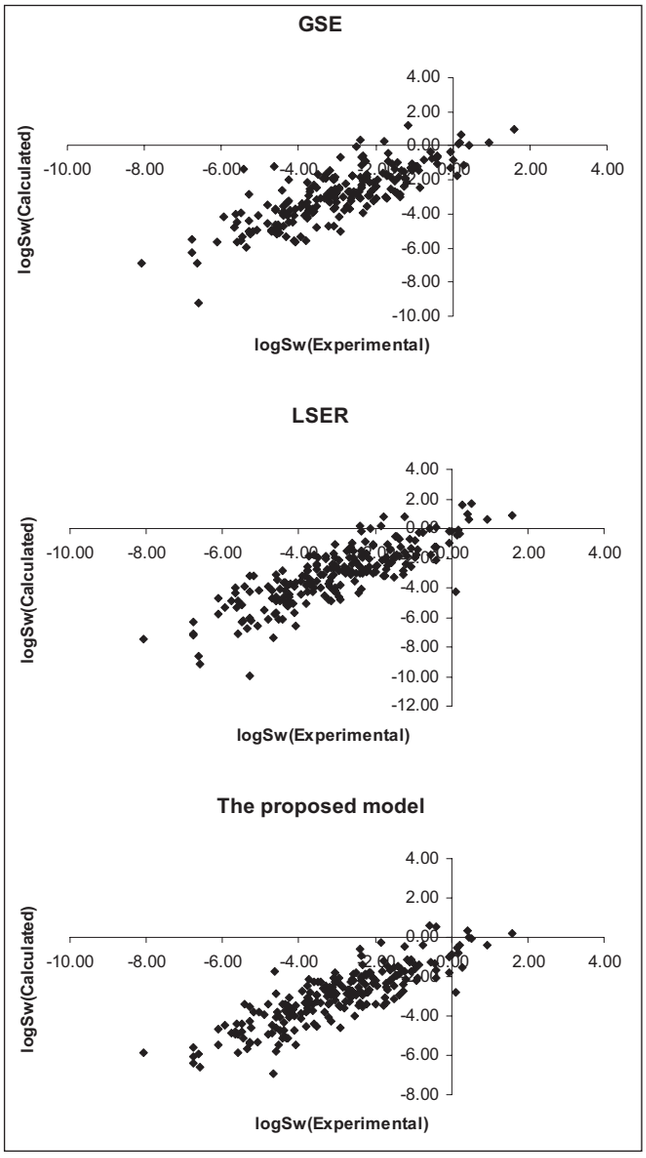


**Figure 2 -** The correlation between experimental values versus predicted values using GSE ($R^2$ = 0.65), LSER ($R^2$ = 0.68), and the proposed model (*Equation 12*) ($R^2$ = 0.73) for train set.

**Table I -** The train set of experimental ($logS_w$) and calculated ($clogS_w$) aqueous solubility data, deviation values (absolute error [AE] and percentage deviation [PD]) of three different models.

| No. | Solute | Ref. | $logS_w$ | GSE | | | LSER | | | Proposed model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD |
| 1 | 1-Naphthol | [18] | -1.98 | -3.04 | 1.06 | 91.3 | -2.59 | 0.61 | 75.3 | -3.37 | 1.39 | 95.9 |
| 2 | 2-Amino-5-bromobenzoic acid | [18] | -3.07 | -3.72 | 0.64 | 77.1 | -1.70 | 1.37 | 2249.2 | -2.94 | 0.14 | 36.6 |
| 3 | 4-Iodophenol | [18] | -1.71 | -3.11 | 1.39 | 95.9 | -2.17 | 0.45 | 64.6 | -3.33 | 1.61 | 97.6 |
| 4 | 1,6-Cleve's acid | [16] | -2.35 | -** | - | - | -1.50 | 0.85 | 606.2 | -1.97 | 0.38 | 139.3 |
| 5 | 2,2',4,5,5'-PCB | [7] | -6.77 | -6.23 | 0.54 | 246.7 | -7.18 | 0.41 | 60.7 | -6.04 | 0.73 | 432.3 |
| 6 | 2,4,5-Trichlorophenol | [16] | -2.22 | -3.56 | 1.34 | 95.5 | -2.92 | 0.70 | 80.0 | -3.38 | 1.16 | 93.1 |
| 7 | 2,4-DB | [16] | -3.73 | -3.74 | 0.01 | 1.5 | -3.38 | 0.36 | 126.7 | -3.14 | 0.59 | 292.8 |
| 8 | 2-Cyclohexyl-4,6-dinitrophenol | [16] | -4.25 | -4.44 | 0.19 | 35.7 | -4.86 | 0.61 | 75.3 | -4.09 | 0.16 | 43.5 |
| 9 | 2-Ethyl-1-hexanol* | [16] | -2.17 | -2.11 | 0.06 | 15.0 | -2.56 | 0.39 | 58.9 | -1.80 | 0.37 | 136.5 |
| 10 | 3,4-Dinitrobenzoic acid | [16] | -1.50 | -2.65 | 1.15 | 92.9 | -1.83 | 0.33 | 53.1 | -2.48 | 0.98 | 89.4 |
| 11 | 4-Amino-2-sulfobenzoic acid | [16] | -1.85 | -** | - | - | 0.15 | 2.00 | 9852.4 | -0.29 | 1.56 | 3509.9 |
| 12 | 4-Hydroxybenzoic acid | [18] | -1.46 | -2.80 | 1.33 | 95.3 | -0.49 | 0.97 | 836.9 | -1.93 | 0.46 | 65.6 |
| 13 | Acequinocyl | [16] | -6.76 | -** | - | - | -7.13 | 0.37 | 57.6 | -6.38 | 0.38 | 138.9 |
| 14 | Acetamide | [16] | 1.58 | 1.00 | 0.58 | 73.7 | 0.86 | 0.72 | 81.0 | 0.17 | 1.41 | 96.1 |
| 15 | Acetamiprid | [16] | -1.72 | -1.75 | 0.02 | 5.4 | -1.94 | 0.21 | 38.5 | -2.34 | 0.61 | 75.7 |
| 16 | Acetanilide | [16] | -1.32 | -1.44 | 0.12 | 23.8 | -1.27 | 0.05 | 13.4 | -1.62 | 0.29 | 49.2 |
| 17 | Acetochlor | [16] | -3.08 | -4.73 | 1.64 | 97.7 | -4.21 | 1.13 | 92.5 | -2.95 | 0.13 | 34.8 |
| 18 | Acetylacetone* | [16] | 0.22 | 0.65 | 0.43 | 168.8 | -0.34 | 0.56 | 72.3 | -0.39 | 0.61 | 75.4 |
| 19 | Acibenzolar-S-methyl | [16] | -4.44 | -2.63 | 1.81 | 6309.0 | -2.83 | 1.60 | 3899.5 | -3.41 | 1.03 | 967.4 |
| 20 | Aconitic acid | [16] | 0.46 | -** | - | - | 0.59 | 0.13 | 34.2 | 0.00 | 0.46 | 65.0 |
| 21 | Acrylamide | [16] | 0.95 | 0.19 | 0.77 | 83.0 | 0.61 | 0.34 | 54.6 | -0.43 | 1.39 | 95.9 |
| 22 | Acrylonitrile | [16] | 0.15 | 0.15 | 0.00 | 0.6 | -0.48 | 0.62 | 76.3 | -0.55 | 0.70 | 80.1 |
| 23 | Adenine | [16] | -2.12 | -2.68 | 0.56 | 72.6 | -0.02 | 2.09 | 12334.1 | -1.80 | 0.32 | 108.0 |
| 24 | Adipic acid | [16] | -0.75 | -0.84 | 0.09 | 18.9 | -0.29 | 0.46 | 188.3 | -0.41 | 0.34 | 117.1 |
| 25 | Aldicarb | [16] | -1.50 | -1.40 | 0.10 | 25.7 | -2.09 | 0.59 | 74.2 | -1.56 | 0.06 | 13.5 |
| 26 | Allidochlor | [16] | -0.95 | -** | - | - | -1.99 | 1.04 | 90.9 | -1.59 | 0.64 | 77.2 |
| 27 | Allobarbital | [16] | -2.06 | -2.13 | 0.07 | 14.8 | -1.17 | 0.89 | 680.0 | -2.05 | 0.01 | 2.1 |
| 28 | Alochlor | [16] | -3.05 | -2.52 | 0.53 | 239.6 | -4.21 | 1.16 | 93.1 | -2.95 | 0.10 | 25.2 |
| 29 | Alpha-acetylbutyrolactone* | [16] | 0.19 | 0.22 | 0.03 | 6.3 | -0.07 | 0.26 | 45.3 | -0.81 | 1.01 | 90.1 |
| 30 | Alprenolol | [18] | -2.63 | -3.27 | 0.64 | 76.9 | -3.11 | 0.48 | 66.6 | -3.07 | 0.44 | 63.6 |
| 31 | Amicarbalide | [16] | -1.77 | -** | - | - | -3.16 | 1.39 | 95.9 | -3.46 | 1.69 | 97.9 |
| 32 | Aminopromazine | [16] | -5.75 | -** | - | - | -4.92 | 0.83 | 580.8 | -4.85 | 0.90 | 698.8 |
| 33 | Amitraz | [16] | -5.47 | -5.34 | 0.13 | 34.1 | -6.24 | 0.78 | 83.3 | -5.13 | 0.33 | 115.3 |
| 34 | Amobarbital | [16] | -2.57 | -2.70 | 0.13 | 25.1 | -1.96 | 0.61 | 310.9 | -2.22 | 0.36 | 128.4 |
| 35 | Ampicillin | [2] | -1.69 | -0.87 | 0.82 | 560.7 | -2.45 | 0.76 | 82.6 | -2.64 | 0.95 | 88.7 |
| 36 | Ancymidol | [16] | -2.60 | -1.75 | 0.85 | 609.2 | -2.89 | 0.29 | 49.2 | -2.75 | 0.15 | 29.1 |
| 37 | Androstanolone | [20] | -4.16 | -4.49 | 0.33 | 53.2 | -5.09 | 0.93 | 88.3 | -3.64 | 0.52 | 227.5 |
| 38 | Aniline* | [16] | -0.41 | -0.64 | 0.23 | 40.7 | -1.21 | 0.80 | 84.1 | -1.64 | 1.23 | 94.1 |
| 39 | Anthraquinone | [20] | -5.19 | -5.03 | 0.16 | 44.5 | -3.23 | 1.96 | 9111.8 | -3.80 | 1.39 | 2348.2 |
| 40 | ANTU | [16] | -2.53 | -2.76 | 0.23 | 41.4 | -3.53 | 1.00 | 89.9 | -4.01 | 1.49 | 96.7 |
| 41 | Ascorbic acid | [16] | 0.28 | -1.12 | 1.40 | 96.0 | 1.62 | 1.34 | 2106.8 | -1.56 | 1.83 | 98.5 |
| 42 | Aspirin | [7] | -1.61 | -1.82 | 0.21 | 38.3 | -1.43 | 0.18 | 51.9 | -1.66 | 0.05 | 11.9 |
| 43 | Asulam | [16] | -1.66 | -0.44 | 1.22 | 1571.8 | -1.09 | 0.57 | 274.6 | -1.47 | 0.19 | 54.7 |
| 44 | Atrazine | [7] | -3.55 | -3.52 | 0.03 | 7.2 | -2.74 | 0.81 | 538.5 | -2.91 | 0.64 | 335.9 |
| 45 | Azidamfenicol | [16] | -1.17 | 1.18 | 2.35 | 22245.6 | -1.87 | 0.70 | 80.1 | -1.15 | 0.02 | 5.1 |
| 46 | Azintamide | [16] | -1.72 | -1.68 | 0.04 | 8.5 | -2.71 | 1.00 | 89.9 | -2.53 | 0.82 | 84.7 |
| 47 | Azoxystrobin | [16] | -4.61 | -3.81 | 0.80 | 524.8 | -4.58 | 0.03 | 6.1 | -4.62 | 0.01 | 2.4 |
| 48 | Badische acid | [16] | -2.57 | -** | - | - | -1.50 | 1.07 | 1085.6 | -1.86 | 0.71 | 414.9 |
| 49 | Barban | [16] | -4.37 | -3.43 | 0.94 | 770.5 | -3.68 | 0.69 | 390.0 | -3.42 | 0.95 | 786.1 |
| 50 | Barbital | [16] | -1.39 | -1.80 | 0.41 | 60.9 | -0.64 | 0.75 | 466.9 | -1.48 | 0.09 | 18.1 |
| 51 | Bendiocarb | [16] | -2.93 | -2.43 | 0.50 | 219.0 | -2.12 | 0.82 | 553.1 | -2.36 | 0.57 | 274.4 |
| 52 | Benzidine | [16] | -2.76 | -2.20 | 0.56 | 263.4 | -3.00 | 0.24 | 42.7 | -3.26 | 0.50 | 68.4 |
| 53 | Benzoic acid | [2] | -1.59 | -2.51 | 0.92 | 88.1 | -1.46 | 0.13 | 36.4 | -2.18 | 0.59 | 74.2 |
| 54 | Benzylimidazole | [16] | -2.26 | -1.73 | 0.53 | 237.1 | -2.72 | 0.46 | 65.3 | -2.41 | 0.15 | 29.0 |
| 55 | Betamethasone | [20] | -3.77 | -3.35 | 0.42 | 163.0 | -4.24 | 0.47 | 66.1 | -3.38 | 0.39 | 145.4 |
| 56 | Bifenox | [16] | -5.93 | -4.19 | 1.74 | 5410.5 | -5.30 | 0.63 | 330.3 | -4.46 | 1.47 | 2841.5 |
| 57 | Bifentrhin | [16] | -6.63 | -6.87 | 0.24 | 43.0 | -8.59 | 1.97 | 98.9 | -5.92 | 0.71 | 409.4 |
| 58 | Bifonazole | [19] | -5.59 | -5.65 | 0.06 | 12.7 | -7.11 | 1.52 | 97.0 | -5.85 | 0.26 | 44.7 |
| 59 | Biotin | [16] | -3.05 | -2.81 | 0.24 | 72.2 | -1.98 | 1.07 | 1064.0 | -2.24 | 0.80 | 533.8 |
| 60 | Bupivacaine | [18] | -3.22 | -4.68 | 1.45 | 96.5 | -4.80 | 1.57 | 97.3 | -4.07 | 0.85 | 85.9 |
| 61 | Caffeine | [20] | -0.95 | -1.19 | 0.24 | 42.5 | -0.82 | 0.13 | 36.1 | -1.38 | 0.43 | 62.5 |
| 62 | Capric acid | [16] | -3.45 | -3.06 | 0.39 | 143.3 | -3.13 | 0.32 | 108.6 | -2.27 | 1.18 | 1402.3 |
| 63 | Caproic acid* | [16] | -1.05 | -1.05 | 0.00 | 0.7 | -1.29 | 0.24 | 42.5 | -1.12 | 0.06 | 13.8 |
| 64 | Carbofuran | [16] | -2.84 | -2.99 | 0.15 | 29.2 | -2.89 | 0.05 | 10.5 | -2.56 | 0.28 | 91.0 |
| 65 | Carbosulfan* | [16] | -6.10 | -5.63 | 0.47 | 196.8 | -5.74 | 0.36 | 130.2 | -5.46 | 0.64 | 336.7 |

**Table I -** The train set of experimental ($logS_w$) and calculated ($clogS_w$) aqueous solubility data, deviation values (absolute error [AE] and percentage deviation [PD]) of three different models (continued).

| No. | Solute | Ref. | $logS_w$ | GSE | | | LSER | | | Proposed model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD |
| 66 | Carboxin | [16] | -1.94 | -2.69 | 1.13 | 1257.2 | -2.59 | 0.38 | 140.9 | -2.82 | 0.26 | 80.4 |
| 67 | Carfentrazone-ethyl* | [16] | -1.97 | -4.67 | 2.30 | 19949.4 | -1.70 | 0.39 | 59.7 | -3.07 | 1.20 | 1493.7 |
| 68 | Carisoprodol | [16] | -2.53 | -2.86 | 0.41 | 156.2 | -2.17 | 0.08 | 18.9 | -2.08 | 0.86 | 620.6 |
| 69 | Carprofen | [18] | -5.06 | -4.72 | 0.36 | 56.0 | -1.50 | 0.02 | 4.8 | -4.84 | 0.14 | 28.0 |
| 70 | Chlordane | [7] | -5.98 | -6.74 | 0.63 | 76.6 | -7.18 | 1.39 | 96.0 | -5.64 | 0.29 | 49.0 |
| 71 | Chlorpropamide | [18] | -2.86 | -2.72 | 0.39 | 144.6 | -2.92 | 0.52 | 234.6 | -3.02 | 0.23 | 69.5 |
| 72 | Chlorprothixene form I | [18] | -5.53 | -6.35 | 1.22 | 1577.5 | -3.38 | 0.40 | 149.2 | -5.63 | 1.12 | 1210.5 |
| 73 | Chlorpyriphos | [7] | -4.75 | -4.36 | 0.92 | 731.8 | -4.86 | 1.31 | 1961.3 | -4.81 | 0.86 | 632.8 |
| 74 | Chlorzoxazone | [18] | -3.72 | -1.46 | 1.06 | 91.3 | -2.56 | 1.19 | 1459.4 | -3.25 | 0.59 | 74.6 |
| 75 | Cholesterol | [20] | -9.23 | -9.18 | 2.62 | 99.8 | -1.83 | 2.57 | 99.7 | -6.60 | 0.01 | 1.4 |
| 76 | Cimetidine | [2] | -0.97 | -1.76 | 0.46 | 188.4 | 0.15 | 0.33 | 52.9 | -1.94 | 0.51 | 68.9 |
| 77 | Citral* | [20] | -2.54 | -3.00 | 0.48 | 66.9 | -0.49 | 0.94 | 88.4 | -2.37 | 0.31 | 50.6 |
| 78 | Codeine | [19] | -2.04 | -3.28 | 0.53 | 70.5 | -7.13 | 1.77 | 98.3 | -3.14 | 1.63 | 97.7 |
| 79 | Corticosterone | [2] | -3.28 | -3.91 | 0.08 | 16.8 | 0.86 | 0.71 | 80.3 | -3.32 | 0.12 | 24.2 |
| 80 | Crotonic acid | [16] | -0.85 | -0.17 | 0.85 | 85.8 | -1.94 | 0.17 | 31.8 | -0.89 | 0.89 | 87.2 |
| 81 | Cumic acid | [16] | -3.62 | -2.53 | 0.58 | 73.6 | -1.27 | 0.51 | 220.3 | -2.78 | 0.26 | 81.2 |
| 82 | Cyanazine | [16] | -2.94 | -2.71 | 0.21 | 61.6 | -4.21 | 0.44 | 176.4 | -2.77 | 0.38 | 138.0 |
| 83 | Cyclizine | [16] | -2.99 | -4.32 | 0.56 | 72.4 | -0.34 | 1.90 | 98.7 | -3.50 | 1.07 | 91.5 |
| 84 | Cyclobarbital | [16] | -2.94 | -2.09 | 0.77 | 83.0 | -2.83 | 0.08 | 20.6 | -2.64 | 0.47 | 66.3 |
| 85 | Cycloleucine | [16] | -0.61 | 0.11 | 0.20 | 36.4 | 0.59 | 0.52 | 231.6 | 0.54 | 0.95 | 796.3 |
| 86 | Cymoxanil | [16] | -0.71 | -0.18 | 1.64 | 4296.0 | 0.61 | 2.17 | 14677.1 | -0.93 | 1.41 | 2499.4 |
| 87 | Cyproconazole | [16] | -3.06 | -4.52 | 0.26 | 83.7 | -0.48 | 1.20 | 93.7 | -3.78 | 0.46 | 65.5 |
| 88 | Cyprodinil | [16] | -3.25 | -4.43 | 0.99 | 876.7 | -0.02 | 0.19 | 35.9 | -3.89 | 0.35 | 124.9 |
| 89 | Danazol | [20] | -5.57 | -6.28 | 0.06 | 12.9 | -0.29 | 0.77 | 83.0 | -4.82 | 0.69 | 392.5 |
| 90 | DDT | [7] | -6.89 | -7.48 | 1.19 | 1448.8 | -2.09 | 0.60 | 298.9 | -5.88 | 2.20 | 15728.1 |
| 91 | Dehydroacetic acid | [16] | 0.37 | 0.17 | 2.76 | 57007.0 | -1.99 | 2.56 | 35849.2 | -0.60 | 1.79 | 6050.9 |
| 92 | Desipramine | [2] | -5.60 | -5.11 | 1.79 | 98.4 | -1.17 | 1.30 | 95.0 | -4.52 | 0.71 | 80.6 |
| 93 | Dexamethasone | [16] | -3.67 | -4.24 | 0.03 | 5.6 | -4.21 | 0.59 | 74.6 | -3.38 | 0.26 | 84.0 |
| 94 | Diallate | [16] | -3.35 | -3.79 | 0.93 | 760.4 | -0.07 | 0.50 | 214.0 | -3.48 | 0.80 | 530.7 |
| 95 | Diazinon | [7] | -4.26 | -3.70 | 0.50 | 68.4 | -3.11 | 0.06 | 13.6 | -3.74 | 0.02 | 5.0 |
| 96 | Dibucaine | [17] | -4.68 | -4.58 | 0.29 | 48.7 | -3.16 | 0.19 | 36.0 | -4.71 | 0.32 | 52.2 |
| 97 | Dicamba | [16] | -2.69 | -2.47 | 0.27 | 45.7 | -4.92 | 0.04 | 9.8 | -2.56 | 0.14 | 27.1 |
| 98 | Dichlobenil | [16] | -3.36 | -3.20 | 0.55 | 257.7 | -6.24 | 0.71 | 408.2 | -2.71 | 1.20 | 1496.7 |
| 99 | Dichlofenthion | [16] | -** | -4.68 | - | - | -1.96 | 1.43 | 2580.7 | -4.66 | 1.45 | 2733.6 |
| 100 | Diclofop-methyl | [16] | -4.01 | -5.31 | 1.62 | 4059.1 | -2.45 | 0.32 | 108.3 | -4.38 | 1.25 | 1675.3 |
| 101 | Difenoconazole | [16] | -3.90 | -6.16 | 0.53 | 240.3 | -2.89 | 1.73 | 98.1 | -5.17 | 0.73 | 81.5 |
| 102 | Digallic acid | [16] | -3.75 | -2.59 | 0.94 | 88.5 | -5.09 | 0.22 | 65.9 | -3.60 | 0.79 | 83.7 |
| 103 | Digitoxin | [20] | -4.39 | -9.98 | 0.89 | 676.2 | -1.21 | 4.70 | 100.0 | -5.39 | 0.11 | 22.9 |
| 104 | Dimethenamid* | [16] | -1.14 | -4.05 | 1.22 | 1565.7 | -3.23 | 1.69 | 97.9 | -2.17 | 0.19 | 54.7 |
| 105 | Dimethirimol | [16] | -1.70 | -1.70 | 0.54 | 248.1 | -3.53 | 0.55 | 251.5 | -1.94 | 0.31 | 102.6 |
| 106 | Dimethomorph | [16] | -5.34 | -5.13 | 1.02 | 90.5 | 1.62 | 0.81 | 84.5 | -5.14 | 0.82 | 85.0 |
| 107 | Dimorpholamine | [16] | -1.74 | -4.30 | 1.83 | 98.5 | -1.43 | 4.40 | 100.0 | -2.77 | 2.87 | 99.9 |
| 108 | Diniconazole | [16] | -** | -5.49 | - | - | -1.09 | 0.58 | 73.8 | -3.91 | 1.00 | 897.8 |
| 109 | Diphenhydramine | [18] | -4.16 | -4.64 | 1.21 | 93.9 | -2.74 | 1.69 | 98.0 | -3.45 | 0.50 | 68.5 |
| 110 | Diphenylhydantoin | [18] | -4.07 | -3.20 | 0.21 | 38.9 | -1.87 | 0.65 | 348.3 | -3.33 | 0.52 | 234.8 |
| 111 | Diuron | [7] | -3.58 | -2.87 | 0.18 | 51.4 | -2.71 | 0.89 | 683.3 | -3.05 | 0.71 | 408.7 |
| 112 | Enrofloxacin | [18] | -1.46 | -3.52 | 1.72 | 5161.5 | -4.58 | 0.34 | 53.9 | -2.51 | 0.67 | 371.1 |
| 113 | EPTC* | [16] | -2.67 | -2.79 | 0.03 | 7.9 | -1.50 | 0.09 | 18.7 | -2.64 | 0.06 | 14.6 |
| 114 | Equilin | [16] | -5.04 | -4.26 | 0.24 | 73.3 | -3.68 | 1.02 | 937.5 | -4.24 | 1.04 | 997.4 |
| 115 | Estrone | [20] | -5.31 | -4.44 | 1.36 | 95.6 | -0.64 | 0.49 | 67.9 | -4.13 | 0.18 | 33.9 |
| 116 | Ethinamate | [16] | -1.58 | -2.13 | 0.25 | 75.9 | -2.12 | 0.30 | 50.4 | -1.72 | 0.10 | 27.0 |
| 117 | Ethirimol | [16] | -2.77 | -4.32 | 0.25 | 77.7 | -3.00 | 1.30 | 95.0 | -2.26 | 0.76 | 472.4 |
| 118 | Ethofumesate | [16] | -2.14 | -2.61 | 1.62 | 4048.5 | -1.46 | 1.15 | 1310.0 | -2.45 | 1.31 | 1926.3 |
| 119 | Ethohexadiol | [16] | -0.78 | -1.74 | 0.24 | 42.2 | -2.72 | 1.19 | 93.6 | -1.23 | 0.69 | 79.7 |
| 120 | Ethoprop* | [16] | -2.74 | -3.04 | 0.23 | 41.2 | -4.24 | 0.53 | 70.4 | -2.91 | 0.40 | 60.1 |
| 121 | Ethyl-p-hydroxybenzoate | [19] | -2.49 | -1.54 | 0.14 | 27.6 | -5.30 | 0.81 | 546.1 | -2.21 | 0.14 | 36.9 |
| 122 | Fenbufen | [16] | -4.10 | -4.20 | 0.96 | 814.8 | -8.59 | 0.86 | 626.5 | -3.78 | 1.28 | 1804.7 |
| 123 | Fenoprofen* | [18] | -2.54 | -3.82 | 1.16 | 1339.5 | -7.11 | 0.12 | 25.0 | -3.37 | 0.33 | 114.6 |
| 124 | Fenoxaprop-ethyl | [16] | -4.50 | -4.89 | 1.10 | 1171.9 | -1.98 | 0.72 | 419.9 | -4.97 | 0.64 | 335.3 |
| 125 | Fenpiclonil | [16] | -4.91 | -4.14 | 0.22 | 39.3 | -4.80 | 0.55 | 254.4 | -4.49 | 0.20 | 59.9 |
| 126 | Fludrocortisone | [16] | -** | -4.02 | - | - | -0.82 | 0.58 | 73.8 | -3.32 | 0.12 | 31.3 |
| 127 | Flufenacet | [16] | -4.02 | -3.85 | 0.21 | 38.0 | -3.13 | 0.04 | 8.9 | -3.80 | 0.01 | 2.1 |
| 128 | Flufenamic acid | [16] | -5.07 | -4.06 | 0.57 | 73.4 | -1.29 | 0.43 | 170.2 | -4.08 | 0.41 | 154.3 |
| 129 | Flumioxazin | [16] | -2.87 | -3.19 | 2.43 | 26903.2 | -2.89 | 2.11 | 12691.3 | -3.53 | 1.76 | 5709.9 |
| 130 | Flurbiprofen | [2] | -3.90 | -4.21 | 0.47 | 191.7 | -5.74 | 0.15 | 40.4 | -3.72 | 0.64 | 334.8 |

**Table I -** The train set of experimental ($logS_w$) and calculated ($clogS_w$) aqueous solubility data, deviation values (absolute error [AE] and percentage deviation [PD]) of three different models (continued).

| No. | Solute | Ref. | $logS_w$ | GSE | | | LSER | | | Proposed model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD |
| 131 | Fluspirilene | [16] | -4.68 | -** | - | - | -7.41 | 2.73 | 99.8 | -6.92 | 2.24 | 99.4 |
| 132 | Fluthiacet-methyl | [16] | -5.68 | -** | - | - | -3.96 | 1.72 | 5122.7 | -4.94 | 0.73 | 441.8 |
| 133 | Folic acid | [16] | -5.44 | -1.39 | 4.05 | 1121490.4 | -3.95 | 1.49 | 2978.1 | -3.39 | 2.05 | 11187.4 |
| 134 | Fumaric acid | [16] | -1.22 | -1.94 | 0.72 | 81.0 | 0.78 | 2.00 | 10003.3 | -0.45 | 0.77 | 490.9 |
| 135 | Furametpyr | [16] | -3.17 | -** | - | - | -4.86 | 1.69 | 97.9 | -4.25 | 1.08 | 91.7 |
| 136 | Furazolidone | [16] | -3.75 | -2.91 | 0.84 | 591.0 | -1.83 | 1.92 | 8174.5 | -2.12 | 1.63 | 4145.0 |
| 137 | Ganciclovir | [16] | -1.77 | 0.25 | 2.02 | 10466.6 | 0.78 | 2.56 | 35887.9 | -1.17 | 0.60 | 301.2 |
| 138 | Glipizide | [18] | -5.49 | -3.97 | 1.52 | 3235.5 | -5.15 | 0.34 | 118.9 | -4.38 | 1.11 | 1189.9 |
| 139 | Gluconolactone | [16] | 0.52 | -** | - | - | 1.71 | 1.19 | 1460.7 | -0.08 | 0.60 | 74.9 |
| 140 | Glyburide | [20] | -5.09 | -4.96 | 0.13 | 34.9 | -6.56 | 1.47 | 96.6 | -5.34 | 0.25 | 44.2 |
| 141 | Heptabarbital | [16] | -3.00 | -3.34 | 0.34 | 54.2 | -2.56 | 0.44 | 177.4 | -2.88 | 0.12 | 33.3 |
| 142 | Hexazinone | [16] | -0.88 | -1.46 | 0.58 | 73.5 | -1.46 | 0.58 | 73.5 | -2.09 | 1.20 | 93.7 |
| 143 | Hexobarbital | [18] | -2.67 | -2.72 | 0.04 | 9.2 | -2.00 | 0.68 | 375.4 | -2.68 | 0.00 | 0.7 |
| 144 | Hydrocortisone | [16] | -3.05 | -3.05 | 0.00 | 1.0 | -4.00 | 0.94 | 88.6 | -3.23 | 0.17 | 32.8 |
| 145 | Hydroflumethiazide | [16] | -3.04 | -1.86 | 1.19 | 1442.5 | -1.05 | 1.99 | 9694.3 | -1.83 | 1.21 | 1532.0 |
| 146 | Hydroxyphenamate | [16] | -0.92 | -1.23 | 0.31 | 50.6 | -1.75 | 0.82 | 85.0 | -2.17 | 1.25 | 94.3 |
| 147 | Hydroxyproline | [16] | 0.44 | 0.01 | 0.43 | 62.8 | 0.99 | 0.55 | 257.1 | 0.34 | 0.10 | 20.1 |
| 148 | Hymexazol | [16] | -0.07 | -0.38 | 0.31 | 50.8 | -0.19 | 0.12 | 24.4 | -1.02 | 0.96 | 88.9 |
| 149 | Idoxuridine | [16] | -2.25 | -0.89 | 1.36 | 2181.0 | -0.85 | 1.40 | 2419.5 | -1.90 | 0.35 | 124.0 |
| 150 | Imazapyr | [16] | -1.36 | -3.01 | 1.65 | 97.7 | -1.85 | 0.49 | 67.6 | -2.96 | 1.60 | 97.5 |
| 151 | Imazaquin | [16] | -3.54 | -4.76 | 1.22 | 94.0 | -3.53 | 0.00 | 1.0 | -4.51 | 0.97 | 89.4 |
| 152 | Imazethapyr | [16] | -2.32 | -3.93 | 1.61 | 97.6 | -2.84 | 0.52 | 69.9 | -3.54 | 1.22 | 94.0 |
| 153 | Imipramine | [18] | -4.11 | -5.58 | 1.47 | 96.6 | -5.73 | 1.62 | 97.6 | -4.76 | 0.65 | 77.7 |
| 154 | Indoprofen | [19] | -4.82 | -3.45 | 1.37 | 2249.8 | -3.87 | 0.95 | 793.5 | -3.38 | 1.44 | 2654.6 |
| 155 | Isoflurophate* | [16] | -1.08 | -1.31 | 0.23 | 41.4 | -0.69 | 0.38 | 142.5 | -1.24 | 0.16 | 31.1 |
| 156 | Isoleucine | [16] | -0.58 | -0.34 | 0.25 | 76.5 | 0.02 | 0.61 | 303.8 | 0.63 | 1.21 | 1515.8 |
| 157 | Isophorone* | [16] | -1.06 | -1.41 | 0.35 | 55.2 | -2.54 | 1.48 | 96.7 | -1.74 | 0.68 | 79.3 |
| 158 | Ketanserin | [16] | -4.60 | -3.70 | 0.90 | 689.0 | -4.85 | 0.25 | 44.2 | -4.04 | 0.56 | 261.0 |
| 159 | Khellin | [16] | -2.40 | -3.05 | 0.65 | 77.4 | -3.00 | 0.60 | 74.9 | -3.34 | 0.94 | 88.5 |
| 160 | Lenacil | [16] | -4.59 | -4.84 | 0.25 | 43.7 | -3.12 | 1.47 | 2858.1 | -2.89 | 1.70 | 4920.2 |
| 161 | Lidocaine | [20] | -1.76 | -3.00 | 1.24 | 94.2 | -3.16 | 1.40 | 96.0 | -3.06 | 1.30 | 94.9 |
| 162 | Linalool* | [18] | -1.99 | -2.38 | 0.39 | 59.3 | -2.96 | 0.97 | 89.3 | -2.21 | 0.22 | 40.4 |
| 163 | Linuron | [16] | -3.52 | -3.18 | 0.34 | 119.0 | -2.43 | 1.09 | 1117.6 | -3.25 | 0.27 | 84.8 |
| 164 | Lomefloxacin | [18] | -2.33 | -0.55 | 1.78 | 5965.0 | -3.03 | 0.70 | 79.9 | -1.42 | 0.91 | 717.1 |
| 165 | Lorazepam | [20] | -3.60 | -3.82 | 0.22 | 39.7 | -4.03 | 0.43 | 63.1 | -4.37 | 0.77 | 83.1 |
| 166 | Malathion | [7] | -3.36 | -1.92 | 1.44 | 2654.2 | -3.13 | 0.23 | 68.6 | -2.61 | 0.75 | 467.5 |
| 167 | Medrogestone | [20] | -5.27 | -5.21 | 0.06 | 14.8 | -6.25 | 0.98 | 89.5 | -4.60 | 0.67 | 367.2 |
| 168 | Methomyl | [16] | -0.45 | -0.78 | 0.33 | 53.5 | -1.20 | 0.75 | 82.4 | -1.31 | 0.86 | 86.1 |
| 169 | Metoprolol | [2] | -1.20 | -2.17 | 0.97 | 89.3 | -2.43 | 1.23 | 94.1 | -2.14 | 0.94 | 88.6 |
| 170 | Morphine | [19] | -3.15 | -2.58 | 0.57 | 271.5 | -3.32 | 0.17 | 31.8 | -2.95 | 0.20 | 59.6 |
| 171 | Nadolol | [2] | -1.57 | -1.09 | 0.48 | 202.0 | -2.32 | 0.75 | 82.4 | -2.15 | 0.58 | 73.5 |
| 172 | Nalidixic acid | [18] | -3.61 | -2.32 | 1.30 | 1873.0 | -1.78 | 1.83 | 6646.6 | -2.27 | 1.34 | 2069.4 |
| 173 | Naphthoic acid | [17] | -3.77 | -4.15 | 0.38 | 57.9 | -2.83 | 0.94 | 779.3 | -3.61 | 0.17 | 47.5 |
| 174 | Naproxen | [2] | -4.21 | -3.79 | 0.42 | 163.0 | -3.71 | 0.50 | 216.3 | -3.62 | 0.59 | 293.1 |
| 175 | Nicotinic acid | [20] | -0.84 | -2.46 | 1.62 | 97.6 | -0.23 | 0.61 | 307.3 | -1.38 | 0.54 | 71.3 |
| 176 | Niflumic acid | [18] | -4.58 | -5.20 | 0.62 | 76.1 | -3.49 | 1.09 | 1141.3 | -3.83 | 0.75 | 464.8 |
| 177 | Ofloxacin | [18] | -1.27 | -1.55 | 0.28 | 48.1 | -2.86 | 1.60 | 97.5 | -2.37 | 1.10 | 92.1 |
| 178 | Oxazepam | [20] | -3.95 | -3.63 | 0.32 | 108.9 | -3.52 | 0.43 | 171.0 | -3.87 | 0.08 | 20.4 |
| 179 | Parathion* | [7] | -4.29 | -3.30 | 0.99 | 877.2 | -3.78 | 0.51 | 222.2 | -3.89 | 0.40 | 152.0 |
| 180 | Perphenazine | [19] | -4.59 | -4.63 | 0.04 | 9.8 | -5.73 | 1.15 | 92.9 | -5.83 | 1.25 | 94.4 |
| 181 | p-Fluorobenzoic acid | [16] | -2.07 | -3.26 | 1.19 | 93.6 | -1.08 | 0.99 | 873.1 | -2.04 | 0.03 | 6.8 |
| 182 | Phenacetin | [20] | -2.35 | -2.28 | 0.07 | 17.5 | -1.84 | 0.51 | 221.4 | -2.07 | 0.28 | 88.9 |
| 183 | Phenol | [20] | -0.06 | -1.25 | 1.19 | 93.5 | -0.95 | 0.89 | 87.1 | -1.82 | 1.76 | 98.3 |
| 184 | Phenolphthalein | [7] | -2.90 | -5.01 | 2.11 | 99.2 | -4.80 | 1.90 | 98.8 | -4.62 | 1.72 | 98.1 |
| 185 | Phenylbutazone | [18] | -4.39 | -3.51 | 0.88 | 659.9 | -3.67 | 0.72 | 425.6 | -4.31 | 0.08 | 19.5 |
| 186 | Phthalazine | [16] | -0.42 | -1.05 | 0.63 | 76.5 | -2.13 | 1.71 | 98.1 | -2.05 | 1.63 | 97.7 |
| 187 | Phthalic acid | [16] | -1.37 | -2.87 | 1.50 | 96.8 | -0.72 | 0.65 | 350.2 | -1.84 | 0.46 | 65.4 |
| 188 | Phthalimide | [16] | -2.61 | -2.85 | 0.24 | 42.2 | -0.95 | 1.66 | 4513.0 | -2.11 | 0.50 | 217.7 |
| 189 | p-Hydroxybenzoicacid | [16] | -1.44 | -2.80 | 1.35 | 95.6 | -0.49 | 0.95 | 788.2 | -1.93 | 0.49 | 67.4 |
| 190 | Picloram | [16] | -2.75 | -3.56 | 0.81 | 84.3 | -1.55 | 1.20 | 1467.0 | -2.88 | 0.13 | 26.3 |
| 191 | Picric acid | [16] | -1.26 | -2.38 | 1.12 | 92.4 | -3.07 | 1.81 | 98.5 | -2.73 | 1.47 | 96.6 |
| 192 | Pirimicarb | [16] | -1.95 | -2.01 | 0.06 | 12.7 | -2.45 | 0.51 | 68.8 | -2.46 | 0.52 | 69.7 |
| 193 | Prednisolone | [20] | -3.21 | -3.22 | 0.01 | 2.3 | -3.78 | 0.57 | 72.9 | -3.37 | 0.16 | 30.8 |
| 194 | Pregnenolone | [20] | -4.65 | -1.17 | 3.48 | 299816.3 | -5.75 | 1.10 | 92.1 | -1.75 | 2.90 | 78516.4 |
| 195 | Primidone | [20] | -2.64 | -3.06 | 0.42 | 62.0 | -1.26 | 1.38 | 2290.3 | -2.28 | 0.36 | 130.1 |

**Table I -** The train set of experimental ($logS_w$) and calculated ($clogS_w$) aqueous solubility data, deviation values (absolute error [AE] and percentage deviation [PD]) of three different models (continued).

| No. | Solute | Ref. | $logS_w$ | GSE | | | LSER | | | Proposed model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD | clog-$S_w$ | AE | PD |
| 196 | Promazine* | [19] | -4.30 | -4.08 | 0.22 | 66.4 | -5.24 | 0.93 | 88.4 | -5.11 | 0.81 | 84.5 |
| 197 | Promethazine | [2] | -4.39 | -4.08 | 0.31 | 104.2 | -4.88 | 0.49 | 67.5 | -4.97 | 0.58 | 73.6 |
| 198 | Propranolol | [2] | -3.62 | -3.93 | 0.31 | 50.5 | -3.66 | 0.04 | 9.2 | -3.74 | 0.12 | 25.0 |
| 199 | Pyrimethamine | [17] | -4.11 | -4.10 | 0.01 | 3.0 | -3.87 | 0.23 | 71.3 | -4.02 | 0.09 | 23.4 |
| 200 | Ranitidine | [18] | -2.50 | -0.08 | 2.43 | 26559.8 | -2.20 | 0.30 | 100.3 | -1.79 | 0.71 | 413.5 |
| 201 | Salbutamol | [20] | -1.22 | -1.91 | 0.69 | 79.6 | -1.56 | 0.34 | 54.1 | -1.73 | 0.51 | 68.9 |
| 202 | Santonin | [20] | -3.09 | -2.02 | 1.07 | 1074.9 | -3.15 | 0.06 | 13.1 | -2.04 | 1.05 | 1032.2 |
| 203 | Sulfadiazine | [20] | -3.51 | -1.71 | 1.80 | 6209.6 | -1.91 | 1.60 | 3865.7 | -2.27 | 1.24 | 1657.8 |
| 204 | Sulfamerazine | [20] | -3.12 | -1.94 | 1.18 | 1413.6 | -2.39 | 0.73 | 435.5 | -2.53 | 0.59 | 286.5 |
| 205 | Sulfamethizole | [20] | -2.41 | -1.86 | 0.55 | 254.8 | -2.57 | 0.16 | 30.1 | -2.73 | 0.32 | 52.2 |
| 206 | Sulfamethoxazole | [20] | -2.62 | -1.78 | 0.84 | 591.8 | -2.41 | 0.21 | 62.0 | -2.73 | 0.11 | 22.0 |
| 207 | Sulfamethoxypyridazine | [20] | -3.28 | -1.68 | 1.60 | 3881.1 | -2.06 | 1.22 | 1565.9 | -2.76 | 0.52 | 234.0 |
| 208 | Sulfisomidine | [20] | -2.24 | -1.84 | 0.40 | 151.2 | -2.81 | 0.57 | 72.8 | -2.81 | 0.57 | 73.3 |
| 209 | Sulindac | [18] | -4.50 | -4.75 | 0.25 | 44.3 | -4.89 | 0.40 | 59.9 | -4.71 | 0.22 | 39.1 |
| 210 | Tetracycline | [18] | -2.93 | -0.64 | 2.29 | 19444.9 | -3.78 | 0.86 | 86.1 | -3.53 | 0.61 | 75.3 |
| 211 | Thionazin* | [16] | -2.34 | -2.05 | 0.29 | 94.0 | -1.93 | 0.41 | 156.6 | -2.92 | 0.58 | 73.7 |
| 212 | Thymol | [18] | -2.19 | -2.92 | 0.73 | 81.3 | -2.84 | 0.65 | 77.6 | -2.82 | 0.64 | 76.8 |
| 213 | Tolmetin | [18] | -4.09 | -3.73 | 0.36 | 130.1 | -3.67 | 0.42 | 164.7 | -3.46 | 0.63 | 326.9 |
| 214 | Triamcinolone | [20] | -3.69 | -2.68 | 1.01 | 923.3 | -3.48 | 0.21 | 62.8 | -2.82 | 0.87 | 635.0 |
| 215 | Triazolam | [20] | -4.08 | -5.66 | 1.58 | 97.4 | -6.62 | 2.54 | 99.7 | -5.48 | 1.40 | 96.1 |
| 216 | Trichloromethiazide | [18] | -3.53 | -2.48 | 1.05 | 1019.3 | -2.46 | 1.07 | 1080.0 | -2.96 | 0.56 | 267.2 |
| 217 | Trifluoperazine* | [19] | -4.52 | -4.65 | 0.13 | 25.4 | -6.11 | 1.59 | 97.4 | -5.50 | 0.98 | 89.4 |
| 218 | Triflupromazine* | [19] | -5.30 | -5.04 | 0.26 | 82.4 | -6.02 | 0.72 | 80.9 | -5.31 | 0.01 | 1.8 |
| 219 | Trimipramine | [18] | -4.79 | -4.54 | 0.25 | 77.4 | -6.11 | 1.32 | 95.3 | -4.92 | 0.13 | 26.5 |
| 220 | Tryptamine | [18] | -3.30 | -1.92 | 1.38 | 2274.7 | -1.89 | 1.41 | 2463.8 | -2.49 | 0.80 | 533.6 |

*Melting point is less than 25 °C. **Melting point has not been reported in the literature.

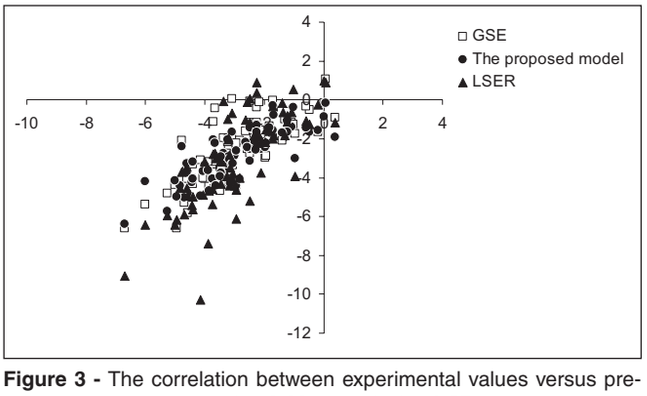


**Figure 3 -** The correlation between experimental values versus predicted value using GSE ($R^2 = 0.62$), LSER ($R^2 = 0.57$) and the proposed (*Equation 12*) ($R^2 = 0.66$) models for test set.

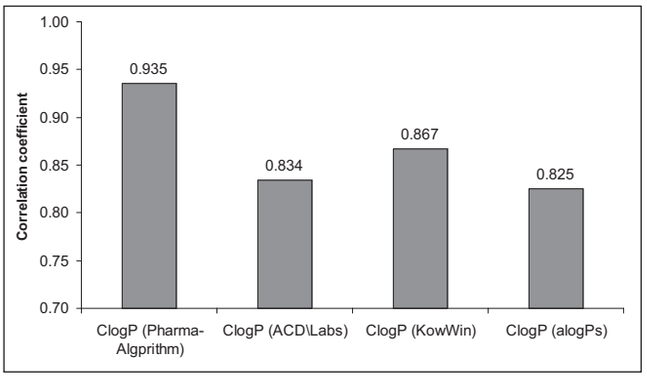


**Figure 4 -** Correlation between experimental *logP* (*ElogP*) and different calculated logPs (*ClogP*) for the test set.

and the obtained AAE values are listed in *Table V*. These results show there is good correlation between *ElogP* and *ClogP* values. The best correlation is observed for the *ClogP* of the Pharma-Algorithms and there are no significant differences between AAE values of aqueous solubilities computed using *ElogP* and *ClogP* values ($p > 0.05$). Also using *ElogP* and different *ClogPs*, the proposed model has lower deviation when compared with that of GSE.

The validity of the proposed model was investigated by cross-validation method. The cross-validation results of the model are investigated and the obtained QSPR models, $R^2$, $Q^2$ (for $F_1$, $F_2$, $F_3$) and AAE of 2, 4 and 10-folds of the cross validations are listed in *Table VI*. The coefficients of QSPRs models are not significantly changed in different cross-validation methods. The $R^2$ values of all models varied between 0.92-0.94 and different $Q^2$ values are greater than 0.6. The overall AAE for 2-fold, 4-fold and 10-fold are 0.709, 0.714 and 0.707, so the evidence shows that *Equation 12* is a robust model.

*Table II* lists details of the proposed test set to compare the accuracies of the aqueous solubility prediction models, experimental and predicted values of solubilities and the references. Normal distribution of this data set was investigated by the Kolmogorov-Smirov and Shapiro-Wilk tests and the results indicated that the experimental solubility, *logP* and *E* of 75 official drugs are normally distributed with the significance level of less than 0.05. The range of the logarithm of the molar solubility is from -6.69 to 0.39, *logP* from -2.00 to 5.90, *E* from 0.17 to 3.67 and *mp* from -118 to 360 ºC.

External validation using the test set was performed according to the above mentioned criteria. The $R^2$, $Q^2$, $[(R^2 - R_0^2)/R^2]$ and k values are 0.66, 0.65, 0.09 and 0.936, respectively, which are in good agreement with the critical values and show the validity of the proposed model.

The results presented in this study show that *E* and the experimental or calculated *logP* are effective descriptors in predicting the aqueous

**Table II -** The proposed test set of experimental ($logS_w$) and calculated ($clogS_w$) aqueous solubility data, deviation values (absolute error [AE] and percentage deviation [PD]) of three different models.

| No. | Solute | Ref. | $logS_w$ | GSE | | | LSER | | | Proposed model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD |
| 1 | Acetaminophen | [18] | -1.06 | -1.18 | 0.116 | 23.5 | -0.63 | 0.430 | 169.3 | -1.39 | 0.328 | 53.0 |
| 2 | Acetazolamide | [30] | -2.49 | -1.180 | 1.315 | 1965.4 | 0.06 | 2.546 | 35080.2 | -1.43 | 1.061 | 1049.5 |
| 3 | Acyclovir | [2] | -2.24 | -0.37 | 1.870 | 7313.1 | 0.91 | 3.145 | 139693.2 | -1.26 | 0.981 | 856.2 |
| 4 | Allopurinol | [20] | -2.26 | -2.19 | 0.070 | 17.5 | 0.38 | 2.639 | 43483.5 | -1.30 | 0.960 | 812.5 |
| 5 | Amiloride | [2] | -3.36 | -1.96 | 1.400 | 2411.9 | -0.07 | 3.292 | 195775.4 | -2.74 | 0.621 | 317.6 |
| 6 | Amoxicilin | [2] | -2.17 | -0.11 | 2.060 | 11381.5 | -1.90 | 0.275 | 88.2 | -2.38 | 0.207 | 37.9 |
| 7 | Antipyrine | [7] | 0.39 | -0.9 | 1.290 | 94.9 | -1.16 | 1.553 | 97.2 | -1.91 | 2.304 | 99.5 |
| 8 | Atenolol | [2] | -1.30 | -0.98 | 0.320 | 108.9 | -1.85 | 0.545 | 71.5 | -1.81 | 0.513 | 69.3 |
| 9 | Atropine | [20] | -2.12 | -1.77 | 0.355 | 126.5 | -3.77 | 1.653 | 97.8 | -2.43 | 0.312 | 51.2 |
| 10 | Azathioprine | [18] | -3.21 | -1.98 | 1.233 | 1609.4 | -1.95 | 1.263 | 1731.2 | -3.16 | 0.044 | 10.6 |
| 11 | Baclofen | [31] | -1.67 | -0.53 | 1.142 | 1286.2 | -1.84 | 0.167 | 31.9 | -0.76 | 0.913 | 718.5 |
| 12 | Benzocaine | [17] | -2.33 | -1.98 | 0.346 | 121.8 | -1.82 | 0.510 | 223.4 | -2.14 | 0.189 | 54.5 |
| 13 | Celecoxib | [30] | -3.74 | -3.73 | 0.011 | 2.6 | -5.38 | 1.646 | 97.7 | -4.55 | 0.812 | 84.6 |
| 14 | Chloramphenicol* | [31] | -2.11 | -1.57 | 0.543 | 248.8 | -2.16 | 0.052 | 11.3 | -2.55 | 0.439 | 63.6 |
| 15 | Chlorpromazine | [2] | -5.27 | -4.82 | 0.450 | 181.8 | -5.97 | 0.696 | 79.9 | -5.72 | 0.448 | 64.3 |
| 16 | Ciprofloxacin | [2] | -3.73 | -1.11 | 2.620 | 41586.9 | -2.78 | 0.954 | 800.4 | -2.04 | 1.685 | 4745.1 |
| 17 | Colchicine | [31] | -0.96 | -1.76 | 0.797 | 84.0 | -3.93 | 2.964 | 99.9 | -3.00 | 2.036 | 99.1 |
| 18 | Cortisone | [2] | -3.00 | -2.72 | 0.285 | 92.8 | -4.43 | 1.427 | 96.3 | -2.88 | 0.117 | 31.0 |
| 19 | Dapsone | [31] | -3.19 | -2.43 | 0.766 | 483.2 | -2.05 | 1.136 | 1268.5 | -2.94 | 0.246 | 76.1 |
| 20 | Diazepam | [33] | -3.76 | -3.34 | 0.420 | 163.0 | -4.58 | 0.818 | 84.8 | -4.06 | 0.304 | 50.4 |
| 21 | Diethylstilbestrol | [17] | -4.57 | -5.81 | 1.239 | 94.2 | -4.94 | 0.376 | 57.9 | -4.82 | 0.254 | 44.3 |
| 22 | Digoxin | [20] | -4.16 | -3.12 | 1.040 | 996.5 | -10.31 | 6.150 | 100.0 | -4.94 | 0.777 | 83.3 |
| 23 | Diltiazem | [2] | -2.95 | -4.39 | 1.440 | 96.4 | -4.64 | 1.694 | 98.0 | -4.41 | 1.456 | 96.5 |
| 24 | Ephedrine | [20] | -0.47 | -0.52 | 0.050 | 10.9 | -1.28 | 0.810 | 84.5 | -1.65 | 1.185 | 93.5 |
| 25 | Estradiol | [20] | -4.84 | -4.95 | 0.110 | 22.4 | -4.53 | 0.306 | 102.3 | -4.43 | 0.408 | 155.8 |
| 26 | Famotidine | [2] | -2.48 | -0.09 | 2.395 | 24731.3 | -1.18 | 1.297 | 1882.7 | -2.53 | 0.054 | 11.6 |
| 27 | Fluorouracil | [31] | -1.03 | -1.24 | 0.212 | 38.7 | 0.55 | 1.578 | 3685.9 | -0.38 | 0.645 | 342.0 |
| 28 | Gemfibrozil | [7] | -3.16 | -4.26 | 1.100 | 92.1 | -4.44 | 1.276 | 94.7 | -3.54 | 0.377 | 58.0 |
| 29 | Griseofulvin | [7] | -4.61 | -3.45 | 1.160 | 1345.4 | -3.47 | 1.140 | 1279.3 | -3.28 | 1.329 | 2032.1 |
| 30 | Guaifenesin | [16] | -0.60 | -0.36 | 0.244 | 75.4 | -1.10 | 0.503 | 68.6 | -1.45 | 0.847 | 85.8 |
| 31 | Haloperidol | [16] | -4.43 | -4.08 | 0.353 | 125.4 | -5.44 | 1.013 | 90.3 | -4.22 | 0.205 | 60.5 |
| 32 | Halothane* | [20] | -1.71 | -1.68 | 0.030 | 7.2 | -1.99 | 0.285 | 48.1 | -1.56 | 0.147 | 40.1 |
| 33 | Hydrochlorothiazide | [2] | -2.63 | -1.61 | 1.020 | 947.1 | -1.04 | 1.590 | 3786.2 | -2.18 | 0.450 | 181.6 |
| 34 | Hydroquinone | [16] | -0.18 | -1.66 | 1.478 | 96.7 | -0.23 | 0.046 | 10.0 | -1.56 | 1.371 | 95.7 |
| 35 | Isoniazid | [16] | 0.01 | -0.15 | 0.163 | 31.3 | 0.97 | 0.961 | 814.7 | -0.85 | 0.856 | 86.1 |
| 36 | Ketoprofen | [2] | -3.25 | -2.73 | 0.520 | 231.1 | -3.95 | 0.700 | 80.0 | -3.27 | 0.019 | 4.2 |
| 37 | Labetalol | [2] | -3.45 | -3.44 | 0.010 | 2.3 | -4.32 | 0.871 | 86.5 | -3.75 | 0.300 | 49.9 |
| 38 | Lamotrigine | [33] | -3.14 | -4.05 | 0.913 | 87.8 | -3.48 | 0.339 | 54.2 | -4.26 | 1.127 | 92.5 |
| 39 | Levodopa | [31] | -1.72 | -0.02 | 1.697 | 4881.7 | -0.35 | 1.370 | 2244.1 | -0.29 | 1.426 | 2565.5 |
| 40 | Lindane | [7] | -4.60 | -4.08 | 0.520 | 231.1 | -4.53 | 0.070 | 17.4 | -3.76 | 0.838 | 588.8 |
| 41 | Lovastatin | [31] | -6.01 | -5.40 | 0.610 | 307.4 | -6.42 | 0.416 | 61.7 | -4.18 | 1.824 | 6565.0 |
| 42 | Manitol | [7] | 0.06 | 1.08 | 1.020 | 947.1 | 0.89 | 0.832 | 578.5 | -0.18 | 0.240 | 42.4 |
| 43 | Maprotiline | [18] | -4.69 | -5.28 | 0.589 | 74.2 | -5.91 | 1.218 | 93.9 | -5.03 | 0.335 | 53.7 |
| 44 | Meprobamate | [31] | -1.82 | -1.36 | 0.460 | 188.7 | -1.62 | 0.202 | 59.3 | -1.44 | 0.384 | 142.3 |
| 45 | Mercaptopurine | [31] | -3.09 | -1.84 | 1.249 | 1674.0 | -0.70 | 2.389 | 24381.6 | -1.66 | 1.430 | 2591.6 |
| 46 | Metoclopramide | [34] | -3.18 | -2.99 | 0.184 | 52.6 | -2.85 | 0.330 | 113.6 | -3.04 | 0.136 | 36.9 |
| 47 | Metronidazole | [18] | -1.22 | -0.59 | 0.636 | 332.9 | -1.14 | 0.086 | 21.9 | -1.09 | 0.129 | 34.5 |
| 48 | Minoxidil | [20] | -1.98 | -2.97 | 0.990 | 89.8 | -2.19 | 0.209 | 38.1 | -2.46 | 0.483 | 67.1 |
| 49 | Mitomycin C | [31] | -2.56 | -2.53 | 0.034 | 8.1 | -0.12 | 2.444 | 27710.1 | -2.46 | 0.101 | 26.2 |
| 50 | Mycophenolic acid | [31] | -4.39 | -3.30 | 1.092 | 1135.1 | -4.99 | 0.602 | 75.0 | -3.19 | 1.203 | 1494.9 |
| 51 | Nifedipine | [31] | -4.78 | -2.10 | 2.676 | 47332.0 | -3.71 | 1.069 | 1071.5 | -2.42 | 2.358 | 22703.0 |
| 52 | Nitrofurantoin | [19] | -3.24 | -2.19 | 1.048 | 1018.2 | -0.98 | 2.254 | 17859.9 | -2.03 | 1.205 | 1502.5 |
| 53 | Nitroglycerin* | [31] | -2.26 | -1.19 | 1.069 | 1073.1 | -2.22 | 0.039 | 9.4 | -1.66 | 0.597 | 295.7 |
| 54 | Omeprazole | [35] | -3.62 | -3.21 | 0.413 | 158.8 | -3.00 | 0.621 | 317.7 | -4.43 | 0.805 | 84.3 |
| 55 | Oxytetracycline | [18] | -3.09 | 0.07 | 3.153 | 142171.5 | -4.04 | 0.950 | 88.8 | -3.30 | 0.215 | 39.1 |
| 56 | PABA | [16] | -1.37 | -1.93 | 0.557 | 72.3 | -0.65 | 0.713 | 416.7 | -1.65 | 0.281 | 47.6 |
| 57 | Papaverine | [18] | -3.87 | -4.43 | 0.558 | 72.3 | -4.66 | 0.791 | 83.8 | -4.67 | 0.802 | 84.2 |
| 58 | Phenobarbital | [18] | -2.29 | -2.39 | 0.096 | 19.9 | -1.90 | 0.390 | 145.6 | -2.59 | 0.292 | 49.0 |
| 59 | Phenytoin | [20] | -3.99 | -4.07 | 0.080 | 16.8 | -3.20 | 0.785 | 510.2 | -3.35 | 0.643 | 339.7 |
| 60 | Progesterone | [2] | -4.40 | -4.35 | 0.050 | 12.2 | -5.64 | 1.242 | 94.3 | -4.08 | 0.323 | 110.2 |
| 61 | Propofol* | [36] | -3.05 | -3.38 | 0.333 | 53.6 | -3.82 | 0.776 | 83.3 | -3.28 | 0.229 | 41.0 |
| 62 | Propoxyphene | [2] | -5.01 | -4.38 | 0.635 | 331.5 | -6.45 | 1.441 | 96.4 | -4.14 | 0.869 | 639.9 |
| 63 | Prostaglandin–E2 | [7] | -2.47 | -2.73 | 0.260 | 45.1 | -5.22 | 2.753 | 99.8 | -3.16 | 0.692 | 79.7 |
| 64 | Quinine | [2] | -2.82 | -2.11 | 0.710 | 412.9 | -4.01 | 1.195 | 93.6 | -4.06 | 1.240 | 94.2 |
| 65 | Riboflavin | [31] | -3.65 | -0.43 | 3.218 | 165026.8 | -2.77 | 0.881 | 659.9 | -2.21 | 1.441 | 2662.5 |

**Table II -** The proposed test set of experimental ($logS_w$) and calculated ($clogS_w$) aqueous solubility data, deviation values (absolute error [AE] and percentage deviation [PD]) of three different models (continued).

| No. | Solute | Ref. | $logS_w$ | GSE | | | LSER | | | Proposed model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD | $clogS_w$ | AE | PD |
| 66 | Salicylic acid | [17] | -1.93 | -2.87 | 0.939 | 88.5 | -1.53 | 0.401 | 152.0 | -2.24 | 0.310 | 51.1 |
| 67 | Sertraline | [14] | -4.94 | -6.59 | 4.650 | 97.8 | -6.17 | 1.231 | 94.1 | -4.98 | 0.045 | 9.80 |
| 68 | Sulfacetamide | [20] | -1.23 | -0.99 | 0.240 | 73.8 | -0.83 | 0.402 | 152.4 | -1.64 | 0.407 | 60.9 |
| 69 | Terfenadine | [2] | -6.69 | -6.63 | 0.065 | 16.1 | -9.05 | 2.365 | 99.6 | -6.39 | 0.300 | 99.5 |
| 70 | Testosterone | [2] | -4.06 | -4.02 | 0.040 | 9.7 | -4.89 | 0.833 | 85.3 | -3.66 | 0.395 | 148.4 |
| 71 | Theophylline | [2] | -1.38 | -2.09 | 0.710 | 80.5 | -0.18 | 1.196 | 1470.4 | -1.71 | 0.327 | 52.9 |
| 72 | Thiabendazole | [17] | -3.48 | -4.68 | 1.196 | 93.6 | -3.21 | 0.279 | 90.1 | -3.94 | 0.458 | 65.1 |
| 73 | Tolbutamide | [17] | -3.46 | -2.93 | 0.533 | 241.5 | -3.13 | 0.334 | 116.0 | -2.93 | 0.536 | 243.8 |
| 74 | Trimethoprim | [18] | -2.95 | -2.22 | 0.731 | 438.4 | -6.11 | 3.163 | 99.9 | -2.61 | 0.338 | 117.6 |
| 75 | Warfarin | [29] | -3.89 | -3.19 | 0.700 | 401.2 | -7.40 | 3.510 | 100.0 | -3.61 | 0.277 | 89.1 |

*Melting point is less than 25 °C.

**Table III -** Comparison of AAE and MPD for three studied predictive models using train set.

| | GSE | LSER | Proposed model |
|---|---|---|---|
| AAE (± SD) | 0.767± 0.683 (N = 205) | 0.875 ± 0.690 (N = 220) | 0.707 ± 0.540 (N = 220) |
| MPD (± SD) | 8300.3 ± 81043.1 (N = 205) | 1117.7 ± 3980.7 (N = 220) | 849.9 ± 5476.9 (N = 220) |
| AAE (excluded 10 % maximum and 10 % minimum values for each model) | 0.676 ± 0.441 (N = 163) | 0.792 ± 0.415 (N = 176) | 0.649 ± 0.344 (N = 176) |
| MPD (excluded 10 % maximum and 10 % minimum values for each model) | 270.2 ± 439.6 (N = 163) | 257.4 ± 370.2 (N = 176) | 195.8 ± 248.6 (N = 176) |

[a]Deviations are calculated for 205 compounds because melting points are not reported for 15 compounds.

**Table IV -** Comparison of three studied predictive models using test set of 75 drugs.

| | GSE | LSER | Proposed model |
|---|---|---|---|
| AAE (± SD) | 0.822 ± 0.834 | 1.179 ± 1.048 | 0.670 ± 0.545 |
| MPD (± SD) | 6234.9 ± 25823.1 | 6813.2 ± 28287.2 | 758.4 ± 2779.0 |
| AAE (excluded 10 % maximum and 10 % minimum values for each model) | 0.706 ± 0.424 | 1.016 ± 0.591 | 0.594 ± 0.356 |
| MPD (excluded 10 % maximum and 10 % minimum values for each model) | 391.9 ± 547.2 | 390.4 ± 666.9 | 192.3 ± 282.6 |

**Table V -** The AAEs of GSE and the proposed model using various logP values for test set of 75 drugs.

| | ElogP | ClogP (Pharma-Algorithms) | ClogP (ACD/Labs) | ClogP (KowWin) | ClogP (aLogPs) |
|---|---|---|---|---|---|
| GSE | 0.882 | 0.822 | 0.821 | 0.903 | 0.813 |
| Prop. model | 0.717 | 0.710 | 0.670 | 0.753 | 0.755 |

**Table VI -** Cross-validation data to investigate the validity of the proposed model.

| | QSPR model | $R^2$ | $Q^2F_1$ | $Q^2F_2$ | $Q^2F_3$ | AAE |
|---|---|---|---|---|---|---|
| | | | 2-fold | | | |
| 1 | $logS_w$ = -1.167E-0.585 ClogP | 0.943 | 0.684 | 0.684 | 0.680 | 0.746 |
| 2 | $logS_w$ = -1.072E-0.614 ClogP | 0.926 | 0.761 | 0.761 | 0.763 | 0.672 |
| | Mean | 0.935 | 0.723 | 0.723 | 0.722 | 0.709 |
| | | | 4-fold | | | |
| 1 | $logS_w$ = -1.149E-0.574 ClogP | 0.931 | 0.751 | 0.750 | 0.724 | 0.687 |
| 2 | $logS_w$ = -1.064E-0.624 ClogP | 0.933 | 0.720 | 0.720 | 0.732 | 0.728 |
| 3 | $logS_w$ = -1.122E-0.614 ClogP | 0.942 | 0.607 | 0.607 | 0.624 | 0.811 |
| 4 | $logS_w$ = -1.144E-0.585 ClogP | 0.929 | 0.795 | 0.794 | 0.789 | 0.631 |
| | Mean | 0.934 | 0.718 | 0.718 | 0.717 | 0.714 |
| | | | 10-fold | | | |
| 1 | $logS_w$ = -1.128E-0.600 ClogP | 0.935 | 0.686 | 0.681 | 0.636 | 0.740 |
| 2 | $logS_w$ = -1.118E-0.599 ClogP | 0.932 | 0.778 | 0.777 | 0.786 | 0.662 |
| 3 | $logS_w$ = -1.111E-0.618 ClogP | 0.931 | 0.810 | 0.809 | 0.814 | 0.560 |
| 4 | $logS_w$ = -1.101E-0.612 ClogP | 0.934 | 0.717 | 0.717 | 0.722 | 0.723 |
| 5 | $logS_w$ = -1.110E-0.602 ClogP | 0.938 | 0.552 | 0.551 | 0.562 | 0.886 |
| 6 | $logS_w$ = -1.114E-0.598 ClogP | 0.930 | 0.865 | 0.865 | 0.867 | 0.492 |
| 7 | $logS_w$ = -1.150E-0.584 ClogP | 0.934 | 0.712 | 0.711 | 0.733 | 0.730 |
| 8 | $logS_w$ = -1.141E-0.590 ClogP | 0.935 | 0.665 | 0.664 | 0.676 | 0.809 |
| 9 | $logS_w$ = -1.108E-0.606 ClogP | 0.936 | 0.669 | 0.667 | 0.670 | 0.801 |
| 10 | $logS_w$ = -1.150E-0.600 ClogP | 0.933 | 0.789 | 0.785 | 0.773 | 0.663 |
| | Mean | 0.934 | 0.724 | 0.723 | 0.724 | 0.707 |

solubility of pharmaceutical compounds. The GSE is very robust model to predict the aqueous solubility using only two parameters, but the experimental melting point is required. The LSER model also can predict the aqueous solubility of drugs using five Abraham solute parameters. The results revealed that the proposed QSPR model is more accurate than GSE and LSER models and can be used to estimate aqueous solubility of drugs/drug candidates in discovery and development investigations. The proposed test set of aqueous solubility of 75 pharmaceutically interested compounds could be recommended for comparison of the aqueous solubility prediction methods.

## REFERENCES

1. Jouyban A. - Handbook of Solubility Data for Pharmaceuticals. - CRC Press, Boca Raton, 2009.
2. Faller B., Ertl P. - Computational approaches to determine drug solubility. - Adv. Drug Delivery Rev., **59**, 533-545, 2007.
3. Taskinen J., Norinder U., John B.T., David J.T. - In silico predictions of solubility. - In: Taylor J.B., Triggle D.J. (Eds.), Comprehensive Medicinal Chemistry, vol. II, Elsevier, Oxford, 2007, pp. 627-648.
4. Jorgensen W.L., Duffy E. M. - Prediction of drug solubility from structure. - Adv. Drug Delivery Rev., **54**, 355-366, 2002.
5. Ran Y., Jain N., Yalkowsky S.H. - Prediction of aqueous solubility of organic compounds by the general solubility equation (GSE). - J. Chem. Inf. Comput. Sci., **41**, 1208-1217, 2001.
6. Delaney J.S. - Predicting aqueous solubility from structure. - Drug Discov. Today., **10**, 289-295, 2005.
7. Delaney J.S. - ESOL: estimating aqueous solubility directly from molecular structure - J. Chem. Inf. Comput. Sci., **44**, 1000-1005, 2004.
8. Machatha S.G., Yalkowsky S.H. - Comparison of the octanol/water partition coefficients calculated by ClogP, ACDlogP and KowWin to experimentally determined values. - Int. J. Pharm., **294**, 185-192, 2005.
9. Jouyban A. - Solubility prediction of drugs in water-polyethylene glycol 400 mixtures using Jouyban-Acree model. - Chem. Pharm. Bull., **54**, 1561-1566, 2006.
10. Ghasemi J., Saaidpour S. - QSPR prediction of aqueous solubility of drug-like organic compounds. - Chem. Pharm. Bull., **55**, 669-674, 2007.
11. Abraham M.H., Le J. - The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. - J. Pharm. Sci., **88**, 868-880, 1999.
12. Stovall D.M., Givens C., Keown S., Hoover K.R., Barnes R., Harris C., Lozano J., Nguyen M., Rodriguez E., Acree Jr. W.E., Abraham M.H. - Solubility of crystalline nonelectrolyte solutes in organic solvents: mathematical correlation of 4-chloro-3-nitrobenzoic acid and 2-chloro-5-nitrobenzoic acid solubilities with the Abraham solvation parameter model. - Phys. Chem. Liq., **43**, 351-360, 2005.
13. Abraham M.H., Ibrahim A., Zissimos A.M. - Determination of sets of solute descriptors from chromatographic measurements - J. Chromatogr. A. 1037, 29-47, 2004.
14. Pharma-Algorithms, ADME Boxes, Version 4.0, PharmaAlgorithms Inc., Toronto, Canada, 2008.
15. Balakin K.V., Savchuk N.P., Tetko I. V. - In silico approaches to prediction of aqueous and DMSO solubility of drug-like compounds: trends, problems and solutions. - Curr. Med. Chem., **13**, 223-241, 2006.
16. Duchowicz P.R., Talevi A., Bruno-Blanch L.E., Castro E.A. - New QSPR study for the prediction of aqueous solubility of drug-like compounds. - Bioorg. Med. Chem., **16**, 7944-7955, 2008.
17. Hopfinger A.J., Esposito E.X., Llinàs A., Glen R.C., Goodman J.M. - Findings of the challenge to predict aqueous solubility. - J. Chem. Inf. Model., **49**, 1-5, 2009.
18. Llinàs A., Glen R.C., Goodman J.M. - Solubility challenge: can you predict solubilities of 32 molecules using a database of 100 reliable measurements? - J. Chem. Inf. Model, **48**, 1289-1303, 2008.
19. Lüder K., Lindfors L., Westergren J., Nordholm S., Persson R., Pedersen M. - In silico prediction of drug solubility. 4. Will simple potentials suffice? - J. Comput. Chem., **30**, 1859-1871, 2009.
20. Jain P., Sepassi K., Yalkowsky S.H. - Comparison of aqueous solubility estimation from Aquafac and the GSE. - Int. J. Pharm., **360**, 122-147, 2008.
21. Division of specialized information services, National Institute of Health. ChemID Plus. http://chem.sis.nlm.nih.gov/chemidplus/ Accessed 20 Sep 2009.
22. Consonni V., Ballabio D., Todeschini R. - Comments on the definition of the Q2 parameter for QSAR validation. - J. Chem. Inf. Model, **49**, 1669-1678, 2009.
23. Golbraikh A., Tropsha A. - Beware of q2! - J. Mol. Graph. Model., **20**, 269- 276, 2002.
24. Golbraikh A., Shen M., Xiao Z., Xiao Y. D., Lee K. H., Tropsha A. - Rational selection of training and test sets for the development of validated QSAR models. - J. Comput. Aided Mol. Des., **17**, 241- 253, 2003.
25. Interactive LogKow (KowWin) demo, http://www.srcinc.com/what-we-do/databaseforms.aspx?id=385 Accessed 20 Sep 2009.
26. ACD/Labs package, release 7.0, Advanced Chemistry Development Inc. Toronto.
27. Virtual Computational Chemistry Laboratory, http://www.vcclab.org/lab/alogps/ Accessed 20 Sep 2009.
28. David B.T. - The Sciences and Practice of Pharmacy. - Remington, Lippincott Williams and Wilkins, Philadelphia, 2006.
29. Dearden J.C., Cronin M.T.D., Kaiser K.L.E. - How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). - SAR QSAR in Environ. Res., **20**, 241-266, 2009.
30. Chen X.Q., Cho S.J., Li Y., Venkatesh S. - Prediction of aqueous solubility of organic compounds using a quantitative structure-property relationship. - J. Pharm. Sci., **91**, 1838-1852, 2003.
31. Yalkowsky S.H., He Y. - Handbook of Aqueous Solubility Data. - CRC Press, Boca Raton, 2003.
32. Seedher N., Bhatia S. - Solubility enhancement of COX-2 inhibitors using various solvent systems. - AAPS PharmSciTech., **4**, E33, 2003.
33. Shayanfar A., Fakhree M.A.A., Acree Jr. W.E., Jouyban A. - Solubility of lamotrigine, diazepam, and clonazepam in ethanol + water mixtures at 298.15 K. - J. Chem. Eng. Data, **54**, 1107-1109, 2009.
34. Budavari S., O'Neil M.J., Smith A. - The Merck Index. - 11th ed., Merck & Co. Inc., New Jersey, 2001.
35. Möschwitzer J., Achleitner G., Pomper H., Müller R.H. - Development of an intravenously injectable chemically stable aqueous omeprazole formulation using nanosuspension technology. - Eur. J. Pharm. Biopharm., **58**, 615-619, 2004.
36. Loftsson T., Hreinsdóttir D., Másson M. - Evaluation of cyclodextrin solubilization of drugs. - Int. J. Pharm., **302**, 18-28, 2005.

## SUPPORTING INFORMATION

The used descriptors for each solute in train and test sets are available from the corresponding author on request.

## ACKNOWLEDGEMENT

The authors would like to acknowledge partial financial support of Drug Applied Research Center (under grant No. 81/87) for conducting this research.

## MANUSCRIPT

Received 26 June 2010, accepted for publication 15 September 2010.