Solubility Prediction of Solute in Non-Aqueous Binary Solvent Mixtures

Abolghasem Jouyban,*a Maryam Khoubnasabjafari,b Somaieh Soltani,c Shahla Soltanpour,d Elnaz Tamizid and William E. Acree Jr.f

*aFaculty of Pharmacy and Drug Applied Research Center, Tabriz University of Medical Sciences, 51664, Iran
bKimia Research Institute, P.O. Box 51665-171, Iran
cResearch Center for Pharmaceutical Nanotechnology, Tabriz University of Medical Sciences, 51664, Iran
dBiotechnology Research Centre, Tabriz University of Medical Sciences, 51664, Iran
eStudent Research Center, Tabriz University of Medical Sciences, 51664, Iran
fDepartment of Chemistry, University of North Texas, Denton, TX 76203-5070, USA

The possibility of replacing theoretically computed Abraham parameters with the experimental Abraham parameters in solubility prediction of solutes in non-aqueous binary solvent mixtures using the Jouyban-Acree model was investigated. The solubilities of 90 data sets collected from the literature were predicted using their Abraham parameters, the solvent coefficients and also the solubilities in mono-solvent systems. The accuracy of the predicted solubilities was evaluated by calculating the mean percentage deviation (MPD) and also individual percentage deviations (IPDs). The overall MPD for the analysis using experimental and computed Abraham parameters were the same and was <14%. A favoured IPD distribution was obtained for these numerical analyses. The data sets investigated in this work were collected at various temperatures and the results confirmed the possibility of solubility prediction in binary solvents at various temperatures. We did explore the possibility of ab initio solubility prediction of solutes in binary mixtures using the calculated solubilities in mono-solvent systems; however, the difference between the predicted and observed values increased to ca. 60% for gas-to-solvent coefficients and ca. 200% for water-to-solvent coefficients, which is too large for many predictive applications.

Keywords: solubility, prediction, non-aqueous mixed solvents, Abraham model, Jouyban-Acree model

Introduction

Solubility of a solute is affected by solvent’s and solute’s chemical structures and could be mathematically represented using meaningful parameters like Abraham solvation parameters. The Abraham solvation parameter models provided numerical methods for prediction of solutes’ solubility in a wide variety of neat organic solvents.1,6 The
Abraham models employ five parameters for each solute and six solvent coefficients that were computed for a number of common solvents. The basic models proposed for process within condensed phases is:

\[
\log \left( \frac{C_s}{C_w} \right) = c + e \cdot E + s \cdot S + a \cdot A + b \cdot B + v \cdot V
\]  

(1)

and for process involving gas-to-condensed phase transfer is:

\[
\log \left( \frac{C_s}{C_g} \right) = c + e \cdot E + s \cdot S + a \cdot A + b \cdot B + l \cdot L
\]  

(2)

where \( C_s \) and \( C_w \) are the solute solubility in the organic solvent and water (in mole per liter), respectively, \( C_g \) is the gas phase concentration of the solute, \( E \) is the excess molar refraction, \( S \) is dipolarity/polarizability of solute, \( A \) denotes the solute’s hydrogen-bond acidity, \( B \) stands for the solute’s hydrogen-bond basicity, \( V \) is the McGowan volume of the solute, and \( L \) is the logarithm of the solute gas-hexadecane partition coefficient at 298.15 K. In equations (1) and (2) the coefficients \( c, e, s, a, b, v \) and \( l \) are the model constants (i.e. solvent’s coefficients), which depend upon the solvent system under consideration. Numerical values of the model constants have been reported in the literature for several water-to-organic solvent and gas-to-organic solvent systems.

Solvent mixing or cosolvency is the most common method to alter the solubility of a solute. There is an infinite number of solvent compositions for a given binary solvent, and for some compounds, both linear and non-linear solubility behavior have been reported in mixed solvent systems. The most accurate model to represent the solubility data in mixed solvent systems is the Jouyban-Acree model. Its general form is:

\[
\ln X_m = f_1 \ln X_1 + f_2 \ln X_2 + \sum_{j=3}^{n/2} B_j \left( f_1 - f_2 \right)^j
\]  

(3)

where \( X \) is the mole fraction solubility of the solute, \( f \) denotes the mole fraction of the solvents 1 and 2 in the solvent mixture, subscripts \( m, 1 \) and \( 2 \) are the mixed solvent and solvents 1 and 2, respectively, \( B \) is the model constant which represent various solute-solute, solvent-solvent and solute-solvent interactions. In a previous study, QSPR models were proposed to calculate the numerical values of the \( B \) terms using the Abraham coefficients for 22 solvents and solute descriptors for 5 solutes. The applicability of the proposed method was checked using 194 solubility data sets of five different solutes in various non-aqueous binary solvents. In this work, the possibility of replacing experimentally obtained Abraham parameters with the computed parameters is examined. The prediction capability of the previously developed QSPR models is checked using 90 solubility data sets of solutes which were not used in training process of the QSPR models. The applicability of the proposed method is also shown for predicting solubility at various temperatures. The main limitation of the Abraham model is that solute solvational parameters are known for only 4,000 organic compounds. In a recently released software, this limitation is overcome and one is able to compute \( E, S, A, B, V \) and \( L \) parameters.

Computational Methods and Experimental Data

The solubilities of the solutes in binary solvent mixtures were collected from the literature. Table S1 listed details of the experimental solubility data. The numerical values of the solvents’ coefficients were listed in Table S2. In addition to the experimental database of solute’s parameters, commercial software is also available to compute the parameters. Table S3 lists the experimental and computed values of solute’s parameters. Since the numerical values of \( A \) term for the solutes studied in the previous paper were equal to zero, the corresponding terms have been omitted from the QSPR models.

The \( B \) constants of the Jouyban-Acree model were computed using equations (4)-(6) and (7)-(9), and these model constants were then used to predict the solubilities.
of solutes in binary solvents. The predictions still required numerical values of the solute solubility in each pure solvent, *i.e.* $X_1$ and $X_2$. In order to provide a predictive model (without any experimentally determined data), $C_s$ values of the solutes in the neat solvents under consideration were computed using Abraham models (using experimental values of $C_w$ or $C_g$). The calculated molar solubilities, $C_s$, were converted to the mole fraction solubilities using density of the pure organic solvent. The calculated $X_1$ and $X_2$ values were then substituted into equation (3), along with the $B_j$ values from equations (4)-(6) (or equations (7)-(9)) to predict the solubility in binary solvents by the Jouyban-Acree model.

Results and Discussion

Validation of the previously derived coefficients for solubility predictions using computed Abraham solute descriptors

The solubilities of the solutes in 194 different binary solvent mixtures (for details see Table 1 of a previous paper¹⁰) were predicted using the Jouyban-Acree model and calculated $B_j$ values based on equations (4)-(6) and (7)-(9). Both experimental and computed Abraham solute descriptors were used in the $B_j$ calculations. Table 2 gives the overall MPD (± SD) values for the four predictive methods employed. There are no significant differences between MPDs for methods I and II that used experimental or computed Abraham parameters and experimental values of $X_1$ and $X_2$ (t-test, $p > 0.05$). This observation is important in that it is possible to use computed solute descriptors instead of experimentally based values for predicting $B_j$ constants of the Jouyban-Acree model. However, significant differences are observed using predicted $X_1$ and $X_2$ by equations (1) and (2) for the same set of data and the coefficients ($p < 0.0005$), revealing that the computed Abraham parameters using PharmaAlgorithm software produced less accurate solubility predictions in mono-solvent systems in comparison with the experimental Abraham parameters. To confirm this hypothesis, readers could refer to the predicted solubilities using equations (1) and (2) employing experimental Abraham parameters. As examples, the IPDs of the predicted solubilities of anthracene using equation (1) in various solvents were listed in Table S4 where the differences between predicted solubilities using experimental and computed Abraham parameters were statistically significant (paired t-test, $p < 0.001$ or $p < 0.0005$, for details see footnote of Table 1.

### Table 1. Details of various numerical analyses carried out using experimental and computed Abraham parameters

<table>
<thead>
<tr>
<th>Method No.</th>
<th>$B_j$ terms computed by</th>
<th>Solubility in mono-solvents</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Water-to-solvent coefficients (equations (4)-(6))</td>
<td>Experimental data</td>
</tr>
<tr>
<td>II</td>
<td>Gas-to-solvent coefficients (equations (7)-(9))</td>
<td>Experimental data</td>
</tr>
<tr>
<td>III</td>
<td>Water-to-solvent coefficients (equations (4)-(6))</td>
<td>Computed by equation (1)</td>
</tr>
<tr>
<td>IV</td>
<td>Gas-to-solvent coefficients (equations (7)-(9))</td>
<td>Computed by equation (2)</td>
</tr>
</tbody>
</table>

### Table 2. Overall MPD (±SD) for solubility prediction of 194 sets from a previous work,¹⁰ and this work using experimental and computed Abraham solute parameters

<table>
<thead>
<tr>
<th>Numerical method</th>
<th>Experimental parameters</th>
<th>Computed parameters</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>194 data set from a previous work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>4.6 ± 3.9</td>
<td>4.6 ± 4.0</td>
<td>Not significant</td>
</tr>
<tr>
<td>II</td>
<td>11.3 ± 13.3</td>
<td>10.5 ± 13.0</td>
<td>Not significant</td>
</tr>
<tr>
<td>III</td>
<td>33.0 ± 20.6</td>
<td>52.0 ± 32.2</td>
<td>$p &lt; 0.0005$</td>
</tr>
<tr>
<td>IV</td>
<td>23.9 ± 21.5</td>
<td>118 ± 258.4</td>
<td>$p &lt; 0.0005$</td>
</tr>
<tr>
<td>This work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>13.7 ± 14.0</td>
<td>13.6 ± 13.8</td>
<td>Not significant</td>
</tr>
<tr>
<td>II</td>
<td>12.7 ± 13.9</td>
<td>12.5 ± 13.7</td>
<td>Not significant</td>
</tr>
<tr>
<td>III</td>
<td>228.4 ± 337.7</td>
<td>168.8 ± 325.4</td>
<td>Not significant</td>
</tr>
<tr>
<td>IV</td>
<td>53.5 ± 30.0</td>
<td>62.7 ± 43.6</td>
<td>$p &lt; 0.016$</td>
</tr>
</tbody>
</table>
Table S4). A possible reason for such deviations could be the non-ideally adjusted water-to-solvent coefficients of some solvents as it was reported slightly different \(c, e, a, b\) and \(v\) values for cyclohexane in an earlier report\(^1\) and a recent one\(^2\) in which the \(IPD\) of anthracene solubility in cyclohexane predicted by equation (1) using earlier\(^1\) and amended\(^6\) \(c, e, a, b\) and \(v\) values were 213.2 and 148.4\%, respectively. We have computed the \(IPD\) of anthracene in cyclohexane using amended\(^6\) \(c, e, a, b\) and \(v\) values equal to 149.6\%. The second reason could be the temperature effects on \(X_i\) and \(X_j\) values which did not considered in the equations (1) and (2).

It is possible to improve our \textit{ab initio} prediction approach by developing better methods to predict the solubility in mono-solvent systems. It is difficult to guesstimate the error that one could reasonably expect from employing predictive methods to estimate the solubility in the neat organic solvents as the published methods have been tested on relatively few of the many possible solute-solvent combinations. Based on our review of the published comparisons, we do not think that it would be unreasonable to assign an expected error in the range of 0.1 to 0.3 log units to solubilities predicted by group contribution and linear free energy correlations for many of the simpler systems.

\textit{Predictions using water-to-solvent process and experimental solubility data in mono-solvents}

The predictive calculations discussed in the preceding section concerned solubility data used in generating equations (4)-(9). A more stringent test of any predictive solubility method is its ability to accurately predict solubilities of additional solute molecules, or solubilities of solutes dissolved in additional binary solvent mixtures. To better assess the applications and limitations of methods I-IV, we have compiled from the published literature experimental solubility data for 90 additional data sets (see Table S1). In the first set of calculations on the new data sets, we computed the \(B_i\) terms calculated using experimental Abraham solute descriptors. The calculated \(B_i\) values were then combined with experimentally measured solubility data in the mono-solvents to predict the mole fraction solubility \(\ln X_i\) values for the 90 additional data sets (numerical method I of Table 1) using equation (3). The prediction accuracy of the data was evaluated using \(MPD\) values for this analysis and reported in column 2 of Table S5. The minimum (0.2\%) and maximum (61.6\%) \(MPDs\) were observed for p-benzoquinone in 2, 2, 4-trimethylpentane + cyclohexane and benzenophene in carbon tetrachloride + dodecane mixtures both at 25 \(^\circ\)C. The overall \(MPD\) (\(\pm SD\)) was 13.7 \(\pm 14.0\%\). A similar set of calculations were performed using computed Abraham parameters by PharmaAlgorithm software (see column 6 of Table S5). The minimum (0.2\%) and maximum (61.6\%) \(MPDs\) were observed for the same data sets and the overall \(MPD\) (\(\pm SD\)) was 13.6 \(\pm 13.8\%\). There was no significant difference between 13.7 and 13.6\% (paired \(t\)-test, \(p > 0.05\)). Figures 1 and 2 showed the relative frequencies of \(IPDs\) sorted in three subgroups, i.e. \(\leq 4\), 4-30 and >30\%, for various numerical methods employing experimental and computed Abraham parameters. There was no significant difference between frequencies of \(IPDs\) of both parameters for numerical method I. Figure S1 depicted the overall \(MPDs\) for various solutes and there was no difference between \(MPDs\) calculated using experimental and computed Abraham parameters. These findings confirm the above results using 194 data sets from a previous work\(^{10}\) and reveal that it is possible to replace the experimentally determined Abraham parameters with the computationally obtained parameters for solubility prediction in mixed solvent system using method I.

The equations (4)-(6) were obtained employing the \(B_i\) terms calculated using solubility data of solutes at 25 and 26 \(^\circ\)C, however, the equations were able to predict the solubility a wider temperature range (20-50 \(^\circ\)C) as is evident (as examples) from set numbers 1-7 or 8-14 of Table S1. This is an oversimplification on the constants of the Jouyban-Acree model where it has been assumed that
the Jouyban-Acree model constants are not temperature dependent. The reason for this simplification was the shortage of the solubility data of solutes in non-aqueous binary solvents at various temperatures. The capability of the Jouyban-Acree model for calculating the solubility of solutes in binary solvents at various temperatures has been shown earlier.26

**Predictions using gas-to-solvent process and experimental solubility data in mono-solvents**

In the second set of calculations on the new data set, we computed the $B_j$ coefficients using equations (7)-(9) and experimentally-based Abraham solute descriptors. The calculated $B_j$ values were then combined with experimental $\ln X_1$ and $\ln X_2$ data to predict the mole fraction solubility $\ln X_m$ values for the 90 additional data sets (numerical method II of Table 1) using equation (3). The obtained MPD values are reported in column 3 of Table S5. The minimum (0.2%) and maximum (67.3%) MPDs were observed for p-benzoquinone in heptane + cyclohexane and for benzophenone in carbon tetrachloride + dodecane mixtures both at 25 °C. The overall MPD (± SD) was 12.7 ± 13.9%.

The same calculations were carried out using the computed Abraham parameters by PharmaAlgorithm software and the MPD values of these computations are listed in column 7 of Table S5. A nearly identical MPD pattern was observed for data predicted by the experimental and computed parameters. The overall MPD was 168.9%. Figure S3 showed overall MPD for various solutes studied for experimental and computed parameters. This particular estimational scheme (method III) requires a prior knowledge of the solute’s aqueous molar solubility, and based on the relatively large IPD and MPD between predicted and observed values the method did not provide a very reasonable prediction of the observed solubility behavior.

**Ab initio predictions using water-to-solvent process and computed solubility data in mono-solvents**

In the third set of predictive calculations we again calculated the $B_j$ terms using equations (4)-(6) and the experimentally-based Abraham solute descriptors; however, in equation (3) the experimental mole fraction solubilities in the two mono-solvents were replaced with estimated $X_1$ and $X_2$ values based on equation (1). Results of these calculations are summarized in the fourth column of Table S5 for 86 of the 90 data sets considered. Predictions could not be made for the four p-tolylacetic acid systems because the molar solubility of p-tolylacetic acid in water, $C_w$, was not known. The molar solubility of the solute in water is a required input parameter in the estimation of solute’s solubility in mono-solvents through equation (1). A minimum MPD of 11.2% was observed for naphthalene in benzene + toluene at 25 °C, and a maximum MPD of 1811.2% was obtained for carbazole in octane + cyclohexane at 25 °C. The overall average MPD was 228.7%. The largest MPDs were observed for data set numbers 15-22 (benzoic acid), 29-44 (carbazole) and 77-82 (phenylacetic acid). Similar computations were performed using computed Abraham parameters for predicting $X_1$, $X_2$ and $B_j$ terms using the relevant equations. The MPD values of these computations are listed in column 8 of Table S5. A nearly identical MPD pattern was observed for data predicted by the experimental and computed parameters. The overall MPD was 168.9%. Figure S3 showed overall MPD for various solutes studied for experimental and computed parameters. This particular estimational scheme (method III) requires a prior knowledge of the solute’s aqueous molar solubility, and based on the relatively large IPD and MPD between predicted and observed values the method did not provide a very reasonable prediction of the observed solubility behavior.

**Ab initio predictions using gas-to-solvent process and computed solubility data in mono-solvents**

Numerical method IV (see Table 1) involved using $B_j$ coefficients based on equations (7)-(9), and estimated
values for the solubility of the solute in both mono-solvents computed from equation (2). The minimum and maximum MPD for method IV (see column 5 of Table S5) were 3.6 and 101.5%, respectively, for naphthalene dissolved in benzene + toluene at 25 °C and for pyrene dissolved in toluene + heptane at 20 °C using experimentally-based Abraham solute descriptors. The overall MPD was 53.5 (± 30.0)%. A slightly larger minimum MPD of 5.6% (for naphthalene in carbon tetrachloride + hexane) and larger maximum MPD of 197.5% (for pyrene in toluene + heptane at 20 °C) were obtained using computed Abraham solute descriptors. The overall MPD was also larger, MPD = 62.7%, for the method IV predictions that used computed solute descriptors as input values (see column 9 of Table S5). As discussed above, the large deviations in binary solvents result mostly from the high IPDs of the solutes in mono-solvents (see Table S6 for details). It is difficult to accurately predict solubility in binary solvent mixtures when the inputted solubility data for the mono-solvents that make up the binary solvent mixtures is poorly predicted. Better estimation methods for solute solubility in mono-solvents should allow one to reduce these deviations significantly.

Conclusions

Published methods for estimating the Bj constants of the Jouyban-Acree model were applied successfully to a data set containing experimental solubility data for 90 additional solute-binary solvent-temperature combinations. None of the binary solvent solubility data was used in the regression analyses used to develop the predictive Bj correlations. The predicted Bj constants, when combined with experimental solubility data for the solute dissolved in the mono-solvents, enabled one to estimate the solubility of crystalline organic solutes in binary solvents using the Jouyban-Acree model. The expected prediction errors were < 14 and < 13%, respectively for water-to-solvent and gas-to-solvent coefficients employing both experimentally determined and theoretically calculated Abraham solute descriptors. The relatively small prediction errors indicate that it is possible to predict the solubility in binary solvents with minimum experimental efforts. Experimental solubility data exists in the published literature for many organic solutes in mono-solvents, and the Jouyban-Acree model allows one to quantitively estimate the extent to which cosolvency increases or decreases solute solubility. Such predictions are important in both solubilization and crystallization processes. Moreover, predictive methods, such as the Jouyban-Acree model, provide a convenient means to screen compiled experimental solubility data in order to detect possible outliers for re-determination. For any solubility datum with very high IPD, the re-measurement is recommended. The proposed methods could also be extended to predict the solubility in mixed solvents at various temperatures. We tried to develop an ab initio prediction method employing Cw or Cg data of the solute (numerical methods III and IV); however, the obtained MPDs were ca. 200 and 60%.

As a practical conclusion, there are a number of possible solutions depending on the availability of the required input data: i) If the experimental solubility data of the solute in mono-solvent systems, i.e. \( X_1 \) and \( X_2 \), are available, the best solution to predict the solubility in mixed solvents is the numerical methods I or II and the expected prediction error is ca. 14%. ii) If \( X_1 \) and \( X_2 \) are not available and the aqueous solubility of the solute is known, one could use the numerical method III and the expected prediction error is relatively high (170%) for computed Abraham parameters. iii) If \( X_i \) and \( X_j \) are not available and \( C_0 \) of the solute is known, the numerical method IV could be a solution and the expected prediction error is slightly high (ca. 60%) for computed Abraham parameters.

Supplenemetary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br, as PDF file.

Acknowledgments

The financial support from Drug Applied Research Center (under grant number 85-64) was gratefully acknowledged. The authors also thank PharmaAlgorithm Inc. for the trial version of their software.

References


Received: August 23, 2007
Web Release Date: February 29, 2008