

Internet Electronic Journal of Molecular Design

July 2003, Volume 2, Number 7, Pages 475–491

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday
Part 1

Guest Editors: Douglas J. Klein and Sonja Nikolić

QSPR with TAU Indices: Water Solubility of Diverse Functional Acyclic Compounds

Kunal Roy¹ and Achintya Saha²

¹ Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry,
Department of Pharmaceutical Technology, Jadavpur University, Calcutta 700 032, India

² Department of Chemical Technology, University of Calcutta, 92 A P C Road, Calcutta 700 009,
India

Received: April 7, 2003; Revised: April 28, 2003; Accepted: May 12, 2003; Published: July 31, 2003

Citation of the article:

K. Roy and A. Saha, QSPR with TAU Indices: Water Solubility of Diverse Functional Acyclic Compounds, *Internet Electron. J. Mol. Des.* **2003**, 2, 475–491, <http://www.biochempress.com>.

QSPR with TAU Indices: Water Solubility of Diverse Functional Acyclic Compounds[#]

Kunal Roy^{1,*} and Achintya Saha²

¹ Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Calcutta 700 032, India

² Department of Chemical Technology, University of Calcutta, 92 A P C Road, Calcutta 700 009, India

Received: April 7, 2003; Revised: April 28, 2003; Accepted: May 12, 2003; Published: July 31, 2003

Internet Electron. J. Mol. Des. 2003, 2 (7), 475–491

Abstract

Motivation. The topochemically arrived unique (TAU) scheme, developed in valence electron mobile (VEM) environment, is unique in that it unravels specific contributions of functionality, branching, shape and size factors to the physicochemical property or biological activity while most other indices give mainly a global contribution of the molecule. QSPR/QSAR studies with TAU indices on different physicochemical properties/biological activities of diverse functional compounds will explore the usefulness of TAU indices in modeling studies.

Method. The present communication attempts to correlate water solubility, $\ln S$, of 193 diverse functional acyclic compounds with different TAU indices, namely T , T_R , F , B , N_V , N_P , N_I , N_X , and N_Y . Sometimes, TAU relations have been improved further upon inclusion of suitable indicator or integer variables. The statistical quality of the QSPR model has been judged by statistical parameters such as predicted variance Q^2 , explained variance R_a^2 , correlation coefficient R , and variance ratio F .

Results. This study shows that TAU indices, along with appropriate indicator variables, can predict up to 91.4% and explain 91.9% of the variance of water solubility. The relations can unravel specific contributions of molecular bulk (size), functionality, branching and shape parameters to the water solubility of diverse functional compounds. In general, water solubility increases with increase in functionality and branching, and decreases with increase in molecular bulk. Further, halocarbons and hydrocarbons specifically show reduced water solubility. Some of the hydrocarbons and halogen compounds act as outliers.

Conclusions. The TAU index is an important tool in exploring structure–property relationships in view of its potential to unravel specific contributions of different structural parameters like molecular bulk, shape factors, branching, functionality and carbon skeletal structure.

Keywords. Quantitative structure–property relationships; QSPR; topochemically arrived unique (TAU) scheme; water solubility; valence electron mobile (VEM) environment; topological indices; structural descriptor.

[#] Dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday.

* Correspondence author; phone: 91–33–2414 6676 (O), 91–33–2435 6547 (R); E-mail: kunalroy_in@yahoo.com, URL: http://www.geocities.com/kunalroy_in.

1 INTRODUCTION

Exploring quantitative structure–property relationships (QSPR) of physico–chemical properties of organic compounds using topological indices is a fertile field of research in theoretical chemistry. Structural attributes responsible for contributing to the physico–chemical properties are identified by various structural descriptors. These indices encode structural information such as atomic connectivity, size, shape, branching, cyclicity, presence of hetero–atoms and unsaturation in numerical form purporting for correlation of chemical structure with various physical properties, chemical reactivity or biological activity [1–7]. Usually, the numerical basis of topological indices is either the adjacency matrix or the topological distance matrix [8]. Often a variety of descriptors are required to generate a satisfactory model as different structural aspects are revealed from different indices. A large number of such descriptors have been described in the last three decades and their usefulness in quantitative structure–activity relationship (QSAR) and quantitative structure–property relationship (QSPR) studies has been extensively studied [4]. Among these, some of the most commonly used indices are the Wiener path number [9], the molecular connectivity indices [8,10,11], the Balaban indices [12], the kappa shape indices [13], the electrotopological state atom index [14,15], the Basak indices [3], *etc.* Very recently, a few more novel indices have been introduced by different groups of workers and their usefulness in QSAR/QSPR studies are being investigated [16–20]. A database of different descriptors may also help in the classification of chemical compounds for a target property or activity [21–23].

Similarly with the lipid–water partition coefficient [24–30], water solubility is a very important physico–chemical parameter that can account for many properties of organic chemicals including biopharmaceutical behavior of drugs [31]. Many attempts have been made to model water solubility using different indices, *e.g.*, the Wiener and connectivity indices [32], PI index [33], quantum chemical descriptors [34,35], dipole moment, surface area, volume, molecular weight, number of hydrogen bond acceptor/donor(s) and number of rotatable bonds [36], *etc.*, and different statistical and QSAR methods, *e.g.*, genetic algorithm and partial least squares [37], principal component analysis [38], comparative molecular field analysis [39], artificial neural network [40], *etc.*

In the late eighties, the topochemically arrived unique (TAU) scheme was described by Pal *et al.* [41,42] in valence electron mobile (VEM) environment and this index was claimed to have the power to decode chemical information unraveling specific contributions of functionality, branching, shape and size factors to the physicochemical property or biological activity. We have recently explored comparative QSAR and QSPR studies with TAU indices along with different topological parameters [43–46]. In the present communication, we report the modeling of water solubility of diverse functional acyclic compounds with TAU indices to unravel the diagnostic feature of TAU indices. Though mostly straight chain aliphatic compounds have been considered in the present study, the subsets of alcohols and hydrocarbons contain a few alicyclic compounds also. In some

cases, attempt has been made to improve TAU relations with appropriate indicator parameters or integer variables.

2 MATERIALS AND METHODS

The physicochemical parameter (ln S) values were taken from the literature [8,47].

TAU [41,42,48,49] are Topochemically Arrived Unique indices developed in VEM (valence electron, mobile) environment as an extension of the connectivity concept [8]. The first order composite topochemical index, T , is defined as:

$$T = \sum_{i < j} E_{ij} = \sum_{i < j} (V_i V_j)^{0.5}$$

where E_{ij} is the VEM edge weight of the edge between i^{th} and j^{th} vertices, V_i represents the VEM vertex weight of the i^{th} vertex, which may be calculated as the ratio of core count of the i^{th} vertex, λ_i , to VEM count of the i^{th} vertex (θ_i). λ_i may be calculated as $(Z - Z^v)/Z^v$ whereas θ_i may be calculated as $8 - (2h + 1.5v + n)$. When unsaturation is present, θ_i should be calculated as $0.5v + 2\pi$. The notations v , n and π represent the numbers of sigma bonds (other than hydrogen), nonbonding electrons and π bonds associated with the atom in that order. Z , Z^v and h represent the atomic number, number of valence electrons and number of hydrogens attached respectively. The values of core count, VEM count and VEM vertex weight of different atoms in different hybridization states and skeletal arrangements are shown in Table 1.

Table 1. Core count, VEM count and VEM vertex weight for different sets of isovertrices

| Atom Feature | Core Count (λ_i) | VEM Count (θ_i) | VEM Vertex Weight (V_i) |
|--------------------|----------------------------|--------------------------|-----------------------------|
| -CH ₃ | 0.500 | 0.500 | 1.000 |
| -CH ₂ - | 0.500 | 1.000 | 0.500 |
| =CH ₂ | 0.500 | 2.500 | 0.200 |
| >CH- | 0.500 | 1.500 | 0.333 |
| =CH- | 0.500 | 3.000 | 0.167 |
| ≡CH | 0.500 | 4.500 | 0.111 |
| >C< | 0.500 | 2.000 | 0.250 |
| =C= | 0.500 | 5.000 | 0.100 |
| =C< | 0.500 | 3.500 | 0.143 |
| ≡C- | 0.500 | 5.000 | 0.100 |
| -OH | 0.333 | 0.500 | 0.667 |
| -O- | 0.333 | 1.000 | 0.333 |
| =O | 0.333 | 2.500 | 0.133 |
| -NH ₂ | 0.400 | 0.500 | 0.800 |
| -NH- | 0.400 | 1.000 | 0.400 |
| =NH | 0.400 | 2.500 | 0.160 |
| >N- | 0.400 | 1.500 | 0.267 |
| =N- | 0.400 | 3.000 | 0.133 |
| ≡N | 0.400 | 4.500 | 0.089 |
| -F | 0.286 | 0.500 | 0.571 |
| -Cl | 1.429 | 0.500 | 2.857 |
| -Br | 4.000 | 0.500 | 8.000 |

Table 1. (Continued)

| Atom Feature | Core Count (λ_i) | VEM Count (θ_i) | VEM Vertex Weight (V_i) |
|--------------|----------------------------|--------------------------|-----------------------------|
| –I | 6.579 | 0.500 | 13.158 |
| –SH | 1.667 | 0.500 | 3.333 |
| –S– | 1.667 | 1.000 | 1.667 |
| >S– | 1.667 | 1.500 | 1.111 |
| >S< | 1.667 | 2.000 | 0.834 |
| =S | 1.667 | 2.500 | 0.667 |

In the case of a heteroatom, the VEM edge weight of an edge incident upon the heteroatom is assigned a negative value (*i.e.*, multiplied by -1). The composite index T can further be divided into two components: the skeletal index T_R and the functionality F . The skeletal index T_R is the topochemical index of the reference alkane that may be obtained by replacing a heteroatom with carbon and removing the multiple bonds that may be present. The derived index F can be easily calculated as $T_R - T$. T_R may further be factored as the branching parameter B and the constitutional parameter vertex count N_V . B is calculated as $T_N - T_R$ where T_N is the topochemical index of the corresponding normal alkane (for acyclic molecules). N_V for the reference alkane may further be partitioned into N_P (number of methyl carbons), N_I (number of methylene carbons), and N_B (number of branched carbons). N_B is composed of N_Y (number of tertiary carbons) and N_X (number of quaternary carbons). Obviously, the value of N_V is equal to sum of N_P , N_I , N_X and N_Y .

The vertex count N_V of a hydrogen-suppressed molecular graph is purely a constitutional parameter because it may be obtained directly from the molecular formula. Even the structural formula is not needed for obtaining the value of N_V . Obviously, any index showing better correlation with physicochemical or biological activity than that shown by N_V will have significance in the context of QSPR/QSAR studies.

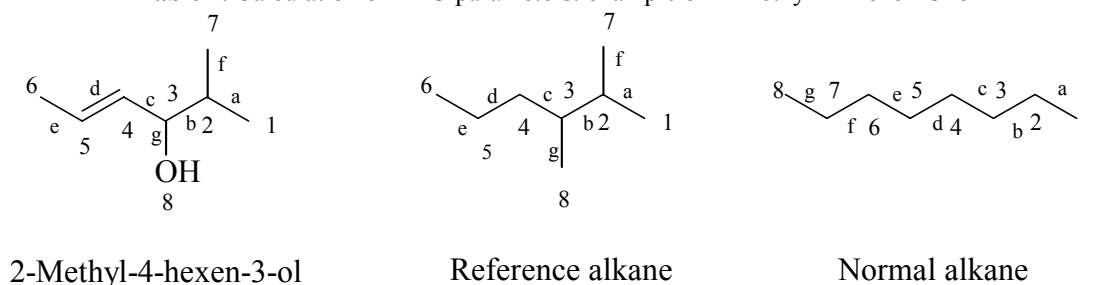
The first order VEM skeletal index T_R is considered as the index for intrinsic lipophilicity while N_B , N_P , N_X and N_Y represent shape parameters. The functionality contribution and bulk parameter are represented by F and N_V , respectively [41,42,48,49]. All TAU indices are basically derived by sequentially partitioning the composite index T into different factors. During the development of QSAR equations with TAU parameters, the above hierarchical relations were followed. For obvious reasons, B and N_B , or N_P and N_B , or N_V and N_I [49] were not used in the same equation.

The calculations of different TAU parameters are illustrated in Table 2 taking as example 2-methyl-4-hexen-3-ol.

Multiple linear regression analyses were done using a software program *RRR98* developed by one of the authors [50]. The statistical quality of the equations [51] was judged by examining the parameters like Q^2 (crossvalidation R^2 or predicted variance), R_a^2 (adjusted R^2 , *i.e.*, explained variance), r or R (correlation coefficient), F (variance ratio) with df (degree of freedom), s (standard error of estimate), $AVRES$ (average of absolute values of residuals) and $SDEP$ (standard deviation

of error of predictions). The significance of the regression coefficients was judged by the *t* test. In case that the intercept of an equation was statistically insignificant and omission of the same did not affect the quality of the equation, the exclusion of the intercept gave statistically more acceptable equation. A compound was considered as an outlier for a particular equation when the residual exceeded twice the standard error of estimate of the equation. PRESS statistics were calculated by the leave-one-out (LOO) technique [52] using the programs *KRPRES1* and *KRPRES2* [50]. Finally, the leave-many-out cross-validation was applied on the final equation of the composite set.

Table 2. Calculation of TAU parameters: example of 2-methyl-4-hexen-3-ol



| Compound | Vertex Count | | | | | | | | Edge Count | | | | | | |
|-----------------------|--------------|-----|-----|-----|-----|---|-----|-----|------------|-------|-------|-------|-------|-------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | a | b | c | d | e | f | g |
| 2-Methyl-4-hexen-3-ol | 1 | 1/3 | 1/3 | 1/6 | 1/6 | 1 | 1 | 2/3 | 0.577 | 0.333 | 0.236 | 0.167 | 0.408 | 0.577 | -0.471 |
| Reference alkane | 1 | 1/3 | 1/3 | 1/2 | 1/2 | 1 | 1 | 1 | 0.577 | 0.333 | 0.408 | 0.5 | 0.707 | 0.577 | 0.577 |
| Normal alkane | 1 | 1/2 | 1/2 | 1/2 | 1/2 | 1 | 1/2 | 1 | 0.707 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.707 |

$$\begin{aligned}
 T &= 1.827 & T_R &= 3.679 \\
 T_N &= 3.914 & F &= 1.852 \\
 B &= 0.235 & N_Y &= 8 \\
 N_B &= 2 & N_X &= 0 \\
 N_Y &= 2 & N_I &= 2 \\
 N_P &= 4 & &
 \end{aligned}$$

3 RESULTS AND DISCUSSION

The calculated topological indices of 193 compounds consisting of 60 alcohols (**1–60**), 43 esters (**61–103**), 16 ethers (**104–119**), 20 ethers (**120–139**), 41 hydrocarbons (**140–180**) and 13 ketones (**181–193**) are given in Table 3. Tables 4 through 10 show relations of water solubility with different topochemical indices. All regression coefficients and variance ratios of the reported equations are significant at 95% and 99% levels respectively unless otherwise stated. Table 3 shows the literature water solubility values of the compounds [8,47] and also the calculated values according to the best equations of individual series and the composite set (*vide* footnote of Table 3).

Table 3. Topological indices and observed and calculated water solubility (ln S) of diverse functional aliphatic compounds

| No | Compound Name | Descriptors | | | ln S | | | | |
|----|----------------------------|-------------|-------|-------|------------------|----------------------|---------------------|-------------------|------------------|
| | | T | T_R | T_N | Obs ^a | Calc | Res | Calc ^h | Res ^h |
| 1 | <i>n</i> -Butanol | 1.130 | 2.414 | 2.414 | 0.006 | 0.210 ^b | -0.204 ^b | -0.147 | 0.153 |
| 2 | 2-Methylpropanol | 0.985 | 2.270 | 2.414 | 0.023 | 0.563 ^b | -0.541 ^b | 0.173 | -0.150 |
| 3 | 2-Butanol | 1.221 | 2.269 | 2.414 | 0.066 | 0.563 ^b | -0.498 ^b | 0.093 | -0.027 |
| 4 | <i>n</i> -Pentanol | 1.630 | 2.914 | 2.914 | -1.347 | -1.089 ^b | -0.258 ^b | -1.333 | -0.014 |
| 5 | 3-Methylbutanol | 1.485 | 2.769 | 2.914 | -1.167 | -0.736 ^b | -0.431 ^b | -1.013 | -0.154 |
| 6 | 2-Methylbutanol | 1.523 | 2.807 | 2.914 | -1.058 | -0.736 ^b | -0.322 ^b | -1.013 | -0.045 |
| 7 | 2-Pentanol | 1.721 | 2.769 | 2.914 | -0.635 | -0.736 ^b | 0.101 ^b | -1.092 | 0.458 |
| 8 | 3-Pentanol | 1.759 | 2.807 | 2.914 | -0.486 | -0.736 ^b | 0.250 ^b | -1.092 | 0.606 |
| 9 | 3-Methyl-2-butanol | 1.593 | 2.641 | 2.914 | -0.405 | -0.382 ^b | -0.023 ^b | -0.772 | 0.367 |
| 10 | 2-Methyl-2-butanol | 1.652 | 2.561 | 2.914 | 0.339 | -0.086 ^b | 0.424 ^b | -0.644 | 0.983 |
| 11 | <i>n</i> -Hexanol | 2.130 | 3.414 | 3.414 | -2.790 | -2.388 ^b | -0.402 ^b | -2.518 | -0.272 |
| 12 | 2-Hexanol | 2.221 | 3.269 | 3.414 | -1.995 | -2.035 ^b | 0.040 ^b | -2.277 | 0.282 |
| 13 | 3-Hexanol | 2.259 | 3.307 | 3.414 | -1.832 | -2.035 ^b | 0.203 ^b | -2.277 | 0.445 |
| 14 | 3-Methyl-3-pentanol | 2.213 | 3.121 | 3.414 | -0.830 | -1.385 ^b | 0.555 ^b | -1.830 | 0.999 |
| 15 | 2-Methyl-2-pentanol | 2.152 | 3.061 | 3.414 | -1.117 | -1.385 ^b | 0.268 ^b | -1.830 | 0.712 |
| 16 | 2-Methyl-3-pentanol | 2.131 | 3.179 | 3.414 | -1.609 | -1.681 ^b | 0.072 ^b | -1.957 | 0.348 |
| 17 | 3-Methyl-2-pentanol | 2.131 | 3.179 | 3.414 | -1.639 | -1.681 ^b | 0.042 ^b | -1.957 | 0.318 |
| 18 | 2,3-Dimethyl-2-butanol | 2.034 | 2.943 | 3.414 | -0.851 | -1.031 ^b | 0.180 ^b | -1.509 | 0.658 |
| 19 | 3,3-Dimethylbutanol | 1.776 | 3.061 | 3.414 | -2.590 | -1.385 ^b | -1.205 ^b | -1.702 | -0.888 |
| 20 | 3,3-Dimethyl-2-butanol | 1.894 | 2.943 | 3.414 | -1.410 | -1.031 ^b | -0.379 ^b | -1.462 | 0.052 |
| 21 | 4-Methylpentanol | 1.985 | 3.269 | 3.414 | -2.282 | -2.035 ^b | -0.247 ^b | -2.198 | -0.084 |
| 22 | 4-Methyl-2-pentanol | 2.076 | 3.124 | 3.414 | -1.814 | -1.681 ^b | -0.133 ^b | -1.957 | 0.143 |
| 23 | 2-Ethylbutanol | 2.061 | 3.345 | 3.414 | -2.787 | -2.035 ^b | -0.752 ^b | -2.198 | -0.589 |
| 24 | Cyclohexanol | 2.345 | 3.393 | 3.414 | -0.960 | -2.035 ^b | 1.075 ^b | -2.277 | 1.318 |
| 25 | <i>n</i> -Heptanol | 2.630 | 3.914 | 3.914 | -4.166 | -3.687 ^b | -0.479 ^b | -3.703 | -0.463 |
| 26 | 2-Methyl-2-hexanol | 2.652 | 3.561 | 3.914 | -2.473 | -2.684 ^b | 0.211 ^b | -3.014 | 0.541 |
| 27 | 3-Methyl-3-hexanol | 2.713 | 3.621 | 3.914 | -2.263 | -2.684 ^b | 0.421 ^b | -3.015 | 0.752 |
| 28 | 3-Ethyl-3-pentanol | 2.774 | 3.682 | 3.914 | -1.917 | -2.684 ^b | 0.767 ^b | -3.015 | 1.098 |
| 29 | 2,3-Dimethyl-2-pentanol | 2.572 | 3.481 | 3.914 | -2.002 | -2.330 ^b | 0.328 ^b | -2.694 | 0.692 |
| 30 | 2,3-Dimethyl-3-pentanol | 2.595 | 3.503 | 3.914 | -1.937 | -2.330 ^b | 0.393 ^b | -2.695 | 0.758 |
| 31 | 2,4-Dimethyl-2-pentanol | 2.507 | 3.416 | 3.914 | -2.145 | -2.330 ^b | 0.185 ^b | -2.694 | 0.549 |
| 32 | 2,4-Dimethyl-3-pentanol | 2.503 | 3.551 | 3.914 | -2.801 | -2.627 ^b | -0.174 ^b | -2.823 | 0.022 |
| 33 | 2,2-Dimethyl-3-pentanol | 2.432 | 3.481 | 3.914 | -2.643 | -2.330 ^b | -0.313 ^b | -2.647 | 0.004 |
| 34 | 3-Heptanol | 2.759 | 3.807 | 3.914 | -3.194 | -3.334 ^b | 0.140 ^b | -3.463 | 0.269 |
| 35 | 4-Heptanol | 2.759 | 3.807 | 3.914 | -3.196 | -3.334 ^b | 0.138 ^b | -3.463 | 0.267 |
| 36 | <i>n</i> -Octanol | 3.130 | 4.414 | 4.414 | -5.401 | -4.987 ^b | -0.414 ^b | -4.888 | -0.513 |
| 37 | 2,2,3-Trimethyl-3-pentanol | 2.902 | 3.811 | 4.414 | -2.931 | -2.979 ^b | 0.048 ^b | -3.384 | 0.453 |
| 38 | 2-Octanol | 3.221 | 4.269 | 4.414 | -4.755 | -4.633 ^b | -0.122 ^b | -4.648 | -0.107 |
| 39 | 2-Ethylhexanol | 3.061 | 4.345 | 4.414 | -4.996 | -4.633 ^b | -0.363 ^b | -4.568 | -0.428 |
| 40 | <i>n</i> -Nonanol | 3.630 | 4.914 | 4.914 | -6.907 | -6.286 ^b | -0.621 ^b | -6.073 | -0.834 |
| 41 | 2-Nonanol | 3.721 | 4.769 | 4.914 | -6.319 | -5.932 ^b | -0.387 ^b | -5.833 | -0.486 |
| 42 | 3-Nonanol | 3.759 | 4.807 | 4.914 | -6.119 | -5.932 ^b | -0.187 ^b | -5.833 | -0.286 |
| 43 | 4-Nonanol | 3.759 | 4.807 | 4.914 | -5.952 | -5.932 ^b | -0.020 ^b | -5.833 | -0.119 |
| 44 | 5-Nonanol | 3.759 | 4.807 | 4.914 | -5.744 | -5.932 ^b | 0.188 ^b | -5.833 | 0.089 |
| 45 | 2,6-Dimethyl-3-heptanol | 3.486 | 4.534 | 4.914 | -5.776 | -5.225 ^b | -0.551 ^b | -5.193 | -0.583 |
| 46 | 3,5-Dimethyl-4-heptanol | 3.579 | 4.627 | 4.914 | -5.298 | -5.225 ^b | -0.073 ^b | -5.193 | -0.105 |
| 47 | 1,1-Diethylpentanol | 3.774 | 4.682 | 4.914 | -5.572 | -5.282 ^b | -0.290 ^b | -5.385 | -0.187 |
| 48 | 7-Methyloctanol | 3.485 | 4.769 | 4.914 | -5.744 | -5.932 ^b | 0.188 ^b | -5.753 | 0.009 |
| 49 | 3,5,5-Trimethylhexanol | 3.169 | 4.454 | 4.914 | -5.769 | -4.929 ^b | -0.840 ^b | -4.938 | -0.831 |
| 50 | <i>n</i> -Decanol | 4.130 | 5.414 | 5.414 | -8.517 | -7.585 ^b | -0.932 ^b | -7.258 | -1.259 |
| 51 | <i>n</i> -Tetradecanol | 6.130 | 7.414 | 7.414 | -12.772 | -12.781 ^b | 0.009 ^b | -11.999 | -0.773 |
| 52 | <i>n</i> -Pentadecanol | 6.630 | 7.914 | 7.914 | -13.796 | -14.081 ^b | 0.285 ^b | -13.184 | -0.612 |
| 53 | <i>n</i> -Hexadecanol | 7.130 | 8.414 | 8.414 | -14.603 | -15.380 ^b | 0.777 ^b | -14.369 | -0.234 |

Table 3. (Continued)

| No | Compound Name | Descriptors | | | | | ln S | | |
|-----|------------------------|-------------|----------------------|----------------------|------------------|----------------------|---------------------|-------------------|------------------|
| | | <i>T</i> | <i>T_R</i> | <i>T_N</i> | Obs ^a | Calc | Res | Calc ^h | Res ^h |
| 54 | 2,2-Dimethyl propanol | 1.276 | 2.561 | 2.914 | -0.889 | -0.086 ^b | -0.803 ^b | -0.517 | -0.372 |
| 55 | 4-Penten-1-ol | 0.894 | 2.914 | 2.914 | -0.355 | -1.089 ^b | 0.734 ^b | -1.084 | 0.729 |
| 56 | 3-Penten-2-ol | 0.917 | 2.769 | 2.914 | 0.127 | -0.736 ^b | 0.863 ^b | -0.821 | 0.948 |
| 57 | 1-Penten-3-ol | 1.062 | 2.807 | 2.914 | 0.035 | -0.736 ^b | 0.771 ^b | -0.857 | 0.892 |
| 58 | 1-Hexen-3-ol | 1.562 | 3.307 | 3.414 | -1.354 | -2.035 ^b | 0.681 ^b | -2.042 | 0.688 |
| 59 | 2-Hexen-4-ol | 1.455 | 3.307 | 3.414 | -0.939 | -2.035 ^b | 1.096 ^b | -2.006 | 1.067 |
| 60 | 2-Methyl-4-penten-3-ol | 1.434 | 3.179 | 3.414 | -1.156 | -1.681 ^b | 0.525 ^b | -1.722 | 0.566 |
| 61 | Methyl formate | -0.962 | 1.914 | 1.914 | 1.015 | 1.321 ^c | -0.306 ^c | 1.575 | -0.560 |
| 62 | Ethyl formate | -0.086 | 2.414 | 2.414 | 0.174 | 0.146 ^c | 2.762 ^c | 0.263 | -0.089 |
| 63 | Ethyl formate | -0.086 | 2.414 | 2.414 | -0.345 | 0.146 ^c | -0.492 ^c | 0.263 | -0.609 |
| 64 | Propyl formate | 0.414 | 2.914 | 2.914 | -1.133 | -1.029 ^c | -0.104 ^c | -0.922 | -0.211 |
| 65 | Propyl formate | 0.414 | 2.914 | 2.914 | -1.174 | -1.029 ^c | -0.146 ^c | -0.922 | -0.253 |
| 66 | Butyl formate | 0.914 | 3.414 | 3.414 | -2.303 | -2.204 ^c | -0.099 ^c | -2.107 | -0.196 |
| 67 | Butyl formate | 0.914 | 3.414 | 3.414 | -2.733 | -2.204 ^c | -0.529 ^c | -2.107 | -0.626 |
| 68 | 1-Pentyl formate | 1.414 | 3.914 | 3.914 | -3.500 | -3.379 ^c | -0.121 ^c | -3.292 | -0.208 |
| 69 | Methyl acetate | -0.555 | 2.269 | 2.414 | 1.191 | 0.544 ^c | 0.647 ^c | 0.693 | 0.498 |
| 70 | Methyl acetate | -0.555 | 2.269 | 2.414 | 0.924 | 0.544 ^c | 0.381 ^c | 0.693 | 0.231 |
| 71 | Ethyl acetate | 0.321 | 2.769 | 2.914 | -0.092 | -0.631 ^c | 0.539 ^c | -0.619 | 0.527 |
| 72 | Ethyl acetate | 0.321 | 2.769 | 2.914 | -0.069 | -0.631 ^c | 0.562 ^c | -0.619 | 0.550 |
| 73 | Isopropyl acetate | 0.843 | 3.124 | 3.414 | -1.194 | -1.409 ^c | 0.215 ^c | -1.541 | 0.347 |
| 74 | Isopropyl acetate | 0.843 | 3.124 | 3.414 | -1.245 | -1.409 ^c | 0.164 ^c | -1.541 | 0.296 |
| 75 | Propyl acetate | 0.821 | 3.269 | 3.414 | -1.704 | -1.807 ^c | 0.103 ^c | -1.805 | 0.101 |
| 76 | Propyl acetate | 0.821 | 3.269 | 3.414 | -1.726 | -1.807 ^c | 0.081 ^c | -1.805 | 0.079 |
| 77 | Isobutyl acetate | 1.176 | 3.624 | 3.914 | -2.849 | -2.584 ^c | -0.265 ^c | -2.670 | -0.179 |
| 78 | Butyl acetate | 1.321 | 3.769 | 3.914 | -3.154 | -2.982 ^c | -0.172 ^c | -2.990 | -0.164 |
| 79 | Isopentyl acetate | 1.676 | 4.124 | 4.414 | -4.398 | -3.759 ^c | -0.639 ^c | -3.855 | -0.543 |
| 80 | Pentyl acetate | 1.821 | 4.269 | 4.414 | -4.283 | -4.157 ^c | -0.126 ^c | -4.175 | -0.108 |
| 81 | Hexyl acetate | 2.321 | 4.769 | 4.914 | -4.721 | -5.332 ^c | 0.611 ^c | -5.360 | 0.639 |
| 82 | Methyl propionate | 0.041 | 2.807 | 2.914 | -0.345 | -0.631 ^c | 0.286 ^c | -0.512 | 0.167 |
| 83 | Methyl propionate | 0.041 | 2.807 | 2.914 | -0.390 | -0.631 ^c | 0.241 ^c | -0.512 | 0.122 |
| 84 | Ethyl propionate | 0.917 | 3.307 | 3.414 | -1.474 | -1.807 ^c | 0.333 ^c | -1.824 | 0.350 |
| 85 | Ethyl propionate | 0.917 | 3.307 | 3.414 | -1.666 | -1.807 ^c | 0.141 ^c | -1.824 | 0.158 |
| 86 | Isopropyl propionate | 1.439 | 3.662 | 3.914 | -2.970 | -2.584 ^c | -0.386 ^c | -2.746 | -0.224 |
| 87 | Propyl propionate | 1.417 | 3.807 | 3.914 | -3.086 | -2.982 ^c | -0.104 ^c | -3.009 | -0.076 |
| 88 | Propyl propionate | 1.417 | 3.807 | 3.914 | -2.992 | -2.982 ^c | -0.010 ^c | -3.009 | 0.018 |
| 89 | Butyl propionate | 1.917 | 4.307 | 4.414 | -4.305 | -4.157 ^c | -0.148 ^c | -4.194 | -0.111 |
| 90 | Isopentyl propionate | 2.272 | 4.662 | 4.914 | -5.088 | -4.934 ^c | -0.154 ^c | -5.060 | -0.028 |
| 91 | Pentyl propionate | 2.417 | 4.807 | 4.914 | -5.181 | -5.332 ^c | 0.150 ^c | -5.379 | 0.198 |
| 92 | Methyl butyrate | 0.541 | 3.307 | 3.414 | -1.945 | -1.807 ^c | -0.138 ^c | -1.697 | -0.248 |
| 93 | Methyl butyrate | 0.541 | 3.307 | 3.414 | -1.988 | -1.807 ^c | -0.181 ^c | -1.697 | -0.291 |
| 94 | Ethyl butyrate | 1.417 | 3.807 | 3.914 | -2.936 | -2.982 ^c | 0.046 ^c | -3.009 | 0.074 |
| 95 | Isopropyl butyrate | 1.939 | 4.162 | 4.414 | -4.465 | -3.759 ^c | -0.706 ^c | -3.931 | -0.535 |
| 96 | Propyl butyrate | 1.917 | 4.307 | 4.414 | -4.423 | -4.157 ^c | -0.266 ^c | -4.194 | -0.228 |
| 97 | Propyl butyrate | 1.917 | 4.307 | 4.414 | -4.390 | -4.157 ^c | -0.233 ^c | -4.194 | -0.196 |
| 98 | Ethyl heptanoate | 2.917 | 5.307 | 5.414 | -6.303 | -6.507 ^c | 0.204 ^c | -6.565 | 0.261 |
| 99 | Ethyl valerate | 1.917 | 4.307 | 4.414 | -4.069 | -4.157 ^c | 0.088 ^c | -4.194 | 0.125 |
| 100 | Ethyl hexanoate | 2.417 | 4.807 | 4.914 | -5.425 | -5.332 ^c | -0.093 ^c | -5.379 | -0.046 |
| 101 | Ethyl octanoate | 3.417 | 5.807 | 5.914 | -7.799 | -7.682 ^c | -0.117 ^c | -7.750 | -0.049 |
| 102 | Ethyl nonanoate | 3.917 | 6.307 | 6.414 | -8.741 | -8.857 ^c | 0.116 ^c | -8.935 | 0.194 |
| 103 | Ethyl decanoate | 4.417 | 6.807 | 6.914 | -9.434 | -10.032 ^c | 0.598 ^c | -10.120 | 0.686 |
| 104 | Dimethyl ether | -1.154 | 1.414 | 1.414 | 1.772 | 1.831 ^d | -0.060 ^d | 2.656 | -0.885 |
| 105 | Isopropyl methyl ether | 0.244 | 2.269 | 2.414 | -0.138 | -0.250 ^d | 0.112 ^d | 0.423 | -0.561 |
| 106 | Isopropyl methyl ether | 0.244 | 2.269 | 2.414 | -0.065 | -0.250 ^d | 0.185 ^d | 0.423 | -0.488 |
| 107 | Diethyl ether | 0.598 | 2.414 | 2.414 | -0.550 | -0.306 ^d | -0.244 ^d | 0.032 | -0.582 |
| 108 | Diethyl ether | 0.598 | 2.414 | 2.414 | -0.254 | -0.306 ^d | 0.052 ^d | 0.032 | -0.286 |

Table 3. (Continued)

| No | Compound Name | Descriptors | | | ln S | | | | |
|-----|-------------------------------|-------------|----------------------|----------------------|------------------|----------------------|---------------------|-------------------|------------------|
| | | <i>T</i> | <i>T_R</i> | <i>T_N</i> | Obs ^a | Calc | Res | Calc ^h | Res ^h |
| 109 | Methyl propyl ether | 0.222 | 2.414 | 2.414 | -0.620 | -0.744 ^d | 0.124 ^d | 0.159 | -0.779 |
| 110 | Methyl propyl ether | 0.222 | 2.414 | 2.414 | -0.877 | -0.744 ^d | -0.133 ^d | 0.159 | -1.036 |
| 111 | Ethyl isopropyl ether | 1.120 | 2.769 | 2.914 | -1.291 | -1.319 ^d | 0.028 ^d | -0.889 | -0.402 |
| 112 | Methyl isobutyl ether | 0.577 | 2.769 | 2.914 | -2.071 | -1.952 ^d | -0.119 ^d | -0.706 | -1.365 |
| 113 | Methyl-sec-butyl ether | 0.782 | 2.807 | 2.914 | -1.704 | -1.757 ^d | 0.054 ^d | -0.762 | -0.941 |
| 114 | Butyl methyl ether | 0.722 | 2.914 | 2.914 | -2.303 | -2.251 ^d | -0.051 ^d | -1.026 | -1.277 |
| 115 | Isopropyl propyl ether | 1.620 | 3.269 | 3.414 | -3.086 | -2.826 ^d | -0.260 ^d | -2.074 | -1.011 |
| 116 | Dipropyl ether | 1.598 | 3.414 | 3.414 | -3.364 | -3.320 ^d | -0.044 ^d | -2.338 | -1.026 |
| 117 | Dibutyl ether | 2.598 | 4.414 | 4.414 | -6.261 | -6.334 ^d | 0.074 ^d | -4.708 | -1.552 |
| 118 | Methyl- <i>t</i> -butyl ether | 0.634 | 2.561 | 2.914 | -0.484 | -0.484 ^d | 0.000 ^d | -0.300 | -0.184 |
| 119 | Ethyl propyl ether | 1.098 | 2.914 | 2.914 | -1.531 | -1.813 ^d | 0.282 ^d | -1.153 | -0.378 |
| 120 | Chloroethane | -0.488 | 1.414 | 1.414 | -2.420 | -1.861 ^e | -0.559 ^e | -3.356 | 0.936 |
| 121 | Chloropropane | 0.012 | 1.914 | 1.914 | -3.516 | -3.004 ^e | -0.512 ^e | -4.541 | 1.025 |
| 122 | 2-Chloropropane | 0.179 | 1.731 | 1.914 | -3.127 | -3.053 ^e | -0.074 ^e | -4.339 | 1.212 |
| 123 | Chlorobutane | 0.512 | 2.414 | 2.414 | -4.934 | -4.147 ^e | -0.787 ^e | -5.726 | 0.792 |
| 124 | Isobutyl chloride | 0.367 | 2.269 | 2.414 | -4.605 | -4.196 ^e | -0.409 ^e | -5.406 | 0.801 |
| 125 | 1,3-Dichloropropane | -1.390 | 2.414 | 2.414 | -3.716 | -4.147 ^e | 0.431 ^e | -5.084 | 1.368 |
| 126 | Chloroform | -2.926 | 1.731 | 1.914 | -2.118 | -3.053 ^e | 0.935 ^e | -3.290 | 1.172 |
| 127 | Bromoethane | -1.293 | 1.414 | 1.414 | -2.429 | -2.849 ^e | 0.420 ^e | -3.084 | 0.655 |
| 128 | Bromopropane | -0.793 | 1.914 | 1.914 | -3.990 | -3.992 ^e | 0.002 ^e | -4.269 | 0.279 |
| 129 | 2-Bromopropane | -0.478 | 1.731 | 1.914 | -3.756 | -4.041 ^e | 0.285 ^e | -4.117 | 0.361 |
| 130 | Bromobutane | -0.293 | 2.414 | 2.414 | -5.448 | -5.135 ^e | -0.313 ^e | -5.454 | 0.006 |
| 131 | Isobutyl bromide | -0.438 | 2.269 | 2.414 | -5.600 | -5.183 ^e | -0.416 ^e | -5.134 | -0.466 |
| 132 | Isoamyl bromide | 0.062 | 2.769 | 2.914 | -6.645 | -6.326 ^e | -0.319 ^e | -6.319 | -0.326 |
| 133 | 1,3-Dibromopropane | -3.000 | 2.414 | 2.414 | -4.792 | -5.135 ^e | 0.343 ^e | -4.540 | -0.252 |
| 134 | Iodomethane | -3.627 | 1.000 | 1.000 | -2.303 | -3.093 ^e | 0.790 ^e | -1.250 | -1.053 |
| 135 | Iodoethane | -1.858 | 1.414 | 1.414 | -3.684 | -4.236 ^e | 0.552 ^e | -2.893 | -0.791 |
| 136 | Iodopropane | -1.358 | 1.914 | 1.914 | -5.273 | -5.378 ^e | 0.105 ^e | -4.078 | -1.195 |
| 137 | Iodobutane | -0.858 | 2.414 | 2.414 | -6.816 | -6.521 ^e | -0.295 ^e | -5.263 | -1.553 |
| 138 | Diiodomethane | -5.130 | 1.414 | 1.414 | -5.388 | -4.236 ^e | -1.152 ^e | -1.788 | -3.600 |
| 139 | Dichloroethsulfide | -3.216 | 3.414 | 3.414 | -5.457 | -6.432 ^e | 0.975 ^e | -6.499 | 1.042 |
| 140 | <i>n</i> -Butane | 1.914 | 1.914 | 1.914 | -6.020 | -6.248 ^f | 0.228 ^f | -5.184 | -0.836 |
| 141 | Isobutane | 1.731 | 1.731 | 1.914 | -5.867 | -5.785 ^f | -0.082 ^f | -4.864 | -1.003 |
| 142 | <i>n</i> -Pentane | 2.414 | 2.414 | 2.414 | -7.530 | -7.595 ^f | 0.065 ^f | -6.369 | -1.161 |
| 143 | 2-Methyl butane | 2.269 | 2.269 | 2.414 | -7.322 | -7.228 ^f | -0.094 ^f | -6.049 | -1.273 |
| 144 | 2,2-Dimethyl propane | 2.000 | 2.000 | 2.414 | -7.198 | -6.548 ^f | -0.650 ^f | -5.554 | -1.644 |
| 145 | 2,2-Dimethyl butane | 2.561 | 2.561 | 2.914 | -8.450 | -8.049 ^f | -0.401 ^f | -6.739 | -1.711 |
| 146 | 2,4-Dimethyl pentane | 3.124 | 3.124 | 3.414 | -10.109 | -9.556 ^f | -0.553 ^f | -8.099 | -2.010 |
| 147 | 2,2,4-Trimethyl pentane | 3.416 | 3.416 | 3.914 | -9.501 | -10.376 ^f | 0.875 ^f | -8.789 | -0.712 |
| 148 | 2,2,5-Trimethyl hexane | 3.916 | 3.916 | 4.414 | -11.624 | -11.723 ^f | 0.099 ^f | -9.974 | -1.650 |
| 149 | Cyclohexane | 3.000 | 3.000 | 2.914 | -7.322 | -7.383 ^f | 0.061 ^f | -7.554 | 0.232 |
| 150 | Methylcyclohexane | 3.393 | 3.393 | 3.414 | -8.867 | -8.460 ^f | -0.407 ^f | -8.419 | -0.448 |
| 151 | 1,2-Dimethylcyclohexane | 3.803 | 3.803 | 3.914 | -9.830 | -9.579 ^f | -0.251 ^f | -9.284 | -0.546 |
| 152 | Cycloheptane | 3.500 | 3.500 | 3.414 | -8.095 | -8.730 ^f | 0.635 ^f | -8.739 | 0.644 |
| 153 | Cyclooctane | 4.000 | 4.000 | 3.914 | -9.560 | -10.077 ^f | 0.517 ^f | -9.924 | 0.364 |
| 154 | <i>n</i> -Hexane | 2.914 | 2.914 | 2.914 | -9.106 | -8.942 ^f | -0.164 ^f | -7.554 | -1.552 |
| 155 | <i>n</i> -Heptane | 3.414 | 3.414 | 3.414 | -10.438 | -10.289 ^f | -0.149 ^f | -8.739 | -1.699 |
| 156 | <i>n</i> -Octane | 3.914 | 3.914 | 3.914 | -12.059 | -11.636 ^f | -0.423 ^f | -9.924 | -2.135 |
| 157 | 2-Methyl pentane | 2.769 | 2.769 | 2.914 | -8.727 | -8.575 ^f | -0.152 ^f | -7.234 | -1.493 |
| 158 | 3-Methyl pentane | 2.807 | 2.807 | 2.914 | -8.819 | -8.672 ^f | -0.147 ^f | -7.234 | -1.585 |
| 159 | 2,2-Dimethylpentane | 3.061 | 3.061 | 3.414 | -8.450 | -9.396 ^f | 0.946 ^f | -7.924 | -0.526 |
| 160 | Cyclopentane | 2.500 | 2.500 | 2.414 | -6.102 | -6.036 ^f | -0.066 ^f | -6.369 | 0.267 |
| 161 | Methylcyclopentane | 2.893 | 2.893 | 2.914 | -7.599 | -7.113 ^f | -0.486 ^f | -7.234 | -0.365 |
| 162 | 1-Pentyne | 1.536 | 2.414 | 2.414 | -3.707 | -4.363 ^f | 0.656 ^f | -6.072 | 2.365 |

Table 3. (Continued)

| No | Compound Name | Descriptors | | | ln S | | | | |
|-----|--------------------------|-------------|----------------------|----------------------|------------------|----------------------|---------------------|-------------------|------------------|
| | | <i>T</i> | <i>T_R</i> | <i>T_N</i> | Obs ^a | Calc | Res | Calc ^h | Res ^h |
| 163 | 1-Hexyne | 2.036 | 2.914 | 2.914 | -5.434 | -5.710 ^f | 0.276 ^f | -7.257 | 1.823 |
| 164 | 1-Heptyne | 2.536 | 3.414 | 3.414 | -6.931 | -7.057 ^f | 0.126 ^f | -8.442 | 1.511 |
| 165 | 1-Octyne | 3.036 | 3.914 | 3.914 | -8.427 | -8.404 ^f | -0.023 ^f | -9.627 | 1.200 |
| 166 | 1-Nonyne | 3.536 | 4.414 | 4.414 | -9.694 | -9.752 ^f | 0.058 ^f | -10.813 | 1.119 |
| 167 | 1,8-Nonadiyne | 2.658 | 4.414 | 4.414 | -6.862 | -6.520 ^f | -0.342 ^f | -10.516 | 3.654 |
| 168 | 1,6-Heptadiyne | 1.658 | 3.414 | 3.414 | -4.030 | -3.825 ^f | -0.205 ^f | -8.146 | 4.116 |
| 169 | 1-Pentene | 1.679 | 2.414 | 2.414 | -6.148 | -6.228 ^f | 0.080 ^f | -6.120 | -0.028 |
| 170 | 2-Pentene | 1.572 | 2.414 | 2.414 | -5.849 | -6.228 ^f | 0.379 ^f | -6.084 | 0.235 |
| 171 | 1-Hexene | 2.179 | 2.914 | 2.914 | -7.437 | -7.575 ^f | 0.138 ^f | -7.306 | -0.131 |
| 172 | 2-Heptene | 2.572 | 3.414 | 3.414 | -8.796 | -8.922 ^f | 0.126 ^f | -8.455 | -0.341 |
| 173 | 1-Octene | 3.179 | 3.914 | 3.914 | -10.638 | -10.269 ^f | -0.369 ^f | -9.676 | -0.962 |
| 174 | 4-Methyl-1-pentene | 2.034 | 2.769 | 2.914 | -7.460 | -7.208 ^f | -0.252 ^f | -6.986 | -0.474 |
| 175 | 1,6-Heptadiene | 1.943 | 3.414 | 3.414 | -7.691 | -7.555 ^f | -0.136 ^f | -8.242 | 0.551 |
| 176 | 1,5-Hexadiene | 1.443 | 2.914 | 2.914 | -6.194 | -6.208 ^f | 0.014 ^f | -7.057 | 0.863 |
| 177 | 1,4-Pentadiene | 0.943 | 2.414 | 2.414 | -4.789 | -4.861 ^f | 0.072 ^f | -5.872 | 1.083 |
| 178 | Cyclopentene | 1.745 | 2.500 | 2.414 | -4.835 | -4.669 ^f | -0.166 ^f | -6.114 | 1.279 |
| 179 | Cyclohexene | 2.245 | 3.000 | 2.914 | -5.941 | -6.016 ^f | 0.075 ^f | -7.299 | 1.358 |
| 180 | Cycloheptene | 2.745 | 3.500 | 3.414 | -7.276 | -7.363 ^f | 0.087 ^f | -8.484 | 1.208 |
| 181 | 2-Butanone | 1.215 | 2.269 | 2.414 | 1.561 | 1.135 ^g | 0.426 ^g | 0.095 | 1.466 |
| 182 | 2-Pentanone | 1.715 | 2.769 | 2.914 | -0.389 | -0.319 ^g | -0.070 ^g | -1.090 | 0.701 |
| 183 | 3-Pentanone | 1.811 | 2.807 | 2.914 | -0.534 | -0.319 ^g | -0.215 ^g | -1.110 | 0.576 |
| 184 | 3-Methyl-2-butanone | 1.613 | 2.641 | 2.914 | -0.286 | -0.223 ^g | -0.063 ^g | -0.779 | 0.493 |
| 185 | 2-Hexanone | 2.215 | 3.269 | 3.414 | -1.794 | -1.772 ^g | -0.022 ^g | -2.275 | 0.481 |
| 186 | 3-Hexanone | 2.311 | 3.307 | 3.414 | -1.904 | -1.772 ^g | -0.132 ^g | -2.295 | 0.391 |
| 187 | 3-Methyl-2-pentanone | 2.151 | 3.179 | 3.414 | -1.545 | -1.676 ^g | 0.131 ^g | -1.964 | 0.419 |
| 188 | 4-Methyl-2-pentanone | 2.070 | 3.124 | 3.414 | -1.637 | -1.676 ^g | 0.039 ^g | -1.955 | 0.318 |
| 189 | 4-Methyl-3-pentanone | 2.209 | 3.179 | 3.414 | -1.870 | -1.676 ^g | -0.194 ^g | -1.984 | 0.114 |
| 190 | 2-Heptanone | 2.715 | 3.769 | 3.914 | -3.274 | -3.226 ^g | -0.048 ^g | -3.461 | 0.187 |
| 191 | 4-Heptanone | 2.811 | 3.807 | 3.914 | -3.325 | -3.226 ^g | -0.099 ^g | -3.480 | 0.155 |
| 192 | 2,4-Dimethyl-3-pentanone | 2.607 | 3.551 | 3.914 | -2.991 | -3.034 ^g | 0.043 ^g | -2.858 | -0.133 |
| 193 | 5-Nonanone | 3.811 | 4.807 | 4.914 | -5.929 | -6.132 ^g | 0.203 ^g | -5.850 | -0.079 |

Obs = Observed; Cal = Calculated; Res = Residual

^a Refs. [8] and [47]; ^b From Eq. (6); ^c From Eq. (10); ^d From Eq. (17); ^e From Eq. (20);

^f From Eq. (26); ^g From Eq. (29); ^h From Eq. (36).

3.1 QSPR for Alcohols (*n* = 60)

Table 4 shows QSPR models for the water solubility of alcohols with topochemical indices. The first order VEM composite topochemical index *T* explains 93.5% and predicts 93.2% of the variance while the skeletal index *T_R* has 97.2% explained variance and 97.1% predicted variance. When *T_R* was split into the branching parameter *B* and the vertex count *N_V*, the resultant equation showed 97.3% explained variance and 97.1% predicted variance. Similar relations were obtained by using shape parameters (*N_P*, or *N_B*, or *N_X* and *N_Y*) along with vertex count. Positive coefficients of *B* and various shape parameters indicate that water solubility increases with increase in branching. Negative coefficients of *N_V* and *T_R* are indicative of negative contributions of bulk and skeletal index to water solubility. The calculated water solubility values according to Eq. (6) are given in Table 3. 3,3-Dimethylbutanol (**19**), cyclohexanol (**24**) and 2-hexen-4-ol (**59**) act as outliers (but not excluded) for Eq. (6).

Table 4. Relations of water solubility (ln S) of alcohols with topochemical indices. Model equation, $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | | Statistics | | | |
|----|---|--------------------------------------|-------------------------------------|----------------|------------------|---------------------|---|---------------|
| | β_1 se | β_2 se | β_3 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 1 | -2.481 <i>T</i> 0.085 | | | 3.213 0.243 | 0.932 (0.834) | 0.935 (0.968) | 0.824 (8.51x10 ² (1, 58)) | 0.690 (60) |
| 2 | -2.599 <i>T_R</i> 0.057 | | | 6.574 0.226 | 0.971 (0.551) | 0.972 (0.986) | 0.541 (2.06x10 ³ (1, 58)) | 0.430 (60) |
| 3 | 2.902 <i>B</i> 0.452 | -1.293 <i>N_V</i> 0.030 | | 6.687 0.278 | 0.971 (0.549) | 0.973 (0.987) | 0.535 (1.05x10 ³ (2, 57)) | 0.417 (60) |
| 4 | -1.300 <i>N_V</i> 0.033 | 0.476 <i>N_B</i> 0.103 | | 6.766 0.320 | 0.963 (0.616) | 0.966 (0.983) | 0.599 (8.33x10 ² (2, 57)) | 0.493 (60) |
| 5 | -1.302 <i>N_V</i> 0.031 | 0.391 <i>N_P</i> 0.068 | | 5.999 0.370 | 0.967 (0.584) | 0.970 (0.985) | 0.560 (9.57x10 ² (2, 57)) | 0.442 (60) |
| 6 | -1.299 <i>N_V</i> 0.029 | 1.004 <i>N_X</i> 0.144 | 0.353 <i>N_Y</i> 0.092 | 6.706 0.260 | 0.972 (0.533) | 0.975 (0.988) | 0.514 (7.61x10 ² (3, 56)) | 0.399 (60) |

se = standard error; *F* values are significant at 99% level [df = np, n – np – i, np = no. of predictor variables; i = 1 if intercept is present; i = 0, otherwise]

[#] *t* values of the regression coefficients and constants are significant at 95% level [df = n – np – i]

3.2 QSPR for Esters (n = 43)

The QSPR models for the water solubility of esters with different topochemical indices are shown in Eqs. (7)–(11) (Table 5). The variation of water solubility of the esters can be explained to the extent of 97.4% (predicted variance 97.2%) by the first order composite topochemical index *T* while it was 98.2% for first order skeletal index *T_R* (predicted variance 98.0%). When *T* was split into different components, the resultant relations Eqs. (9)–(11) were similar in quality to that of the skeletal index. However, these relations show specific contributions of branching, shape factors and bulk. From the sign of the regression coefficients Eqs. (9)–(10), it is evident that water solubility of esters increase with increase in branching and decrease with increase of bulk. In Eq. (11), *N_B* shows negative coefficient as it is a component of *N_V* which has negative contribution to water solubility. The effects of branching or shape parameters will be evident in equations containing *N_V* term. The calculated solubility values according to Eq. (10) are given in Table 3. Isopropyl butyrate **95** acts as outlier (but not excluded) for Eq. (10).

Table 5. Relations of water solubility (ln S) of esters with topochemical indices. Model equation, $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | Statistics | | | |
|----|---|--------------------------------------|-----------------|------------------|---------------------|---|---------------|
| | β_1 se | β_2 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 7 | -2.128 <i>T</i> 0.054 | | -0.172 0.091 | 0.972 (0.410) | 0.974 (0.987) | 0.401 (1.57x10 ³ (1, 41)) | 0.310 (43) |
| 8 | -2.320 <i>T_R</i> 0.048 | | 5.808 0.187 | 0.980 (0.345) | 0.982 (0.991) | 0.332 (2.31x10 ³ (1, 41)) | 0.265 (43) |
| 9 | 2.747 <i>B</i> 0.635 | -1.163 <i>N_V</i> 0.025 | 5.975 0.199 | 0.979 (0.354) | 0.982 (0.991) | 0.334 (1.14x10 ³ (2, 40)) | 0.259 (43) |
| 10 | -1.175 <i>N_V</i> 0.025 | 0.397 <i>N_B</i> 0.091 | 6.022 0.197 | 0.979 (0.354) | 0.982 (0.991) | 0.334 (1.14x10 ³ (2, 40)) | 0.257 (43) |
| 11 | -1.175 <i>N_I</i> 0.025 | -1.953 <i>N_B</i> 0.089 | 3.672 0.155 | 0.979 (0.354) | 0.982 (0.991) | 0.334 (1.14x10 ³ (2, 40)) | 0.257 (43) |

3.3 QSPR for Ethers ($n = 16$)

Table 6 shows the QSPR models for the water solubility of ethers with different topochemical indices, Eqs. (12)–(17). The first order composite topochemical index T can explain 86.0% variance (predicted variance 81.8%) of the water solubility of ethers. On splitting of T into skeletal index T_R and functionality F , considerable rise in statistical quality (explained variance 98.6% and predicted variance 98.3%) was obtained. When T_R was further split into different shape and size terms, the resultant relations could explain up to 99.1% of the variance Eq. (17). However, due to insufficient occurrence of quaternary type vertex (N_X), leave-one-out could not be applied for Eq. (17). The relations showed positive impact of branching and shape factors and negative impact of functionality and bulk. Eq. (17) shows that the effect of quaternary type of vertex is more pronounced than that of the tertiary type (regression coefficient 1.458 for N_X vs. 0.299 for N_Y). The calculated water solubility values according to Eq. (17) are shown in Table 3, which shows that all ether compounds fitted well in this equation.

Table 6. Relations of water solubility (ln S) of ethers with topochemical indices. Model equation, $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | | | Statistics | | | |
|----|---|-----------------|-----------------|-----------------|----------------|----------------------------------|---------------------|--|---------------|
| | β_1 se | β_2 se | β_3 se | β_4 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 12 | -2.007 0.157 | | | | | 0.818 (0.746) | 0.860 (0.927) | 0.676 (1.63x10 ² (1, 15)) | 0.576 (16) |
| 13 | -3.011 0.105 | -1.129 0.277 | | | 8.984 0.755 | 0.983 (0.226) | 0.986 (0.994) | 0.211 (5.45 x10 ² (2, 13)) | 0.144 (16) |
| 14 | -1.063 0.295 | 3.429 0.557 | -1.495 0.055 | | 9.025 0.668 | 0.974 (0.283) | 0.986 (0.994) | 0.214 (3.52x10 ² (3, 12)) | 0.140 (16) |
| 15 | -1.131 0.484 | -1.496 0.090 | 0.466 0.181 | | 9.216 1.098 | 0.955 (0.371) | 0.962 (0.985) | 0.350 (1.29x10 ² (3, 12)) | 0.193 (16) |
| 16 | -1.078 0.340 | -1.497 0.063 | 0.519 0.103 | | 8.020 0.816 | 0.967 (0.316) | 0.981 (0.992) | 0.247 (2.63x10 ² (3, 12)) | 0.160 (16) |
| 17 | -1.167 0.236 | -1.507 0.044 | 1.458 0.181 | 0.299 0.092 | 9.349 0.535 | — ⁺ — ⁺ | 0.991 (0.997) | 0.171 (4.17x10 ² (4, 11)) | 0.114 (16) |

⁺ LOO could not be applied because of insufficient occurrence of N_X type vertex.

3.4 QSPR for Halocarbons ($n = 20$)

Eqs. (18)–(20) describing QSPR for water solubility of halocarbons with topochemical indices are given in Table 7, showing that only 45.0% of the variance could be explained by the first order skeletal index (predicted variance 38.1%). The composite topochemical index gave further inferior relation. However, by incorporating two indicator variables I_{Br} and I_I (indicating the presence or absence of Br and I, respectively) along with integer variables, statistically acceptable relation (explained variance 78.4%, predicted variance 66.8%) was obtained. The regression coefficients of I_{Br} and I_I suggest that presence of Br and I decreases water solubility. Further, increase in branching (as shown by the coefficient of N_P) also reduces water solubility. All compounds fitted well in Eq. (20). The calculated water solubility values according to Eq. (20) are shown in Table 3.

Table 7. Relations of water solubility $\ln S$ of halocarbons with topochemical indices. Model equation, $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | | | Statistics | | | |
|----|---|-----------------|-----------------|-----------------|----------------|-----------------|---------------------|--------------------------------|--------------|
| | β_1 se | β_2 se | β_3 se | β_4 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 18 | -2.100 T_R | | | | | 0.381 | 0.450 | 1.048 | 0.789 |
| | 0.112 | | | | | (1.084) | (0.671) | (3.52x10 ² (1, 19)) | (20) |
| 19 | -1.247 N_P | -0.844 N_I | | | | 0.321 | 0.445 | 1.053 | 0.798 |
| | 0.144 | 0.157 | | | | (1.134) | (0.688) | (1.75x10 ² (2, 18)) | (20) |
| 20 | -2.334 N_P | -1.143 N_I | -0.988 I_{Br} | -2.374 I_I | 3.950 | 0.668 | 0.784 | 0.657 | 0.484 |
| | 0.441 | 0.144 | 0.341 | 0.432 | 1.070 | (0.794) | (0.911) | (18.200 (4, 15)) | (20) |

I_{Br} and I_I denote presence or absence of bromine and iodine respectively.

3.5 QSPR for Hydrocarbons ($n = 41$)

Table 8 shows that variance of water solubility of hydrocarbons is explained to the extent of 75.2% (predicted variance 73.0%) by the composite topochemical index. When the composite index was split into T_R and F , there was a slight reduction in quality. In the case of hydrocarbons, F signifies the presence of multiple bonds (electron richness). The relation involving F and N_V explained to the extent of 78.4% (predicted variance 75.7%). On introduction of the indicator variable I_{cyclo} (indicating the presence or absence of cyclicity) and integer variables N_{trip} and N_{doub} (indicating the number of triple and double bonds, respectively), the QSPR quality increases up to 96.1% of explained variance, Eq. (26). These relations show that the water solubility of hydrocarbons increases with branching, cyclicity and presence of multiple bonds while it decreases with an increase in bulk. The effect of triple bond is more pronounced than that of double bond (regression coefficient 3.232 for triple bond vs. 1.367 for double bond in Eq. (26)). 2,2,4-Trimethylpentane (**147**) and 2,2-dimethylpentane (**159**) act as outliers (but not excluded) for Eq. (26). The calculated solubility values according to Eq. (26) are given in Table 3.

Table 8. Relations of water solubility ($\ln S$) of hydrocarbons with topochemical indices. Equation $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | | | Statistics | | | | |
|----|---|-----------------|-------------------|------------------|------------------|----------------|-----------------|---------------------|--------------------------------|--------------|
| | β_1 se | β_2 se | β_3 se | β_4 se | β_5 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 21 | -2.210 T | | | | | -1.980 | 0.730 | 0.752 | 0.976 | 0.836 |
| | 0.200 | | | | | 0.542 | (1.006) | (0.871) | (1.22x10 ² (1, 39)) | (41) |
| 22 | -2.153 T_R | 2.295 F | | | | -2.195 | 0.717 | 0.746 | 0.987 | 0.824 |
| | 0.242 | 0.282 | | | | 0.742 | (1.030) | (0.871) | (59.867 (2, 38)) | (41) |
| 23 | 2.074 F | -1.094 N_V | | | | -1.655 | 0.757 | 0.784 | 0.910 | 0.749 |
| | 0.256 | 0.110 | | | | 0.715 | (0.954) | (0.892) | (73.637 (2, 38)) | (41) |
| 24 | 2.280 F | -1.114 N_V | 1.061 I_{Cyclo} | | | -1.881 | 0.811 | 0.834 | 0.797 | 0.590 |
| | 0.232 | 0.096 | 0.300 | | | 0.641 | (0.841) | (0.920) | (68.153 (3, 37)) | (41) |
| 25 | -1.352 N_V | 0.381 N_P | 2.244 I_{Cyclo} | 3.212 N_{trip} | 1.352 N_{doub} | -1.559 | 0.947 | 0.957 | 0.407 | 0.284 |
| | 0.055 | 0.091 | 0.278 | 0.161 | 0.122 | 0.417 | (0.425) | (0.981) | (1.78x10 ² (5, 35)) | (41) |
| 26 | 2.531 B | -1.347 N_V | 1.559 I_{Cyclo} | 3.232 N_{trip} | 1.367 N_{doub} | -0.860 | 0.952 | 0.961 | 0.385 | 0.269 |
| | 0.518 | 0.051 | 0.158 | 0.151 | 0.115 | 0.346 | (0.425) | (0.983) | (2.00x10 ² (5, 35)) | (41) |

I_{cyclo} denotes presence or absence of cyclicity. N_{doub} and N_{trip} denote number of double and triple bonds respectively

3.6 QSPR for Ketones ($n = 13$)

The equations Eqs. (27)–(30) relating water solubility of ketones with topochemical indices are shown in Table 9. The first order composite topochemical index can explain 98.5% of the variance (predicted variance 97.1%) while the first order skeletal index explained 98.1% of the variance (predicted variance 96.5%). With integer variables (N_I and N_P or N_B), relations explaining 98.9% of variance (predicted variance 97.9%) were obtained. The calculated solubility values according to Eq. (29) are shown in Table 3. 2–Butanone (**181**) behaves as an outlier (but not excluded) for Eq. (29).

Table 9. Relations of water solubility (ln S) of ketones with topochemical indices. Model equation, $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | Statistics | | | |
|----|---|-----------------------|-----------------|------------------|---------------------|--|---------------|
| | β_1 se | β_2 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 27 | -2.789 T 0.100 | | 4.436 0.235 | 0.971 (0.299) | 0.985 (0.993) | 0.227 (7.71 x10 ² (1, 11)) | 0.163 (13) |
| 28 | -2.846 T_R 0.114 | | 7.460 0.380 | 0.965 (0.331) | 0.981 (0.991) | 0.252 (6.20x10 ² (1, 11)) | 0.165 (13) |
| 29 | -1.453 N_I 0.044 | -2.811 N_B 0.118 | 5.399 0.255 | 0.979 (0.257) | 0.989 (0.995) | 0.191 (5.44x10 ² (2, 10)) | 0.130 (13) |
| 30 | -1.453 N_I 0.044 | -2.811 N_P 0.118 | 11.021 0.484 | 0.979 (0.257) | 0.989 (0.995) | 0.191 (5.44x10 ² (2, 10)) | 0.130 (13) |

3.7 QSPR for All Compounds ($n = 193$)

Table 10 lists the relations of water solubility values of all compounds (composite set) with topochemical indices.

Table 10. Relations of water solubility (ln S) of the composite set with topochemical indices. Equation $\ln S = \sum \beta_i x_i + \alpha$

| Eq | Regression coefficient(s) and constant [#] | | | | | Statistics | | | | |
|----|---|-----------------------|----------------------|----------------------|----------------------------|-----------------|------------------|---------------------|--|----------------|
| | β_1 se | β_2 se | β_3 se | β_4 se | β_5 se | α se | Q^2 (SDEP) | R_a^2 (r or R) | s (F (df)) | AVRES (n) |
| 31 | -1.089 T 0.118 | | | | | -2.086 0.279 | 0.280 (2.727) | 0.305 (0.556) | 2.686 (85.296 (1, 191)) | 2.249 (193) |
| 32 | -1.519 T_R 0.085 | 0.693 F 0.153 | | | | | 0.302 (2.684) | 0.325 (0.573) | 2.648 (2.61x10 ² (2, 191)) | 2.265 (193) |
| 33 | 0.938 F 0.171 | -0.766 N_V 0.083 | 2.143 N_X 0.540 | 1.591 N_Y 0.256 | | -1.446 0.695 | 0.392 (2.506) | 0.422 (0.659) | 2.449 (36.106 (4, 188)) | 1.976 (193) |
| 34 | 0.790 F 0.165 | -0.764 N_V 0.082 | 1.162 N_P 0.167 | | | -3.137 0.733 | 0.404 (2.480) | 0.430 (0.663) | 2.432 (49.300 (3, 189)) | 1.940 (193) |
| 35 | 0.901 F 0.167 | -0.767 N_V 0.083 | 1.660 N_B 0.246 | | | -1.368 0.653 | 0.397 (2.496) | 0.422 (0.657) | 2.449 (47.803 (3, 189)) | 1.991 (193) |
| 36 | 0.338 F 0.067 | -1.185 N_V 0.033 | 0.815 N_X 0.205 | 0.320 N_Y 0.103 | -5.787 I_{HC-X} 0.170 | 5.344 0.282 | 0.914 (0.943) | 0.919 (0.960) | 0.916 (4.38x10 ² (5, 187)) | 0.650 (193) |
| 37 | 0.404 F 0.057 | -1.205 N_V 0.027 | 0.922 N_X 0.166 | 0.404 N_Y 0.084 | -5.794 I_{HC-X} 0.142 | 5.294 0.232 | 0.944 (0.750) | 0.946 (0.974) | 0.735 (6.59x10 ² (5, 181)) | 0.567 (187) |

I_{HC-X} denotes hydrocarbon or halocarbon compounds

The composite topochemical index T singularly explained only 30.5% of the variance (predicted variance 28.0%). When T was partitioned into T_R and F , the relation could explain 32.5% of the variance (predicted variance 30.2%). On further partitioning of T_R , different relations, Eqs. (33)–

(35), showing about 42% explained variance (about 40% predicted variance) were obtained. When an indicator variable I_{HC-X} (denoting hydrocarbon or halocarbon compounds) was included, there was considerable rise in the statistical quality: both explained variance and predicted variances increased by 50 units.

Table 11. Results of leave-many-out cross-validation applied on Eq. (36). Model equation, $\ln S = \sum \beta_i x_i + \alpha$

| Type of cross-validation | Number of cycles | Average regression coefficients (standard deviations) | Q^2 (Av. Pres) |
|--------------------------|------------------|---|---------------------|
| Leave-10%-out | 10 ^a | 0.338 (0.036) F -1.184 (0.009) N_V +0.815 (0.068) N_X + 0.320 (0.030) N_Y -5.787 (0.056) I_{HC-X} +5.342 (0.106) | 0.917 (0.664) |
| Leave-25%-out | 4 ^b | 0.340 (0.067) F -1.186 (0.018) N_V +0.812 (0.096) N_X + 0.321 (0.036) N_Y -5.783 (0.124) I_{HC-X} +5.340 (0.198) | 0.916 (0.665) |

Q^2 denotes cross-validated R^2 . Average Pres means average of absolute values of predicted residuals

^a Compounds were deleted in 10 cycles in the following manner: (1, 11, 21, ..., 191), (2, 12, 22, ..., 192), ..., (10, 20, 30, ..., 190)

^b Compounds were deleted in 4 cycles in the following manner: (1, 5, 9, ..., 193), (2, 6, 10, ..., 190), ..., (4, 8, 12, ..., 192)

The calculated water solubility values according to the Eq. (36) are given in Table 3. Leave-many-out cross-validation (leave-10%-out and leave-25%-out) has been applied on Eq. (36) and the results are given in Table 11, which shows the robustness of Eq. (36). The scatter plot of observed vs. calculated, Eq. (36), water solubility values are shown in Figure 1. Diidomethane **138**, 2,4-dimethylpentane **146**, *n*-octane **156**, 1-pentyne **162**, 1,8-nonadiyne **167** and 1,6-heptadiyne **168** act as outliers (but not excluded) for Eq. (36). When these outliers were excluded, a much better QSPR (explained variance 94.6% and predicted variance 94.4%) was obtained. The relations show that water solubility increases with functionality and branching (as evidenced from the coefficients of F and shape factors) while it decreases with an increase in bulk. Further, water solubility is specifically smaller for hydrocarbons and halocarbon compounds (as evidenced from large negative coefficient of I_{HC-X}).

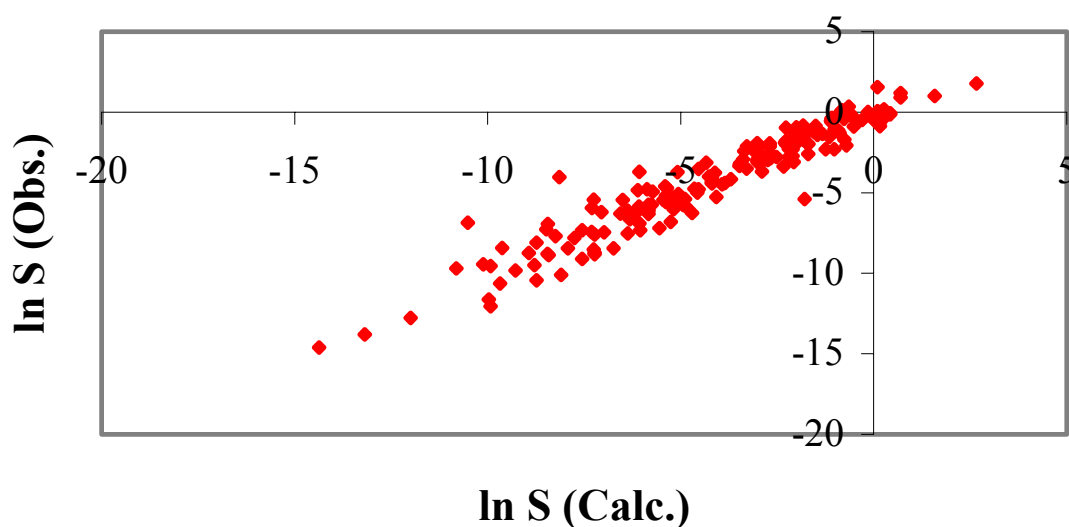


Figure 1. Scatter plot of observed vs. calculated values Eq. (36) of water solubility.

4 CONCLUSIONS

This study shows that although the first order composite topochemical index T does not always provide acceptable QSPR models for water solubility of heterofunctional compounds, the TAU scheme can generate statistically acceptable relations when the first order composite index is partitioned into different components, such as skeletal index, size and shape factors, branching and functionality. For 193 heterofunctional compounds, the TAU descriptors, along with indicator variables, can explain up to 91.9% and predict up to 91.4% of the variance of water solubility. Moreover, TAU indices can decode specific contributions of molecular bulk (size), functionality, branching and shape parameters to the water solubility of diverse functional compounds. In general, water solubility increases with an increase in functionality and branching, and decreases with an increase in molecular bulk. Moreover, halocarbons and hydrocarbons show less water solubility for reasons not explained by the used topochemical parameters. The study suggests that the TAU scheme merits further assessment in justifying its usefulness in QSPR/QSAR studies.

Acknowledgment

The authors acknowledge Sri Dipak Kumar Pal for constant help and inspiration. A financial grant from J. U. Research Fund is also thankfully acknowledged.

5 REFERENCES

- [1] P. J. Hansen and P. C. Jurs, Chemical Applications of Graph Theory, *J. Chem. Educ.* **1988**, *65*, 574–580.
- [2] D. H. A. Rouvray, Rationale for the Topological Approaches in Chemistry, *J. Mol. Struct. (Theochem)* **1995**, *336*, 101–104.
- [3] S. C. Basak and B. D. Gute, Characterization of Molecular Structures Using Topological Indices, *SAR QSAR Environ. Res.* **1997**, *7*, 1–21.
- [4] J. Devillers and A. T. Balaban, Eds., *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, Netherlands, 1999.
- [5] I. Rios–Santamarina, R. García–Domenech, J. Cortijo, P. Santamaria, E. J. Morcillo, and J. Gálvez, Natural Compounds with Bronchodilator Activity Selected by Molecular Topology, *Internet Electron. J. Mol. Des.* **2002**, *1*, 70–79, <http://www.biochempress.com>.
- [6] E. Estrada, The Structural Interpretation of the Randić Index, *Internet Electron. J. Mol. Des.* **2002**, *1*, 360–366, <http://www.biochempress.com>.
- [7] B. D. Gute, S. C. Basak, D. Mills, and D. M. Hawkins, Tailored Similarity Spaces for the Prediction of Physicochemical Properties, *Internet Electron. J. Mol. Des.* **2002**, *1*, 374–387, <http://www.biochempress.com>.
- [8] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- [9] H. Wiener, Structural Determination of Paraffin Boiling Point, *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- [10] L. B. Kier and L. H. Hall, General Definition of Valence δ Values for Molecular Connectivity, *J. Pharm. Sci.* **1983**, *72*, 1170–1173.
- [11] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, England, 1986.
- [12] A. T. Balaban, C. Catana, M. Dawson, and I. Niculescu–Duvaz, Application of Weighted Topological Index J for QSAR of Carcinogenesis Inhibitors (Retinoic Acid Derivatives), *Rev. Roum. Chim.* **1990**, *35*, 997–1003.
- [13] L. B. Kier, An index of Flexibility from Molecular Shape Descriptors, *Prog. Clin. Biol. Res.* **1989**, *291*, 105–109.
- [14] L. B. Kier and L. H. Hall, An Electrotopological State Index for Atoms in Molecules, *Pharm. Res.* **1990**, *7*, 801–807.
- [15] L. H. Hall, B. K. Mohny, and L. B. Kier, Comparison of Electrotopological State Index with Molecular Orbital Parameters: Inhibition of MAO by Hydrazides, *Quant. Struct.–Act. Relat.* **1993**, *12*, 44–48.

- [16] F. Torrens, A New Topological Index to Elucidate Apolar Hydrocarbons, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 709–719.
- [17] H. Patel and M. T. Cronin, A Novel Index for the Description of Molecular Linearity, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1228–1236.
- [18] P. V. Khadikar, S. Karmakar, S. Singh, and A. Shrivastava, Use of PI index in Predicting Toxicity of Nitrobenzene Derivatives, *Bioorg. Med. Chem.* **2002**, *10*, 3163–3170.
- [19] B. Ren, Novel Atomic-Level-Based AI topological Descriptors: Application to QSPR/QSAR Modeling, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 858–868.
- [20] E. Estrada and E. Molina, Novel Local (Fragment-Based) Topological Molecular Descriptors for QSPR/QSAR and Molecular Design, *J. Mol. Graphics Mod.* **2001**, *50*, 54–64.
- [21] G. W. Kauffman and P. C. Jurs, QSAR and K-Nearest Neighbour Classification Analysis of Selective Cyclooxygenase-2 Inhibitors Using Topologically-Based Numerical Descriptors, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1553–1560.
- [22] R. Garcia-Domenech, J. V. de Julian-Ortiz, M. J. Duart, J. M. Garcia-Torrecillas, G. M. Anton-Fos, I. Rios-Santamarina, C. de Gregorio-Alapont, and J. Galvez, Search of A Topological Pattern to Evaluate Toxicity of Heterogeneous Compounds, *SAR QSAR Environ. Res.* **2001**, *12*, 237–254.
- [23] G. A. Bakken and P. C. Jurs, Classification of Multidrug-Resistance Reversal Agents Using Structure-Based Descriptors and Linear Discriminant Analysis, *J. Med. Chem.* **2000**, *43*, 4534–4541.
- [24] H. Kubinyi, In *Burger's Medicinal Chemistry and Drug Discovery*, M. E. Wolff, Ed., 5th ed., vol. I, John Wiley & Sons, N. Y., 1995, pp. 497–571.
- [25] A. K. Ghose and G. M. Crippen, Atomic Physicochemical Parameters for Three Dimensional-Structure-Directed Quantitative Structure-Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions, *J. Chem. Inf. Comput. Sci.*, **1987**, *27*, 21–35.
- [26] A. K. Ghose, V. N. Viswanadhan, and J. J. Wendoloski, Prediction of Hydrophobic Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods, *J. Phys. Chem.* **1998**, *102*, 3762–3772.
- [27] N. Bodor, Z. Gabanyi, and C. -K. Wong, A New Method for The Estimation of Partition Coefficient, *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786.
- [28] G. Klopman and S. Wang, A Computer Automated Structure Evaluation (CASE) Approach to the Calculation of Partition Coefficients. *J. Comput. Chem.* **1991**, *12*, 1025–1032.
- [29] I. Moriguchi, S. Hirono, Q. Liu, I. Nakagome, and Y. Matsushita, Simple Method of Calculating Octanol/Water Partition Coefficient, *Chem. Pharm. Bull. (Tokyo)* **1992**, *40*, 127–130.
- [30] A. K. Saxena, Physicochemical Significance of Topological Parameters: Molecular Connectivity Index and Information Content. Part 2. Correlation Studies with Molar Refractivity and Lipophilicity, *Quant. Struct.-Act. Relat.* **1995**, *14*, 142–150.
- [31] L. Z. Benet, D. L. Kroetz, and L. B. Sheiner, In *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, J. G. Hardman, L. E. Limbard, P. B. Molinoff, R. W. Ruddon and A. Goodman Gilman, Eds., McGraw-Hill, New York, 1996, pp. 3–27.
- [32] M. M. Ferreira, Polycyclic Aromatic Hydrocarbons: A QSPR Study. Quantitative Structure-Property Relationships, *Chemosphere* **2001**, *44*, 125–146.
- [33] P. V. Khadikar, F. Mandloi, A. V. Bajaj, and S. Joshi, QSAR Study on Solubility of Alkanes in Water and Their Partition Coefficients in Different Solvent Systems Using PI Index, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 419–422.
- [34] A. R. Katritzky, Y. Wang, S. Sild, and T. Tamm, QSPR Studies on Vapour Pressure, Aqueous Solubility, and the Prediction of Water-Air Partition Coefficients, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720–725.
- [35] C. Yin, X. Liu, W. Guo, T. Lin, X. Wang, and L. Wang, Prediction and Application in QSPR of Aqueous Solubility of Sulfur-Containing Aromatic Esters Using GA-Based MLR with Quantum Descriptors, *Water Res.* **2002**, *36*, 2975–2982.
- [36] X. Q. Chen, S. J. Cho, Y. Li, and S. Venkatesh, Prediction of Aqueous Solubility of Organic Compounds Using A Quantitative Structure-Property Relationship, *J. Pharm. Sci.* **2002**, *91*, 1838–1852.
- [37] S. Wanchana, F. Yamashita, and M. Hashida, Quantitative Structure/Property Relationship Analysis on Aqueous Solubility Using Genetic Algorithm-Combined Partial Least Squares Method, *Pharmazie* **2002**, *57*, 127–129.
- [38] H. Gao, V. Shanmugasundaram, and P. Lee, Estimation of Aqueous Solubility of Organic Compounds with QSPR Approach, *Pharm. Res.* **2002**, *19*, 497–503.
- [39] S. Puri, J. S. Chickos, and W. J. Welsh, Three Dimensional Quantitative Structure-Property Relationship (3D-QSPR) Model for Prediction of Thermodynamic Properties of Polychlorinated Biphenyls (PCBs): Enthalpies of Fusion and Their Application to Estimates of Enthalpies of Sublimation and Aqueous Solubilities, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 55–62.
- [40] R. Liu and S. S. So, Development of Quantitative Structure-Property Relationship Models for Early ADME Evaluation in Drug Discovery. 1. Aqueous Solubility, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1633–1639.

- [41] D. K. Pal, C. Sengupta, and A. U. De, A New Topochemical Descriptor (TAU) in Molecular Connectivity Concept: Part I – Aliphatic Compounds, *Indian J. Chem.* **1988**, *27B*, 734–739.
- [42] D. K. Pal, C. Sengupta, and A. U. De, Introduction of A Novel Topochemical Index and Exploitation of Group Connectivity Concept to Achieve Predictability in QSAR and RDD, *Indian J. Chem.* **1989**, *28B*, 261–267.
- [43] K. Roy, D. K. Pal, A. U. De, and C. Sengupta, Comparative QSAR with Molecular Negentropy, Molecular Connectivity, STIMS and TAU Indices: Part I. Tadpole Narcosis of Diverse Functional Acyclic Compounds, *Indian J. Chem.* **1999**, *38B*, 664–671.
- [44] K. Roy, D. K. Pal, A. U. De, and C. Sengupta, Comparative QSAR Studies with Molecular Negentropy, Molecular Connectivity, STIMS and TAU Indices. Part II: General Anaesthetic Activity of Aliphatic Hydrocarbons, Halocarbons and Ethers, *Indian J. Chem.* **2001**, *40B*, 129–135.
- [45] K. Roy and A. Saha, Comparative QSPR Studies with Molecular Connectivity, Molecular Negentropy and TAU Indices. Part I: Molecular Thermochemical Properties of Diverse Functional Acyclic Compounds, *J. Mol. Model.* **2003**, *9*, in press.
- [46] K. Roy and A. Saha, Comparative QSPR Studies with Molecular Connectivity, Molecular Negentropy and TAU Indices. Part 2. Lipid–Water Partition Coefficient of Diverse Functional Acyclic Compounds, *Internet Electron. J. Mol. Des.* **2003**, *2*, 288–305, <http://www.biochempress.com>.
- [47] C. Hansch, J. E. Quinlan, and G. L. Lawrence, The Linear Free–Energy Relationship Between Partition Coefficient and the Aqueous Solubility of Organic Liquids, *J. Org. Chem.* **1968**, *33*, 347–350.
- [48] D. K. Pal, M. Sengupta, C. Sengupta, and A. U. De, QSAR with TAU (τ) Indices: Part I – Polymethylene Primary Diamines as Amebicidal Agents, *Indian J. Chem.* **1990**, *29B*, 451–454.
- [49] D. K. Pal, S. K. Purkayastha, C. Sengupta, and A. U. De, Quantitative Structure–Property Relationships with TAU indices: Part I – Research Octane Numbers of Alkane Fuel Molecules, *Indian J. Chem.* **1992**, *31B*, 109–114.
- [50] The GW–BASIC programs *RRR98* (multiple regression), *KRPRES1* and *KRPRES2* (PRESS statistics) were developed by Kunal Roy (1998) and standardized using known data sets.
- [51] G. W. Snedecor and W. G. Cochran, *Statistical Methods*; Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi, 1967, pp. 381–418.
- [52] S. Wold and L. Eriksson, In *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, Ed., VCH, Weinheim, 1995, p. 312.