Inter*net* Electronic Journal of Molecular Design

May 2003, Volume 2, Number 5, Pages 288–305

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Haruo Hosoya on the occasion of the 65$^{th}$ birthday
Part 9

Guest Editor: Jun–ichi Aihara

# Comparative QSPR Studies with Molecular Connectivity, Molecular Negentropy and TAU Indices. Part 2. Lipid–Water Partition Coefficient of Diverse Functional Acyclic Compounds

Kunal Roy[1] and Achintya Saha[2]

[1] Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Calcutta 700032, India
[2] Department of Chemical Technology, University of Calcutta, 92 A P C Road, Calcutta 700009, India

Received: September 19, 2002; Revised: November 30, 2002; Accepted: January 29, 2003; Published: May 31, 2003

**Citation of the article:**
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.

**Inter*net* Electronic Journal of Molecular Design**

**BioChem Press**
http://www.biochempress.com

# Comparative QSPR Studies with Molecular Connectivity, Molecular Negentropy and TAU Indices. Part 2. Lipid–Water Partition Coefficient of Diverse Functional Acyclic Compounds[#]

Kunal Roy[1,*] and Achintya Saha[2]

[1] Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Calcutta 700032, India
[2] Department of Chemical Technology, University of Calcutta, 92 A P C Road, Calcutta 700009, India

Received: September 19, 2002; Revised: November 30, 2002; Accepted: January 29, 2003; Published: May 31, 2003

*Internet Electron. J. Mol. Des.* **2003**, *2* (5), 288–305

**Abstract**

**Motivation**. In late eighties, topochemically arrived unique (TAU) scheme was described by Pal *et al*. in valence electron mobile (VEM) environment and this index was claimed to have power to decode chemical information in somewhat better way than molecular connectivity index (MCI). The TAU scheme is unique in that it unravels specific contributions of functionality, branching, shape and size factors to the physicochemical property or biological activity while other indices give mainly a global contribution of the molecule. Subsequently, several papers described QSAR and QSPR with TAU index to show the diagnostic potential of the parameter, but much work has not been done on the index. Thus, a comparison among the relations involving these indices may explore relative suitability of the schemes in describing physicochemical parameters.

**Method**. The present communication attempts to correlate lipid–water partition coefficient (log *P*) of 168 diverse functional acyclic compounds with TAU indices and to compare those with relations involving molecular negentropy (*I*) and first order valence molecular connectivity ($^1\chi^v$) indices to explore the diagnostic feature of TAU scheme.

**Results**. This study shows that TAU indices can unravel specific contributions of molecular bulk (size), functionality, branching and shape parameters to the lipophilicity of diverse functional compounds. In general, lipophilicity increases with increase in molecular bulk and skeletal index value, and decreases with increase in branching and functionality.

**Conclusions**. The TAU index is an important tool in exploring structure–property relationship studies in view of its potential to unravel specific contribution 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; molecular connectivity index; structural descriptor; topological index; molecular graph; molecular negentropy; octanol–water partition coefficient; log *P*.

[#] Dedicated to Professor Haruo Hosoya on the occasion of the 65$^{th}$ 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

Physico–chemical properties of organic compounds change in a systematic way with change in chemical structure, which actually determine the type and magnitude of intramolecular and intermolecular interactions [1]. The chemical information coded in the two dimensional structural representation (topology) are decoded to some extent by assigning a value (index) to a compound or fragment of a compound considering features like adjacency, branching, unsaturation, hetero–atom variation, cyclicity, perturbation effects etc [2–18]. Such indices have been found to be helpful in exploring quantitative structure–property, structure–activity and activity–property relationship studies [5–18]. Quantitative relations generated from such studies help in hypothesizing important contributions of specific structural aspects or chemical interactions in modifying physicochemical properties and biological activities and also in predicting properties or activities of untested (and even not synthesized) compounds.

Among the topological indices, molecular connectivity index of Kier and Hall has received maximum popularity [19, 20]. It has been correlated with several physicochemical properties and biological activities of diverse organic compounds. In late eighties, topochemically arrived unique (TAU) scheme was described by Pal *et al*. [21, 22] in valence electron mobile (VEM) environment and this index was claimed to have power to decode chemical information in somewhat better way than molecular connectivity index (MCI). Subsequently, a few papers [23–27] described QSAR and QSPR with TAU index to show the diagnostic potential of the parameter, but much work has not been done on the index. TAU scheme is unique in that it unravels specific contributions of functionality, branching, shape and size factors to the physicochemical property or biological activity while other indices give mainly a global contribution of the molecule. Thus, a comparison among the relations involving these indices may explore relative suitability of the schemes in describing physicochemical parameters.

The octanol–water partition coefficient (log *P*) representing the overall lipophilicity of a molecule has been extensively used in QSAR and QSPR models as hydrophobicity parameter [28]. Since experimental determination of partition coefficient values of large set of compounds is a tedious job, there are many approaches of calculating log *P* values, *e.g*., Ghose and Crippen's atom contribution method [29, 30], Bodor's quantum chemical method [31], Klopman's Multi–CASE method [32], Moriguchi's method [33] etc. Topological schemes also have been used to model log *P* values [34, 35]. The present communication attempts to correlate lipid–water partition coefficient (log *P*) of 168 diverse functional acyclic compounds (51 alcohols, 25 amines, 25 carboxylic acids and esters, 12 ethers, 6 halocarbons, 34 hydrocarbons and 15 ketones) with TAU indices and to compare those with relations involving molecular negentropy (*I*) [36] and first order valence molecular connectivity ($^1\chi^v$) indices to explore the diagnostic feature of TAU scheme.

## 2 MATERIALS AND METHODS

The physicochemical parameter (log *P*) values were taken from the literature [37,38]. Molecular connectivity [1,19,20] and molecular negentropy [36,39,40] values were calculated according to the original references. Molecular negentropy (MN) is based on the information theory of Shannon [39,40] and obtained from total molecular graph. Atoms (vertices) of a molecular graph are partitioned into disjoint subsets of equivalent vertices and MN is obtained from the formula:

$$I = -N\sum_j p_j \log_{10} p_j \tag{1}$$

where $N$ is the total number of vertices in total molecular graph and $p$ is the probability of random selection of an element (vertex) belonging to subset $j$ members of which are equivalent. The value of $p$ is equal to $n_j / N$, $n_j$ being the number of elements of subset $j$. Alternatively, MN can be conveniently calculated from the following:

$$I = N\log_{10} N - \sum n_j \log_{10} n_j \tag{2}$$

Molecular connectivity indices are obtained from hydrogen suppressed graphs of molecules and can be calculated according to the method of Kier and Hall [1,19,20]. $^1\chi$ and $^1\chi^v$ are two important and simplest molecular connectivity indices. These are defined as follows:

$$^1\chi = \sum_{e=1}^{N_e}[(\delta_i\delta_j)_e]^{-0.5} \tag{3}$$

where $\delta_i$ and $\delta_j$ are adjacency counts (number of atoms joined to the particular atom in the hydrogen–suppressed graph) of the vertices $i$ and $j$ joining the edge (bond) $e$ and $N_e$ is the total number of edges

$$^1\chi^v = \sum_{e=1}^{N_e}[(\delta_i^v\delta_j^v)_e]^{-0.5} \tag{4}$$

where $\delta_i^v$ and $\delta_j^v$ are valence δ values of the $i^{th}$ and $j^{th}$ vertices forming the edge $e$:

$$\delta_i^v = (Z^v - h)/(Z - Z^v - 1) \tag{5}$$

In Eq. (5), $Z^v$ is the number of valence electrons, $Z$ is the total number of electrons and $h$ is the number of hydrogens attached to the atom. TAU [21–27] are *T*opochemically *A*rrived *U*nique indices developed in VEM (valence electron, mobile) environment. These include $T$ (composite topochemical index), $T_R$ (skeletal index), $F$ (functionality index) and $B$ (simple branching index). The topochemical composite index ($T$) is defined as:

$$T = \sum_{i<j} E_{ij} = \sum_{i<j}(V_iV_j)^{0.5} \tag{6}$$

where $E_{ij}$ is VEM edge weight of the edge between $i^{th}$ and $j^{th}$ vertices. $V_i$ represents VEM vertex

weight of the $i^{th}$ vertex, which may be calculated as the ratio of core count of the $i^{th}$ vertex $\lambda_i$ to VEM count of the $i^{th}$ vertex $\theta_i$. $\lambda_i$ may be calculated as $(Z - Z^v) / Z^v$ whereas $\theta_i$ may be calculated as $8 - (2h + 1.5\nu + n)$. When unsaturation is present, $\theta_i$ should be calculated as $0.5\nu + 2\pi$. The notations $\nu$, $n$ and $\pi$ represent the numbers of sigma bonds (other than hydrogen), nonbonded electrons and pi bonds associated with the atom in that order.

In case of a heteroatom, VEM edge weight of edge incident upon the heteroatom is assigned a negative value. The skeletal index $T_R$ is the topochemical index of the reference alkane that may be obtained by replacing heteroatom with carbon and removing the multiple bonds that may be present. The derived indices $F$ and $B$ are easily calculated as $T_R - T$ and $T_N - T_R$ respectively where $T_N$ is topochemical index of the corresponding normal alkane (for acyclic molecules).

STIMS are Simplest Topological Integers from Molecular Structures, that were derived by Pal *et al*. [23–26], as a subset of TAU to obtain an easy tool for predicting various properties and activities of simple molecules directly from the molecular graphs of their reference alkanes. These include $N_P$ (number of methyl carbons), $N_I$ (number of methylene carbons), $N_Y$ (number of tertiary carbons), $N_X$ (number of quaternary carbons) and $N_B$ (number of branched carbons).

The vertex count ($N_V$) of the hydrogen–suppressed molecular formula is purely a constitutional parameter because it may be obtained directly from the molecular formula. Even 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 molecular index $T_R$ is considered as the index for lipophilicity while $N_B$, $N_X$ and $N_Y$ represent shape parameters. The functionality contribution and bulk parameter are represented by $F$ and $N_V$ respectively [21]. All TAU indices are basically derived by sequentially partitioning the composite index $T$ into different factors. $T$ may initially be factored into two components, $T_R$ (skeletal index) and $F$ (functionality). Subsequently, $T_R$ may be partitioned into $B$ (branching) and $N_V$ (bulk). $N_V$ can be partitioned into $N_P$, $N_I$ and $N_B$. $N_B$ may further be factored into $N_X$ and $N_Y$. During development of QSAR equations with TAU parameters, these hierarchical relations were followed. For obvious reasons, $B$ and $N_B$ (both represent branching) or $N_P$ and $N_B$ (both have interrelation) [24] or $N_V$ and $N_I$ ($N_I$ may be considered as trimmed counterpart of $N_V$) [24] were not used in the same equation.

Mutiple linear regression analyses were done using a software program *RRR98* developed by one of the authors [41]. Statistical quality of the equations [42] was judged by examining the parameters like $R_a^2$ (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) and *AVRES* (average of absolute values of residuals). Significance of the regression coefficients was judged by the $t$ test. In case that intercept of an equation was statistically insignificant and omission of the same did not affect the

quality of the equation, 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. The robustness of the best equations under different series was checked with leave–one–out (LOO) technique [43,44] using programs *KRPRES1* and *KRPRES2* [41]. Two LOO parameters, $Q^2$ (crossvalidation $R^2$ or predicted variance) and SDEP (standard deviation of error of predictions), were used to compare the equations.

**Table 1.** Topological indices of diverse functional aliphatic compounds

| No | Compound Name | $^1\chi^v$ | $I$ | $T$ | $T_R$ | $T_N$ |
|---|---|---|---|---|---|---|
| **1** | Methanol | 0.447 | 3.238 | –0.817 | 1.000 | 1.000 |
| **2** | Ethanol | 1.023 | 6.555 | 0.130 | 1.414 | 1.414 |
| **3** | *n*–Propanol | 1.523 | 10.315 | 0.630 | 1.914 | 1.914 |
| **4** | *n*–Butanol | 2.023 | 14.404 | 1.130 | 2.414 | 2.414 |
| **5** | *n*–Pentanol | 2.523 | 18.755 | 1.630 | 2.914 | 2.914 |
| **6** | *n*–Hexanol | 3.023 | 23.325 | 2.130 | 3.414 | 3.414 |
| **7** | *n*–Heptanol | 3.523 | 28.081 | 2.630 | 3.914 | 3.914 |
| **8** | *n*–Octanol | 4.023 | 33.001 | 3.130 | 4.414 | 4.414 |
| **9** | *n*–Nonanol | 4.523 | 38.006 | 3.630 | 4.914 | 4.914 |
| **10** | Isopropanol | 1.413 | 7.679 | 0.683 | 1.731 | 1.914 |
| **11** | Isobutanol | 1.879 | 11.768 | 0.985 | 2.270 | 2.414 |
| **12** | *tert*–Butanol | 1.724 | 7.622 | 1.092 | 2.000 | 2.414 |
| **13** | Isopentanol | 2.379 | 16.120 | 1.485 | 2.769 | 2.914 |
| **14** | 2–Methyl butanol | 2.417 | 18.528 | 1.523 | 2.807 | 2.914 |
| **15** | 1–Methyl butanol | 2.451 | 18.528 | 1.721 | 2.769 | 2.914 |
| **16** | 3–Pentanol | 2.489 | 14.314 | 1.759 | 2.807 | 2.914 |
| **17** | 3–Methyl–2–butanol | 2.324 | 15.893 | 1.593 | 2.641 | 2.914 |
| **18** | 2–Methyl–2–butanol | 2.284 | 15.291 | 1.652 | 2.561 | 2.914 |
| **19** | 2,2–Dimethyl–1–propanol | 2.170 | 11.973 | 1.276 | 2.561 | 2.914 |
| **20** | 2–Hexanol | 2.951 | 23.098 | 2.221 | 3.269 | 3.414 |
| **21** | 3–Hexanol | 2.989 | 23.098 | 2.259 | 3.307 | 3.414 |
| **22** | 3–Methyl–3–pentanol | 2.845 | 18.054 | 2.213 | 3.121 | 3.414 |
| **23** | 2–Methyl–2–pentanol | 2.784 | 19.860 | 2.152 | 3.061 | 3.414 |
| **24** | 2–Methyl–3–pentanol | 2.862 | 20.462 | 2.131 | 3.179 | 3.414 |
| **25** | 3–Methyl–2–pentanol | 2.862 | 22.870 | 2.131 | 3.179 | 3.414 |
| **26** | 4–Methyl–2–pentanol | 2.807 | 20.462 | 2.076 | 3.124 | 3.414 |
| **27** | 2,3–Dimethyl–2–butanol | 2.667 | 17.225 | 2.034 | 2.943 | 3.414 |
| **28** | 3,3–Dimethyl–1–butanol | 2.670 | 16.543 | 1.776 | 3.061 | 3.414 |
| **29** | 3,3–Dimethyl–2–butanol | 2.624 | 16.316 | 1.894 | 2.943 | 3.414 |
| **30** | 2–Methyl–2–hexanol | 3.284 | 24.617 | 2.652 | 3.561 | 3.914 |
| **31** | 3–Methyl–3–hexanol | 3.345 | 27.025 | 2.713 | 3.621 | 3.914 |
| **32** | 3–Ethyl–3–pentanol | 3.406 | 17.005 | 2.774 | 3.682 | 3.914 |
| **33** | 2,3–Dimethyl–2–pentanol | 3.205 | 24.389 | 2.572 | 3.481 | 3.914 |
| **34** | 2,3–Dimethyl–3–pentanol | 3.228 | 24.389 | 2.595 | 3.503 | 3.914 |
| **35** | 2,4–Dimethyl–2–pentanol | 3.140 | 21.981 | 2.507 | 3.416 | 3.914 |
| **36** | 2,4–Dimethyl–3–pentanol | 3.234 | 16.563 | 2.503 | 3.551 | 3.914 |
| **37** | 2,2–Dimethyl–3–pentanol | 3.162 | 21.072 | 2.432 | 3.481 | 3.914 |
| **38** | 2,2,3–Trimethyl–3–pentanol | 3.534 | 25.162 | 2.902 | 3.811 | 4.414 |
| **39** | Cyclohexanol | 2.575 | 17.674 | 2.345 | 3.393 | 3.414 |
| **40** | 4–Penten–1–ol | 2.133 | 16.858 | 0.894 | 2.914 | 2.914 |
| **41** | 3–Penten–2–ol | 2.080 | 16.403 | 0.917 | 2.769 | 2.914 |
| **42** | 1–Penten–3–ol | 2.115 | 16.630 | 1.062 | 2.807 | 2.914 |
| **43** | 1–Hexen–3–ol | 2.615 | 21.059 | 1.562 | 3.307 | 3.414 |
| **44** | 2–Hexen–4–ol | 2.618 | 20.832 | 1.455 | 3.307 | 3.414 |

**Table 1.** (Continued)

| No | Compound Name | $^1\chi^v$ | $I$ | $T$ | $T_R$ | $T_N$ |
|---|---|---|---|---|---|---|
| **45** | 2–Methyl–4–penten–3–ol | 2.488 | 18.423 | 1.434 | 3.179 | 3.414 |
| **46** | 2,2,2–Trifluoroethanol | 0.103 | 6.555 | –1.357 | 2.561 | 2.914 |
| **47** | Ethylene glycol | 1.132 | 5.786 | –0.655 | 1.914 | 1.914 |
| **48** | Allyl alcohol | 1.133 | 8.796 | –0.106 | 1.914 | 1.914 |
| **49** | *sec*–Butanol | 1.951 | 14.177 | 1.221 | 2.269 | 2.414 |
| **50** | 2,3–Butanediol | 2.004 | 11.587 | 0.545 | 2.641 | 2.914 |
| **51** | 1–Ethynyl cyclohexanol | 2.969 | 21.144 | 2.562 | 3.471 | 4.414 |
| **52** | Methyl amine | 0.577 | 3.882 | –0.894 | 1.000 | 1.000 |
| **53** | Ethyl amine | 1.115 | 7.365 | 0.075 | 1.414 | 1.414 |
| **54** | *n*–Propyl amine | 1.615 | 11.244 | 0.575 | 1.914 | 1.914 |
| **55** | *n*–Butyl amine | 2.115 | 15.426 | 1.075 | 2.414 | 2.414 |
| **56** | *n*–Pentyl amine | 2.615 | 19.855 | 1.575 | 2.914 | 2.914 |
| **57** | *n*–Hexyl amine | 3.115 | 24.490 | 2.075 | 3.414 | 3.414 |
| **58** | *n*–Heptyl amine | 3.615 | 29.303 | 2.575 | 3.914 | 3.914 |
| **59** | Isobutyl amine | 1.971 | 12.791 | 0.930 | 2.269 | 2.414 |
| **60** | *sec*–Butyl amine | 2.026 | 15.199 | 1.176 | 2.269 | 2.414 |
| **61** | 2–Amino octane | 4.026 | 34.045 | 3.176 | 4.269 | 4.414 |
| **62** | Cyclohexyl amine | 2.650 | 18.796 | 2.300 | 3.393 | 3.414 |
| **63** | Isopropyl amine | 1.488 | 8.608 | 0.638 | 1.731 | 1.914 |
| **64** | Methyl ethylamine | 1.561 | 11.016 | –0.373 | 1.914 | 1.914 |
| **65** | Di–*n*–propyl amine | 3.121 | 18.242 | 1.520 | 3.414 | 3.414 |
| **66** | Triethylamine | 3.070 | 13.413 | 1.025 | 3.345 | 3.414 |
| **67** | Di–*n*–butyl amine | 4.121 | 26.219 | 2.520 | 4.414 | 4.414 |
| **68** | Diethylamine | 2.121 | 10.985 | 0.520 | 2.414 | 2.414 |
| **69** | *n*–Propyl–*n*–butyl amine | 3.621 | 29.075 | 2.020 | 3.914 | 3.914 |
| **70** | Methyl–*n*–butyl amine | 2.561 | 19.627 | 0.627 | 2.914 | 2.914 |
| **71** | Piperidine | 2.207 | 14.295 | 1.106 | 3.000 | 2.914 |
| **72** | Ethyl–isopropyl amine | 2.504 | 16.992 | 1.049 | 2.769 | 2.914 |
| **73** | *n*–Propyl–*sec*–butylamine | 3.542 | 28.848 | 2.087 | 3.807 | 3.914 |
| **74** | *n*–Propyl–isobutylamine | 3.477 | 26.440 | 1.875 | 3.769 | 3.914 |
| **75** | Trimethylamine | 1.342 | 4.462 | –1.550 | 1.731 | 1.914 |
| **76** | Dimethyl–*n*–butylamine | 2.918 | 21.025 | 0.308 | 3.269 | 3.414 |
| **77** | Acetic acid | 0.928 | 5.793 | –0.069 | 1.731 | 1.914 |
| **78** | Propionic acid | 1.488 | 9.422 | 0.528 | 2.269 | 2.414 |
| **79** | Butyric acid | 1.988 | 13.410 | 1.028 | 2.769 | 2.914 |
| **80** | Hexanoic acid | 2.988 | 22.181 | 2.028 | 3.769 | 3.914 |
| **81** | Decanoic acid | 4.988 | 41.917 | 4.028 | 5.769 | 5.914 |
| **82** | Ethyl formate | 1.467 | 9.422 | –0.086 | 2.414 | 2.414 |
| **83** | *n*–Propyl formate | 1.967 | 13.410 | 0.414 | 2.914 | 2.914 |
| **84** | Methyl acetate | 1.316 | 8.593 | –0.555 | 2.269 | 2.414 |
| **85** | Ethyl acetate | 1.904 | 12.581 | 0.321 | 2.769 | 2.914 |
| **86** | *n*–Propyl acetate | 2.404 | 16.851 | 0.821 | 3.269 | 3.414 |
| **87** | Isopropyl acetate | 2.299 | 14.215 | 0.843 | 3.124 | 3.414 |
| **88** | *n*–Butyl acetate | 2.904 | 21.352 | 1.321 | 3.769 | 3.914 |
| **89** | *sec*–Butyl acetate | 2.837 | 21.124 | 1.381 | 3.662 | 3.914 |
| **90** | Methyl propionate | 1.877 | 12.581 | 0.041 | 2.807 | 2.914 |
| **91** | Methyl butyrate | 2.377 | 16.851 | 0.541 | 3.307 | 3.414 |
| **92** | Ethyl hexanoate | 3.965 | 30.916 | 2.417 | 4.807 | 4.914 |
| **93** | Ethyl heptanoate | 4.465 | 35.934 | 2.917 | 5.307 | 5.414 |
| **94** | Ethyl octanonate | 4.965 | 41.088 | 3.417 | 5.807 | 5.914 |
| **95** | Ethyl nonanonate | 5.465 | 46.363 | 3.917 | 6.307 | 6.414 |
| **96** | Ethyl decanonate | 5.965 | 51.751 | 4.417 | 6.807 | 6.914 |
| **97** | Ethyl propionate | 2.465 | 16.851 | 0.917 | 3.307 | 3.414 |
| **98** | Ethyl butyrate | 2.965 | 21.352 | 1.417 | 3.807 | 3.914 |
| **99** | Ethyl isobutyrate | 2.847 | 21.124 | 1.315 | 3.679 | 3.914 |
| **100** | Pentyl acetate | 3.404 | 26.049 | 1.821 | 4.269 | 4.414 |

**Table 1.** (Continued)

| No | Compound Name | $^1\chi^v$ | $I$ | $T$ | $T_R$ | $T_N$ |
|---|---|---|---|---|---|---|
| **101** | Butyl pentanoate | 4.465 | 35.934 | 2.917 | 5.307 | 5.414 |
| **102** | Diethyl ether | 1.992 | 9.360 | 0.598 | 2.414 | 2.414 |
| **103** | Methyl butyl ether | 2.404 | 17.926 | 0.722 | 2.914 | 2.914 |
| **104** | Methyl–sec–butyl ether | 2.337 | 17.699 | 0.782 | 2.807 | 2.914 |
| **105** | Methyl isobutyl ether | 2.260 | 15.291 | 0.577 | 2.769 | 2.914 |
| **106** | Methyl tert–butyl ether | 2.112 | 11.144 | 0.634 | 2.561 | 2.914 |
| **107** | Ethyl propyl ether | 2.492 | 17.926 | 1.098 | 2.914 | 2.914 |
| **108** | Ethyl isopropyl ether | 2.386 | 15.291 | 1.120 | 2.769 | 2.914 |
| **109** | Di–*n*–propyl ether | 2.992 | 16.475 | 1.598 | 3.414 | 3.414 |
| **110** | *n*–Propyl isopropyl ether | 2.886 | 19.860 | 1.620 | 3.269 | 3.414 |
| **111** | Methyl *n*–propyl ether | 1.904 | 13.575 | 0.222 | 2.414 | 2.414 |
| **112** | Methyl isopropyl ether | 1.799 | 10.939 | 0.244 | 2.269 | 2.414 |
| **113** | Ethyl cyclopropyl ether | 2.048 | 14.222 | 1.282 | 2.931 | 2.914 |
| **114** | Chloroform | 1.974 | 2.063 | –2.926 | 1.731 | 1.914 |
| **115** | Methyl iodide | 2.582 | 2.063 | –3.627 | 1.000 | 1.000 |
| **116** | Ethyl iodide | 2.533 | 5.191 | –1.858 | 1.414 | 1.414 |
| **117** | 1–Bromopropane | 2.621 | 8.820 | –0.793 | 1.914 | 1.914 |
| **118** | 1–Chlorobutane | 2.513 | 12.808 | 0.512 | 2.414 | 2.414 |
| **119** | 1–Fluoropentane | 1.940 | 17.078 | 1.673 | 2.914 | 2.914 |
| **120** | *n*–Pentane | 2.414 | 12.034 | 2.414 | 2.414 | 2.414 |
| **121** | 2–Methylbutane | 2.270 | 13.613 | 2.269 | 2.269 | 2.414 |
| **122** | 2–Methylpentane | 2.770 | 18.114 | 2.769 | 2.769 | 2.914 |
| **123** | 3–Methylpentane | 2.808 | 16.308 | 2.807 | 2.807 | 2.914 |
| **124** | *n*–Hexane | 2.914 | 14.729 | 2.914 | 2.914 | 2.914 |
| **125** | *n*–Heptane | 3.414 | 19.426 | 3.414 | 3.414 | 3.414 |
| **126** | 2,4–Dimethyl pentane | 3.126 | 14.155 | 3.124 | 3.124 | 3.414 |
| **127** | *n*–Octane | 3.914 | 22.487 | 3.914 | 3.914 | 3.914 |
| **128** | Cyclopentane | 2.000 | 4.147 | 2.500 | 2.500 | 2.414 |
| **129** | Cyclohexane | 2.500 | 4.976 | 3.000 | 3.000 | 2.914 |
| **130** | Methylcyclopentane | 2.394 | 15.143 | 2.893 | 2.893 | 2.914 |
| **131** | Cycloheptane | 3.000 | 5.805 | 3.500 | 3.500 | 3.414 |
| **132** | Methylcyclohexane | 2.894 | 19.713 | 3.393 | 3.393 | 3.414 |
| **133** | Cyclooctane | 3.500 | 6.634 | 4.000 | 4.000 | 3.914 |
| **134** | 1,2–Dimethylcyclohexane | 3.305 | 20.629 | 3.803 | 3.803 | 3.914 |
| **135** | 1–Pentene | 2.024 | 14.404 | 1.679 | 2.414 | 2.414 |
| **136** | 2–Pentene | 2.026 | 14.177 | 1.572 | 2.414 | 2.414 |
| **137** | 1–Hexene | 2.524 | 18.755 | 2.179 | 2.914 | 2.914 |
| **138** | 2–Heptene | 3.026 | 23.098 | 2.572 | 3.414 | 3.414 |
| **139** | 1–Octene | 3.524 | 28.081 | 3.179 | 3.914 | 3.914 |
| **140** | 4–Methyl–1–pentene | 2.379 | 16.120 | 2.034 | 2.769 | 2.914 |
| **141** | 1,6–Heptadiene | 2.633 | 16.470 | 1.943 | 3.414 | 3.414 |
| **142** | 1,5–Hexadiene | 2.133 | 12.041 | 1.443 | 2.914 | 2.914 |
| **143** | 1,4–Pentdiene | 1.633 | 9.665 | 0.943 | 2.414 | 2.414 |
| **144** | Cyclopentene | 1.650 | 9.665 | 1.745 | 2.500 | 2.414 |
| **145** | Cyclohexene | 2.150 | 12.041 | 2.245 | 3.000 | 2.914 |
| **146** | Cycloheptene | 2.650 | 16.470 | 2.745 | 3.500 | 3.414 |
| **147** | 1–Pentyne | 1.849 | 1.846 | 1.536 | 2.414 | 2.414 |
| **148** | 1–Hexyne | 2.349 | 16.028 | 2.036 | 2.914 | 2.914 |
| **149** | 1–Heptyne | 2.849 | 20.457 | 2.536 | 3.414 | 3.414 |
| **150** | 1–Octyne | 3.349 | 25.092 | 3.036 | 3.914 | 3.914 |
| **151** | 1–Nonyne | 3.849 | 29.905 | 3.536 | 4.414 | 4.414 |
| **152** | 1,8–Nonadiyne | 3.284 | 19.338 | 2.658 | 4.414 | 4.414 |
| **153** | 1,6–Heptadiyne | 2.284 | 12.223 | 1.658 | 3.414 | 3.414 |
| **154** | Acetone | 1.204 | 4.729 | 0.618 | 1.731 | 1.914 |
| **155** | 2–Butanone | 1.765 | 11.016 | 1.215 | 2.269 | 2.414 |

**Table 1.** (Continued)

| No | Compound Name | $^1\chi^v$ | $I$ | $T$ | $T_R$ | $T_N$ |
|---|---|---|---|---|---|---|
| **156** | 2–Pentanone | 2.265 | 15.199 | 1.715 | 2.769 | 2.914 |
| **157** | 3–Pentanone | 2.325 | 10.985 | 1.811 | 2.807 | 2.914 |
| **158** | 3–Methyl–2–butanone | 2.147 | 12.564 | 1.613 | 2.641 | 2.914 |
| **159** | 2–Hexanone | 2.765 | 19.627 | 2.215 | 3.269 | 3.414 |
| **160** | 3–Hexanone | 2.825 | 19.627 | 2.311 | 3.307 | 3.414 |
| **161** | 3–Methyl–2–petanone | 2.686 | 19.400 | 2.151 | 3.179 | 3.414 |
| **162** | 4–Methyl–2–petanone | 2.621 | 16.992 | 2.070 | 3.124 | 3.414 |
| **163** | 2–Methyl–3–petanone | 2.708 | 16.992 | 2.209 | 3.179 | 3.414 |
| **164** | 2–Heptanone | 3.265 | 24.262 | 2.715 | 3.769 | 3.914 |
| **165** | 3–Heptanone | 3.325 | 24.262 | 2.811 | 3.807 | 3.914 |
| **166** | 2,4–Dimethyl–3–pentanone | 3.091 | 12.971 | 2.607 | 3.551 | 3.914 |
| **167** | 5–Nonanone | 4.325 | 26.219 | 3.811 | 4.807 | 4.914 |
| **168** | 2–Nonanone | 4.265 | 34.045 | 3.715 | 4.769 | 4.914 |

# 3 RESULTS AND DISCUSSION

The calculated topological indices of 168 compounds are given in Table 1. Tables 2–9 show relations of hydrophobicity (log *P*) with different topological indices. All regression coefficients and variance ratios of the reported equations are significant at 95% and 99% levels respectively unless otherwise stated (marked with *). Table 10 shows the literature log *P* values of the compounds [37,38] and also the calculated values according to the best equations of individual series and the composite set (*vide* foot note of Table 10).

**Table 2.** Relations of hydrophobicity (log *P*) of alcohols with various indices. Model equation, $\log P = \sum\beta_i x_i + \alpha$

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ se | $\beta_2$ se | $\beta_3$ se | $\beta_4$ se | $\alpha$ se | $Q^2$ (SDEP) | $R_a^2$ (r or R) | s (F (df)) | AVRES (n) |
| 7 | MCI | 0.907 $^1\chi^v$ | | | | –1.118 | 0.709 | 0.752 | 0.435 | 0.199 |
| | | 0.073 | | | | 0.194 | (0.467) | (0.870) | (152.827 (1, 49)) | (51) |
| 8 | MN(I) | 0.112 *I* | | | | –0.843 | 0.738 | 0.759 | 0.429 | 0.252 |
| | | 0.009 | | | | 0.169 | (0.443) | (0.874) | (158.762 (1, 49)) | (51) |
| 9 | TAU | 0.755 *T* | | | | –0.097 | 0.711 | 0.764 | 0.424 | 0.273 |
| | | 0.059 | | | | 0.115 | (0.466) | (0.877) | (163.039 (1, 49)) | (51) |
| 10 | TAU | 0.973 $T_R$ | –0.416 *F* | | | –1.186 | 0.744 | 0.815 | 0.376 | 0.203 |
| | | 0.078 | 0.103 | | | 0.303 | (0.438) | (0.907) | (111.313 (2, 48)) | (51) |
| 11 | TAU | –0.436 *F* | –1.162 *B* | 0.489 $N_V$ | | –1.209 | 0.699 | 0.814 | 0.377 | 0.190 |
| | | 0.107 | 0.304 | 0.039 | | 0.281 | (0.475) | (0.908) | (73.713 (3, 47)) | (51) |
| 12 | TAU | –0.434 *F* | 0.475 $N_V$ | –0.292 $N_B$ | | –1.075 | 0.735 | 0.812 | 0.379 | 0.233 |
| | | 0.108 | 0.038 | 0.078 | | 0.280 | (0.446) | (0.907) | (72.886 (3, 47)) | (51) |
| 13 | TAU | –0.409 *F* | 0.463 $N_V$ | –0.144 $N_P$ | | –0.857 | 0.692 | 0.786 | 0.404 | 0.217 |
| | | 0.114 | 0.041 | 0.056 | | 0.314 | (0.480) | (0.894) | (62.253 (3, 47)) | (51) |
| 14 | TAU | –0.445 *F* | 0.488 $N_V$ | –0.403 $N_X$ | –0.274 $N_Y$ | –1.116 | 0.689 | 0.813 | 0.378 | 0.202 |
| | | 0.108 | 0.040 | 0.125 | 0.079 | 0.312 | (0.483) | (0.910) | (55.308 (4, 46)) | (51) |

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]

## 3.1 QSPR of Alochols (*n* = 51)

Eqs. (7)–(14) (Table 2) show the relations of log *P* of alcohols with different indices. First order valence molecular connectivity and molecular negentropy could explain the variance of the data set

to the same extent (75.2% and 75.9% respectively) while the latter could predict the variance (73.8%) better than the former (70.9%). TAU indices, Eqs. (9)–(14), could explain up to 81.5% of the variance vis–a–vis predict up to 74.4% of variance and show specific contributions of carbon skeleton ($T_R$), functionality ($F$), bulk ($N_V$), branching ($B$) and shape parameters ($N_X$ and $N_Y$). The relations show positive impact of $T_R$ and bulk ($N_V$) and negative impact of functionality, branching and shape parameters. Ethylene glycol and 2,3–butanediol show outlier behavior in all the models. Though equation 10 is of the best statistical quality, Eq. (14) carries more information about specific contributions of different shape and size factors, and hence, it has been used to calculate the log $P$ values of the alcohols as reported in Table 10.

**Table 3.** Relations of hydrophobicity (log $P$) of amines with various indices. Model equation, log $P = \sum\beta_i x_i + \alpha$

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ se | $\beta_2$ se | $\beta_3$ se | $\beta_4$ se | $\alpha$ se | $Q^2$ (SDEP) | $R_a^2$ (r or R) | s (F (df)) | AVRES (n) |
| 15 | MCI | 0.943 $^1\chi^v$ | | | | –1.167 | 0.960 | 0.964 | 0.171 | 0.136 |
| | | 0.037 | | | | 0.100 | (0.177) | (0.983) | (645.042 (1, 23)) | (25) |
| 16 | MN(I) | 0.104 $I$ | | | | –0.628 | 0.878 | 0.894 | 0.295 | 0.217 |
| | | 0.007 | | | | 0142 | (0.310) | (0.948) | (202.840 (1, 23)) | (25) |
| 17 | TAU | 0.702 $T$ | | | | 0.425 | 0.672 | 0.734 | 0.466 | 0.371 |
| | | 0.086 | | | | 0.134 | (0.507) | (0.863) | (67.307 (1, 23)) | (25) |
| 18 | TAU | 0.951 $T_R$ | | | | –1.512 | 0.942 | 0.951 | 0.200 | 0.154 |
| | | 0.044 | | | | 0.132 | (0.206) | (0.976) | (468.424 (1, 23)) | (25) |
| 19 | TAU | 0.479 $N_V$ | –0.144* $N_B$ | | | –1.600 | 0.952 | 0.958 | 0.184 | 0.143 |
| | | 0.020 | 0.075 | | | 0.127 | (0.194) | (0.981) | (277.630 (2, 22)) | (25) |
| 20 | TAU | 0.420 $N_I$ | 0.717 $N_Y$ | | | –0.481 | 0.788 | 0.825 | 0.379 | 0.228 |
| | | 0.039 | 0.161 | | | 0.183 | (0.408) | (0.916) | (57.383 (2, 22)) | (25) |

## 3.2 QSPR of Amines ($n$ = 25)

The relations of log $P$ of amines with different indices are shown in Eqs. (15)–(20) (Table 3). The variation of lipophilicity of the amines could be explained to the extent of 96.4% by first order valence connectivity index (predicted variance 96.0%) while it was only 89.4% for molecular negentropy (predicted variance 87.8%). In case of TAU indices, the composite topochemical index $T$ could explain 73.4% (predicted variance 67.2%) while skeletal index $T_R$ could explain 95.1% of the variance (predicted variance 94.6%). The functionality index $F$ did not show any importance. Further, $T_R$ was partitioned into size and shape factors and the best relation obtained, Eq. (19), could explain 95.8% of the variance (predicted variance 95.2%). However, the regression coefficient of $N_B$ in Eq. (19) was significant at 90% level. Piperidine showed outlier behavior for this equation. For amines, bulk and $T_R$ show positive impact and branching shows negative impact on lipophilicity.

## 3.3 QSPR of Carboxylic Acids and Esters ($n$ = 25)

Table 4 shows the relations of log $P$ of carboxylic acids and esters with different indices, Eqs. (21)–(29). First order molecular connectivity index could explain 99.6% variance (predicted

variance 99.5%) while molecular negentropy explained 99.4% variance (predicted variance 99.3%). Under the TAU scheme, composite index *T* showed somewhat inferior relation which could explain 96.3% of the variance (predicted variance 95.9%). However, when the composite index was partitioned into skeletal, functionality, shape and size factors, the relations could explain up to 99.7% of the variance (predicted variance up to 99.6%). The relations showed positive impact of skeletal index $T_R$ and bulk ($N_V$), and negative impact of functionality (*F*) and branching (*B*) factors. This means that lipophilicity of carboxylic acids and esters increases with increase in molecular size, and it decreases as the molecule becomes more branched and as the functionality value rises. Decanoic acid showed outlier behavior in almost all the relations.

**Table 4.** Relations of hydrophobicity (log *P*) of carboxylic acids and esters with various indices

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ se | $\beta_2$ se | $\beta_3$ se | $\beta_4$ se | $\alpha$ se | $Q^2$ (SDEP) | $R_a^2$ (r or R) | s (F (df)) | AVRES (n) |
| 21 | MCI | 0.998 $^1\chi^v$ | | | | –1.190 | 0.995 | 0.996 | 0.087 | 0.061 |
| | | 0.013 | | | | 0.042 | (0.093) | (0.998) | (6042.458 (1, 23)) | (25) |
| 22 | MN(I) | 0.108 *I* | | | | –0.651 | 0.993 | 0.994 | 0.106 | 0.085 |
| | | 0.002 | | | | 0.044 | (0.116) | (0.997) | (4095.987 (1, 23)) | (25) |
| 23 | TAU | 0.972 *T* | | | | 0.314 | 0.959 | 0.963 | 0.265 | 0.188 |
| | | 0.039 | | | | 0.079 | (0.275) | (0.982) | (632.016 (1, 23)) | (25) |
| 24 | TAU | 1.009 $T_R$ | –0.214 *F* | | | –1.586 | 0.995 | 0.997 | 0.079 | 0.060 |
| | | 0.012 | 0.051 | | | 0.126 | (0.092) | (0.998) | (3630.220 (2, 22)) | (25) |
| 25 | TAU | –0.211 *F* | –0.951 *B* | 0.505 $N_V$ | | –1.689 | 0.995 | 0.997 | 0.081 | 0.060 |
| | | 0.054 | 0.275 | 0.006 | | 0.142 | (0.097) | (0.998) | (2315.067 (3, 21)) | (25) |
| 26 | TAU | –0.183 *F* | 0.509 $N_V$ | –0.143 $N_B$ | | –1.769 | 0.996 | 0.997 | 0.075 | 0.053 |
| | | 0.049 | 0.006 | 0.034 | | 0.126 | (0.091) | (0.999) | (2699.400 (3, 21)) | (25) |
| 27 | TAU | –0.183 *F* | 0.509 $N_V$ | –0.143 $N_P$ | | –1.482 | 0.996 | 0.997 | 0.075 | 0.053 |
| | | 0.049 | 0.006 | 0.034 | | 0.157 | (0.091) | (0.999) | (2699.400 (3, 21)) | (25) |
| 28 | TAU | –0.183 *F* | 0.509 $N_I$ | 0.875 $N_Y$ | | –0.750 | 0.996 | 0.997 | 0.075 | 0.053 |
| | | 0.049 | 0.006 | 0.035 | | 0.123 | (0.091) | (0.999) | (2699.402 (3, 21)) | (25) |
| 29 | TAU | –0.183 *F* | 0.509 $N_I$ | 0.875 $N_P$ | | –2.501 | 0.996 | 0.997 | 0.075 | 0.053 |
| | | 0.049 | 0.006 | 0.035 | | 0.160 | (0.091) | (0.999) | (2699.402 (3, 21)) | (25) |

## 3.4 QSPR of Ethers (*n* = 12)

Eqs. (30)–(37) describing relations of log *P* of ethers with topological indices are shown in Table 5 which shows that 95.1% of the variance could be explained by first order molecular connectivity index (predicted variance 93.7%). However, only 63.3% could be explained by molecular negentropy (predicted variance 55.6%). Similarly, the composite topochemical index *T* could explain only 68.1% of the variance (predicted variance 58.2%). However, the topochemical skeletal index $T_R$ explained 93.1% of the variance (predicted variance 91.6%). This relation was improved further by partitioning $T_R$ into branching and size parameters. The best relation, Eq. (37), could explain 99.3% of the variance. However, due to insufficient occurrence of quaternary carbon fragment ($N_X$ type) in the compounds, LOO could not be applied for this equation. Ethyl cyclopropyl ether was an outlier for this equation. $T_R$ and bulk showed positive contributions, while branching and shape factors showed negative impact. Functionality did not show any contribution.

Specific contributions of tertiary and quaternary type carbons, and molecular bulk are evident from the respective regression coefficients.

**Table 5.** Relations of hydrophobicity (log *P*) of ethers with various indices

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ se | $\beta_2$ se | $\beta_3$ se | $\beta_4$ se | $\alpha$ se | $Q^2$ (SDEP) | $R_a^2$ (r or R) | s (F (df)) | AVRES (n) |
| 30 | MCI | 0.926 $^1\chi^v$ | | | | –0.788 | 0.937 | 0.951 | 0.077 | 0.061 |
| | | 0.064 | | | | 0.148 | (0.083) | (0.977) | (212.339 (1, 10)) | (12) |
| 31 | MN(I) | 0.089 *I* | | | | | 0.556 | 0.633 | 0.211 | 0.142 |
| | | 0.004 | | | | | (0.222) | (0.795) | (505.572 (1, 11)) | (12) |
| 32 | TAU | 0.623 *T* | | | | 0.797 | 0.582 | 0.681 | 0.196 | 0.152 |
| | | 0.126 | | | | 0.124 | (0.215) | (0.843) | (24.490 (1, 10)) | (12) |
| 33 | TAU | 0.988 $T_R$ | | | | –1.411 | 0.916 | 0.931 | 0.091 | 0.056 |
| | | 0.081 | | | | 0.227 | (0.097) | (0.968) | (149.718 (1, 10)) | (12) |
| 34 | TAU | –1.308 *B* | 0.495 $N_V$ | | | –1.465 | 0.963 | 0.967 | 0.063 | 0.035 |
| | | 0.182 | 0.029 | | | 0.170 | (0.063) | (0.986) | (162.196 (2, 9)) | (12) |
| 35 | TAU | 0.500 $N_V$ | –0.251 $N_B$ | | | –1.470 | 0.932 | 0.943 | 0.083 | 0.043 |
| | | 0.038 | 0.049 | | | 0.224 | (0.086) | (0.976) | (91.236 (2, 9)) | (12) |
| 36 | TAU | 0.427 $N_I$ | 0.475 $N_P$ | | | –1.057 | 0.735 | 0.837 | 0.140 | 0.108 |
| | | 0.057 | 0.090 | | | 0.371 | (0.172) | (0.931) | (29.181 (2, 9)) | (12) |
| 37 | TAU | 0.500 $N_V$ | –0.470 $N_X$ | –0.215 $N_Y$ | | –1.470 | —$^+$ | 0.993 | 0.029 | 0.012 |
| | | 0.013 | 0.032 | 0.018 | | 0.080 | —$^+$ | (0.997) | (522.515 (3, 8)) | (12) |

$^+$ LOO could not be applied due to insufficient occurrence of $N_X$ fragment (quaternary carbon)

**Table 6.** Relations of hydrophobicity (log *P*) of halocarbons with various indices

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ se | $\beta_2$ se | $\beta_3$ se | $\beta_4$ se | $\alpha$ se | $Q^2$ (SDEP) | $R_a^2$ (r or R) | s (F (df)) | AVRES (n) |
| 38 | MCI | –0.222* $^1\chi^v$ | | | | 2.605 | –1.010 | –0.157 | 0.276 | 0.193 |
| | | 0.392 | | | | 0.932 | (0.333) | (0.273) | (0.322 (1, 4)) | (6) |
| 39 | MN(I) | 0.038 *I* | | | | 1.778 | 0.424 | 0.748 | 0.129 | 0.090 |
| | | 0.009 | | | | 0.092 | (0.178) | (0.894) | (15.843 (1, 4)) | (6) |
| 40 | TAU | 0.118 *T* | | | | 2.219 | 0.640 | 0.849 | 0.100 | 0.070 |
| | | 0.022 | | | | 0.048 | (0.141) | (0.938) | (29.122 (1, 4)) | (6) |
| 41 | TAU | 0.347 $T_R$ | | | | 1.421 | 0.564 | 0.831 | 0.105 | 0.078 |
| | | 0.069 | | | | 0.137 | (0.155) | (0.930) | (25.652 (1, 4)) | (6) |
| 42 | TAU | –0.208 *F* | 1.449 *B* | | | 2.675 | –$^+$ | 0.897 | 0.083 | 0.039 |
| | | 0.032 | 0.594 | | | 0.095 | –$^+$ | (0.968) | (22.663 (2, 3)) | (6) |
| 43 | TAU | –0.208 *F* | 0.265 $N_B$ | | | 2.675 | –$^+$ | 0.897 | 0.083 | 0.039 |
| | | 0.032 | 0.109 | | | 0.095 | –$^+$ | (0.968) | (22.663 (2, 3)) | (6) |

***** Insignificant at 90% level; $^+$ LOO could not be applied due to insufficient occurrence of branching

## 3.5 QSPR of Halocarbons (*n* = 6)

In case of halocarbons, inferior relations were obtained. These relations, Eqs. (38)–(43) are listed in Table 6. Molecular connectivity could not give any acceptable relation (explained variance 27.2%, insignificant β–coefficient) while molecular negentropy explained 74.8% variance (predicted variance 42.4%). The composite topochemical index *T* could explain 84.9% of the variance (predicted variance 64.0%). The topochemical skeletal index $T_R$ explained almost similar to *T* (83.1% of the variance), but the predicted variance reduced to 56.4%. On further partitioning of

$T_R$, TAU scheme generated equations 42 and 43, both explaining 89.7% of the variance. However, due to insufficient occurrence of branching, PRESS statistics could not be obtained for these two equations. There was no outlier for the equations. Functionality was found to have negative contribution while branching showed positive impact. However, the number of data points of halocarbon compounds was insufficient for regression with two predictor variables. The results obtained are treated only as preliminary ones.

**Table 7.** Relations of hydrophobicity (log *P*) of hydrocarbons with various indices

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ se | $\beta_2$ se | $\beta_3$ se | $\beta_4$ se | $\alpha$ se | $Q^2$ (SDEP) | $R_a^2$ (r or R) | s (F (df)) | AVERES (n) |
| 44 | MCI | 0.936 $^1\chi^v$ | | | | 0.244 | 0.949 | 0.953 | 0.125 | 0.103 |
| | | 0.036 | | | | 0.099 | (0.130) | (0.977) | (673.244 (1, 32)) | (34) |
| 45 | MN(I) | 0.064 *I* | | | | 1.772 | 0.455 | 0.507 | 0.407 | 0.311 |
| | | 0.011 | | | | 0.181 | (0.421) | (0.722) | (34.920 (1, 32)) | (34) |
| 46 | TAU | 0.564 *T* | | | | 1.300 | 0.511 | 0.550 | 0.389 | 0.314 |
| | | 0.088 | | | | 0.237 | (0.399) | (0.751) | (41.312 (1, 32)) | (34) |
| 47 | TAU | 0.856 $T_R$ | –0.245 *F* | | | 0.198 | 0.742 | 0.768 | 0.279 | 0.235 |
| | | 0.082 | 0.085 | | | 0.261 | (0.290) | (0.884) | (55.615 (2, 31)) | (34) |
| 48 | TAU | –0.270 *F* | –0.854 *B* | 0.453 $N_V$ | | | 0.788 | 0.816 | 0.248 | 0.211 |
| | | 0.083 | 0.654 | 0.011 | | | (0.263) | (0.910) | (1449.764 (3, 31)) | (34) |
| 49 | TAU | –0.477 *F* | 0.722 $N_P$ | 0.456 $N_I$ | | –0.254 | 0.977 | 0.981 | 0.080 | 0.061 |
| | | 0.026 | 0.020 | 0.012 | | 0.083 | (0.086) | (0.991) | (560.968 (3, 30)) | (34) |

## 3.6 QSPR of Hydrocarbons (*n* = 34)

Eqs. (44)–(49) relating log *P* of hydrocarbons with topological indices are shown in Table 7. First order molecular connectivity index could explain 95.3% of the variance (predicted variance 94.9%) while molecular negentropy explained only 50.7% of the variance (predicted variance 45.5%). Though the composite topochemical index *T* did not give satisfactory results (explained variance 55.0%, predicted variance 51.1%), on partitioning into different factors like functionality, branching and size, TAU scheme gave an highly acceptable relation, Eq. (49), with 98.1% explained variance and 97.7% predicted variance. This relation contains $N_P$ as the shape parameter and $N_I$ as the size parameter. 1,6–Heptadiyne showed outlier behaviour for the equation. Another relation (eq. 48), which is somewhat inferior in quality (explained variance 81.6%, predicted variance 78.8%) than that of Eq. (49), does not have any outlier. From the relations, it is evident that bulk ($N_V$) of the molecules has positive impact while branching and functionality have negative impact on log *P* values.

## 3.7 QSPR of Ketones (*n* = 15)

Table 8 shows the relations of log *P* values of ketones with topological parameters. First order molecular connectivity index and composite topochemical index *T* explained nearly to the same extent (96.7% and 95.8% respectively). However, MCI predicted somewhat better than *T* (94.3% *vs*. 92.9% predicted variance). On the other hand, molecular negentropy gave inferior relation,

explaining 83.0% of the variance (predicted variance 79.0%) and having intercept significant at 90% level. When $T$ was factored into $F$, $B$ and $N_I$, Eq. (53), explained variance rose to 98.2% (predicted variance 96.8%). This equation carries information about specific contributions of the parameters $F$ (functionality), $B$ (branching) and $N_I$ (size factor). Intercept of this equation is significant at 90% level. This relation shows that log $P$ value of ketones increases with increase in size and decreases as branching and functionality value rise. Acetone showed outlier behavior in all the equations.

**Table 8.** Relations of hydrophobicity (log $P$) of ketones with various indices

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\alpha$ | $Q^2$ | $R_a^2$ | s | AVERES |
| | | se | se | se | se | se | (SDEP) | (r or R) | (F (df)) | (n) |
| 50 | MCI | 0.912 $^1\chi^v$ | | | | –1.257 | 0.943 | 0.967 | 0.141 | 0.098 |
| | | 0.045 | | | | 0.131 | (0.177) | (0.984) | (405.934 (1, 13)) | (15) |
| 51 | MN(I) | 0.096 $I$ | | | | –0.458* | 0.790 | 0.830 | 0.317 | 0.229 |
| | | 0.012 | | | | 0.223 | (0.341) | (0.918) | (69.277 (1, 13)) | (15) |
| 52 | TAU | 0.896 $T$ | | | | –0.735 | 0.929 | 0.958 | 0.158 | 0.111 |
| | | 0.050 | | | | 0.119 | (0.198) | (0.980) | (319.367 (1, 13)) | (15) |
| 53 | TAU | –2.241 $F$ | 5.404 $B$ | 0.476 $N_I$ | | 1.508* | 0.968 | 0.982 | 0.103 | 0.069 |
| | | 0.682 | 0.500 | 0.019 | | 0.711 | (0.133) | (0.993) | (254.377 (3, 11)) | (15) |

* Significant at 90% level

**Table 9.** Relations of hydrophobicity (log $P$) of the composite set with various indices

| Eq | Index Type | Regression coefficient(s) and constant | | | | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\alpha$ | $Q^2$ | $R_a^2$ | s | AVRES |
| | | se | se | se | se | se | (SDEP) | (r or R) | (F (df)) | (n) |
| 54 | MCI | 0.956 $^1\chi^v$ | | | | –0.866 | 0.633 | 0.639 | 0.639 | 0.501 |
| | | 0.055 | | | | 0.154 | (0.642) | (0.801) | (296.937 (1, 166)) | (168) |
| 55 | MN(I) | 0.080 $I$ | | | | 0.238 | 0.386 | 0.397 | 0.826 | 0.654 |
| | | 0.008 | | | | 0.147 | (0.831) | (0.633) | (111.103 (1, 166)) | (168) |
| 56 | TAU | 0.583 $T$ | | | | 0.684 | 0.464 | 0.497 | 0.755 | 0.568 |
| | | 0.045 | | | | 0.094 | (0.776) | (0.707) | (165.896 (1, 166)) | (168) |
| 57 | TAU | 0.827 $T_R$ | –0.286 $F$ | | | –0.502 | 0.559 | 0.586 | 0.684 | 0.510 |
| | | 0.057 | 0.064 | | | 0.213 | (0.704) | (0.769) | (119.242 (2, 165)) | (168) |
| 58 | TAU | –0.318 $F$ | –2.513 $B$ | 0.420 $N_V$ | | –0.344 | 0.602 | 0.634 | 0.643 | 0.465 |
| | | 0.060 | 0.347 | 0.027 | | 0.201 | (0.669) | (0.801) | (97.622 (3, 164)) | (168) |
| 59 | TAU | –0.234 $F$ | 1.111 $B$ | 0.376 $N_I$ | | 0.669 | 0.542 | 0.594 | 0.678 | 0.497 |
| | | 0.064 | 0.396 | 0.026 | | 0.156 | (0.718) | (0.775) | (82.454 (3, 164)) | (168) |
| 60 | TAU | –0.298 $F$ | 0.437 $N_V$ | –0.568 $N_B$ | | –0.370 | 0.626 | 0.654 | 0.626 | 0.443 |
| | | 0.058 | 0.027 | 0.071 | | 0.195 | (0.649) | (0.812) | (106.169 (3, 164)) | (168) |
| 61 | TAU | –0.234 $F$ | 0.420 $N_V$ | –0.327 $N_P$ | | 0.094 | 0.601 | 0.627 | 0.650 | 0.448 |
| | | 0.061 | 0.027 | 0.047 | | 0.220 | (0.670) | (0.796) | (94.392 (3, 164)) | (168) |
| 62 | TAU | –0.279 $F$ | 0.437 $N_I$ | 0.335 $N_P$ | | –0.223 | 0.631 | 0.657 | 0.623 | 0.432 |
| | | 0.058 | 0.026 | 0.053 | | 0.224 | (0.644) | (0.815) | (107.797 (3, 164)) | (168) |
| 63 | TAU | –0.221 $F$ | 0.404 $N_I$ | 0.510 $N_X$ | 0.337 $N_Y$ | 0.425 | 0.594 | 0.619 | 0.656 | 0.463 |
| | | 0.062 | 0.027 | 0.163 | 0.083 | 0.147 | (0.676) | (0.793) | (68.955 (4, 163)) | (168) |
| 64 | TAU | –0.310 $F$ | 0.439 $N_V$ | –0.809 $N_X$ | –0.529 $N_Y$ | –0.365 | 0.631 | 0.659 | 0.621 | 0.439 |
| | | 0.058 | 0.026 | 0.146 | 0.073 | 0.207 | (0.644) | (0.817) | (81.749 (4, 163)) | (168) |
| 65 | TAU | –0.572 $F$ | 0.500 $N_V$ | –0.852 $N_X$ | –0.512 $N_Y$ | –0.463 | 0.803 | 0.812 | 0.467 | 0.354 |
| | | 0.050 | 0.021 | 0.110 | 0.055 | 0.164 | (0.475) | (0.904) | (176.662 (4, 159)) | (164) |

## 3.8 QSPR of Composite Set ($n$ = 168)

Table 9 lists the relations of log $P$ values of all compounds (composite set) with topological indices. For the composite set, MCI and molecular negentropy could explain only 63.9% and 39.7% respectively of the variance (predicted variance 63.3% and 38.6% respectively). The composite topochemical index $T$ singularly explained 49.7% of the variance (predicted variance 46.4%). When $T$ was partitioned into $T_R$ and $F$, then the relation could explain 58.6% of the variance (predicted variance 55.9%). On further partitioning of $T_R$, different relations, Eqs. (58), (60), (61), (62) showing 63–65% explained variance (60–63% predicted variance) were obtained.

**Tables 10.** Observed and calculated molecular hydrophobicity (log $P$) data

| No | Compound | log $P$ | | | No | Compound | log $P$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Exp [a] | Calc | Calc [i] | | | Exp [a] | Calc | Calc [i] |
| **1** | Methanol | –0.66 | –0.948[b] | –0.049 | **2** | Ethanol | –0.32 | –0.223[b] | 0.555 |
| **3** | *n*–Propanol | 0.34 | 0.265[b] | 0.995 | **4** | *n*–Butanol | 0.88 | 0.754[b] | 1.434 |
| **5** | *n*–Pentanol | 1.40 | 1.242[b] | 1.873 | **6** | *n*–Hexanol | 1.84 | 1.730[b] | 2.312 |
| **7** | *n*–Heptanol | 2.34 | 2.218[b] | 2.752 | **8** | *n*–Octanol | 2.84 | 2.707[b] | 3.191 |
| **9** | *n*–Nonanol | 3.15 | 3.195[b] | 3.630 | **10** | Isopropanol | 0.14 | 0.097[b] | 0.538 |
| **11** | Isobutanol | 0.61 | 0.479[b] | 0.904 | **12** | *tert*–Butanol | 0.37 | 0.518[b] | 0.741 |
| **13** | Isopentanol | 1.14 | 0.968[b] | 1.344 | **14** | 2–Methyl butanol | 1.14 | 0.968[b] | 1.344 |
| **15** | 1–Methyl butanol | 1.14 | 1.073[b] | 1.417 | **16** | 3–Pentanol | 1.14 | 1.073[b] | 1.417 |
| **17** | 3–Methyl–2–butanol | 0.91 | 0.799[b] | 0.887 | **18** | 2–Methyl–2–butanol | 0.89 | 1.006[b] | 1.180 |
| **19** | 2,2–Dimethyl–1–propanol | 1.36 | 0.839[b] | 1.064 | **20** | 2–Hexanol | 1.61 | 1.561[b] | 1.856 |
| **21** | 3–Hexanol | 1.61 | 1.561[b] | 1.856 | **22** | 3–Methyl–3–pentanol | 1.39 | 1.495[b] | 1.620 |
| **23** | 2–Methyl–2–pentanol | 1.39 | 1.494[b] | 1.620 | **24** | 2–Methyl–3–pentanol | 1.41 | 1.288[b] | 1.327 |
| **25** | 3–Methyl–2–pentanol | 1.41 | 1.288[b] | 1.327 | **26** | 4–Methyl–2–pentanol | 1.41 | 1.288[b] | 1.327 |
| **27** | 2,3–Dimethyl–2–butanol | 1.17 | 1.220[b] | 1.090 | **28** | 3,3–Dimethyl–1–butanol | 1.86 | 1.327[b] | 1.503 |
| **29** | 3,3–Dimethyl–2–butanol | 1.19 | 1.158[b] | 1.047 | **30** | 2–Methyl–2–hexanol | 1.87 | 1.982[b] | 2.059 |
| **31** | 3–Methyl–3–hexanol | 1.87 | 1.983[b] | 2.059 | **32** | 3–Ethyl–3–pentanol | 1.87 | 1.983[b] | 2.059 |
| **33** | 2,3–Dimethyl–2–pentanol | 1.67 | 1.709[b] | 1.530 | **34** | 2,3–Dimethyl–3–pentanol | 1.67 | 1.709[b] | 1.530 |
| **35** | 2,4–Dimethyl–2–pentanol | 1.67 | 1.709[b] | 1.530 | **36** | 2,4–Dimethyl–3–pentanol | 1.71 | 1.502[b] | 1.237 |
| **37** | 2,2–Dimethyl–3–pentanol | 1.69 | 1.647[b] | 1.487 | **38** | 2,2,3–Trimethyl–3–pentanol | 1.99 | 2.068[b] | 1.689 |
| **39** | Cyclohexanol | 1.23 | 1.561[b] | 1.856 | **40** | 4–Penten–1–ol | 1.04 | 0.914[b] | 1.645 |
| **41** | 3–Penten–2–ol | 0.81 | 0.715[b] | 1.168 | **42** | 1–Penten–3–ol | 0.81 | 0.763[b] | 1.201 |
| **43** | 1–Hexen–3–ol | 1.31 | 1.251[b] | 1.640 | **44** | 2–Hexen–4–ol | 1.31 | 1.203[b] | 1.607 |
| **45** | 2–Methyl–4–penten–3–ol | 1.11 | 0.977[b] | 1.111 | **46** | 2,2,2–Trifluoroethanol | 0.41 | –0.334[b] | 0.248 |
| **47** | Ethylene glycol | –1.93 | –0.307[b] | 0.597 | **48** | Allyl alcohol | 0.17 | –0.062[b] | 0.767 |
| **49** | sec–Butanol | 0.61 | 0.585[b] | 0.978 | **50** | 2,3–Butanediol | –0.92 | 0.333[b] | 0.563 |
| **51** | 1–Ethynylcyclohexanol | 1.73 | 2.471[b] | 2.498 | **52** | Methyl amine | –0.57 | –0.642[c] | –0.073 |
| **53** | Ethyl amine | –0.13 | –0.162[c] | 0.538 | **54** | *n*–Propyl amine | 0.48 | 0.317[c] | 0.978 |
| **55** | *n*–Butyl amine | 0.75 | 0.796[c] | 1.417 | **56** | *n*–Pentyl amine | 1.49 | 1.276[c] | 1.856 |
| **57** | *n*–Hexyl amine | 1.98 | 1.755[c] | 2.295 | **58** | *n*–Heptyl amine | 2.57 | 2.234[c] | 2.735 |
| **59** | Isobutyl amine | 0.73 | 0.652[c] | 0.887 | **60** | *sec*–butyl amine | 0.74 | 0.652[c] | 0.964 |
| **61** | 2–Amino octane | 2.82 | 2.569[c] | 2.721 | **62** | Cyclohexyl amine | 1.49 | 1.610[c] | 1.842 |
| **63** | Isopropyl amine | 0.26 | 0.173[c] | 0.524 | **64** | Methyl ethylamine | 0.15 | 0.317[c] | 0.684 |
| **65** | Di–*n*–propyl amine | 1.67 | 1.755[c] | 2.123 | **66** | Triethylamine | 1.44 | 1.610[c] | 1.462 |
| **67** | Di–*n*–butyl amine | 2.68 | 2.713[c] | 3.002 | **68** | Diethylamine | 0.57 | 0.796[c] | 1.245 |
| **69** | *n*–Propyl–*n*–butyl amine | 2.12 | 2.234[c] | 2.563 | **70** | Methyl–*n*–butyl amine | 1.33 | 1.276[c] | 1.562 |
| **71** | Piperidine | 0.85 | 1.276[c] | 1.684 | **72** | Ethyl–isopropyl amine | 0.93 | 1.131[c] | 1.209 |
| **73** | *n*–Propyl–*sec*–butylamine | 1.91 | 2.090[c] | 2.087 | **74** | *n*–Propyl–isobutylamine | 2.07 | 2.090[c] | 2.033 |
| **75** | Trimethylamine | 0.27 | 0.173[c] | –0.153 | **76** | Dimethyl–*n*–butylamine | 1.70 | 1.610[c] | 1.264 |
| **77** | Acetic acid | –0.17 | –0.204[d] | 0.305 | **78** | Propionic acid | 0.25 | 0.316[d] | 0.763 |
| **79** | Butyric acid | 0.79 | 0.825[d] | 1.202 | **80** | Hexanoic acid | 1.88 | 1.844[d] | 2.081 |
| **81** | Decanoic acid | 4.09 | 3.881[d] | 3.838 | **82** | Ethyl formate | 0.23 | 0.321[d] | 1.057 |

**Tables 10.** (Continued)

| No | Compound | log *P* | | | No | Compound | log *P* | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Exp [a] | Calc | Calc [i] | | | Exp [a] | Calc | Calc [i] |
| **83** | *n*–Propyl formate | 0.73 | 0.830[d] | 1.496 | **84** | Methyl acetate | 0.23 | 0.118[d] | 0.427 |
| **85** | Ethyl acetate | 0.73 | 0.696[d] | 0.983 | **86** | *n*–Propyl acetate | 1.23 | 1.205[d] | 1.422 |
| **87** | Isopropyl acetate | 1.03 | 1.092[d] | 0.945 | **88** | *n*–Butyl acetate | 1.73 | 1.715[d] | 1.862 |
| **89** | *sec*–Butyl acetate | 1.53 | 1.602[d] | 1.384 | **90** | Methyl propionate | 0.73 | 0.638[d] | 0.885 |
| **91** | Methyl butyrate | 1.23 | 1.147[d] | 1.324 | **92** | Ethyl hexanoate | 2.73 | 2.744[d] | 2.758 |
| **93** | Ethyl heptanoate | 3.23 | 3.253[d] | 3.197 | **94** | Ethyl octanonate | 3.73 | 3.762[d] | 3.637 |
| **95** | Ethyl nonanonate | 4.23 | 4.271[d] | 4.076 | **96** | Ethyl decanonate | 4.73 | 4.781[d] | 4.515 |
| **97** | Ethyl propionate | 1.23 | 1.216[d] | 1.440 | **98** | Ethyl butyrate | 1.73 | 1.725[d] | 1.880 |
| **99** | Ethyl isobutyrate | 1.53 | 1.587[d] | 1.358 | **100** | Pentyl acetate | 2.23 | 2.224[d] | 2.301 |
| **101** | Butyl pentanoate | 3.23 | 3.253[d] | 3.197 | **102** | Diethyl ether | 1.03 | 1.030[e] | 1.269 |
| **103** | Methyl butyl ether | 1.53 | 1.530[e] | 1.592 | **104** | Methyl–*sec*–butyl ether | 1.33 | 1.315[e] | 1.114 |
| **105** | Methyl isobutyl ether | 1.33 | 1.315[e] | 1.062 | **106** | Methyl *tert*–butyl ether | 1.06 | 1.060[e] | 0.865 |
| **107** | Ethyl propyl ether | 1.53 | 1.530[e] | 1.708 | **108** | Ethyl isopropyl ether | 1.33 | 1.315[e] | 1.231 |
| **109** | Di–*n*–propyl ether | 2.03 | 2.030[e] | 2.148 | **110** | *n*–propyl isopropyl ether | 1.83 | 1.815[e] | 1.670 |
| **111** | Methyl *n*–propyl ether | 1.03 | 1.030[e] | 1.153 | **112** | Methyl isopropyl ether | 0.83 | 0.815[e] | 0.675 |
| **113** | Ethyl cyclopropyl ether | 1.24 | 1.315[e] | 1.231 | **114** | Chloroform | 1.97 | 1.970[f] | –0.580 |
| **115** | Methyl iodide | 1.69 | 1.711[f] | –0.919 | **116** | Ethyl iodide | 2.00 | 1.993[f] | –0.060 |
| **117** | 1–Bromopropane | 2.10 | 2.111[f] | 0.554 | **118** | 1–Chlorobutane | 2.39 | 2.279[f] | 1.242 |
| **119** | 1–Fluoropentane | 2.33 | 2.416[f] | 1.886 | **120** | *n*–Pentane | 2.50 | 2.557[g] | 1.832 |
| **121** | 2–Methylbutane | 2.30 | 2.367[g] | 1.302 | **122** | 2–Methylpentane | 2.80 | 2.823[g] | 1.741 |
| **123** | 3–Methylpentane | 2.80 | 2.823[g] | 1.741 | **124** | *n*–Hexane | 3.00 | 3.013[g] | 2.271 |
| **125** | *n*–Heptane | 3.50 | 3.469[g] | 2.710 | **126** | 2,4–Dimethylpentane | 3.10 | 3.089[g] | 1.651 |
| **127** | *n*–Octane | 4.00 | 3.926[g] | 3.149 | **128** | Cyclopentane | 2.05 | 2.026[g] | 1.832 |
| **129** | Cyclohexane | 2.46 | 2.482[g] | 2.271 | **130** | Methylcyclopentane | 2.35 | 2.292[g] | 1.741 |
| **131** | Cycloheptane | 2.87 | 2.938[g] | 2.710 | **132** | Methylcyclohexane | 2.76 | 2.748[g] | 2.181 |
| **133** | Cyclooctane | 3.28 | 3.394[g] | 3.149 | **134** | 1,2–Dimethylcyclohexane | 3.06 | 3.013[g] | 2.091 |
| **135** | 1–Pentene | 2.20 | 2.207[g] | 1.604 | **136** | 2–Pentene | 2.20 | 2.156[g] | 1.571 |
| **137** | 1–Hexene | 2.70 | 2.663[g] | 2.043 | **138** | 2–Heptene | 3.20 | 3.068[g] | 2.449 |
| **139** | 1–Octene | 3.70 | 3.575[g] | 2.922 | **140** | 4–Methyl–1–pentene | 2.50 | 2.472[g] | 1.514 |
| **141** | 1,6–Heptadiene | 2.90 | 2.768[g] | 2.254 | **142** | 1,5–Hexadiene | 2.40 | 2.312[g] | 1.815 |
| **143** | 1,4–Pentdiene | 1.90 | 1.856[g] | 1.376 | **144** | Cyclopentene | 1.75 | 1.666[g] | 1.598 |
| **145** | Cyclohexene | 2.16 | 2.122[g] | 2.037 | **146** | Cycloheptene | 2.57 | 2.578[g] | 2.476 |
| **147** | 1–Pentyne | 1.98 | 2.138[g] | 1.560 | **148** | 1–Hexyne | 2.48 | 2.595[g] | 1.999 |
| **149** | 1–Heptyne | 2.98 | 3.051[g] | 2.438 | **150** | 1–Octyne | 3.48 | 3.507[g] | 2.877 |
| **151** | 1–Nonyne | 3.98 | 3.963[g] | 3.317 | **152** | 1,8–Nonadiyne | 3.46 | 3.544[g] | 3.045 |
| **153** | 1,6–Heptadiyne | 2.46 | 2.632[g] | 2.166 | **154** | Acetone | 0.21 | –0.002[h] | 0.518 |
| **155** | 2–Butanone | 0.29 | 0.405[h] | 0.976 | **156** | 2–Pentanone | 0.79 | 0.881[h] | 1.415 |
| **157** | 3–Pentanone | 0.79 | 0.806[h] | 1.433 | **158** | 3–Methyl–2–butanone | 0.59 | 0.679[h] | 0.894 |
| **159** | 2–Hexanone | 1.29 | 1.357[h] | 1.854 | **160** | 3–Hexanone | 1.29 | 1.282[h] | 1.872 |
| **161** | 3–Methyl–2–petanone | 1.09 | 0.949[h] | 1.333 | **162** | 4–Methyl–2–petanone | 1.09 | 1.188[h] | 1.325 |
| **163** | 2–Methyl–3–petanone | 1.09 | 1.079[h] | 1.351 | **164** | 2–Heptanone | 1.79 | 1.834[h] | 2.293 |
| **165** | 3–Heptanone | 1.79 | 1.758[h] | 2.311 | **166** | 2,4–Dimethyl–3–pentanone | 1.39 | 1.353[h] | 1.269 |
| **167** | 5–Nonanone | 2.79 | 2.711[h] | 3.190 | **168** | 2–Nonanone | 2.79 | 2.786[h] | 3.172 |

[a] Taken from refs. [37] and [38] ; [b] As per Eq. (14); [c] As per Eq. (19); [d] As per Eq. (26); [e] As per Eq. (37); [f] As per Eq. (42); [g] As per Eq. (49); [h] As per Eq. (53); [i] As per Eq. (64).

The best TAU relation, Eq. (64), could explain 65.9% of the variance (predicted variance 63.1%). Ethylene glycol, 2,3–butanediol, 2,4–dimethylpentane, chloroform, methyl iodide, ethyl iodide and 1–bromopropane showed outlier behaviour for most of the equations. When the four

halogen compounds from this list of outliers were deleted, the statistical quality of the equation, Eq. (65), rose significantly (explained variance 81.2% and predicted variance 80.3%). Positive impact of carbon skeleton ($T_R$) and bulk ($N_V$) factors and negative impacts of functionality, branching and shape factors were found with specific quantitative contribution pattern.

## 4 CONCLUSIONS

This study shows that though composite topochemical index *T* does not always provide better model for log *P* of heterofunctional acyclic compounds in comparison to molecular connectivity and negentropy, TAU scheme can generate statistically superior relations when the composite index is partitioned into different components like skeletal index, size and shape factors, branching and functionality. Moreover, TAU indices can unravel specific contributions of molecular bulk (size), functionality, branching and shape parameters to the lipophilicity of diverse functional compounds. In general, lipophilicity increases with increase in molecular bulk and skeletal index value, and decreases with increase in branching and functionality. However, the halocarbons show some aberrant behavior and behave as outliers in the composite set.

**Acknowledgment**. The authors acknowledge Sri Dipak Kumar Pal for constant help and inspiration.

## 5 REFERENCES

[1] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
[2] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 1983.
[3] P. J. Hansen and P. C. Jurs, Chemical Applications of Graph Theory, *J. Chem. Educ.* **1988**, *65*, 574–580.
[4] D. H. A. Rouvray, Rationale For The Topological Approaches in Chemistry, *J. Mol. Struct.* (*Theochem)* **1995**, *336*, 101–104.
[5] I. Motoc and A. T. Balaban, Topological Indices: Intercorrelation, Physical Meaning and Correlational Ability. *Rev. Roum. Chim.* **1981**, *26*, 593–600.
[6] S. C. Basak and B. D. Gute, Characterization of Molecular Structures Using Topological Indices, *SAR QSAR Environ. Res.* **1997**, *7*, 1–21.
[7] K. Balasubramanian and S. C. Basak, Characterization of Isospectral Graph Using Graph Invariants and Derived Orthogonal Parameters, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 367–373.
[8] J. Devillers and A. T. Balaban, Eds., *Topological Indices and Related Decriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, Netherlands, 1999.
[9] M. Randic, A. T. Balaban and S. C. Basak, On Structural Interpretation of Several Distance Related Topological Indices, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 593–601.
[10] A. Mercader, E. A. Castro and A. A. Toropov, Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 4. Modeling the Enthalpy of Formation of Hydrocarbons from Elements, *Int. J. Mol. Sci.* **2001**, *2*, 121–132, http://www.mdpi.org/ijms.
[11] 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.
[12] A. A. Toropov and A. P. Toropova, QSAR Modeling of Mutagenicity Based on Graphs of Atomic Orbitals, *Internet Electron. J. Mol. Des.* **2002**, *1*, 108–114, http://www.biochempress.com.
[13] D. J. G. Marino, P. J. Peruzzo, E. A. Castro, and A. A. Toropov, QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and

Correlation Weights of Local Graph Invariants, *Internet Electron. J. Mol. Des.* **2002**, *1*, 115–133, http://www.biochempress.com.
[14] S.–S. Liu, H.–L. Liu, Y.–Y. Shi, and L.–S. Wang, QSAR of Cyclooxygenase–2 (COX–2) Inhibition by 2,3–Diarylcyclopentenones Based on MEDV–13, *Internet Electron. J. Mol. Des.* **2002**, *1*, 310–318, http://www.biochempress.com.
[15] E. Estrada, The Structural Interpretation of the Randić Index, *Internet Electron. J. Mol. Des.* **2002**, *1*, 360–366, http://www.biochempress.com.
[16] 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.
[17] I. Lukovits, A. Miličević, S. Nikolić, and N. Trinajstić, On Walk Counts and Complexity of General Graphs, *Internet Electron. J. Mol. Des.* **2002**, *1*, 388–400, http://www.biochempress.com.
[18] E. A. Castro, I. Gutman, D. Marino and P. Peruzzo, Upgrading the Wiener Index, *J. Serb. Chem. Soc.* **2002**, *67*, 647–651.
[19] L. B. Kier and L. H. Hall, General Definition of Valence δ Values for Molecular Connectivity, *J. Pharm. Sci.* **1983**, *72*, 1170–1173.
[20] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, England, 1986.
[21] 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.
[22] 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.
[23] 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.
[24] D. K. Pal, S. K. Purkayastha, C. Sengupta and A. U. De, Quantitative Structure–Property Relationships wuth TAU indices: Part I – Research Octane Numbers of Alkane Fuel Molecules, *Indian J. Chem.*, **1992**, *31B*, 109–114.
[25] 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.
[26] 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.
[27] 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, *Proc. 90th Indian Sci. Congr. 2003*, *Bangalore*, *India* (*3 - 7 January 2003)*, abstract no. 152, pp. 125-126.
[28] H. Kubinyi, In *Burger's Medicinal Chemistry and Drug Discovery*, M. E. Wolff, Ed., 5$^{th}$ ed., vol. I, John Wiley & Sons, N. Y., 1995, pp. 497–571.
[29] 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.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] M. Mandloi, A. Sikarwar, N. S. Sapre, S. Karmakar and P. V. Khadikar, A Comparative QSAR Study Using Wiener, Szeged and Molecular Connectivity Indices, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57–62.
[36] L. B. Kier, Use of Molecular Negentropy to Encode Structure Governing Biological Activity, *J. Pharm. Sci.* **1980**, *69*, 807–810.
[37] A. Leo, C. Hansch and D. Elkins, Partition Coefficients and Their Uses, *Chem. Rev.* **1971**, *71*, 525–616
[38] C. Hansch, J. E. Quinlan and G. L. Lawrence, The Linear Free Energy Relationship Between Partition–Coefficients and The Aqueous Solubility of Organic Liquids. *J. Org. Chem.* **1968**, *33*, 347–350.

[39] C. E. Shannon, A Mathematical Theory of Communication, *Bell*. *Syst.Tech*. *J*. **1948**, *27*, 379–423.
[40] C. E. Shannon and W. Weaver, W, *The Mathematical Theory of Communications*, University of Illinois Press, Urbana III, 1949.
[41] The GW–BASIC programs *RRR98*, *KRPRES1* and *KRPRES2* were developed by Kunal Roy (1998) and standardized using known data sets.
[42] G. W. Snedecor and W. G. Cochran, *Statistical Methods*; Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi, 1967, pp. 381 – 418.
[43] L. B. Kier and L. H. Hall, In *Advances in Drug Research*, volume 22, B. Testa, Ed., Academic Press, New York, 1992, pp. 1–38.
[44] S. Wold and L. Eriksson, In *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, Ed., VCH, Weinheim, 1995, p. 312.

## Biographies

**Kunal Roy** is a faculty member of the Division of Medicinal and Pharmaceutical Chemistry, Jadavpur University, India. Dr. Roy obtained his B. Pharm (Gold Medalist), M. Pharm (Gold Medalist) and Ph.D. degrees from Jadavpur University, Calcutta, India. He is a member of the Board of Associate Editors (2002–03) of the *Journal of the Indian Chemical Society*, India.

**Achintya Saha** is a faculty member of the Department of Chemical Technology, University of Calcutta, Calcutta, India. Dr. Saha obtained his B. Pharm, M. Pharm and Ph.D. degrees from Jadavpur University, India. Dr. Saha is a member of the Board of Associate Editors (2003–04) of the *Journal of the Indian Chemical Society*, India.