CROATICA CHEMICA ACTA CCACAA **68** (3) 417–434 (1995)

ISSN 0011-1643
CCA–2243

Original Scientific Paper

A Structure-Property Study of the Solubility of Aliphatic Alcohols in Water

Bono Lučić, Sonja Nikolić and Nenad Trinajstić

The Rugjer Bošković Institute,
HR-10 001 Zagreb, The Republic of Croatia

Albin Jurić

Department of Chemistry, Agricultural Institute,
HR-43 260 Križevci, The Republic of Croatia

Zlatko Mihalić

Faculty of Natural Sciences, The University of Zagreb,
HR-10 001 Zagreb, The Republic of Croatia

Received November 15, 1994; revised December 19, 1994; accepted December 20, 1994

This report presents the best possible QSPR models for predicting the solubility of aliphatic alcohols in water that can be obtained with non-orthogonalized valence-connectivity basis. The corresponding models with orthogonalized basis, ordered in the usual manner, are also given. However, both models are statistically equivalent. Therefore, we proposed a novel approach to the QSPR modelling, which is based on the consideration of all possible orthogonalization orderings of the valence-connectivity basis and dominant descriptor analysis. This novel procedure produced models that are better than the empirical models of Amidon *et al.* and the Kier-Hall models when considering the same source of experimental data.

INTRODUCTION

The QSPR (quantitative structure-property relationships) studies, based on topological indices, are increasingly being used in several areas of chemistry, biochemistry and environmental research.[1–17] In this report we present the structure-property study of the aqueous solubility of aliphatic alcohols using a recently developed fully automatic procedure for selecting the optimum structure-property models based on the orthogonalization ordering[18,19] of the connectivity basis[20,21] and the dominant descriptor analysis.[22] As the molecular property to study, we selected the solubility of aliphatic alcohols in water because there are already several QSPR models available in the literature[23–28] and thus we can compare our results with the existing models. The solubility of aliphatic alcohols in water is a useful parameter in many applications including biochemical research. This is the first step in understanding alcohol transport in the living organisms.

The strategy of our approach is as follows. The multiple linear regression, based on ordered orthogonalized connectivity indices, is used.[29–31] Experimental data are taken from Amidon *et al.*[23] In order to simplify the presentation, aliphatic alcohols are represented by weighted graphs[32] to account for the presence of oxygen in the molecule and the valence-connectivity indices[33] are used as molecular descriptors.

COMPUTATIONAL PROCEDURE

The valence-connectivity indices of different order are defined as:[33]

$$^{\ell}\chi^{v} = \sum_{\text{path}} \left[\delta(i)\,\delta(j)\,\ldots\,\delta(\ell+1)\right]^{-0.5} \tag{1}$$

where $\delta(i)$, $\delta(j)$, ..., $\delta(\ell + 1)$ are weights (valence-delta values) of vertices (atoms) $i, j, \ldots, \ell + 1$ making up the path of length ℓ in a vertex-weighted tree (aliphatic alcohol). The weights or the valence-delta values are given by:

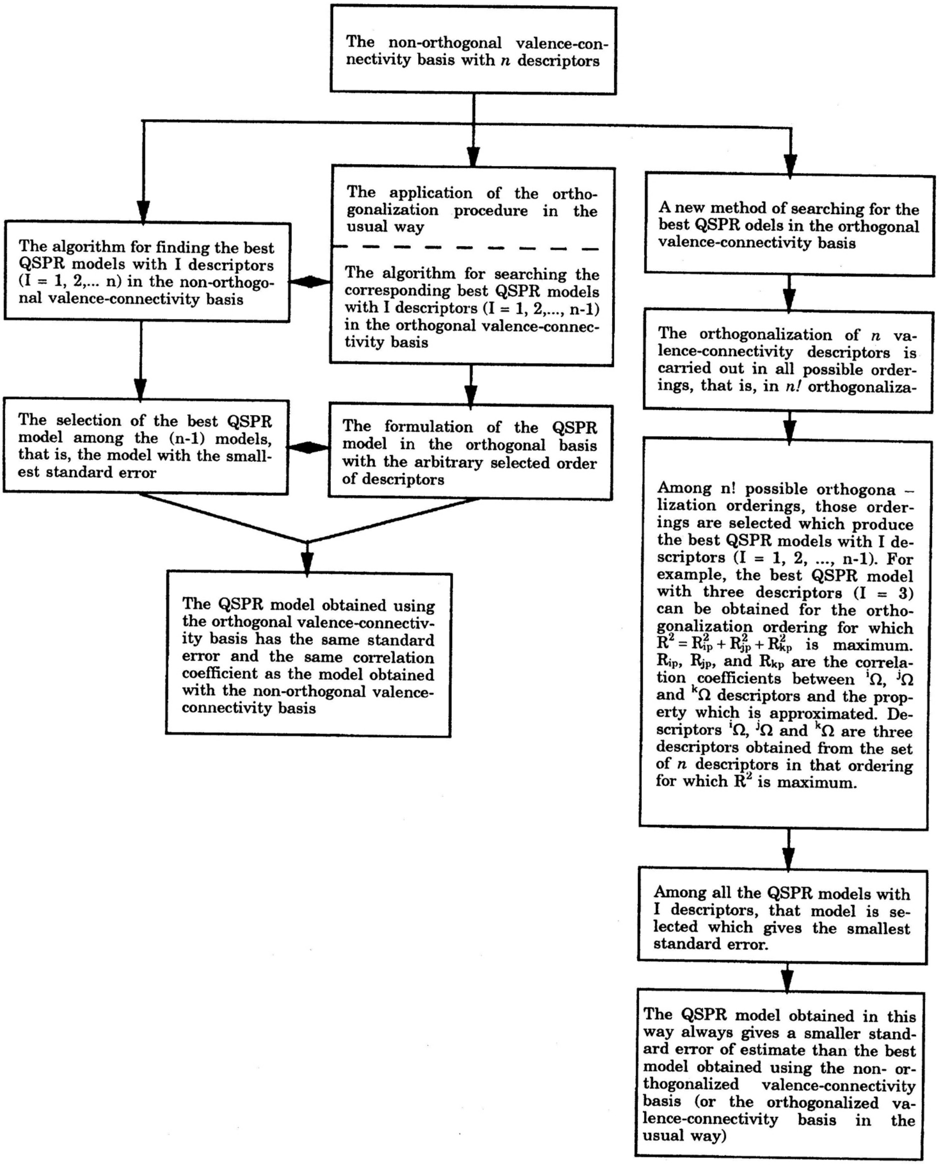
$$\delta(i) = (Z_i^v - \mathrm{H}_i) \,/\, (Z_i - Z_i^v - 1) \tag{2}$$

where Z_i^v stands for the number of valence electrons in atom i, Z_i is its atomic number and H_i is the number of hydrogen atoms attached to atom i. The valence-delta value for oxygen in the bonding situation characteristic of aliphatic alcohols is 6. It can be easily seen that Eq. (1) reduces to standard expression for computing the connectivity indices, when only hydrocarbons are considered, because valence-delta values then become valencies of carbon atoms.

From Eq. (1), one can compute the valence-connectivity indices of various order. They represent the non-orthogonal valence-connectivity basis.

TABLE I

Flow-chart of the computational procedure



The non-orthogonal valence-connectivity basis with *n* descriptors

The algorithm for finding the best QSPR models with I descriptors (I = 1, 2,... n) in the non-orthogonal valence-connectivity basis

The selection of the best QSPR model among the (n-1) models, that is, the model with the smallest standard error

The application of the orthogonalization procedure in the usual way

The algorithm for searching the corresponding best QSPR models with I descriptors (I = 1, 2,..., n-1) in the orthogonal valence-connectivity basis

The formulation of the QSPR model in the orthogonal basis with the arbitrary selected order of descriptors

The QSPR model obtained using the orthogonal valence-connectivity basis has the same standard error and the same correlation coefficient as the model obtained with the non-orthogonal valence-connectivity basis

A new method of searching for the best QSPR odels in the orthogonal valence-connectivity basis

The orthogonalization of *n* valence-connectivity descriptors is carried out in all possible orderings, that is, in *n!* orthogonaliza-

Among n! possible orthogonalization orderings, those orderings are selected which produce the best QSPR models with I descriptors (I = 1, 2, ..., n-1). For example, the best QSPR model with three descriptors (I = 3) can be obtained for the orthogonalization ordering for which $R^2 = R_{ip}^2 + R_{jp}^2 + R_{kp}^2$ is maximum. R_{ip}, R_{jp}, and R_{kp} are the correlation coefficients between ${}^i\Omega$, ${}^j\Omega$ and ${}^k\Omega$ descriptors and the property which is approximated. Descriptors ${}^i\Omega$, ${}^j\Omega$ and ${}^k\Omega$ are three descriptors obtained from the set of *n* descriptors in that ordering for which R^2 is maximum.

Among all the QSPR models with I descriptors, that model is selected which gives the smallest standard error.

The QSPR model obtained in this way always gives a smaller standard error of estimate than the best model obtained using the non-orthogonalized valence-connectivity basis (or the orthogonalized valence-connectivity basis in the usual way)

The next step is to orthogonalize the valence-connectivity basis to obtain the orthogonalized set of valence-connectivity indices ${}^{\ell}\Omega^{v}$ ($\ell = 0,1,\ldots, \ell + 1$). Computation of the orthogonalized indices has been outlined in several recent papers.[21,29–31,34] The most important step in our procedure is computing the QSPR models for different orderings of orthogonalized descriptors. In our recent work,[18,19] we found that some orthogonalization orderings lead to QSPR models with smaller values of the standard error of estimate than the corresponding models with non-orthogonalized descriptors. A special feature of our procedure is the dominant descriptor analysis, which consists of selecting the dominant descriptors and removing insignificant descriptors from the model. The procedure is sufficiently described in our other report[19] in this issue of the journal, so in Table I we give only a flow chart of the computational procedure applied in the present work.

RESULTS AND DISCUSSION

The truncated non-orthogonal valence-connectivity basis ${}^{\ell}\chi^{v}$ ($\ell = 0,1,\ldots,6$) of 54 aliphatic alcohols is given in Table II. Experimental solubilities of aliphatic alcohols in water are taken from Amidon *et al.*[23]

The orthogonalized valence connectivity indices ${}^{l}\Omega^{v}$ ($l = 0,1,\ldots,6$) are given in Table III.

In Table IV we give the best possible QSPR models that can be obtained with non-orthogonal valence-connectivity indices used in multiple linear regression *vs.* ln of experimental aqueous solubility of aliphatic alcohols.

Since we consider the standard error of estimate S as the critical quantity for determining the quality of a regression, the model with six descriptors ($S = 0.347$) appears to be of highest quality of all the models given in Table IV.

In Table V, we give the best possible QSPR models wih I-tuples of orthogonalized valence-connectivity indices used in a multiple linear regression *vs.* ln of experimental aqueous solubility of aliphatic alcohols.

Since the regression model with orthogonalized descriptors is more stable (that is, the coefficients of the descriptors do not change if one descriptor is omitted from the regression), but not better than the model with non-orthogonalized descriptors (because the values of the correlation coefficient R, the standard error S and the F-test remain the same for both models),[21,22,29–31,34] the same is confirmed here. The QSPR models in Tables IV and V do not differ in their statistical characteristics. However, some orthogonalization orderings followed by the dominant descriptors analysis (*e.g.*, considering only significant descriptors and neglecting insignificant descriptors) do lead to better models than those obtained with non-orthogonalized descriptors. This is shown in Table VI.

The best model, according to its statistical characteristics, is the third model in this table:

TABLE II

The non-orthogonal valence-connectivity indices of aliphatic alcohols

Aliphatic alcohol	$^0\chi^v$	$^1\chi^v$	$^2\chi^v$	$^3\chi^v$	$^4\chi^v$	$^5\chi^v$	$^6\chi^v$
1-Decanol	7.811	5.023	3.198	2.012	1.246	0.756	0.446
1-Dodecanol	9.225	6.023	3.906	2.512	1.599	1.006	0.623
1-Tetradecanol	10.640	7.023	4.613	3.012	1.953	1.256	0.800
1-Pentadecanol	11.347	7.523	4.966	3.262	2.130	1.381	0.888
1-Hexadecanol	12.054	8.023	5.320	3.512	2.306	1.506	0.976
2,3,3-Trimethyl-3-pentanol	7.154	3.534	3.877	2.304	0.530	0.000	0.000
2,2-Dimethyl-1-propanol	4.654	2.170	2.719	0.474	0.000	0.000	0.000
2,2-Dimethyl-3-pentanol	6.232	3.162	3.290	1.386	0.612	0.000	0.000
2,3-Dimethyl-2-butanol	5.525	2.667	2.808	1.413	0.000	0.000	0.000
2,3-Dimethyl-2-pentanol	6.232	3.205	2.963	1.818	0.500	0.000	0.000
2,3-Dimethyl-3-pentanol	6.232	3.228	2.865	1.960	0.408	0.000	0.000
2,4-Dimethyl-2-pentanol	6.232	3.140	3.410	0.908	0.999	0.000	0.000
2,4-Dimethyl-3-pentanol	6.179	3.234	2.979	1.366	0.770	0.000	0.000
2,6-Dimethyl-4-heptanol	7.593	4.201	3.913	1.487	1.185	0.385	0.385
3-Ethyl-1-butanol	5.361	2.784	2.562	0.862	0.612	0.000	0.000
2-Ethyl-1-hexanol	6.560	3.955	2.599	1.801	0.917	0.498	0.209
2-Methyl-1-propanol	3.732	1.879	1.576	0.365	0.000	0.498	0.209
2-Methyl-1-butanol	4.439	2.417	1.696	1.009	0.129	0.000	0.209
2-Methyl-2-butanol	4.654	2.284	2.166	0.865	0.000	0.000	0.209
2-Methyl-2-pentanol	5.361	2.784	2.562	0.862	0.612	0.000	0.000
2-Methyl-3-pentanol	5.309	2.862	2.220	1.188	0.471	0.000	0.000
2-Methyl-2-hexanol	6.096	3.284	2.916	1.142	0.609	0.433	0.000
3,3-Dimethyl-1-butanol	5.361	2.670	3.034	0.862	0.335	0.000	0.000
3,3-Dimethyl-2-butanol	5.525	2.624	3.042	1.253	0.000	0.000	0.000
3,5,5-Trimethyl-1-hexanol	7.646	4.064	4.382	1.583	1.368	0.352	0.137
3,5-Dimethyl-4-heptanol	7.593	4.310	3.262	2.454	1.044	0.465	0.096
3-Ethyl-3-pentanol	6.096	3.406	2.285	1.974	0.750	0.000	0.000
3-Ethyl-3-heptanol	7.483	4.406	3.035	2.312	1.109	0.579	0.250
3-Methyl-1-butanol	4.439	2.379	1.906	0.706	0.258	0.000	0.250
3-Methyl-2-butanol	4.602	2.324	1.985	0.965	0.000	0.000	0.250
3-Methyl-2-pentanol	5.309	2.862	2.126	1.467	0.341	0.000	0.000
3-Methyl-3-pentanol	5.361	2.845	2.204	1.523	0.250	0.000	0.000
3-Methyl-3-hexanol	6.069	3.345	2.600	1.550	0.715	0.177	0.000
1-Butanol	3.569	2.023	1.077	0.512	0.158	0.177	0.000
2-Butanol	3.732	1.951	1.257	0.591	0.000	0.177	0.000
4-Methyl-1-pentanol	5.146	2.879	2.260	0.940	0.500	0.183	0.000
4-Methyl-2-pentanol	5.309	2.807	2.479	0.813	0.682	0.000	0.000
1-Pentanol	4.276	2.523	1.431	0.762	0.362	0.112	0.000
2-Pentanol	4.439	2.451	1.638	0.706	0.418	0.000	0.000
3-Pentanol	4.439	2.489	1.470	0.942	0.289	0.000	0.000
1-Hexanol	4.983	3.023	1.784	1.012	0.539	0.256	0.079
2-Hexanol	5.146	2.951	1.991	0.975	0.500	0.295	0.000
3-Hexanol	5.146	2.989	1.851	1.093	0.537	0.204	0.000
1-Heptanol	5.690	3.523	2.138	1.262	0.715	0.381	0.181
3-Heptanol	5.853	3.489	2.204	1.362	0.644	0.380	0.144
4-Heptanol	5.853	3.489	2.231	1.244	0.811	0.289	0.144
7-Methyl-1-octanol	7.267	4.379	3.320	1.690	1.018	0.595	0.332
1-Octanol	6.397	4.023	2.491	1.512	0.892	0.506	0.269
2-Octanol	6.560	3.951	2.698	1.475	0.867	0.488	0.250
1-Nonanol	7.104	4.523	2.845	1.762	1.069	0.631	0.358
2-Nonanol	7.267	4.451	3.052	1.725	1.043	0.613	0.345
3-Nonanol	7.267	4.489	2.911	1.862	1.011	0.590	0.322
4-Nonanol	7.267	4.489	2.938	1.763	1.108	0.557	0.341
5-Nonanol	7.267	4.489	2.938	1.782	1.024	0.675	0.276

TABLE III

The orthogonalized valence-connectivity indices of aliphatic alcohols given in the increasing order of the ℓ value

Aliphatic alcohol	$^0\Omega^v$	$^1\Omega^v$	$^2\Omega^v$	$^3\Omega^v$	$^4\Omega^v$	$^5\Omega^v$	$^6\Omega^v$
1-Decanol	1.625	–0.297	0.086	–0.086	0.032	0.029	–0.008
1-Dodecanol	3.040	–0.263	0.091	–0.077	–0.009	0.023	0.010
1-Tetradecanol	4.454	–0.228	0.097	–0.069	–0.049	0.017	0.027
1-Pentadecanol	5.161	–0.211	0.099	–0.065	–0.069	0.014	0.036
1-Hexadecanol	5.868	–0.194	0.102	–0.061	–0.089	0.011	0.045
2,3,3-Trimethyl-3-pentanol	0.969	0.712	–0.121	–0.351	0.067	–0.018	–0.016
2,2-Dimethyl-1-propanol	–1.531	0.248	0.438	–0.089	–0.172	0.030	–0.084
2,2-Dimethyl-3-pentanol	0.046	0.409	0.081	–0.010	0.037	0.069	–0.033
2,3-Dimethyl-2-butanol	–0.661	0.387	–0.044	–0.149	–0.270	0.012	–0.063
2,3-Dimethyl-2-pentanol	0.046	0.366	–0.195	–0.122	0.022	0.031	–0.002
2,3-Dimethyl-3-pentanol	0.046	0.343	–0.266	–0.182	–0.027	0.041	–0.014
2,4-Dimethyl-2-pentanol	0.046	0.431	0.175	0.360	0.156	0.053	0.022
2,4-Dimethyl-3-pentanol	–0.006	0.299	–0.075	0.170	0.060	0.088	–0.001
2,6-Dimethyl-4-heptanol	1.408	0.367	0.117	0.341	–0.045	–0.006	0.169
3-Ethyl-1-butanol	–0.824	0.150	0.066	0.209	0.037	0.075	–0.001
2-Ethyl-1-hexanol	0.374	–0.144	–0.109	0.099	0.142	–0.108	0.007
2-Methyl-1-propanol	–2.454	–0.137	0.182	–0.027	–0.059	–0.488	0.131
2-Methyl-1-butanol	–1.747	–0.157	–0.005	–0.198	0.007	0.081	0.181
2-Methyl-2-butanol	–1.531	0.133	0.021	0.002	–0.227	0.019	0.171
2-Methyl-2-pentanol	–0.824	0.150	0.066	0.209	0.037	0.075	–0.001
2-Methyl-3-pentanol	–0.877	0.035	–0.115	0.072	–0.010	0.103	–0.015
2-Methyl-2-hexanol	–0.117	0.168	0.069	0.183	–0.141	–0.191	–0.177
3,3-Dimethyl-1-butanol	–0.824	0.265	0.403	–0.178	0.038	0.074	–0.073
3,3-Dimethyl-2-butanol	–0.661	0.430	0.139	–0.200	–0.239	0.019	–0.084
3,5,5-Trimethyl-1-hexanol	1.460	0.542	0.355	–0.007	0.405	–0.154	0.029
3,5-Dimethyl-4-heptanol	1.408	0.257	–0.404	–0.024	0.097	–0.192	–0.039
3-Ethyl-3-pentanol	–0.117	0.046	–0.418	–0.087	0.207	0.093	0.040
3-Ethyl-3-heptanol	1.297	0.081	–0.370	0.034	0.087	–0.206	0.060
3-Methyl-1-butanol	–1.747	–0.119	0.161	–0.086	0.050	0.087	0.221
3-Methyl-2-butanol	–1.584	0.056	–0.044	–0.044	–0.200	0.033	0.216
3-Methyl-2-pentanol	–0.877	0.035	–0.209	–0.099	–0.007	0.076	–0.010
3-Methyl-3-pentanol	–0.824	0.090	–0.220	–0.121	–0.076	0.068	–0.022
3-Methyl-3-hexanol	–0.117	0.107	–0.175	0.056	0.063	–0.018	–0.053
1-Butanol	–2.617	–0.400	0.071	–0.110	0.126	–0.116	–0.009
2-Butanol	–2.454	–0.208	–0.052	0.017	–0.089	–0.169	–0.008
4-Methyl-1-pentanol	–1.040	–0.102	0.163	–0.065	0.083	0.012	–0.099
4-Methyl-2-pentanol	–0.877	0.090	0.079	0.223	0.089	0.091	0.004
1-Pentanol	–1.910	–0.383	0.073	–0.106	0.132	0.063	–0.053
2-Pentanol	–1.747	–0.191	–0.023	0.125	0.055	0.092	0.010
3-Pentanol	–1.747	–0.229	–0.145	0.030	–0.005	0.101	–0.002
1-Hexanol	–1.203	–0.366	0.076	–0.102	0.112	0.041	–0.056
2-Hexanol	–1.040	–0.174	–0.020	0.110	–0.046	–0.059	–0.121
3-Hexanol	–1.040	–0.212	–0.116	0.102	–0.005	0.021	–0.086
1-Heptanol	–0.496	–0.349	0.078	–0.098	0.092	0.038	–0.034
3-Heptanol	–0.333	–0.195	–0.113	0.087	–0.082	–0.017	–0.048
4-Heptanol	–0.333	–0.195	–0.086	0.174	0.020	0.054	–0.001
7-Methyl-1-octanol	1.082	–0.051	0.171	–0.053	0.010	–0.031	–0.014
1-Octanol	0.211	–0.331	0.081	–0.094	0.072	0.035	–0.025
2-Octanol	0.374	–0.140	–0.015	0.118	–0.073	–0.011	–0.018
1-Nonanol	0.918	–0.314	0.084	–0.090	0.052	0.032	–0.017
2-Nonanol	1.082	–0.122	–0.013	0.122	–0.093	–0.014	–0.003
3-Nonanol	1.082	–0.160	–0.108	0.095	–0.109	0.013	–0.020
4-Nonanol	1.082	–0.160	–0.081	0.164	–0.064	0.041	0.019
5-Nonanol	1.082	–0.160	–0.081	0.145	–0.134	–0.057	–0.087

TABLE IV

The best possible QSPR models with I-tuples non-orthogonal valence-connectivity indices in a regression *vs.* ln of the solubility of aliphatic alcohols in water

---

***I* = 1**

$\ln \mathrm{SOL} = (5.846 \pm 0.225) + (-2.689 \pm 0.060)\ {}^1\chi^v$

$n = 54$, $R = 0.9874$, $S = 0.568$, $F = 2023$

***I* = 2**

$\ln \mathrm{SOL} = (6.132 \pm 0.186) + (-3.270 \pm 0.114)\ {}^1\chi^v + (1.225 \pm 0.220)\ {}^3\chi^v$

$n = 54$, $R = 0.9922$, $S = 0.452$, $F = 1612$

***I* = 3**

$\ln \mathrm{SOL} = (1.912 \pm 0.455) + (3.626 \pm 0.396)\ {}^0\chi^v + (-6.099 \pm 0.376)\ {}^1\chi^v +$
$+ (-2.280 \pm 0.279)\ {}^2\chi^v$

$n = 54$, $R = 0.9953$, $S = 0.354$, $F = 1765$

***I* = 4**

$\ln \mathrm{SOL} = (0.937 \pm 0.933) + (4.410 \pm 0.764)\ {}^0\chi^v + (-6.667 \pm 0.605)\ {}^1\chi^v +$
$+ (-2.851 \pm 0.482)\ {}^2\chi^v + (-0.406 \pm 0.340)\ {}^3\chi^v$

$n = 54$, $R = 0.9954$, $S = 0.352$, $F = 1336$

***I* = 5**

$\ln \mathrm{SOL} = (-0.284 \pm 1.306) + (5.086 \pm 0.914)\ {}^0\chi^v + (-6.944 \pm 0.636)\ {}^1\chi^v +$
$+ (-3.236 \pm 0.560)\ {}^2\chi^v + (-0.793 \pm 0.446)\ {}^3\chi^v + (-0.541 \pm 0.408)\ {}^4\chi^v$

$n = 54$, $R = 0.9956$, $S = 0.350$, $F = 1085$

***I* = 6**

$\ln \mathrm{SOL} = (-0.649 \pm 1.329) + (4.968 \pm 0.913)\ {}^0\chi^v + (-6.535 \pm 0.709)\ {}^1\chi^v +$
$+ (-3.180 \pm 0.558)\ {}^2\chi^v + (-0.926 \pm 0.455)\ {}^3\chi^v + (-0.813 \pm 0.458)\ {}^4\chi^v +$
$+ (-0.761 \pm 0.598)\ {}^5\chi^v$ $n = 54$, $R = 0.9958$, $S = 0.347$, $F = 917$

***I* = 7**

$\ln \mathrm{SOL} = (-0.869 \pm 1.348) + (4.903 \pm 0.915)\ {}^0\chi^v + (-6.263 \pm 0.760)\ {}^1\chi^v +$
$+ (-3.188 \pm 0.558)\ {}^2\chi^v + (-1.015 \pm 0.464)\ {}^3\chi^v + (-0.908 \pm 0.468)\ {}^4\chi^v +$
$+ (-0.466 \pm 0.470)\ {}^5\chi^v + (-0.650 \pm 0.609)\ {}^6\chi^v$

$n = 54$, $R = 0.9958$, $S = 0.348$, $F = 785$

$$\ln \mathrm{SOL} = (-3.666 \pm 0.046) + (3.626 \pm 0.377)\ {}^0\Omega^v + (-2.689 \pm 0.036)\ {}^1\Omega^v + (-1.015 \pm 0.450)\ {}^3\Omega^v \qquad (3)$$

$$n = 54,\ R = 0.9957,\ S = 0.337,\ F = 1944,\ Q = 2.951$$

TABLE V

The best possible QSPR models with *I*-tuples of orthogonalized valence-connectivity indices used in a regression *vs.* ln of the solubility of aliphatic alcohols in water

---

$\boldsymbol{I = 1}$; orthogonalization ordering: ${}^1\chi^v$
$\ln \mathrm{SOL} = (5.846 \pm 0.225) + (-\ 2.689 \pm 0.060)\ {}^1\Omega^v$
$n = 54,\ R = 0.9874,\ S = 0.568,\ F = 2023$

$\boldsymbol{I = 2}$; orthogonalization ordering: ${}^1\chi^v,\ {}^3\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.062) + (-2.689 \pm 0.048)\ {}^1\Omega^v + (-1.225 \pm 0.220)\ {}^3\Omega^v$
$n = 54,\ R = 0.9922,\ S = 0.452,\ F = 1612$

$\boldsymbol{I = 3}$; orthogonalization ordering: ${}^0\chi^v,\ {}^1\chi^v,\ {}^2\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.048) + (-1.947 \pm 0.028)\ {}^0\Omega^v + (3.284 \pm 0.181)\ {}^1\Omega^v +$
$+ (-2.380 \pm 0.279)\ {}^2\Omega^v$
$n = 54,\ R = 0.9953,\ S = 0.354,\ F = 1765$

$\boldsymbol{I = 4}$; orthogonalization ordering: ${}^0\chi^v,\ {}^1\chi^v,\ {}^2\chi^v,\ {}^3\chi^v$
$\ln \mathrm{SOL} = (-3.666 + 0.0480) + (-1.947 \pm 0.028)\ {}^0\Omega^v + (3.284 \pm 0.180)\ {}^1\Omega^v +$
$+ (-2.380 \pm 0.278)\ {}^2\Omega^v + (0.406 \pm 0.340)\ {}^3\Omega^v$
$n = 54,\ R = 0.9954,\ S = 0.352,\ F = 1336$

$\boldsymbol{I = 5}$; orthogonalization ordering: ${}^0\chi^v,\ {}^1\chi^v,\ {}^2\chi^v,\ {}^3\chi^v,\ {}^4\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.048) + (-1.947 \pm 0.028)\ {}^0\Omega^v + (3.284 \pm 0.178)\ {}^1\Omega^v +$
$+ (-2.380 \pm 0.276)\ {}^2\Omega^v + (0.406 \pm 0.337)\ {}^3\Omega^v + (-0.541 \pm 0.408)\ {}^4\Omega^v$
$n = 54,\ R = 0.9956,\ S = 0.350,\ F = 1085$

$\boldsymbol{I = 6}$; orthogonalization ordering: ${}^0\chi^v,\ {}^1\chi^v,\ {}^2\chi^v,\ {}^3\chi^v,\ {}^4\chi^v,\ {}^6\chi^v$
$\ln \mathrm{SOL} = (-\ 3.666 \pm 0.047) + (-1.947 \pm 0.027)\ {}^0\Omega^v + (3.284 \pm 0.177)\ {}^1\Omega^v +$
$+ (-2.380 \pm 0.274)\ {}^2\Omega^v + (0.406 \pm 0.335)\ {}^3\Omega^v + (-0.541 \pm 0.406)\ {}^4\Omega^v +$
$+ (0.761 \pm 0.598)\ {}^6\Omega^v$
$n = 54,\ R = 0.9958,\ S = 0.347,\ F = 917$

$\boldsymbol{I = 7}$; orthogonalization ordering: ${}^0\chi^v,\ {}^1\chi^v,\ {}^2\chi^v,\ {}^3\chi^v,\ {}^4\chi^v,\ {}^5\chi^v,\ {}^6\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.047) + (-1.947 \pm 0.027)\ {}^0\Omega^v + (3.284 \pm 0.177)\ {}^1\Omega^v +$
$+ (-2.380 \pm 0.274)\ {}^2\Omega^v + (0.406 \pm 0.335)\ {}^3\Omega^v + (-0.541 \pm 0.406)\ {}^4\Omega^v +$
$+ (0.559 \pm 0.462)\ {}^5\Omega^v + (-0.650 \pm 0.609)\ {}^6\Omega^v$
$n = 54,\ R = 0.9958,\ S = 0.348,\ F = 785$

---

where Q is equal to R/S and is called the quality factor.[35] The corresponding orthogonalization ordering is given by: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^2\chi^v$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^4\chi^v$, ${}^5\chi^v$, ${}^6\chi^v$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$. The dominant valence-connectivity indices are denoted in bold print, while the rest of indices represent the set of insignificant descriptors that were omitted from the model.

TABLE VI

The best possible QSPR models with *I*-tuples of orthogonalized valence-connectivity indices obtained by selecting the optimum orthogonalization ordering of descriptors. Bold letters denote those descriptors (*e.g.*, dominant descriptors) which, after the orthogonalization in the indicated order, take part in the construction of the model

---

$\boldsymbol{I} = \mathbf{1}$; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{0}\chi^{v}$, ${}^{2}\chi^{v}$, ${}^{3}\chi^{v}$, ${}^{4}\chi^{v}$, ${}^{5}\chi^{v}$, ${}^{6}\chi^{v}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.077) + (-2.6890 \pm 0.060)\ {}^{1}\Omega^{v}$
$n = 54, R = 0.9874, S = 0.568, F = 2023$

$\boldsymbol{I} = \mathbf{2}$; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{2}\chi^{v}$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{3}\chi^{v}$, ${}^{4}\chi^{v}$, ${}^{5}\chi^{v}$, ${}^{6}\chi^{v}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.062) + (3.626 \pm 0.392)\ {}^{0}\Omega^{v} + (-2.689 \pm 0.048)\ {}^{1}\Omega^{v}$
$n = 54, R = 0.9953, S = 0.351, F = 2697$

$\boldsymbol{I} = \mathbf{3}$; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{2}\chi^{v}$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{4}\chi^{v}$, ${}^{5}\chi^{v}$, ${}^{6}\chi^{v}$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.046) + (3.626 \pm 0.377)\ {}^{0}\Omega^{v} + (-2.689 \pm 0.036)\ {}^{1}\Omega^{v} +$
$+ (-1.015 \pm 0.450)\ {}^{3}\Omega^{v}$
$n = 54, R = 0.9957, S = 0.337, F = 1944$

$\boldsymbol{I} = \mathbf{4}$; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{2}\chi^{v}$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{4}\chi^{v}$, ${}^{\mathbf{6}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{5}\chi^{v}$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.046) + (3.626 \pm 0.377)\ {}^{0}\Omega^{v} + (-2.689 \pm 0.036)\ {}^{1}\Omega^{v} +$
$+ (-1.015 \pm 0.452)\ {}^{3}\Omega^{v} + (-0.482 \pm 0.567)\ {}^{6}\Omega^{v}$
$n = 54, R = 0.9958, S = 0.338, F = 1450$

$\boldsymbol{I} = \mathbf{5}$; orthogonalization ordering: ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{0}\chi^{v}$, ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{2}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{6}\chi^{v}$, ${}^{\mathbf{4}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.046) + (3.270 \pm 0.086)\ {}^{1}\Omega^{v} + (-2.785 \pm 0.475)\ {}^{2}\Omega^{v} +$
$+ (-4.480 \pm 0.069)\ {}^{3}\Omega^{v} + (-0.908 \pm 0.459)\ {}^{4}\Omega^{v} + (0.794 \pm 0.423)\ {}^{5}\Omega^{v}$
$n = 54, R = 0.9958, S = 0.341, F = 1145$

$\boldsymbol{I} = \mathbf{6}$; orthogonalization ordering: ${}^{\mathbf{4}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{6}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{3}\chi^{v}$, ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{2}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.046) + (-0.734 \pm 0.083)\ {}^{0}\Omega^{v} + (-3.013 \pm 0.343)\ {}^{1}\Omega^{v} +$
$+ (3.180 \pm 0.552)\ {}^{2}\Omega^{v} + (-6.398 \pm 0.090)\ {}^{4}\Omega^{v} + (-0.466 \pm 0.465)\ {}^{5}\Omega^{v} +$
$+ (6.239 \pm 0.333)\ {}^{6}\Omega^{v}$
$n = 54, R = 0.9958, S = 0.344, F = 936$

In Table V, the orthogonal valence-connectivity indices are obtained by the following orthogonalization order: ${}^{o}\chi^{v}$, ${}^{1}\chi^{v}$, ${}^{2}\chi^{v}$, ${}^{3}\chi^{v}$, ${}^{4}\chi^{v}$, ${}^{5}\chi^{v}$, ${}^{6}\chi^{v}$. This orthogonalization order is used by the majority of authors who use orthogonalized descriptors in their work. However, if the set of descriptors is unhomogeneous, that is, consisting of different sorts of descriptors, then orthogonalization ordering is usually selected according to some arbitrary criterion. The best QSPR model with three orthogonal descriptors in Table V, with the orthogonalization order: ${}^{o}\chi^{v}$, ${}^{1}\chi^{v}$, ${}^{2}\chi^{v}$ is:

$$\ln \mathrm{SOL} = (-3.666 \pm 0.048) + (-1.947 \pm 0.028)\, {}^{0}\Omega^{v} + \\ + (3.284 \pm 0.181)\, {}^{1}\Omega^{v} + (-2.380 \pm 0.279)\, {}^{2}\Omega^{v} \qquad (4)$$

$$n = 54,\ R = 0.9953,\ S = 0.354,\ F = 1765,\ Q = 2.812$$

The model presented before, *e.g.*, Eq. (3), is better than the above one. This is so because the three listed orthogonalized valence-connectivity indices in Eq. (4) are less dominant toward the rest of descriptors that are insignificant than in the case of the three orthogonalized descriptors in the QSPR model (3). The correlation coefficients between the orthogonalized descriptors in Table III for model (4) and ln SOL are: $R_{\mathrm{ov}} = 0.9567$, $R_{\mathrm{1v}} = 0.2487$, $R_{\mathrm{2v}} = 0.1166$, $R_{\mathrm{3v}} = 0.0179$, $R_{\mathrm{5v}} = 0.0162$ and $R_{\mathrm{6v}} = 0.0143$. The correlation coefficient for the QSPR model (4) can be computed, taking into account only the three dominant descriptors that appear in Eq. (4), as:

$$R = \left[R_{\mathrm{ov}}^{2} + R_{\mathrm{1v}}^{2} + R_{\mathrm{2v}}^{2}\right]^{1/2} = 0.9953 \qquad (5)$$

which is identical to the value of *R* given above. In the case of model (3), the values of the correlation coefficients between the orthogonalized valence-connectivity indices and ln SOL (the order of *R*'s follows the orthogonalization ordering given in Table VI for model (3) are the following: $R_{\mathrm{1v}} = 0.9874$, $\mathrm{R}_{\mathrm{2v}} = 0.0035$, $R_{\mathrm{ov}} = 0.1253$, $R_{\mathrm{4v}} = 0.0029$, $R_{\mathrm{5v}} = 0.0101$, $R_{\mathrm{6v}} = 0.0091$ and $R_{\mathrm{3v}} = 0.0294$. The correlation coefficient with three dominant descriptors in Eq. (3) is given by:

$$R = \left[R_{\mathrm{1v}}^{2} + R_{\mathrm{ov}}^{2} + R_{\mathrm{3v}}^{2}\right]^{1/2} = 0.9957 \qquad (6)$$

One can clearly see that in this case the insignificant descriptors have smaller values than in the case of model (4). The difference is caused by different selection of the orthogonalization ordering, that is, the usual orthogonalization ordering always produces individual correlation coefficients with higher values than the optimum orthogonalization ordering used, for example, to obtain the QSPR models in Table VI.

In Table VII, we give the best possible QSPR models wih *I*-tuples of non-orthogonalized valence-connectivity indices and c_{OH} descriptor used in a regression *vs.* ln of experimental aqueous solubility of aliphatic alcohols.

In Table VIII we give the possible QSPR models with *I*-tuples of six orthogonalized valence-connectivity indices and c_{OH} descriptor obtained by selecting the optimum orthogonalization orderings of descriptors.

If we compare the corresponding models (the corresponding model means the model with the same number of descriptors) in Table IV and Table VI, and the corresponding models in Table VII and Table VIII, we see that in each case the model with the orthogonalized descriptors possesses a lower value of S than the model with non-orthogonalized descriptors. This

TABLE VII

The best possible QSPR models with I-tuples of non-orthogonal valence-connectivity indices and c_{OH} descriptor used in a regression *vs.* ln of the solubility of aliphatic alcohols in water

---

$I = 1$

$\ln SOL = (5.846 \pm 0.225) + (-2.689 \pm 0.060)\ {}^1\chi^v$

$n = 54,\ R = 0.9874,\ S = 0.568,\ F = 2023$

$I = 2$

$\ln SOL = (9.182 \pm 0.230) + (-2.593 \pm 0.032)\ {}^1\chi^v + (-6.023 \pm 0.499)\ c_{OH}$

$n = 54,\ R = 0.9967,\ S = 0.292,\ F = 3895$

$I = 3$

$\ln SOL = (8.830 \pm 0.309) + (-2.813 \pm 0.085)\ {}^1\chi^v + (0.437 \pm 0.158)\ {}^3\chi^v +$
$+ (-5.205 \pm 0.555)\ c_{OH}$

$n = 54,\ R = 0.9972,\ S = 0.275,\ F = 2938$

$I = 4$

$\ln SOL = (9.774 \pm 0.394) + (-0.390 \pm 0.114)\ {}^0\chi^v + (-2.369 \pm 0.151)\ {}^1\chi^v +$
$+ (0.596 \pm 0.150)\ {}^3\chi^v + (-5.747 \pm 0.528)\ c_{OH}$

$n = 54,\ R = 0.9977,\ S = 0.249,\ F = 2678$

$I = 5$

$\ln SOL = (10.051 \pm 0.436) + (-0.435 \pm 0.117)\ {}^0\chi^v + (-2.503 \pm 0.177)\ {}^1\chi^v +$
$+ (0.713 \pm 0.170)\ {}^3\chi^v + (0.352 \pm 0.248)\ {}^4\chi^v + (-5.654 \pm 0.527)\ c_{OH}$

$n = 54,\ R = 0.9978,\ S = 0.247,\ F = 2187$

$I = 6$

$\ln SOL = (12.741 \pm 2.042) + (-2.023 \pm 1.184)\ {}^0\chi^v + (-1.303 \pm 0.907)\ {}^1\chi^v +$
$+ (0.951 \pm 0.706)\ {}^2\chi^v + (1.235 \pm 0.422)\ {}^3\chi^v + (0.650 \pm 0.331)\ {}^4\chi^v +$
$+ (-6.708 \pm 0.940)\ c_{OH}$

$n = 54,\ R = 0.9979,\ S = 0.245,\ F = 1853$

$I = 7$

$\ln SOL = (12.414 \pm 2.102) + (-1.963 \pm 1.193)\ {}^0\chi^v + (-1.215 \pm 0.920)\ {}^1\chi^v +$
$+ (0.892 \pm 0.714)\ {}^2\chi^v + (1.150 \pm 0.440)\ {}^3\chi^v + (0.564 \pm 0.353)\ {}^4\chi^v +$
$+ (-0.238 \pm 0.331)\ {}^5\chi^v + (-6.612 \pm 0.954)\ c_{OH}$

$n = 54,\ R = 0.9979,\ S = 0.246,\ F = 1573$

---

is not so only in the case of the models with descriptors which all possess the same statistical characteristics. However, in some cases, as we have shown in our paper on predicting the physicochemical properties of the α-amino acids,[19] it is possible to obtain the QSPR model with a single orthogo-

TABLE VIII

The best possible QSPR models with I-tuples of six orthogonalized valence-connectivity indices (${}^0\chi^v$, ${}^1\chi^v$, ${}^2\chi^v$, ${}^3\chi^v$, ${}^4\chi^v$, ${}^5\chi^v$) and c_{OH} descriptor obtained by selecting the optimum orthogonalization orderings of descriptors. Bold letters denote those descrriptors (*e.g.*, dominant descriptors) which after the orthogonalization in the indicated order, take part in the construction of the model

---

$\boldsymbol{I}$ = **1**; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^0\chi^v$, ${}^2\chi^v$, ${}^3\chi^v$, ${}^4\chi^v$, ${}^5\chi^v$, c_{OH}
$\ln \mathrm{SOL} = (-3.366 \pm 0.077) + (-2.689 \pm 0.060)\ {}^1\Omega^v$
$n = 54$, $R = 0.9874$, $S = 0.568$, $F = 2023$

$\boldsymbol{I}$ = **2**; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^2\chi^v$, $\mathbf{c_{OH}}$, ${}^0\chi^v$, ${}^3\chi^v$, ${}^4\chi^v$, ${}^5\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.037) + (-2.689 \pm 0.029)\ {}^1\Omega^v + (-6.336 \pm 0.479)\ \Omega_{cOH}$
$n = 54$, $R = 0.9972$, $S = 0.273$, $F = 4479$

$\boldsymbol{I}$ = **3**; orthogonalization ordering: ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$, c_{OH}, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^3\chi^v$, ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^2\chi^v$, ${}^4\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.033) + (-1.438 \pm 0.038)\ {}^0\Omega^v + (-2.134 \pm 0.231)\ {}^1\Omega^v +$
$+ (-8.892 \pm 0.090)\ {}^5\Omega^v$
$n = 54$, $R = 0.9978$, $S = 0.244$, $F = 3722$

$\boldsymbol{I}$ = **4**; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^0\chi^v$, $\mathbf{c_{OH}}$, ${}^2\chi^v$, ${}^{\mathbf{4}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^5\chi^v$
$\ln \mathrm{SOL} = (-3.666 \pm 0.033) + (-2.689 \pm 0.025)\ {}^1\Omega^v + (-\ 1.225 \pm 0.116)\ {}^3\Omega^v +$
$+ (-0.650 \pm 0.324)\ {}^4\Omega^v + (5.747 \pm 0508)\ \Omega_{cOH}$
$n = 54$, $R = 0.9979$, $S = 0.240$, $F = 2897$

$\boldsymbol{I}$ = **5**; orthogonalization ordering: ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^0\chi^v$, $\mathbf{c_{OH}}$, ${}^2\chi^v$, ${}^{\mathbf{4}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.033) + (-2.689 \pm 0025)\ {}^1\Omega^v + (-1.225 \pm 0.117)\ {}^3\Omega^v +$
$+ (-0.650 \pm 0.276)\ {}^4\Omega^v + (-0.238 \pm 0.324)\ {}^5\Omega^v + (5.747 \pm 0.510)\ \Omega_{cOH}$
$n = 54$, $R = 0.9979$, $S = 0.241$, $F = 2296$

$\boldsymbol{I}$ = **6**; orthogonalization ordering: $\mathbf{c_{OH}}$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^2\chi^v$, ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{4}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.033) + (0.390 \pm 0.111)\ {}^0\Omega^v + (-2.813 \pm 0.075)\ {}^1\Omega^v +$
$+ (4.452 \pm 0.049)\ {}^3\Omega^v + (0.564 \pm 0.350)\ {}^4\Omega^v + (0.416 \pm 0.308)\ {}^5\Omega^v +$
$+ (16.260 \pm 0.403)\ \Omega_{cOH}$
$n = 54$, $R = 0.9979$, $S = 0.243$, $F = 11874$

$\boldsymbol{I}$ = **7**; orthogonalization ordering: $\mathbf{c_{OH}}$, ${}^{\mathbf{3}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{1}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{0}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{2}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{5}}\boldsymbol{\chi}^{\mathbf{v}}$, ${}^{\mathbf{4}}\boldsymbol{\chi}^{\mathbf{v}}$
$\ln \mathrm{SOL} = (-3.666 \pm 0.034) + (0.390 \pm 0.113)\ {}^0\Omega^v + (-2.813 \pm 0.076)\ {}^1\Omega^v +$
$+ (0.024 \pm 0.528)\ {}^2\Omega^v + (4.452 \pm 0.050)\ {}^3\Omega^v + (0.564 \pm 0.353)\ {}^4\Omega^v +$
$+ (0.416 \pm 0.311)\ {}^5\Omega^v + (16.260 \pm 0.407)\ \Omega_{cOH}$
$n = 54$, $R = 0.9979$, $S = 0.246$, $F = 1573$

---

nal descriptor which possesses a lower value of S than a model with a single non-orthogonal descriptor. In the present case, this is not possible because descriptor ${}^1\chi^v$ is by far the most dominant descriptor in the model.

A comparison between the experimental and computed values (using Eq. (3)) for ln SOL of aliphatic alcohols is given in Table IX. The plot between the experimental and computed values, that is, $(\ln \mathrm{SOL})_{\mathrm{exp}}$ *vs.* $\ln (\mathrm{SOL})_{\mathrm{calc}}$, is shown in Figure 1.

Model (3) compares well with other models in the literature, although this comparison is not quite adequate because the authors have used different numbers of alcohols and different descriptors. Nevertheles, we list these models because all of them could be improved by using the procedure described in this paper. Amidon *et al.*[23] produced three models between the aqueous solubility of aliphatic alcohols and their surface areas. These authors adopted the definition of molecular surface area from Hermann:[36] A molecule is considered as a collection of spheres with each radius located at the nuclear centre. To each radius on the solute molecule (aliphatic alcohol), a radius for the solvent (water) is added to give a surface. This approach has the convenient property of eliminating from the total surface area of a molecule those areas which are not exposed (or accessible) to the solvent. There is another advantage of Hermann's procedure, *viz.* it allows individual atom contributions to the total surface area to be computed. This permits the total surface area of a molecule to be divided into group contributions, *e.g.*, hydrocarbon and hydroxyl group parts of alcohols, and the group contribution to the solubility can be estimated. Standard bond angles

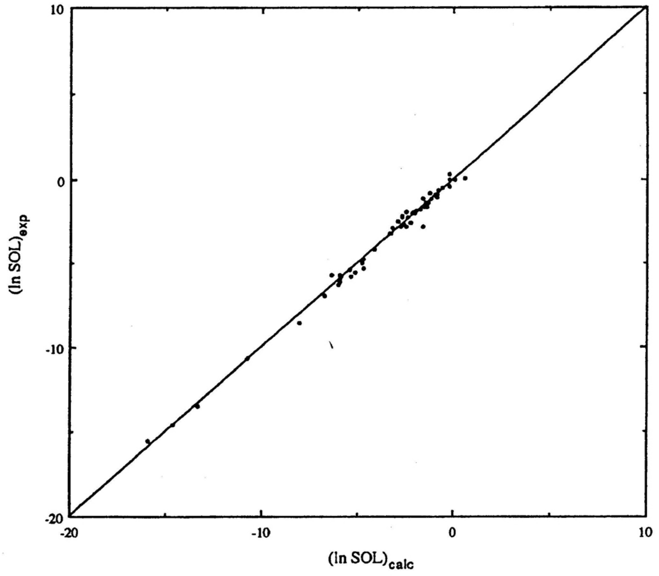


Figure 1. A plot of $(\ln \mathrm{SOL})_{\mathrm{exp}}$ *versus* $(\ln \mathrm{SOL})_{\mathrm{calc}}$ computed using Eq. (3).

TABLE IX

Experimental and computed values of ln SOL of aliphatic alcohols. The computed values are obtained from the best possible structure-water solubility models based on Eq. (3) and Eq. (12)

Aliphatic alcohol	$(\ln \mathrm{SOL})_{\mathrm{exp}}$	(ln SOL)calc Eq. (3)	$(\ln \mathrm{SOL})_{\mathrm{calc}}$ Eq. (12)
1-Decanol	–8.517	–8.058	–8.043
1-Dodecanol	–10.680	–10.694	–10.665
1-Tetradecanol	–13.437	–13.330	–13.286
1-Pentadecanol	–14.614	–14.648	–14.597
1-Hexadecanol	–15.587	–15.966	–15.908
2,3,3-Trimethyl-3-pentanol	–2.932	–3.179	–3.234
2,2-Dimethyl-1-propanol	–0.889	–0.857	–0.923
2,2-Dimethyl-3-pentanol	–2.644	–2.628	–2.670
2,3-Dimethyl-2-butanol	–0.851	–0.929	–0.853
2,3-Dimethyl-2-pentanol	–2.002	–2.169	–2.211
2,3-Dimethyl-3-pentanol	–1.938	–2.072	–2.244
2,4-Dimethyl-2-pentanol	–2.146	–2.670	–2.156
2,4-Dimethyl-3-pentanol	–2.802	–2.428	–2.774
2,6-Dimethyl-4-heptanol	–5.776	–5.342	–5.366
3-Ethyl-1-butanol	–2.787	–1.624	–2.395
2-Ethyl-1-hexanol	–4.997	–4.757	–5.279
2-Methyl-1-propanol	0.023	0.108	0.059
2-Methyl-1-butanol	–1.058	–0.895	–1.291
2-Methyl-2-butanol	0.339	–0.212	0.237
2-Methyl-2-pentanol	–1.118	–1.624	–1.084
2-Methyl-3-pentanol	–1.609	–1.498	–1.681
2-Methyl-2-hexanol	–2.473	–2.887	–2.394
3,3-Dimethyl-1-butanol	–2.590	–2.250	–2.224
3,3-Dimethyl-2-butanol	–1.411	–1.268	–1.291
3,5,5-Trimethyl-1-hexanol	–5.770	–5.881	–5.963
3,5-Dimethyl-4-heptanol	–5.298	–4.708	–5.483
3-Ethyl-3-pentanol	–1.917	–2.431	–2.545
3-Ethyl-3-heptanol	–5.573	–5.136	–5.176
3-Methyl-1-butanol	–1.168	–1.143	–1.246
3-Methyl-2-butanol	–0.405	–0.252	–0.305
3-Methyl-2-pentanol	–1.640	–1.373	–1.659
3-Methyl-3-pentanol	–0.830	–1.240	–1.149
3-Methyl-3-hexanol	–2.263	–2.668	–2.469
1-Butanol	0.006	–0.185	–0.178
2-Butanol	0.066	0.588	0.781
4-Methyl-1-pentanol	–2.283	–2.417	–2.557
4-Methyl-2-pentanol	–1.814	–1.768	–1.607
1-Pentanol	–1.347	–1.463	–1.489
2-Pentanol	–0.635	–0.772	–0.536
3-Pentanol	–0.486	–0.613	–0.590
1-Hexanol	–2.790	–2.782	–2.800
2-Hexanol	–1.995	–2.055	–1.847
3-Hexanol	–1.833	–1.965	–1.907
1-Heptanol	–4.167	–4.103	–4.111
3-Heptanol	–3.194	–3.277	–3.218
4-Heptanol	–3.197	–3.347	–3.224
7-Methyl-1-octanol	–5.745	–6.387	–6.490
1-Octanol	–5.401	–5.421	–5.422
2-Octanol	–4.756	–4.707	–4.469
1-Nonanol	–6.908	–6.740	–6.732
2-Nonanol	–6.320	–6.027	–5.779
3-Nonanol	–6.119	–5.915	–5.840
4-Nonanol	–5.952	–5.971	–5.846
5-Nonanol	–5.745	–5.940	–5.486

and bond lengths were used for all aliphatic alcohols considered. Amidon *et al.* have examined only 51 aliphatic alcohols, omitting the solid alcohols. Their first model relates ln SOL and the total surface area (TSA) in Å:[2]

$$\ln \text{SOL} = -0.043\ \text{TSA} + 11.78$$

$$n = 51,\ R = 0.974,\ S = 0.499,\ Q = 1.952 \tag{7}$$

Their second model relates ln SOL and the hydrocarbon surface area HYSA (HYSA = TSA – OHSA, where OHSA is the hydroxyl group surface area):

$$\ln \text{SOL} = -0.0396\ \text{HYSA} + 8.94$$

$$n = 51,\ R = 0.94,\ S = 0.706,\ Q = 1.331 \tag{8}$$

This model is poorer than above the model with TSA.

Their third model relates ln SOL and both HYSA and OHSA:

$$\ln \text{SOL} = -0.043\ \text{HYSA} - 0.060\ \text{OHSA} + 12.41$$

$$n = 51,\ R = 0.978,\ S = 0.462,\ Q = 2.117 \tag{9}$$

This model is better than model (8), but not significantly better than model (7). Our model from above (*e.g.*, QSPR model (3)) is superior to model (9), since its value of S (0.337) is much lower than the corresponding value ($S = 0.462$) for model (9).

There are several Kier and Hall models available in the literature which relate the solubility of aliphatic alcohols in water, and connectivity indices and empirical parameters. The first Kier and Hall model is given by:[37]

$$\ln \text{SOL} = 6.702 - 2.666\ {}^1\chi$$

$$n = 51,\ R = 0.978,\ S = 0.455,\ Q = 2.150 \tag{10}$$

where ${}^1\chi$ is the original Randić connectivity index (the first-order connectivity index).[38] Model (10) is improved considerably by the use of the empirical parameter c_{OH}. This parameter varies with the position of substitution: 0.707 for primary, 0.577 for secondary and 0.500 for tertiary carbon atoms.[37] With the introduction of c_{OH}, the structure-solubility model (10) becomes :

$$\ln \mathrm{SOL} = 9.204 - 2.630\ {}^{1}\chi - 4.390\ c_{\mathrm{OH}} \tag{11}$$

$$n = 50,\ R = 0.991,\ S = 0.289,\ Q = 3.430$$

This model is better than any of the models with two descriptors published in the literature. However, our model with two orthogonalized descriptors from Table VIII is somewhat better than the above Kier-Hall model:

$$\ln \mathrm{SOL} = (-3.666 \pm 0.037) + (-2.689 \pm 0.479)\ {}^{1}\Omega^{v} + (-6.336 \pm 0.479)\ \Omega_{c_{\mathrm{OH}}} \tag{12}$$

$$n = 54,\ R = 0.997,\ S = 0.273,\ F = 4479,\ Q = 3.652$$

For comparison, the values of $(\ln \mathrm{SOL})_{\mathrm{calc}}$ computed by this model are also given in Table IX. The plot between the experimental and computed values by means of Eq. (12), that is, $(\ln \mathrm{SOL})_{\mathrm{exp}}$ *vs.* $\ln (\mathrm{SOL})_{\mathrm{calc}}$, is shown in Figure 2.

CONCLUDING REMARKS

A novel approach for designing the QSPR models for predicting the solubility of aliphatic alcohols in water is proposed. The starting point of this approach is the non-orthogonal valence-connectivity basis with n descriptors

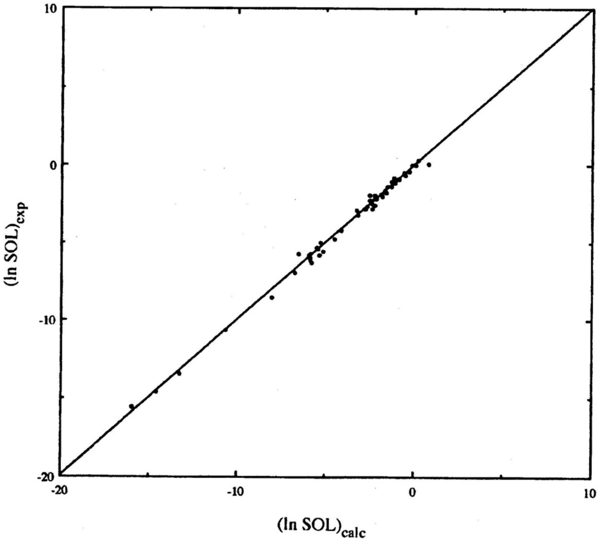


Figure 2. A plot of $(\ln \mathrm{SOL})_{\mathrm{exp}}$ *versus* $(\ln \mathrm{SOL})_{\mathrm{calc}}$ computed using Eq. (12).

(in the present study we considered the basis of 7 indices). The next step is the orthogonalization procedure as proposed by Randić.[29,30] The novelty of our approach is in allowing all the orthogonalization of valence-connectivity basis is carried out in all possible orderings, that is, in $n!$ orthogonalization orderings. This results in the appearance of QSPR models with orthogonalized valence-connectivity indices with better statistical parameters than the models based on the usual orthogonalization ordering based on decreasing the numerical value of the index. Thus, our models appear to be superior to the empirical models of Amidon *et al.*[23] and the Kier-Hall models.[24,25,37]

We also wish to point out that, due to the of problems in determining accurately the solubility of alcohols in water (as one of the referees pointed out), it can happen that the S value is lower than the experimental error. Therefore, it is advisable to use only models with two or, at most ,with three descriptors.

Acknowledgment. – This work was supported by the Ministry of Science and Technology of the Republic of Croatia through Grants 1-07-159 and 1-07-185. We thank the reviewers for their constructive comments which lead to the considerable improvement of the revised version of our paper.

REFERENCES

1. M. Randić and C.L. Wilkins, *J. Phys. Chem.* **83** (1979) 1525.
2. A. Sabljić and N. Trinajstić, *Acta Pharm.* **31** (1981) 189.
3. A. T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, *Topics Curr. Chem.* **114** (1983) 21.
4. D. H. Rouvray, *Sci. Am.* **254** (1986) 40.
5. D. E. Needham, I.-C. Wei, and P. G. Seybold, *J. Amer. Chem. Soc.* **110** (1988) 4186.
6. N. Trinajstić, S. Nikolić and S. Carter, *Kem. Ind. (Zagreb)* **38** (1989) 469.
7. S. C. Basak, G. J. Niemi, and G. D. Veith, in: *Computational Chemical Graph Theory*, D. H. Rouvray, Ed., Nova Science Publishers, Commack, NY, 1990, p. 235.
8. W. Karcher and J. Devillers, in: *Practical Applications of Quantitative Structure Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, W. Karcher and J. Devillers, Eds., Kluwer Academic Press, Dordrecht, 1990, p. 1.
9. A. Sabljić, *Sci. Total Environm.* **109/110** (1991) 197.
10. S. C. Basak, G. J. Niemi, and G. D. Veith, *J. Math. Chem.* **7** (1991) 243.
11. Z. Mihalić and N. Trinajstić, *J. Chem. Educ.* **69** (1992) 701.
12. N. Trinajstić, *Chemical Graph Theory*, 2nd revised edition, CRC Press, Boca Raton, FL, 1992, chapter 10.
13. A. Sabljić, H. Güsten, J. Hermens and A. Opperhuizen, *Environm. Sci. Technol.* **27** (1993) 1394.
14. M. Randić and P. G. Seybold, *SAR and QSAR* **1** (1993) 77.
15. M. Randić, Z. Mihalić, S. Nikolić and N. Trinajstić, *Croat. Chem. Acta* **66** (1993) 411.
16. E. C. Kirby, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1030.
17. J. Galvez, R. Garcia-Domenech, V. De Julian-Ortiz and R. Soler, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1198.

18. B. Lučić, S. Nikolić, N. Trinajstić and D. Juretić, *J. Chem. Inf. Comput. Sci.*, in press.
19. B. Lučić, S. Nikolić, N. Trinajstić, D. Juretić and A. Jurić, *Croat. Chem. Acta* **35** (1995) 532.
20. M. Randić, *J. Chem. Inf. Comput. Sci.* **32** (1992) 57.
21. M. Randić and N. Trinajstić, *New J. Chem.* **18** (1994) 179.
22. M. Randić and N. Trinajstić, *J. Mol. Struct. (Theochem)* **284** (1993) 209.
23. G. L. Amidon, S. H. Yalkowsky and S. Leung, *J. Pharm. Sci.* **63** (1974) 1858.
24. L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976, p. 151.
25. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure Activity Analysis.* Wiley, New York, 1986, p. 29.
26. P. G. Seybold, M. May and U. A. Bagal, *J. Chem. Educ.* **64** (1987) 575.
27. S. Nikolić, N. Trinajstić and Z. Mihalić, *J. Math. Chem.* **12** (1993) 251.
28. T. M. Nelson and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **34** (1994) 601.
29. M. Randić, *J. Chem. Inf. Comput. Sci.* **31** (1991) 311.
30. M. Randić, *New J. Chem.* **15** (1991) 517.
31. M. Randić, *J. Chem. Educ.* **69** (1992) 713.
32. Ref. 12, p. 35.
33. L. B. Kier and L. H. Hall, *J. Pharm. Sci.* **65** (1976) 1806.
34. D. Amić, D. Davidović-Amić and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **35** (1995) 136.
35. L. Pogliani, Amino Acids **6** (1994) 141.
36. R. B. Hermann, *J. Phys. Chem.* **76** (1972) 2754.
37. L. H. Hall, L. B. Kier and W. J. Murray, *J. Pharm. Sci.* **64** (1975) 1974.
38. M. Randić, *J. Amer. Chem. Soc.* **97** (1975) 6609.

SAŽETAK

Studij odnosa strukture i topljivosti alifatskih alkohola u vodi

Bono Lučić, Sonja Nikolić, Nenad Trinajstić, Albin Jurić
i Zlatko Mihalić

Prikazani su najbolji mogući modeli odnosa strukture i topljivosti alifatskih alkohola u vodi, koji se mogu dobiti upotrebom neortogonalizanih indeksa valentne povezanosti. Također su prikazani i odgovarajući modeli, koji se dobiju upotrebom ortogonaliziranih indeksa valentne povezanosti. Obje vrste modela su statistički identični. Zato smo predložili novi pristup modeliranju odnosa strukture i svojstava molekula, koji se temelji na razmatranju svih mogućih poredaka ortogonalizacije indeksa valentne povezanosti i na odabiranju dominantnih deskriptora pri gradnji modela. Taj novi pristup daje modele odnosa strukture i topljivosti alifatskih alkohola u vodi, koji su bolji od postojećih empirijskih modela i Kier-Hall-ovih modela.