

Artificial Neural Network Approach to Predict the Solubility of C<sub>60</sub> in Various Solvents

István Z. Kiss\*,† Géza Mándi, and Mihály T. Beck\*,‡

Institute of Physical Chemistry, University of Debrecen, H-4010 Debrecen, P.O. Box 7, Hungary

Received: February 24, 2000; In Final Form: May 19, 2000

A multiparameter artificial neural network (ANN) approach was successfully utilized to predict the solubility of C<sub>60</sub> in different solvents. Molar volume, polarizability parameter, LUMO energy, saturated surface, and average polarizability molecular properties were chosen to be the most important factors determining the solubilities. The results show that in a large number of solvents (126) the solubility decreases with increasing molar volumes of the solvents and increases with their polarizability and saturated surface areas. A method is suggested to the approximate determination of experimentally not easily measurable solubility related thermodynamic parameters, e.g., the Hildebrand parameter, based on reliable solubility measurements.

## I. Introduction

The solubility of C<sub>60</sub> in different solvents has been extensively studied in the past few years.<sup>1–4</sup> The experimental studies were mainly motivated to improve the efficiency of the separation process of fullerenes. Moreover, C<sub>60</sub>, being a globular molecule, provides an excellent example for theoretical studies. Single solvent properties were applied to account for the extremely different solubilities,<sup>1–3</sup> but these attempts were not successful. Multiparameter studies based on linear free energy relationship (LFER) provided better results.<sup>5,6</sup> These studies suffer from two drawbacks. First, they assume that the chosen solvent property contribute to solubility with same weight for all solvents. This assumption may be valid for a family of solvents (e.g., for a homologous series) but not for widely different ones. Second, the solvent properties (especially the Hildebrand parameter) required for LFER studies are not easily available, therefore only a subset of the solvents (61 in ref 5, 47 in ref 6) were involved in the calculations. Our study was focused on finding molecular parameters that are readily available meanwhile providing satisfactory results for 134 solvents. To obtain better results ANNs were applied to introduce nonlinearity in the numerical treatment. ANNs provide a very efficient way of extracting important information from large quantities of data in chemistry.<sup>7,8</sup>

The Hildebrand parameter  $\delta^9$  is an important thermodynamic parameter widely used in solubility studies. However, calculation of the Hildebrand parameter involves the precise measurement of evaporation enthalpy. Since the Hildebrand parameter (among other parameters as well) is related to solubility, in theory, it is possible to determine the Hildebrand parameter based on reliable solubility measurements. We present a procedure to calculate the Hildebrand parameters of different solvents based on the solubilities of C<sub>60</sub> and I<sub>2</sub> in them.

The paper is structured as follows. Experimental and numerical methods are described in section II. Selection procedure of important solvent properties is presented in section III. An ANN algorithm was applied to check the chemists' old "similar dissolves similar" rule of thumb in section IV. The calculation

of C<sub>60</sub> solubility and Hildebrand parameter are presented in sections V and VI, respectively. The results are discussed and compared to the LFER studies in section VII.

## II. Solvent Properties

The solubilities of C<sub>60</sub> in different solvents (see Table 1) were taken from the comprehensive review of Beck and Mándi.<sup>4</sup> The solubility was usually measured using photometry, but for some solvents <sup>13</sup>C NMR method and gravimetry was applied by different investigators. For some solvents (e.g., benzene) different solubilities were measured using different methods. In these cases the solvents were not used directly in the fitting procedures therefore we could compare the different experimental values with the predicted ones.

Since the experimental solvent properties are usually difficult to acquire we restricted our study on easily measurable physical properties (e.g., density) or physical properties that can be calculated numerically. The following properties were applied in the numerical treatment (see Table 1, Table 2 and Table 1S): molar volume ( $V_m$ ), polarizability parameter ( $[n^2 - 1]/[n^2 + 2]$ ), molar refraction ( $R_m = V_m[n^2 - 1]/[n^2 + 2]$ ), experimental polarizability ( $\alpha = 3\epsilon_0 R_m/N_A$ ), experimental polarizability volume ( $\alpha' = \alpha/(4\pi\epsilon_0)$ ), HOMO energy, LUMO energy, exchange energy, total electrostatic interaction, electron–electron repulsion, electron–nuclear attraction, total energy, saturated surface area, unsaturated surface area, polar surface area, calculated dipole moment, average polarizability, and polarizability volume (for a definition of parameters see MOPAC 6.0 manual<sup>10</sup> and references therein).  $V_m$  was calculated from density and molecular weight, polarizability parameter from the refraction index ( $n$ ). The geometry of a solvent molecule was preoptimized by PCModel (commercially available) program based on van der Waals radii of the atoms. The geometry was further optimized by subsequent application of MMX (or Hückel method for aromatic molecules) and semi-empiric AM1 methods. The parameters (using the optimized geometry) were calculated using MOPAC 6.0<sup>10</sup> and PCModel software packages on a PC (Table 2 lists the specific program used for the calculation). Note that there is correlation between some of the experimental ( $R_m$ ,  $\alpha$ , and  $\alpha'$ ) and the calculated

\* Corresponding author.

† E-mail: ikiss@delfin.klte.hu.

‡ E-mail: beckmt@tigris.klte.hu.

TABLE 1: The Applied Solvents in the Numerical Treatment<sup>a</sup>

solvents	$G_{\text{exp}}$	$G_{\text{ANN}}$	Ig $S_{\text{exp}}$	Ig $S_{\text{ANN}}$	$\delta_{\text{exp}}/\text{MPa}^{0.5}$	$\delta(\text{C}_{60})/\text{MPa}^{0.5}$	$\delta(\text{I}_2)/\text{MPa}^{0.5}$
pentane	0	0	-6.1	-6.9	14.3	16	15
hexane	1	1	-5.1	-5.9	14.9	18	17
octane	1	1	-5.2	-5.2	15.6	15	
iso-octane			-5.2	-5.9		15	14
decane	1	1	-4.7	-4.7	13.5	14	
<b>dodecane</b>	2	1	-3.5	-4.8	16.2	17	
tetradecane	1	1	-4.3	-4.6		13	
cyclopentane	0	0	-6.6	-6.3	17.8	17	16
cyclohexane	1	1	-5.3	-5.1	16.8	19	15
cis-decahydronaphthalene	3	1	-3.3	-3.4	18	19	14
trans-decahydronaphthalene			-3.5	-3.6		19	14
1,5,9-cyclododecatriene			-2.7	-3.3		20	
cyclopentyl bromide	2	2	-4.2	-3.7		20	
cyclohexyl chloride	2	3	-4.1	-3.9		20	
cyclohexyl bromide	3	3	-3.4	-3.5		20	
cyclohexyl iodide	4	4	-2.8	-3.3		20	
1,2-dibromocyclohexane	4	*	-2.6	-3.3		20	
cyclohexene	2	2	-3.8	-4.0	15.1	20	
1-methyl-1-cyclohexene	2	3	-3.8	-3.8		20	
methylcyclohexane	1	1	-4.5	-4.9	15.9	19	
cis-1,2-dimethylcyclohexane	1	1	-4.6	-4.6		18	
trans-1,2-dimethylcyclohexane			-4.6	-4.8		18	
ethylcyclohexane	1	1	-4.3	-4.5		19	
<b>dichloromethane</b>	1	1	-4.6	-5.7	19.8	20	22
chloroform	1	*	-4.8	-4.9	19	20	20
carbon tetrachloride	1	1	-4.4	-4.6	17.6	20	17
dibromomethane	1	1	-4.5	-4.2		20	
bromoform	3	*	-3.2	-3.6		20	21
<b>iodomethane</b>	2	2	-4.2	-5.5	20.9	22	
diiodomethane			-4.8	-4.9	24.1	24	
bromochloromethane	2	1	-4.2	-4.6		20	
bromoethane	1	1	-5.2	-5.3	19.6	20	
iodoethane	1	1	-4.5	-4.5	19.2	20	
1,1,2,2-tetrachloroethane	3	3	-3.1	-4.1	19.8	20	
1,1,2-trichlorotrifluoroethane	0	0	-5.6	-5.1	14.9	20	
1,2-dichloroethane	1	1	-5.0	-4.7	20	20	
1,2-dibromoethane	2	2	-4.2	-3.7	19.8	20	
1,1,1-trichloroethane	1	1	-4.7	-4.7		20	
1-chloropropane	0	0	-5.6	-5.6	17.4	21	
1-bromopropane	1	1	-5.2	-4.8	18.2	20	
1-iodopropane	1	1	-4.6	-4.3		20	
2-chloropropane	0	0	-5.9	-5.9	16.6	21	
2-bromopropane	0	1	-5.4	-5.0		20	
2-iodopropane <sup>6</sup>	1	1	-4.8	-4.4		20	
1,2-dichloropropane	1	1	-4.9	-4.6	18.4	20	
1,3-dichloropropane	1	1	-4.8	-4.5		20	
(+)-1,2-dibromopropane	2	*	-4.3	-3.7	21.3	20	
1,3-dibromopropane	2	2	-4.2	-3.6		20	
1,3-diiodopropane	3	*	-3.4	-3.2		20	
1,2,3-tribromopropane	4	4	-2.9	-3.2		20	
1,2,3-trichloropropane	2	*	-4.0	-4.1		20	
1-bromo-2-methylpropane	1	1	-4.9	-4.8		20	
1-chloro-2-methylpropane	0	0	-5.4	-5.4		21	
1-iodo-2-methylpropane	2	2	-4.3	-4.4		20	
2-chloro-2-methylpropane	0	0	-5.7	-5.6		21	
2-bromo-2-methylpropane	1	1	-5.0	-5.1		20	
2-iodo-2-methylpropane	1	*	-4.4	-4.6		20	
1,2-dibromoethylene	2	2	-3.7	-4.0	20.7	20	
trichloroethylene	2	2	-3.8	-4.4	18.8	20	
tetrachloroethylene	2	*	-3.8	-4.0	19	20	
1-chloro-2-methylpropene	1	1	-4.5	-4.7		20	
benzene			-4.0	-4.2	18.8	20	19
toluene	3	3	-3.4	-3.7	18.2	20	16
1,2-dimethylbenzene	4	3	-2.9	-3.4		20	
1,3-dimethylbenzene	3	3	-3.3	-3.3		20	
1,4-dimethylbenzene	3	3	-3.3	-3.3		20	
1,2,3-trimethylbenzene	3	2	-3.1	-3.0		20	
1,2,4-trimethylbenzene	4	4	-2.5	-2.9		20	
1,3,5-trimethylbenzene			-3.5	-3.1	18	20	
1,3,5-trimethylbenzene			-3.5	-3.1		20	
1,2,3,4-tetramethylbenzene	4	4	-2.9	-2.8		20	
1,2,3,5-tetramethylbenzene	4	4	-2.4	-2.8		20	
tetralin	4	4	-2.5	-2.7		20	
ethylbenzene	3	3	-3.4	-3.5	18	20	

TABLE 1 (Continued)

solvents	$G_{\text{exp}}$	$G_{\text{ANN}}$	$\lg S_{\text{exp}}$	$\lg S_{\text{ANN}}$	$\delta_{\text{exp}}/\text{MPa}^{0.5}$	$\delta(\text{C}_{60})/\text{MPa}^{0.5}$	$\delta(\text{I}_2)/\text{MPa}^{0.5}$
<i>n</i> -propylbenzene	2	2	-3.5	-3.4	17.6	20	
iso-propylbenzene			-3.6	-3.5		20	
<i>n</i> -butylbenzene	3	2	-3.4	-3.3		20	
<i>sec</i> -butylbenzene	2	2	-3.6	-3.4		20	
<i>tert</i> -butylbenzene	2	2	-3.7	-3.5		20	
fluorobenzene	2	<i>b</i>	-4.1	-4.3		20	
chlorobenzene	3	3	-3.0	-3.6	19.4	20	
bromobenzene	3	<i>b</i>	-3.3	-3.3	20.2	20	
iodobenzene	3	3	-3.5	-3.2	20.7	20	
1,2-dichlorobenzene	4	4	-2.4	-3.2	20.5	20	
1,3-dichlorobenzene	3	4	-3.4	-3.2		20	
1,2-dibromobenzene	4	4	-2.6	-2.9		20	
1,3-dibromobenzene	4	4	-2.6	-2.8		20	
1,2-Br-Cl-benzene			-2.4	-3.0		20	
1,3-Br-Cl-benzene			-3.0	-3.0		20	
1,2,4-trichlorobenzene	4	4	-2.8	-2.8		20	
styrene	3	3	-3.2	-3.3		20	
<b><i>o</i>-cresol</b>	0	0	-5.7	-3.3		20	
nitrobenzene	2	2	-3.9	-4.8	20.5	20	
benzonitrile	2	2	-4.2	-4.1	17.2	20	
anisole	3	<i>b</i>	-3.1	-3.3		20	
benzaldehyde	2	2	-4.2	-3.8	19.2	20	
phenyl isocyanate			-3.4	-3.5		20	
2-nitrotoluene	3	3	-3.4	-3.5		20	
3-nitrotoluene	3	3	-3.4	-3.5		20	
thiophenol			-3.0	-2.9		20	
benzyl chloride	3	3	-3.4	-3.1		20	
benzyl bromide			-3.1	-2.9		20	
$\alpha,\alpha,\alpha$ -trichlorotoluene			-3.0	-2.9		20	
1-methylnaphthalene			-2.2	-2.4		20	
dimethylnaphthalenes	4	4	-2.1	-2.3		20	
1-phenylnaphthalene	4	4	-1.9	-2.1		20	
1-chloronaphthalene	4	4	-2.0	-2.4		20	
1-bromo-2-methylnaphthalene	4	4	-2.1	-2.3		20	
methanol	4	4	< -8	-7.6	29.6	29	
ethanol	0	0	-7.1	-7.5	26	26	20
1-propanol	0	0	-6.4	-7.1	24.3	24	
1-butanol	0	0	-5.9	-6.8	23.3	22	
1-pentanol	0	0	-5.3	-6.1	22.3	21	
1-hexanol	0	0	-5.1	-5.3	21.9	20	
1-octanol	1	<i>b</i>	-5.0	-4.6	21.1	18	
nitromethane	1	1	< -8	-7.1	26	23	
nitroethane	0	0	-6.7	-5.8	22.7	21	
acetone	0	<i>b</i>	-7.0	-6.4	20.2	22	
acetonitrile	0	0	< -8	-7.7	24.3	27	
<b><i>n</i>-butylamine</b>			-3.3	-7.0	17.8	18	
<b>acrylonitrile</b>			-6.4	-7.6	21.5	21	
2-methoxyethyl ether	0	0	-5.2	-4.7		20	
<i>N,N</i> -dimethylformamide	1	1	-5.3	-5.3	20.5	21	19
<b>tetrahydrofuran</b>	0	0	< -8	-6.4	18.6	21	
<b>tetrahydrothiophene</b>	0	0	-5.4	-3.4		20	
thiophene	0	<i>b</i>	-4.4	-4.3	20.1	20	
2-methylthiophene	1	1	-3.0	-3.3		20	
<i>N</i> -methyl-2-pyrrolidone	3	3	-3.9	-3.7	23.1	20	
pyridine	2	2	-4.0	-4.4	21.9	20	
quinoline	2	2	-2.9	-2.7	22.1	20	
aniline	4	4	-3.9	-3.5	21.1	20	
<i>N</i> -methylaniline			-3.8	-3.0		20	
<i>N,N</i> -dimethylaniline			-3.2	-2.8	19.8	20	

<sup>a</sup> The experimental ( $\lg S_{\text{exp}}$ ) and calculated ( $\lg S_{\text{ANN}}$ ) solubility of C<sub>60</sub> in them and the corresponding group numbers ( $G_{\text{exp}}$  and  $G_{\text{ANN}}$ , see text for definition). The experimental ( $\delta_{\text{exp}}$ ) and ANN predicted (based on the solubility of C<sub>60</sub> and I<sub>2</sub>,  $\delta(\text{C}_{60})$  and  $\delta(\text{I}_2)$ , respectively) Hildebrand parameters are also presented. The outlier solvents are boldfaced. <sup>b</sup> Not classified.

(average polarizability, polarizability volume) solvent properties. These values were used as counterchecks during the numerical procedures (i.e., similar results must be obtained for these parameters provided that both the ANN and the quantum mechanical calculations are properly done).

The selection of solvent properties is always somewhat arbitrary and with increasing the number of parameters along with the number of solvents the quality of fitting should improve. In our study we tried to find some important solvent parameters

with which satisfactory results can be obtained while keeping the number of parameters relatively small.

The solubilities are not given in weight units (e.g., mg/mL), but in terms of logarithmic values of molar fractions ( $\log S$ ) because the  $\log S$  values correspond to the free energy changes in the solvation process. For some solvents zero solubility values were reported therefore the respective  $\log S$  is undeterminable. To compare the measured and numerically fitted values we used the experimentally detectable < -8 notation in the tables.

**TABLE 2: The Applied Solvent Properties, Their Origin (experimental (exptl) or calculated by the given programs), and the Number of Successfully Identified Group Number of Solvents Using One-Parameter Approach**

parameter	origin	identified solvents
molar volume of the solvent ( $V_m$ , cm <sup>3</sup> mol <sup>-1</sup> )	exptl	72
polarizability parameter ( $(n^2 - 1)/(n^2 + 2)$ )	exptl	75
molar refraction ( $R_m$ )	exptl	69
exptl polarizability $\alpha$ ( $\times 10^{33}$ , C <sup>2</sup> m <sup>2</sup> J <sup>-1</sup> )	exptl	67
exptl polarizability volume $\alpha'$ ( $\times 10^{23}$ , cm <sup>3</sup> )	exptl	69
$E_{\text{HOMO}}$ (eV)	MOPAC	67
$E_{\text{LUMO}}$ (eV)	MOPAC	69
exchange energy $E_{\text{exchange}}$ (eV)	MOPAC	70
total electrostatic interaction $E_{\text{tot-stat}}$ (eV)	MOPAC	72
electron-electron repulsion $E_{\text{el-el}}$ (eV)	MOPAC	67
electron-nuclear attraction $E_{\text{el-nu}}$ (eV)	MOPAC	69
total energy $E_{\text{tot}}$ (eV)	MOPAC	71
saturated surface $S_{\text{sat}}$ (Å <sup>2</sup> )	PCMODEL	69
unsaturated surface $S_{\text{unsat}}$ (Å <sup>2</sup> )	PCMODEL	67
polar surface $E_{\text{pol}}$ (Å <sup>2</sup> )	PCMODEL	69
calculated dipole moment $D_c$ (debye)	MOPAC	66
average polarizability $P_a$ ( $\times 10^{23}$ ESU)	MOPAC	66
polarizability volume $V_{\text{pol}}$ (Å <sup>3</sup> )	MOPAC	69

The solubilities of iodine<sup>11</sup> (see Table 3S) in some solvents were also used in this study to calculate the Hildebrand parameters of the solvents (see Table 1).<sup>12</sup>

Following a standard ANN methodology,<sup>7</sup> all parameters (involving log  $S$  values) were linearly rescaled in the region of 0.1–0.9 based on the minimum and maximum values of the specific solvent property.

### III. Choosing the Most Important Solvent Properties

One hundred fourteen solvents were classified into five different groups according to the solubility of C<sub>60</sub> in them. The borderlines of the groups were selected so that approximately the same number of solvents can be found in each group. The respective ranges are as follows: group #0:  $0-5 \times 10^{-6}$ , group #1:  $5 \times 10^{-6}-5 \times 10^{-5}$ , group #2:  $5 \times 10^{-5}-3 \times 10^{-4}$ , group #3:  $3 \times 10^{-4}-1 \times 10^{-3}$ , group #4:  $1 \times 10^{-3}-1$ . The group numbers of the solvents are given in Table 1 ( $G_{\text{exp}}$ ). We applied Geremia Neural Network Development System 2.0 by Active Record Software System, Ltd. (Hungary) to train an ANN to

classify the solvents into the correct groups. Using only a single solvent property (see Table 2) about 58%–66% of the solvents can be successfully classified (the best results with polarizability parameter, 75 correct response, while the worst results with average polarizability and calculated dipole moment, 66 correct response, are obtained). The accuracy of the success rate can be estimated from the correlated parameters ( $R_m$ ,  $\alpha$ ,  $\alpha'$ ,  $P_a$ ,  $V_{\text{pol}}$ ), and was found to be 1%. Combination of two solvent properties (all possible combinations were checked, the best pairs are shown in Table 3) greatly enhances the performance of the network: with polarizability parameter and the average polarizability 75% of the solvents are classified correctly. Note that the average polarizability alone gives poor predictions (58%) but the prediction is much better if it is combined with other parameters, e.g., with the polarizability parameter. Similar synergetic-like effect can be observed with LUMO energy and the molar volume. The LUMO energy alone has a success rate of 60%; however, the rate increases to 73% with molar volume (the combination of molar volume and polarizability parameter results in 72% success). On the basis of the results of one- and two-parameter combinations the five most successful solvent properties were chosen: molar volume, polarizability parameter, LUMO energy, saturated surface area, and average polarizability. With these parameters the network classified correctly 91 solvents, classified incorrectly 10 solvents and could not classify 13 solvents (80% success rate, see Table 1). Adding a sixth parameter (out of the remaining thirteen) does not increase the performance of the network (see Table 4), therefore we concluded that the listed five parameters are the most dominant factors determining the solubility of C<sub>60</sub>.

### IV. Kohonen Network Approach

Kohonen ANNs have been widely applied to map a multi-dimensional parameter space to a two-dimensional array of neurons. The projection is made so that the topological relations are preserved. The “similar dissolves similar” rule of thumb means that solvents with similar parameter combinations will result in similar solubility. Therefore, provided that proper molecular parameters were chosen, if two solvents are mapped onto the same (or nearby) neuron the solubility should be close in them. We note that the inverse statement is not necessarily

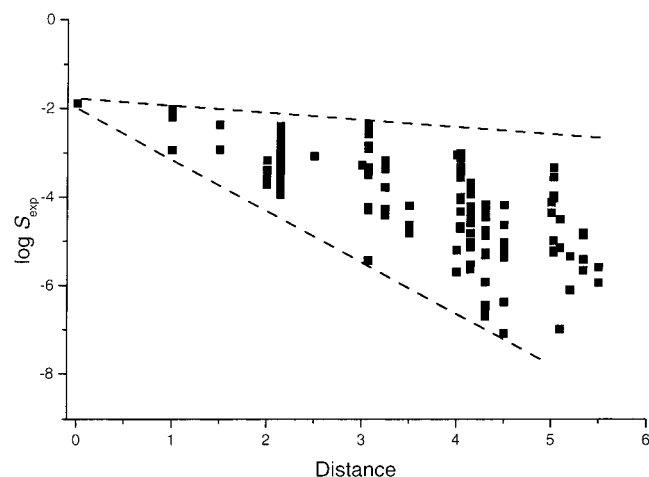
**TABLE 3: The Number of Successfully Identified Group Number of Solvents Using Two- and Three-Parameter Combinations**

parameters	identified solvents	parameters	identified solvents
$V_m, (n^2 - 1)/(n^2 + 2)$	82	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}$	80
$V_m, E_{\text{LUMO}}$	83	$V_m, (n^2 - 1)/(n^2 + 2), S_{\text{sat}}$	80
$V_m, S_{\text{sat}}$	76	$V_m, (n^2 - 1)/(n^2 + 2), P_a$	87
$V_m, P_a$	77	$V_m, E_{\text{LUMO}}, S_{\text{sat}}$	83
$(n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}$	75	$V_m, E_{\text{LUMO}}, P_a$	81
$(n^2 - 1)/(n^2 + 2), S_{\text{sat}}$	78	$V_m, S_{\text{sat}}, P_a$	84
$(n^2 - 1)/(n^2 + 2), P_a$	85	$(n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}$	84
$E_{\text{LUMO}}, S_{\text{sat}}$	79	$(n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, P_a$	90
$E_{\text{LUMO}}, P_a$	82	$(n^2 - 1)/(n^2 + 2), S_{\text{sat}}, P_a$	86
$S_{\text{sat}}, P_a$	82	$E_{\text{LUMO}}, S_{\text{sat}}, P_a$	89

**TABLE 4: The Number of Successfully Identified Group Number of Solvents Using Four and Six-parameter Combinations**

parameters	identified solvents	parameters	identified solvents
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}$	84	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{tot-stat}}$	90
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, P_a$	83	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{el-el}}$	84
$V_m, (n^2 - 1)/(n^2 + 2), S_{\text{sat}}, P_a$	87	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{el-nu}}$	84
$V_m, E_{\text{LUMO}}, S_{\text{sat}}, P_a$	86	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{tot}}$	79
$(n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a$	91	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, S_{\text{unsat}}$	87
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, R_m$	87	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{pol}}$	89
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, \alpha$	85	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, D_c$	78
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, \alpha'$	86	$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, V_{\text{pol}}$	84
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{HOMO}}$	91		
$V_m, (n^2 - 1)/(n^2 + 2), E_{\text{LUMO}}, S_{\text{sat}}, P_a, E_{\text{exchange}}$	88		





**Figure 1.** The solubility of C<sub>60</sub> in the solvent as a function of the distance between the ANN projection of C<sub>60</sub> and that of the specific solvent on the Kohonen map (Figure 1S). Dashed lines show the approximate area of the points.

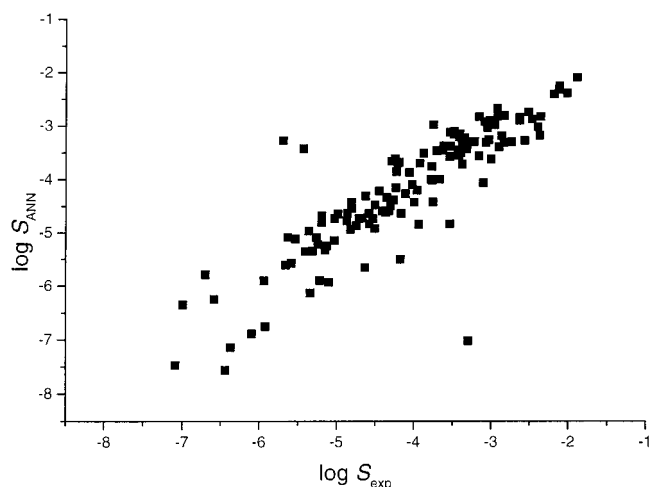
true. Solvents in which the solubility is similar might have different molecular parameters: the change of one parameter might compensate the change of other parameter(s). We expect the ANN to reveal hidden relationships between the provided data sets.

A 10 × 10 two-dimensional neuron-map was constructed to project the 134 solvents with the five solvent parameters determined in section III. The projection was carried out following standard ANN procedures.<sup>7</sup> The neuron map is toroidal, maximum signal criterion was applied with triangular function for scaling corrections on neighbor weights.

The obtained two-dimensional map is presented in the supporting material (Figure 1S). Interestingly, the solvents with largest solubilities are mapped onto the same neuron as C<sub>60</sub> (since C<sub>60</sub> is not a solvent, it has not been used in the construction of the ANN) and the solvents with small solubilities are mapped further from this point. In Figure 1 the solubility is shown as a function of the distance from the C<sub>60</sub> center. The correlation is remarkable, especially if we keep in mind that the actual solubility values were not used explicitly to make the projection. The further the solvent from the position of C<sub>60</sub> is being mapped, the smaller solubility is measured. We do not expect a straight line (i.e., linear correlation) in this graph since the variation of the solubility in *x* and *y* direction on the neuron map might be different and nonlinear.

## V. Calculating the Solubility of C<sub>60</sub>

Since the classification and Kohonen approach ensured that the chosen parameters are suitable we constructed an ANN to estimate the solubility of C<sub>60</sub>. An ANN of five input units, five hidden units, and one output unit was applied. Jetnet 3.0 freeware software<sup>13</sup> was used for the training (Langevin type network with bold driver method for dynamically changing the learning parameters). Twenty thousand epochs were used during a training procedure (with further increasing the length of training the predictive power of the network seriously decreased). The architecture of the network was determined using a bottom-top approach, i.e., increasing the number of hidden units until overfitting occurs and the quality of the fit becomes worse. The results are given in Table 1. Correlation between the experimental and calculated solubility values are shown in Figure 2. The standard deviation was found to be  $\sigma = 0.58$ . Excluding the solvents that have deviation of more than  $2\sigma$



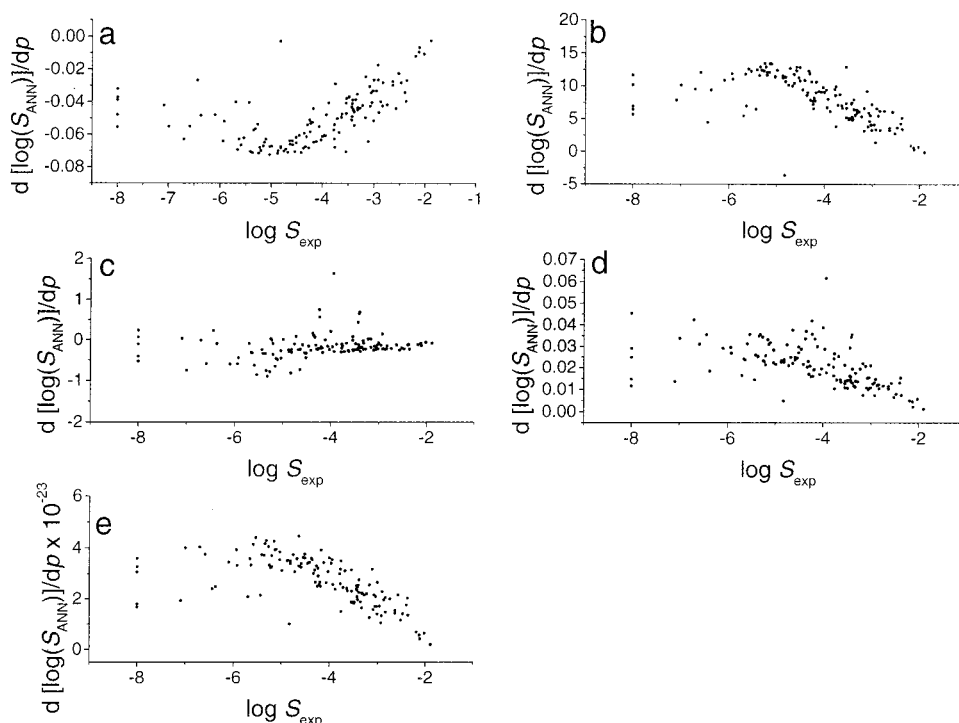
**Figure 2.** Calculated vs experimental logarithmic solubilities.

(outliers), the standard deviation decreased to 0.45. To evaluate the effect of a molecular parameter on the solubility we determined the  $[d \log S_{\text{ANN}}/d \text{ solvent parameter}]$  derivatives ("weights") numerically for each solvent (Table 2S). The derivative represents a coefficient that is optimal for a linear fit of the considered solvent. Because of the nonlinear numerical treatment of the problem, we expected to obtain different derivatives for different solvents. The derivatives are shown as a function of the experimental solubility in Figure 3. Generally (except the graph for  $E_{\text{LUMO}}$ ) the derivatives get closer to zero as the solubility increases. The sign of derivatives determines whether solubility is increased or decreased with increasing the parameter. On the basis of the ANN prediction, the solubility decreases with increasing molar volume, increases with polarizability parameter, saturated surface area and average polarizability, and does not have a definite tendency with LUMO energy. The peculiar feature of LUMO energy is probably related to the fact that it is not a dominant factor for all the solvents (as it was found in section III), but can be an important cofactor for a certain group of solvents.

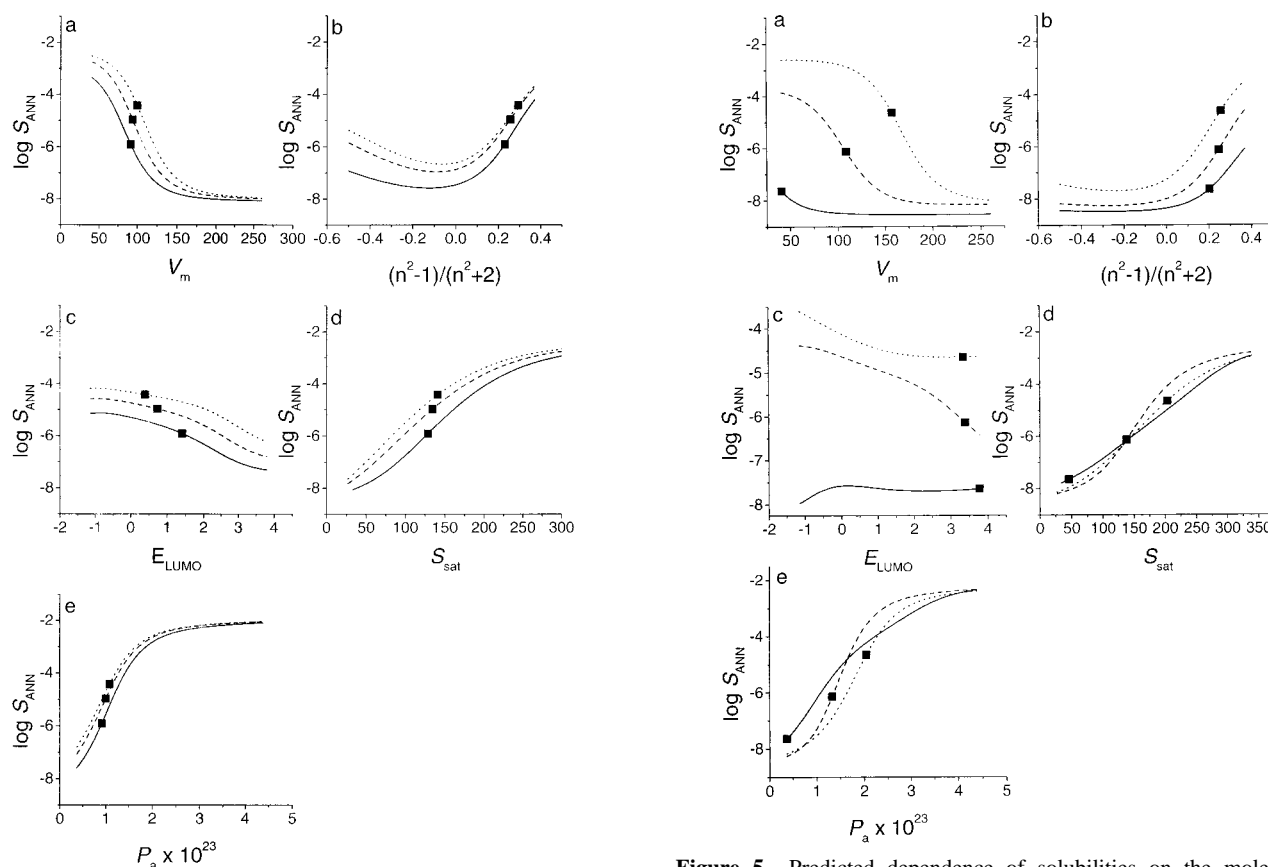
One may expect that for a family of solvents derivatives will be approximately the same. To check this assumption, predicted solubility vs molecular parameter graphs were created. These graphs show what would be the solubility in a specific solvent if it had different parameter. The graphs can be similar for a family of solvents. Figure 4 shows the graphs for halogen substituted alkanes. The curves are usually slightly shifted but have the same shape. However, alcohols (Figure 5) exhibit qualitatively different variations and derivatives indicating that the solubility properties can be relatively different even in a homologous series. These figures also justify the application of nonlinear treatment: if linear relationships had been observed the application of ANNs would have been irrelevant.

## VI. Calculating Hildebrand Parameters

Once solubility is known for a given solvent one might relate it to other important thermodynamic quantities. Therefore, an ANN can be trained, for example, to predict the Hildebrand parameter ( $\delta$ ) of the solvent if the solubility is known. As it was shown by Ruoff et al.,<sup>3</sup> the solubility vs  $\delta$  parameter graph consists of largely scattered points with a maximum at about the Hildebrand parameter of C<sub>60</sub>. Both the large error and the observed maximum (in agreement with the theoretical predictions<sup>1,9</sup>) imply that additional parameters should be considered to calculate the Hildebrand parameter from a reliable solubility



**Figure 3.** The calculated numerical derivative of the solubilities with respect to the specific molecular property (a–e: molar volume, polarizability parameter, LUMO energy, saturated surface, average polarizability, respectively) vs experimental solubilities for the solvents.



**Figure 4.** Predicted dependence of solubilities on the molecular parameters for different solvents. Solvents: 2-X-propane, (solid) X = Cl, (dashed) Br, (dotted) I. a–e for different parameters: molar volume, polarizability parameter, LUMO energy, saturated surface, average polarizability, respectively.

measurement. As a demonstration of the strength of the ANN approach we present a calculation of Hildebrand parameters of

**Figure 5.** Predicted dependence of solubilities on the molecular parameters for different solvents. Homologous series of alcohols. (solid) methanol, (dashed) 1-pentanol, (dotted) 1-octanol. a–e for different parameters: molar volume, polarizability parameter, LUMO energy, saturated surface, average polarizability, respectively.

the solvents using four parameters (polarizability parameter, LUMO energy, average polarizability, and the solubility of  $C_{60}$ ) with a 4–2–1 network (see Figure 6). Note that in this case

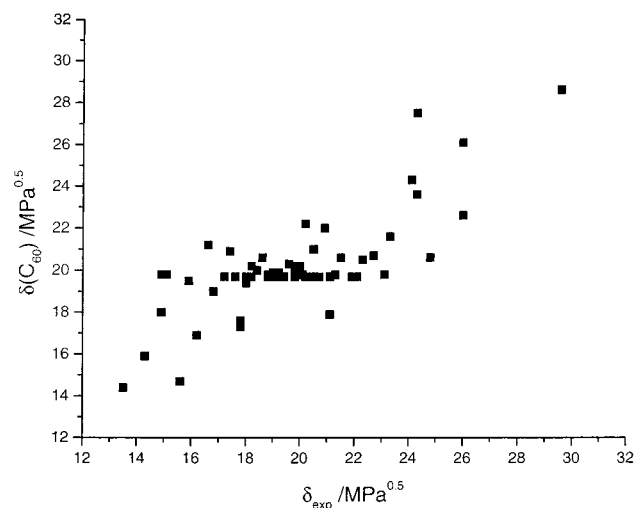


Figure 6. Calculated vs experimental Hildebrand parameters.

the size of the network was reduced due the lack of available data in the training data set and the applied parameters were not optimized. Similar calculations were carried out using the solubilities of I<sub>2</sub> instead of C<sub>60</sub> in different solvents. The standard deviations of the predicted and experimental Hildebrand parameters were 2.0 in both cases. Hence a rough estimate ( $\pm 10\%$ ) can be obtained for the Hildebrand parameter of the solvents (see Table 1).

## VII. Discussion

We demonstrated that ANNs can be trained to predict the solubility of C<sub>60</sub> in various solvents. The parameters required for the procedure are either calculated numerically or easily available experimentally (density and refraction index measurements). The “similar dissolves similar” rule of thumb was reassured using a Kohonen-type network. Similarity is expressed as similar molecular parameter combinations. However, when other properties are considered, it is likely that the criterion of similarity will be different.

The ANN appeared to be very efficient in finding hidden similarity relations in a relatively large solubility related data set. The relations are chemically correct: the derivatives of the solubility with respect to the different molecular parameters are approximately the same for certain group of solvents. The ANN predicts the logarithmic solubility of the solute with good accuracy of 0.45 for 126 solvents out of the applied 134. This standard deviation is comparable to the previous works ( $\sigma = 0.40$  in ref 5 and  $\sigma = 0.47$  in ref 6) but in our treatment 134 solvents were involved. The trend of the nonlinear fit of the ANN is the same as in the LFER studies: the logarithmic solubility decreases with molar volume of the solvent<sup>5</sup> and increases with increasing polarizability parameter,<sup>6</sup> saturated surface area, and average polarizability.

Different polyenes are reported to react with C<sub>60</sub>.<sup>14</sup> We measured the (logarithmic) solubility in cyclopentadiene, cycloheptatriene, and cyclooctatriene experimentally (−2.6, −3.2, and −2.8, respectively). The numerical predictions are −4.9, −3.4, and −3.2, respectively. These values are smaller than the experimental ones; however, the difference is near the error limit (0.45) for cycloheptatriene and cyclooctatetraene. There is a remarkable difference between the predicted and experimental values for cyclopentadiene. This is an indication of chemical reaction between the solvent and C<sub>60</sub>. If the difference between the predicted and experimental solubility is large (as in the case of cyclopentadiene) chemical reaction is possible either in the

TABLE 5: Outlier Solvents (“+” Sign) with Different Approaches

solvents	ANN	Flunt et al. <sup>6</sup>	Marcus <sup>5</sup>
pentane			+
dodecane	+	+	
cis-decahydronaphthalene		+	
dichloromethane	+		
iodomethane	+		
trichloroethylene			+
1-iodopropane			+
1,2-dimethylbenzene			+
1,3,5-trimethylbenzene			+
chlorobenzene			+
iodobenzene			+
1,2-dichlorobenzene			+
2-methylphenol			+
<i>o</i> -cresol	+		
anisole			+
1-pentanol			+
1,4-dioxane			+
acrylonitrile	+		
<i>N,N</i> -dimethylformamide			+
tetrahydrofuran	+		
tetrahydrothiophene	+		+
<i>N</i> -methylpyrrolidinone			+
<i>n</i> -butylamine	+		

liquid or in the solid phase. However, the agreement of the predicted and experimentally found solubility does not prove the absence of chemical reaction.

There are several reported different values of solubility of C<sub>60</sub> in benzene: −3.7<sup>2</sup>, −3.7<sup>3</sup>, −4.0,<sup>4</sup> −4.0.<sup>14</sup> The ANN algorithm predicts −4.2 therefore the numerical result is consistent with the experimental data of Beck and Mándi<sup>4</sup> and Letcher et al.<sup>14</sup>

Table 5 lists the outlier solvents and compares them with the results of LFER studies. The large solubility of some outliers (e.g., *n*-butylamine, cyclopentadiene) can be related to chemical reaction. The prediction of ANN and LFER studies was rather poor for long-chain alkanes (e.g., *n*-dodecane).

Solubility is usually easily measurable thermodynamic quantity. Therefore, solubility related thermodynamic quantities could be determined by simple solubility measurements using proper fitting techniques such as ANNs. We proposed an algorithm to determine approximately the Hildebrand parameter of a solvent by measuring the solubility of C<sub>60</sub> and/or I<sub>2</sub> in them. The obtained results can be regarded as a first evaluation and can be useful when no other data is available. The safest is to measure the solubility of both C<sub>60</sub> and I<sub>2</sub> and accept the values if they are close to each other (e.g., the Hildebrand parameter of CHBr<sub>3</sub> is predicted to be approximately 20–21 MPa<sup>0.5</sup>).

**Acknowledgment.** This work was supported by the following Hungarian Grants: OTKA 16694, OTKA 25222.

**Supporting Information Available:** The applied solvent properties (Table 1S), the numerical derivatives of the logarithmic solubility with respect to the solvent properties (Table 2S), and the solubility of I<sub>2</sub> (Table 3S), as well as the Kohonen map of the solvents (Figure 1S). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. G.; Mathews, C. K. *J. Org. Chem.* **1992**, *57*, 6077.
- (2) Scrivens, W. A.; Tour, J. M. *J. Chem. Soc., Chem. Commun.* **1993**, 1207.
- (3) Ruoff, R. F.; Tse, D. S.; Malhotra, R.; Lorents, D. C. *J. Phys. Chem.* **1993**, *97*, 3379.
- (4) Beck, M. T.; Mándi, G. *Fullerene Sci. Technol.* **1997**, *5*.

- (5) Marcus, Y. *J. Phys. Chem. B* **1997**, 101, 8617.
- (6) Flunt, R.; Makitra, R.; Makogon, O.; Pyrih, J. *Org. React.* **1997**, 31, 21.
- (7) Zupan, J.; Gasteiger, J. *Neural Network for Chemists: an Introduction*; VCH: Weinheim, 1993.
- (8) Sumpter, B. G.; Getino, C.; Noid, D. W. *Annu. Rev. Phys. Chem.* **1994**, 45, 439.
- (9) Hildebrand, J. H.; Scott, R. L. *Regular Solutions*; Prentice-Hall: Eglewood Cliffs, NJ, 1962.
- (10) Coolidge, M. B.; Stewart, J. P. J. *MOPAC 6.0*; 1990.
- (11) Linke, W. F. *Solubilities Inorganic and Metal-Organic Compounds*, 4th ed.; Van Nostrand: Princeton, NJ, 1958.
- (12) Burrell, H. *Polymer Handbook*, 2nd ed.; Brandrup, J., Immergut, E. H., Eds.; John Wiley & Sons: New York, 1975; p IV-337.
- (13) Peterson, C.; Rögnvaldsson, R.; Lönnblad, L. *Compt. Phys. Commun.* **1994**, 81, 85.
- (14) Letcher, T. M.; Crosby, P. B.; Domanska, U.; Fowler, P. W.; Legon, A. C. *S. Afr. J. Chem.* **1993**, 46, 41.