

Predicting infinite dilution activity coefficients of organic compounds in water by quantum-connectivity descriptors

Ernesto Estrada · Gerardo A. Díaz ·
Eduardo J. Delgado

Received: 24 January 2006 / Accepted: 17 September 2006 / Published online: 19 October 2006
© Springer Science+Business Media B.V. 2006

Abstract Quantitative structure-property relationship (QSPR) models are developed to predict the logarithm of infinite dilution activity coefficient of hydrocarbons, oxygen containing organic compounds and halogenated hydrocarbons in water at 298.15 K. The description of the molecular structure in terms of quantum-connectivity descriptors allows to obtain more simple QSPR models because of the quantum-chemical and topological information coded in this type of descriptors. The models developed in this paper have fewer descriptors and better statistics than other models reported in literature. The current models allow a more transparent physical interpretation of the phenomenon in terms of intermolecular interactions which occur in solution and which explain the respective deviations from ideality.

Keywords Infinite dilution activity coefficients · Organic compounds · Aqueous solution · QSPR · Quantum-connectivity descriptors

Introduction

Infinite dilution activity coefficients (γ^∞) characterize the behavior of the solute molecule when it is com-

pletely surrounded by the solvent. It is extremely valuable for both theoretical and practical purposes. For theoreticians, limiting activity coefficients provide incisive information regarding solute–solvent interactions in the absence of solute–solute interactions, in consequence is a measurement of the deviation from the ideal solution. From a practical point of view, γ^∞ data allow for the prediction of retention and selectivity for chromatographic systems, provide essential data in the design of separation equipment, and in environmental sciences infinite dilution activity coefficients allow the study of the fate and distribution of organic chemicals in the environment [1–4].

The experimental determination of γ^∞ of nonelectrolytes in aqueous solutions is often difficult due to limitations of existing experimental techniques, mainly for solutes that are sparingly soluble in water, and no single method can be used for the many potential solutes.

Therefore, for reasons of cost, safety, and availability of techniques to measure γ^∞ , it is desirable to be able to predict it than measure it. Thus, many predictive models have been reported among which the MOSCED [5], ASOG [6] and UNIFAC [7] are known to be very powerful in terms of their accuracy of prediction. However, due to the strong non-ideality existing in aqueous solutions, these traditional predictive methods tend to be only moderately successful in predicting γ^∞ of organic solutes in water. Clearly, in order to parameterize and extend these models to aqueous solutions, a more extensive and accurate data are required along with a correct calculation of the entropic contribution resulting from water's rich hydrogen-bonding network.

E. Estrada
Complex Systems Research Group, X-rays Unit, RIAIDT,
Edificio CACTUS, University of Santiago of Compostela,
15782 Santiago de Compostela, Spain

G. A. Díaz · E. J. Delgado (✉)
Theoretical and Computational Chemistry Group (QTC),
Faculty of Chemical Sciences, Universidad de Concepción,
Casilla 160-C, Concepción, Chile
e-mail: edelgado@udec.cl

The development of a quantitative structure-property relationship (QSPR) model provides an alternative method of estimating the aqueous infinite dilution activity coefficient directly from molecular structure without making an experimental measurement. Recently, several articles on correlations between structural features and γ^∞ have been reported [8–11]. In these studies different types of molecular descriptors are used to develop multiple linear regression and computational neural network models to predict γ^∞ . The molecular descriptors involved encode topological, geometric, electronic and quantum chemical molecular features. Even though these models do very well predicting γ^∞ , the number of descriptors involved in the models is rather high limiting in this way their application. It would be highly desirable to have models with fewer descriptors which allow their application in a more straightforward way, and, on the other hand, to preserve the principle of parsimony [12].

Quantum-connectivity indices [13, 14] are defined by using a combination of topological invariants, such as interatomic connectivity, and quantum chemical information, such as atomic charges and bond orders. These indices also contain important three-dimensional information incorporated by the quantum-chemical parameters used in their definition. Therefore, these indices are more rich in chemical information than traditional molecular descriptors since encode, at the same time, both topological and quantum chemical features of molecules.

In consequence, we postulate the following hypothesis: it is possible to find good QSPR models, based on quantum-connectivity indices, for the prediction of infinite dilution activity coefficients of organic compounds in water. The models so developed should contain fewer parameter than other reported QSPR models, using traditional molecular descriptors, because of the information encoded by quantum-connectivity indices.

In this article we report QSPR models for the prediction of infinite dilution activity coefficients, γ^∞ , of organic compounds in water. The data sets contain 103 hydrocarbons, 108 oxygen containing compounds and 39 halogenated hydrocarbons. The model developed for hydrocarbons involves only two descriptors, the model developed for oxygen containing compounds includes only one descriptor, and the model for halogenated hydrocarbons involves two descriptors. The three models predict infinite dilution activity coefficients with fewer descriptors and better performance than other QSPR models reported in literature.

Chemical data

The data sets of infinite dilution activity coefficients for hydrocarbons and oxygen containing compounds were taken from the data reported by He and Zhong [10]. The set of hydrocarbons contains 103 compounds covering a $\ln(\gamma^\infty)$ range from about 6 to 19 ln units. The set of oxygen containing compounds contains 108 structures covering a $\ln(\gamma^\infty)$ range from about 1 to 23 ln units. The set of halogenated hydrocarbons contains 39 structures covering a $\ln(\gamma^\infty)$ range from about 6 to 12 ln units.

Computational methods

Full-geometry optimization was carried out for all compounds in the data sets using the AM1 semiempirical quantum chemical method as implemented in the Mopac version 6.0. The keyword Bonds was always used to generate the bond order and valence matrices. The Mopac output files were then used for calculating quantum-connectivity indices using the software Modelab for Windows version 1.0 (<http://www.modelab.com>). The multiple linear regressions models and statistics were carried out using the softwares Sigmatat 3.0. (<http://www.sigmatat.com>) and Codessa.

Results and discussion

Hydrocarbons

A total of 68 quantum-connectivity were calculated for all 103 compounds. The best regression equation found involve only two descriptors as follows:

$$\ln\gamma^\infty = 3.97 + 15.15 \ ^1\Omega_P - 14.87 \ ^1\Omega_P^C(q)$$
$$R^2 = 0.93, \quad s^2 = 0.53, \quad F = 622, \quad (R_{cv})^2 = 0.92$$

where $^1\Omega_P$ is the path bond-order-based connectivity index of order 1, and $^1\Omega_P^C(q)$ is the path hydrogen-corrected charge-based connectivity index of order 2. The statistical parameters are the following: R is the correlation coefficient, s^2 is the variance, F is the Fisher test value, and R_{cv} is the one-leave-out crossvalidated correlation coefficient. This last parameter provides an estimation of the stability of the obtained regression model by focusing on the sensitivity of the model to the elimination of any single data point. The obtained value for R_{cv} in this study, very close to the correlation coefficient, assures the stability of the model.

Unfortunately, both descriptors involved in the model are highly correlated with a correlation coefficient of 0.99. It is well-known that the collinearity of descriptor may often lead to statistically meaningless regression equations without any predictive power [15]. Therefore, we proceeded to orthogonalize both descriptors according to the procedure reported by Randić [16]. Briefly, we started selecting the first orthogonal descriptor, i.e. the more relevant descriptor in the regression. One parameter regressions show that the ${}^1\Omega_P$ descriptor has a square correlation coefficient of 0.76, against 0.72 for ${}^1\Omega_P^C(q)$. In consequence, we choose ${}^1\Omega_P$ as the first orthogonal descriptor. Next, we make a regression of ${}^1\Omega_P^C(q)$ against ${}^1\Omega_P$, and then the calculated descriptor value is subtracted from the original observed value. The residual, the difference between calculated and observed descriptor value, becomes the new orthogonal descriptor which will be labeled D_2 . As result, the descriptor D_2 is orthogonal to ${}^1\Omega_P$.

Then, we correlate the property, $\ln(\gamma^\infty)$, with these two orthogonal descriptors. The resulting regression equation is the following:

$$\ln \gamma^\infty = 6.49 + 1.67 {}^1\Omega_P - 14.86 D_2$$

The statistics is the same of the above regression using non-orthogonal descriptors, as expected, since the descriptor D_2 is obtained as a combination of the former and cannot have more information content than the former. The calculated values, by this equation, and the experimental values of the logarithm of infinite dilution activity coefficient are compared in Table 1, and the respective scatter plot is shown in Fig. 1.

Oxygen containing compounds

For the 108 oxygen containing compounds the best correlation equation involves only one descriptors as follows:

$$\ln \gamma^\infty = -1.97 + 1.83 {}^1\varepsilon_P(\rho)$$

$$R^2 = 0.98; s^2 = 0.29; F = 5059; (R_{cv}) = 0.98$$

where ${}^1\varepsilon_P(\rho)$ is the path bond-order-based bond connectivity index of order 1, and the statistical parameters have the same meaning given above. The value of the cross-validated correlation coefficient assures the stability of the model. The calculated values, by this equation, and the experimental values of the logarithm of infinite dilution activity coefficient are compared in Table 2, and the respective scatter plot is shown in Fig. 2.

Halogenated hydrocarbons

For the 39 halogenated hydrocarbons the best correlation equation contains only two descriptor, one quantum-connectivity index and one electrostatic descriptor (charged partial surface area (CPSA) descriptor). The equation obtained is:

$$\ln \gamma^\infty = 5.50 + 37.67 \text{FNSA}^{(3)} + 2.34 {}^2\varepsilon_P(\rho)$$

$$R^2 = 0.92; s^2 = 0.20; F = 205.27; (R_{cv}) = 0.91$$

where $\text{FNSA}^{(3)}$ is the fractional negative surface area descriptor (negative partial surface area/total molecular surface area) and ${}^2\varepsilon_P(\rho)$ is the path bond-order-based bond connectivity index of order 2. The inter-correlation between these two descriptors is low (0.18) indicating the low collinearity of them, i.e., they encode different type of information. The statistical parameters have the same meaning given above. The value of the cross-validated correlation coefficient assures the stability of the model. The calculated values, by this equation, and the experimental values of the logarithm of infinite dilution activity coefficient are compared in Table 3, and the respective scatter plot is shown in Fig. 3.

Molecular interpretation

It has been well established that intermolecular forces are responsible for the deviation from ideality in liquid solutions. Consequently activity coefficients have been related to the extent of the forces acting between solute and solvent molecules. The main interactions existing between uncharged molecules are classified as van der Waals and hydrogen-bond interactions.

Van der Waals interactions between neighboring molecules are always attractive and non-specific. This is true no matter how different in polarity the interacting molecules are. Van der Waals interactions comprise the following components: dipole–dipole interactions, dipole–induced dipole interactions, and induced dipole–induced dipole interactions (London dispersion interactions). In general, the London type interactions are the dominant contribution.

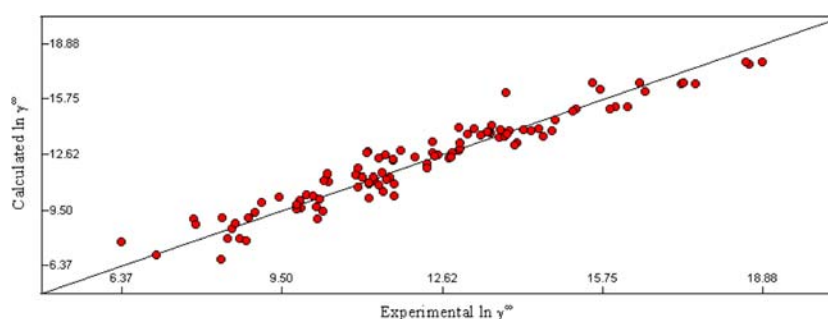
Hydrogen-bond interactions may not always be present. They only occur between interaction partners with complementary properties, i.e.: between a H-donor and a H-acceptor species. Thus, these interactions are specific and like van der Waals interactions are always attractive.

Table 1 Calculated and experimental values of $\ln(\gamma^\infty)$ for hydrocarbons

	Name	CAS number	${}^1\Omega_P$	D_2	$\ln(\gamma^\infty)$ (calc.)	$\ln(\gamma^\infty)$ (exp.)	Diff.
1	1,2,3-Trimethylbenzene	526-73-8	3.4002	0.0148	12.38	11.39	0.99
2	1,2,4-Trimethylbenzene	95-36-3	3.3928	0.0106	12.30	11.67	0.63
3	1,2-Benzanthracene	56-55-3	6.6314	-0.063	16.60	17.56	-0.96
4	1,2-Benzofluorene	238-84-6	6.4012	-0.0582	16.29	15.72	0.57
5	1,3,5-Cycloheptatriene	544-25-2	2.5778	-0.0987	9.32	8.98	0.34
6	1,3,5-Trimethylbenzene	108-67-8	3.3847	0.0003	12.24	11.67	0.57
7	1,3-Butadiene	106-99-0	1.1786	-0.118	6.70	8.32	-1.62
8	1,3-Dimethylnaphthalene	575-41-7	4.4724	-0.0126	13.75	13.9	-0.15
9	1,4-Cyclohexadiene	628-41-1	2.3429	-0.1112	8.74	8.6	0.14
10	1,4-Dimethylnaphthalene	571-58-4	4.4817	0.0078	13.84	13.54	0.3
11	1,4-Pentadiene	591-93-5	1.6548	-0.1014	7.74	8.82	-1.08
12	1,5-Dimethylnaphthalene	571-61-9	4.4804	0.0081	13.83	13.52	0.31
13	1,5-Hexadiene	592-42-7	2.1608	-0.0736	9.00	10.20	-1.2
14	1,6-Heptadiene	3070-53-9	2.667	-0.046	10.25	11.7	-1.45
15	1-Butene	106-98-9	1.536	-0.0428	8.42	8.53	-0.11
16	1-Butyne	107-00-6	1.3713	-0.072	7.71	6.37	1.34
17	1-Ethyl-naphthalene	1127-76-0	4.6137	0.0005	14.23	13.6	0.63
18	1-Heptyne	628-71-7	2.8886	0.011	11.47	10.95	0.52
19	1-Hexene	592-41-6	2.5483	0.0118	10.91	11.39	-0.48
20	1-Hexyne	693-02-7	2.3824	-0.0163	10.22	9.45	0.77
21	1-Methylcyclohexene	591-49-1	3.1058	-0.0292	11.23	11.54	-0.31
22	1-Methylnaphthalene	90-12-0	4.0502	-0.0447	12.57	12.55	0.02
23	1-Nonyne	3452-09-3	3.9011	0.0652	13.96	13.77	0.19
24	1-Octyne	629-05-0	3.3948	0.038	12.71	12.45	0.26
25	1-Pentene	109-67-1	2.0422	-0.0153	9.67	10.18	-0.51
26	1-Pentyne	627-19-0	1.8768	-0.0423	8.99	7.79	1.2
27	2,2,4-Trimethylpentane	540-84-1	3.4858	0.1085	13.91	14.77	-0.86
28	2,2-Dimethylbutane	75-83-2	2.608	0.0675	11.84	12.33	-0.49
29	2,2-Dimethylpentane	590-35-2	3.1135	0.0949	13.09	14.05	-0.96
30	2,2-Dimethylpropane	463-82-1	2.296	0.0553	11.14	11.32	-0.18
31	2,3,4-Trimethylpentane	565-75-3	3.617	0.1358	14.54	14.83	-0.29
32	2,3-Benzofluorene	243-17-4	6.3894	-0.0666	16.14	16.59	-0.45
33	2,3-Dimethylbutane	79-29-8	2.6847	0.0751	12.08	12.34	-0.26
34	2,3-Dimethylnaphthalene	581-40-8	4.4712	-0.0134	13.74	13.13	0.61
35	2,3-Dimethylpentane	565-59-3	4.2399	0.1724	16.12	13.87	2.25
36	2,4-Dimethylpentane	108-08-7	3.174	0.0973	13.23	14.08	-0.85
37	2,6-Dimethylnaphthalene	581-42-0	4.4635	-0.017	13.67	13.39	0.28
38	2-Methyl propene	115-11-7	1.3769	-0.0628	7.85	8.44	-0.59
39	2-Methyl-1,3-butadiene	78-79-5	1.5909	-0.0879	7.84	8.68	-0.84
40	2-Methyl-1-pentene	763-29-1	2.4471	0.0133	10.77	11.00	-0.23
41	2-Methyl-2-butene	513-35-9	1.8998	0.0091	9.52	9.79	-0.27
42	2-Methylantracene	613-12-7	5.546	-0.0394	15.14	15.24	-0.1
43	2-Methylbutane	78-78-4	2.0424	0.0172	10.15	11.21	-1.06
44	2-Methylhexane	591-76-4	3.3082	0.1099	13.64	14.60	-0.96
45	2-Methylnaphthalene	91-57-6	4.0412	-0.0492	12.49	12.49	0
46	2-Methylpentane	107-83-5	2.8019	0.0826	12.39	12.76	-0.37
47	2-Methylpropane	75-28-5	1.7527	0.0128	9.6	9.88	-0.28
48	2-Pentene	109-68-2	2.0474	0.0005	10	9.86	0.14
49	3,3-Dimethylpentane	562-49-2	3.1733	0.1177	13.53	13.75	-0.22
50	3,4-Benzopyrene	50-32-8	7.4402	-0.0786	17.72	18.61	-0.89
51	3-Methyl-1-butene	563-45-1	1.9312	-0.0215	9.39	10.3	-0.91
52	3-Methylheptane	589-81-1	3.8512	0.1524	15.17	15.9	-0.73
53	3-Methylhexane	589-34-4	3.3451	0.1249	13.92	13.93	-0.01
54	3-Methylpentane	96-14-0	2.8393	0.0983	12.68	12.82	-0.14
55	4-Ethenylcyclohexene	100-40-3	3.274	-0.0685	10.93	11.7	-0.77
56	4-Methyl-1-pentene	691-37-2	2.4159	0.001	10.5	11.49	-0.99
57	9,10-Dimethylantracene	781-43-1	6.0021	0.0113	16.66	15.57	1.09
58	9-Methylantracene	779-02-2	5.5631	-0.0302	15.31	16.25	-0.94
59	Acenaphthene	83-32-9	4.6905	-0.0727	13.22	12.98	0.24
60	Anthracene	120-12-7	5.1233	-0.0726	13.95	14.36	-0.41
61	Benzene	71-43-2	2.1175	-0.0895	8.69	7.82	0.87

Table 1 continued

	Name	CAS number	${}^1\Omega_P$	D_2	$\ln(\gamma^\infty)$ (calc.)	$\ln(\gamma^\infty)$ (exp.)	Diff.
62	Biphenyl	92-52-4	4.3248	-0.053	12.91	12.98	-0.07
63	Chrysene	218-01-9	6.6371	-0.0614	16.63	17.34	-0.71
64	<i>cis</i> -1,2-Dimethylcyclohexane	2207-01-4	3.8682	0.047	13.63	13.85	-0.22
65	Cycloheptane	291-64-5	3.5467	0.0008	12.46	12.11	0.35
66	Cycloheptene	628-92-2	3.2073	-0.0333	11.34	11.3	0.04
67	Cyclohexane	110-82-7	3.0368	-0.0134	11.35	11.29	0.06
68	Cyclohexene	110-83-8	2.6909	-0.0601	10.08	10.25	-0.17
69	Cyclooctane	292-64-8	4.0566	0.0363	13.79	13.58	0.21
70	Cyclopentane	287-92-3	2.5349	-0.0301	10.27	10.12	0.15
71	Cyclopentene	292-64-8	2.1878	-0.0723	9.06	8.86	0.2
72	Ethylbenzene	100-41-4	3.102	0.0091	11.52	10.4	1.12
73	Indan	496-11-7	3.6897	-0.0542	11.83	11	0.83
74	Isopropylbenzene	98-82-8	3.5055	0.014	12.54	11.53	1.01
75	Methylcyclohexane	108-87-2	3.4444	0.0133	12.43	12.8	-0.37
76	Methylcyclopentane	96-37-7	2.9416	0.003	11.35	11.6	-0.25
77	<i>m</i> -Methylstyrene	100-80-1	3.1728	-0.0535	10.98	11.21	-0.23
78	<i>m</i> -Xylene	108-38-3	2.9626	-0.0245	11.06	10.41	0.65
79	Naphthacene	92-24-0	6.6332	-0.0583	16.67	16.48	0.19
80	Naphthalene	91-20-3	3.6192	-0.0819	11.3	11.07	0.23
81	<i>n</i> -Butane	106-97-8	1.9234	0.0412	10.31	9.99	0.32
82	<i>n</i> -Butylbenzene	104-51-8	4.1143	0.0459	14.03	13.25	0.78
83	<i>n</i> -Decane	124-18-5	4.9609	0.2048	17.8	18.88	-1.08
84	<i>n</i> -Heptane	142-82-5	3.4423	0.1231	14.06	14.51	-0.45
85	<i>n</i> -Hexane	110-54-3	2.9358	0.0959	12.81	12.96	-0.15
86	<i>n</i> -Nonane	111-84-2	4.4549	0.1777	16.55	17.29	-0.74
87	<i>n</i> -Octane	111-65-9	3.9484	0.1503	15.3	16.02	-0.72
88	<i>n</i> -Pentane	109-66-0	2.4297	0.0685	11.56	11.46	0.1
89	<i>n</i> -propylbenzene	103-65-1	3.6088	0.0194	12.79	11.82	0.97
90	<i>o</i> -Ethyltoluene	611-14-3	3.5329	0.025	12.75	11.18	1.57
91	<i>o</i> -Xylene	95-47-6	2.9703	-0.0208	11.13	10.32	0.81
92	Perylene	198-55-0	7.451	-0.0718	17.83	18.56	-0.73
93	<i>p</i> -Ethyltoluene	622-96-8	3.5248	0.023	12.71	11.16	1.55
94	Phenanthrene	85-01-8	5.1282	-0.0715	13.97	14.22	-0.25
95	<i>p</i> -Methylstyrene	622-97-9	3.1734	-0.0529	10.99	11.21	-0.22
96	Propane	74-98-6	1.4164	0.0131	9.05	8.35	0.7
97	Propene	115-07-1	0.9937	-0.0832	6.91	7.05	-0.14
98	<i>p</i> -Xylene	106-42-3	2.963	-0.0242	11.07	10.41	0.66
99	Pyrene	129-00-0	5.9316	-0.0882	15.06	15.17	-0.11
100	<i>sec</i> -Butylbenzene	135-98-8	4.0482	0.0577	14.09	12.96	1.13
101	Styrene	100-42-5	2.7505	-0.086	9.8	9.8	0
102	<i>tert</i> -Butylbenzene	98-06-6	3.8304	0.0292	13.31	12.44	0.87
103	Toluene	108-88-3	2.5402	-0.0566	9.88	9.12	0.76

Fig. 1 Scatter plot of the calculated vs. experimental $\ln(\gamma^\infty)$ for hydrocarbons

In hydrocarbons, wherein hydrogen bonding is not an issue, the intermolecular interactions are reduced to dispersion and electrostatic interactions. This expected

behavior is confirmed by the information encoded by the descriptors involved in the model. These descriptors are the path bond-order-based connectivity index

Table 2 Calculated and experimental values of $\ln(\gamma^\infty)$ for oxygen containing compounds

	Name	CAS number	$^1\varepsilon_P(\rho)$	$\ln(\gamma^\infty)$ (calc.)	$\ln(\gamma^\infty)$ (exp.)	Diff.
1	2,2,3-Trimethyl-3-pentanol	7294-05-5	4.8924	6.98	6.95	0.03
2	2,2-Dimethyl-1-butanol	1185-33-7	4.2248	5.75	6.61	- 0.86
3	2,2-Dimethyl-1-propanol	75-84-3	3.2835	4.03	4.91	- 0.88
4	2,2-Dimethyl-3-pentanol	3970-62-5	4.5409	6.33	6.66	- 0.33
5	2,3-Dimethyl-2-butanol	594-60-5	3.6557	4.71	4.88	- 0.17
6	2,3-Dimethyl-2-pentanol	4911-70-0	4.5354	6.32	6.02	0.30
7	2,3-Dimethyl-3-pentanol	595-41-5	4.6033	6.45	5.96	0.49
8	2,4-Dimethyl-2-pentanol	625-06-9	4.2804	5.86	6.16	- 0.30
9	2,4-Dimethyl-3-pentanol	600-36-2	4.6089	6.46	6.82	- 0.36
10	2,4-Dimethyl-3-pentanone	565-80-0	4.5546	6.36	7.01	- 0.65
11	2,6-Dimethyl-4-heptanone	108-83-8	5.7812	8.60	9.08	- 0.48
12	2-Butanol	78-92-2	3.0314	3.57	3.26	0.31
13	2-Butanone	78-93-3	2.8669	3.27	3.24	0.03
14	2-Heptanone	110-43-0	5.2272	7.59	7.24	0.35
15	2-Hexanol	626-93-7	4.4466	6.16	5.64	0.52
16	2-Hexanone	591-78-6	4.2701	5.84	5.87	- 0.03
17	2-Methyl <i>sec</i> -butyl methyl ether	6795-87-5	4.2691	5.84	6.11	- 0.27
18	2-Methyl-1-butanol	137-32-6	3.9101	5.18	5.08	0.10
19	2-Methyl-1-propanol	78-83-1	3.0275	3.56	3.89	- 0.33
20	2-Methyl-2-hexanol	625-23-0	4.7018	6.63	6.49	0.14
21	2-Methyl-2-pentanol	590-36-3	3.9894	5.32	5.14	0.18
22	2-Methyl-3-pentanol	565-67-3	4.2649	5.83	5.63	0.20
23	2-Methyl-3-pentanone	565-69-5	4.1884	5.69	5.89	- 0.20
24	2-Nonanone	821-55-6	6.4042	9.74	9.70	0.04
25	2-Pentanol	6032-29-7	3.7345	4.86	4.57	0.29
26	2-Pentanone	107-87-9	3.5584	4.54	4.54	0.00
27	2-Propanol	67-63-0	2.143	1.95	2.03	- 0.08
28	3,3-Dimethyl-2-butanol	464-07-3	3.6531	4.71	5.43	- 0.72
29	3,3-Dimethyl-2-butanone	75-97-8	3.5214	4.47	5.66	- 1.19
30	3-Ethyl-3-pentanol	597-49-9	5.188	7.52	5.94	1.58
31	3-Hexanol	623-37-0	4.6251	6.49	5.85	0.64
32	3-Hexanone	589-38-8	4.515	6.29	6.02	0.27
33	3-Methyl-1-butanol	123-51-3	3.7322	4.85	5.34	- 0.49
34	3-Methyl-2-Butanone	563-80-4	3.2319	3.94	4.43	- 0.49
35	3-Methyl-2-pentanol	565-60-6	4.2562	5.81	5.66	0.15
36	3-Methyl-2-pentanone	565-61-7	4.108	5.54	5.56	- 0.02
37	3-Methyl-3-hexanol	597-96-6	4.9384	7.06	6.28	0.78
38	3-Methyl-3-pentanol	77-74-7	4.2393	5.78	4.85	0.93
39	3-Pentanone	96-22-0	3.8246	5.02	4.67	0.35
40	4-Heptanone	123-19-3	5.2053	7.55	7.41	0.14
41	4-Methyl-2-pentanol	108-11-2	4.0262	5.39	5.86	- 0.47
42	4-Methyl-2-pentanone	108-10-1	3.8459	5.06	5.68	- 0.62
43	5-Methyl-2-Hexanone	110-12-3	4.5637	6.37	7.33	- 0.96
44	5-Nonanone	502-56-7	6.6299	10.15	9.98	0.17
45	Acetaldehyde	75-07-0	1.7421	1.21	1.37	- 0.16
46	Acetone	67-64-1	1.9102	1.52	1.95	- 0.43
47	Butanol	71-36-3	3.4398	4.32	3.92	0.40
48	Butyl methyl ether	628-28-4	4.1599	5.64	6.30	- 0.66
49	Butyl pentanoate	591-68-4	7.3379	11.45	9.87	1.58
50	Butyraldehyde	123-72-8	3.3496	4.15	3.88	0.27
51	Decanol	112-30-1	7.7084	12.13	12.38	- 0.25
52	Diethyl ether	60-29-7	3.4531	4.34	4.23	0.11
53	Diisopropyl ether	108-20-3	4.0432	5.42	6.44	- 1.02
54	di- <i>n</i> -Butyl ether	142-96-1	6.2975	9.55	10.76	- 1.21
55	di- <i>n</i> -Propyl ether	111-43-3	4.8793	6.95	7.74	- 0.79
56	Dodecanol	112-53-8	9.1314	14.73	15.31	- 0.58
57	Ethanol	64-17-5	2.0074	1.70	1.32	0.38
58	Ethyl acetate	141-78-6	3.5687	4.55	4.18	0.37
59	Ethyl butyrate	105-54-4	5.2187	7.57	6.59	0.98
60	Ethyl formate	109-94-4	3.3199	4.10	3.86	0.24
61	Ethyl isopropyl ether	625-54-7	3.7488	4.88	5.30	- 0.42

Table 2 continued

	Name	CAS number	$^1\varepsilon_P(\rho)$	$\ln(\gamma^\infty)$ (calc.)	$\ln(\gamma^\infty)$ (exp.)	Diff.
62	Ethyl pentanoate	539-82-2	5.9297	8.87	7.96	0.91
63	Ethyl propanoate	105-37-3	4.5272	6.31	5.54	0.77
64	Ethyl propenoate	140-88-5	4.4528	6.17	5.60	0.57
65	Ethyl propyl ether	628-32-0	4.1662	5.65	5.55	0.10
66	Ethylene oxide	75-21-8	2.1438	1.95	1.83	0.12
67	Heptadecanol	1454-85-9	12.689	21.24	21.30	-0.06
68	Heptanal	111-71-7	5.4804	8.05	8.34	-0.29
69	Heptanol	111-70-6	5.5739	8.22	8.09	0.13
70	Hexadecanol	36653-82-4	11.9774	19.94	19.77	0.17
71	Hexanal	66-25-1	4.769	6.75	6.70	0.05
72	Hexanoic acid	142-62-1	4.9775	7.13	6.40	0.73
73	Hexanol	111-27-3	4.8625	6.92	6.68	0.24
74	Isobutyl acetate	110-19-0	4.5538	6.36	6.74	-0.38
75	Isobutyl methyl ether	625-44-5	3.7414	4.87	6.09	-1.22
76	Isopentyl acetate	123-92-2	5.2673	7.66	8.00	-0.34
77	Isopropyl acetate	108-21-4	3.8595	5.09	5.28	-0.19
78	Isopropyl butyrate	638-11-9	5.5096	8.11	8.03	0.08
79	Isopropyl propyl ether	627-08-7	4.4626	6.19	7.09	-0.90
80	Methyl acetate	79-20-9	2.8183	3.18	3.12	0.06
81	Methyl butyrate	623-42-7	4.469	6.20	5.80	0.40
82	Methyl formate	107-31-3	2.5835	2.75	2.74	0.01
83	Methyl hexanoate	106-70-7	5.891	8.80	8.29	0.51
84	Methyl isobutyrate	547-63-7	4.1397	5.60	5.73	-0.13
85	Methyl pentanoate	624-24-8	5.1801	7.50	7.14	0.36
86	Methyl propanoate	554-12-1	3.777	4.94	4.47	0.47
87	Methyl propyl ether	557-17-5	3.4506	4.34	4.88	-0.54
88	<i>n</i> -Butyl acetate	123-86-4	4.9767	7.13	6.70	0.43
89	<i>n</i> -Hexyl acetate	142-92-7	6.3986	9.73	9.43	0.30
90	Nonanal	124-19-6	6.9038	10.66	11.23	-0.57
91	Nonanol	3452-97-9	6.9968	10.83	11.03	-0.20
92	<i>n</i> -Pentyl acetate	628-63-7	5.6871	8.43	8.08	0.35
93	<i>n</i> -Propyl acetate	109-60-4	4.2669	5.83	5.49	0.34
94	<i>n</i> -Propyl formate	110-74-7	4.0237	5.39	5.13	0.26
95	<i>n</i> -Propyl propanoate	106-36-5	5.2254	7.59	6.99	0.60
96	Octadecanol	112-92-5	13.4004	22.54	23.34	-0.80
97	Octanal	124-13-0	6.1917	9.35	9.02	0.33
98	Octanol	111-87-5	6.2854	9.52	9.56	-0.04
99	Pentadecanol	629-76-5	11.2659	18.64	18.77	-0.13
100	Pentanal	110-62-3	4.0574	5.45	5.39	0.06
101	Pentanoic acid	109-52-4	4.2663	5.83	4.84	0.99
102	Pentanol	71-41-0	4.1509	5.62	5.29	0.33
103	Propanol	71-23-8	2.7311	3.02	2.60	0.42
104	Propionaldehyde	123-38-6	2.6507	2.88	2.56	0.32
105	<i>sec</i> -Butyl methyl ether	6795-87-5	3.9253	5.21	5.71	-0.50
106	<i>tert</i> -Butanol	75-65-0	2.3446	2.32	2.48	-0.16
107	<i>tert</i> -Butyl methyl ether	1634-04-4	3.2968	4.06	4.73	-0.67
108	Tetradecanol	112-72-1	10.5545	17.33	17.50	-0.17

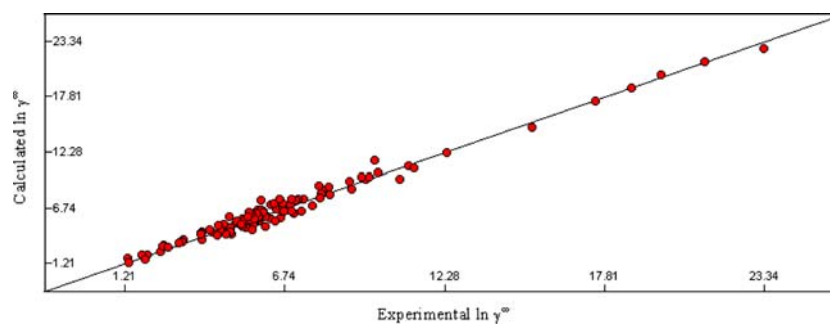
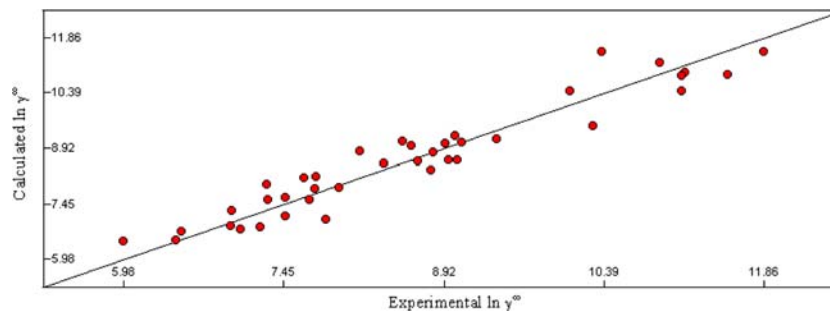
Fig. 2 Scatter plot of the calculated vs. experimental $\ln(\gamma^\infty)$ for oxygen containing organic compounds

Table 3 Calculated and experimental values of $\ln(\gamma^\infty)$ for halogenated hydrocarbons

	Name	CAS number	${}^2_{\text{EP}}(\rho)$	FNSA ⁽³⁾	$\ln(\gamma^\infty)$ (calc.)	$\ln(\gamma^\infty)$ (exp.)	Diff.
1	1,1,1,2-Tetrachloroethane	630-20-6	2.4502	-0.0580	9.06	9.09	-0.03
2	1,1,1-Trichloroethane	71-55-6	2.0684	-0.0471	8.57	8.68	-0.11
3	1,1,2,2-Tetrabromoethane	79-27-6	2.4651	-0.0469	9.51	10.29	-0.78
4	1,1,2,2-Tetrachloroethane	79-34-5	2.4862	-0.0660	8.84	8.15	0.69
5	1,1,2-Trichloroethane	79-00-5	1.9445	-0.0664	7.55	7.31	0.24
6	1,1-Dichloroethane	75-34-3	1.5378	-0.0490	7.26	6.98	0.28
7	1,2,3-Trichloropropane	96-18-4	2.3650	-0.0669	8.52	8.37	0.15
8	1,2-Dibromoethane	106-93-4	1.4324	-0.0488	7.02	7.84	-0.82
9	1,2-Dibromopropane	78-75-1	1.9287	-0.0379	8.59	8.97	-0.38
10	1,2-Dichlorobenzene	95-50-1	2.9663	-0.0402	10.93	11.13	-0.20
11	1,2-Dichloroethane	107-06-2	1.4448	-0.0641	6.47	6.46	0.01
12	1,2-Dichloropropane	78-87-5	1.9412	-0.0499	8.17	7.75	0.42
13	1,3-Dibromopropane	109-64-8	1.9424	-0.0451	8.35	8.80	-0.45
14	1,3-Dichlorobenzene	541-73-1	2.9930	-0.0448	10.82	11.10	-0.28
15	1,3-Dichloropropane	142-28-9	1.9514	-0.0590	7.85	7.74	0.11
16	1,3-Dichloropropene	542-75-6	1.5183	-0.0593	6.82	7.24	-0.42
17	1,4-Dichlorobenzene	106-46-7	2.9873	-0.0431	10.87	11.53	-0.66
18	1-Bromo-2-chloroethane	107-04-0	1.4388	-0.0559	6.77	7.06	-0.29
19	1-Bromobutane	109-65-9	1.9387	-0.0236	9.15	9.41	-0.26
20	1-Bromopentane	110-53-2	2.4450	-0.0212	10.43	11.10	-0.67
21	1-Bromopropane	106-94-5	1.4307	-0.0262	7.86	7.96	-0.10
22	1-Chloro-2-methylpropane	513-36-0	1.9319	-0.0279	8.97	8.62	0.35
23	1-Chlorobutane	109-69-3	1.9403	-0.0266	9.04	8.94	0.10
24	1-Chlorohexane	544-10-5	2.9558	-0.0251	11.48	11.86	-0.38
25	1-Chloropropane	540-54-5	1.4365	-0.0331	7.62	7.47	0.15
26	1-Iodopropane	107-08-4	1.4284	-0.0068	8.59	9.05	-0.46
27	2-Bromobutane	78-76-2	1.9301	-0.0208	9.24	9.03	0.21
28	2-Bromopropane	75-26-3	1.5268	-0.0251	8.13	7.64	0.49
29	2-Chlorobutane	78-86-4	1.9389	-0.0249	9.11	8.54	0.57
30	2-Chloropropane	75-29-6	1.5352	-0.0300	7.97	7.30	0.67
31	2-Iodopropane	75-30-9	1.5231	-0.0067	8.82	8.82	0.00
32	3-Bromo-1-propene	106-95-6	1.1667	-0.0294	7.13	7.47	-0.34
33	3-Chloro-1-propene	107-05-1	1.1707	-0.0364	6.87	6.97	-0.10
34	Benzyl chloride	100-44-7	2.9842	-0.0273	11.46	10.37	1.09
35	Bromoethane	74-96-4	1.0096	-0.0307	6.71	6.52	0.19
36	Chloroethane	75-00-3	1.0163	-0.0379	6.45	5.98	0.47
37	Iodobenzene	591-50-4	2.5513	-0.0078	11.18	10.90	0.28
38	Iodoethane	75-03-6	1.0084	-0.0078	7.57	7.69	-0.12
39	Pentachloroethane	76-01-7	3.0181	-0.0565	10.44	10.08	0.36

Fig. 3 Scatter plot of the calculated vs. experimental $\ln(\gamma^\infty)$ for halogenated hydrocarbons

of order 1, ${}^1\Omega_P$, and the orthogonalized descriptor D_2 . The first one shows high correlation ($R^2 = 0.96$) with the molecular volume which, in turn, is highly correlated to molecular polarizability, a measure of the ease with which the molecular electron cloud can distort in an electric field [17, 18]. This deformation in the

molecular charge distribution generates an induced dipole moment which is proportional to the strength of the electric field, being the molecular polarizability the proportionality constant between them. Thus, polarizability of the molecule is closely related to the strength of the London forces and consequently ${}^1\Omega_P$ encodes

information related to the extent of dispersion interactions.

On the other hand, it is found that the remaining information contained in the orthogonalized descriptor D_2 is related to electrostatic interactions since this descriptor shows a good correlation ($R^2 = 0.80$) with the topographic electronic index, a complex function of the atomic charges and molecular geometry. This index, which is calculated as the sum of the charge differences over all molecular bonds, encodes electrostatic features of the molecules [19].

The calculation of molecular volume and the topographic electronic index, as well as their respective correlations to ${}^1\Omega_p$ and D_2 , respectively, were performed with the Codessa software [20] using the Ampac output files as input files.

Therefore, both descriptors ${}^1\Omega_p$ and D_2 encode information concerning dispersion and electrostatic interactions responsible for the deviation from ideality and consequently for the value of the infinite dilution activity coefficient.

The model for oxygen containing organic compounds contains only one descriptor, the path bond-order-based bond connectivity index of order 1, ${}^1\varepsilon_p(\rho)$. In order to identify the type of information encoded by this descriptor we have correlated it to the diverse molecular descriptors defined in the Codessa software. It is found that ${}^1\varepsilon_p(\rho)$ is highly correlated ($R^2 = 0.97$) to the molecular polarizability, and consequently this descriptor is involved in the dispersion interactions between solute and solvents molecules. This finding shows dispersion forces are the dominant interactions between water and oxygen containing organic compounds.

The model for halogenated hydrocarbons contains two descriptors, one quantum-connectivity index, ${}^2\varepsilon_p(\rho)$, and one charged partial surface area (CPSA) descriptor, $FNSA^{(3)}$. CPSA descriptors combine information about atomic charges and solvent-accessible surface areas thereby describing polar surface interactions of the molecules [21].

The more relevant descriptor in the model is the path bond-order-based bond connectivity index of order 2, ${}^2\varepsilon_p(\rho)$. From the above finding for oxygen containing compounds it is clear this family of descriptors, path bond-order based bond connectivity index, is a measurement of the molecular polarizability and in consequence these descriptors are encoding information related to dispersion interactions. For the case of halogenated hydrocarbons, the regression between ${}^2\varepsilon_p(\rho)$ and the molecular polarizability has an high regression coefficient ($R^2 = 0.98$), as expected. The regression analysis between these two molecular de-

scriptors was performed using the Codessa software as it was indicated for the other families of compounds.

On the other hand, the $FNSA^{(3)}$ descriptor (fractional partial negatively charged surface area/total molecular surface area) encodes information related to polar surface interactions between halogen atoms and the solvent. This descriptor adds information that is not present in the quantum-connectivity description of the molecular structure.

Thus, the two descriptors involved in the model, ${}^2\varepsilon_p(\rho)$ and $FNSA^{(3)}$, altogether account for polar and dispersion interactions responsible for the non-ideality shown in solutions of halogenated hydrocarbons.

Conclusions

In this article it is shown that quantum-connectivity indices are good molecular descriptors for the prediction of infinite dilution activity coefficients in water of hydrocarbons, oxygen containing organic compounds and halogenated hydrocarbons. The models developed in this paper have fewer descriptors and better statistics than other models reported in literature. This best performance of quantum-connectivity indices is due to they contain structural information in a more condensed way than in other series of molecular descriptors, such as constitutional, geometric, electrostatic, quantum-chemical, and topological ones; since quantum-connectivity indices are based on graph-theoretic invariants applied to quantum-chemical molecular parameters, such as atomic charge density and bond orders.

Acknowledgments The authors acknowledge the financial support received from FONDECYT, project No 7020464. EE thanks “Ramón y Cajal” program, Spain for partial financial support.

References

1. Sandler SI (1996) Fluid Phase Equilib 116:343
2. Sandler SI (1999) J Chem Thermodynamics 31:3
3. Kojima K, Zhang S, Hiaki T (1997) Fluid Phase Equilib 131:145
4. Eckert CA, Newman BA, Nicolaidis GL, Long TC (1981) AIChE J 27:33
5. Thomas ER, Eckert CA (1984) Ind Eng Chem Process Des Dev 23:194
6. Kojima K, Tochigi K (1979) Prediction of vapor liquid equilibria by the Asog method Physical Sciences Data 3. Elsevier Scientific Publishing Company, Amsterdam
7. Fredenslund Aa, Jones RL, Prausnitz JM (1975) AIChE J 21:1086
8. Mitchell BE, Jurs PC (1998) J Chem Inf Comput Sci 38:200
9. Delgado EJ, Alderete JB (2001) J Comput Chem 22:1851

10. He J, Zhong C (2003) *Fluid Phase Equilib* 205:303
11. Giralt F, Espinosa G, Arenas A, Ferre-Gine J, Amat L, Girones X, Carbo-Dorca R, Cohen Y (2004) *AIChE J* 50:1315
12. Hawkins DM (2004) *J Chem Inf Comput Sci* 44:1
13. Estrada E, Molina E (2001) *J Chem Inf Comput Sci* 41:791
14. Estrada E, Delgado EJ, Alderete JB, Jaña GA (2004) *J Comput Chem* 25:1787
15. Katritzky AR, Lobanov VS, Karelson M (1995) *Chem Soc Rev* 24:279
16. Randic M (1991) *J Chem Inf Comput Sci* 31:311
17. Berry RS, Rice SA, Ross J (1980) *Physical chemistry*. John Wiley & Sons, New York
18. McQuarrie DA, Simon JD (1997) *Physical chemistry. A molecular approach*. University Sciences Books, Sausalito
19. Liu HX, Hu RJ, Zhang RS, Yao XJ, Liu MC, Hu ZD, Fan BT (2005) *J Comput Aid Mol Des* 19:33
20. Katritzky AR, Lobanov VS, Karelson M (1994) *Codessa version 2.0 Reference Manual*
21. Stanton DT, Jurs PC (1990) *Anal Chem* 62:2323