

A molecular structure based model for predicting surface tension of organic compounds

E. J. DELGADO* and G. A. DIAZ

Faculty of Chemical Sciences, Theoretical and Computational
Chemistry Group (QTC), Universidad de Concepción,
Casilla 160-C, Concepción, Chile

(Received 13 December 2005; in final form 15 June 2006)

A Quantitative Structure-Property Relationship (QSPR) model for the prediction of surface tension of organic compounds was derived from a data set of 320 chemicals including N, O, F, Cl, Br, and/or S atoms and covering a range of about 14–45 dyn cm⁻¹. The model, only involving six molecular descriptors obtained solely from the chemical structures, yielded an r^2 of 0.96. Its predictive capability was estimated from an external test set containing 55 structures not considered in the training set ($r^2 = 0.94$). It was shown that the selected molecular descriptors presented a physical meaning corresponding to the different intermolecular interactions occurring in the bulk solution. The model is applicable to a wider variety of compounds, includes less parameters and correlates better than other QSPR models reported in literature.

Keywords: Surface tension; QSPR; Organic compounds; Molecular structure

1. Introduction

A liquid surface tends to contract to the minimum area as a result of unbalanced forces of molecular attraction at the surface. The molecules at the surface are attracted into the body of the liquid because the attraction of the underlying molecules is greater than the attraction by the vapor molecules on the other side of the surface. This inward attraction causes the surface contraction and gives rise to a force in the plane of the surface.

The surface tension of a liquid, γ , is defined as the force per unit length on the surface that opposes the expansion of the surface area. Surface tension is an important physical property about which accurate and detailed experimental information are of primary importance. Knowledge about surface tension, and the mechanisms underlying its origin, is critical in addressing such basic surface science issues as wetting, adhesion, friction, spreading and detergency [1]. Accordingly, experimental data for surface

*Corresponding author. Email: edelgado@udec.cl

tension are abundant at present, however, there are some justifications to develop models that can predict it. This is especially important in those cases when it is not practical to measure the surface tension due to the lack of the appropriate equipment or to the immediate need for a value for an engineering estimate.

The different approaches for the prediction of surface tension can be classified into the following categories: (1) correlations with other experimentally determined physicochemical properties such as density, viscosity; (2) correlations based on the corresponding states theorem; (3) models based on group contributions; (4) quantitative-structure-property relationships (QSPR).

The first approach requires a sufficient quantity of purified compound and therefore is not applicable for compounds not isolated or in development. In the second approach, due to the lack of critical properties for most substances and the difficulty of their accurate measurements, the estimation errors are considerably high and consequently its application is limited. The third approach provides good results for a large number of compounds. However, difficulties can arise in decomposing some structures into appropriate fragments whose constants are available or for compounds with fragments for which no group contributions have been fitted before. The fourth approach, based on molecular structure, is important not only from a fundamental physical point of view, since it allows a more transparent interpretation of the phenomenon on physical ground, but it is also technologically useful for the efficient production of materials with specific properties for a given application [1]. Thus, predictive models based on molecular structure are important for the design of novel chemicals since properties can be thus predicted prior to synthesis. In this way, the design of novel compounds may be guided by the calculation results.

In this article we report a QSPR model for the prediction of surface tension of organic compounds based on six molecular descriptors having definite physical meaning corresponding to the different intermolecular interactions occurring in the bulk solution.

2. Materials and methods

2.1 Data

The data set of surface tension, at 298 K and atmospheric pressure, of organic compounds was taken from literature [2]. A total of 320 structurally diverse organic compounds were selected to develop the model. The number and types of structures was considered enough to reach an adequate compromise between chemical variety and computation time. The compound data set, grouped according to the compound family, namely, alkanes, alkenes, aldehydes, ketones, ethers, carboxylic acids, esters, alcohols, amides, amines, nitriles, aromatics, nitro-, sulfured-, fluorinated-, chlorinated-, and brominated-hydrocarbons and covering a range of about 14–45 dyn cm⁻¹ is listed in table 1.

2.2 Computational methods

The QSPR model was developed using the Microsoft Windows version of the Codessa program [3]. This program performs the calculation of molecular descriptors and

Table 1. Experimental and calculated surface tension values (dyn cm⁻¹) for the training set.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd</i>	<i>Res.</i>
Alkanes				
Decane	124-18-5	23.4	22.5	-0.9
2,3-Dimethylbutane	79-29-8	16.9	19.5	2.6
2,4-Dimethylheptane	2213-23-2	20.9	21.3	0.4
2,5-Dimethylheptane	2216-30-0	20.9	22.2	1.3
2,6-Dimethylheptane	1072-05-5	20.6	21	0.4
2,2-Dimethylpentane	590-35-2	17.6	18.6	1.0
2,3-Dimethylpentane	565-59-3	19.5	21.3	1.8
3,3-Dimethylpentane	562-49-2	19.1	21.8	2.7
Dodecane	112-40-3	24.9	24.4	-0.5
3-Ethylpentane	617-78-7	20	21.1	1.1
Heptane	142-82-5	19.7	19.3	-0.4
Hexadecane	544-76-3	27.1	28.2	1.1
Hexane	110-54-3	17.9	18.2	0.3
2-Methylhexane	591-76-4	18.8	19.1	0.3
3-Methylhexane	589-34-4	19.3	20.3	1.0
2-Methyloctane	3221-61-2	21.4	21.2	-0.2
4-Methyloctane	2216-34-4	21.9	21.9	0.0
2-Methylpentane	107-83-5	16.9	17.8	0.9
3-Methylpentane	96-14-0	17.6	19.7	2.1
Nonane	111-84-2	22.4	22.5	0.1
Octane	111-65-9	21.2	20.4	-0.8
Pentadecane	629-62-9	26.7	27.2	0.5
Pentane	109-66-0	15.5	16.9	1.4
Tetradecane	629-59-4	26.2	26.3	0.1
Tridecane	629-50-5	25.6	25.4	-0.2
2,2,3-Trimethylpentane	564-02-3	20.2	22.6	2.4
2,2,4-Trimethylpentane	540-84-1	18.4	18.9	0.5
Undecane	1120-21-4	24	23.5	-0.5
Alkenes				
Cyclohexene	110-83-8	26.2	23.9	-2.3
Cyclopentene	142-29-0	22.2	23.7	1.5
1-Decene	872-05-9	23.6	23.2	-0.4
1-Heptene	592-76-7	19.8	20.6	0.8
1-Hexene	592-41-6	17.9	19.7	1.8
2-Methyl-2-butene	513-35-9	16.5	19.1	2.6
1-Nonene	124-11-8	22.6	22.3	-0.3
1-Octene	111-66-0	21.3	21.5	0.2
<i>cis</i> -2-Pentene	627-20-3	16.8	18.7	1.9
<i>trans</i> -2-Pentene	646-04-8	16.4	18.7	2.3
1-Tridecene	2437-56-1	25.8	25.8	0.0
Aldehydes				
Benzaldehyde	100-52-7	38	39.8	1.8
Butyraldehyde	123-72-8	24.4	25.3	0.9
2-Furaldehyde	98-01-1	43.1	43.1	0.0
Heptanaldehyde	111-71-7	26.3	26.8	0.5
<i>o</i> -Methoxybenzaldehyde	135-02-4	42.6	39.8	-2.8
<i>p</i> -Methoxybenzaldehyde	123-11-5	42.1	39.6	-2.5
Paraldehyde	123-63-7	25.6	26.8	1.2
Pentanaldehyde	110-62-3	25.4	25.7	0.3
Ketones				
Acetone	67-64-1	23.5	25.1	1.6
Acetophenone	98-86-2	39.1	38.1	-1.0
2-Butanone	78-93-3	24	25.2	1.2
Carvone	99-49-0	34.2	34.1	-0.1
Cyclopentanone	120-92-3	32.8	31.1	-1.7

(Continued)

Table 1. Continued.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd</i>	<i>Res.</i>
2-Heptanone	110-43-0	26.1	25.4	-0.7
3-Heptanone	106-35-4	25.7	25.5	-0.2
4-Heptanone	123-19-3	25.5	25	-0.5
2-Hexanone	591-78-6	25.5	24.7	-0.8
2,4-Hexanedione	3002-24-2	29.7	32	2.3
2-Methylcyclohexanone	583-60-8	31.5	31	-0.5
3-Methylcyclohexanone	591-24-2	30.8	30	-0.8
4-Methylcyclohexanone	589-92-4	30.5	31.4	0.9
2,4-Pentanedione	123-54-6	30.4	32.1	1.7
2-Pentanone	107-87-9	23.3	24.2	0.9
3-Pentanone	96-22-0	24.8	25.6	0.8
Ethers				
Butyl ethyl ether	628-81-9	20.2	19.8	-0.4
Butyl methyl ether	628-28-4	19.6	19.3	-0.3
Dibutyl ether	142-96-1	22.5	21.4	-1.1
1,1-Diethoxyethane	105-57-7	20.9	21.5	0.6
Diethoxymethane	462-95-3	20.7	21	0.3
Diethyl ether	60-29-7	16.7	18.2	1.5
Diisopentyl ether	544-01-4	22.6	22.8	0.2
Dipropyl ether	111-43-3	20	19.3	-0.7
Diisopropyl ether	108-20-3	17.3	18.8	1.5
1,1-Dimet-oxyethane	534-15-6	21	21.6	0.6
Dimetoxymethane	109-87-5	20.6	20.7	0.1
1,2-Dimet-oxybenzene	91-16-7	32.8	32.3	-0.5
Dipentyl ether	693-65-2	24.4	23.1	-1.3
1,1-Dipropoxyethane	105-82-8	22.6	22.2	-0.4
Ethylene oxide	75-21-8	23.5	22.8	-0.7
Ethyl methyl ether	540-67-0	15.3	17.7	2.4
Ethyl pentyl ether	17952-11-3	21.7	20.7	-1.0
Ethyl propyl ether	628-32-0	19.3	18.7	-0.6
Ethoxybenzene	103-73-1	32.4	29.9	-2.5
<i>o</i> -Methoxyphenol	90-05-1	38.9	37.2	-1.7
Phenyl propyl ether	622-85-5	31.7	29.3	-2.4
Amines				
Butylamine	109-73-9	23.4	23.8	0.4
<i>sec</i> -Butylamine	13952-84-6	21.1	23.6	2.5
Cyclohexylamine	108-91-8	31.2	29	-2.2
Dibenzylamine	103-49-1	40.6	39.6	-1.0
Dibutylamine	111-92-2	24.1	24.6	0.5
Diethylamine	109-89-7	19.9	21.7	1.8
<i>N,N</i> -Diethylaniline	91-66-7	34	33.1	-0.9
Diisobutylamine	110-96-3	21.7	21.9	0.2
Diisopentylamine	544-00-3	23.9	25.7	1.8
Diisopropylamine	108-18-9	19.2	21.5	2.3
<i>N,N</i> -Dimethylaniline	121-69-7	35.5	34.2	-1.3
Diphenylamine	122-39-4	42.8	40.4	-2.4
Dipropylamine	142-84-7	22.3	22.6	0.3
Isobutylamine	78-81-9	21.8	23.7	1.9
<i>N</i> -Methylaniline	100-61-8	36.9	35.1	-1.8
2-Methylpropylamine	78-81-9	21.8	23.7	1.9
Phenylhydrazine	100-63-0	44.9	42.6	-2.3
Propylamine	107-10-8	21.8	23.4	1.6
<i>m</i> -Toluidine	108-44-1	37.9	35.2	-2.7
<i>p</i> -Toluidine	106-49-0	37.2	35.6	-1.6
Triphenylamine	603-34-9	43.8	43.6	-0.2
Tripropylamine	102-69-2	22.4	24.6	2.2

(Continued)

Table 1. Continued.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd</i>	<i>Res.</i>
Nitriles				
Benzonitrile	100-47-0	38.8	41.4	2.6
Butyronitrile	109-74-0	26.9	29.6	2.7
Ethylcyanoacetate	105-56-6	36.1	37.5	1.4
Hexanenitrile	628-73-9	27.4	28.8	1.4
Methylcyanoacetate	105-34-0	38.7	40	1.3
4-Methylpentanenitrile	542-54-1	26.6	27.8	1.2
Octanenitrile	124-12-9	27.6	29.1	1.5
Phenylacetoneitrile	140-29-4	41.7	41.7	0.0
Succinonitrile	110-61-2	50.6	49.1	-1.5
Nitrate hydrocarbons				
Nitrobenzene	98-95-3	43.5	42.7	-0.8
Nitroethane	79-24-3	32.2	31.1	-1.1
Nitromethane	75-52-5	36.6	35.9	-0.7
<i>o</i> -Nitrophenol	88-75-5	44.4	45.3	0.9
1-Nitropropane	108-03-2	30.1	28.6	-1.5
2-Nitropropane	79-46-9	29.3	28.7	-0.6
<i>o</i> -Nitrotoluene	88-72-2	41.2	39.9	-1.3
<i>m</i> -Nitrotoluene	99-08-1	40.8	39.6	-1.2
<i>p</i> -Nitrotoluene	99-99-0	39.8	39.7	-0.1
Piperidine	110-89-4	28.9	27.4	-1.5
Pyrrole	109-97-7	37.1	37.3	0.2
Pyrrolidine	123-75-1	29.2	27.6	-1.6
Quinoline	91-22-5	42.6	41.2	-1.4
Sulfured hydrocarbons				
1-Butanethiol	109-79-5	25.2	22.4	-2.8
Diethyl sulfide	352-93-2	24.6	22.2	-2.4
Dipentyl sulfide	872-10-6	27.4	28.6	1.2
Ethyl phenyl sulfide	622-38-8	36.5	34.3	-2.2
1-Propanethiol	107-03-9	24.2	21	-3.2
Thiophene	110-02-1	30.7	31.6	0.9
Aromatics				
Biphenyl	92-52-4	39.2	37.9	-1.3
Butylbenzene	104-51-8	28.7	28.5	-0.2
<i>sec</i> -Butylbenzene	135-98-8	28.1	29.6	1.5
<i>tert</i> -Butylbenzene	98-06-6	27.7	28.4	0.7
4- <i>tert</i> -Butylpyridine	3978-81-2	33.1	33.3	0.2
<i>p</i> -Cymene	99-87-6	26.7	28.7	2.0
Ethylbenzene	100-41-4	28.8	28.9	0.1
<i>o</i> -Ethyltoluene	611-14-3	29.7	29.2	-0.5
<i>p</i> -Ethyltoluene	622-96-8	28.3	28.8	0.5
Isobutylbenzene	538-93-2	27	27.8	0.8
Isopropylbenzene	98-82-8	27.7	28.6	0.9
1-Methylnaphthalene	90-12-0	37.6	36.9	-0.7
4-Methylpyridine	108-89-4	34.9	37.4	2.5
Naphthalene	91-20-3	40.1	37.8	-2.3
Propylbenzene	103-65-1	28.5	28.3	-0.2
Pyridine	110-86-1	36.6	35.9	-0.7
Toluene	108-88-3	28	29.3	1.3
1,2,3-Trimethylbenzene	526-73-8	28.3	30	1.7
1,2,4-Trimethylbenzene	95-63-6	29.2	29.3	0.1
1,3,5-Trimethylbenzene	108-67-8	27.6	28.2	0.6
<i>o</i> -Xylene	95-47-6	29.8	29.5	-0.3
<i>m</i> -Xylene	108-38-3	28.5	28.6	0.1
<i>p</i> -Xylene	106-42-3	28	28.8	0.8

(Continued)

Table 1. Continued.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd</i>	<i>Res.</i>
Carboxylic acids				
Acetic acid	64-19-7	29.4	31.5	2.1
Formic acid	64-18-6	37.2	37.7	0.5
Heptanoic acid	111-14-8	27.8	29.7	1.9
Isobutyric acid	79-31-2	24.6	26.6	2.0
3-Methylbutyric acid	503-74-2	25.1	26.8	1.7
Pentanoic acid	109-52-4	26.7	26.5	-0.2
Tetradecanoic acid	544-63-8	31.6	33.8	2.2
Fluorinated hydrocarbons				
1-Fluorohexane	373-14-8	20.9	19.1	-1.8
1-Fluoropentane	592-50-7	19.5	18	-1.5
<i>m</i> -Fluorotoluene	352-70-5	29.2	30.6	1.4
<i>p</i> -Fluorotoluene	352-32-9	27.7	30.7	3.0
Chlorinated hydrocarbons				
Chlorobenzene	108-90-7	33	34.5	1.5
2-Chlorobutane	78-86-4	21.6	19.6	-2.0
1-Chlorododecane	112-52-7	29.3	26.3	-3.0
Chloroform	67-66-3	26.7	28.5	1.8
Chloromethane	74-87-3	15.4	14.9	-0.5
1-Chloro-2-methylpropane	513-36-0	21.7	19.5	-2.2
1-Chloronaphthalene	90-13-1	41.6	41.1	-0.5
2-Chloropropane	75-29-6	19.2	16.1	-3.1
3-Chloro-1-propene	107-05-1	23.2	22	-1.2
<i>p</i> -Chlorotoluene	106-43-4	32.2	32.7	0.5
<i>m</i> -Dichlorobenzene	541-73-1	35.5	38.3	2.8
1,1,2,2-Tetrachloroethane	79-34-5	35.6	34.4	-1.2
1,1,1-Trichloroethane	71-55-6	25.2	24.2	-1.0
Brominated hydrocarbons				
1-Bromobutane	109-65-9	25.9	23.1	-2.8
1-Bromodecane	112-29-8	29.1	26.9	-2.2
1-Bromododecane	143-15-7	30.4	28.5	-1.9
Bromoethane	74-96-4	23.7	21.2	-2.5
1-Bromo-hexadecane	112-82-3	31.2	31.8	0.6
1-Bromohexane	111-25-1	27.4	24.1	-3.3
Bromomethane	74-83-9	23.7	23.7	0.0
1-Bromo-3-methylbutane	107-82-4	25.6	23	-2.6
1-Bromo-2-methylpropane	78-77-3	24.3	25	0.7
1-Bromonaphthalene	90-11-9	43.9	44.8	0.9
1-Bromononane	693-58-3	29.1	26.2	-2.9
1-Bromopropane	106-94-5	25.3	23.5	-1.8
1-Bromotetradecane	112-71-0	30.8	30.1	-0.7
<i>p</i> -Bromotoluene	106-38-7	33.9	36.6	2.7
1-Bromoundecane	693-67-4	29.8	27.7	-2.1
Dibromomethane	74-95-3	39.1	39.3	0.2
1,2-Dibromopropane	78-75-1	33.9	35.4	1.5
1,2,3-Tribromopropane	96-11-7	44.8	49.1	4.3
Esters				
Allyl acetate	591-87-7	25.8	26.9	1.1
Butyl acetate	123-86-4	24.9	23.9	-1.0
<i>tert</i> -Butyl acetate	540-88-5	21.9	22.7	0.8
Butyl butyrate	109-21-7	25.3	24.4	-0.9
Butyl formate	592-84-7	24.5	24.2	-0.3
Butyl propionate	590-01-2	24.9	24.2	-0.7
Diethyl carbonate	105-58-8	25.9	25.6	-0.3
Diethyl maleate	141-05-9	32.1	30.8	-1.3

(Continued)

Table 1. Continued.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd</i>	<i>Res.</i>
Diethyl malonate	105-53-3	31.3	29.2	-2.1
Diethyl oxalate	95-92-1	31.6	29.9	-1.7
Dimethyl carbonate	616-38-6	28.6	28.3	-0.3
Dipropyl carbonate	623-96-1	26.4	25.1	-1.3
Ethyl acetate	141-78-6	23.4	24.1	0.7
Ethyl acetoacetate	141-97-9	31.9	31.2	-0.7
Ethyl benzoate	43-89-0	34.5	33.9	-0.6
Ethyl butyrate	105-54-4	24	23.8	-0.2
Ethyl crotonate	623-70-1	26.7	26.1	-0.6
Ethyl dodecanoate	106-33-2	27.9	28.6	0.7
Ethyl formate	109-94-4	23.2	25.2	2.0
Ethyl fumarate	623-91-6	31.3	30.8	-0.5
Ethyl hexadecanoate	628-97-7	30.7	31.7	1.0
Ethyl hexanoate	123-66-0	25.3	24.7	-0.6
Ethyl isobutyrate	97-62-1	22.7	23.7	1.0
Ethyl lactate	97-64-3	28.3	29.6	1.3
Ethyl-3-methyl butyrate	108-64-5	23.3	23.6	0.3
Ethyl pentanoate	539-82-2	24.7	24.2	-0.5
Ethyl propionate	105-37-3	23.8	23.9	0.1
Hexyl acetate	142-92-7	26	24.7	-1.3
Isobutyl acetate	110-19-0	23.1	23.3	0.2
Isobutyl butyrate	539-90-2	22.4	23.8	1.4
Isobutyl formate	542-55-2	23.3	23.6	0.3
Isobutyl propionate	540-42-1	26.1	23.6	-2.5
Isopentyl acetate	123-92-2	24.3	24.1	-0.2
Isopentyl butyrate	106-27-4	25.1	24.8	-0.3
Isopropyl acetate	108-21-4	21.8	23.2	1.4
Isopropyl formate	625-55-8	21.7	24	2.3
Methyl acetate	79-20-9	24.8	25.4	0.6
Methyl acetoacetate	105-45-3	32.6	33.3	0.7
Methyl benzoate	93-58-3	37.2	35.7	-1.5
2-Methylbutyl acetate	624-41-9	24.3	25	0.7
Methyl butyrate	623-42-7	24.6	24.2	-0.4
Methyl decanoate	110-42-9	28.1	27.1	-1.0
Methyl dodecanoate	111-82-0	29.2	28.5	-0.7
Methyl heptanoate	106-73-0	26.5	25.3	-1.2
Methyl hexadecanoate	112-39-0	29.6	31.5	1.9
Methyl hexanoate	106-70-7	25.9	24.8	-1.1
Methyl isobutyrate	547-63-7	23.2	24.2	1.0
Methyl octanoate	111-11-5	27.4	25.8	-1.6
Methyl pentanoate	624-24-8	25.3	24.5	-0.8
Methyl propionate	554-12-1	24.5	24.6	0.1
1-Methylpropyl acetate	105-46-4	23.1	24.4	1.3
2-Methylpropyl acetate	110-19-0	23.1	23.3	0.2
2-Methylpropyl formate	542-55-2	23.3	23.6	0.3
Methyl salicylate	119-36-8	39.2	38.8	-0.4
Methyl tetradecanoate	124-10-7	29	29.9	0.9
Pentyl acetate	628-63-7	25.2	24.2	-1.0
Pentyl formate	638-49-3	25.5	24.4	-1.1
Propyl acetate	109-60-4	23.8	23.6	-0.2
Propyl benzoate	2315-68-6	33.9	32.9	-1.0
Propyl butyrate	105-66-8	24.6	23.8	-0.8
Propyl formate	110-74-7	24	24.2	0.2
Propyl isobutyrate	644-49-5	23.3	23.8	0.5
Propyl pentoate	141-06-0	25.3	24.4	-0.9
Propyl propionate	106-36-5	24.2	23.8	-0.4

(Continued)

Table 1. Continued.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd</i>	<i>Res.</i>
Alcohols				
Benzyl alcohol	100-51-6	34.8	35.7	0.9
1-Butanol	71-36-3	25	23.2	-1.8
<i>o</i> -Cresol	95-48-7	36.9	35.8	-1.1
<i>m</i> -Cresol	108-39-4	35.7	35.1	-0.6
<i>p</i> -Cresol	106-44-5	36.2	35.2	-1.0
1-Decanol	112-30-1	28.5	28.5	0.0
2,4-Dimethylphenol	105-67-9	32.4	33.4	1.0
2,5-Dimethylphenol	95-87-4	34.6	33.3	-1.3
3,4-Dimethylphenol	95-65-8	33.5	35.5	2.0
3,5-Dimethylphenol	108-68-9	32.1	34.2	2.1
1-Dodecanol	112-53-8	29.4	30.3	0.9
Ethanol	64-17-5	22.1	21.2	-0.9
1-Hexanol	111-27-3	25.8	25	-0.8
Methanol	67-56-1	22.1	21	-1.1
2-Methyl-2-butanol	75-85-4	22.3	23.1	0.8
3-Methyl-1-butanol	123-51-3	23.7	23.8	0.1
<i>cis</i> -2-Methylcyclohexanol	7443-70-1	30.5	29.4	-1.1
2-Methyl-1-pentanol	105-30-6	25	25.4	0.4
3-Methyl-1-pentanol	589-35-5	25	25.9	0.9
4-Methyl-1-pentanol	626-89-1	24.1	24.6	0.5
2-Methyl-2-pentanol	590-36-3	22.9	23.1	0.2
3-Methyl-2-pentanol	565-60-6	24.9	25.5	0.6
4-Methyl-2-pentanol	108-11-2	22.6	23.5	0.9
2-Methyl-3-pentanol	565-67-3	24.2	23.6	-0.6
3-Methyl-3-pentanol	77-74-7	23.3	24.6	1.3
2-Methyl-1-propanol	78-83-1	22.6	22.3	-0.3
1-Nonanol	143-08-8	27.8	27.6	-0.2
1-Octanol	111-87-5	27.1	26.8	-0.3
2-Octanol	123-96-6	25.9	26.1	0.2
2-Pentanol	6032-29-7	23.5	22.8	-0.7
1-Propanol	71-23-8	23.3	21.9	-1.4
2-Propanol	67-63-0	21	20.5	-0.5
2-Propen-1-ol	107-18-6	25.3	25.9	0.6
1-Tetradecanol	112-72-1	31	32	1.0
Thymol	89-83-8	31.9	34.2	2.3
Amides				
Acetanilide	103-84-4	43.9	41.7	-2.2
Benzamide	55-21-0	45.5	44.7	-0.8
Formanilide	103-70-8	42.1	43.3	1.2
2-Phenylacetamide	103-81-1	44.3	43.3	-1.0
Propionamide	79-05-0	36.8	34.1	-2.7
Miscellaneous				
Benzoylbromide	618-32-6	42.4	43.4	1.0
Benzoylchloride	98-88-4	38.6	39.4	0.8
<i>p</i> -Bromophenol	106-41-2	46.2	44.8	-1.4
<i>o</i> -Chloroaniline	95-51-2	41.2	41	-0.2
<i>o</i> -Chloronitrobenzene	88-73-3	45.2	46.5	1.3
<i>m</i> -Chloronitrobenzene	121-73-3	46.2	46.1	-0.1
<i>p</i> -Chloronitrobenzene	100-00-5	43.2	46.2	3.0
<i>o</i> -Chlorophenol	95-57-8	39.7	39.9	0.2
<i>m</i> -Chlorophenol	108-43-0	41.2	40.4	-0.8
2,4-Dichlorophenol	120-83-2	43.5	43.2	-0.3
Ethyl chloroformate	541-41-3	26.2	28.8	2.6
Ethyl dichloroacetate	535-15-9	32	30.6	-1.4
Ethyl thiocyanate	542-90-5	34.2	36.8	2.6
Methyl dichloroacetate	116-54-1	34	32.1	-1.9
Phenyl isothiocyanate	103-72-0	40	40.9	0.9
<i>p</i> -Toluenesulfonylchloride	98-59-9	40.2	40.7	0.5
Triethyl phosphate	78-40-0	29.5	29.9	0.4

searches for the best multiple linear relationships between predicted and experimental property data.

Prior to the descriptor calculation and fitting of these descriptors to the experimental data, modeling was performed in order to set the chemicals in their lowest energy 3D conformations. To do this, initial three-dimensional geometries of the chemical structures were generated using the Hyperchem 7.0 molecular modeling package [4]. These 3D structures were refined later using Ampac 5.0, a semiempirical molecular modeling program [5], using AM1 parametrization. The Ampac output files containing the refined geometries and electron wave functions of individual compounds, along with the experimental values of surface tension, were loaded into the Codessa program to calculate the molecular descriptors and fitting of these descriptors to the experimental data.

3. Results

The correlation analysis to find the best QSPR model was carried out using the best multilinear regression analysis method available in the Codessa program. The procedure begins with the computation of the molecular descriptors, in this study a total of 650 molecular descriptors were calculated for all 320 compounds. Then, the pool of molecular descriptors was reduced by removing descriptors that could not be calculated for every structure in the data set, and those descriptors with an essentially constant value for all the structures. Thereafter, the intercorrelations between all the remaining descriptors were calculated, and orthogonal pairs of descriptors i and j (with $r_{i,j}^2 < 0.1$) were selected. From the pairs of descriptors which did not meet this criterion only one descriptor was retained, that which correlated better with the property. From this set of orthogonal pairs, the best descriptors pairs defined as those with the highest two parameter regression correlation coefficients were selected for further development to higher order regressions. Next, for each of these pairs selected, an orthogonal descriptor was added, and the three-parameter regression was calculated. This procedure was repeated with all orthogonal descriptors to a given pair of descriptors, and the best triplets, defined as those with highest correlation coefficients, was selected for the next higher order regression analysis. In this way, the number of orthogonal descriptors in the model was incrementally increased up to the optimum as determined by the Fisher criterion at a given probability level and the cross-validated correlation coefficient. The model obtained with this procedure was expected to yield maximum predictive ability.

The best correlation found involved six descriptors, four constitutional parameters (relative number of carbon atoms (N_C^R), relative number of oxygen atoms (N_O^R), relative number of N atoms (N_N^R), relative molecular weight (M_W^R)), one topological index (Kier and Hall index order 3, (${}^3\chi^Y$)), and one descriptor accounting for hydrogen bonding (HA dependent HDSA-1):

$$\gamma = -6.45 + 56.60 N_C^R + 48.40 N_O^R + 83.09 N_N^R + 0.98 M_W^R + 3.47 {}^3\chi^Y + 0.16 \text{HA}$$

where the relative parameters ($N_C^R, N_O^R, N_N^R, M_W^R$) are obtained dividing the respective figures by the number of atoms in the molecule. The statistics for the fit was the following: $r^2 = 0.96$, $F = 1317$, $s = 1.43$ and $r_{CV}^2 = 0.95$. Where r^2 is the squared

Table 2. Correlation coefficients for the six descriptors involved in the model.

Descriptor	Coefficient	<i>t</i> -test
Intercept	-6.45 ± 0.43	-14.89
Relative number of C atoms	56.60 ± 1.11	51.14
Relative number of O atoms	48.40 ± 1.42	34.22
Relative number of N atoms	83.09 ± 3.08	26.99
Relative molecular weight	0.98 ± 0.03	37.86
Kier & Hall index (order 3)	3.47 ± 0.12	28.60
HA dependent HDSA-1	0.16 ± 0.01	23.01

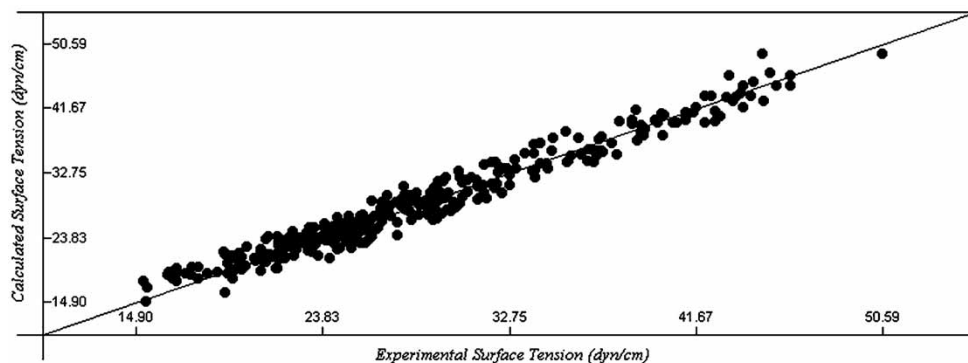


Figure 1. Scatter plot of the calculated vs. experimental surface tension values.

correlation coefficient, F is the Fisher test value, s is the standard deviation, and r_{CV}^2 is the squared cross-validated correlation coefficient. This last coefficient provides an estimation of the stability of the obtained regression model, i.e. the sensitivity of the model to the elimination of any single data point. For each experimental data point, the regression is recalculated with the same descriptors but for the data set without this point. The obtained regression is used to predict the value of this point, and the set of calculated surface tension in this way is correlated with the experimental values. For this model, the squared cross-validated coefficient has the value 0.95, as compared to the value of 0.96 of the squared correlation coefficient, indicating a good stability of the regression model.

Addition of more descriptors in the regression equation resulted in higher correlation coefficients, but lower F statistics values, suggesting that the additional descriptors were not contributing to improve the fit to the actual property but rather to the error in the measurements. On the other hand, it is important to develop regressions with as few parameters as possible. Although, the model found includes six parameters, four of them (constitutional descriptors) are easily obtained from the molecular formula.

In table 2, the descriptors involved in the best correlation equation along with their respective correlation coefficients, standard errors and the t -test values are shown. The calculated and experimental values of surface tension are compared in table 1, and the scatter plot is shown in figure 1.

The predictive performance of the model was estimated from an external test set of chemicals not included in the training set. The validation set included 55 compounds with a diverse selection of chemical structures. Table 3 lists the experimental and

Table 3. Experimental and calculated surface tension values (dyn cm^{-1}) for the test set.

<i>Name</i>	<i>CAS number</i>	<i>Exp.</i>	<i>Calcd.</i>	<i>Res.</i>
1,1,2,2-Tetrachloroethane	79-34-5	35.6	34.2	-1.4
1,1-Dichloroethane	75-34-3	24.1	22.4	-1.7
1-Bromo-4-chlorobenzene	106-39-8	37.5	37.4	-0.2
1-Bromopentane	110-53-2	26.9	25.1	-1.8
1-Chloro-3-methylbutane	107-84-6	22.8	21.9	-0.9
1-Chlorobutane	109-69-3	23.2	22.3	-0.9
1-Chloropropane	540-54-5	21.3	20.9	-0.4
1-Iodo-2-methylpropane	513-38-2	29.8	30.5	0.7
1-Iodo-3-methylbutane	541-28-6	28.1	27.3	-0.8
1-Iodoheptane	4282-40-0	30.0	28.4	-1.6
1-Iodohexadecane	544-77-4	32.3	33.8	1.6
1-Iodohexane	638-45-9	29.5	28.0	-1.5
1-Iodooctane	629-27-6	30.2	28.8	-1.4
1-Iodopentane	628-17-1	28.9	27.9	-1.0
1-Iodopropane	107-08-4	28.8	27.9	-0.9
2-Bromopropane	75-26-3	23.3	22.1	-1.1
2-Ethoxyethanol	110-80-5	28.4	30.4	2.0
2-Iodobutane	513-48-4	27.7	28.6	0.9
2-Iodopropane	75-30-9	26.6	28.6	2.0
2-Methoxyethanol	109-86-4	30.9	31.2	0.4
2-Propanethiol	75-33-2	21.3	23.6	2.3
3-Methylpyridine	108-99-6	34.5	36.1	1.6
4-Oxopentanoic acid	123-76-2	39.8	38.6	-1.2
Benzylamine	100-46-9	39.3	35.9	-3.4
Bromobenzene	108-86-1	35.2	34.2	-1.0
Carbendisulfide	75-15-0	31.6	32.3	0.7
Chloroacetic acid	79-11-8	40.5	41.1	0.6
Cyclohexane	110-82-7	24.7	23.8	-0.9
Cyclohexanol	108-93-0	32.9	31.5	-1.5
Cycloheptanol	502-41-0	32.7	32.5	-0.2
Cyclohexanone	108-94-1	34.6	32.9	-1.7
Cyclopentane	287-92-3	21.9	23.0	1.1
Cyclopentanol	96-41-3	32.5	31.5	-1.0
Diethylsulfate	64-67-5	33.1	32.2	-0.8
Dimethylamine	124-40-3	26.4	26.2	-0.1
Dimethylsulfide	75-18-3	24.1	24.1	0.0
Dipropoxymethane	505-84-0	22.8	24.2	1.4
Ethanethiol	75-08-1	23.1	24.2	1.1
Ethylcyclohexane	1678-91-7	25.2	26.3	1.2
Ethylmethylsulfide	624-89-5	24.4	25.5	1.0
Fluorobenzene	462-06-6	26.7	28.7	2.1
Iodobenzene	591-50-4	38.7	37.7	-1.0
Iodoethane	75-03-6	28.5	27.7	-0.8
Iodomethane	74-88-4	30.4	32.9	2.6
Methanethiol	74-93-1	23.9	23.8	-0.1
Methylcyclohexane	108-87-2	23.3	25.0	1.7
Methylcyclopentane	96-37-7	21.7	24.2	2.5
Methylphenylsulfide	100-68-5	39.7	36.5	-3.3
<i>o</i> -Bromotoluene	95-46-5	34.2	34.2	0.0
<i>p</i> -Dibromobenzene	106-37-6	39.3	41.0	1.7
<i>p</i> -Dichlorobenzene	106-46-7	32.5	33.7	1.2
Phenylsalicylate	118-55-8	42.8	43.5	0.7
<i>p</i> -Iodotoluene	624-31-7	36.8	35.8	-1.0
<i>p</i> -Tolunitrile	104-85-8	37.0	37.9	0.9
Tribenzylamine	620-40-6	40.0	41.7	1.7

calculated surface tension values for the validation set. The respective statistical analysis for calculated *versus* experimental values for validation is as follows: $r^2 = 0.94$; $F = 133.4$; $s = 1.52$. This result confirms the predictive capability of the model.

4. Discussion

Surface tension is closely related to the forces of intermolecular attraction. The stronger the intermolecular forces are, the more tightly the molecules are held together in the liquid phase and, therefore the higher the surface tension will be. The main attractive interactions between uncharged species are the 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, known also as London forces. In general, the London dispersion energy is the dominant contribution of the van der Waals interactions.

In contrast to van der Waals interactions, 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. Thus, hydrogen-bond are specific interactions and, like van der Waals interactions, are always attractive.

In order to allow a more transparent physical-chemical interpretation of the descriptors involved in the model, the different types of intermolecular interactions which may occur in the bulk are discussed separately below. The relative constitutional descriptors in the model account for the significance of the respective interactions in a molecule in particular.

4.1 Descriptors involved in dispersion interactions

Experimental findings [6, 7] show that surface tension increases as a function of the number of carbon atoms and molecular weight for a set of congeners. Light-scattering and heats of mixing measurements have shown that in the higher *n*-alkanes, adjacent chains are oriented parallel to one another, giving rise to an appreciable enhancement in adhesion as consequence of the much greater polarizability parallel to the chains.

It is well known that molecular polarizability is directly proportional to the number of electrons in the molecule. On the other hand, molecular weight is roughly related to the number of electrons in the molecule, the higher the molecular weight the higher the size of the electron cloud, and in consequence the higher the polarizability of the molecule. Therefore, molecular weight encodes information related to molecular polarizability.

The Kier and Hall index (order 3), ${}^3\chi^V$, encodes information related to geometrical features of the molecules, namely, the molecular van der Waals volume [8]. The excellent correlation between the molecular volume and ${}^3\chi^V$ is not surprising because the connectedness of halogens, introduced by Kier and Hall, was calculated using molecular refraction, a well-known measure of the van der Waals volume. Since it has been well established that the polarizability of an atom or a molecule is proportional to

its volume, it may be concluded, in consequence, that ${}^3\chi^V$ represents a measure of the molecular polarizability.

Therefore, it is evident from the above discussion that the above three descriptors (N_C^R , M_W^R , ${}^3\chi^V$) encode information related to the polarizability of the molecules. This affirmation is supported with a squared correlation coefficient of 0.75 for the multilinear correlation between these three descriptors and the molecular polarizability, the ease with which the molecular electron cloud can distort by an electric field. 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, consequently the above three descriptors account for dispersion interaction between molecules.

4.2 Descriptors involved in polar interactions

The number of atoms of oxygen and nitrogen in the molecules encodes information concerning polar interactions among molecules in the bulk. The introduction of substituents into organic compounds with increasing differences in electronegativity with respect to carbon, e.g. nitrogen or oxygen atoms, produces a charge separation in the bond generating a dipole moment, and provided the dipole moment orientations of neighboring molecules. There will be an attractive interaction among them which will be seen reflected in an enhanced value of surface tension. The positive values of the model correlation coefficients for these descriptors support this interpretation.

4.3 Descriptors involved in hydrogen-bond interactions

The hydrogen acceptor dependent hydrogen donor surface area descriptor, HA dependent HDSA-1, is connected with the hydrogen-bonding ability of the molecule. It is expected that the adhesion of the molecules in the surface layer to those in the underlying bulk liquid would be strongly enhanced by the presence of hydrogen bonding interactions. Accordingly, there have been many generalized statements made which associate hydrogen-bonding intermolecular interactions with increased surface tension of pure liquids.

5. Conclusions

The model reported in this article allows the prediction of surface tension of a wider variety of organic compounds, with less parameters and with better statistics than other QSPR models reported in literature. Thus, a ten-parameter model having a squared correlation coefficient of 0.983 for a dataset of 146 structures, including alkanes, esters and alcohols; has been published by Stanton and Jurs [9]. Kauffman and Jurs [10] have designed an eight-descriptor model for a data set of 159 structures. Their model predicts surface tension with a squared correlation coefficient of 0.83. On the other hand, Freitas *et al.* [11] report a six-parameter model, developed from a linear free energy analysis, which predicts surface tension of a data set of 299 compounds with a squared correlation coefficient of 0.88. The Parsimony Principle (Occam's Razor Principle) calls for using models and procedures that contain all that is necessary for the modeling

but nothing more, i.e. given a number of models with nearly the same predictive error, that containing fewer parameters should be preferred because simplicity is desirable in itself [12].

The six descriptors involved in our model, which can be calculated from the molecular structure, have definite physical meaning corresponding to the different intermolecular interactions which take place in the bulk solution, namely, dispersion, polar, and hydrogen-bond interactions. Since the model is based only on molecular descriptors, it is applicable to new and developing compounds for which no group contributions have been fitted before. Therefore, the prediction of surface tension can be made prior to synthesis and the design of novel compounds with certain desired value of surface tension may, in this way, be guided by the results of calculations.

Acknowledgment

The authors acknowledge financial support from Fondecyt under project No. 1020464.

References

- [1] P. Mach, C.C. Huang, T. Stoebe, E.D. Wedell, T. Nguyen, W.H. De Jeu, F. Guittard, J. Naciri, R. Shashidhar, N. Clark, I.M. Jiang, F.J. Kao, H. Liu, H. Nohira. *Langmuir*, **14**, 4330 (1998).
- [2] J.A. Dean. *Lange's Handbook of Chemistry*, 14th Edn, McGraw-Hill, New York (1992).
- [3] A.R. Katritzky, V.S. Lobanov and M. Karelson, Codessa version 2.0, Reference Manual (1994).
- [4] *Hyperchem 7.0*; Hypercube, Inc., 1115 NW 4th Street, Gainesville, FL 32601, USA (2002).
- [5] *Ampac 5.0*; Semichem, 7128 Summit, Shawnee, KS 66216, USA (1994).
- [6] F.M. Fowkes. *J. Phys. Chem.*, **84**, 510 (1980).
- [7] D.G. Legrand, G.L. Gaines Jr. *J. Colloid Interface Sci.*, **42**, 181 (1973).
- [8] J.K. Labanowski, I. Motoc, R.A. Dammkoehler. *Comput. Chem.*, **15**, 47 (1991).
- [9] D.T. Stanton, P.C. Jurs. *J. Chem. Inf. Comput. Sci.*, **32**, 109 (1992).
- [10] G.W. Kauffman, P.C. Jurs. *J. Chem. Inf. Comput. Sci.*, **41**, 408 (2001).
- [11] A.A. Freitas, F.H. Quina, F.A. Caroll. *Langmuir*, **16**, 6689 (2000).
- [12] E. Estrada, E.J. Delgado, J.B. Alderete, G.A. Jaña. *J. Comput. Chem.*, **25**, 1787 (2004).