

QSPR MODELING OF NORMAL BOILING POINT OF ALDEHYDES, KETONES, AND ESTERS BY MEANS OF NEAREST NEIGHBORING CODES CORRELATION WEIGHTING

DUCHOWICZ, P.¹; CASTRO, E.A.^{1,*}; TOPOROV, A.A.²

¹CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas,
Universidad Nacional de La Plata, C.C. 962, La Plata 1900, Argentina.
e-mail: castro@dalton.quimica.unlp.edu.ar or jubert@arnet.com.ar

²Vostok Innovation Company, 4th Sidik Azim Street, Home 3, Tashkent 700047, Uzbekistan

Received: March 14, 2002. In final form: June 13, 2002

ABSTRACT

A particular sort of variable descriptor is employed to correlate structure and normal boiling points of acyclic carbonyl compounds comprising mono- and dialdehydes, mono- and diketones, keto aldehydes, and esters of monocarboxylic acids with various degrees of branching but devoid of hydrogen bonding. Results are compared with other data derived from a study made in terms of five "rigid" topological indices and they show the advantages of resorting to flexible molecular descriptors to attain accurate enough predictions. Some possible extensions of the method are pointed out.

RESUMEN

Una variedad particular de variable descriptiva se utiliza para correlacionar la estructura con la temperatura normal de ebullición de compuestos carbonílicos acíclicos, incluyendo mono y dialdehídos, mono y dicetonas, ceto-aldehídos y ésters de ácidos monocarboxílicos con varios grados de ramificación pero carentes de enlaces por puente de hidrógeno. Los resultados se comparan con los datos obtenidos de un estudio reciente realizado en función de 5 índices topológicos "rígidos" mostrándose las ventajas de utilizar descriptores moleculares flexibles para obtener predicciones con exactitud satisfactoria. Se sugieren algunas posibles extensiones del método.

INTRODUCTION

The correlation, rationalization and prediction of physicochemical properties of pure liquids and of mixtures, such as boiling point, density, viscosity, critical temperatures, flash points, static dielectric constant, and refractive index, is of practical (process design and control) and theoretical (role of the molecular structure in determining the macroscopic properties of the solvent) relevance to both chemists and chemical engineers. The boiling point of a compound is predetermined by the intermolecular interactions in the liquid and by the difference in the molecular internal partition function in the gas phase and in the liquid at the boiling temperature. Then, it is directly related to the molecular chemical structure and several methods have been indeed developed for calculating the normal boiling point (nbp) of a compound from its molecular structure [1].

Other physical chemistry properties, such as critical temperatures [2] and flash points [3] can be estimated from nbp. Various rules and formulae were proposed early on to correlate nbp of homolo-

gous hydrocarbons with the number of carbon atoms or molecular weight [4]. Later on, other methods have employed physical parameters such as parachor and the molar refractivity [5]. Previous methods to calculate nbp have been reviewed [6]. More recently, the Quantitative Structure-Property Relationships (QSPR) approach [7,8] based on calculated molecular descriptors has been applied to predict nbp [9,10].

The employment of calculated molecular descriptors in QSPR analysis has two main advantages: (1) the descriptors can be univocally defined for any molecular structure or/and fragment; (2) thanks to the high and well-defined physical information content encoded into many theoretical descriptors, they can clarify the mechanism relating the studied property with the chemical structure. Moreover, QSPR models based on calculated molecular descriptors help to understanding of the inter- and intramolecular interactions that are mainly responsible for the behavior of complex chemical systems and processes [11]. The basis of QSPR is the assumption that compounds of similar structure will exhibit similar properties. The key to the success of QSPR is accurate measurement of the structural measurement of the structural features that modulate the observed property. The structural characteristics of a molecule can be measured on either a substructure or a whole-molecule basis. These measures, or *descriptors*, encode structural properties that fall broadly into three main classes: (a) topological, (b) geometric, and (c) electronic. Methods that apply multiple linear regressions for selection of descriptors and generation of predictive equations for nbp of large sets of heterocyclic organic compounds have been described [12,13]. Later, it was shown that the accuracy of similar models could be improved by employing variable descriptor techniques [14]. These QSPR models were also shown to provide greater accuracy than previous methods based on fixed descriptor techniques [15].

In this paper we present a comparative study of correlation between structure and nbp of acyclic carbonyl compounds on the basis of a particular sort of a variable descriptor: the nearest neighboring codes correlation weighting. Results are compared with other previous ones derived from five fixed topological indices. Next section deals with the description of this new index, illustrating its calculation for a prototype molecule. Then we give results for a set of 200 carbonyl compounds, and we analyze the quality of the data obtained for this particular molecular group, making the suitable comparisons with previous estimations. Finally, we discuss the meaning of the present approach as well as the convenience and possibility to extend calculations to calculate other physical chemistry properties and/or different molecular sets.

Nearest Neighboring Codes Correlation Weighting

Most difficulties authors have encountered with multivariate regression analysis (MRA) arose from either the use of too many descriptors or descriptors that are highly interrelated. This very well known feature includes the "nightmares" of the regression equations, the "nightmares" of the chaotic selection of suitable descriptors, as well as ambiguities of the criteria employed to choose optimal descriptors and uncertainties when selecting the order in which descriptors are to be orthogonalized. None of these difficulties exists for simple regression based on a unique molecular descriptor, particularly if the regression is linear. This is one of the major reasons why researchers are striving to find or to design novel descriptors that would produce good correlation for a single molecular property of a set of compounds. However, one must take into account that not many molecular properties can be sufficiently well described by a single descriptor, although it would be rather desirable. Are there such sorts of "optimal" molecular descriptors? The answer is positive, as we will see rather soon [16].

In order to come up with much better regression correlations, we need an important improvement in the design of topological indices and other molecular descriptors. One such promising direction was outlined about ten years ago, although it has been overlooked for too long [17,18]. Recently there has been a sort of renaissance to these promising descriptors by publishing a series of papers in

which their abilities were well illustrated [19-25]. What characterizes these novel descriptors is their dependence on a variable, or several variables, that determined during the search for best correlation. In this regard, these newly defined descriptors are fundamentally different from the myriad of available topological indices [26] in that they are variables themselves rather than numerically fixed quantities.

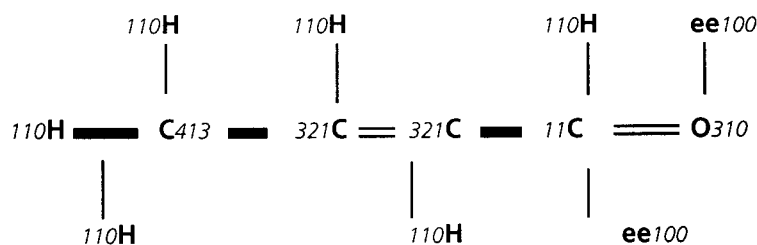
Contrasting to the usual employment of topological indices, which one can calculate after selecting of a set of compounds to be studied and then proceeding with the corresponding statistical analysis, the variable indices are initially non-numerical. Hence, they cannot be calculated in advance for the set of compounds. Instead, one starts with an arbitrary set of values for the yet undetermined variables and through an iterative procedure, varies these initial values seeking for those values that will yield the smallest standard error for the property under consideration. From these general remarks concerning variable topological indices it is clear that their use can only improve correlations over the use of simple indices because in case the variables takes on a zero value (a very improbable fact), we would obtain the results that coincide with those results based on the traditional fixed topological indices. Thus, this method leads us to a natural improvement. However, it is natural to ask: How much better? The use of variable molecular descriptors improves regression statistics significantly.

A correlation weight of local graph invariants (CWLI) is a new kind of variable descriptor, which has proved to be quite successful to predict physical chemistry properties [28]. The definition of this index is based on the nearest neighboring code (NNC) of a vertex in a molecular graph, which is calculated as follows

$$NNC_k = 100(N) + 10(N_c) + N_H \quad [I]$$

where N is the total number of all neighbors of the k -th vertex, N_c is the number of neighbors of the k -th vertex which are images of carbon atoms, and N_H is the number of neighbors of the k -th vertex which are images of the hydrogen atoms. The following scheme shows the NNC_k values for the $CH_3CH=CH=O$, where Smiles notation is used and ee denotes lone pairs.

SCHEME



The basic steps to calculate the optimized CWLI are the following [27]:

- Choose a molecular descriptor, which is a function of the CWLI. In the present case such a descriptor is calculated as

$$DCW = \sum_{\text{vertices}} [CW(A_k) + CW(NNC_k)] \quad [II]$$

where A_k is the chemical element which is image of the k -th vertex of the graph, $CW(A_k)$ stands for the correlation weights of A_k ; NNC_k is calculated via Eq.[I], and $CW(NNC_k)$ are the correlation weights

corresponding to the molecular graph associated to the NNC_k .

- CW's are computed by means of an optimization procedure performed in such a way to produce as large as possible correlation coefficient between descriptor (2) and the physical chemistry property under consideration for the molecules included in the training set.
- After finding CW's, DCW's are computed for the components of the training set through equation 2.
- Finally, the physical chemistry property is computed through a general mathematical relationship

$$\text{Property} = f(\text{DCW}) \quad \text{[III]}$$

Usually, f is a polynomial function, although other analytical structures may be employed.

Calculation of normal boiling points of aldehydes, ketones, and esters

We have the method described above to a set of 200 acyclic saturated and unsaturated carbonyl compounds (mono- and dialdehydes, mono- y diketones, keto aldehydes, and esters of monocarboxylic) with various degrees of branching but devoid of hydrogen bonding to study nbp's. We have chosen the same molecular set employed by Balaban *et al* [28] since it is a well representative molecular group and at the same time it allows us to make a direct comparison with theoretical calculations performed on the basis of five fixed topological descriptors. The experimental data are presented in Table 1 together with the structures indicated by Smiles notation [29]. For acyclic compounds, this notation is simpler to follow than the usual IUPAC nomenclature since there is no ambiguity in how double bonds are denoted. Besides, triple bonds are indicated by the symbol #. We also include in the last column for comparative purposes theoretical results derived from the application of fixed five topological descriptors [28].

TABLE 1. Data for 200 carbonyl compounds.

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [28]
1	C ₂ H ₄ O*	CC=O	21	45
2	C ₃ H ₂ O*	O=CC#C	56	61
3	C ₃ H ₄ O*	C=CC=O	53	73
4	C ₃ H ₆ O*	CCC=O	48	54
5	C ₃ H ₆ O*	CC(=O)C	56	44
6	C ₄ H ₄ O*	CC#CC=O	107	104
7	C ₄ H ₄ O*	CC(=O)C#C	84	67
8	C ₄ H ₆ O^	CC=CC=O	105	110
9	C ₄ H ₆ O*	CC(=C)C=O	68	83
10	C ₄ H ₆ O*	CC(=O)C=C	81	83
11	C ₄ H ₈ O*	CCCC=O	75	79
12	C ₄ H ₈ O*	CC(C)C=O	64	71
13	C ₄ H ₈ O^	CC(=O)CC	80	73
14	C ₅ H ₆ O^	C#CC(C)C=O	91	84
15	C ₅ H ₆ O*	CCC(=O)C#C	106	82
16	C ₅ H ₈ O*	C=CCCC=O	99	100
17	C ₅ H ₈ O*	CC=CCC=O	106	118
18	C ₅ H ₈ O*	CCC=CC=O	124	125

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [2a]
19	C ₅ H ₈ O [^]	CCC(=O)C=C	103	97
20	C ₅ H ₈ O*	CC=CC(C)=O	122	122
21	C ₅ H ₈ O*	CC(C)=CC=O	133	115
22	C ₅ H ₈ O*	CC=C(C)C=O	117	114
23	C ₅ H ₈ O*	CC(=C)C(C)=O	98	97
24	C ₅ H ₈ O ₂ *	CCC(=O)C(C)=O	108	115
25	C ₅ H ₁₀ O*	CCCCC=O	103	107
26	C ₅ H ₁₀ O [^]	CC(C)CC=O	93	93
27	C ₅ H ₁₀ O [^]	CC(=O)CCC	102	99
28	C ₅ H ₁₀ O*	CCC(C)=O	91	96
29	C ₅ H ₁₀ O*	CCC(=O)CC	102	90
30	C ₅ H ₁₀ O [^]	CC(=O)C(C)C	94	91
31	C ₅ H ₁₀ O*	CC(C)(C)C=O	75	83
32	C ₆ H ₈ O*	CC=CC=CC=O	174	172
33	C ₆ H ₈ O [^]	C#CC(=O)C(C)C	118	98
34	C ₆ H ₈ O*	CC=C(C=C)C=O	144	136
35	C ₆ H ₁₀ O*	CCC=CCC=O	121	140
36	C ₆ H ₁₀ O [^]	CCCC=CC=O	146	143
37	C ₆ H ₁₀ O*	CC(=O)CCC=C	130	117
38	C ₆ H ₁₀ O [^]	CC=CCC(C)=O	127	133
39	C ₆ H ₁₀ O [^]	CCC=CC(C)=O	140	142
40	C ₆ H ₁₀ O [^]	CC(=C)CC(C)=O	124	113
41	C ₆ H ₁₀ O*	CCC(=O)C=CC	139	134
42	C ₆ H ₁₀ O*	CCC=C(C)C=O	136	135
43	C ₆ H ₁₀ O [^]	CCC(=O)C(C)=C	119	114
44	C ₆ H ₁₀ O [^]	CC(C)=CC(C)=O	130	132
45	C ₆ H ₁₀ O ₂ *	CCC(=O)C(=O)CC	130	124
46	C ₆ H ₁₂ O*	CCCCCC=O	131	129
47	C ₆ H ₁₂ O*	CC(C)CCC=O	122	119
48	C ₆ H ₁₂ O [^]	CC(=O)CCCC	127	125
49	C ₆ H ₁₂ O*	CCC(C)CC=O	122	120
50	C ₆ H ₁₂ O*	CCCC(C)C=O	116	120
51	C ₆ H ₁₂ O*	CCC(=O)CCC	123	121
52	C ₆ H ₁₂ O [^]	CC(=O)CC(C)C	117	114
53	C ₆ H ₁₂ O [^]	CCC(CC)C=O	118	112
54	C ₆ H ₁₂ O*	CC(C)C(C)C=O	116	116
55	C ₆ H ₁₂ O [^]	CC(=O)C(C)CC	118	115
56	C ₆ H ₁₂ O*	CC(C)C(=O)CC	115	113
57	C ₆ H ₁₂ O [^]	CC(=O)C(C)(C)C	106	102
58	C ₇ H ₁₂ O [^]	CCCCCC=CC=O	166	162
59	C ₇ H ₁₂ O*	CCCC=CCC=O	157	161
60	C ₇ H ₁₂ O*	CC=CCCC(C)=O	153	151
61	C ₇ H ₁₂ O*	CCC(=O)CCC=C	142	139
62	C ₇ H ₁₂ O*	CC(=C)CCC(C)=O	150	135
63	C ₇ H ₁₂ O [^]	CC(CC(C)=O)C=C	138	132
64	C ₇ H ₁₂ O [^]	CC(CC=C)C(C)=O	137	132
65	C ₇ H ₁₂ O*	CC=CC(=O)C(C)C	148	148

(cont.)

(Table 1 cont.)

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [28]
66	C ₇ H ₁₄ O*	CCCCCCC=O	153	150
67	C ₇ H ₁₄ O^	CC(=O)CCCCC	151	147
68	C ₇ H ₁₄ O^	CCC(C)CCC=O	144	144
69	C ₇ H ₁₄ O*	CCC(=O)CCCC	149	146
70	C ₇ H ₁₄ O*	CCCC(C)CC=O	144	143
71	C ₇ H ₁₄ O^	CC(=O)CCC(C)C	144	138
72	C ₇ H ₁₄ O^	CCCC(=O)CCC	144	139
73	C ₇ H ₁₄ O^	CC(=O)CC(C)CC	140	139
74	C ₇ H ₁₄ O^	CCC(=O)CC(C)C	136	137
75	C ₇ H ₁₄ O^	CC(=O)C(C)CCC	143	138
76	C ₇ H ₁₄ O*	CC(C)C(=O)CCC	135	136
77	C ₇ H ₁₄ O^	CCC(=O)C(C)CC	135	134
78	C ₇ H ₁₄ O*	CC(C)C(=O)C(C)C	124	126
79	C ₇ H ₁₄ O*	CCC(=O)C(C)(C)C	125	125
80	C ₉ H ₁₄ O*	CC(C)=CC(=O)C=C(C)C	198	186
81	C ₈ H ₁₄ O*	CC(C)C=CCC(C)=O	163	174
82	C ₈ H ₁₄ O^	CC(C)CC=CC(C)=O	179	174
83	C ₈ H ₁₄ O*	CC=CC(=O)CC(C)C	170	162
84	C ₈ H ₁₄ O^	CC=CCC(=O)C(C)C	162	162
85	C ₈ H ₁₄ O*	CCC(C)C(=O)C=CC	171	164
86	C ₈ H ₁₄ O*	CCC(C)=CC(=O)CC	167	169
87	C ₈ H ₁₄ O*	CC=C(C)C(C)C(C)=O	154	157
88	C ₈ H ₁₄ O^	CCC(C)=C(C)C(C)=O	158	174
89	C ₈ H ₁₄ O ₂ *	CCCC(=O)C(=O)CCC	168	165
90	C ₈ H ₁₆ O^	CCCCCCCC=O	171	170
91	C ₈ H ₁₆ O*	CCCCCCC(C)=O	173	167
92	C ₈ H ₁₆ O^	CCCCCC(C)C=O	162	165
93	C ₈ H ₁₆ O*	CCCCCC(=O)CC	168	167
94	C ₈ H ₁₆ O*	CC(C)CCCC(=O)C	170	161
95	C ₈ H ₁₆ O*	CCCC(=O)CCCC	165	163
96	C ₈ H ₁₆ O^	CC(=O)CCC(C)CC	167	162
97	C ₈ H ₁₆ O*	CC(C)CCC(=O)CC	164	160
98	C ₈ H ₁₆ O^	CCCC(CC)C=O	163	162
99	C ₈ H ₁₆ O*	CCCC(C)CC(=O)C	156	161
100	C ₈ H ₁₆ O^	CC(C)CC(=O)CCC	156	159
101	C ₈ H ₁₆ O^	CC(C)C(=O)CCCC	160	159
102	C ₈ H ₁₆ O*	CCCC(CCC)C=O	160	156
103	C ₈ H ₁₆ O^	CCC(C)CC(=O)CC	161	158
104	C ₈ H ₁₆ O^	CCCC(=O)C(C)CC	154	157
105	C ₈ H ₁₆ O^	CC(C)CC(=O)C(C)C	147	151
106	C ₈ H ₁₆ O*	CCC(=O)C(C)(C)CC	151	147
107	C ₈ H ₁₆ O*	CC(C)C(=O)C(C)(C)C	135	142
108	C ₉ H ₁₆ O*	CCC=CC(=O)CC(C)C	183	186
109	C ₉ H ₁₈ O^	CCCCCCCC=O	191	187
110	C ₉ H ₁₈ O^	CC(=O)CCCCCCC	195	186
111	C ₉ H ₁₈ O^	CCC(=O)CCCCCC	190	186

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [28]
112	C ₉ H ₁₈ O*	CCCC(=O)CCCC	187	183
113	C ₉ H ₁₈ O*	CCCCC(=O)CCCC	186	183
114	C ₉ H ₁₈ O^	CC(C)CCCC(=O)CC	183	182
115	C ₉ H ₁₈ O*	CCCC(=O)CCC(C)C	178	179
116	C ₉ H ₁₈ O^	CCCCC(=O)C(C)C	183	179
117	C ₉ H ₁₈ O^	CCCC(C)CC(=O)C	183	182
118	C ₉ H ₁₈ O*	CCCC(C)CC(=O)CC	179	178
119	C ₉ H ₁₈ O*	CCC(C)CC(=O)CCC	161	179
120	C ₉ H ₁₈ O^	CCCCC(=O)C(C)CC	174	179
121	C ₉ H ₁₈ O*	CC(C)CC(=O)CC(C)C	165	168
122	C ₉ H ₁₈ O*	CCC(C)C(=O)C(C)CC	162	168
123	C ₉ H ₁₈ O*	CC(C)(C)C(=O)C(C)(C)C	152	149
124	C ₁₀ H ₂₀ O*	CCCCCCCCC=O	208	204
125	C ₁₀ H ₂₀ O*	CC(=O)CCCCCCCC	210	203
126	C ₁₀ H ₂₀ O^	CCC(=O)CCCCCCC	211	203
127	C ₁₀ H ₂₀ O*	CCCC(=O)CCCCC	207	201
128	C ₂ H ₄ O ₂ *	COC=O	32	30
129	C ₃ H ₆ O ₂ *	CCOC=O	54	55
130	C ₃ H ₆ O ₂ ^	COC(C)=O	58	53
131	C ₄ H ₆ O ₂ *	C=CCOC=O	83	78
132	C ₄ H ₆ O ₂ *	CC(=O)OC=C	72	78
133	C ₄ H ₆ O ₂ ^	COC(=O)C=C	85	79
134	C ₄ H ₈ O ₂ *	CCCCOC=O	81	79
135	C ₄ H ₈ O ₂ ^	CC(C)OC=O	68	76
136	C ₄ H ₈ O ₂ ^	CCOC(C)=O	77	77
137	C ₄ H ₈ O ₂ *	CCC(=O)OC	80	77
138	C ₅ H ₈ O ₂ *	CC(=O)OCC=C	104	99
139	C ₅ H ₈ O ₂ *	CCOC(=O)C=C	101	99
140	C ₅ H ₈ O ₂ ^	COC(=O)C=CC	120	118
141	C ₅ H ₈ O ₂ ^	COC(=O)C(C)=C	101	97
142	C ₅ H ₁₀ O ₂ ^	CCCCOC=O	107	106
143	C ₅ H ₁₀ O ₂ ^	CC(C)COC=O	98	100
144	C ₅ H ₁₀ O ₂ *	CCOC(C)=O	102	101
145	C ₅ H ₁₀ O ₂ *	CCC(C)OC=O	97	102
146	C ₅ H ₁₀ O ₂ *	CCCC(=O)OC	102	103
147	C ₅ H ₁₀ O ₂ ^	CCOC(=O)CC	99	100
148	C ₅ H ₁₀ O ₂ *	CC(C)OC(C)=O	90	97
149	C ₆ H ₁₀ O ₂ ^	CC(=O)OCCC=C	127	118
150	C ₆ H ₁₀ O ₂ *	CCOC(=O)C=C	123	118
151	C ₆ H ₁₀ O ₂ *	CCOC(=O)CC=C	119	118
152	C ₆ H ₁₀ O ₂ *	CCOC(=O)C=CC	136	134
153	C ₆ H ₁₀ O ₂ ^	CC(OC(C)=O)C=C	113	119
154	C ₆ H ₁₀ O ₂ *	CC(C)OC(=O)C=C	110	117
155	C ₆ H ₁₀ O ₂ *	CCOC(=O)C(C)=C	118	117
156	C ₆ H ₁₂ O ₂ *	CCCCCOC=O	132	131

(cont.)

(Table 1 cont.)

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [28]
157	C ₆ H ₁₂ O ₂ [^]	CCCCOC(C)=O	127	126
158	C ₆ H ₁₂ O ₂ [*]	CCCCC(=O)OC	130	129
159	C ₆ H ₁₂ O ₂ [^]	CCOC(=O)CC	122	122
160	C ₆ H ₁₂ O ₂ [*]	CC(C)COC(C)=O	117	121
161	C ₆ H ₁₂ O ₂ [^]	CCCC(=O)OCC	120	125
162	C ₆ H ₁₂ O ₂ [^]	CCC(C)OC(C)=O	112	122
163	C ₆ H ₁₂ O ₂ [*]	CCC(=O)OC(C)C	110	120
164	C ₆ H ₁₂ O ₂ [*]	CC(=O)OC(C)(C)C	98	110
165	C ₇ H ₁₀ O ₂ [*]	CC=CC(=O)OC(C)=C	137	146
166	C ₇ H ₁₂ O ₂ [^]	CC(=O)OCCCC=C	151	140
167	C ₇ H ₁₂ O ₂ [*]	CCOC(=O)CCC=C	145	139
168	C ₇ H ₁₂ O ₂ [^]	CCCCC(=O)OC=C	133	143
169	C ₇ H ₁₂ O ₂ [*]	CC(C)COC(=O)C=C	132	136
170	C ₇ H ₁₂ O ₂ [*]	CC(CC=C)OC(C)=O	134	137
171	C ₇ H ₁₂ O ₂ [*]	CCOC(=O)C=CCC	158	157
172	C ₇ H ₁₂ O ₂ [^]	CC=CC(=O)OC(C)C	146	150
173	C ₇ H ₁₄ O ₂ [*]	CCCCCOC=O	156	153
174	C ₇ H ₁₄ O ₂ [^]	CCCCCOC(C)=O	149	150
175	C ₇ H ₁₄ O ₂ [*]	CCCCC(=O)OC	151	151
176	C ₇ H ₁₄ O ₂ [^]	CCCCOC(=O)CC	146	147
177	C ₇ H ₁₄ O ₂ [^]	CCCCC(=O)OCC	145	148
178	C ₇ H ₁₄ O ₂ [^]	CCOC(=O)CCC	143	146
179	C ₇ H ₁₄ O ₂ [*]	CCC(=O)OCC(C)C	137	141
180	C ₇ H ₁₄ O ₂ [*]	CCCC(=O)OC(C)C	131	143
181	C ₇ H ₁₄ O ₂ [^]	CCC(C)OC(=O)CC	132	142
182	C ₇ H ₁₄ O ₂ [*]	CC(C)C(C)OC(C)=O	129	139
183	C ₇ H ₁₄ O ₂ [^]	CC(C)OC(=O)C(C)C	134	138
184	C ₈ H ₁₂ O ₂ [^]	CCOC(=O)C=CC=CC	196	191
185	C ₈ H ₁₄ O ₂ [*]	CC(C)C=CCOC(C)=O	172	176
186	C ₈ H ₁₄ O ₂ [*]	CCOC(=O)CC=CCC	167	171
187	C ₈ H ₁₄ O ₂ [^]	CCCC=CC(=O)OCC	175	178
188	C ₇ H ₁₀ O ₂ [*]	COC(=O)C=CC=CC	180	179
189	C ₈ H ₁₄ O ₂ [*]	CCC(OC(C)=O)C=CC	156	165
190	C ₈ H ₁₄ O ₂ [*]	CC(=O)OC(C)(C)CC=C	138	150
191	C ₈ H ₁₆ O ₂ [^]	CCCCCOC(C)=O	169	170
192	C ₈ H ₁₆ O ₂ [*]	CCCCCCC(=O)OC	172	170
193	C ₈ H ₁₆ O ₂ [^]	CCCCCOC(=O)CC	169	169
194	C ₈ H ₁₆ O ₂ [*]	CCCCC(=O)OCC	167	168
195	C ₈ H ₁₆ O ₂ [*]	CCCCOC(=O)CCC	165	166
196	C ₈ H ₁₆ O ₂ [^]	CCCCC(=O)OCCC	168	168
197	C ₈ H ₁₆ O ₂ [*]	CCCC(=O)OCC(C)C	157	162
198	C ₈ H ₁₆ O ₂ [*]	CC(C)COC(=O)C(C)C	147	158
199	C ₈ H ₁₆ O ₂ [*]	CCC(C)C(=O)OC(C)C	144	158
200	C ₉ H ₁₈ O ₂ [*]	CCCCCCCCOC=O	178	190

* Member of the training set

^ Member of the test set

Among physicochemical experimentally determined properties, the boiling points at normal pressure (*i.e.* nbp's) are probably the most widely available and precisely measured data for lower molecular weight substances. For compounds with larger molecules, boiling points at lower pressures are also available, but since the data are available for different pressures, comparisons cannot be made. Although the main objective for developing new molecular descriptors is to use them for drug design and for being able to predict biological activities to propose valid topological indices, one must rely on intramolecular comparisons for the novel local vertex invariants and on intermolecular comparisons for the resulting molecular descriptors. These comparisons and the corresponding correlations should be made for a molecular property that is measured with sufficient accuracy. The nbp's fulfill this expectation best [7-10, 23, 30-38]. In addition, nbp's are suitable because they are not practically identical for isomeric compounds with the same composition like many thermodynamic properties. Although nbp's usually increase with increasing molecular weight, they also depend on constitution and differ significantly among isomers.

Regarding the way to work with the original molecular set there are various options. We have chosen what we have considered the more rational one. In fact, we have divided the whole set into two subsets: a training set and a test set. The first one was employed to determine coefficients in the regression equations while the second one served for true predictive purposes, *i.e.* the equations derived with the help of the molecules composing the training set were applied to the tests set to derive nbp's. The training set is composed of 121 molecules while the remaining 79 are included into the test set (see Table 1). Naturally, there are many ways to select the components of each set. However, in this case the particular possible choices do not influence significantly the results. This feature is significant since this cross-validation method is a further verification fit of the regression models.

Regarding the analytical form of function *f* in equation 3, we have selected the simplest ones: first-, second- and third-order polynomials, *i.e.*

$$\text{nbp} = b(\text{DCW}) + a \quad \text{[IV]}$$

$$\text{nbp} = c(\text{DCW})^2 + b(\text{DCW}) + a \quad \text{[V]}$$

$$\text{nbp} = d(\text{DCW})^3 + c(\text{DCW})^2 + b(\text{DCW}) + a \quad \text{[VI]}$$

Clearly, simple regression involving but a single descriptor restricts regression analysis considerably. Many, correlations, particularly when involving molecules of different size, need not be linear. However, even if we have molecules of the same or similar size, a quadratic or/and cubic regression may result in a better description of the relationship between property and molecular descriptor than a simple linear model. In general, one should test single descriptor regression for quadratic dependence and, if warranted, for higher polynomial relationships or other functional dependence [16].

Table 2 lists local invariants and their corresponding correlation weights, which have been obtained by means of the Monte Carlo optimization procedure described previously [27].

The fitting polynomials obtained from the training set are the following:

$$\begin{aligned} \text{nbp} &= 3.7871 (\pm 0.0597) (\text{DCW}) - 15.6393 (\pm 2.3851) & \text{[VII]} \\ r &= 0.9855, \text{ SE} = 6.5175 \end{aligned}$$

$$\begin{aligned} \text{nbp} &= -0.0230 (\pm 0.0043) (\text{DCW})^2 + 5.5248 (\pm 0.3322) (\text{DCW}) - 46.1488 (\pm 6.1437) & \text{[VIII]} \\ r &= 0.9883, \text{ SE} = 5.8820 \end{aligned}$$

$$\text{nbp} = 0.0006 (\pm 0.0003) (\text{DCW})^3 - 0.0888 (\pm 0.0372) (\text{DCW})^2 + 7.8578 (\pm 1.3494) (\text{DCW}) - 71.6714 (\pm 15.5567) \quad [\text{IX}]$$

$$r = 0.9887, \text{SE} = 5.8285$$

where r stands for the regression coefficient and SE is the standard error of the estimate.

TABLE 2. Correlation weights of local graph invariants.

Atom(A)	CW(A)	NNC	CW(NNC)
ee	0.3881	0100	0.3658
H	1.6433	0110	0.1612
C	0.4739	0211	0.8799
O	0.5087	0220	8.7157
		0301	0.0682
		0310	0.9060
		0311	2.8897
		0312	0.2032
		0320	3.9356
		0321	5.2943
		0330	5.8669
		0403	2.0830
		0412	2.6391
		0413	-0.0001
		0420	-0.0761
		0421	1.4050
		0422	1.6856
		0430	-0.8475
		0431	0.9551
		0440	-0.0702

Table 3 displays the results for the calculation of nbp derived from equations. 7 - 9, together with deviations as well as the deviations obtained from the theoretical calculation inserted in Table 2.

TABLE 3. Theoretical results of the nbp of 200 carbonyl compounds (deviations are given between parenthesis). * Member of the training set. ^ Member of the test set. AAD: Average absolute deviation.

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
1 *	37 (-16)	27 (-6)	22 (-1)	45 (-24)
2 *	62 (-6)	57 (-1)	57 (-1)	61 (-5)
3 *	60 (-7)	55 (-2)	54 (-1)	73 (-20)
4 *	59 (-11)	54 (-6)	53 (-5)	54 (-6)
5 *	57 (-1)	51 (5)	50 (6)	44 (12)
6 *	107 (0)	109 (-2)	109 (-2)	104 (3)
7 *	81 (3)	80 (4)	81 (3)	67 (17)

(Table 3 cont.)

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
8 ^	95 (10)	95 (10)	97 (8)	110(5)
9 *	78 (-10)	76 (-8)	77 (-9)	83 (-15)
10 *	79 (2)	78 (3)	79 (2)	83 (-2)
11 *	81 (-6)	80 (-5)	81 (-6)	79 (-4)
12 *	72 (-8)	69 (-5)	70 (-6)	71 (-7)
13 ^	79 (1)	77 (3)	78 (2)	73 (7)
14 ^	96 (-5)	97 (-6)	99 (-8)	84 (7)
15 *	103 (3)	104 (2)	105 (1)	82 (24)
16 *	104 (-5)	105 (-6)	106 (-7)	100 (-1)
17 *	117 (-11)	119 (-13)	119 (-13)	118 (-12)
18 *	117 (7)	119 (5)	119 (5)	125 (-1)
19 ^	101 (2)	102 (1)	104 (-1)	97 (6)
20 *	114 (8)	116 (6)	117 (5)	122 (0)
21 *	112 (21)	114 (19)	115 (18)	115 (18)
22 *	112 (5)	114 (3)	115 (2)	114 (3)
23 *	97 (1)	98 (0)	99 (-1)	97 (1)
24 *	106 (2)	108 (0)	109 (-1)	115 (-7)
25 *	103 (0)	104 (-1)	105 (-2)	107 (-4)
26 ^	94 (-1)	94 (-1)	96 (-3)	93 (0)
27 ^	100 (2)	102 (0)	103 (-1)	99 (3)
28 *	94 (-3)	94 (-3)	95 (-4)	96 (-5)
29 *	98 (4)	99 (3)	100 (2)	90 (12)
30 ^	91 (3)	91 (3)	93 (1)	91 (3)
31 *	83 (-8)	83 (-8)	84 (-9)	83 (-8)
32 *	152 (22)	153 (21)	153 (21)	172 (2)
33 ^	116 (2)	118 (0)	119 (-1)	98 (20)
34 *	135 (9)	137 (7)	137 (7)	136 (8)
35 *	138 (-17)	140 (-19)	140 (-19)	140 (-19)
36 ^	138 (8)	140 (6)	141 (5)	143 (3)
37 *	123 (7)	125 (5)	126 (4)	117 (13)
38 ^	136 (-9)	138 (-11)	139 (-12)	133 (-6)
39 ^	136 (4)	138 (2)	139 (1)	142 (-2)
40 ^	119 (5)	121 (3)	122 (2)	113 (11)
41 *	136 (3)	138 (1)	138 (1)	134 (5)
42 *	134 (2)	136 (0)	136 (0)	135 (1)
43 ^	119 (0)	121 (-2)	122 (-3)	114 (5)
44 ^	132 (-2)	134 (-4)	135 (-5)	132 (-2)
45 *	128 (2)	130 (0)	131 (-1)	124 (6)
46 *	125 (6)	127 (4)	127 (4)	129 (2)
47 *	116 (6)	118 (4)	118 (4)	119 (3)
48 ^	122 (5)	125 (3)	126 (1)	125 (2)
49 *	116 (6)	118 (4)	118 (4)	120 (2)
50 *	116 (0)	118 (-2)	118 (-2)	120 (-4)
51 *	122 (1)	124 (-1)	125 (-2)	121 (2)
52 ^	113 (4)	115 (2)	116 (1)	114 (3)

(cont.)

(Table 3 cont.)

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
53 ^	116 (2)	118 (0)	119 (-1)	112 (6)
54 *	106 (10)	108 (8)	109 (7)	116 (0)
55 ^	113 (5)	115 (3)	116 (2)	115 (3)
56 *	113 (2)	115 (0)	116 (-1)	113 (2)
57 ^	103 (3)	104 (2)	106 (0)	102 (4)
58 ^	160 (6)	161 (5)	162 (4)	162 (4)
59 *	160 (-3)	161 (-4)	160 (-3)	161 (-4)
60 *	158 (-5)	159 (-6)	158 (-5)	151 (2)
61 *	145 (-3)	147 (-5)	146 (-4)	139 (3)
62 *	141 (9)	143 (7)	142 (8)	135 (15)
63 ^	136 (2)	138 (0)	139 (-1)	132 (6)
64 ^	136 (1)	138 (-1)	139 (-2)	132 (5)
65 *	149 (-1)	150 (-2)	150 (-2)	148 (0)
66 *	147 (6)	148 (5)	148 (5)	150 (3)
67 ^	144 (7)	146 (5)	147 (4)	147 (4)
68 ^	137 (7)	140 (4)	140 (4)	144 (0)
69 *	144 (5)	146 (3)	145 (4)	146 (3)
70 *	135 (9)	137 (7)	137 (7)	143 (1)
71 ^	137 (7)	140 (4)	140 (4)	138 (6)
72 ^	144 (0)	146 (-2)	147 (-3)	139 (5)
73 ^	135 (5)	137 (3)	138 (2)	139 (1)
74 ^	135 (1)	137 (-1)	138 (-2)	137 (-1)
75 ^	135 (8)	137 (6)	138 (5)	138 (5)
76 *	135 (0)	137 (-2)	137 (-2)	136 (-1)
77 ^	135 (0)	137 (-2)	138 (-3)	134 (1)
78 *	126 (-2)	128 (-4)	128 (-4)	126 (-2)
79 *	125 (0)	127 (-2)	127 (-2)	125 (0)
80 *	207 (-9)	199 (-1)	201 (-3)	186 (12)
81 *	170 (-7)	170 (-7)	169 (-6)	174 (-11)
82 ^	170 (9)	170 (9)	171 (8)	174 (5)
83 *	170 (0)	170 (0)	169 (1)	162 (8)
84 ^	170 (-8)	170 (-8)	171 (-9)	162 (0)
85 *	170 (1)	170 (1)	169 (2)	164 (7)
86 *	175 (-8)	174 (-7)	173 (-6)	169 (-2)
87 *	166 (-12)	166 (-12)	165 (-11)	157 (-3)
88 ^	171 (-13)	170 (-12)	172 (-14)	174 (-16)
89 *	166 (2)	171 (-3)	170 (-2)	165 (3)
90 ^	168 (3)	168 (3)	169 (2)	170 (1)
91 *	166 (7)	166 (7)	165 (8)	167 (6)
92 ^	159 (3)	160 (2)	161 (1)	165 (-3)
93 *	166 (2)	166 (2)	165 (3)	167 (1)
94 *	157 (13)	158 (12)	157 (13)	161 (9)
95 *	166 (-1)	166 (-1)	165 (0)	163 (2)
96 ^	154 (13)	155 (12)	156 (11)	162 (5)
97 *	157 (7)	158 (6)	157 (7)	160 (4)

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
98 ^	159 (4)	160 (3)	161 (2)	162 (1)
99 *	157 (-1)	158 (-2)	157 (-1)	161 (-5)
100 ^	157 (-1)	158 (-2)	159 (-3)	159 (-3)
101 ^	157 (3)	158 (2)	159 (1)	159 (1)
102 *	159 (1)	160 (0)	159 (1)	156 (4)
103 ^	157 (4)	158 (3)	159 (2)	158 (3)
104 ^	157 (-3)	157 (-4)	159 (-5)	157 (-3)
105 ^	148 (-1)	149 (-2)	150 (-3)	151 (-4)
106 *	147 (4)	148 (3)	148 (3)	147 (4)
107 *	137 (-2)	140 (-5)	139 (-4)	142 (-7)
108 *	179 (4)	178 (5)	176 (7)	186 (-3)
109 ^	190 (1)	186 (5)	189 (2)	187 (4)
110 ^	188 (7)	184 (11)	187 (8)	186 (9)
111 ^	188 (2)	184 (6)	187 (3)	186 (4)
112 *	188 (-1)	184 (3)	184 (3)	183 (4)
113 *	188 (-2)	184 (2)	184 (2)	183 (3)
114 ^	179 (4)	177 (6)	179 (4)	182 (1)
115 *	179 (-1)	177 (1)	176 (2)	179 (-1)
116 ^	179 (4)	177 (6)	179 (4)	179 (4)
117 ^	179 (4)	177 (6)	179 (4)	182 (1)
118 *	179 (0)	177 (2)	176 (3)	178 (1)
119 *	179 (-18)	177 (-16)	176 (-15)	179 (-18)
120 ^	179 (-5)	177 (-3)	179 (-5)	179 (-5)
121 *	169 (4)	169 (-4)	168 (-3)	168 (-3)
122 *	169 (-7)	169 (-7)	168 (-6)	168 (-6)
123 *	149 (3)	151 (1)	150 (2)	149 (3)
124 *	212 (-4)	203 (5)	206 (2)	204 (4)
125 *	210 (0)	201 (9)	204 (6)	203 (7)
126 ^	210 (1)	201 (10)	208 (3)	203 (8)
127 *	210 (-3)	201 (6)	204 (3)	201 (6)
128 *	42 (-10)	32 (0)	29 (3)	30 (2)
129 *	59 (-5)	54 (0)	54 (0)	55 (-1)
130 ^	60 (-2)	56 (2)	55 (3)	53 (5)
131 *	82 (1)	81 (2)	82 (1)	78 (5)
132 *	66 (6)	62 (10)	62 (10)	78 (-6)
133 ^	83 (2)	82 (3)	83 (2)	79 (6)
134 *	75 (6)	73 (8)	74 (7)	79 (2)
135 ^	70 (-2)	67 (1)	68 (0)	76 (-8)
136 ^	78 (-1)	76 (1)	77 (0)	77 (0)
137 *	82 (-2)	81 (-1)	82 (-2)	77 (3)
138 *	101 (3)	102 (2)	103 (1)	99 (5)
139 *	101 (0)	102 (-1)	103 (-2)	99 (2)
140 ^	118 (2)	120 (0)	121 (-1)	118 (2)
141 ^	101 (0)	102 (-1)	103 (-2)	97 (4)
142 ^	103 (4)	104 (3)	106 (1)	106 (1)

(cont.)

(Table 3 cont.)

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
143 ^	94 (4)	94 (4)	96 (2)	100 (-2)
144 *	100 (2)	101 (1)	102 (0)	101 (1)
145 *	92 (5)	92 (5)	93 (4)	102 (-5)
146 *	104 (-2)	106 (-4)	106 (-4)	103 (-1)
147 ^	100 (-1)	101 (-2)	102 (-3)	100 (-1)
148 *	89 (1)	89 (1)	90 (0)	97 (-7)
149 ^	123 (4)	125 (2)	126 (1)	118 (9)
150 *	123 (0)	125 (-2)	125 (-2)	118 (5)
151 *	123 (-4)	125 (-6)	125 (-6)	118 (1)
152 *	135 (1)	138 (-2)	137 (-1)	134 (2)
153 ^	111 (2)	113 (0)	115 (-2)	119 (-6)
154 *	111 (-1)	113 (-3)	114 (-4)	117 (-7)
155 *	118 (0)	120 (-2)	121 (-3)	117 (1)
156 *	125 (7)	127 (5)	127 (5)	131 (1)
157 ^	122 (5)	124 (3)	125 (2)	126 (1)
158 *	126 (4)	128 (2)	128 (2)	129 (1)
159 ^	122 (0)	124 (-2)	125 (-3)	122 (0)
160 *	113 (4)	115 (2)	115 (2)	121 (-4)
161 ^	122 (-2)	124 (-4)	125 (-5)	125 (-5)
162 ^	111(1)	113 (-1)	114 (-2)	122 (-10)
163 *	111 (-1)	113 (-3)	113 (-3)	120 (-10)
164 *	96 (2)	96 (2)	97 (1)	110 (-12)
165 *	143 (-6)	145 (-8)	144 (-7)	146 (-9)
166 ^	144 (7)	146 (5)	147 (4)	140 (11)
167 *	144 (1)	146 (-1)	146 (-1)	139 (6)
168 ^	132 (1)	134 (-1)	135 (-2)	143 (-10)
169 *	135 (-3)	137 (-5)	137 (-5)	136 (-4)
170 *	133 (1)	136 (-2)	135 (-1)	137 (-3)
171 *	157 (1)	158 (0)	157 (1)	157 (1)
172 ^	146 (0)	148 (-2)	149 (-3)	150 (-4)
173 *	147 (9)	148 (8)	148 (8)	153 (3)
174 ^	144 (5)	145 (4)	146 (3)	150 (-1)
175 *	148 (3)	149 (2)	149 (2)	151 (0)
176 ^	144 (2)	145 (1)	146 (0)	147 (-1)
177^	144 (1)	145 (0)	146 (-1)	148 (-3)
178 ^	144 (-1)	145 (-2)	146 (-3)	146 (-3)
179 *	134 (3)	137 (0)	136 (1)	141 (-4)
180 *	133 (-2)	135 (-4)	135 (-4)	143 (-12)
181 ^	133 (-1)	135 (-3)	136 (-4)	142 (-10)
182 *	123 (6)	126 (3)	126 (3)	139 (-10)
183 ^	123 (11)	126 (8)	127 (7)	138 (-4)
184 ^	193 (3)	188 (8)	192 (4)	191 (5)
185 *	170 (2)	169 (3)	169 (3)	176 (-4)
186 *	179 (-12)	177 (-10)	177 (-10)	171 (-4)
187 ^	179 (-4)	177 (-2)	179 (-4)	178 (-3)

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
188 *	175 (5)	174 (6)	173 (7)	179 (1)
189 *	168 (-12)	168 (-12)	167 (-11)	165 (-9)
190 *	140 (-2)	142 (-4)	142 (-4)	150 (-12)
191 ^	165 (4)	165 (4)	167 (2)	170 (-1)
192 *	170 (2)	169 (3)	168 (4)	170 (2)
193 ^	165 (4)	165 (4)	167 (2)	169 (0)
194 *	165 (2)	165 (2)	165 (2)	168 (-1)
195 *	165 (0)	165 (0)	165 (0)	166 (-1)
196 ^	165 (3)	165 (3)	167 (1)	168 (0)
197 *	156 (1)	157 (0)	156 (1)	162 (-5)
198 *	147 (0)	149 (-2)	148 (-1)	158 (-11)
199 *	145 (-1)	147 (-3)	146 (-2)	158 (-14)
200 *	190 (-12)	186 (-8)	187 (-9)	190 (-12)
AAD (training set)	4.67	4.30	4.22	-
AAD (test set)	3.67	3.58	3.23	-
AAD (complete set)	4.25	4.03	3.78	5.00

The analysis of results depicted in Table 3 shows that the use of flexible descriptors affords better predictions than those derived from fixed descriptors (*i.e.* compare absolute average deviations given in the last three rows). In order to judge suitably the goodness degree of our results, one must take into account that present predictions are determined based on just one descriptor, while previous theoretical results [28] were computed with the use of five descriptors. Besides, the employment of second and third order polynomials gives better results than the employment of linear equations. This finding illustrates the need to resort to higher order polynomial relationships in order to improve results. This fact is in line with several previous analyses [39-46]. Particularly noteworthy are the results obtained from the members of the test set, which are true predictions. In fact, it is rather surprising that deviations for this set are lower than those computed for the training set, when the usual is the opposite behavior. Probably, this peculiar behavior is due to chance reasons. Another feature that shows the better quality of our results are the corresponding maximum absolute deviations, which are 24°C and 21°C for previous calculations [28] and present computations via equation 9, respectively.

CONCLUSIONS

We have shown that using flexible enough topological descriptors one can obtain good enough predictions of physical chemistry molecular properties. In particular, results improve remarkably when comparing results derived from fixed descriptors. The iterative numerical procedure needed to produce the smallest standard error for the property considered is simple enough and the improvements obtained in the regression statistics when using them warrants its use. It is especially important to point out the fact that the employment of variable molecular descriptors makes it unnecessary to resort the application of several variables fitting equations, as it is the case when using fixed variable molecular descriptors.

It remains to be seen how well this approximation works for predicting other physical chemistry properties and/or other molecular species. Work along these lines is under development in our laboratories and results will be presented elsewhere in the forthcoming future.

ACKNOWLEDGMENT

Authors thank very much some valuable suggestions pointed out by two referees, which have been useful to improve the final version of this paper.

REFERENCES

- [1] Rechsteiner, C.E. In *Handbook of Chemical Property Estimation Methods*. Lyman, W.J.; Reehl, W.F.; Rosenblatt, D.H. Eds. McGraw-Hill, New York, **1982**, Chapter 12.
- [2] Fisher, C. H. *Chem. Eng.* **1989**, *96*, 157
- [3] Satyanarayana, K.; Kakati, M.C. *Fire Mater.* **1991**, *15*, 97
- [4] Walker, J. J. *Chem. Soc.* **1894**, *65*, 193
- [5] Meissner, H.P. *Chem. Eng. Prog.* **1949**, *45*, 149
- [6] Horvath, A.L. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*. Elsevier, Amsterdam, **1992**, Chapter 2
- [7] Katritzky, A.R.; Lobanov, W. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, 279.
- [8] Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027.
- [9] Katritzky, A. R.; Mu, L.; Lobanov, V.S.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400.
- [10] Katritzky, A.R.; Lobanov, V.S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28.
- [11] Cocchi, M.; De Benedetti, P.G.; Seeber, R.; Tassi, L.; Ulrici, A. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1190.
- [12] Stanton, D.T.; Jurs, P.C.; Hicks, M.G. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301.
- [13] Stanton, D.T.; Engolf, L.M.; Jurs, P.C.; Hicks, M.G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306.
- [14] Randic, M.; Trinajstic, N. *J. Mol. Struct.* **1993**, *300*, 551.
- [15] Randic, M. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 233.
- [16] Randic, M.; Basak, S.C. *Variable Molecular Descriptors*. In *Some Aspects of Mathematical Chemistry*. Sinha, D.K.; Basak, S.C.; Mohany, R.K., Busa, I.N.; Mallick, Eds., Visva-Bharati University Press, Santiniketan, India, **2000**.
- [17] Randic, M. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213.
- [18] Randic, M. *J. Comput. Chem.* **1991**, *12*, 970.
- [19] Randic, M. *New J. Chem.* **2000**, *24*, 165.
- [20] Randic, M.; Dobrowolski, J.Cz. *Int. J. Quantum Chem.* **1998**, *70*, 1209.
- [21] Peruzzo, P.; Marino, D.; Castro, E.A.; Toropov, A.A. *J. Mol. Struct. THEOCHEM* **2001**, *572(1-3)*, 53.
- [22] Mercader, A.; Castro, E.A.; Toropov, A.A. *J. Mol. Model.* **2001**, *7(1-3)*, 1391.
- [23] Krenkel, G.; Castro, E.A.; Toropov, A.A. *J. Mol. Struct. THEOCHEM* **2001**, *542(1-3)*, 107.
- [24] Mercader, A.; Castro, E.A.; Toropov, A.A. *Chem. Phys. Lett.* **2000**, *330*, 612.
- [25] Krenkel, G.; Castro, E.A.; Toropov, A.A. *Int. J. Mol. Sci.* **2001**, *2*, 57.
- [26] Todeschini, R. Consonni, V. *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, **2000**.
- [27] Toropov, A.A.; Toropova, A.P. *Russ. J. Coord. Chem.* **1998**, *24(2)*, 89.
- [28] Balaban, A.T.; Mills, D.; Basak, S. C. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 758 Errata for this paper in *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1477.
- [29] Weininger, D.; Weininger, J.L. In *Comprehensive Medicinal Chemistry*, Hansch, C. Ed., Pergamon Press, Oxford, UK, **1990**, Vol.4, pp.59.
- [30] Engolf, L.M.; Wessel, M.D.; Jurs, P.C. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947.
- [31] Turner, B.E.; Costello, C.L.; Jurs, P.C. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 639.
- [32] Stanton, D.T.; Jurs, P.C. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301.
- [33] Murray, J.S.; Lane, P.; Politzer, P. *J. Mol. Struct. THEOCHEM* **1995**, *342*, 15.
- [34] Rücker, G.; Rücker, C. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 710.
- [35] Lukovits, I; Linert, W. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 715.

- [36] Trinajstić, N.; Nikolić, S.; Lucić, B. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 631.
- [37] Firpo, F.; Gavernet, L.; Castro, E.A.; Toropov, A.A. *J. Mol. Struct. THEOCHEM* **2000**, 501-502, 419.
- [38] Katritzky, A.R.; Gordeeva, E.V. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 758.
- [39] Castro, E.A. *Comp. Chem.* **1997**, *21(5)*, 305.
- [40] Firpo, M.; Gavernet, L.; Castro, E.A. *Polish J. Chem.* **1999**, *73*, 1041.
- [41] Firpo, M.; Gavernet, L.; Castro, E.A. *Rom. J. Phys.* **1999**, *44(1,2)*, 181.
- [42] Romanelli, G.P.; Cafferata, L.F.R.; Castro, E.A. *J. Mol. Struct. THEOCHEM* **2000**, *504*, 261.
- [43] Duchowicz, P.; Castro, E.A. *Acta Chem. Slov.* **2000**, *47*, 281.
- [44] Romanelli, G.P.; Cafferata, L.F.R. *Khim. Fizika* **2000**, *19(9)*, 105.
- [45] Duchowicz, P.; Castro, E.A. *J. Korean Chem. Soc.* **2000**, *44(6)*, 501.
- [46] Duchowicz, P.; Castro, E.A. *Arkivoc* **2001**, *2(6)*, U39.