– The Journal of the Argentine Chemical Society - Vol. 90 - N $^{\circ}$ 1/3 - 91/107 (2002) —

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 carbonilicos 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 interand 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 renascence 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

$$NNC_{\nu} = 100(N) + 10(N_{\nu}) + N_{\mu}$$
[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 CH₃CH=CH-CH=O, where Smiles notation is used and ee denotes lone pairs.



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

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

$$Property = f(DCW)$$
[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].

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	C₃H₄O*	C=CC=O	53	73
4	C₃H ₆ O*	CCC=O	48	54
5	C₃H ₆ O*	CC(=O)C	56	44
6	C ₄ H ₄ O*	CC#CC=O	107	104
7	C_4H_4O*	CC(=O)C#C	84	67
8	C₄H ₆ O^	CC=CC=O	105	110
9	C₄H ₆ O*	CC(=C)C=O	68	83
10	C ₄ H ₆ O*	CC(=O)C=C	81	83
11	C₄H ₈ O*	CCCC=O	75	79
12	C₄H8O*	CC(C)C=O	64	71
13	C₄H ₈ O^	CC(=O)CC	80	73
14	C₅H ₆ O^	C#CC(C)C=O	91	84
15	C₅H ₆ O*	CCC(=O)C#C	106	82
16	C5H8O*	C=CCCC=O	99	100
17	C ₅ H ₈ O*	CC=CCC=O	106	118
18	C₅H ₈ O*	CCC=CC=O	124	125

TABLE 1. Data for 200 carbonyl compounds.

94

19 $C_{ch}0^{ch}$ $CCCC(ch)C(c=0$ 113 114 21 $C_{ch}0^{ch}$ $CCCC(ch)C=0$ 113 115 22 $C_{ch}0^{ch}$ $CCCC(ch)C=0$ 133 115 23 $C_{ch}0^{ch}$ $CCC(cc)C=0$ 133 115 24 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 133 115 25 $C_{ch}0^{ch}$ $CCC(cc)C=0$ 133 115 27 $C_{ch}0^{ch}$ $CCC(cc)C=0$ 133 115 28 $C_{ch}0^{ch}$ $CCC(cc)C=0$ 133 136 29 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 134 135 29 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 134 136 31 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 136 131 32 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 136 133 33 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 136 133 33 $C_{ch}0^{ch}$ $CCCC(cc)C=0$ 136 134 41	Number	Formula	SMILES	Do) (oC)	nbp (theor.) [2 ₈]
$ \begin{array}{ccccc} 20 & C_{11} C_{12} & CC_{12} C_{12} & CC_{12} C_{12} & CC_{12} C_{12} & CC_{12} & CC$				103	<u> </u>
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	20	C ₅ H ₈ O*	CC=CC(C)=O	122	122
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	21	C ₅ H ₈ O*	CC(C)=CC=O	133	115
$ \begin{array}{ccccccc} 23 & \mathbb{C}_{9} \mathbb{N}_{0}^{\circ} & \mathbb{C}_{1} \mathbb{C}\mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb{C}_{1} \mathbb$	22	C ₅ H ₈ O*	CC=C(C)C=O	117	114
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	23	C ₅ H ₈ O*	CC(=C)C(C)=O	98	97
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	24	C ₅ H ₈ O ₂ *	CCC(=0)C(C)=0	108	115
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	25	C ₅ H ₁₀ O*	0=0000	103	107
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	26	C ₅ H ₁₀ O^	CC(C)CC=O	63	63
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	27	$C_5H_{10}OA$	CC(=0)CCC	102	66
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	28	C ₅ H ₁₀ O*	CCC(C)=O	91	96
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	29	C5H10O*	CCC(=0)CC	102	06
31 $G_{c}H_{0}O^{*}$ $CCG(G)(C,C=0$ 75 83 32 $G_{c}H_{0}O^{*}$ $CCC(C)(C)(C=0$ 174 172 33 $G_{c}H_{0}O^{*}$ $CCC(C)(C)(C=0$ 174 173 35 $G_{c}H_{10}O^{*}$ $CCC(C)(C)(C=0$ 144 136 37 $G_{c}H_{10}O^{*}$ $CCCC(C)(C)(C=0$ 121 143 38 $G_{c}H_{10}O^{*}$ $CCCCC(C)(C)(C=0$ 124 113 40 $CC=CCCC=0$ 124 133 113 41 $G_{c}H_{10}O^{*}$ $CCCCCC(C)=0$ 124 133 43 $G_{c}H_{10}O^{*}$ $CCCCCC(C)=0$ 124 133 44 $G_{c}H_{10}O^{*}$ $CCCCCCC(C)=0$ 124 134 45 $G_{c}H_{10}O^{*}$ $CCCCCCC(C)=0$ 133 124 46 $G_{c}H_{12}O^{*}$ $CCCC(C)C)=0$ 134 135 47 $G_{c}H_{10}O^{*}$ $CCCCCCC)=0$ 134 124 48 $G_{c}H_{12}O^{*}$ $CCCCCCC)=0$ 124 <t< td=""><td>30</td><td>C₅H₁₀O^</td><td>CC(=0)C(C)C</td><td>94</td><td>91</td></t<>	30	C ₅ H ₁₀ O^	CC(=0)C(C)C	94	91
32 $C_{e}H_{0}O^{*}$ $CC=CC=CC=0$ 174 172 33 $C_{e}H_{0}O^{*}$ $CC=CC=CC=0$ 124 136 35 $C_{e}H_{0}O^{*}$ $CC=CC(C=C)$ 144 136 36 $C_{e}H_{10}O^{*}$ $CC=CC(C=C)$ 144 136 37 $C_{e}H_{10}O^{*}$ $CCC=CCC=C$ 130 117 38 $C_{e}H_{10}O^{*}$ $CCC=CCC(C=C)$ 134 133 41 $C_{e}H_{10}O^{*}$ $CCC=CC(C)=O$ 124 113 42 $C_{e}H_{10}O^{*}$ $CCC=CC(C)=O$ 133 133 43 $C_{e}H_{10}O^{*}$ $CCC=CC(C)=O$ 134 113 44 $C_{e}H_{10}O^{*}$ $CCC=CC(C)=O$ 136 133 45 $C_{e}H_{10}O^{*}$ $CCC(=C)C(C)=O$ 136 132 46 $C_{e}H_{10}O^{*}$ $CCC(=C)C(C)=O$ 136 132 47 $C_{e}H_{10}O^{*}$ $CCC(=C)C(C)=O$ 136 132 48 $C_{e}H_{10}O^{*}$ $CCC(C)CCC=O$ 136	31	C5H10O*	CC(C)(C)C=O	75	83
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	32	C ₆ H ₈ O*	CC=CC=CC=O	174	172
34 $C_6H_0O^*$ $CCC=C(C=O)$ 144 136 35 $C_6H_1O^*$ $CCC=C(C=O)$ 144 133 37 $C_6H_1O^*$ $CCC=C(C=O)$ 146 143 38 $C_6H_1O^*$ $CCC=C(C=O)$ 146 143 38 $C_6H_1O^*$ $CCC=C(C=O)$ 121 143 39 $C_6H_1O^*$ $CCC=C(C)=O$ 124 143 41 $C_6H_1O^*$ $CCC=C(C)=O$ 124 143 42 $C_6H_1O^*$ $CCC=C(C)=O$ 139 113 43 $C_6H_1O^*$ $CCC=C(C)=O$ 139 113 44 $C_6H_1O^*$ $CCCC=C(C)=O$ 139 113 45 $C_6H_1O^*$ $CCCC=C(C)=O$ 139 124 47 $C_6H_1O^*$ $CCCCC=O$ 130 124 48 $C_6H_1O^*$ $CCCCCCCC=O$ 130 124 49 $C_6H_1O^*$ $CCCCCCCCC=O$ 130 124 51 $C_6H_1O^*$ $CCCCCCCCCC=O$	ŝ	C ₆ H ₈ O^	C#CC(=0)C(C)C	118	98
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	34	C ₆ H ₈ O*	CC=C(C=C)C=O	144	136
36 C_{eH_1O} CCCC=CC=O 146 143 37 C_{eH_1O} CCCC=CC(C)=O 146 143 38 C_{eH_1O} CCCC=CC(C)=O 127 133 39 C_{eH_1O} CCCC=CC(C)=O 127 133 40 C_{eH_1O} CCC=CC(C)=O 127 133 41 C_{eH_1O} CCC=CC(C)=O 124 113 42 C_{eH_1O} CCC=CC(C)=O 139 114 43 C_{eH_1O} CCC(=O(C)C)=O 136 133 44 C_{eH_1O} CCC(=O(C)C)=O 139 114 45 C_{eH_1O} CCC(=O(C)C)=O 130 123 46 C_{eH_1O} CCC(=O(C)C)=O 130 124 47 C_{eH_1O} CCC(=O(C)C)=O 130 124 48 C_{eH_1O} CCC(=O(C)C)=O 130 124 49 C_{eH_1O} CCC(C)C)=O 131 129 51 C_{eH_1O} CCC(C)C)=O	35	C ₆ H ₁₀ O*	0=000=000	121	140
37 $C_6H_{10}O^*$ $CCC(=O)CCC=C$ 130 117 38 $C_{6H_{10}O^*$ $CCC=CC(C)=O$ 122 133 40 $C_{6H_{10}O^*$ $CCC=CC(C)=O$ 124 113 41 $C_{6H_{10}O^*$ $CCC=CC(C)=O$ 139 133 42 $C_{6H_{10}O^*$ $CCC=C(C)C=O$ 139 134 43 $C_{6H_{10}O^*$ $CCC=C(C)C=O$ 136 135 44 $C_{6H_{10}O^*$ $CCC(=O)C(C)=O$ 136 135 45 $C_{6H_{10}O^*$ $CCC(=O)C(C)=O$ 136 135 47 $C_{6H_{12}O^*$ $CCC(=O)C(C)=O$ 136 135 48 $C_{6H_{12}O^*$ $CCC(C)C=O$ 136 114 49 $C_{6H_{12}O^*$ $CCCC(C)C=O$ 131 129 49 $C_{6H_{12}O^*$ $CCC(C)CC=O$ 131 129 51 $C_{6H_{12}O^*$ $CCC(C)CC=O$ 132 129 52 $C_{6H_{12}O^*$ $CCC(C)CCC=O$ 122 129	36	C ₆ H ₁₀ O^	0=00=000	146	143
38 $G_6 H_{10} O^A$ $CC = CCC(C) = O$ 127 133 39 $G_6 H_{10} O^A$ $CCC = CCC(C) = O$ 124 133 41 $G_6 H_{10} O^A$ $CCC = CC(C) = O$ 136 133 42 $G_6 H_{10} O^A$ $CCC = CC(C) = O$ 136 133 43 $G_6 H_{10} O^A$ $CCC = C(C) = O$ 136 133 44 $C_6 H_{10} O^A$ $CCC(= C(C) = O$ 136 134 45 $C_6 H_{10} O^A$ $CCC(= C(C) = O$ 136 135 47 $C_6 H_{10} O^A$ $CCC(= O) (C = C)$ 131 129 48 $C_6 H_{10} O^A$ $CCC(C) = O$ 131 129 49 $C_6 H_{10} O^A$ $CCC(C) = O$ 131 129 50 $C_6 H_{10} O^A$ $CCC(C) = O$ 131 129 51 $C_6 H_{10} O^A$ $CCC(C) = O$ 131 129 52 $C_6 H_{10} O^A$ $CCC(C) = O$ 122 119 53 $C_6 H_{10} O^A$ $CCC(C) = O$ 122	37	C ₆ H ₁₀ O*	CC(=0)CCC=C	130	117
39 $C_6H_{10}O^A$ $CCC=CC(C)=0$ 140 142 41 $C_6H_{10}O^A$ $CCC=CC(C)=0$ 134 113 42 $C_6H_{10}O^A$ $CCC=CC(C)=0$ 136 133 43 $C_6H_{10}O^A$ $CCCC=CC(C)=0$ 136 133 45 $C_6H_{10}O^A$ $CCC(=O)(C=CC)$ 139 133 45 $C_6H_{10}O^A$ $CCC(=O)(C=CC)$ 139 133 46 $C_6H_{12}O^A$ $CCC(=O)(C=CC)$ 131 129 47 $C_6H_{12}O^A$ $CCCC(C)=CO$ 131 129 48 $C_6H_{12}O^A$ $CCCC(C)C=O$ 131 129 51 $C_6H_{12}O^A$ $CCCC(C)C=O$ 131 129 53 $C_6H_{12}O^A$ $CCCC(C)C=O$ 131 120 53 $C_6H_{12}O^A$ $CCCC(C)C=O$ 131 123 54 $C_6H_{12}O^A$ $CCCC(C)C=O$ 132 121 55 $C_6H_{12}O^A$ $CCCC(C)C=O$ 122 123 55 <td>38</td> <td>C₆H₁₀O^</td> <td>CC=CCC(C)=O</td> <td>127</td> <td>133</td>	38	C ₆ H ₁₀ O^	CC=CCC(C)=O	127	133
40 CG(=C)CC(C)=0 124 113 41 $C_{6}H_{10}O^{*}$ $CC(=C)CC(C)=0$ 134 133 42 $C_{6}H_{10}O^{*}$ $CCC(=O)C=CC$ 139 134 43 $C_{6}H_{10}O^{*}$ $CCC(=O)C(=C)$ 139 134 45 $C_{6}H_{10}O^{*}$ $CCC(=O)C(=C)C(=)$ 130 124 46 $C_{6}H_{10}O^{*}$ $CCCC(C)=CO$ 130 124 47 $C_{6}H_{10}O^{*}$ $CCCC(=O)C(=O)C(=)CC$ 131 129 47 $C_{6}H_{12}O^{*}$ $CCCCCCCC=O$ 131 129 48 $C_{6}H_{12}O^{*}$ $CCCCCCCCCC$ 131 129 51 $C_{6}H_{12}O^{*}$ $CCCCCCCCCCC$ 127 127 52 $C_{6}H_{12}O^{*}$ $CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	39	C ₆ H ₁₀ O^	CCC=CC(C)=O	140	142
41 $C_{6}H_{10}O^*$ $CCC(=O)C=CC$ 139 134 42 $C_{6}H_{10}O^*$ $CCC(=O)C(C)=C$ 139 134 43 $C_{6}H_{10}O^*$ $CCC(=O)C(C)=C$ 136 137 45 $C_{6}H_{10}O^*$ $CCC(=O)C(C)=O$ 136 133 46 $C_{6}H_{12}O^*$ $CCC(=O)C(C)=O$ 131 122 47 $C_{6}H_{12}O^*$ $CCC(C)CC=O$ 131 129 48 $C_{6}H_{12}O^*$ $CCC(C)CC=O$ 131 129 49 $C_{6}H_{12}O^*$ $CCC(C)CC=O$ 122 119 50 $C_{6}H_{12}O^*$ $CCC(C)CC=O$ 122 129 129 51 $C_{6}H_{12}O^*$ $CCC(C)CC=O$ 122 129 120 51 $C_{6}H_{12}O^*$ $CCCC(C)CC=O$ 122 123 123 52 $C_{6}H_{12}O^*$ $CCCC(C)C=O$ 133 123 53 $C_{6}H_{12}O^*$ $CCC(C)C=O$ 116 112 54 $C_{6}H_{12}O^*$ $CCCCCCC=C$ 138 112 55 $C_{6}H_{12}O^*$ $CCCC$	40	C ₆ H ₁₀ O^	CC(=C)CC(C)=O	124	113
42 $G_6H_{10}O^*$ $CCC=C(C)=0$ 136 135 43 $G_6H_{10}O^*$ $CCC(=O(C)=C$ 139 114 45 $G_6H_{10}O^*$ $CCC(=O(C)=C$ 130 123 46 $G_6H_{12}O^*$ $CCC(=O)C(C)=C$ 130 124 47 $G_6H_{12}O^*$ $CCC(=O)C(C)=O$ 130 129 48 $G_6H_{12}O^*$ $CCC(C)=O$ 120 129 119 49 $C_6H_{12}O^*$ $CCC(C)=O$ 120 122 119 50 $C_6H_{12}O^*$ $CCC(C)C=O$ 121 123 123 51 $C_6H_{12}O^*$ $CCC(C)C=O$ 116 123 111 53 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 1112 54 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 1112 55 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 1112 55 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 1112 56 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 112 57 $C_6H_{12}O^*$ $CCC(C)C)C=O$ <t< td=""><td>41</td><td>C₆H₁₀O*</td><td>CCC(=0)C=CC</td><td>139</td><td>134</td></t<>	41	C ₆ H ₁₀ O*	CCC(=0)C=CC	139	134
43 $G_{6}H_{10}O^{A}$ $CCC(=O)C(C)=C$ 119 114 45 $G_{6}H_{10}O_{2}^{*}$ $CCC(=O)C(=O)$ 130 122 46 $G_{6}H_{10}O_{2}^{*}$ $CCC(C=O)$ 131 129 47 $G_{6}H_{10}O_{2}^{*}$ $CCC(C=O)$ 131 129 48 $G_{6}H_{12}O^{*}$ $CCC(CC=O)$ 121 129 49 $G_{6}H_{12}O^{*}$ $CCC(C)CC=O$ 122 125 50 $G_{6}H_{12}O^{*}$ $CCC(C)CC=O$ 122 126 51 $G_{6}H_{12}O^{*}$ $CCC(C)C=O$ 116 122 53 $G_{6}H_{12}O^{*}$ $CCC(C)C=O$ 116 123 121 53 $G_{6}H_{12}O^{*}$ $CCC(C)C=O$ 116 116 116 54 $G_{6}H_{12}O^{*}$ $CCC(C)C=O$ 118 1112 114 55 $G_{6}H_{12}O^{*}$ $CCC(C)C(C)C=O$ 117 114 112 55 $G_{6}H_{12}O^{*}$ $CCC(C)C(C)C=O$ 116 116 116 116 56 $G_{6}H_{12}O^{*}$ $CCC(C)C(C)C=O$ 118 11	42	C ₆ H ₁₀ O*	CCC=C(C)C=O	136	135
44 $C_{6}H_{10}O^{A}$ $CC(C)=CC(C)=0$ 130 132 45 $C_{6}H_{10}O_{2}^{*}$ $CCC(C=0$ 131 129 46 $C_{6}H_{10}O_{2}^{*}$ $CCC(C=0$ 131 129 47 $C_{6}H_{10}O_{2}^{*}$ $CCC(CC=0$ 131 129 48 $C_{6}H_{10}O_{2}^{*}$ $CCC(CC=0$ 122 119 49 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)CC=0$ 131 122 50 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)CC=0$ 131 122 51 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)CC=0$ 116 122 53 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)CC=0$ 116 123 53 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)C)CC=0$ 117 114 53 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)C)CC=0$ 116 123 54 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)C)C=0$ 116 123 55 $C_{6}H_{10}O_{2}^{*}$ $CCC(C)C)C=0$ 116 123 55 $C_{6}H_{10}O_{2}^{*}$ CCC	43	C ₆ H ₁₀ O^	CCC(=0)C(C)=C	119	114
45 $C_{6}H_{10}O_2^*$ $CCC(=0)C(=0)C(=)C(=)C(=)$ 120 129 46 $C_{6}H_{12}O^*$ $CCCCCC=0$ 131 129 48 $C_{6}H_{12}O^*$ $CCCCCC=0$ 131 129 49 $C_{6}H_{12}O^*$ $CCCCCC=0$ 127 129 50 $C_{6}H_{12}O^*$ $CCC(C)CC=0$ 127 129 51 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 127 129 53 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 116 121 53 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 118 114 53 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 118 112 54 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 118 113 55 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 156 116 55 $C_{6}H_{12}O^*$ $CCCCCC=C=0$ 156 116 56 $C_{7}H_{12}O^*$ $CCC(C)C=0$ 157 111 57 $C_{6}H_{12}O^*$ $CCCCCCCC=0$ 158 116	44	C ₆ H ₁₀ O^	CC(C)=CC(C)=O	130	132
46 $C_6H_{12}O^*$ $CCCCCC=O$ 131 129 47 $C_6H_{12}O^*$ $CCC(C)CC=O$ 131 129 48 $C_6H_{12}O^*$ $CCC(C)CC=O$ 122 129 49 $C_6H_{12}O^*$ $CCC(C)CC=O$ 122 120 50 $C_6H_{12}O^*$ $CCC(C)C=O$ 122 120 51 $C_6H_{12}O^*$ $CCC(C)C=O$ 123 121 53 $C_6H_{12}O^*$ $CCC(C)C=O$ 116 123 53 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 112 54 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 113 55 $C_6H_{12}O^*$ $CCC(C)C=O$ 118 113 55 $C_6H_{12}O^*$ $CCC(C)C=O$ 138 113 56 $C_7H_{12}O^*$ $CCCCCC=C$ 138 113 57 $C_6H_{12}O^*$ $CCC(C)C=O$ 157 113 58 $C_7H_{12}O^*$ $CCCCCC=CC=O$ 156 166 61 $C_7H_{12}O$	45	C ₆ H ₁₀ O ₂ *	CCC(=0)C(=0)CC	130	124
47 $C_{6}H_{12}O^{*}$ $CC(C)CCC=0$ 122 119 48 $C_{6}H_{12}O^{*}$ $CC(C)CCC=0$ 122 120 50 $C_{6}H_{12}O^{*}$ $CCC(C)CC=0$ 122 121 51 $C_{6}H_{12}O^{*}$ $CCC(C)CC=0$ 123 121 52 $C_{6}H_{12}O^{*}$ $CCC(C)CC=0$ 123 121 53 $C_{6}H_{12}O^{*}$ $CCC(C)C=0$ 116 121 54 $C_{6}H_{12}O^{*}$ $CCC(C)C=0$ 118 1114 55 $C_{6}H_{12}O^{*}$ $CCC(C)C=0$ 118 1113 55 $C_{6}H_{12}O^{*}$ $CCC(C)C(C=0)$ 116 116 57 $C_{6}H_{12}O^{*}$ $CCC(C)C(C=0)$ 118 1113 58 $C_{7}H_{12}O^{*}$ $CCCCCCCC=0$ 157 161 60 $C_{7}H_{12}O^{*}$ $CCCCCCCC=0$ 153 161 61 $C_{7}H_{12}O^{*}$ $CCCCCCCCC=0$ 153 151 62 $C_{7}H_{12}O^{*}$ $CCCCCCCCC=0$ 153 151 63 $C_{7}H_{12}O^{*}$ $CCCCCCCCC=0$ 153	46	C ₆ H ₁₂ O*	0=00000	131	129
48 $C_6H_1O^A$ $CC(=O)CCCC$ 127 125 49 $C_6H_1O^A$ $CCC(=O)CCCC$ 122 120 50 $C_6H_1O^A$ $CCCC(C)C=O$ 116 121 51 $C_6H_1O^A$ $CCCC(C)C=O$ 116 121 53 $C_6H_1O^A$ $CCC(C)C=O$ 117 114 53 $C_6H_1O^A$ $CCC(C)C=O$ 118 111 54 $C_6H_1O^A$ $CCC(C)C(=O)$ 118 111 55 $C_6H_1O^A$ $CCC(C)C(=O)$ 118 111 55 $C_6H_1O^A$ $CCC(C)C(=O)$ 116 116 56 $C_6H_1O^A$ $CCC(C)C(=O)$ 116 113 57 $C_6H_1O^A$ $CCC(C)C(=O)$ 116 113 58 $C_7H_1O^A$ $CCCCCCC=CC = O$ 156 161 60 $C_7H_1O^A$ $CCCCCCCCCCCCCC 157 161 61 C_7H_1O^A CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	47	C ₆ H ₁₂ O*	CC(C)CCC=0	122	119
49 $C_{6}H_{12}O^*$ $CCC(C)CC=0$ 122 120 50 $C_{6}H_{12}O^*$ $CCCC(C)C=0$ 116 121 51 $C_{6}H_{12}O^*$ $CCCC(C)C=0$ 116 121 53 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 117 114 53 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 118 112 54 $C_{6}H_{12}O^*$ $CCC(C)C=0$ 118 113 55 $C_{6}H_{12}O^*$ $CC(C)C(C)C=0$ 116 116 55 $C_{6}H_{12}O^*$ $CC(C)C(C)C=0$ 118 113 56 $C_{6}H_{12}O^*$ $CC(C)C(C)C=0$ 166 163 57 $C_{6}H_{12}O^*$ $CC(C)C(C)C=0$ 156 113 58 $C_{7}H_{12}O^*$ $CCCCCCCCC=0$ 157 161 60 $C_{7}H_{12}O^*$ $CCCCCCCCCCC=0$ 153 151 61 $C_{7}H_{12}O^*$ $CCCCCCCCCCCCCC 136 153 61 C_{7}H_{12}O^* CCCCCCCCCCCCCCCCCCCC 136 153 $	48	C ₆ H ₁₂ O^	CC(=0)CCCC	127	125
50 $C_6H_1O^*$ $CCCC(C)C=0$ 11612051 $C_6H_1O^*$ $CCCC(C)C=0$ 11711452 $C_6H_1O^*$ $CCC(=0)CC(C)$ 11711453 $C_6H_1O^*$ $CCC(=0)CC(C)C=0$ 11811254 $C_6H_1O^*$ $CCC(C)C=0$ 11811555 $C_6H_1O^*$ $CC(C)C(C)C=0$ 11611656 $C_6H_1O^*$ $CC(C)C(C)C=0$ 11811557 $C_6H_1O^*$ $CC(C)C(C)C=0$ 11611658 $C_7H_1O^*$ $CC(C)C(C)C=0$ 15716160 $C_7H_1O^*$ $CCCCC=CC=0$ 15716161 $C_7H_1O^*$ $CCCCCCCCCC=0$ 15315162 $C_7H_1O^*$ $CCCCCCCCCC=0$ 15315163 $C_7H_1O^*$ $CCCCCCCCC=0$ 13313264 $C_7H_1O^*$ $CC(CC(C)=0)CCC=C13713263C_7H_12O^*CC(CCC)=0)CCC13713264C_7H_12O^*CC(CCC)=0)CCC=C137132$	49	C ₆ H ₁₂ O*	CCC(C)CC=O	122	120
51 $C_6H_{12}O^*$ $CCC(=O)CCC$ 123 121 52 $C_6H_{12}O^*$ $CCC(=O)CC(C)C = 0$ 117 114 53 $C_6H_{12}O^*$ $CCC(=O)CC(C)C = 0$ 118 112 54 $C_6H_{12}O^*$ $CCC(C)C=0$ 116 116 55 $C_6H_{12}O^*$ $CC(C)C(C)C = 0$ 118 115 55 $C_6H_{12}O^*$ $CC(C)C(C)C = 0$ 116 115 56 $C_6H_{12}O^*$ $CC(C)C(C)C = 0$ 116 113 57 $C_6H_{12}O^*$ $CC(C)C(C)C = 0$ 116 113 58 $C_7H_{12}O^*$ $CCCCCCCCC = 0$ 157 161 60 $C_7H_{12}O^*$ $CCCCCCCCCC = 0$ 153 151 61 $C_7H_{12}O^*$ $CCCCCCCCCC = 0$ 153 151 62 $C_7H_{12}O^*$ $CCC(C)=0)CCCC = 0$ 153 151 63 $C_7H_{12}O^*$ $CCC(C)=0)CCCC = 0$ 153 132 63 $C_7H_{12}O^*$ $CCCCCCCCCC = 0$ 137 132 64 $C_7H_{12}O^*$ $CC(CCCCC)C = 0$ 137	20	C ₆ H ₁₂ O*	CCCC(C)C=O	116	120
52 $C_6H_{12}O^{A}$ $CC(=O)CC(C)C = 0$ 117 114 53 $C_6H_{12}O^{A}$ $CCC(C)C=0$ 118 112 54 $C_6H_{12}O^{A}$ $CCC(C)C=0$ 116 116 55 $C_6H_{12}O^{A}$ $CCC(C)C=0$ 118 115 55 $C_6H_{12}O^{A}$ $CCC(C)C=0$ 116 116 57 $C_6H_{12}O^{A}$ $CC(C)C(C)C=0$ 115 113 57 $C_6H_{12}O^{A}$ $CC(C)C(C)C=0$ 116 116 58 $C_7H_{12}O^{A}$ $CCCCCC=CC=0$ 166 162 59 $C_7H_{12}O^{A}$ $CCCCCCCCCC=0$ 153 161 60 $C_7H_{12}O^{A}$ $CCCCCCCCCC=0$ 153 151 61 $C_7H_{12}O^{A}$ $CCCCCCCCCC=0$ 153 151 63 $C_7H_{12}O^{A}$ $CC(CCCCC)=0$ 137 132 63 $C_7H_{12}O^{A}$ $CC(CCCC)=0$ 137 132 64 $C_7H_{12}O^{A}$ $CC(CCCC)=0$ 137 132	51	C ₆ H ₁₂ O*	CCC(=0)CCC	123	121
53 $C_6H_{12}O^A$ $CCC(CC)C=0$ 11811254 $C_6H_{12}O^*$ $CCC(C)C(C=0$ 11611655 $C_6H_{12}O^*$ $CC(C)C(C)C=0$ 11611556 $C_6H_{12}O^*$ $CC(C)C(C=0)CC$ 11811357 $C_6H_{12}O^*$ $CC(C)C(=0)CC$ 11610258 $C_7H_{12}O^*$ $CC(C)C(C=0)CC$ 10610259 $C_7H_{12}O^*$ $CCCCCCCCCCC=0$ 15716161 $C_7H_{12}O^*$ $CCCCCCCCCC=0$ 15315163 $C_7H_{12}O^*$ $CCC(C)=0)CCCC=C$ 13813264 $C_7H_{12}O^*$ $CC(CCCCCCC)=0$ 13713264 $C_7H_{12}O^*$ $CC(CCCCC)=0)C=C$ 13713264 $C_7H_{12}O^*$ $CC(CCCCCCC)=0$ 13713264 $C_7H_{12}O^*$ $CC(CCCCCCC)=0$ 137132	52	C ₆ H ₁₂ O^	CC(=0)CC(C)C	117	114
54 $C_6H_{12}O^*$ $CC(C)C(C)C=0$ 116 116 55 $C_6H_{12}O^*$ $CC(C)C(C)C=0$ 118 115 56 $C_6H_{12}O^*$ $CC(C=0)C(C)C=0$ 118 113 57 $C_6H_{12}O^*$ $CC(C)C(C)C=0$ 115 113 58 $C_7H_{12}O^*$ $CC(C=0)C(C)(C)C$ 106 102 59 $C_7H_{12}O^*$ $CCCC=CCC=0$ 157 161 60 $C_7H_{12}O^*$ $CCCC=CCCC=0$ 153 151 61 $C_7H_{12}O^*$ $CCCC=CCCC=0$ 153 151 63 $C_7H_{12}O^*$ $CCC(C)=0)C=C$ 138 132 63 $C_7H_{12}O^*$ $CC(CCC)=0)C=C$ 137 132 64 $C_7H_{12}O^*$ $CC(CCC)=0)C=C$ 137 132	53	C ₆ H ₁₂ O^	CCC(CC)C=0	118	112
55 $C_6H_{12}O^A$ $CC(f=O)C(C)CC$ 118 115 56 $C_6H_{12}O^A$ $CC(C)C(=O)CC$ 118 113 57 $C_6H_{12}O^A$ $CC(C)C(=O)CC$ 115 113 58 $C_7H_{12}O^A$ $CC(C=O)C(C)(C)C$ 106 102 58 $C_7H_{12}O^A$ $CCCCC=CC=O$ 166 162 60 $C_7H_{12}O^A$ $CCCCC=CCC=O$ 153 151 61 $C_7H_{12}O^A$ $CCCCC=CCC=O$ 153 151 63 $C_7H_{12}O^A$ $CCC(C=C)CCC=O$ 153 135 63 $C_7H_{12}O^A$ $CC(CCC(C)=O)C=C$ 137 132 64 $C_7H_{12}O^A$ $CC(CCCC)=O$ 137 132	54	C ₆ H ₁₂ O*	CC(C)C(C)C=O	116	116
56 $C_{6}H_{12}O^{*}$ $CC(C)C(=O)CC$ 115 113 57 $C_{6}H_{12}O^{*}$ $CC(C)C(=O)C(C)$ 106 102 58 $C_{7}H_{12}O^{*}$ $CC(=O)C(C)(C)C$ 106 102 59 $C_{7}H_{12}O^{*}$ $CCCCC=CCC=O$ 156 161 60 $C_{7}H_{12}O^{*}$ $CCCCC=CCC=O$ 153 151 61 $C_{7}H_{12}O^{*}$ $CCCCC=CCC=O$ 153 151 62 $C_{7}H_{12}O^{*}$ $CCC(C)=O)C=C$ 132 139 63 $C_{7}H_{12}O^{*}$ $CC(CCC(C)=O)C=C$ 138 132 64 $C_{7}H_{12}O^{*}$ $CC(CCCC)=O(C)=O)C=C$ 137 132	55	C ₆ H ₁₂ O^	CC(=0)C(C)CC	118	115
57 $C_6H_{12}O^A$ $CC(=O)C(C)(C)$ 106 102 58 $C_7H_{12}O^A$ $CCCCC=CC=O$ 166 162 59 $C_7H_{12}O^A$ $CCCCC=CC=O$ 156 161 60 $C_7H_{12}O^A$ $CCCCC=CCC=O$ 157 161 61 $C_7H_{12}O^A$ $CCCCC=CCC=O$ 153 151 62 $C_7H_{12}O^A$ $CCC(=O)CCC=C$ 142 139 63 $C_7H_{12}O^A$ $CC(CCC(C)=O)C=C$ 138 132 64 $C_7H_{12}O^A$ $CC(CCCC)=O)C=C$ 137 132	56	C ₆ H ₁₂ O*	CC(C)C(=0)CC	115	113
58 $C_7H_{12}O^{\wedge}$ CCCCCC=CC=0 166 162 59 $C_7H_{12}O^{*}$ CCCC=CCC=0 157 161 60 $C_7H_{12}O^{*}$ CCCC=CCCC=0 153 151 61 $C_7H_{12}O^{*}$ CCCC=CCC(C)=0 153 151 61 $C_7H_{12}O^{*}$ CCC(=0)CCC=C 142 139 62 $C_7H_{12}O^{*}$ CC(C=C)CCC(C)=0 150 135 63 $C_7H_{12}O^{*}$ CC(CC(C)=0)C=C 138 132 64 $C_7H_{12}O^{*}$ CC(CCC)=0 137 132	57	C ₆ H ₁₂ O^	CC(=0)C(C)(C)C	106	102
59 $C_7H_{12}O^*$ CCCC=CCC=O 157 161 60 $C_7H_{12}O^*$ CC=CCCC(C)=O 153 151 61 $C_7H_{12}O^*$ CC=CCCC(C)=O 153 151 62 $C_7H_{12}O^*$ CCC(=O)CCC=C 142 139 63 $C_7H_{12}O^*$ CC(CC(C)=O)C=C 138 132 64 $C_7H_{12}O^*$ CC(CC=C)C(C)=O)C=C 137 132	58	C ₇ H ₁₂ O^	0=00=0000	166	162
60 $C_7H_{12}O^*$ $CC=CCCC(C)=O$ 153 151 151 139 151 139 151 139 151 139 151 139 139 151 139 151 139 151 139 151 139 135 135 135 135 132 135 132 <td>59</td> <td>C₇H₁₂O*</td> <td>0=000=0000</td> <td>157</td> <td>161</td>	59	C ₇ H ₁₂ O*	0=000=0000	157	161
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^A CC(CC(C)=O)C=C 138 132 64 C ₇ H ₁₂ O^A CC(CC=C)C(C)=O 137 132	60	C ₇ H ₁₂ O*	CC=CCCC(C)=O	153	151
62 $C_7H_{12}O^*$ $CC(=C)CCC(C)=O$ 150 135 63 $C_7H_{12}O^{\wedge}$ $CC(CC(C)=O)C=C$ 138 132 64 $C_7H_{12}O^{\wedge}$ $CC(CC=C)C(C)=O$ 137 132	61	C ₇ H ₁₂ O*	CCC(=0)CCC=C	142	139
63 C ₇ H ₁₂ O ^A CC(CC(C)=0)C=C 138 132 64 C ₇ H ₁₂ O ^A CC(CC=C)C(C)=O 137 132	62	C ₇ H ₁₂ O*	CC(=C)CCC(C)=O	150	135
64 C ₇ H ₁₂ O ^A CC(CC=C)C(C)=O 137 132	63	C ₇ H ₁₂ O^	CC(CC(C)=0)C=C	138	132
	64	C ₇ H ₁₂ O^	CC(CC=C)C(C)=O	137	132

.

•

(cont.)

(Table 1 cont.)

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [28]
66	C ₇ H ₁₄ O*	CCCCCCC=0	153	150
67	C ₇ H ₁₄ O^	CC(=0)CCCCC	151	147
68	C ₇ H ₁₄ O^	CCC(C)CCC=O	144	144
69	C7H14O*	CCC(=0)CCCC	149	146
70	C7H14O*	CCCC(C)CC=O	144	143
7.1	C ₇ H ₁₄ O^	CC(=0)CCC(C)C	144	138
72	C7H14O^	CCCC(=0)CCC	144	139
73	C ₇ H ₁₄ O^	CC(=0)CC(C)CC	140	139
74	C ₇ H ₁₄ O^	CCC(=0)CC(C)C	136	137
75	C ₇ H ₁₄ O^	CC(=0)C(C)CCC	143	138
76	C7H14O*	CC(C)C(=O)CCC	135	136
77	C ₇ H ₁₄ O^	CCC(=0)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(=0)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) = 0	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(=0)C(=0)CCC	168	165
90	C ₈ H ₁₆ O^	CCCCCCC=0	171	170
91	C ₈ H ₁₆ O*	CCCCCC(C)=0	173	167
92	C ₈ H ₁₆ O^	CCCCCC(C)C=O	162	165
93	C ₈ H ₁₆ O*	CCCCCC(=0)CC	168	167
94	C ₈ H ₁₆ O*	CC(C)CCCC(=O)C	170	161
95	C ₈ H ₁₆ O*	CCCC(=0)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^	CCCCC(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(=0)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(=0)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=0	191	187
110	C ₉ H ₁₈ O^	CC(=0)CCCCCCC	195	186
111	C ₉ H ₁₈ O^	CCC(=0)CCCCCC	190	186

Number	Formula	SMILES	nbp (exp) (°C)	nbp (theor.) [28]
112	C ₉ H ₁₈ O*	CCCC(=0)CCCCC	187	183
113	C ₉ H ₁₈ O*	CCCCC(=0)CCCC	186	183
114	C ₉ H ₁₈ O^	CC(C)CCCC(=O)CC	183	182
115	C ₉ H ₁₈ O*	CCCC(=0)CCC(C)C	178	179
116	C ₉ H ₁₈ O^	CCCCCC(=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)(C)(C)(C)(C)(C)(C)(C)(C)(C)(C	152	149
124	C ₁₀ H ₂₀ O*	CCCCCCCCC=0	208	204
125	C ₁₀ H ₂₀ O*	CC(=0)CCCCCCC	210	203
126	C ₁₀ H ₂₀ O^	CCC(=0)CCCCCC	211	203
127	C ₁₀ H ₂₀ O*	CCCC(=0)CCCCCC	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	$C_4H_6O_2^{1}$	COC(=O)C=C	85	79
134	C ₄ H ₈ O ₂ *	CCCOC=0	81	79
135	C ₄ H ₈ O ₂ ^	CC(C)OC=O	68	76
136	$C_4H_8O_2^{1}$	CCOC(C)=O	77	77
137	C ₄ H ₈ O ₂ *	CCC(=O)OC	80	77
138	C5H8O2*	CC(=0)0CC=C	104	99
139	C ₅ H ₈ O ₂ *	CCOC(=0)C=C	101	99
140	C₅H ₈ O₂^	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=0	107	106
143	C ₅ H ₁₀ O ₂ ^	CC(C)COC=O	98	100
144	C ₅ H ₁₀ O ₂ *	CCCOC(C)=O	102	101
145	C ₅ H ₁₀ O ₂ *	CCC(C)OC=0	97	102
146	C ₅ H ₁₀ O ₂ *	CCCC(=0)0C	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)=0	90	97
149	C ₆ H ₁₀ O ₂ ^	CC(=0)OCCC=C	127	118
150	C ₆ H ₁₀ O ₂ *	CCCOC(=0)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	$C_6H_{10}O_2*$	CCOC(=O)C(C)=C	118	117
156	C ₆ H ₁₂ O ₂ *	CCCCCOC=O	132	131

(Tal	ble	1	con	t.)
1.000				•••

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(=0)0C	130	129
159	C ₆ H ₁₂ O ₂ ^	CCCOC(=0)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(=0)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(=0)OC(C)C	110	120
164	C ₆ H ₁₂ O ₂ *	CC(=0)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(=0)0CCCC=C	151	140
167	C ₇ H ₁₂ O ₂ *	CCOC(=0)CCC=C	145	139
168	C ₇ H ₁₂ O ₂ ^	CCCCC(=0)0C=C	133	143
169	C ₇ H ₁₂ O ₂ *	CC(C)COC(=0)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(=0)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=0	156	153
174	C ₇ H ₁₄ O ₂ ^	CCCCCOC(C)=O	149	150
175	C ₇ H ₁₄ O ₂ *	CCCCCC(=0)0C	151	151
176	C ₇ H ₁₄ O ₂ ^	CCCCOC(=0)CC	146	147
177	C ₇ H ₁₄ O ₂ ^	CCCCC(=0)0CC	145	148
178	C ₇ H ₁₄ O ₂ ^	CCCOC(=0)CCC	143	146
179	C ₇ H ₁₄ O ₂ *	CCC(=0)OCC(C)C	13 ₇	141
180	C ₇ H ₁₄ O ₂ *	CCCC(=0)OC(C)C	131	143
181	C ₇ H ₁₄ O ₂ ^	CCC(C)OC(=0)CC	132	142
182	C7H14O2*	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(=0)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(=0)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(=0)C=CC=CC	180	179
189	C ₈ H ₁₄ O ₂ *	CCC(OC(C)=0)C=CC	156	165
190	C ₈ H ₁₄ O ₂ *	CC(=0)0C(C)(C)CC=C	138	150
191	C ₈ H ₁₆ O ₂ ^	CCCCCCOC(C)=0	169	170
192	C ₈ H ₁₆ O ₂ *	CCCCCCC(=0)0C	172	170
193	C ₈ H ₁₆ O ₂ ^	CCCCCOC(=0)CC	169	169
194	C ₈ H ₁₆ O ₂ *	CCCCCC(=0)0CC	167	168
195	C ₈ H ₁₆ O ₂ *	CCCCOC(=0)CCC	165	166
196	C ₈ H ₁₆ O ₂ ^	CCCCC(=0)0CCC	168	168
197	C ₈ H ₁₆ O ₂ *	CCCC(=0)0CC(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	$C_9H_{18}O_2*$	CCCCCCCCCC=0	178	190

* Member of the training set

^ Member of the test set

Same and the

A LOUIS CONTRACTOR

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 significative 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.*

$$nbp = b(DCW) + a$$
 [IV]

$$nbp = c(DCW)^2 + b(DCW) + a$$
[V]

$$nbp = d(DCW)^3 + c(DCW)^2 + b(DCW) + a$$
[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:

nbp =
$$-0.0230 (\pm 0.0043) (DCW)^2 + 5.5248 (\pm 0.3322) (DCW) - 46.1488 (\pm 6.1437) [VIII] r = 0.9883, SE = 5.8820$$

$$\label{eq:nbp} \begin{split} nbp &= 0.0006 \ (\pm 0.0003) \ (DCW)^3 - 0.0888 \ (\pm 0.0372) \ (DCW)^2 + 7.8578 \ (\pm 1.3494) \ (DCW) - \\ &- 71.6714 \ (\pm 15.5567) \ [IX] \\ r &= 0.9887, \ SE = 5.8285 \end{split}$$

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

Atom(A)	CW(A)	NNC	CW(NNC)	-
ee	0.3881	0100	0.3658	
Н	1.6433	0110	0.1612	
С	0.4739	0211	0.8799	
0	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 2.	Correlation	weights of	local graph	invariants.
--	----------	-------------	------------	-------------	-------------

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)	

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 (0)	125 (5)	126 (4)	117 (13)
38 V	125 (7)	138 (-11)	139 (-12)	133 (-6)
20 /	136 (4)	138 (2)	139 (1)	142 (-2)
39	100 (4) 110 (5)	121 (3)	122 (2)	113 (11)
40 **	136 (3)	138 (1)	138 (1)	134 (5)
41 40 *	124 (2)	136 (0)	136 (0)	135 (1)
42 ~	119 (0)	121 (-2)	122 (-3)	114 (5)
45 /	132 (-2)	134 (-4)	135 (-5)	132 (-2)
44 ^	132 (-2)	130 (0)	131 (-1)	124 (6)
40 ~ AG *	120 (2)	127 (Δ)	127 (4)	129 (2)
40 ~ 47 *	116 (6)	118 (<i>A</i>)	118 (4)	119 (3)
4/ ^	1 10 (0) 1 7 7 (E)	125 (3)	126 (1)	125 (2)
48 ^	122 (3)	118 (1)	118 (4)	120 (2)
49 *	116 (0)	110 (4)	118 (-7)	120 (-4)
50 *	1 10 (U)	174 (1)	125 (-2)	121 (2)
51 *	122(1)	124 (-1)	116 (1)	11/ (2)
52 ^	113 (4)	115(2)	110(1)	114(2)

(Table 3 cont.)

Molecule	nbp Eq.(7)	nbp Eq.(8)	Nbp Eq.(9)	AD [28]
53 ^	116 (2)		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)

k

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)	171 (-4)
186 *	1/9 (-12)	177 (-10)	170 (10)	179 (3)
187 ^	179 (-4)	177 (-2)	179 (-4)	1/8 (-3)

A NOT A DESCRIPTION OF

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 form 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] Katritizky, 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] Egolf, 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] Trinajstic, N.; Nikolic, S.; Lucic, 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. Koren Chem. Soc. 2000, 44(6), 501.
- [46] Duchowicz, P.; Castro, E.A. Arkivoc 2001, 2(6), U39.