

## Correlation between Structure and Normal Boiling Points of Acyclic Carbonyl Compounds

Alexandru T. Balaban,<sup>\*,†</sup> Denise Mills, and Subhash C. Basak

Natural Resources Research Institute, University of Minnesota–Duluth, 5013 Miller Trunk Highway,  
Duluth, Minnesota 55811

Received March 17, 1999

Acyclic saturated and unsaturated carbonyl compounds (mono- and dialdehydes, mono- and diketones, keto aldehydes, and esters of monocarboxylic acids) with various degrees of branching but devoid of hydrogen bonding were included in a quantitative structure–property relationship study for correlating chemical structure with normal boiling points. Three separate correlations are presented in terms of five topological indices: for 127 mono- and dialdehydes and -ketones, for 73 esters, and for the combined set of all 200 carbonyl compounds. Standard deviations are lower than 7 °C.

### INTRODUCTION

Among physicochemical experimentally determined properties, the boiling points at normal pressure (normal boiling points, NBPs) 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. A more universal property would be the variation of vapor pressure versus temperature up to the critical point; however, such data are scarce, but some quantitative structure–property relationships (QSPRs) and correlations with topological indices have been reported both for the parameters of the Antoine equation of vapor pressures and for critical data of alkanes.<sup>1</sup>

Topological indices (TIs) are numbers associated with chemical formulas via mathematical operations on graph-theoretical invariants of the corresponding molecular (constitutional) graphs.<sup>2–11</sup> These TIs have the advantage that they can be easily and rapidly computed for any constitutional formula, making it possible to scan huge combinatorial libraries such as the virtual libraries resulting from combining various substituents on a molecular scaffold in drug design. A recent example is provided by the optimization of a peptidic fragment in an immunosuppressive protein used for organ transplants.<sup>12</sup> Another advantage resides in the fact that TIs have good correlational abilities<sup>13</sup> and a sufficient diversity for allowing meaningful statistics to come into play. So far, the main drawbacks of TIs (difficulty to encode stereochemical information, degeneracy, and hence impossibility to retrieve structures from the values of TIs) have only partly been overcome.<sup>6,14,15</sup>

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 TIs, one must rely on intramolecular comparisons for the novel local vertex invariants and on intermolecular comparisons for the result-

ing molecular descriptors, i.e., the TIs. These comparisons and the corresponding correlations should be made for a molecular property that is measured with sufficient accuracy. The NBPs fulfill this expectation best.<sup>13,16–39</sup> Also, NBPs are convenient because they are not practically identical for isomeric compounds with the same composition (same molecular formula) like many thermodynamic properties. The NBPs usually increase with increasing molecular weight (fluoroalkanes are an exception), but at the same time they depend also on constitution and differ significantly among isomers.

### QSPR FOR NBPS OF ACYCLIC CARBONYL COMPOUNDS

Recently, we reported that aliphatic dinitriles presented the highest increment in the NBP of a homologous series on addition of a methylene group: whereas oxalonitrile (cyanogen) has an NBP of –22 °C, the next homologue, malononitrile, has an NBP of 220 °C; the reasons for this exceptionally high increment were examined and discussed.<sup>40</sup> Earlier papers had explored the use of TIs for QSPR studies in aliphatic molecules possessing heteroatoms such as halogens,<sup>37,38</sup> ethereal oxygen, or divalent sulfur atoms, excluding systems with intermolecular hydrogen bonds.<sup>35,39</sup>

It was of interest to continue these studies, extending them to carbonyl compounds and also avoiding molecules with inter- or intramolecular hydrogen bonding. Since the problem of accurate descriptors for heteroatoms in topological indices is a critical one in QSAR/QSPR (QSAR = quantitative structure–activity relationship) studies,<sup>41</sup> it was interesting to determine whether TIs could yield useful correlations for oxygen heteroatoms in carbonyl and carbalkoxy groups. A set of 200 NBP values for acyclic monoaldehydes, dialdehydes, monoketones, diketones, keto aldehydes, and esters (carbalkoxy derivatives of monocarboxylic acids) was collected. Saturated systems, as well as systems with one or two double bonds or one triple bond, were included, but not ketenes. For most compounds, two information sources were used.<sup>42,43</sup> The experimental data are assumed to be correct with an error margin of less than 4 °C. Differences in NBP among diastereoisomers fall within this error margin.

<sup>†</sup> Permanent address: Organic Chemistry Department, Polytechnic University, 77206 Bucharest, Romania.

The experimental data are presented in Table 1 together with the structures indicated by Smiles notation.<sup>44</sup> For acyclic compounds, this notation is simpler to follow than the usual IUPAC nomenclature; there is no ambiguity in how double bonds are denoted; triple bonds are indicated by the symbol #. As in a previous paper,<sup>40</sup> values of topological indices (A) in correlations are computed in the form  $\ln(A + 1)$ .

### SELECTION OF MOLECULAR DESCRIPTORS

It was considered that indicator variables (such as the number of multiple bonds, carbon atoms, or the total number of non-hydrogen atoms) are not needed if one uses a computer program that calculates for each structure a sufficient number of TIs and if a statistical treatment is then applied to select a given number of such molecular descriptors. Actually, when including the number of carbon atoms along with topological indices, poor correlations were obtained. With our fairly large numbers of compounds in the two sets (esters in one set and aldehydes plus ketones in the other set) and a fortiori in the combined set, we decided to use the in-house program POLLY,<sup>45</sup> which computes over 100 TIs, and then to select five descriptors on the basis of statistical results. A comparable but different set of TIs is computed by the program CODESSA.<sup>46,47</sup>

We decided to start with the following 13 indices:  $s_0$ ,  $v_0$ ,  $v_1$ ,  $b_0$ ,  $J_b$ ,  $J_x$ ,  $J_y$ ,  $J$ ,  $IC_0$ ,  $IC_1$ ,  $IC_2$ ,  $CIC_1$ , and  $CIC_2$ . After eliminating intercorrelated indices from this group by way of the CORR procedure of the SAS program,<sup>48</sup> we were left with the following 8 indices:  $s_0$ ,  $J$ ,  $J_y$ ,  $IC_0$ ,  $IC_1$ ,  $IC_2$ ,  $CIC_1$ , and  $CIC_2$ . Using the REG procedure of the same statistical program, we obtained an optimal five-parameter model for the combined data set. These same five parameters were used to develop equations for the two separate data sets.

The five molecular descriptors selected by the statistical package<sup>48</sup> are the following ones: (1) the average distance-based molecular connectivities index  $J$ <sup>49</sup> and (2) its modification  $J_y$  for the presence of heteroatoms and multiple bonding [the nature of heteroatoms is introduced according to a periodically varying atomic property, namely, the covalent radius  $y_i$  (relative to that of carbon) calculated according to a formula<sup>50</sup> in terms of the numbers of the group and period of Mendeleev's table of elements; multiple bonds are taken into account by compressing the topological distances to  $1/2$  for a double bond and  $1/3$  for a triple bond],<sup>35,36,51,52</sup> (3) information content indices  $IC_1$  and  $IC_2$ ;<sup>53</sup> (4) complementary information content indices  $CIC_1$  and  $CIC_2$ ;<sup>53</sup> and (5) index  $s_0$ , which is the sum of square roots of vertex degrees ( $\delta_i$ ).<sup>11</sup>

The equations defining the above indices are the following ones:

$$J_y = q(\mu + 1)^{-1} \sum_{ij} (d_i y_i \times d_j y_j)^{-0.5}$$

$$J = q(\mu + 1)^{-1} \sum_{ij} (d_i \times d_j)^{-0.5}$$

In the above equations,  $q$  is the number of edges (covalent bonds) in the hydrogen-depleted molecular graph,  $\mu$  is the cyclomatic number (zero in the present case), and  $d_i$  denotes

the distance-sum of vertex  $i$  (or atom  $i$ ); this distance-sum is obtained from the distance matrix of the graph as the sum over row or column  $i$  which contains the topological distances from vertex  $i$  to all other vertices in the graph. The summation denoted by  $ij$  extends to all edges  $ij$ , i.e., to all covalent bonds between atoms  $i$  and  $j$ .

$$s_0 = \sum_i (\delta_i)^{-0.5}$$

$$IC_r = - \sum_{i=1}^r p_i \log_2 p_i$$

### RESULTS AND DISCUSSION

Because we expected that aldehydes and ketones as well as aldoketones devoid of hydrogen-bonded derivatives (such hydrogen-bonded derivatives are hydroxy, amino, or amido derivatives or enolic tautomers in measurable amounts) would behave similarly, we investigated a single correlation for all such carbonyl compounds. However, for esters we suspected possible deviations from the behavior of the previous carbonyl compounds, and therefore we investigated first a separate correlation for the esters and only afterward a single correlation for all carbonyl derivatives.

In Table 1 we present only the results of the latter correlation which leads to the poorest statistical parameters, but in the following we give the data for all three correlations.

ketones and aldehydes:

$$NBP = 210(\pm 21)J_y - 326(\pm 18)J + 251(\pm 7)s_0 + 61(\pm 12)IC_2 - 134(\pm 25)IC_1 - 160(\pm 14) \quad (1)$$

$$N = 127; \quad s = 6.49 \text{ }^\circ\text{C}; \quad r^2 = 0.9705; \quad F = 796$$

esters:

$$NBP = 320(\pm 20)J_y - 418(\pm 22)J + 217(\pm 9)s_0 + 150(\pm 15)IC_2 - 281(\pm 26)IC_1 - 99(\pm 15) \quad (2)$$

$$N = 73; \quad s = 4.0 \text{ }^\circ\text{C}; \quad r^2 = 0.9866; \quad F = 984$$

combined set:

$$NBP = 302(\pm 13)J_y - 372(\pm 15)J + 223(\pm 6)s_0 + 116(\pm 9)IC_2 - 272(\pm 13)IC_1 - 109(\pm 11) \quad (3)$$

$$N = 200; \quad s = 6.93 \text{ }^\circ\text{C}; \quad r^2 = 0.9640; \quad F = 1039$$

The intercorrelation factors between the five TIs are presented in Table 2 in matrix form. It can be seen that these factors are not exceedingly high (all are lower than 0.9). In addition, all parameters are statistically significant according to the  $t$ -score, being significant at less than 0.0001 for each parameter in all three models.

Correlations 1–3 are acceptable on the basis of the values  $r^2 \geq 0.964$  and of the standard deviation, which is lower than 7 °C for the first and last sets, and equal to 4 °C for the set of esters. No outliers with deviations larger than 25 °C can be seen in Table 1.

As seen from eqs 1–3 the coefficients of the five indices and the free term (indicating the algebraic signs and relative weights) are rather similar for all three sets. Actually, all terms of eq 3 are intermediate between those of the first two

**Table 1.** Data for 173 Aldehydes and Ketones, Followed by Data for 73 Esters<sup>a</sup>

	formula	SMILES	$J_y$	$J$	$s_0$	IC <sub>1</sub>	IC <sub>2</sub>	expt	calc	expt-calc
Aldehydes + Ketones										
1	C <sub>2</sub> H <sub>4</sub> O	CC=O	1.1734	0.9681	1.3102	0.9798	1.1404	21	45	-24
2	C <sub>3</sub> H <sub>2</sub> O	O=CC#C	1.3757	1.0902	1.4848	1.1793	1.2768	56	61	-5
3	C <sub>3</sub> H <sub>4</sub> O	C=CC=O	1.3260	1.0902	1.4848	1.0986	1.3218	53	73	-20
4	C <sub>3</sub> H <sub>6</sub> O	CCC=O	1.2021	1.0902	1.4848	1.0192	1.2936	48	54	-6
5	C <sub>3</sub> H <sub>6</sub> O	CC(=O)C	1.3463	1.2012	1.5210	0.9443	0.9443	56	44	12
6	C <sub>4</sub> H <sub>4</sub> O	CC#CC=O	1.4310	1.1603	1.6333	1.1181	1.2925	107	104	3
7	C <sub>4</sub> H <sub>4</sub> O	CC(=O)C#C	1.4780	1.2641	1.6647	1.1881	1.2925	84	67	17
8	C <sub>4</sub> H <sub>6</sub> O	CC=CC=O	1.3737	1.1603	1.6333	1.0536	1.3468	105	110	-5
9	C <sub>4</sub> H <sub>6</sub> O	CC(=C)C=O	1.4429	1.2641	1.6647	1.1148	1.3468	68	83	-15
10	C <sub>4</sub> H <sub>6</sub> O	CC(=O)C=C	1.4443	1.2641	1.6647	1.1148	1.3468	81	83	-2
11	C <sub>4</sub> H <sub>8</sub> O	CCCC=O	1.2308	1.1603	1.6333	0.9933	1.3135	75	79	-4
12	C <sub>4</sub> H <sub>8</sub> O	CC(C)C=O	1.3483	1.2641	1.6647	0.9933	1.2102	64	71	-7
13	C <sub>4</sub> H <sub>8</sub> O	CC(=O)CC	1.3553	1.2641	1.6647	0.9933	1.2102	80	73	7
14	C <sub>5</sub> H <sub>6</sub> O	C#CC(C)C=O	1.4637	1.3228	1.7904	1.1915	1.4325	91	84	7
15	C <sub>5</sub> H <sub>6</sub> O	CCC(=O)C#C	1.4720	1.3228	1.7904	1.1915	1.3918	106	82	24
16	C <sub>5</sub> H <sub>8</sub> O	C=CCCC=O	1.3013	1.2057	1.7627	1.0818	1.4093	99	100	-1
17	C <sub>5</sub> H <sub>8</sub> O	CC=CCC=O	1.3527	1.2057	1.7627	1.0818	1.4308	106	118	-12
18	C <sub>5</sub> H <sub>8</sub> O	CCC=CC=O	1.3760	1.2057	1.7627	1.0818	1.4308	124	125	-1
19	C <sub>5</sub> H <sub>8</sub> O	CCC(=O)C=C	1.4500	1.3228	1.7904	1.1291	1.4308	103	97	6
20	C <sub>5</sub> H <sub>8</sub> O	CC=CC(C)=O	1.4651	1.2884	1.7904	1.0321	1.2689	122	122	0
21	C <sub>5</sub> H <sub>8</sub> O	CC(C)=CC=O	1.4877	1.2884	1.7904	1.0818	1.2689	133	115	18
22	C <sub>5</sub> H <sub>8</sub> O	CC=C(C)C=O	1.5103	1.3228	1.7904	1.0818	1.3083	117	114	3
23	C <sub>5</sub> H <sub>8</sub> O	CC(=C)C(C)=O	1.5252	1.3845	1.8173	1.0818	1.2689	98	97	1
24	C <sub>5</sub> H <sub>8</sub> O <sub>2</sub>	CCC(=O)C(C)=O	1.5319	1.4217	1.9260	1.0671	1.2826	108	115	-7
25	C <sub>5</sub> H <sub>10</sub> O	CCCCC=O	1.2545	1.2057	1.7627	0.9658	1.3172	103	107	-4
26	C <sub>5</sub> H <sub>10</sub> O	CC(C)CC=O	1.3436	1.2884	1.7904	1.0293	1.3300	93	93	0
27	C <sub>5</sub> H <sub>10</sub> O	CC(=O)CCC	1.3494	1.2884	1.7904	0.9836	1.2616	102	99	3
28	C <sub>5</sub> H <sub>10</sub> O	CCC(C)C=O	1.3828	1.3228	1.7904	1.0293	1.3625	91	96	-5
29	C <sub>5</sub> H <sub>10</sub> O	CCC(=O)CC	1.3900	1.3228	1.7904	0.9836	1.1881	102	90	12
30	C <sub>5</sub> H <sub>10</sub> O	CC(=O)C(C)C	1.4549	1.3845	1.8173	0.9658	1.1288	94	91	3
31	C <sub>5</sub> H <sub>10</sub> O	CC(C)(C)C=O	1.4974	1.4274	1.8257	0.9658	1.0716	75	83	-8
32	C <sub>6</sub> H <sub>8</sub> O	CC=CC=CC=O	1.4547	1.2375	1.8773	1.0203	1.3669	174	172	2
33	C <sub>6</sub> H <sub>8</sub> O	C#CC(=O)C(C)C	1.5405	1.4217	1.9260	1.1550	1.3191	118	98	20
34	C <sub>6</sub> H <sub>8</sub> O	CC=C(C=C)C=O	1.5825	1.3843	1.9020	1.1550	1.4639	144	136	8
35	C <sub>6</sub> H <sub>10</sub> O	CCC=CCC=O	1.3699	1.2375	1.8773	1.0640	1.4115	121	140	-19
36	C <sub>6</sub> H <sub>10</sub> O	CCCC=CC=O	1.3704	1.2375	1.8773	1.0640	1.4398	146	143	3
37	C <sub>6</sub> H <sub>10</sub> O	CC(=O)CCC=C	1.3830	1.3024	1.9020	1.1039	1.4398	130	117	13
38	C <sub>6</sub> H <sub>10</sub> O	CC=CCC(C)=O	1.4255	1.3024	1.9020	1.0640	1.3712	127	133	-6
39	C <sub>6</sub> H <sub>10</sub> O	CCC=CC(C)=O	1.4551	1.3024	1.9020	1.0640	1.3712	140	142	-2
40	C <sub>6</sub> H <sub>10</sub> O	CC(=C)CC(C)=O	1.4676	1.3745	1.9260	1.1039	1.3712	124	113	11
41	C <sub>6</sub> H <sub>10</sub> O	CCC(=O)C=CC	1.4805	1.3434	1.9020	1.0640	1.3712	139	134	5
42	C <sub>6</sub> H <sub>10</sub> O	CCC=C(C)C=O	1.5063	1.3434	1.9020	1.1039	1.4004	136	135	1
43	C <sub>6</sub> H <sub>10</sub> O	CCC(=O)C(C)=C	1.5269	1.4217	1.9260	1.1039	1.3712	119	114	5
44	C <sub>6</sub> H <sub>10</sub> O	CC(C)=CC(C)=O	1.5482	1.3745	1.9260	1.0487	1.1915	130	132	-2
45	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CCC(=O)C(=O)CC	1.5450	1.4568	2.0241	1.0578	1.2293	130	124	6
46	C <sub>6</sub> H <sub>12</sub> O	CCCCCC=O	1.2734	1.2375	1.8773	0.9412	1.2890	131	129	2
47	C <sub>6</sub> H <sub>12</sub> O	CC(C)CCC=O	1.3416	1.3024	1.9020	1.0199	1.3623	122	119	3
48	C <sub>6</sub> H <sub>12</sub> O	CC(=O)CCCC	1.3457	1.3024	1.9020	0.9662	1.2851	127	125	2
49	C <sub>6</sub> H <sub>12</sub> O	CCC(C)CC=O	1.3855	1.3434	1.9020	1.0199	1.3890	122	120	2
50	C <sub>6</sub> H <sub>12</sub> O	CCCC(C)C=O	1.3868	1.3434	1.9020	1.0199	1.3890	116	120	-4
51	C <sub>6</sub> H <sub>12</sub> O	CCC(=O)CCC	1.3925	1.3434	1.9020	0.9662	1.2556	123	121	2
52	C <sub>6</sub> H <sub>12</sub> O	CC(=O)CC(C)C	1.4221	1.3745	1.9260	1.0054	1.2630	117	114	3
53	C <sub>6</sub> H <sub>12</sub> O	CCC(CC)C=O	1.4310	1.3843	1.9020	1.0199	1.3350	118	112	6
54	C <sub>6</sub> H <sub>12</sub> O	CC(C)C(C)C=O	1.4699	1.4217	1.9260	0.9662	1.2211	116	116	0
55	C <sub>6</sub> H <sub>12</sub> O	CC(=O)C(C)CC	1.4738	1.4217	1.9260	1.0054	1.2923	118	115	3
56	C <sub>6</sub> H <sub>12</sub> O	CC(C)C(=O)CC	1.4766	1.4217	1.9260	1.0054	1.2630	115	113	2
57	C <sub>6</sub> H <sub>12</sub> O	CC(=O)C(C)(C)C	1.5719	1.5131	1.9569	0.9412	1.0054	106	102	4
58	C <sub>7</sub> H <sub>12</sub> O	CCCCC=CC=O	1.3661	1.2613	1.9800	1.0413	1.4372	166	162	4
59	C <sub>7</sub> H <sub>12</sub> O	CCCC=CCC=O	1.3745	1.2613	1.9800	1.0413	1.4132	157	161	-4
60	C <sub>7</sub> H <sub>12</sub> O	CC=CCCC(C)=O	1.4036	1.3126	2.0023	1.0547	1.3977	153	151	2
61	C <sub>7</sub> H <sub>12</sub> O	CCC(=O)CCC=C	1.4173	1.3512	2.0023	1.0760	1.4372	142	139	3
62	C <sub>7</sub> H <sub>12</sub> O	CC(=C)CCC(C)=O	1.4348	1.3681	2.0241	1.0889	1.3977	150	135	15
63	C <sub>7</sub> H <sub>12</sub> O	CC(CC(C)=O)C=C	1.4807	1.4107	2.0241	1.1220	1.4693	138	132	6
64	C <sub>7</sub> H <sub>12</sub> O	CC(CC=C)C(C)=O	1.5003	1.4282	2.0241	1.1220	1.4693	137	132	5
65	C <sub>7</sub> H <sub>12</sub> O	CC=CC(=O)C(C)C	1.5433	1.4282	2.0241	1.0413	1.3070	148	148	0
66	C <sub>7</sub> H <sub>14</sub> O	CCCCCCC=O	1.2887	1.2613	1.9800	0.9199	1.2576	153	150	3
67	C <sub>7</sub> H <sub>14</sub> O	CC(=O)CCCCC	1.3450	1.3126	2.0023	0.9486	1.2734	151	147	4
68	C <sub>7</sub> H <sub>14</sub> O	CCC(C)CCC=O	1.3825	1.3512	2.0023	1.0043	1.3985	144	144	0
69	C <sub>7</sub> H <sub>14</sub> O	CCC(=O)CCCC	1.3880	1.3512	2.0023	0.9486	1.2734	149	146	3
70	C <sub>7</sub> H <sub>14</sub> O	CCCC(C)CC=O	1.3985	1.3661	2.0023	1.0043	1.3985	144	143	1
71	C <sub>7</sub> H <sub>14</sub> O	CC(=O)CCCC(C)C	1.4029	1.3681	2.0241	1.0043	1.3105	144	138	6
72	C <sub>7</sub> H <sub>14</sub> O	CCCC(=O)CCC	1.4049	1.3661	2.0023	0.9486	1.2211	144	139	5
73	C <sub>7</sub> H <sub>14</sub> O	CC(=O)CC(C)CC	1.4479	1.4107	2.0241	1.0043	1.3347	140	139	1

Table 1. (Continued)

	formula	SMILES	$J_y$	$J$	$s_0$	IC <sub>1</sub>	IC <sub>2</sub>	expt	calc	expt-calc
Aldehydes + Ketones (Continued)										
74	C <sub>7</sub> H <sub>14</sub> O	CCC(=O)CC(C)C	1.4507	1.4107	2.0241	1.0043	1.3105	136	137	-1
75	C <sub>7</sub> H <sub>14</sub> O	CC(=O)C(C)CCC	1.4672	1.4282	2.0241	1.0043	1.3347	143	138	5
76	C <sub>7</sub> H <sub>14</sub> O	CC(C)C(=O)CCC	1.4704	1.4282	2.0241	1.0043	1.3105	135	136	-1
77	C <sub>7</sub> H <sub>14</sub> O	CCC(=O)C(C)CC	1.4998	1.4568	2.0241	1.0043	1.3105	135	134	1
78	C <sub>7</sub> H <sub>14</sub> O	CC(C)C(=O)C(C)C	1.5422	1.4960	2.0454	0.9486	1.0849	124	126	-2
79	C <sub>7</sub> H <sub>14</sub> O	CCC(=O)C(C)C(C)	1.5784	1.5310	2.0521	0.9832	1.1622	125	125	0
80	C <sub>8</sub> H <sub>12</sub> O	CC(C)=CC(=O)C=C(C)C	1.6409	1.4672	2.2133	1.0522	1.1660	198	186	12
81	C <sub>8</sub> H <sub>14</sub> O	CC(C)C=CCC(C)=O	1.4766	1.3648	2.1134	1.0688	1.3918	163	174	-11
82	C <sub>8</sub> H <sub>14</sub> O	CC(C)CC=CC(C)=O	1.4768	1.3648	2.1134	1.0688	1.3918	179	174	5
83	C <sub>8</sub> H <sub>14</sub> O	CC=CC(=O)CC(C)C	1.5112	1.4233	2.1134	1.0688	1.3918	170	162	8
84	C <sub>8</sub> H <sub>14</sub> O	CC=CCC(=O)C(C)C	1.5112	1.4243	2.1134	1.0688	1.3918	162	162	0
85	C <sub>8</sub> H <sub>14</sub> O	CCC(C)C(=O)C=CC	1.5579	1.4644	2.1134	1.0688	1.4132	171	164	7
86	C <sub>8</sub> H <sub>14</sub> O	CCC(C)=CC(=O)CC	1.5713	1.4405	2.1134	1.0688	1.3476	167	169	-2
87	C <sub>8</sub> H <sub>14</sub> O	CC=C(C)C(C)C(C)=O	1.6210	1.5208	2.1329	1.0501	1.2955	154	157	-3
88	C <sub>8</sub> H <sub>14</sub> O	CCC(C)=C(C)C(C)=O	1.6683	1.5208	2.1329	1.0192	1.2467	158	174	-16
89	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	CCCC(=O)C(=O)CCC	1.5341	1.4761	2.1953	1.0228	1.2590	168	165	3
90	C <sub>8</sub> H <sub>16</sub> O	CCCCCCCC=O	1.3013	1.2795	2.0732	0.9014	1.2276	171	170	1
91	C <sub>8</sub> H <sub>16</sub> O	CCCCCCCC(C)=O	1.3460	1.3210	2.0935	0.9318	1.2539	173	167	6
92	C <sub>8</sub> H <sub>16</sub> O	CCCCCCC(C)C=O	1.3803	1.3551	2.0935	0.9873	1.3820	162	165	-3
93	C <sub>8</sub> H <sub>16</sub> O	CCCCCCC(=O)CC	1.3835	1.3551	2.0935	0.9318	1.2627	168	167	1
94	C <sub>8</sub> H <sub>16</sub> O	CC(C)CCCC(=O)C	1.3913	1.3648	2.1134	0.9944	1.3355	170	161	9
95	C <sub>8</sub> H <sub>16</sub> O	CCCC(=O)CCCC	1.4059	1.3750	2.0935	0.9318	1.2398	165	163	2
96	C <sub>8</sub> H <sub>16</sub> O	CC(=O)CCC(C)CC	1.4294	1.4014	2.1134	0.9944	1.3564	167	162	5
97	C <sub>8</sub> H <sub>16</sub> O	CC(C)CCC(=O)CC	1.4315	1.4014	2.1134	0.9944	1.3355	164	160	4
98	C <sub>8</sub> H <sub>16</sub> O	CCCCC(CC)C=O	1.4375	1.4090	2.0935	0.9873	1.3820	163	162	1
99	C <sub>8</sub> H <sub>16</sub> O	CCCC(C)CC(=O)C	1.4526	1.4233	2.1134	0.9944	1.3564	156	161	-5
100	C <sub>8</sub> H <sub>16</sub> O	CC(C)CC(=O)CCC	1.4563	1.4233	2.1134	0.9944	1.3355	156	159	-3
101	C <sub>8</sub> H <sub>16</sub> O	CC(C)C(=O)CCCC	1.4572	1.4243	2.1134	0.9944	1.3355	160	159	1
102	C <sub>8</sub> H <sub>16</sub> O	CCCC(CCC)C=O	1.4586	1.4291	2.0935	0.9873	1.3410	160	156	4
103	C <sub>8</sub> H <sub>16</sub> O	CCC(C)CC(=O)CC	1.4727	1.4405	2.1134	0.9944	1.3355	161	158	3
104	C <sub>8</sub> H <sub>16</sub> O	CCCC(=O)C(C)CC	1.4994	1.4644	2.1134	0.9944	1.3355	154	157	-3
105	C <sub>8</sub> H <sub>16</sub> O	CC(C)CC(=O)C(C)C	1.5114	1.4764	2.1329	0.9873	1.2363	147	151	-4
106	C <sub>8</sub> H <sub>16</sub> O	CCC(=O)C(C)C(C)C	1.6078	1.5690	2.1391	0.9873	1.2363	151	147	4
107	C <sub>8</sub> H <sub>16</sub> O	CC(C)C(=O)C(C)C(C)C	1.6249	1.5847	2.1581	0.9628	1.1086	135	142	-7
108	C <sub>9</sub> H <sub>16</sub> O	CCC=CC(=O)CC(C)C	1.5160	1.4255	2.1953	1.0654	1.4212	183	186	-3
109	C <sub>9</sub> H <sub>18</sub> O	CCCCCCCCC=O	1.3118	1.2942	2.1584	0.8858	1.1997	191	187	4
110	C <sub>9</sub> H <sub>18</sub> O	CC(=O)CCCCCCC	1.3481	1.3279	2.1770	0.9167	1.2329	195	186	9
111	C <sub>9</sub> H <sub>18</sub> O	CCC(=O)CCCCCC	1.3800	1.3574	2.1770	0.9167	1.2459	190	186	4
112	C <sub>9</sub> H <sub>18</sub> O	CCCC(=O)CCCCC	1.4032	1.3783	2.1770	0.9167	1.2329	187	183	4
113	C <sub>9</sub> H <sub>18</sub> O	CCCCC(=O)CCCC	1.4115	1.3858	2.1770	0.9167	1.2329	186	183	3
114	C <sub>9</sub> H <sub>18</sub> O	CC(C)CCCC(=O)CC	1.4183	1.3945	2.1953	0.9825	1.3507	183	182	1
115	C <sub>9</sub> H <sub>18</sub> O	CCCC(=O)CCC(C)C	1.4429	1.4168	2.1953	0.9825	1.3318	178	179	-1
116	C <sub>9</sub> H <sub>18</sub> O	CCCCCC(=O)C(C)C	1.4441	1.4183	2.1953	0.9825	1.3318	183	179	4
117	C <sub>9</sub> H <sub>18</sub> O	CCCCC(C)CC(=O)C	1.4488	1.4255	2.1953	0.9825	1.3689	183	182	1
118	C <sub>9</sub> H <sub>18</sub> O	CCCC(C)CC(=O)CC	1.4770	1.4514	2.1953	0.9825	1.3507	179	178	1
119	C <sub>9</sub> H <sub>18</sub> O	CCC(C)CC(=O)CCC	1.4786	1.4514	2.1953	0.9825	1.3507	161	179	-18
120	C <sub>9</sub> H <sub>18</sub> O	CCCCC(=O)C(C)CC	1.4891	1.4607	2.1953	0.9825	1.3507	174	179	-5
121	C <sub>9</sub> H <sub>18</sub> O	CC(C)CC(=O)CC(C)C	1.4956	1.4672	2.2133	0.9925	1.2485	165	168	-3
122	C <sub>9</sub> H <sub>18</sub> O	CCC(C)C(=O)C(C)CC	1.5759	1.5444	2.2133	0.9925	1.2887	162	168	-6
123	C <sub>9</sub> H <sub>18</sub> O	CC(C)C(C)C(=O)C(C)C(C)C	1.6903	1.6546	2.2594	0.9167	0.9167	152	149	3
124	C <sub>10</sub> H <sub>20</sub> O	CCCCCCCCC=O	1.3207	1.3059	2.2369	0.8721	1.1743	208	204	4
125	C <sub>10</sub> H <sub>20</sub> O	CC(=O)CCCCCCCC	1.3507	1.3339	2.2541	0.9030	1.2119	210	203	7
126	C <sub>10</sub> H <sub>20</sub> O	CCC(=O)CCCCCCC	1.3780	1.3594	2.2541	0.9030	1.2273	211	203	8
127	C <sub>10</sub> H <sub>20</sub> O	CCCC(=O)CCCCC	1.3997	1.3795	2.2541	0.9030	1.2199	207	201	6
Esters										
128	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	COC=O	1.2202	1.0902	1.4848	1.0986	1.2255	32	30	2
129	C <sub>3</sub> H <sub>6</sub> O <sub>2</sub>	CCOC=O	1.2462	1.1603	1.6333	1.1148	1.3468	54	55	-1
130	C <sub>3</sub> H <sub>6</sub> O <sub>2</sub>	COC(C)=O	1.3694	1.2641	1.6647	1.1148	1.2795	58	53	5
131	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	C=CCOC=O	1.3140	1.2057	1.7627	1.1915	1.4474	83	78	5
132	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	CC(=O)OC=C	1.4171	1.2884	1.7904	1.1915	1.3918	72	78	-6
133	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	COC(=O)C=C	1.4614	1.3228	1.7904	1.1915	1.3918	85	79	6
134	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	CCOC=O	1.2669	1.2057	1.7627	1.1291	1.4308	81	79	2
135	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	CC(C)OC=O	1.3566	1.2884	1.7904	1.0818	1.2689	68	76	-8
136	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	CCOC(C)=O	1.3623	1.2884	1.7904	1.0818	1.2689	77	77	0
137	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	CCC(=O)OC	1.4014	1.3228	1.7904	1.1291	1.3825	80	77	3
138	C <sub>5</sub> H <sub>8</sub> O <sub>2</sub>	CC(=O)OCC=C	1.3943	1.3024	1.9020	1.1960	1.4639	104	99	5
139	C <sub>5</sub> H <sub>8</sub> O <sub>2</sub>	CCOC(=O)C=C	1.4450	1.3434	1.9020	1.1960	1.4639	101	99	2
140	C <sub>5</sub> H <sub>8</sub> O <sub>2</sub>	COC(=O)C=CC	1.4898	1.3434	1.9020	1.1550	1.4207	120	118	2
141	C <sub>5</sub> H <sub>8</sub> O <sub>2</sub>	COC(=O)C(C)=C	1.5362	1.4217	1.9260	1.1960	1.4207	101	97	4
142	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CCCCOC=O	1.2837	1.2375	1.8773	1.1039	1.4398	107	106	1
143	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CC(C)COC=O	1.3520	1.3024	1.9020	1.1039	1.3712	98	100	-2
144	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CCOC(C)=O	1.3569	1.3024	1.9020	1.1039	1.3712	102	101	1



Table 1. (Continued)

	formula	SMILES	$J_y$	$J$	$s_0$	IC <sub>1</sub>	IC <sub>2</sub>	expt	calc	expt-calc
Esters (Continued)										
145	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CCC(C)OC=O	1.3965	1.3434	1.9020	1.1039	1.4004	97	102	-5
146	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CCCC(=O)OC	1.4019	1.3434	1.9020	1.1039	1.4004	102	103	-1
147	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CCOC(=O)CC	1.4034	1.3434	1.9020	1.1039	1.3712	99	100	-1
148	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	CC(C)OC(C)=O	1.4337	1.3745	1.9260	1.0487	1.1915	90	97	-7
149	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CC(=O)OCCC=C	1.3818	1.3126	2.0023	1.2012	1.5200	127	118	9
150	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CCCOC(=O)C=C	1.4284	1.3512	2.0023	1.2012	1.5200	123	118	5
151	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CCOC(=O)CC=C	1.4446	1.3661	2.0023	1.2012	1.5200	119	118	1
152	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CCOC(=O)C=CC	1.4807	1.3661	2.0023	1.1320	1.4076	136	134	2
153	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CC(OC(C)=O)C=C	1.4909	1.4107	2.0241	1.1672	1.4344	113	119	-6
154	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CC(C)OC(=O)C=C	1.4938	1.4107	2.0241	1.1672	1.4076	110	117	-7
155	C <sub>6</sub> H <sub>10</sub> O <sub>2</sub>	CCOC(=O)C(C)=C	1.5162	1.4282	2.0241	1.1672	1.4076	118	117	1
156	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCCCCOC=O	1.2975	1.2613	1.9800	1.0760	1.4372	132	131	1
157	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCCCOC(C)=O	1.3545	1.3126	2.0023	1.0889	1.3977	127	126	1
158	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCCCC(=O)OC	1.3960	1.3512	2.0023	1.0760	1.4039	130	129	1
159	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCCCOC(=O)CC	1.3980	1.3512	2.0023	1.0889	1.3727	122	122	0
160	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CC(C)COC(C)=O	1.4124	1.3681	2.0241	1.0760	1.3070	117	121	-4
161	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCCC(=O)OCC	1.4142	1.3661	2.0023	1.0889	1.3977	120	125	-5
162	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCC(C)OC(C)=O	1.4579	1.4107	2.0241	1.0760	1.3337	112	122	-10
163	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CCC(=O)OC(C)C	1.4605	1.4107	2.0241	1.0760	1.3070	110	120	-10
164	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	CC(=O)OC(C)(C)C	1.5295	1.4791	2.0521	1.0192	1.0760	98	110	-12
165	C <sub>7</sub> H <sub>10</sub> O <sub>2</sub>	CC=CC(=O)OC(C)=C	1.5486	1.4233	2.1134	1.1906	1.4401	137	146	-9
166	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CC(=O)OCCCC=C	1.3750	1.3210	2.0935	1.1783	1.5248	151	140	11
167	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CCOC(=O)CCC=C	1.4367	1.3750	2.0935	1.1783	1.5248	145	139	6
168	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CCCCC(=O)OC=C	1.4370	1.3750	2.0935	1.1369	1.4660	133	143	-10
169	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CC(C)COC(=O)C=C	1.4651	1.4014	2.1134	1.1783	1.4743	132	136	-4
170	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CC(CC=C)OC(C)=O	1.4864	1.4233	2.1134	1.1783	1.4958	134	137	-3
171	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CCOC(=O)C=CCC	1.4877	1.3750	2.0935	1.1484	1.4743	158	157	1
172	C <sub>7</sub> H <sub>12</sub> O <sub>2</sub>	CC=CC(=O)OC(C)C	1.5200	1.4233	2.1134	1.1059	1.3463	146	150	-4
173	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCCCOC=O	1.3086	1.2795	2.0732	1.0501	1.4107	156	153	3
174	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCCCOC(C)=O	1.3543	1.3210	2.0935	1.0688	1.4081	149	150	-1
175	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCCC(=O)OC	1.3903	1.3551	2.0935	1.0501	1.3813	151	151	0
176	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCCOC(=O)CC	1.3923	1.3551	2.0935	1.0688	1.3865	146	147	-1
177	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCCC(=O)OCC	1.4137	1.3750	2.0935	1.0688	1.4081	145	148	-3
178	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCCC(=O)CCC	1.4144	1.3750	2.0935	1.0688	1.3865	143	146	-3
179	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCC(=O)OCC(C)C	1.4405	1.4014	2.1134	1.0983	1.3918	137	141	-4
180	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCCC(=O)OC(C)C	1.4649	1.4233	2.1134	1.0688	1.3476	131	143	-12
181	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CCC(C)OC(=O)CC	1.4816	1.4405	2.1134	1.0688	1.3476	132	142	-10
182	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CC(C)C(C)OC(C)=O	1.5162	1.4764	2.1329	1.0501	1.2714	129	139	-10
183	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	CC(C)OC(=O)C(C)C	1.5202	1.4764	2.1329	1.0501	1.2467	134	138	-4
184	C <sub>8</sub> H <sub>12</sub> O <sub>2</sub>	CCOC(=O)C=CC=CC	1.5271	1.3783	2.1770	1.1148	1.4372	196	191	5
185	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	CC(C)C=CCOC(C)=O	1.4619	1.3633	2.1953	1.1259	1.4221	172	176	-4
186	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	CCOC(=O)CC=CCC	1.4683	1.3783	2.1770	1.1359	1.4713	167	171	-4
187	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	CCCC=CC(=O)OCC	1.4827	1.3783	2.1770	1.1359	1.4904	175	178	-3
188	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	COC(=O)C=CC=CC	1.5300	1.3551	2.0935	1.1243	1.4401	180	179	1
189	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	CCC(OC(C)=O)C=CC	1.5607	1.4796	2.1953	1.1259	1.4422	156	165	-9
190	C <sub>8</sub> H <sub>14</sub> O <sub>2</sub>	CC(=O)OC(C)(C)CC=C	1.5896	1.5317	2.2191	1.1525	1.4221	138	150	-12
191	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCCCOC(C)=O	1.3553	1.3279	2.1770	1.0487	1.3930	169	170	-1
192	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCCCCC(=O)OC	1.3860	1.3574	2.1770	1.0268	1.3543	172	170	2
193	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCCCOC(=O)CC	1.3878	1.3574	2.1770	1.0487	1.3930	169	169	0
194	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCCCC(=O)OCC	1.4100	1.3783	2.1770	1.0487	1.3930	167	168	-1
195	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCCC(=O)OCC	1.4110	1.3783	2.1770	1.0487	1.3737	165	166	-1
196	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCCC(=O)OCCC	1.4192	1.3858	2.1770	1.0487	1.3930	168	168	0
197	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCCC(=O)OCC(C)C	1.4509	1.4168	2.1953	1.0916	1.4212	157	162	-5
198	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CC(C)COC(=O)C(C)C	1.4943	1.4588	2.2133	1.0487	1.2765	147	158	-11
199	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	CCC(C)C(=O)OC(C)C	1.5422	1.5047	2.2133	1.0753	1.3589	144	158	-14
200	C <sub>9</sub> H <sub>18</sub> O <sub>2</sub>	CCCCCCCCOC=O	1.3263	1.3059	2.2369	1.0061	1.3522	178	190	-12

<sup>a</sup> The last column indicates the difference between experimental and calculated NBPs (rounded off to integer values). Values of indices represent  $\ln(\text{index} + 1)$ .

equations. An interesting feature of these three equations is that the two pairs of related indices ( $J_y$  and  $J$ ; IC<sub>2</sub> and IC<sub>1</sub>) have opposite signs.

Whereas the experimental NBPs reported in Table 1 vary between 21 and 211 °C, none of the five TIs has such a wide range of variation. Since all values of TIs in Table 1 are of comparable magnitude, varying within narrow ranges for all 200 compounds, while NBPs have a much wider range of variation, one is entitled to compare the weights of the five indices by noting that the coefficient of  $J$  has the largest

absolute value and that of IC<sub>2</sub> the lowest one.

A few comments are necessary with regard to the data presented in Table 1. Whereas no degeneracy can be observed for  $J_y$ , all other descriptors do present degeneracies; i.e., the same value of the index can be seen for more than one compound. Among these last four indices, the highest degeneracy is observed for  $s_0$  followed by IC<sub>1</sub> and IC<sub>2</sub>, and the lowest degeneracy is for  $J$ . This last type of degeneracy is due to the fact that index  $J$  is the same for compounds sharing the same skeleton but differing in the position of

**Table 2.** Intercorrelation Matrices between the Five Molecular Descriptors for the Three Correlations 1–3

	$J_y$	$J$	$s_0$	$IC_1$	$IC_2$
Aldehydes/Ketones: Pearson Correlation Coefficients/Prob > $ R $ under $H_0$ ; $\rho = 0/N = 127$					
$J_y$	1.00000	0.81149	0.42796	0.26303	-0.10978
	0.0	0.0001	0.0001	0.0028	0.2192
$J$	0.81149	1.00000	0.76569	-0.19316	-0.17974
	0.0001	0.0	0.0001	0.0296	0.0432
$s_0$	0.42796	0.76569	1.00000	-0.44314	-0.01233
	0.0001	0.0001	0.0	0.0001	0.8906
$IC_1$	0.26303	-0.19316	-0.44314	1.00000	0.58452
	0.0028	0.0296	0.0001	0.0	0.0001
$IC_2$	-0.10978	-0.17974	-0.01233	0.58452	1.00000
	0.2192	0.0432	0.8906	0.0001	0.0
Esters: Pearson Correlation Coefficients/Prob > $ R $ under $H_0$ ; $\rho = 0/N = 73$					
$J_y$	1.00000	0.89628	0.55472	0.20528	0.04131
	0.0	0.0001	0.0001	0.0815	0.7286
$J$	0.89628	1.00000	0.74339	-0.09106	-0.05132
	0.0001	0.0	0.0001	0.4436	0.6663
$s_0$	0.55472	0.74339	1.00000	-0.28846	0.20875
	0.0001	0.0001	0.0	0.0133	0.0763
$IC_1$	0.20528	-0.09106	-0.28846	1.00000	0.65543
	0.0815	0.4436	0.0133	0.0	0.0001
$IC_2$	0.04131	-0.05132	0.20875	0.65543	1.00000
	0.7286	0.6663	0.0763	0.0001	0.0
Aldehydes/Ketones/Esters: Pearson Correlation Coefficients/Prob > $ R $ under $H_0$ ; $\rho = 0/N = 200$					
$J_y$	1.00000	0.82892	0.44863	0.15981	-0.09061
	0.0	0.0001	0.0001	0.0238	0.2020
$J$	0.82892	1.00000	0.75274	-0.13560	-0.13338
	0.0001	0.0	0.0001	0.0556	0.0597
$s_0$	0.44863	0.75274	1.00000	-0.24734	0.09797
	0.0001	0.0001	0.0	0.0004	0.1675
$IC_1$	0.15981	-0.13560	-0.24734	1.00000	0.68665
	0.0238	0.0556	0.0004	0.0	0.0001
$IC_2$	-0.09061	-0.13338	0.09797	0.68665	1.00000
	0.2020	0.0597	0.1675	0.0001	0.0

the multiple bond(s) or in the nature of heteroatoms; only  $J_y$  removes this type of degeneracy.

Trends in NBPs with structural details parallel the trends that are familiar for other types of compounds, such as acyclic hydrocarbons, namely:

(a) The NBP decreases with increasing molecular branching.

(b) Unsaturation increases the NBP, especially when the carbon-carbon multiple bond is conjugated with the carbonyl bond. This latter effect is undoubtedly due to the enhanced polarity of the molecule. Probably, with an extra indicator variable denoting the presence or absence of such conjugation, the correlation would be improved, but as stated at the beginning we did not wish to include indicator variables and thus increase the number of descriptors for a small gain in statistical parameters.

(c) On comparing aldehydes or ketones with the same skeleton but differing in the type of unsaturation, one sees that gradual depletion of hydrogen atoms raises slightly the NBP, as stated above, i.e., a stepwise increase on replacing a single C-C bond by a double C=C bond and then again by a triple bond.

(d) Generally, aldehydes have lower NBPs than isomeric ketones, despite the fact that the skeleton of a ketone is more branched; the lower polarity of aldehydes probably accounts for this fact.

(e) With isomeric ketones, the NBP decreases when the centricity of the carbonyl group increases, in agreement with the increase in molecular branching.

(f) An ester has a similar NBP as an aldehyde or ketone with a similar skeleton and the same number of carbon atoms, despite the presence of an extra oxygen atom. This is probably due to the lower polarity of the esteric carbonyl group.

## CONCLUSIONS

The present study indicates that five topological indices afford reasonable correlations between chemical structure and NBPs of 200 acyclic carbonyl compounds (aldehydes, ketones, and esters). Despite some degeneracy of four of these indices, a satisfactory discrimination is achieved. This discrimination and the corresponding correlation originates in the positive and negative coefficients of congeneric (but not intercorrelated) descriptors, namely,  $J_y$  and  $J$ , as well as  $IC_1$  and  $IC_2$ .

The two separate correlations for 127 saturated or unsaturated carbonyl compounds on one hand, as well as for 73 saturated and unsaturated carbalkoxy compounds, on the other hand, gave slightly better standard deviations than the combined correlation for all 200 carbonyl compounds.

Although no previous study was dedicated exclusively to NBPs of carbonyl compounds, such molecules were included in more diverse sets of substances, such as industrially important chemicals.<sup>16–18</sup> Naturally, the descriptors were different in these papers, but comparable statistical results were obtained.

## ACKNOWLEDGMENT

This is contribution no. 248 from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this paper was supported in part by Grants F49620-94-1-0401 and F49620-96-1-0330 from the United States Air Force. The assistance of Mr. Jason Dagit in handling the Smiles codes is acknowledged.

## REFERENCES AND NOTES

- Balaban, A. T.; Feroiu, V. Correlation between structure and critical data or vapor pressures of alkanes by means of topological indices. *Rep. Mol. Theory* **1990**, *1*, 133–139.
- Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. Topological indices for structure-activity correlations. *Top. Curr. Chem.* **1983**, *114*, 21–55.
- Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, FL, 1992; pp 225–274.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Wiley and Research Studies Press: New York and Letchworth, U.K., 1986.
- Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures*; Research Studies Press-Wiley: Chichester, U.K., 1993.
- Balaban, A. T., Ed. *From Chemical Topology to Three-Dimensional Geometry*; Plenum Press: New York, 1997.
- Balaban, A. T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199–206.
- Devillers, J.; Balaban, A. T., Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: Reading U.K.; Chapters by Balaban, A. T. and Ivanciuc, O.; Nikolic, S.; Trinajstić, N. and Mihalic, Z.; Bonchev, D.; Estrada, E.; Basak, S. C., in press.
- Balaban, A. T.; Balaban, T. S. Correlations using topological indices based on real graph invariants. *J. Chim. Phys.-Chim. Biol.* **1992**, *89*, 1735–1745.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- Randic, M. Topological indices. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: Chichester, U.K., 1998; pp 3018–3032.

- (12) Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-assisted rational design of immunosuppressive compounds. *Nature Biotechnol.* **1998**, *16*, 748–752.
- (13) Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
- (14) Seybold, P. G.; May, M.; Bagal, U. A. Molecular structure–property relationships. *J. Chem. Educ.* **1987**, *64*, 575–581.
- (15) Zefirov, N. S.; Palyulin, V. A.; Radchenko, E. V. Problem of generation of structures with definite properties. Solution of inverse problem for Balaban centric index. *Dokl. Akad. Nauk SSSR* **1991**, *316*, 921–924. Baskin, I. I.; Gordeeva, E. V.; Devdariani, R. O.; Zefirov, N. S.; Palyulin, V. A.; Stankevich, M. I. Methodology for solving the inverse problem of structure–property relationships for the case of topological indices. *Dokl. Akad. Nauk SSSR* **1989**, *307*, 613–617. Skvortsova, M. L.; Baskin, I. I.; Slovokhotova, O. L.; Palyulin, V. A.; Zefirov, N. S. Inverse problem in QSAR/QSPR studies for the case of topological indices characterizing molecular shape (Kier indices). *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 630–634. Skvortsova, M. L.; Stankevich, I. V.; Zefirov, N. S. Generation of molecular structures of polycondensed benzenoid hydrocarbons from Randic index. *Zh. Strukt. Khim.* **1992**, *33* (3), 99–104. Skvortsova, M. L.; Stankevich, I. V.; Zefirov, N. S. Topological properties of catacondensed benzenoid hydrocarbons: Randic index and its relation to chemical structure. In *Proceedings of the Conference "Molecular Graphs in Chemical Studies"*; Kalinin State University: Kalinin, USSR, 1990; p 84 (in Russian); Gordeeva, E. V.; Molchanova, M. S.; Zefirov, N. S. General methodology and computer program for the exhaustive restoring of chemical structures by molecular connectivity indices. Solution to the inverse problem in QSAR/QSPR. *Tetrahedron Comput. Methodol.* **1990**, *3*, 389–415.
- (16) Egolf, L. M.; Wessel, M. D.; Jurs, P. C. Prediction of boiling points and critical temperatures of industrially important organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947–956.
- (17) Wessel, M. D.; Jurs, P. C. Prediction of normal boiling points for a diverse set of industrially important organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 841–850.
- (18) Egolf, L. M.; Jurs, P. C. Estimation of autoignition temperatures of hydrocarbons, alcohols, and esters from molecular structure. *Ind. Eng. Chem. Res.* **1992**, *31*, 1792. Wessel, M. D.; Jurs, P. C. Prediction of normal boiling points of hydrocarbons from molecular structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 68–76.
- (19) Turner, B. E.; Costello, C. L.; Jurs, P. C. Prediction of critical temperatures and pressures of industrially important organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 639–645.
- (20) Stanton, D. T.; Jurs, P. C. Computer assisted prediction of normal boiling points of furans, tetrahydrofurans, and thiophenes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301–310.
- (21) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple organics. *J. Phys. Chem.* **1996**, *100*, 10400–10407.
- (22) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal boiling points for organic compounds: Correlation and prediction by a quantitative structure–property relationship. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28–41.
- (23) Murray, J. S.; Lane, P.; Politzer, P. Special relationships of fluorinated methane/ethane boiling points and heats of vaporization to molecular properties. *THEOCHEM* **1995**, *342*, 15–21.
- (24) Rücker, G.; Rücker, C. Symmetry-aided computation of the detour matrix and the detour index. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 710–714.
- (25) Lukovits, I.; Linert, W. Polarity-numbers of cycle-containing structures. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 715–719.
- (26) Trinajstić, N.; Nikolic, S.; Lucic, B. The detour matrix in chemistry. *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 631–638.
- (27) Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. Charge indexes. New topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.
- (28) Liu, S.; Cao, C.; Li, Z. Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance-edge (MDE) vector. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 387–394.
- (29) Hall, L. H.; Kier, L. B. Electrotopylogical state indices for atom types: A novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039–1045.
- (30) Wang, S.; Milne, G. W. A. Graph theory and group contributions in the estimation of boiling points. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1242–1250.
- (31) Gakh, A. A.; Gakh, E. G.; Sumpter, B. G.; Noid, D. W. Neural network-graph theory approach to the prediction of the physical properties of organic compounds. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 832–839.
- (32) Mihalic, Z.; Nikolic, S.; Trinajstić, N. Comparative study of molecular descriptors derived from the distance matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28–37.
- (33) Randic, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (34) Basak, S. C.; Gute, B. D.; Grunwald, G. D. A comparative study of topological and geometrical parameters in estimating normal boiling point and octanol/water partition coefficient. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054–1060.
- (35) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative structure–property relationship study of normal boiling points for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program. *Tetrahedron* **1998**, *54*, 9129–9142.
- (36) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Design of topological indices. 10. Parameters based on electronegativity and covalent radius for the computation of molecular graph descriptors for heteroatom-containing molecules. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 395–401.
- (37) Balaban, A. T.; Kier, L. B.; Joshi, N. Correlations between chemical structure and normal boiling points of halogenated alkanes C<sub>1</sub>–C<sub>4</sub>. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 233–237.
- (38) Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. D. Correlation between structure and normal boiling points of haloalkanes C<sub>1</sub>–C<sub>4</sub> using neural networks. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118–1121.
- (39) Balaban, A. T.; Joshi, N.; Kier, L. B.; Hall, L. H. Correlations between chemical structure and normal boiling points of acyclic ethers, peroxides, acetals, and their sulfur analogues. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 237–244.
- (40) Balaban, A. T.; Mills, D.; Basak, S. C. Normal boiling points of 1, $\omega$ -alkanedinitriles: The highest increment in a homologous series. *J. Chem. Inf. Comput. Sci.*, submitted for publication.
- (41) Dreisbach, R. R., Ed. *Physical Properties of Organic Compounds—II*; Advances in Chemistry Series No. 22; American Chemical Society: Washington, DC, 1959.
- (42) Weast, R. C.; Astle, M. J. *Handbook of Data on Organic Compounds*; CRC Press: Boca Raton, FL, 1985.
- (43) Daubert, T. E.; Danner, R. P., Eds. Design Institute for Physical Property Data (DIPR). *Physical and Thermodynamic Properties of Pure Chemicals: Data Compilation*; Hemisphere: New York, 1989; Vols. 1–4.
- (44) Weininger, D.; Weininger, J. L. Chemical structures and computers. In *Comprehensive Medicinal Chemistry*; Hansch, C., Ed.; Pergamon Press: Oxford, U.K., 1990; Vol. 4, pp 59–83.
- (45) Basak, S. C.; Harris, D. K.; Magnuson, V. R. "POLLY 2.3", Copyright of the University of Minnesota, 1988.
- (46) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA Manual*; University of Florida: Gainesville, FL, 1995.
- (47) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR study. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (48) *SAS/STAT User's Guide*, Release 6.03 ed.; SAS Institute, Inc.: Cary, NC, 1988.
- (49) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *80*, 399–404.
- (50) Balaban, A. T. Chemical graphs. 48. Topological index *J* for heteroatom-containing molecules taking into account periodicities of element properties. *Math. Chem. (MATCH)* **1986**, *21*, 115–122.
- (51) Balaban, A. T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199–206.
- (52) Balaban, A. T.; Filip, P. Computer program for topological index *J* (average distance sum connectivity). *Math. Chem. (MATCH)* **1984**, *16*, 163–190.
- (53) Basak, S. C.; Magnuson, V. R. Determining structural similarity of chemicals using graph-theoretic indices. *Discrete Appl. Math.* **1988**, *19*, 17–44.

C1990024T