J Math Chem (2010) 47:355–369
DOI 10.1007/s10910-009-9574-9

ORIGINAL PAPER

Use of the international chemical identifier for constructing QSPR-model of normal boiling points of acyclic carbonyl substances

Andrey A. Toropov · Alla P. Toropova · Emilio Benfenati · Danuta Leszczynska · Jerzy Leszczynski

Received: 29 October 2008 / Accepted: 10 July 2009 / Published online: 26 July 2009

Abstract Optimal descriptors calculated with international chemical identifier have been used to construct one-variable model of the normal boiling points of acyclic carbonyl substances. Attempts to calculate the model for three splits into training and test sets gave stable results. Statistical quality of the model is $n = 150$, $r^2 = 0.9825$, $s = 4.96\,^{\circ}\mathrm{C}$, $F = 8{,}312$ (training set) and $n = 50$, $r^2 = 0.9791$, $s = 4.68\,^{\circ}\mathrm{C}$, $F = 2{,}249$ (test set).

Keywords QSPR · InChI · Normal boiling point · Acyclic carbonyl compounds

1 Introduction

Modern technology that offers an easy access to Internet has transformed the way that scientists carry out research and share data. In order to take an advantage of existing resources new approaches are being currently developed. In some cases they emerge as a creative combination of the existing procedures. Simplified molecular input line entry system (SMILES [1–6]) combined with the international chemical identifier (InChI [7,8]) become widely used methodology to elucidate molecular structures in databases which are available via internet [9,10].

Electronic supplementary material The online version of this article (doi:10.1007/s10910-009-9574-9) contains supplementary material, which is available to authorized users.

A. A. Toropov (✉) · A. P. Toropova · E. Benfenati
Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy
e-mail: aatoropov@yahoo.com

D. Leszczynska · J. Leszczynski
Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry,
Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA

A logical step in extending research capabilities is to merge these two approaches. This results in development of the quantitative structure-property/activity relationships (QSPR/QSAR) which are based on the SMILES method augmented by the InChI obtained (i.e., directly from internet databases) molecular characteristics. Though the SMILES-based QSPR/QSAR have been already established [1–6], the InChI-based QSPR/QSAR method probably represents both novelty and perspective approach.

This paper reports the using of the proposed approach in order to predict normal boiling points of acyclic carbonyl compounds.

2 Methods

Figure 1 shows an example of the InChI. There are three layers for the InChI: formula, connectivity, and hydrogen atoms.

The formula is provided using the following InChI-attributes:

Cdd, i.e., C1, C2, …, C10, …C50;
Hdd, i.e., H1, H2, …H50;
Odd i.e., O and O2.

The connectivity is evaluated as follows:

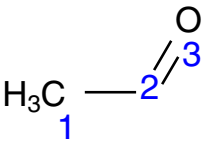
dd, i.e., 1, 2, …11;
−dd, i.e., −1, −2, …−11.

We proposed a new algorithm for building the InChI-based model. It includes following steps:

1. Definition of the split (i.e., list of the training set and list of the test set);
2. Definition of the total list of the InChI attributes;
3. Calculation of the correlation weights for the InChI attributes which give maximum of the correlation coefficient between the DCW and normal boiling points. The DCW is calculated as

$$\mathrm{DCW(InChI)} = \Sigma W(I_k) \tag{1}$$

where the I_k is InChI attribute and $W(I_k)$ is the correlation weight for the I_k.



InChI=1/C2H4O/c1-2-3/h2H,1H3

Description	Layers
Formula	C2H4O
Connectivity (c)	1-2-3
Hydrogen atoms (h)	2H,1H3

Fig. 1 Example of the InChI layers (CH_3–CH=O)

4. By calculating the $W(I_k)$ which produce maximum of the correlation coefficient for the training set (the Monte Carlo optimization), one can apply the data from the training set to develop a model for the normal boiling points

$$\mathrm{BP} = C_0 + C_1 \times \mathrm{DCW(InChI)} \tag{2}$$

5. Predictive potential of the Eq. 2 can be estimated using the data from the test set.

3 Results

Table 1 shows the statistical quality of the InChI-based models. Three different splits into the training and test set have been examined. These splits have been done randomly, but in the manner that provides representation of each InChI attribute in the training sets. The total list of the ID and structures of acyclic carbonyl compounds as well as lists of the test sets for three splits into training and test sets are presented in the "Electronic supplementary materials".

Table 2 contains the numerical data on the correlation weights for calculation of the DCW with Eq. 1 (split 1).

One-variable model obtained in the first run of the Monte Carlo optimization is presented by Eq. 3:

$$\mathrm{NBP} = -187.003(\pm 0.2734) + 23.8641(\pm 0.0195) \times \mathrm{DCW(InChI)} \tag{3}$$

Table 1 Statistical quality of the models

Probe	Training set, $n = 150$			Test set, $n = 50$		
	r^2	s (°C)	F	r^2	s (°C)	F
Split 1						
1	0.9825	4.960	8,312	0.9791	4.684	2,249
2	0.9825	4.958	8,317	0.9791	4.681	2,249
3	0.9825	4.958	8,319	0.9789	4.702	2,232
Average	0.9825	4.959	8,316	0.9791	4.689	2,243
Split 2						
1	0.9817	4.957	7,923	0.9795	5.006	2,295
2	0.9816	4.960	7,913	0.9794	5.024	2,280
3	0.9817	4.953	7,933	0.9794	5.020	2,281
Average	0.9817	4.957	7,923	0.9794	5.017	2,285
Split 3						
1	0.9830	4.915	8,553	0.9743	4.927	1,819
2	0.9830	4.915	8,552	0.9742	4.935	1,813
3	0.9830	4.915	8,553	0.9743	4.930	1,817
Average	0.9830	4.915	8,553	0.9743	4.931	1,816

Table 2 Correlation weights of InChI attributes (I_k) for calculation with Eq. 1 (split 1)

I_k	CW (I_k) in probe 1	CW (I_k) in probe 2	CW (I_k) in probe 3
(10	−0.5004076	−0.4546696	−0.5458930
(11	1.1276483	1.1889077	1.2635092
(2	0.1164717	0.1406312	0.0588950
(3	0.0168812	0.0394082	−0.0304492
(4	0.1309041	0.1447433	0.0718253
(5	−0.1169965	−0.0888549	−0.1659703
(6	−0.2646353	−0.2258246	−0.3058773
(7	−0.1227086	−0.0962723	−0.1731667
(8	−0.4172319	−0.3751912	−0.4611138
(9	−0.2194212	−0.1891073	−0.2694821
(	−0.8400031	−0.8702095	−0.8320413
+	0.9402182	0.9030358	0.9189753
,1	0.7221000	0.4750230	0.7384015
,2	−0.4494587	−0.4343066	−0.4580674
,3	−0.4736658	−0.4601273	−0.4720167
,4	−0.0832371	−0.0825969	−0.0871929
,5	−0.3442347	−0.3336629	−0.3414627
,6	−0.1353575	−0.1299887	−0.1308145
,7	−0.2678706	−0.2531487	−0.2619367
,8	−0.5601085	−0.5407882	−0.5553224
−10	−0.1339907	−0.1295610	−0.1310786
−11	−0.6530373	−0.6310113	−0.6518874
−2	0.1423675	0.1396336	0.1339927
−3	0.0607224	0.0517911	0.0570930
−4	0.1589245	0.1480897	0.1514930
−5	−0.0906264	−0.0905008	−0.0900246
−6	−0.0376432	−0.0373798	−0.0357214
−7	−0.0447664	−0.0464481	−0.0435986
−8	−0.1424821	−0.1391135	−0.1393596
−9	−0.0935953	−0.0950979	−0.0945712
−	0.6247169	0.5999432	0.6101379
/	−1.0236106	−1.1392569	−1.1133232
0	−0.6198414	−0.6176090	−0.6872184
1	0.7784196	0.8050595	0.8886746
3	0.4807324	0.4943048	0.5197012
4	0.6029377	0.6099258	0.6335558
5	0.3326880	0.3531548	0.3707008
6	0.4498228	0.4683370	0.4913319
7	0.4152888	0.4319783	0.4533021
8	0.3044078	0.3273508	0.3469306
9	0.3245043	0.3430782	0.3609223
C10	2.9218261	2.5801170	3.2068049
C2	−4.5756989	−4.4975921	−4.5351508
C3	−2.6876895	−2.6675707	−2.6442188
C4	−1.1852849	−1.2242095	−1.1529105
C5	0.5563385	0.4539203	0.5817122
C6	2.2115061	2.0514711	2.2360021
C7	3.8534752	3.6320125	3.8751468
C8	5.2016409	4.9291148	5.2195070
C9	7.3558224	7.0032593	7.3582822
H10	2.0013381	1.9435594	1.8570560
H12	1.5711523	1.5297015	1.4188281
H14	1.0772285	1.0527103	0.9183440

Table 2 continued

I_k	CW (I_k) in probe 1	CW (I_k) in probe 2	CW (I_k) in probe 3
H16	0.6080240	0.6056312	0.4495906
H18	−0.2739840	−0.2440846	−0.4243721
H20	3.6068468	3.5758117	2.9890739
H2	0.2127852	0.2094544	0.2117508
H3	0.7440124	0.7153664	0.7407645
H4	3.4991800	3.3847554	3.3517408
H6	3.2135102	3.1061154	3.0654361
H8	2.7047797	2.6199402	2.5609168
H	−0.0084371	−0.0009134	−0.0074401
O2	−0.0917285	−0.0643192	−0.0688208
O	−0.2019423	−0.1736037	−0.1778297
b3	−0.4078824	−0.2355347	−0.2812069
b4	0.0013793	0.1658007	0.1126245
b5	0.2325261	0.3818139	0.3447712
b6	0.3677442	0.5117197	0.4754316
b7	0.2914556	0.4418840	0.3970063
c1	8.6574103	10.1902324	9.9560082
h10	3.3201766	3.3723130	3.5223494
h1	2.3224274	2.1235452	2.3585457
h2	1.8214884	1.8579826	1.8283108
h3	1.4623508	1.5070376	1.4816411
h4	1.6761145	1.7162705	1.6901392
h5	1.5153762	1.5598367	1.5373881
h6	1.7349612	1.7708748	1.7544210
h7	1.7216089	1.7568473	1.7399343
h8	1.4709487	1.5187093	1.4893875
h9	1.4502817	1.4955153	1.4685271

Eight symbols at the beginning of the InChI string (i.e., the "InChI = 1/") are not used in the Monte Carlo optimization

Table 3 Example of the DCW calculation with Eq. 1 (split 1, probe 1) "InChI = 1/C2H4O/c1-2-3/h2H, 1H3"; DCW = 8.8139818

I_k	CW (I_k) in probe 1
C2	−4.5756989
H4	3.4991800
O	−0.2019423
/	−1.0236106
c1	8.6574103
−2	0.1423675
−3	0.0607224
/	−1.0236106
h2	1.8214884
H	−0.0084371
,1	0.7221000
H3	0.7440124

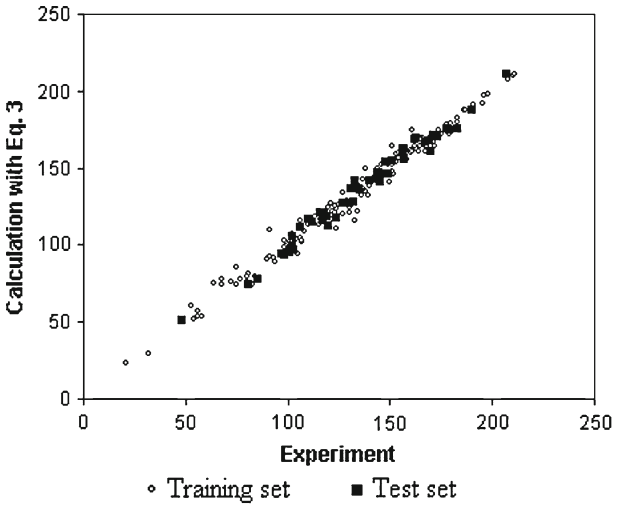


Fig. 2 Experimental versus calculated with Eq. 3 values of normal boiling points (°C)

$n = 150$, $r^2 = 0.9825$, $s = 4.96\,^{\circ}\mathrm{C}$, $F = 8{,}312$ (training set)
$n = 50$, $r^2 = 0.9791$, $s = 4.68\,^{\circ}\mathrm{C}$, $F = 2{,}249$ (test set)

Table 3 contains an example of the DCW(InChI) calculation. Table 4 shows experimental and calculated with Eq. 3 normal boiling points. Graphically the model is demonstrated in Fig. 2.

4 Discussion

Statistical quality of the model of normal boiling points of the acyclic carbonyl compounds described in [11] is the following: $n = 200$, $s = 6.93\,^{\circ}\mathrm{C}$, $r^2 = 0.9640$, $F = 984$. It can be compared with the results obtained using the proposed approach. The SMILES-based model for the acyclic carbonyl compounds described in [2] is statistically characterized by: $n = 100$, $r^2 = 0.9795$, $s = 5.35\,^{\circ}\mathrm{C}$, $F = 4{,}673$ (training set); $n = 100$, $r^2 = 0.9764$, $s = 5.38\,^{\circ}\mathrm{C}$, $F = 4{,}055$ (test set).

This comparison reveals the accuracy of the proposed method. We believe that the statistical quality of the model calculated with Eq. 3 is quite satisfactory.

5 Conclusions

We proposed new InChI-based QSPR approach. The optimal InChI-based descriptors which are calculated by the Monte Carlo optimization can be used as a tool for prediction of the normal boiling points of acyclic carbonyl substances. Thus InChI can be useful in the QSPR/QSAR analyses. This fact, taking into account the increase of number of databases available via internet with the InChI-based representation of molecular structure, is interesting from heuristic point of view.

Table 4 Experimental and calculated with Eq. 3 values of normal boiling points

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
Training set					
1	InChI = 1/C2H4O/c1-2-3/h2H, 1H3	8.8139818	21.000	23.335	−2.335
2	InChI = 1/C3H4/c1-3-2/h1H, 2H3	10.2333137	56.000	57.206	−1.206
3	InChI = 1/C3H4O/c1-2-3-4/h2-3H, 1H2	10.3904108	53.000	60.955	−7.955
4	InChI = 1/C4H4O/c1-2-3-4-5/h4H, 1H3	12.1273199	107.000	102.405	4.595
5	InChI = 1/C4H4O/c1-3-4(2)5/h1H, 2H3	11.1594896	84.000	79.308	4.692
6	InChI = 1/C4H6O/c1-2-3-4-5/h2-4H, 1H3/b3-2+	11.7970412	105.000	94.523	10.477
7	InChI = 1/C4H6O/c1-4(2)3-5/h3H, 1H2, 2H3	10.9453239	68.000	74.197	−6.197
8	InChI = 1/C4H6O/c1-4(2)3-5/h3H, 1H2, 2H3	10.9453239	75.000	74.197	0.803
9	InChI = 1/C4H8O/c1-4(2)3-5/h3-4H, 1-2H3	10.9745588	64.000	74.895	−10.895
10	InChI = 1/C5H6O/c1-3-5(2)4-6/h1, 4-5H, 2H3	12.4246352	91.000	109.500	−18.500
11	InChI = 1/C5H8O/c1-2-3-4-5-6/h2, 5H, 1, 3-4H2	11.9920704	99.000	99.177	−0.177
12	InChI = 1/C5H8O/c1-2-3-4-5-6/h2-3, 5H, 4H2, 1H3/b3-2+	12.6794022	106.000	115.580	−9.580
13	InChI = 1/C5H8O/c1-5(2)3-4-6/h3-4H, 1-2H3	12.6785390	133.000	115.559	17.441
14	InChI = 1/C5H8O/c1-4(2)5(3)6/h1H2, 2-3H3	11.9546359	98.000	98.284	−0.284
15	InChI = 1/C5H8O2/c1-3-5(7)4(2)6/h3H2, 1-2H3	12.3998083	108.000	108.907	−0.907
16	InChI = 1/C5H10O/c1-5(2)3-4-6/h4-5H, 3H2, 1-2H3	11.6784298	93.000	91.692	1.308
17	InChI = 1/C6H8O/c1-2-3-4-5-6-7/h2-6H, 1H3/b3-2+, 5-4+	15.1610323	174.000	174.801	−0.801
18	InChI = 1/C6H8O/c1-4-6(7)5(2)3/h1, 5H, 2-3H3	12.8980230	118.000	120.797	−2.797
19	InChI = 1/C6H8O/c1-3-6(4-2)5-7/h3-5H, 1H2, 2H3/b6-4+	14.1048046	144.000	149.595	−5.595
20	InChI = 1/C6H10O/c1-2-3-4-5-6-7/h3-4, 6H, 2, 5H2, 1H3/b4-3+	13.1514637	121.000	126.845	−5.845
21	InChI = 1/C6H10O/c1-2-3-4-5-6-7/h4-6H, 2-3H2, 1H3/b5-4+	14.0383233	146.000	148.009	−2.009
22	InChI = 1/C6H10O/c1-3-4-5-6(2)7/h4-5H, 3H2, 1-2H3/b5-4+	13.6369344	140.000	138.430	1.570
23	InChI = 1/C6H10O/c1-3-4-6(2)5-7/h4-5H, 3H2, 1-2H3/b6-4-	13.4199104	136.000	133.251	2.749
24	InChI = 1/C6H12O/c1-2-3-4-5-6-7/h6H, 2-5H2, 1H3	13.2452203	131.000	129.082	1.918
25	InChI = 1/C6H12O/c1-6(2)4-3-5-7/h5-6H, 3-4H2, 1-2H3	12.9338176	122.000	121.651	0.349
26	InChI = 1/C6H12O/c1-3-6(2)4-5-7/h5-6H, 3-4H2, 1-2H3	12.9338176	122.000	121.651	0.349

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
27	InChI=1/C6H12O/c1-5(2)6(3)4-7/h4-6H, 1-3H3	12.7184888	116.000	116.512	−0.512
28	InChI=1/C6H12O/c1-4-5(2)6(3)7/h5H, 4H2, 1-3H3	12.7409838	118.000	117.049	0.951
29	InChI=1/C6H12O/c1-5(7)6(2,3)4/h1-4H3	12.2184947	106.000	104.580	1.420
30	InChI=1/C7H12O/c1-2-3-4-5-6-7-8/h5-7H, 2-4H2, 1H3/b6-5+	14.9236324	166.000	169.136	−3.136
31	InChI=1/C7H12O/c1-3-4-5-6-7(2)8/h3-4H, 5-6H2, 1-2H3/b4-3+	14.4912965	153.000	158.819	−5.819
32	InChI=1/C7H12O/c1-3-5-6-7(8)4-2/h3H, 1, 4-6H2, 2H3	13.7711033	142.000	141.632	0.368
33	InChI=1/C7H12O/c1-4-6(2)5-7(3)8/h4, 6H, 1, 5H2, 2-3H3	13.4780786	138.000	134.639	3.361
34	InChI=1/C7H12O/c1-4-5-6(2)7(3)8/h4, 6H, 1, 5H2, 2-3H3	13.5148193	137.000	135.516	1.484
35	InChI=1/C7H14O/c1-3-4-5-6-7(2)8/h3-6H2, 1-2H3	14.2124106	151.000	152.163	−1.163
36	InChI=1/C7H14O/c1-3-4-7(2)5-6-8/h6-7H, 3-5H2, 1-2H3	13.8815931	144.000	144.269	−0.269
37	InChI=1/C7H14O/c1-4-5-7(8)6(2)3/h6H, 4-5H2, 1-3H3	13.6045518	135.000	137.657	−2.657
38	InChI=1/C7H14O/c1-5(2)7(8)6(3)4/h5-6H, 1-4H3	13.0068181	124.000	123.393	0.607
39	InChI=1/C9H14O/c1-7(2)5-9(10)6-8(3)4/h5-6H, 1-4H3	16.1531714	198.000	198.478	−0.478
40	InChI=1/C8H14O/c1-4-5-8(9)6-7(2)3/h4-5, 7H, 6H2, 1-3H3/b5-4+	14.9372666	170.000	169.461	0.539
41	InChI=1/C8H14O/c1-5-6(2)7(3)8(4)9/h5, 7H, 1-4H3/b6-5+	14.4051264	154.000	156.762	−2.762
42	InChI=1/C8H14O/c1-5-6(2)7(3)8(4)9/h5H2, 1-4H3/b7-6-	14.5554124	158.000	160.349	−2.349
43	InChI=1/C8H14O2/c1-3-5-7(9)8(10)6-4-2/h3-6H2, 1-2H3	14.6243201	168.000	161.993	6.007
44	InChI=1/C8H16O/c1-3-4-5-6-8(2)7-9/h7-8H, 3-6H2, 1-2H3	14.6456149	162.000	162.501	−0.501
45	InChI=1/C8H16O/c1-3-5-6-7-8(9)4-2/h3-7H2, 1-2H3	14.8878465	168.000	168.282	−0.282
46	InChI=1/C8H16O/c1-3-5-7-8(9)6-4-2/h3-7H2, 1-2H3	14.9312993	165.000	169.319	−4.319
47	InChI=1/C8H16O/c1-4-7(2)5-6-8(3)9/h7H, 4-6H2, 1-3H3	14.5943951	167.000	161.279	5.721
48	InChI=1/C8H16O/c1-3-5-6-8(4-2)7-9/h7-8H, 3-6H2, 1-2H3	14.6434903	163.000	162.451	0.549
49	InChI=1/C8H16O/c1-4-5-6-8(9)7(2)3/h7H, 4-6H2, 1-3H3	14.5510615	160.000	160.245	−0.245
50	InChI=1/C8H16O/c1-3-5-8(7-9)6-4-2/h7-8H, 3-6H2, 1-2H3	14.6209794	160.000	161.914	−1.914
51	InChI=1/C8H16O/c1-4-7(3)6-8(9)5-2/h7H, 4-6H2, 1-3H3	14.5638315	161.000	160.550	0.450
52	InChI=1/C8H16O/c1-6(2)5-8(9)7(3)4/h6-7H, 5H2, 1-4H3	13.9579686	147.000	146.091	0.909

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
53	InChI = 1/C8H16O/c1-6(2)7(9)8(3, 4)5/h6H, 1-5H3	13.6453488	135.000	138.631	−3.631
54	InChI = 1/C9H16O/c1-4-5-6-9(10)7-8(2)3/h5-6, 8H, 4, 7H2, 1-3H3/b6-5+	15.4901751	183.000	182.656	0.344
55	InChI = 1/C9H18O/c1-3-4-5-6-7-8-9(2)10/h3-8H2, 1-2H3	15.8767993	195.000	191.883	3.117
56	InChI = 1/C9H18O/c1-3-5-7-9(10)8-6-4-2/h3-8H2, 1-2H3	15.6905992	186.000	187.439	−1.439
57	InChI = 1/C9H18O/c1-4-5-6-7-9(10)8(2)3/h8H, 4-7H2, 1-3H3	15.1777047	183.000	175.199	7.801
58	InChI = 1/C9H18O/c1-4-6-9(10)7-8(3)5-2/h8H, 4-7H2, 1-3H3	15.1762290	161.000	175.164	−14.164
59	InChI = 1/C9H18O/c1-4-6-7-9(10)8(3)5-2/h8H, 4-7H2, 1-3H3	15.1630638	174.000	174.850	−0.850
60	InChI = 1/C9H18O/c1-8(2, 3)7(10)9(4, 5)6/h1-6H3	13.9750740	152.000	146.500	5.500
61	InChI = 1/C10H20O/c1-3-4-5-6-7-8-9-10(2)11/h3-9H2, 1-2H3	16.6367909	210.000	210.019	−0.019
62	InChI = 1/C10H20O/c1-3-5-6-7-8-9-10(11)4-2/h3-9H2, 1-2H3	16.6775090	211.000	210.991	0.009
63	InChI = 1/C2H4O2/c1-4-2-3/h2H, 1H3	9.0831201	32.000	29.757	2.243
64	InChI = 1/C3H6O2/c1-2-5-3-4/h3H, 2H2, 1H3	9.9990221	54.000	51.615	2.385
65	InChI = 1/C4H6O2/c1-3-6-4(2)5/h3H, 1H2, 2H3	11.0211990	72.000	76.008	−4.008
66	InChI = 1/C4H8O2/c1-2-3-6-4-5/h4H, 2-3H2, 1H3	11.2295391	81.000	80.980	0.020
67	InChI = 1/C4H8O2/c1-3-4(5)6-2/h3H2, 1-2H3	11.1764090	80.000	79.712	0.288
68	InChI = 1/C5H8O2/c1-3-4-7-5(2)6/h3H, 1, 4H2, 2H3	12.1902400	104.000	103.906	0.094
69	InChI = 1/C5H8O2/c1-3-5(6)7-4-2/h3H, 1, 4H2, 2H3	11.9617329	101.000	98.453	2.547
70	InChI = 1/C5H10O2/c1-3-4-7-5(2)6/h3-4H2, 1-2H3	12.3292233	102.000	107.223	−5.223
71	InChI = 1/C5H10O2/c1-3-4-5(6)7-2/h3-4H2, 1-2H3	12.1007162	102.000	101.770	0.230
72	InChI = 1/C5H10O2/c1-3-5(6)7-4-2/h3-4H2, 1-2H3	12.1007162	99.000	101.770	−2.770
73	InChI = 1/C6H10O2/c1-3-4-5-8-6(2)7/h3H, 1, 4-5H2, 2H3	12.8814467	127.000	120.401	6.599
74	InChI = 1/C6H10O2/c1-3-5-8-6(7)4-2/h4H, 2-3, 5H2, 1H3	12.9174733	123.000	121.261	1.739
75	InChI = 1/C6H10O2/c1-3-5-6(7)8-4-2/h3, 5H, 4H2, 1-2H3/b5-3+	13.3643092	136.000	131.924	4.076
76	InChI = 1/C6H10O2/c1-4-5(2)8-6(3)7/h4-5H, 1H2, 2-3H3	12.8022157	113.000	118.510	−5.510
77	InChI = 1/C6H10O2/c1-4-8-6(7)5(2)3/h2, 4H2, 1, 3H3	12.7807646	118.000	117.998	0.002
78	InChI = 1/C6H12O2/c1-3-4-5-8-6(2)7/h3-5H2, 1-2H3	13.1347611	127.000	126.446	0.554

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
79	InChI = 1/C6H12O2/c1-3-5-8-6(7)4-2/h3-5H2, 1-2H3	13.0666728	122.000	124.821	−2.821
80	InChI = 1/C6H12O2/c1-3-5-6(7)8-4-2/h3-5H2, 1-2H3	13.0695495	120.000	124.890	−4.890
81	InChI = 1/C6H12O2/c1-4-6(7)8-5(2)3/h5H, 4H2, 1-3H3	12.5939932	110.000	113.541	−3.541
82	InChI = 1/C7H10O2/c1-4-5-7(8)9-6(2)3/h4-5H, 2H2, 1, 3H3/b5-4+	13.8245154	137.000	142.907	−5.907
83	InChI = 1/C7H12O2/c1-3-4-5-6-9-7(2)8/h3H, 1, 4-6H2, 2H3	14.0394526	151.000	148.036	2.964
84	InChI = 1/C7H12O2/c1-4-7(8)9-5-6(2)3/h4, 6H, 1, 5H2, 2-3H3	13.2316936	132.000	128.759	3.241
85	InChI = 1/C7H12O2/c1-4-5-6(2)9-7(3)8/h4, 6H, 1, 5H2, 2-3H3	13.4894823	134.000	134.911	−0.911
86	InChI = 1/C7H12O2/c1-4-5-7(8)9-6(2)3/h4-6H, 1-3H3/b5-4+	14.2183743	146.000	152.306	−6.306
87	InChI = 1/C7H14O2/c1-3-4-5-6-9-7(2)8/h3-6H2, 1-2H3	14.2290292	149.000	152.560	−3.560
88	InChI = 1/C8H16O2/c1-3-4-5-6-7-8(9)10-2/h3-7H2, 1-2H3	14.7126253	151.000	164.101	−13.101
89	InChI = 1/C7H14O2/c1-3-5-6-9-7(8)4-2/h3-6H2, 1-2H3	13.9772984	146.000	146.553	−0.553
90	InChI = 1/C7H14O2/c1-4-7(8)9-5-6(2)3/h6H, 4-5H2, 1-3H3	13.5518039	137.000	136.399	0.601
91	InChI = 1/C7H14O2/c1-4-6(3)9-7(8)5-2/h6H, 4-5H2, 1-3H3	13.5371630	132.000	136.049	−4.049
92	InChI = 1/C7H14O2/c1-5(2)6(3)9-7(4)8/h5-6H, 1-4H3	3.2310873	129.000	128.745	0.255
93	InChI = 1/C7H14O2/c1-5(2)7(8)9-6(3)4/h5-6H, 1-4H3	12.9540702	134.000	122.134	11.866
94	InChI = 1/C8H14O2/c1-7(2)5-4-6-10-8(3)9/h4-5, 7H, 6H2, 1-3H3/b5-4+	14.9294124	172.000	169.274	2.726
95	InChI = 1/C7H10O2/c1-3-4-5-6-7(8)9-2/h3-6H, 1-2H3/b4-3+, 6-5+	15.3472158	180.000	179.244	0.756
96	InChI = 1/C8H14O2/c1-5-6-8(3,4)10-7(2)9/h5H, 1, 6H2, 2-4H3	14.0949492	138.000	149.360	−11.360
97	InChI = 1/C8H16O2/c1-3-4-5-6-7-10-8(2)9/h3-7H2, 1-2H3	14.9380860	169.000	169.481	−0.481
98	InChI = 1/C8H16O2/c1-3-5-7-10-8(9)6-4-2/h3-7H2, 1-2H3	14.9075224	165.000	168.752	−3.752
99	InChI = 1/C8H16O2/c1-6(2)5-10-8(9)7(3)4/h6-7H, 5H2, 1-4H3	13.9341917	147.000	145.524	1.476
100	InChI = 1/C9H18O2/c1-2-3-4-5-6-7-8-11-9-10/h9H, 2-8H2, 1H3	15.2949735	178.000	177.998	0.002
102	InChI = 1/C3H6O/c1-3(2)4/h1-2H3	10.0830035	56.000	53.619	2.381
104	InChI = 1/C4H8O/c1-3-4(2)5/h3H2, 1-2H3	11.0401611	80.000	76.461	3.539
106	InChI = 1/C5H8O/c1-2-3-4-5-6/h3-5H, 2H2, 1H3/b4-3+	12.4745454	124.000	110.691	13.309
108	InChI = 1/C5H8O/c1-3-4-5(2)6/h3-4H, 1-2H3/b4-3+	12.7247044	122.000	116.661	5.339

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
110	InChI=1/C5H10O/c1-2-3-4-5-6/h5H, 2-4H2, 1H3	12.0949709	103.000	101.633	1.367
112	InChI=1/C5H10O/c1-3-5(2)4-6/h4-5H, 3H2, 1-2H3	11.7024330	91.000	92.265	−1.265
114	InChI=1/C5H10O/c1-4(2)5(3)6/h4H, 1-3H3	11.5552180	94.000	88.752	5.248
115	InChI=1/C5H10O/c1-5(2, 3)4-6/h4H, 1-3H3	11.4379067	75.000	85.952	−10.952
116	InChI=1/C6H10O/c1-3-4-5-6(2)7/h3H, 1, 4-5H2, 2H3	12.9137149	130.000	121.171	8.829
117	InChI=1/C6H10O/c1-3-4-5-6(2)7/h3-4H, 5H2, 1-2H3/b4-3+	13.4728037	127.000	134.513	−7.513
119	InChI=1/C6H10O/c1-3-5-6(7)4-2/h3, 5H, 4H2, 1-2H3/b5-3+	13.3937007	139.000	132.626	6.374
122	InChI=1/C6H10O2/c1-3-5(7)6(8)4-2/h3-4H2, 1-2H3	13.1191227	130.000	126.073	3.927
125	InChI=1/C6H12O/c1-3-5-6(7)4-2/h3-5H2, 1-2H3	13.0989410	123.000	125.591	−2.591
128	InChI=1/C6H12O/c1-4-6(7)5(2)3/h5H, 4H2, 1-3H3	12.6026859	115.000	113.749	1.251
130	InChI=1/C7H12O/c1-6(2)4-5-7(3)8/h1, 4-5H2, 2-3H3	13.7371559	150.000	140.822	9.178
132	InChI=1/C7H14O/c1-2-3-4-5-6-7-8/h7H, 2-6H2, 1H3	14.2904145	153.000	154.025	−1.025
135	InChI=1/C7H14O/c1-6(2)4-5-7(3)8/h6H, 4-5H2, 1-3H3	13.8188876	144.000	142.772	1.228
138	InChI=1/C7H14O/c1-4-7(8)5-6(2)3/h6H, 4-5H2, 1-3H3	13.5404002	136.000	136.126	−0.126
141	InChI=1/C7H14O/c1-5-6(8)7(2, 3)4/h5H2, 1-4H3	13.1276736	125.000	126.277	−1.277
143	InChI=1/C8H14O/c1-7(2)5-4-6-8(3)9/h4, 6-7H, 5H2, 1-3H3/b6-4+	15.0579034	179.000	172.340	6.660
146	InChI=1/C8H14O/c1-4-7(3)6-8(9)5-2/h6H, 4-5H2, 1-3H3/b7-6-	14.8483239	167.000	167.339	−0.339
147	InChI=1/C8H16O/c1-2-3-4-5-6-7-8-9/h8H, 2-7H2, 1H3	14.8179969	171.000	166.615	4.385
150	InChI=1/C8H16O/c1-4-8(9)6-5-7(2)3/h7H, 4-6H2, 1-3H3	14.5784724	164.000	160.899	3.101
153	InChI=1/C8H16O/c1-4-6-8(9)7(3)5-2/h7H, 4-6H2, 1-3H3	14.5364206	154.000	159.896	−5.896
155	InChI=1/C9H18O/c1-2-3-4-5-6-7-8-9-10/h9H, 2-8H2, 1H3	15.8377970	191.000	190.952	0.048
157	InChI=1/C9H18O/c1-3-5-6-8-9(10)7-4-2/h3-8H2, 1-2H3	15.7037644	187.000	187.753	−0.753
160	InChI=1/C9H18O/c1-4-5-6-8(2)7-9(3)10/h8H, 4-7H2, 1-3H3	15.3860046	183.000	180.170	2.830
162	InChI=1/C9H18O/c1-7(2)5-9(10)6-8(3)4/h7-8H, 5-6H2, 1-4H3	14.7342600	165.000	164.617	0.383
163	InChI=1/C9H18O/c1-5-7(3)9(10)8(4)6-2/h7-8H, 5-6H2, 1-4H3	14.7297974	162.000	164.510	−2.510
164	InChI=1/C10H20O/c1-2-3-4-5-6-7-8-9-10-11/h10H, 2-9H2, 1H3	16.5503759	208.000	207.957	0.043

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
166	InChI = 1/C3H6O2/c1-3(4)5-2/h1-2H3	10.0797676	58.000	53.542	4.458
167	InChI = 1/C4H6O2/c1-2-3-6-4-5/h2, 4H, 1, 3H2	10.9714644	83.000	74.821	8.179
169	InChI = 1/C4H8O2/c1-4(2)6-3-5/h3-4H, 1-2H3	11.1145855	68.000	78.237	−10.237
170	InChI = 1/C4H8O2/c1-3-6-4(2)5/h3H2, 1-2H3	11.1127317	77.000	78.192	−1.192
173	InChI = 1/C5H10O2/c1-2-3-4-7-5-6/h5H, 2-4H2, 1H3	12.1604183	107.000	103.194	3.806
176	InChI = 1/C5H10O2/c1-4(2)7-5(3)6/h4H, 1-3H3	11.6574062	90.000	91.191	−1.191
177	InChI = 1/C6H10O2/c1-3-5-6(7)8-4-2/h3H, 1, 4-5H2, 2H3	12.8162351	119.000	118.845	0.155
180	InChI = 1/C6H12O2/c1-3-4-5-6(7)8-2/h3-5H2, 1-2H3	13.0695495	130.000	124.890	5.110
181	InChI = 1/C6H12O2/c1-5(2)4-8-6(3)7/h5H, 4H2, 1-3H3	12.6652627	117.000	115.242	1.758
183	InChI = 1/C6H12O2/c1-5(7)8-6(2, 3)4/h1-4H3	12.1456504	98.000	102.842	−4.842
186	InChI = 1/C7H12O2/c1-3-5-6-7(8)9-4-2/h5-6H, 3-4H2, 1-2H3/b6-5+	14.3688811	158.000	155.897	2.103
187	InChI = 1/C7H14O2/c1-2-3-4-5-6-9-7-8/h7H, 2-6H2, 1H3	14.3070330	156.000	154.421	1.579
189	InChI = 1/C7H14O2/c1-3-5-7(8)9-6-4-2/h3-6H2, 1-2H3	13.9513848	143.000	145.934	−2.934
191	InChI = 1/C8H12O2/c1-3-5-6-7-8(9)10-4-2/h3, 5-7H, 4H2, 1-2H3/b5-3+, 7-6+	16.0844054	196.000	196.837	−0.837
193	InChI = 1/C8H14O2/c1-3-5-6-7-8(9)10-4-2/h6-7H, 3-5H2, 1-2H3/b7-6+	15.0521311	175.000	172.203	2.797
194	InChI = 1/C8H14O2/c1-4-6-8(5-2)10-7(3)9/h4, 6, 8H, 5H2, 1-3H3/b6-4+	14.3875646	156.000	156.343	−0.343
195	InChI = 1/C8H16O2/c1-3-4-5-6-7-8(9)10-2/h3-7H2, 1-2H3	14.7126253	172.000	164.101	7.899
197	InChI = 1/C8H16O2/c1-3-5-6-7-8(9)10-4-2/h3-7H2, 1-2H3	14.7126253	167.000	164.101	2.899
198	InChI = 1/C8H16O2/c1-3-5-6-8(9)10-7-4-2/h3-7H2, 1-2H3	14.7126253	168.000	164.101	3.899
200	InChI = 1/C8H16O2/c1-5-7(4)8(9)10-6(2)3/h6-7H, 5H2, 1-4H3	13.7820986	144.000	141.894	2.106
Test set					
101	InChI = 1/C3H6O/c1-2-3-4/h3H, 2H2, 1H3	9.9794346	48.000	51.147	−3.147
103	InChI = 1/C4H6O/c1-3-4(2)5/h3H, 1H2, 2H3	10.9486284	81.000	74.276	6.724
105	InChI = 1/C5H6O/c1-3-5(6)4-2/h1H, 4H2, 2H3	12.5269506	106.000	111.941	−5.941
107	InChI = 1/C5H8O/c1-3-5(6)4-2/h3H, 1, 4H2, 2H3	11.8802435	103.000	96.508	6.492

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
109	InChI = 1/C5H8O/c1-3-5(2)4-6/h3-4H, 1-2H3/b5-3+	12.9123985	117.000	121.140	−4.140
111	InChI = 1/C5H10O/c1-3-4-5(2)6/h3-4H2, 1-2H3	12.2637758	102.000	105.661	−3.661
113	InChI = 1/C5H10O/c1-3-5(6)4-2/h3-4H2, 1-2H3	12.0192268	102.000	99.825	2.175
118	InChI = 1/C6H10O/c1-5(2)4-6(3)7/h1, 4H2, 2-3H3	12.7716463	124.000	117.781	6.219
120	InChI = 1/C6H10O/c1-4-6(7)5(2)3/h2, 4H2, 1, 3H3	12.8130329	119.000	118.768	0.232
121	InChI = 1/C6H10O/c1-5(2)4-6(3)7/h4H, 1-3H3	13.1589070	130.000	127.022	2.978
123	InChI = 1/C6H12O/c1-3-4-5-6(2)7/h3-5H2, 1-2H3	13.1670294	127.000	127.216	−0.216
124	InChI = 1/C6H12O/c1-3-4-6(2)5-7/h5-6H, 3-4H2, 1-2H3	12.9131188	116.000	121.157	−5.157
126	InChI = 1/C6H12O/c1-5(2)4-6(3)7/h5H, 4H2, 1-3H3	12.6975310	117.000	116.012	0.988
127	InChI = 1/C6H12O/c1-3-6(4-2)5-7/h5-6H, 3-4H2, 1-2H3	12.9109943	118.000	121.106	−3.106
129	InChI = 1/C7H12O/c1-2-3-4-5-6-7-8/h4-5, 7H, 2-3, 6H2, 1H3/b5-4+	14.6514132	157.000	162.640	−5.640
131	InChI = 1/C7H12O/c1-4-5-7(8)6(2)3/h4-6H, 1-3H3/b5-4+	14.2711223	148.000	153.564	−5.564
133	InChI = 1/C7H14O/c1-3-7(2)5-4-6-8/h6-7H, 3-5H2, 1-2H3	13.8815931	144.000	144.269	−0.269
134	InChI = 1/C7H14O/c1-3-5-6-7(8)4-2/h3-6H2, 1-2H3	13.9606799	149.000	146.156	2.844
136	InChI = 1/C7H14O/c1-3-5-7(8)6-4-2/h3-6H2, 1-2H3	14.0041327	144.000	147.193	−3.193
137	InChI = 1/C7H14O/c1-4-6(2)5-7(3)8/h6H, 4-5H2, 1-3H3	13.7981888	140.000	142.278	−2.278
139	InChI = 1/C7H14O/c1-4-5-6(2)7(3)8/h6H, 4-5H2, 1-3H3	13.8349296	143.000	143.155	−0.155
140	InChI = 1/C7H14O/c1-4-6(3)7(8)5-2/h6H, 4-5H2, 1-3H3	13.5625000	135.000	136.654	−1.654
142	InChI = 1/C8H14O/c1-7(2)5-4-6-8(3)9/h4-5, 7H, 6H2, 1-3H3/b5-4+	14.9531893	163.000	169.841	−6.841
144	InChI = 1/C8H14O/c1-4-5-6-8(9)7(2)3/h4-5, 7H, 6H2, 1-3H3/b5-4+	14.9098557	162.000	168.807	−6.807
145	InChI = 1/C8H14O/c1-4-6-8(9)7(3)5-2/h4, 6-7H, 5H2, 1-3H3/b6-4+	14.9999289	171.000	170.957	0.043
148	InChI = 1/C8H16O/c1-3-4-5-6-7-8(2)9/h3-7H2, 1-2H3	14.9618629	173.000	170.048	2.952
149	InChI = 1/C8H16O/c1-7(2)5-4-6-8(3)9/h7H, 4-6H2, 1-3H3	14.5943951	170.000	161.279	8.721
151	InChI = 1/C8H16O/c1-4-5-7(2)6-8(3)9/h7H, 4-6H2, 1-3H3	14.6585468	156.000	162.810	−6.810
152	InChI = 1/C8H16O/c1-4-5-8(9)6-7(2)3/h7H, 4-6H2, 1-3H3	14.5784724	156.000	160.899	−4.899
154	InChI = 1/C8H16O/c1-5-7(9)8(3,4)6-2/h5-6H2, 1-4H3	14.3288890	151.000	154.943	−3.943

Table 4 continued

No.	InChI	DCW	Expr.	Calc.	Expr. − Calc.
156	InChI = 1/C9H18O/c1-3-5-6-7-8-9(10)4-2/h3-8H2, 1-2H3	15.6877225	190.000	187.370	2.630
158	InChI = 1/C9H18O/c1-4-9(10)7-5-6-8(2)3/h8H, 4-7H2, 1-3H3	15.1908699	183.000	175.513	7.487
159	InChI = 1/C9H18O/c1-4-5-9(10)7-6-8(2)3/h8H, 4-7H2, 1-3H3	15.1908699	178.000	175.513	2.487
161	InChI = 1/C9H18O/c1-4-6-8(3)7-9(10)5-2/h8H, 4-7H2, 1-3H3	15.1762290	179.000	175.164	3.836
165	InChI = 1/C10H20O/c1-3-5-6-7-9-10(11)8-4-2/h3-9H2, 1-2H3	16.6803857	207.000	211.059	−4.059
168	InChI = 1/C4H6O2/c1-3-4(5)6-2/h3H, 1H2, 2H3	11.0848763	85.000	77.528	7.472
171	InChI = 1/C5H8O2/c1-3-4-5(6)7-2/h3-4H, 1-2H3/b4-3+	12.5616448	120.000	112.769	7.231
172	InChI = 1/C5H8O2/c1-4(2)5(6)7-3/h1H2, 2-3H3	11.8095216	101.000	94.821	6.179
174	InChI = 1/C5H10O2/c1-5(2)3-7-4-6/h4-5H, 3H2, 1-2H3	11.7438772	98.000	93.254	4.746
175	InChI = 1/C5H10O2/c1-3-5(2)7-4-6/h4-5H, 3H2, 1-2H3	11.7839224	97.000	94.210	2.790
178	InChI = 1/C6H10O2/c1-4-6(7)8-5(2)3/h4-5H, 1H2, 2-3H3	12.7280694	110.000	116.741	−6.741
179	InChI = 1/C6H12O2/c1-2-3-4-5-8-6-7/h6H, 2-5H2, 1H3	13.2129521	132.000	128.312	3.688
182	InChI = 1/C6H12O2/c1-4-5(2)8-6(3)7/h5H, 4H2, 1-3H3	12.6681395	112.000	115.311	−3.311
184	InChI = 1/C7H12O2/c1-3-5-6-7(8)9-4-2/h3H, 1, 4-6H2, 2H3	13.7618082	145.000	141.410	3.590
185	InChI = 1/C7H12O2/c1-3-5-6-7(8)9-4-2/h4H, 2-3, 5-6H2, 1H3	13.7752968	133.000	141.732	−8.732
188	InChI = 1/C7H14O2/c1-3-5-6-7(8)9-4-2/h3-6H2, 1-2H3	13.9513848	145.000	145.934	−0.934
190	InChI = 1/C7H14O2/c1-4-5-7(8)9-6(2)3/h6H, 4-5H2, 1-3H3	13.5518039	131.000	136.399	−5.399
192	InChI = 1/C8H14O2/c1-3-5-6-7-8(9)10-4-2/h5-6H, 3-4, 7H2, 1-2H3/b6-5+	14.8446549	167.000	167.251	−0.251
196	InChI = 1/C8H16O2/c1-3-5-6-7-10-8(9)4-2/h3-7H2, 1-2H3	14.8640696	169.000	167.715	1.285
199	InChI = 1/C8H16O2/c1-4-5-8(9)10-6-7(2)3/h7H, 4-6H2, 1-3H3	14.3597984	157.000	155.681	1.319

Acknowledgments The authors thank the Marie Curie Fellowship for financial support (the contract ID 39036, CHEMPREDICT). At JSU this work was supported by the NSF CREST grant No. HRD-0833178.

References

1. D. Vidal, M. Thormann, M. Pons, J. Chem. Inf. Model. **45**, 386–393 (2005)
2. A.A. Toropov, A.P. Toropova, D.V. Mukhamedzhanova, I. Gutman, Indian J. Chem. A **44**, 1545–1552 (2005)

3. A.A. Toropov, D. Leszczynska, J. Leszczynski, Chem. Phys. Lett. **441**, 119–122 (2007)
4. A.A. Toropov, A.P. Toropova, I. Raska Jr., Eur. J. Med. Chem. **43**, 714–740 (2008)
5. A.A. Toropov, E. Benfenati, Bioorg. Med. Chem. **16**, 4801–4809 (2008)
6. A.A. Toropov, A.P. Toropova, E. Benfenati, Chem. Phys. Lett. **461**, 343–347 (2008)
7. http://wwmm.ch.cam.ac.uk/inchifaq/
8. http://www.warr.com/inchi.pdf
9. http://toxnet.nlm.nih.gov/
10. http://webbook.nist.gov/chemistry/
11. A.T. Balaban, D. Mills, S.C. Basak, J. Chem. Inf. Comput. Sci. **39**, 758–764 (1999)