

$$\text{bp} = 6.29(\pm 1.12)\phi - 53.46(\pm 1.50)^{\text{D}}\chi^0 + 71.76(\pm 3.10)^1\chi - 19.43(\pm 0.70)N_{\text{Br}} - 22.15(\pm 1.33)N_{\text{I}} + 10.03(\pm 2.16)[^1\chi^{\text{v}} - ^0\chi^{\text{v}}] - 81.57(\pm 2.03) \quad (3)$$

$$s = 11.20, n = 532, r^2 = 0.97, F = 2814$$

An alternate expression in terms of N_{F} , N_{Cl} , N_{Br} , N_{I} , and $^1\chi$ gives a slightly inferior correlation but makes the interpretation of the equation much easier:

$$\text{bp} = 138.62(\pm 5.63)^1\chi - 15.35(\pm 1.33)N_{\text{H}} - 76.35(\pm 3.45)N_{\text{F}} - 37.21(\pm 3.48)N_{\text{Cl}} - 14.29(\pm 3.54)N_{\text{Br}} + 14.45(\pm 3.73)N_{\text{I}} - 53.78(\pm 3.27) \quad (4)$$

$$s = 11.59, n = 532, r^2 = 0.97, F = 2620$$

In the halomethanes where there are no constitutional or steric isomers possible, the boiling point in general increases with addition of Cl, Br, or I atoms. This increase is in the order $\text{I} > \text{Br} > \text{Cl}$, F. In cases where $N_{\text{C}} > 1$, the constitutional isomerism and topology further complicate the picture. We have used topological indexes to explain some of these variations. Although these indexes are somewhat complicated and may present some difficulty in their interpretation, they can be used in predicting the boiling points of proposed compounds. This can make these equations useful in the design of new halocarbons with boiling points in a certain range.

SUMMARY

Here we have accumulated the largest data set of haloalkanes in the open literature that we are aware of (see table in supplementary material). The data is presented as SMILES strings to permit optical scanning and loading onto a computer diskette for further study. The boiling point data presented here was collated from the literature. In some cases the precision of the data was not known since this information was often not stated in early literature.

To predict the boiling points, we have derived modestly good linear equation models in six easily calculated nonempirical variables. In three of the equations the counts of I and Br appear as equation variables. Thus, the deletion of molecules containing Br and I permits the analysis of the boiling point QSAR of fluorocarbons with the remaining fewer variables. The analysis of these halocarbons, 280 in number, with the four remaining variables gives an equation of statistical quality comparable to the six-variable equation.

Although the linear models we have derived here explain the variance in the data set reasonably well, it may be that the structure-activity relationships operating here are completely nonlinear. We continue our studies on this data set to explore this possibility.

We have also studied the effects of constitutional isomerism and the replacement of one halogen by another on the boiling points. These effects we have summarized in the Figure 6A-D.

Supplementary Material Available: Table containing SMILES strings and boiling points of compounds used in this study (6 pages). Ordering information is given on any current masthead page.

REFERENCES AND NOTES

- (1) Jones, R. R.; Wigley, T., Eds.; *Ozone Depletion: Health and Environmental Consequences*; Wiley: New York, 1989.
- (2) *CRC Handbook of Physics and Chemistry*, 68th ed.; CRC: Boca Raton, FL. *Dictionary of Organic Compounds*, 5th ed.; Chapman and Hall: New York, 1982. Smith, B. D.; Srivastava, R. *Thermodynamic Data for Pure Compounds, Part B*; Halogenated Hydrocarbons and Alcohols, Physical Sciences Data No. 25; Elsevier: Amsterdam, 1986.
- (3) Lindgren, F.; Eriksson, L.; Hellberg, S.; Johnson, J.; Sjostrom, M.; Wold, S. *Quant. Struct.-Act. Relat.* **1991**, *10*, 36.
- (4) Seybold, P. G.; May, M. A.; Gargas, M. L. *Acta Pharm. Jugosl.* **1986**, *36*, 253.
- (5) Bindal, M. C.; Singh, P.; Gupta, S. P. *Arzneim.-Forsch.* **1980**, *30*, 234.
- (6) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- (7) Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. *J. Pharm. Sci.* **1976**, 1226.
- (8) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *80*, 399.
- (9) Kier, L. B. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109.
- (10) Kier, L. B.; Hall, L. H. *Quant. Struct.-Act. Relat.* **1991**, *10*, 134.
- (11) The SAS Institute Inc., P.O. Box 8000, Cary, NC 27511-8000.

Correlations between Chemical Structure and Normal Boiling Points of Acyclic Ethers, Peroxides, Acetals, and Their Sulfur Analogues

ALEXANDRU T. BALABAN,*[†] LEMONT B. KIER,*[‡] and NIKHIL JOSHI[‡]

Organic Chemistry Department, Polytechnic Institute Bucharest, Splaiul Independentei 313, 77206 Bucharest, Romania, and Department of Medicinal Chemistry, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298-0540

Received September 10, 1991

Boiling points at normal pressure for a set of 185 acyclic saturated compounds with one or two divalent oxygen or sulfur heteroatoms devoid of hydrogen bonding were correlated with chemical structures using two or three topological descriptors. The standard error is 8.2 °C, and the correlation coefficient r^2 is 0.971. For subsets (only 72 monoethers or only 44 monosulfides) the standard error is 4-5 °C and r^2 is about 0.98.

INTRODUCTION

Along with the accompanying paper¹ which deals with correlations involving univalent halogens, we also investigated

correlations between boiling points at normal pressure (BPs) and chemical structures of 185 saturated acyclic compounds possessing divalent oxygen or sulfur (chalcogen) atoms, and devoid of hydrogen bonding, namely, 73 ethers, 17 diethers (with heteroatoms at least three bonds apart), 21 acetals, 6 peroxides, as well as their sulfur analogues: 45 sulfides (thioethers), 6 bis-sulfides, 4 thioacetals, and 13 disulfides,

[†] Polytechnic Institute Bucharest.

[‡] Virginia Commonwealth University.

Table I. Comparison of BPs for Three Classes of Compounds Having the Same Skeleton

no. in Table II	C	H	O	S	name	BP, °C
26	6	14			<i>n</i> -hexane	69.0
27	6	12	1		2-oxahexane	70.3
18	6	12	1		3-oxahexane	63.6
16	6	10	2		2,5-dioxahexane	84.7
15	6	10	2		2,4-dioxahexane	67.0
38	6	10	2		3,4-dioxahexane	63.0
40	6	12		1	2-thiahexane	123.2
25	6	10		1	3-thiahexane	118.5
24	6	10		2	2,5-dithiahexane	183.0
22	6	10		2	2,4-dithiahexane	166.0
				2	3,4-dithiahexane	154.0
	7	14			<i>n</i> -heptane	98.4
45	7	12	1		2-oxaheptane	99.5
46	7	12	1		3-oxaheptane	92.3
47	7	12	1		2,6-dioxaheptane	104.5
35	7	10	2		2,5-dioxaheptane	102.0
32	7	10	2		3,5-dioxaheptane	88.0
65	7	12		1	2-thiaheptane	145.0
66	7	12		1	3-thiaheptane	144.2
67	7	12		1	4-thiaheptane	142.8
42	7	10		2	3,4-dithiaheptane	173.7
44	7	10		2	3,5-dithiaheptane	181.0
	8	18			<i>n</i> -octane	125.7
83	8	16	1		2-oxaoctane	125.0
84	8	16	1		3-oxaoctane	118.0
85	8	16	1		4-oxaoctane	117.1
62	8	14	2		2,7-dioxaoctane	132.5
63	8	14	2		3,6-dioxaoctane	123.5
105	8	16			2-thiaoctane	171.0
106	8	16		1	4-thiaoctane	166.0
82	8	14		2	3,5-dithiaoctane	211.0
76	8	14		2	4,5-dithiaoctane	195.8

respectively. All such systems with up to $N = 11$ non-hydrogen atoms, for which reliable literature data could be found, were included.

The purpose of this paper is (i) to explore how heteroatoms included within acyclic molecular skeletons influence a fairly precisely measured physical property, namely, the normal boiling point; (ii) to show that topological descriptors can account satisfactorily for the observed trends of boiling points; and (iii) to present quantitative structure-property relationships (QSPRs) that can validate databanks and that may be used for predicting boiling points of unknown compounds belonging to the set of compounds being included in this study.

A recent paper² described a computer-assisted prediction of normal boiling points of cyclic compounds containing ethereal oxygen atoms and of their sulfur analogues (furans, tetrahydrofurans, and thiophenes); with 12 descriptors and 236 compounds, a correlation with $r^2 = 0.974$ and with a standard deviation of 9.6 °C was obtained. Statistical data were somewhat better for subsets or after eliminating outliers, but a little worse for the complete set of 343 compounds.

DATA COLLECTION

In addition to Beilstein literature data extending till 1960, boiling points at normal pressure (BPs) were also collected from the *CRC Handbook of Physics and Chemistry*¹⁹ and the *Dictionary of Organic Compounds*,²⁰ which both contain more recent literature data.

The precision of most data is probably within 1 °C, but in several cases when BPs differing by 1–4 °C were reported, their arithmetic mean was included so that in these cases the imprecision is about twice as large.

In one instance where reported BPs differed by 10 °C, namely, butyl propyl sulfide (no. 106 in Table II: reported

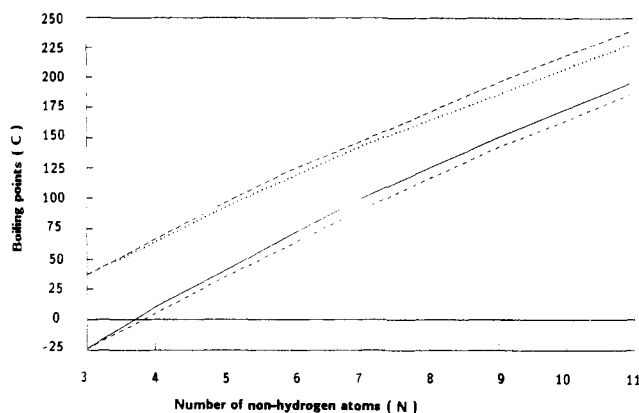


Figure 1. Normal boiling points of ethers and sulfides vs the number of non-hydrogen atoms. (—) AlkOMe; (---) Alk₂O; (···) AlkSMMe; (— · —) Alk₂S.

BPs of 157–158 °C³ and 166–168 °C⁴), we considered that the higher BP was more likely by comparison with compound no. 105, having the same skeleton.

TOPOLOGICAL DESCRIPTORS

The following constitutional parameters were included as variables in the correlations tested by means of the SAS statistical package⁵ and MOLCONN program:⁶ numbers N_c , N_o , and N_s of C, O, and S atoms, respectively; number N of non-hydrogen atoms; number N_{Me} of methyl groups.

A comprehensive set of topological indexes was tested; it included molecular connectivity indexes (Randić's $^1\chi$;⁷ Kier and Hall's extended connectivities $^2\chi$, $^3\chi$,^{8,9} and valence $\chi^{8,9}$ indexes $^0\chi^v$, $^1\chi^v$, $^v\chi^2$) and shape indexes (Balaban's average distance sum connectivity J ,^{10,11} as well as J_{het} modified by electronegativity or steric parameters for the presence of heteroatoms;¹² Kier's κ indexes;^{13,14} Kier's α - κ indexes¹⁵).

A more detailed description of the J_{het} topological index follows. From the distance matrix of the graph, one computes the distance sum (or distance degree) S_i by summing entries over row i or column i ; this constitutes a local vertex invariant for vertex i , with a wider range of variation than the vertex degree, although both of them are integer numbers.

The chemical nature of the atom symbolized by this vertex is introduced by multiplying S_i with a chemical coefficient C_i , which is 1.00 for carbon atoms, but which is different from 1 for heteroatoms. Alternatively, the hybridization state of the atoms might be introduced if one wants to finely tune the nature of the chemical bonds in this fashion. The resulting numbers $S'_i = S_i C_i$ are then introduced into the formula for J in the place of the distance sums:

$$J_{het} = \left(\frac{q}{\mu + 1} \right) \sum_{edges\ ij} (S'_i S'_j)^{-1/2}$$

where q is the number of edges and $\mu = N + q$ is the cyclo-matic number of the graph.

Of particular interest is the electrotopological state of the O or S heteroatom (S_{het}),^{16,17} when nonequivalent heteroatoms are present, their arithmetic mean is employed.

Both J_{het} and S_{het} parameters are completely nondegenerate in the set of compounds studied.

TRENDS IN BOILING POINTS OF CHALCOGEN-CONTAINING COMPOUNDS

The boiling points of chalcogen atom-containing chains present some interesting trends.

Table II. Chalcogen Atom-Containing Compounds with Their Boiling Points, J_{het} Values, Calculated (calc) Boiling Points (Using eq 2), and Corresponding Residuals (res)

obs	C	H	O	S	name	J_{het}	BP	calcd	res
1	2	6	1	0	dimethyl ether	5.164	-23.7	5.2	-28.9
2	2	6	2	0	dimethyl peroxide	11.373	14.0	9.1	4.9
3	2	6	0	1	dimethyl sulfide	5.164	37.3	50.2	-12.9
4	2	6	0	2	dimethyl disulfide	11.373	109.7	99.1	10.6
5	3	8	1	0	ethyl methyl ether	4.921	10.8	27.8	-17.0
6	3	8	2	0	ethyl methyl peroxide	10.114	39.0	34.6	4.4
7	3	8	2	0	dimethoxymethane	6.927	42.0	43.9	-1.9
8	3	8	0	1	ethyl methyl sulfide	4.921	66.6	72.8	-6.2
9	3	8	0	2	ethyl methyl disulfide	10.114	135.0	124.7	10.3
10	3	8	0	2	bis(methylthio)methane	6.927	148.5	133.9	14.6
11	4	10	1	0	methyl propyl ether	4.559	40.0	50.7	-10.7
12	4	10	1	0	diethyl ether	4.860	34.6	49.9	-15.3
13	4	10	1	0	isopropyl methyl ether	5.296	32.0	42.3	-10.3
14	4	10	2	0	diethyl peroxide	9.512	63.0	58.2	4.8
15	4	10	2	0	isopropyl methyl peroxide	9.924	53.5	50.7	2.8
16	4	10	2	0	ethoxymethoxymethane	6.555	67.0	66.8	0.2
17	4	10	2	0	1,1-dimethoxyethane	7.477	64.4	59.5	4.9
18	4	10	2	0	1,2-dimethoxyethane	6.196	84.7	67.9	16.8
19	4	10	0	1	methyl propyl sulfide	4.559	95.5	95.8	-0.3
20	4	10	0	1	diethyl sulfide	4.860	92.0	94.9	-2.9
21	4	10	0	1	isopropyl methyl sulfide	5.296	84.4	87.3	-2.9
22	4	10	0	2	diethyl disulfide	9.512	154.0	148.3	5.7
23	4	10	0	2	1,1-bis(methylthio)ethane	7.477	156.0	149.6	6.4
24	4	10	0	2	ethylthiomethylthiomethane	6.555	166.0	156.9	9.1
25	4	10	0	2	1,2-bis(methylthio)ethane	6.196	183.0	157.9	25.1
26	5	12	1	0	butyl methyl ether	4.267	70.3	73.5	-3.2
27	5	12	1	0	ethyl propyl ether	4.627	63.6	72.4	-8.8
28	5	12	1	0	ethyl isopropyl ether	5.187	52.5	64.5	-12.0
29	5	12	1	0	isobutyl methyl ether	4.750	59.0	65.8	-6.8
30	5	12	1	0	<i>s</i> -butyl methyl ether	5.116	59.5	66.4	-6.9
31	5	12	1	0	<i>t</i> -butyl methyl ether	5.832	55.2	53.5	1.7
32	5	12	2	0	diethoxymethane	6.324	88.0	89.4	-1.4
33	5	12	2	0	2,2-dimethoxypropane	8.082	83.0	71.5	11.5
34	5	12	2	0	1,3-dimethoxypropane	5.662	104.5	91.3	13.2
35	5	12	2	0	1-ethoxy-2-methoxyethane	5.993	102.0	90.4	11.6
36	5	12	2	0	1,2-dimethoxypropane	6.659	92.0	83.8	8.2
37	5	12	0	1	ethyl isopropyl sulfide	5.187	107.4	109.5	-2.1
38	5	12	0	1	butyl methyl sulfide	4.267	123.2	118.5	4.7
39	5	12	0	1	isobutyl methyl sulfide	4.750	112.5	110.8	1.7
40	5	12	0	1	ethyl propyl sulfide	4.627	118.5	117.5	1.0
41	5	12	0	1	<i>t</i> -butyl methyl sulfide	5.832	101.5	98.5	3.0
42	5	12	0	2	ethyl propyl disulfide	8.709	173.7	172.5	1.2
43	5	12	0	2	ethyl isopropyl disulfide	9.475	165.5	164.0	1.5
44	5	12	0	2	bis(ethylthio)methane	6.324	181.0	179.5	1.5
45	6	14	1	0	methyl pentyl ether	4.055	99.5	96.0	3.5
46	6	14	1	0	ethyl butyl ether	4.386	92.3	95.0	-2.7
47	6	14	1	0	dipropyl ether	4.525	90.1	94.6	-4.5
48	6	14	1	0	isopropyl propyl ether	4.937	80.2	87.1	-6.9
49	6	14	1	0	ethyl isobutyl ether	4.775	82.0	87.6	-5.6
50	6	14	1	0	isopentyl methyl ether	4.394	91.2	88.7	2.5
51	6	14	1	0	methyl 2-methylbutyl ether	4.818	91.5	89.1	2.4
52	6	14	1	0	ethyl <i>s</i> -butyl ether	5.114	81.2	88.3	-71.4
53	6	14	1	0	methyl 2-methylbutyl ether	4.673	93.0	89.5	3.5
54	6	14	1	0	diisopropyl ether	5.427	69.0	79.4	-10.4
55	6	14	1	0	methyl <i>t</i> -pentyl ether	5.721	86.3	78.3	8.0
56	6	14	1	0	1,2-dimethylpropyl methyl ether	5.301	82.0	82.1	-0.1
57	6	14	2	0	1,1-diethoxyethane	6.881	103.0	105.0	-2.0
58	6	14	2	0	1,1-dimethoxy-2-methylpropane	7.417	103.5	99.6	3.9
59	6	14	2	0	2-ethoxy-2-methoxy-2,2-dimethylpropane	7.688	96.0	94.5	1.5
60	6	14	2	0	1,1-dimethoxybutane	6.873	112.0	106.7	5.3
61	6	14	2	0	1-methoxy-1-propoxyethane	6.636	104.0	105.7	-1.7
62	6	14	2	0	1,4-dimethoxybutane	5.271	132.5	114.4	18.1
63	6	14	2	0	1,2-diethoxyethane	5.835	123.5	112.7	10.8
64	6	14	2	0	1,3-dimethoxybutane	6.054	120.3	107.4	12.9
65	6	14	0	1	methyl pentyl sulfide	4.055	145.0	141.0	4.0
66	6	14	0	1	butyl ethyl sulfide	4.386	144.2	140.1	4.1
67	6	14	0	1	dipropyl sulfide	4.525	142.8	139.7	3.1
68	6	14	0	1	isopropyl propyl sulfide	4.937	132.0	132.2	-0.2
69	6	14	0	1	ethyl isobutyl sulfide	4.775	134.2	132.6	1.6
70	6	14	0	1	isopentyl methyl sulfide	4.394	137.0	133.7	3.3
71	6	14	0	1	methyl 2-methylbutyl sulfide	4.673	139.0	134.6	4.4
72	6	14	0	1	<i>s</i> -butyl ethyl sulfide	5.114	133.6	133.3	0.3
73	6	14	0	1	<i>t</i> -butyl ethyl sulfide	5.660	120.4	120.9	-0.5
74	6	14	0	1	diisopropyl sulfide	5.427	120.0	124.4	-4.4
75	6	14	0	1	1-ethylpropyl methyl sulfide	5.149	137.0	134.9	2.1
76	6	14	0	2	dipropyl disulfide	8.251	195.8	195.8	0.0

Table II (Continued)

obs	C	H	O	S	name	J_{het}	BP	calcd	res
77	6	14	0	2	diisopropyl disulfide	9.450	177.2	179.7	-2.5
78	6	14	0	2	s-butyl ethyl disulfide	8.999	181.0	189.0	-8.0
79	6	14	0	2	isopropyl propyl disulfide	8.812	185.9	187.8	-1.9
80	6	14	0	2	t-butyl ethyl disulfide	9.684	175.7	176.1	-0.4
81	6	14	0	2	1,1-bis(ethylthio)ethane	6.881	186.0	195.1	-9.1
82	6	14	0	2	1,2-bis(ethylthio)ethane	5.835	211.0	202.8	8.2
83	7	16	1	0	hexyl methyl ether	3.901	125.0	118.3	6.7
84	7	16	1	0	ethyl pentyl ether	4.183	118.0	117.5	0.5
85	7	16	1	0	butyl propyl ether	4.368	117.1	117.0	0.1
86	7	16	1	0	butyl isopropyl ether	4.673	107.0	109.8	-2.8
87	7	16	1	0	isobutyl propyl ether	4.673	102.5	109.8	-7.3
88	7	16	1	0	ethyl isopentyl ether	4.465	112.0	110.4	1.6
89	7	16	1	0	t-butyl propyl ether	5.360	97.4	98.6	-1.2
90	7	16	1	0	2,2-dimethylpropyl ethyl ether	5.092	91.5	99.4	-7.9
91	7	16	1	0	t-butyl isopropyl ether	5.818	87.6	91.0	-3.4
92	7	16	1	0	ethyl 1-methylbutyl ether	4.900	106.5	110.8	-4.3
93	7	16	1	0	ethyl t-pentyl ether	5.657	101.0	100.4	0.6
94	7	16	1	0	1,2-dimethylpropyl ethyl ether	5.292	99.3	104.1	-4.8
95	7	16	1	0	ethyl 1-ethylpropyl ether	5.196	90.0	111.6	-21.6
96	7	16	2	0	dipropoxymethane	5.824	137.0	134.6	2.4
97	7	16	2	0	2,2-diethoxypropane	7.463	114.0	117.1	-3.1
98	7	16	2	0	1-ethoxy-1-propoxyethane	6.545	126.0	127.9	-1.9
99	7	16	2	0	1,1-diethoxypropane	6.981	124.0	128.3	-4.3
100	7	16	2	0	1,3-diethoxypropane	5.448	140.5	135.7	4.8
101	7	16	2	0	1,5-dimethoxyhexane	4.977	157.5	137.1	20.4
102	7	16	2	0	1-ethoxy-4-methoxybutane	5.212	146.0	136.4	9.6
103	7	16	2	0	1,4-dimethoxyhexane	5.602	145.0	130.6	14.4
104	7	16	2	0	1,3-dimethoxyhexane	6.135	141.0	130.8	10.2
105	7	16	0	1	hexyl methyl sulfide	3.901	171.0	163.4	7.6
106	7	16	0	1	butyl propyl sulfide	4.368	166.0	162.0	4.0
107	7	16	0	1	isobutyl propyl sulfide	4.673	155.0	154.8	0.2
108	7	16	0	1	isobutyl isopropyl sulfide	5.020	145.0	147.5	-2.5
109	7	16	0	1	ethyl 2-methylbutyl sulfide	4.720	159.0	156.3	2.7
110	7	16	0	1	t-butyl propyl sulfide	5.360	138.0	143.6	-5.6
111	7	16	0	1	s-butyl isopropyl sulfide	5.347	142.0	148.2	-6.2
112	7	16	0	1	ethyl isopentyl sulfide	4.465	159.0	155.4	3.6
113	7	16	0	1	butyl isopropyl sulfide	4.673	163.5	154.8	8.7
114	7	16	0	2	1,3-bis(ethylthio)propane	5.448	229.5	225.8	3.7
115	8	18	1	0	dibutyl ether	4.283	142.0	139.1	2.9
116	8	18	1	0	isopentyl propyl ether	4.441	125.0	132.3	-7.3
117	8	18	1	0	butyl isobutyl ether	4.522	132.0	132.1	-0.1
118	8	18	1	0	butyl s-butyl ether	4.747	130.5	133.1	-2.6
119	8	18	1	0	butyl t-butyl ether	5.051	125.0	121.4	3.6
120	8	18	1	0	s-butyl isobutyl ether	5.039	122.0	126.0	-4.0
121	8	18	1	0	1,3-dimethylpentyl methyl ether	4.802	121.0	128.3	-7.3
122	8	18	1	0	diisobutyl ether	4.785	122.2	125.0	-2.8
123	8	18	1	0	isobutyl t-butyl ether	5.378	112.0	114.1	-2.1
124	8	18	1	0	di-t-butyl ether	6.134	106.0	102.8	3.2
125	8	18	1	0	isopropyl t-pentyl ether	5.823	114.5	115.5	-1.0
126	8	18	1	0	heptyl methyl ether	3.786	151.0	140.6	10.4
127	8	18	1	0	1-ethylpropyl propyl ether	5.079	128.5	133.8	-5.3
128	8	18	2	0	di-t-butyl peroxide	9.775	109.5	114.1	-4.6
129	8	18	2	0	1,1-diisopropoxyethane	6.985	126.0	135.9	-9.9
130	8	18	2	0	1,1-dipropoxyethane	6.331	147.0	150.4	-3.4
131	8	18	2	0	1,3-dimethoxyhexane	6.035	158.0	152.9	5.1
132	8	18	2	0	2,4-dimethoxy-2-methylpentane	6.788	147.0	136.3	10.7
133	8	18	2	0	1,4-diethoxybutane	5.142	165.0	158.5	6.5
134	8	18	0	1	dibutyl sulfide	4.283	188.9	184.1	4.8
135	8	18	0	1	diisobutyl sulfide	4.785	170.0	170.1	-0.1
136	8	18	0	1	butyl isobutyl sulfide	4.552	178.0	177.1	0.9
137	8	18	0	1	di-t-butyl sulfide	6.134	148.5	147.8	0.7
138	8	18	0	1	di-s-butyl sulfide	5.321	165.0	171.8	-6.8
139	8	18	0	1	butyl s-butyl sulfide	4.747	177.0	178.2	-1.2
140	8	18	0	1	s-butyl isobutyl sulfide	5.039	167.0	171.0	-4.0
141	8	18	0	1	heptyl methyl sulfide	3.786	195.0	185.6	9.4
142	8	18	0	2	dibutyl disulfide	7.386	226.0	242.1	-16.1
143	8	18	0	2	diisobutyl disulfide	8.054	215.0	227.5	-12.5
144	8	18	0	2	di-t-butyl disulfide	9.775	201.0	204.2	-3.2
145	8	18	0	2	1,1-bis(isopropylthio)ethane	6.985	205.0	226.0	-21.0
146	9	20	1	0	1-ethyl-1,3-dimethylbutyl methyl ether	5.597	151.5	140.7	10.8
147	9	20	1	0	ethyl heptyl ether	3.895	165.5	162.1	3.4
148	9	20	1	0	butyl isopentyl ether	4.363	157.0	154.5	2.5
149	9	20	1	0	t-butyl isopentyl ether	5.018	139.0	137.1	1.9
150	9	20	1	0	butyl pentyl ether	4.175	163.0	161.3	1.7
151	9	20	1	0	1,5-dimethylhexyl methyl ether	4.335	153.5	149.9	3.6
152	9	20	1	0	isobutyl isopentyl ether	4.567	139.0	147.6	-8.6
153	9	20	1	0	methyl 1-methylheptyl ether	4.145	162.0	156.8	5.2

Table II (Continued)

obs	C	H	O	S	name	J_{het}	BP	calcd	res
154	9	20	1	0	methyl octyl ether	3.699	173.0	162.7	10.3
155	9	20	1	0	2-ethylhexyl methyl ether	4.553	159.5	157.2	2.3
156	9	20	1	0	methyl 1,1,4-trimethylpentyl ether	5.035	159.5	139.7	19.8
157	9	20	1	0	3,5-dimethylhexyl methyl ether	4.379	155.5	149.8	5.7
158	9	20	1	0	ethyl 1,1,3-trimethylbutyl ether	5.473	141.0	138.4	2.6
159	9	20	2	0	<i>t</i> -butyl <i>t</i> -pentyl peroxide	9.549	126.0	139.3	-13.3
160	9	20	2	0	1,1-dimethoxy-2,2-dimethylpentane	7.249	164.0	159.7	4.3
161	9	20	2	0	1,1-diethoxypentane	6.540	163.0	173.4	-10.4
162	9	20	2	0	1,1-dipropoxypropane	6.543	166.5	173.4	-6.9
163	9	20	2	0	1,1-diisopropoxypropane	7.143	146.0	159.0	-13.0
164	9	20	2	0	1,3-dipropoxypropane	5.211	165.0	180.2	-15.2
165	9	20	2	0	1,3-diisopropoxypropane	5.586	159.5	166.5	-7.0
166	9	20	0	1	ethyl heptyl sulfide	3.895	195.0	207.2	-12.2
167	9	20	0	1	methyl octyl sulfide	3.699	218.0	207.7	10.3
168	9	20	0	2	bis(butylthio)methane	5.441	250.0	269.6	-19.6
169	9	20	0	2	2,2-bis(propylthio)propane	6.857	235.0	252.7	-17.7
170	10	22	1	0	ethyl octyl ether	3.795	186.5	184.3	2.2
171	10	22	1	0	ethyl 1,1,3,3-tetramethylbutyl ether	5.638	156.5	150.7	5.8
172	10	22	1	0	bis(1-ethylpropyl) ether	5.504	162.0	173.4	-11.4
173	10	22	1	0	bis(1-methylbutyl) ether	5.083	162.0	171.3	-9.3
174	10	22	1	0	butyl 1-methylpentyl ether	4.593	170.0	177.4	-7.4
175	10	22	1	0	diisopentyl ether	4.427	173.2	169.9	3
176	10	22	1	0	dipentyl ether	4.109	186.8	183.4	3.4
177	10	22	1	0	isopropyl heptyl ether	4.094	173.0	177.1	-4.1
178	10	22	1	0	heptyl propyl ether	3.951	187.0	183.9	3.1
179	10	22	1	0	isopentyl pentyl ether	4.263	174.0	176.7	-2.7
180	10	22	1	0	methyl 1-methyloctyl ether	4.006	188.5	179.1	9.4
181	10	22	0	1	di- <i>t</i> -pentyl sulfide	6.126	199.0	196.9	2.1
182	10	22	0	1	dipentyl sulfide	4.109	228.0	228.4	-0.4
183	10	22	0	1	diisopentyl sulfide	4.427	215.0	214.9	0.1
184	10	22	0	1	isobutyl 4-methylpentyl sulfide	4.378	216.0	215.0	1.0
185	10	22	0	1	methyl nonyl sulfide	3.631	240.0	229.8	10.2

(1) As shown in Table I by the hydrocarbons and their mono and dioxa congeners, for equal numbers of non-hydrogen atoms the ethers, diethers, peroxides, and acetals have BP's comparable to hydrocarbons. In other words, replacement of one or two methylene groups by an oxygen heteroatom does not appreciably change the BP.

(2) Deviations from linearity with increasing numbers N of non-hydrogen atoms (graph vertexes) in compounds having the same skeleton are slight in the range $N = 3-11$, characterizing the compounds under discussion. For linear chains this is illustrated in Figure 1 with symmetrical dialkyl ethers and with alkyl methyl ethers.

(3) For isomeric ethers with the same skeleton having a more centrally situated ethereal oxygen, the BP is lower than for systems with a more marginal one, e.g. by 11 °C for compounds 48 and 50 from Table II. This can also be seen from Figure 1.

(4) For isomeric diethers, acetals, and peroxides having the same skeleton, the BP is higher when the two heteroatoms are farther apart, e.g., by 11.7 °C for compounds 14 and 18 or 20.5 °C for compounds 96 and 101.

The previous observation 3 would suggest that an oxygen heteroatom behaves like a pseudobranch, but observation 4 contradicts this model.

(5) Sulfur heteroatoms behave similarly to oxygen heteroatoms, but the increment in BPs on replacing a CH_2 group by a sulfur atom is substantial, so that comment 1 above does not apply to S heteroatoms. However, in comments 2-4 one may replace O by S; the difference in BPs for isomeric compounds having the same skeleton is, however, larger: e.g., for compounds 22 and 25, this difference is 29 °C.

Previously it had been reported¹⁸ that straight-chain sulfides had lower BPs when the sulfur heteroatoms were closer to the middle of the chain, in agreement with comment 3 above when applied to sulfur heteroatoms.

It is interesting to observe that the topological index J_{het} modified for heteroatoms as will be mentioned below reflects

qualitatively all trends 3-5 mentioned above.

QUANTITATIVE STRUCTURE-PROPERTY CORRELATIONS

A two-variable regression (BP versus the molecular connectivity $^1\chi$ and the sulfur atom count N_s) affords satisfactory results (correlation coefficient, $r^2 = 0.964$; standard error, $s = 9.0$ °C; Fisher value, $F = 2390$; $n = 185$)

$$\text{BP} = -59.10(\pm 3.11) + 44.30(\pm 0.76)^1\chi + 42.88(\pm 0.95)N_s$$

The t factors are -19.0, 58.2, and 45.2, respectively. Parameters $^1\chi$ and N_s have an intercorrelation coefficient of -0.037 so that they are practically not intercorrelated.

Inclusion of a third parameter with topological descriptors that do not discriminate among isomers differing in the position of the heteroatom(s), such as N_{Me} or J , improve the correlation slightly.

The major drawback of these correlations is the fact that they provide the same calculated BP for isomeric compounds with the same skeleton, but differing in the position of the heteroatom(s).

Two parameters, each of which eliminates this degeneracy, are S_{het} and J_{het} . The first one presents a high-correlation coefficient with N_s (-0.898); therefore, it cannot be used as an additional parameter along with N_s . It can, however, be used along with other descriptors:

$$\text{BP} = -11.23 - 7.21S_{\text{het}} + 35.04^\circ\chi^v - 18.30T_{\text{Me}} \quad (1)$$

with $r^2 = 0.954$, $F = 1019$, $s = 10.5$ °C, and $n = 185$.

Since the oxygen and sulfur heteroatoms differ by both their steric and their electronic parameters, they can be modeled topologically by multiplying the distance sums (over rows or columns in the topological distance matrix) by coefficients corresponding to these differences; thus, instead of integer-number local vertex invariants (LOVIs) corresponding to

Table III. Boiling Points of Ethers Calculated (calc) Using eq 3 Along with Corresponding Residuals (res) and E-State Values

obs	C	H	O	name	ES	BP	calc	res
1	3	8	1	ethyl methyl ether	4.542	10.8	22.0	-11.2
2	4	10	1	methyl propyl ether	4.691	40.0	45.0	-5.0
3	4	10	1	diethyl ether	4.833	34.6	41.3	-6.7
4	4	10	1	isopropyl methyl ether	4.750	32.0	37.1	-5.1
5	5	12	1	butyl methyl ether	4.782	70.3	69.4	0.9
6	5	12	1	ethyl propyl ether	4.983	63.6	64.2	-0.6
7	5	12	1	ethyl isopropyl ether	5.042	52.5	56.4	-3.9
8	5	12	1	isobutyl methyl ether	4.803	59.0	62.6	-3.6
9	5	12	1	s-butyl methyl ether	4.899	59.5	60.1	-0.6
10	5	12	1	t-butyl methyl ether	4.938	55.2	54.7	0.5
11	6	14	1	methyl pentyl ether	4.844	99.5	94.6	4.9
12	6	14	1	ethyl butyl ether	5.074	92.3	88.7	3.6
13	6	14	1	dipropyl ether	5.132	90.1	87.2	2.9
14	6	14	1	isopropyl propyl ether	5.191	80.2	79.3	0.9
15	6	14	1	ethyl isobutyl ether	5.095	82.0	81.8	0.2
16	6	14	1	isopentyl methyl ether	4.853	91.2	88.1	3.1
17	6	14	1	methyl 1-methylbutyl ether	4.991	91.5	84.5	7.0
18	6	14	1	ethyl s-butyl ether	5.191	81.2	79.3	1.9
19	6	14	1	methyl 2-methylbutyl ether	4.894	93.0	87.0	6.0
20	6	14	1	diisopropyl ether	5.250	69.0	71.5	-2.5
21	6	14	1	methyl t-pentyl ether	5.087	86.3	77.7	8.6
22	6	14	1	1,2-dimethylpropyl methyl ether	5.012	82.0	77.6	4.4
23	7	16	1	hexyl methyl ether	4.888	125.0	120.3	4.7
24	7	16	1	ethyl pentyl ether	5.136	118.0	113.9	4.1
25	7	16	1	butyl propyl ether	5.223	117.1	111.6	5.5
26	7	16	1	butyl isopropyl ether	5.282	107.0	103.8	3.2
27	7	16	1	isobutyl propyl ether	5.244	102.5	104.7	-2.2
28	7	16	1	ethyl isopentyl ether	5.144	112.0	107.3	4.7
29	7	16	1	t-butyl propyl ether	5.378	97.4	97.0	0.4
30	7	16	1	2,2-dimethylpropyl ethyl ether	5.198	91.5	101.6	-10.1
31	7	16	1	t-butyl isopropyl ether	5.438	87.6	89.1	-1.5
32	7	16	1	ethyl 1-methylbutyl ether	5.282	106.5	103.8	2.7
33	7	16	1	ethyl t-pentyl ether	5.378	101.0	97.0	4.0
34	7	16	1	1,2-dimethylpropyl ethyl ether	5.303	99.3	96.9	2.4
35	7	16	1	ethyl 1-ethylpropyl ether	5.340	90.0	102.3	-12.3
36	8	18	1	dibutyl ether	5.314	142.0	136.0	6.0
37	8	18	1	isopentyl propyl ether	5.294	125.0	130.2	-5.2
38	8	18	1	butyl isobutyl ether	5.335	132.0	129.2	2.8
39	8	18	1	butyl s-butyl ether	5.432	130.5	126.7	3.8
40	8	18	1	butyl t-butyl ether	5.470	125.0	121.4	3.6
41	8	18	1	s-butyl isobutyl ether	5.453	122.0	119.8	2.2
42	8	18	1	1,3-dimethylpentyl methyl ether	5.123	121.0	128.4	-7.4
43	8	18	1	diisobutyl ether	5.356	122.2	122.3	-0.1
44	8	18	1	isobutyl t-butyl ether	5.491	112.0	114.5	-2.5
45	8	18	1	di-t-butyl ether	5.625	106.0	106.7	-0.7
46	8	18	1	isopropyl t-pentyl ether	5.586	114.5	112.1	2.4
47	8	18	1	heptyl methyl ether	4.922	151.0	146.2	4.8
48	8	18	1	1-ethylpropyl propyl ether	5.490	128.5	125.2	3.3
49	9	20	1	1-ethyl-1,3-dimethylbutyl methyl ether	5.398	151.5	143.7	7.8
50	9	20	1	ethyl heptyl ether	5.214	165.5	165.4	0.1
51	9	20	1	butyl isopentyl ether	5.385	157.0	154.7	2.3
52	9	20	1	t-butyl isopentyl ether	5.540	139.0	140.0	-1.0
53	9	20	1	butyl pentyl ether	5.376	163.0	161.2	1.8
54	9	20	1	1,5-dimethylhexyl methyl ether	5.132	153.5	154.9	-1.4
55	9	20	1	isobutyl isopentyl ether	5.406	139.0	147.8	-8.8
56	9	20	1	methyl 1-methylheptyl ether	5.130	162.0	161.3	0.7
57	9	20	1	methyl octyl ether	4.948	173.0	172.3	0.7
58	9	20	1	2-ethylhexyl methyl ether	5.092	159.5	162.3	-2.8
59	9	20	1	methyl 1,1,4-trimethylpentyl ether	5.288	159.5	146.6	12.9
60	9	20	1	3,5-dimethylhexyl methyl ether	4.994	155.5	158.5	-3.0
61	9	20	1	ethyl 1,1,3-trimethylbutyl ether	5.540	141.0	140.0	1.0
62	10	22	1	ethyl octyl ether	5.240	186.5	191.6	-5.1
63	10	22	1	ethyl 1,1,3,3-tetramethylbutyl ether	5.605	156.5	160.8	-4.3
64	10	22	1	bis(1-ethylpropyl) ether	5.847	162.0	163.2	-1.2
65	10	22	1	bis(1-methylbutyl) ether	5.731	162.0	166.2	-4.2
66	10	22	1	butyl 1-methylpentyl ether	5.584	170.0	176.3	-6.3
67	10	22	1	diisopentyl ether	5.455	173.2	173.4	-0.2
68	10	22	1	dipentyl ether	5.438	186.8	186.4	0.4
69	10	22	1	isopropyl heptyl ether	5.422	173.0	180.5	-7.5
70	10	22	1	heptyl propyl ether	5.363	187.0	188.4	-1.4
71	10	22	1	isopentyl pentyl ether	5.447	174.0	179.9	-5.9
72	10	22	1	methyl 1-methyloctyl ether	5.157	188.5	187.4	1.1

carbon atoms, one obtains for heteroatoms real-number LO-VIs, which are then introduced into the formula for calculating J similarly to the integer distance sums. A simplified pro-

cedure was adopted taking 0.1 as the coefficient both for O and S heteroatoms. Since in the data set no compound has more than one type of heteroatom and since the number of

Table IV. Boiling Points of Sulfides Calculated (calc) Using eq 6 Along with Corresponding Residuals (res), *E*-State Values, and First-Order χ Shape Index

obs	C	H	S	name	ES	K α 1	BP	calc	res
1	3	8	1	ethyl methyl sulfide	1.856	4.350	66.6	74.6	-8.0
2	4	10	1	methyl propyl sulfide	1.902	5.350	95.5	97.2	-1.7
3	4	10	1	diethyl sulfide	1.963	5.350	92.0	95.2	-3.2
4	4	10	1	isopropyl methyl sulfide	1.880	5.350	84.4	88.1	-3.7
5	5	12	1	ethyl isopropyl sulfide	1.986	6.350	107.4	108.7	-1.3
6	5	12	1	butyl methyl sulfide	1.926	6.350	123.2	120.5	2.7
7	5	12	1	isobutyl methyl sulfide	1.910	6.350	112.5	111.2	1.3
8	5	12	1	ethyl propyl sulfide	2.008	6.350	118.5	117.8	0.7
9	5	12	1	<i>t</i> -butyl methyl sulfide	1.882	6.350	101.5	102.3	-0.8
10	6	14	1	methyl pentyl sulfide	1.942	7.350	145.0	144.1	0.9
11	6	14	1	butyl ethyl sulfide	2.033	7.350	144.2	141.1	3.1
12	6	14	1	dipropyl sulfide	2.053	7.350	142.8	140.4	2.4
13	6	14	1	isopropyl propyl sulfide	2.031	7.350	132.0	131.1	0.7
14	6	14	1	ethyl isobutyl sulfide	2.016	7.350	134.2	131.8	2.4
15	6	14	1	isopentyl methyl sulfide	1.930	7.350	137.0	134.6	2.4
16	6	14	1	methyl 2-methylbutyl sulfide	1.934	7.350	139.0	134.5	4.5
17	6	14	1	<i>s</i> -butyl ethyl sulfide	2.031	7.350	133.6	131.3	2.3
18	6	14	1	<i>t</i> -butyl ethyl sulfide	1.988	7.350	120.4	122.9	-2.5
19	6	14	1	diisopropyl sulfide	2.009	7.350	120.0	122.2	-2.2
20	6	14	1	1-ethylpropyl methyl sulfide	1.970	7.350	137.0	133.3	3.7
21	7	16	1	hexyl methyl sulfide	1.952	8.350	171.0	167.9	3.1
22	7	16	1	butyl propyl sulfide	2.078	8.350	166.0	163.7	2.3
23	7	16	1	isobutyl propyl sulfide	2.061	8.350	155.0	154.4	0.6
24	7	16	1	isobutyl isopropyl sulfide	2.039	8.350	145.0	145.3	-0.3
25	7	16	1	ethyl 2-methylbutyl sulfide	2.041	8.350	159.0	155.1	3.9
26	7	16	1	<i>t</i> -butyl propyl sulfide	2.034	8.350	138.0	145.5	-7.5
27	7	16	1	<i>s</i> -butyl isopropyl sulfide	2.054	8.350	142.0	144.8	-2.8
28	7	16	1	ethyl isopentyl sulfide	2.036	8.350	159.0	155.2	3.8
29	7	16	1	butyl isopropyl sulfide	2.056	8.350	163.5	154.6	8.9
30	8	18	1	dibutyl sulfide	2.102	9.350	188.9	187.0	1.9
31	8	18	1	diisobutyl sulfide	2.069	9.350	170.0	168.4	1.6
32	8	18	1	butyl isobutyl sulfide	2.086	9.350	178.0	177.7	0.3
33	8	18	1	di- <i>t</i> -butyl sulfide	2.014	9.350	148.5	150.5	-2.0
34	8	18	1	di- <i>s</i> -butyl sulfide	2.100	9.350	165.0	167.4	-2.4
35	8	18	1	butyl <i>s</i> -butyl sulfide	2.101	9.350	177.0	177.2	-0.2
36	8	18	1	<i>s</i> -butyl isobutyl sulfide	2.084	9.350	167.0	167.9	-0.9
37	8	18	1	heptyl methyl sulfide	1.960	9.350	195.0	191.7	3.3
38	9	20	1	ethyl heptyl sulfide	2.066	10.350	195.0	212.3	-17.3
39	9	20	1	methyl octyl sulfide	1.965	10.350	218.0	215.6	2.4
40	10	22	1	di- <i>t</i> -pentyl sulfide	2.104	11.350	199.0	195.7	3.3
41	10	22	1	dipentyl sulfide	2.133	11.350	228.0	234.2	-6.2
42	10	22	1	diisopentyl sulfide	2.110	11.350	215.0	215.2	-0.2
43	10	22	1	isobutyl 4-methylpentyl sulfide	2.103	11.350	216.0	215.5	0.5
44	10	22	1	methyl nonyl sulfide	1.970	11.350	240.0	239.6	0.4

sulfur atoms is a separate parameter, this parameter also accounts for the nature of the heteroatom(s): oxygen for $N_s = 0$, and sulfur for $N_s > 0$. Thus, it was not necessary to discriminate in J_{het} between the two types of heteroatoms. This results in satisfactory qualitative agreement in calculated BPs with observations 3–5 above on how the presence of one or two heteroatoms in molecular chains influences the BP.

The correlational equation is

$$BP = -41.75(\pm 3.90) + 43.79(\pm 0.69)^1\chi + 45.03(\pm 0.90)N_s - 2.90(\pm 0.41)J_{het} \quad (2)$$

The t values are -10.7, 63.3, 50.3, and -7.1, respectively. The correlation factor is $r^2 = 0.9714$, $F = 2048$, and $s = 8.2$ °C, and $n = 185$. With an average BP of 133 °C, this corresponds to a fit error of 6.2 °C. No significant intercorrelation exists between the three descriptors.

All results, including BPs calculated with this equation, and residuals (difference between experimental and calculated BPs) are shown in Table II.

One can observe that there are only two large outliers (compounds 1 and 25, with residuals between 25 and 30 °C) and two less large ones (nos. 95 and 145 with residuals between 21 and 22 °C). While for dimethyl ether a fairly large deviation is expected because it is the lowest-boiling compound (with a BP difference of more than 30 °C from the next

homologue), it may be possible that the experimental BPs of the other three outlier compounds are in error.

SUBSET CORRELATIONS

In order to test the topological descriptors for more accurate predictions, we ran two separate correlations, one for monoethers and another for monosulfides. We thus eliminated some of the inhomogeneities of the database. We also discarded the first term in these sets (Me_2O and Me_2S , respectively) because otherwise the intercept would have too large a standard error. Two variable correlations were tested, and three-variable correlations with S_0 (electrotopological state for oxygen) as one of the variables were also investigated.

The results are the following: For 72 ethers (with mean BP = 117.2):

$$BP = -51.08(\pm 9.13) + 46.99(\pm 0.95)^1\chi - 3.93(\pm 1.45)J_{het} \quad (3)$$

The t values are -5.6, 49.3, and -2.7, respectively; $r^2 = 0.978$; $s = 6.4$ °C; and $F = 1501$.

$$BP = 46.45(\pm 16.02) - 25.86(\pm 3.81)S_0 - 12.50(\pm 0.66)N_{\text{Me}} + 37.89(\pm 0.70)^0\chi^v \quad (4)$$

The t values are 2.9, -6.8, -18.9, and 54.5, respectively for the coefficients; $r^2 = 0.9872$, $s = 4.89$ °C, $F = 1749$, and $n = 72$.

Calculated values and residuals for the latter correlation are presented in Table III. The two outliers with absolute residuals larger than 12 are the same as the most serious ones in correlation 1 above.

For 44 Sulfides (with mean BP = 150.7):

$$BP = 20.01(\pm 7.49) + 43.93(\pm 0.83)^1\chi - 6.34(\pm 1.20)J_{\text{het}} \quad (5)$$

The t values are 2.7, 52.7, and -5.3, respectively; $r^2 = 0.989$; $s = 4.28$ °C, and $F = 1807$.

$$BP = 50.59(\pm 23.95) - 32.91(\pm 13.45)ES_s - 9.86(\pm 0.68)N_{\text{Me}} + 24.10(\pm 0.54)K\alpha_1 \quad (6)$$

The t values are 2.1, -2.4, -14.5, and 44.3, respectively; $r^2 = 0.989$; $s = 4.3$ °C; and $n = 44$.

Although in this case both types of correlations yield practically the same statistical results, we reproduce in Table IV only the second batch of data. In both correlations there is just one outlier with a boiling point whose experimental accuracy is subject to doubt (ethyl heptyl sulfide).

One may conclude this section by stating that for ethers the best correlation affords a fit error of 4.2 °C and for sulfides a fit error of 2.8 °C.

In conclusion, we have found that four parameters (the molecular connectivity $^1\chi$, the topological index J modified for the presence of heteroatoms, the electrotopological state S of the heteroatoms, and the number N_s of sulfur atoms) give a good correlation with BPs for unknown compounds belonging to these classes of compounds and this range of carbon atoms and heteroatoms.

ACKNOWLEDGMENT

A.T.B. thanks Sterling Drug Inc. for support. The assistance of Mr. B. Brown for computer programming was much appreciated.

REFERENCES AND NOTES

- (1) Balaban, A. T.; Joshi, N.; Kier, L. B.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **1992**, *32* (preceding paper in this issue).
- (2) Stanton, D. T.; Jurs, P. C.; Hicks, M. G. *J. Am. Chem. Soc.* **1991**, *31*, 301.
- (3) Bordwell, F. G.; Anderson, H. M.; Pitt, B. M. *J. Am. Chem. Soc.* **1954**, *76*, 1082.
- (4) Jurecek, M.; Vecera, M. *Chem. Listy* **1954**, *48*, 542.
- (5) SAS Institute Inc. *SAS User's Guide: Statistics, Version 5*; SAS: Cary, NC; 1985.
- (6) The programs MOLCONN and MOLCONN2 can be obtained from Prof. L. H. Hall, Hall Associates Consulting, 2 Davis Street, Quincy, MA 02170.
- (7) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- (8) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (9) Kier, L. B.; Hall, L. H. *"Molecular Connectivity in Structure-Activity Analysis"* Research Studies Press and Wiley: New York, 1986.
- (10) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *80*, 399.
- (11) Balaban, A. T. *Pure Appl. Chem.* **1983**, *55*, 199.
- (12) Balaban, A. T. *Math. Chem.* **1986**, *21*, 115.
- (13) Kier, L. B. *Quant. Struct. Act. Relat.* **1985**, *4*, 109.
- (14) Kier, L. B. *Med. Res. Rev.* **1987**, *7*, 417.
- (15) Kier, L. B. *Quant. Struct.-Act. Relat.* **1986**, *5*, 7.
- (16) Kier, L. B.; Hall, L. H. *Pharm. Res.* **1990**, *7*, 801.
- (17) Hall, L. H.; Mohnney, B.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76.
- (18) White, P. T.; Barnard-Smith, D. G.; Fidler, F. A. *Ind. Eng. Chem.* **1952**, *44*, 1430.
- (19) *CRC Handbook of Physics and Chemistry*, 68th ed.; CRC: Boca Raton, FL.
- (20) *Dictionary of Organic Compounds*, 5th ed.; Chapman and Hall: New York, 1982.

Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited

ARTHUR DALBY, JAMES G. NOURSE,* W. DOUGLAS HOUNSHELL, ANN K. I. GUSHURST, DAVID L. GRIER, BURTON A. LELAND, and JOHN LAUFER

Molecular Design Limited, 2132 Farallon Drive, San Leandro, California 94577

Received January 23, 1992

A series of file formats used for storing and transferring chemical structure information that have evolved over several years at Molecular Design Limited are described. These files are built using one or more connection table (Ctab) blocks. The Ctab block format is described in detail. The file formats described are the MOLfile for a single (multifragment) molecule, the RGfile for a generic query, the SDfile for multiple structures and data, the RXNfile for a single reaction, and the RDfile for multiple reactions and data. The relationships of these files are given as well as examples.

1. INTRODUCTION

This paper describes the chemical table file (CTfile) formats currently used in a wide variety of chemical structure-manipulating computer programs. These file formats were developed by a large number of people at Molecular Design Limited (MDL) over the past 13 years.¹ While the formats were developed for use with the various MDL programs, their use has gone well beyond this role.

The evolution of the CTfile formats did not proceed by a well-defined plan. Changes were frequently made over the years to accommodate new program features or application needs. While the changes often added new data fields or appendices, they were done such that older files would remain valid and older programs could read the portion of new files

they could interpret. Because of this policy, as well as the widespread use of programs which read and write CTfiles, there is probably more valid chemical structure information in existence in these formats than in any other format, current or proposed.

While the evolution of these CTfile formats was closely tied to the development of the various MDL computer programs, the purpose of this paper is to describe just the file formats and not the various computer programs and their features. In general, references to the various programs which use these file formats will be made only when necessary to describe the purpose of a particular part of a file format. Literature references will be given in place of detailed descriptions of various program features.