

CHROM. 7690

## GAS CHROMATOGRAPHIC RETENTION AND MOLECULAR STRUCTURE

### AN APPROACH TO SYSTEMATIC IDENTIFICATION OF FUNCTIONAL GROUPS AND CARBON NUMBER

ANDRZEJ W. LADON

*The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario (Canada)*

---

#### SUMMARY

On the basis of an equation employing retentions  $t_1$  and  $t_2$  on two stationary phases:

$$\log t_1 = a \cdot \log t_2 + a_f$$

where  $a$  is a constant for all solutes and  $a_f$  is a term characteristic of the particular association of functional groups, one can determine the class to which a given solute belongs. The carbon number  $n$  can then be calculated from another equation:

$$n = b_1 \cdot \log t_1 + b_2 \cdot \log t_2 + b_f$$

where  $b_1$  and  $b_2$  are a set of coefficients that is the same for all solutes, and  $b_f$  is a term characteristic of the functional groups.

---

#### INTRODUCTION

James and Martin<sup>1</sup> in 1952 drew attention to the explicit relationship between gas chromatographic retention and molecular structure. Since that time little has been done to develop systematic procedures for elucidation of the structure of unknown substances. The multiplicity of compounds revealed by gas chromatography precludes any possibility of identifying structures on a regular basis through a comparison with a synthesized standard. At present, little structural information is obtained from high-resolution chromatograms of complex mixtures. The value of gas chromatography would be enormously enhanced if retention times were applied for routine structure elucidation.

Smith *et al.*<sup>2</sup> plotted retentions of large numbers of hydrocarbons on two stationary phases. They found distinct zones on their diagrams, which enclosed the members of particular classes of substances such as alkanes, alkenes, cycloalkanes, cycloalkenes, alkadienes, cycloalkadienes and alkynes. Similarly, plotting retentions at two temperatures on one stationary phase, Bricteux and Duyckaerts<sup>3</sup> distinguished the cycloparaffins from the paraffins.

Walraven *et al.*<sup>4</sup> found that the zones of particular classes of compounds could be further divided into smaller segments representing isomers of particular carbon numbers. Moreover, other zones exist representing compounds with more detailed structural similarity (like substances having the same content of primary, secondary, tertiary and quaternary carbon atoms). Subsequently these fine-structure zones were treated mathematically<sup>5</sup>.

A formula was then derived<sup>6</sup> for determining the carbon number  $n$  within classes of substances. This formula is a straightforward extension of the James–Martin<sup>1</sup> rule for the homologous series and employs three isothermal retentions  $t_1, t_2, t_3$ :

$$n = C_0 + C_1 \cdot \log t_1 + C_2 \cdot \log t_2 + C_3 \cdot \log t_3$$

where  $C_0, C_1, C_2, C_3$  are constants. Using groups of hydrocarbons of known structure this formula has been successfully tested for three stationary phases<sup>6,7</sup>, for three different temperatures on one stationary phase, as well as for one stationary phase at a particular temperature combined with another stationary phase at two different temperatures<sup>6</sup>.

From all these findings a system of correlations emerges between gas chromatographic retention and (1) the functional groups of compounds (that is, the class); (2) the carbon number of a molecule when the class is known; (3) the fine structure, including the branching.

If we now apply these correlations, we can identify unknown peaks on gas chromatographic tracings to a high degree of certainty. To do so we must measure their retentions on two or three stationary phases and at several temperatures. We must also have retention data for known compounds of corresponding classes. Then we have the framework for the regression equations used subsequently in the identification scheme.

## THEORY

The considerations in the present work are limited to isothermal gas chromatographic operation but a similar concept of identification could be applied with the temperature programming.

### *The class-distinction equation*

The logarithm of adjusted retention time  $t$  for all substances, on a given stationary phase, can be represented as a sum:

$$\log t = f + r \tag{1}$$

where  $f$  is the contribution due to functional groups and  $r$  is the contribution of the rest of the molecule.

For  $m$  retention times  $t_1, t_2, \dots, t_i, \dots, t_m$  of any solute at certain temperature on different stationary phases with subscripts  $1, 2, \dots, i, \dots, m$ , (or on one stationary phase at different temperatures denoted by  $1, 2, \dots, i, \dots, m$ )  $m$  equations of type 1 can be written:

$$\begin{aligned}
 \log t_1 &= f_1 + r_1 \\
 \log t_2 &= f_2 + r_2 \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 \log t_i &= f_i + r_i \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 \log t_m &= f_m + r_m
 \end{aligned} \tag{2}$$

where  $f_i$  and  $r_i$  are the corresponding contributions. If a set of numbers  $R_0, R_1, R_2, \dots, R_i, \dots, R_m$  exists such that the criterion:

$$\sum_{i=1}^m R_i \cdot r_i = R_0 \tag{3}$$

is fulfilled, then the residual term  $r_1, r_2, \dots, r_i, \dots, r_m$  values derived from the system of eqns. 2 can be substituted in eqn. 3. The result is:

$$\sum_{i=1}^m R_i \cdot (\log t_i - f_i) = R_0 \tag{4}$$

or

$$\sum_{i=1}^m R_i \cdot \log t_i = R_0 + \sum_{i=1}^m R_i \cdot f_i = R_f \tag{5}$$

where  $R_f$  is a term characteristic of a given class.

#### *The carbon number equation*

Let us assume that the residual term  $r$ , in eqn. 1, is composed of two parts:

$$r = g + s \cdot n \tag{6}$$

where  $g$  corresponds to the geometry of a molecule and varies from one molecule to the other,  $s$  is a constant increment corresponding to a single carbon atom and  $n$  is the number of carbon atoms in the molecule. The geometry is viewed in the present work as a spatial distribution of constitutional elements and thus includes the connectivity of atoms, the angles between the bonds (conformation), as well as the distinctive arrangement of atoms about either asymmetric atoms or rigid parts of the molecule (configuration).

For  $m$  adjusted retention times  $t_1, t_2, \dots, t_i, \dots, t_m$  of any solute on different stationary phases denoted by 1, 2,  $\dots, i, \dots, m$ , (or on one stationary phase at different temperatures with corresponding subscripts)  $m$  equations can be written:

$$\begin{aligned}
 \log t_1 &= f_1 + g_1 + s_1 \cdot n \\
 \log t_2 &= f_2 + g_2 + s_2 \cdot n \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 \log t_i &= f_i + g_i + s_i \cdot n \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 \log t_m &= f_m + g_m + s_m \cdot n
 \end{aligned} \tag{7}$$

where  $f_i$ ,  $g_i$  and  $s_i$  are the contributions of corresponding aspects of the structure of a given molecule. Let us assume that the geometrical terms  $g_1, g_2, \dots, g_i, \dots, g_m$  for all the molecules conform to the condition:

$$\sum_{i=1}^m G_i \cdot g_i = G_0 \quad (8)$$

where  $G_0, G_1, G_2, \dots, G_i, \dots, G_m$  are constants.

The  $g_1, g_2, \dots, g_i, \dots, g_m$ , values obtained from the system of eqns. 7 can be substituted in eqn. 8 to give:

$$\sum_{i=1}^m G_i \cdot (\log t_i - f_i - s_i \cdot n) = G_0 \quad (9)$$

If eqn. 9 is rearranged to yield  $n$  and auxiliary parameters introduced:

$$b_f = -\left(G_0 + \sum_{i=1}^m G_i \cdot f_i\right) / \sum_{i=1}^m G_i \cdot s_i \quad (10)$$

as well as for  $i$  varying from 1 to  $m$ ,

$$b_i = G_i / \sum_{i=1}^m G_i \cdot s_i \quad (11)$$

then

$$n = b_f + \sum_{i=1}^m b_i \cdot \log t_i \quad (12)$$

Coefficient  $b_f$  is dependent upon the class to which the molecule belongs while coefficients  $b_i$  remain constant for all of the molecules, irrespective of the class.

## TWO STATIONARY PHASES

Several pairs of stationary phases have been investigated mathematically in

TABLE I

PARAMETERS OF THE CLASS DISTINCTION EQUATION OBTAINED THROUGH A REGRESSION PROCEDURE, USING SQ vs. DMS

Class of hydrocarbons	Parameters	
	$a$	$a_f$
Alkanes	1.1586	-1.2109
Cycloalkanes	1.1586	-1.3565
Alkenes	1.1586	-1.5284
Cycloalkenes	1.1586	-1.7107
Alkadienes	1.1586	-1.9372
Common residual standard deviation		0.0456

the present study. Dimethylsulpholane (DMS) and squalane (SQ) have been selected to illustrate the approach to the identification of functional groups and carbon number. Relative retention data for alkanes, cycloalkanes, alkenes, cycloalkenes and alkadienes were taken from those measured by Cramers<sup>8</sup> on DMS at 25° and by Hively and Hinton<sup>9</sup> on SQ at 27°.

TABLE II

COMPARISON OF MEASURED RETENTIONS ON SQ WITH THOSE CALCULATED FROM THE CLASS DISTINCTION EQUATION USING RETENTIONS ON DMS, FOR ALKANES, CYCLOALKANES, ALKENES, CYCLOALKENES AND ALKADIENES

No.	Name of compound	Logarithms of relative retention			Difference between calculated and measured values for SQ, 27°
		Measured		Calcd.	
		DMS, 25°	SQ, 27°	SQ, 27°	
1	Ethane	-2.0458	-3.5346	-3.5811	-0.0464
2	Propane	-1.5376	-2.9867	-2.9923	-0.0056
3	2-Methylpropane	-1.2441	-2.6300	-2.6523	-0.0223
4	Butane	-1.0506	-2.4401	-2.4281	0.0120
5	2,2-Dimethylpropane	-1.0655	-2.3840	-2.4454	-0.0613
6	2-Methylbutane	-0.7399	-2.0670	-2.0682	-0.0012
7	Pentane	-0.6162	-1.9359	-1.9248	0.0111
8	2,2-Dimethylbutane	-0.4921	-1.7642	-1.7811	-0.0169
9	2,3-Dimethylbutane	-0.3382	-1.6146	-1.6027	0.0118
10	2-Methylpentane	-0.3382	-1.5964	-1.6027	-0.0064
11	3-Methylpentane	-0.2588	-1.5286	-1.5108	0.0177
12	Hexane	-0.1979	-1.4464	-1.4402	0.0061
13	2,2-Dimethylpentane	-0.1215	-1.3298	-1.3517	-0.0219
14	2,4-Dimethylpentane	-0.1068	-1.3063	-1.3346	-0.0284
15	2,2,3-Trimethylbutane	-0.0362	-1.2692	-1.2529	0.0163
16	3,3-Dimethylpentane	0.0445	-1.1759	-1.1593	0.0166
17	2,3-Dimethylpentane	0.1092	-1.1090	-1.0844	0.0247
18	2-Methylhexane	0.0603	-1.1273	-1.1410	-0.0138
19	3-Methylhexane	0.1139	-1.0814	-1.0789	0.0025
20	3-Ethylpentane	0.1761	-1.0376	-1.0069	0.0307
21	Heptane	0.2117	-0.9650	-0.9657	-0.0007
22	2,2,4-Trimethylpentane	0.1219	-1.0227	-1.0697	-0.0470
23	2,2-Dimethylhexane	0.2586	-0.8768	-0.9113	-0.0345
24	2,5-Dimethylhexane	0.2956	-0.8336	-0.8685	-0.0349
25	2,4-Dimethylhexane	0.3214	-0.8159	-0.8386	-0.0227
26	2,2,3-Trimethylpentane	0.3720	-0.8044	-0.7799	0.0244
27	3,3-Dimethylhexane	0.3844	-0.7724	-0.7656	0.0067
28	2,3,4-Trimethylpentane	0.4428	-0.7291	-0.6979	0.0312
29	2,3,3-Trimethylpentane	0.4814	-0.7042	-0.6531	0.0511
30	2,3-Dimethylhexane	0.4670	-0.6811	-0.6699	0.0112
31	2-Methyl-3-ethylpentane	0.4804	-0.6828	-0.6543	0.0285
32	2-Methylheptane	0.4591	-0.6580	-0.6790	-0.0211
33	4-Methylheptane	0.4823	-0.6463	-0.6521	-0.0059
34	3,4-Dimethylhexane	0.5236	-0.6371	-0.6043	0.0329

(Continued on p. 208)

TABLE II (continued)

No.	Name of compound	Logarithms of relative retention			Difference between calculated and measured values for SQ, 27°
		Measured		Calcd.	
		DMS, 25°	SQ, 27°	SQ, 27°	
35	3-Methyl-3-ethylpentane	0.5411	-0.6326	-0.5840	0.0486
36	3-Ethylhexane	0.5184	-0.6260	-0.6103	0.0156
37	3-Methylheptane	0.5038	-0.6187	-0.6273	-0.0085
38	Octane	0.6284	-0.4824	-0.4829	-0.0005
39	Cyclopentane	-0.1469	-1.6284	-1.5267	0.1017
40	Methylcyclopentane	0.0565	-1.3288	-1.2910	0.0379
41	Cyclohexane	0.2106	-1.1649	-1.1125	0.0525
42	1,1-Dimethylcyclopentane	0.2063	-1.1129	-1.1175	-0.0045
43	1-cis-3-Dimethylcyclopentane	0.2143	-1.0640	-1.1082	-0.0442
44	1-trans-3-Dimethylcyclopentane	0.2405	-1.0448	-1.0778	-0.0330
45	1-trans-2-Dimethylcyclopentane	0.2601	-1.0339	-1.0551	-0.0213
46	1-cis-2-Dimethylcyclopentane	0.4354	-0.8854	-0.8521	0.0333
47	Methylcyclohexane	0.4257	-0.8658	-0.8633	0.0026
48	Ethylcyclopentane	0.4878	-0.8228	-0.7913	0.0315
49	1,1,3-Trimethylcyclopentane	0.3400	-0.8703	-0.9625	-0.0922
50	1-trans-2-cis-4-Trimethylcyclopentane	0.4071	-0.7833	-0.8849	-0.1016
51	1-trans-2-cis-3-Trimethylcyclopentane	0.4586	-0.7503	-0.8251	-0.0748
52	1-cis-2-trans-4-Trimethylcyclopentane	0.5799	-0.6336	-0.6846	-0.0510
53	Cycloheptane	0.8000	-0.5499	-0.4296	0.1203
54	Isopropylcyclopentane	0.7896	-0.4504	-0.4417	0.0087
55	Propylcyclopentane	0.8808	-0.3565	-0.3360	0.0206
56	Ethylcyclohexane	0.8831	-0.3468	-0.3333	0.0134
57	Propene	-1.2757	-3.0595	-3.0064	0.0531
58	2-Methylpropene	-0.8041	-2.5321	-2.4600	0.0721
59	1-Butene	-0.8386	-2.5230	-2.5000	0.0230
60	trans-2-Butene	-0.7190	-2.4056	-2.3614	0.0442
61	cis-2-Butene	-0.6459	-2.3565	-2.2767	0.0798
62	3-Methyl-1-butene	-0.5918	-2.1918	-2.2140	-0.0222
63	1-Pentene	-0.4237	-2.0315	-2.0193	0.0123
64	2-Methyl-1-butene	-0.3507	-1.9965	-1.9347	0.0618
65	trans-2-Pentene	-0.3279	-1.9303	-1.9083	0.0220
66	cis-2-Pentene	-0.2899	-1.9112	-1.8643	0.0469
67	2-Methyl-2-butene	-0.2233	-1.8681	-1.7871	0.0809
68	3,3-Dimethyl-1-butene	-0.3872	-1.9090	-1.9770	-0.0680
69	4-Methyl-1-pentene	-0.1739	-1.6977	-1.7299	-0.0323
70	2,3-Dimethyl-1-butene	-0.0894	-1.6538	-1.6320	0.0219
71	4-Methyl-cis-2-pentene	-0.1379	-1.6629	-1.6882	-0.0252
72	4-Methyl-trans-2-pentene	-0.1238	-1.6319	-1.6718	-0.0399
73	2-Methyl-1-pentene	0.0166	-1.5450	-1.5092	0.0358
74	1-Hexene	0.0000	-1.5346	-1.5284	0.0062
75	2-Ethyl-1-butene	0.0741	-1.4856	-1.4426	0.0430
76	cis-3-Hexene	0.0445	-1.4828	-1.4768	0.0060
77	trans-3-Hexene	0.0220	-1.4785	-1.5029	-0.0244
78	2-Methyl-2-pentene	0.0986	-1.4561	-1.4141	0.0419
79	3-Methyl-trans-2-pentene	0.1729	-1.3840	-1.3281	0.0559
80	trans-2-Hexene	0.0535	-1.4592	-1.4665	-0.0073

TABLE II (continued)

No.	Name of compound	Logarithms of relative retention			Difference between calculated and measured values for SQ, 27°
		Measured		Calcd.	
		DMS, 25°	SQ, 27°	SQ, 27°	
81	<i>cis</i> -2-Hexene	0.1092	-1.4295	-1.4019	0.0276
82	2-Methyl- <i>cis</i> -2-pentene	0.1358	-1.4355	-1.3711	0.0643
83	4,4-Dimethyl-1-pentene	0.0162	-1.4309	-1.5097	-0.0788
84	2,3-Dimethyl-2-butene	0.2610	-1.3261	-1.2260	0.1000
85	4,4-Dimethyl- <i>trans</i> -2-pentene	0.0561	-1.3716	-1.4634	-0.0918
86	3,3-Dimethyl-1-pentene	0.1176	-1.3307	-1.3922	-0.0615
87	2,3,3-Trimethyl-1-butene	0.1764	-1.3224	-1.3241	-0.0017
88	4,4-Dimethyl- <i>cis</i> -2-pentene	0.1772	-1.2857	-1.3231	-0.0374
89	3,4-Dimethyl-1-pentene	0.1664	-1.2782	-1.3356	-0.0574
90	2,4-Dimethyl-1-pentene	0.2087	-1.2716	-1.2866	-0.0150
91	2,4-Dimethyl-2-pentene	0.2079	-1.2472	-1.2876	-0.0404
92	3-Methyl-1-hexene	0.2055	-1.2388	-1.2904	-0.0515
93	3-Ethyl-1-pentene	0.1940	-1.2277	-1.3037	-0.0760
94	2,3-Dimethyl-1-pentene	0.2615	-1.2118	-1.2255	-0.0136
95	5-Methyl-1-hexene	0.2646	-1.2090	-1.2219	-0.0129
96	2-Methyl- <i>trans</i> -3-hexene	0.1875	-1.2147	-1.3112	-0.0965
97	3-Methyl-2-ethyl-1-butene	0.3047	-1.1643	-1.1754	-0.0111
98	4-Methyl- <i>cis</i> -2-hexene	0.2499	-1.1864	-1.2389	-0.0524
99	4-Methyl-1-hexene	0.2891	-1.1713	-1.1934	-0.0221
100	4-Methyl- <i>trans</i> -2-hexene	0.2420	-1.1759	-1.2480	-0.0721
101	5-Methyl- <i>trans</i> -2-hexene	0.2676	-1.1586	-1.2183	-0.0597
102	3,4-Dimethyl- <i>trans</i> -2-pentene	0.3959	-1.0691	-1.0698	-0.0008
103	2-Methyl-1-hexene	0.4198	-1.0721	-1.0421	0.0300
104	3-Methyl- <i>trans</i> -3-hexene	0.4445	-1.0376	-1.0134	0.0242
105	1-Heptene	0.4077	-1.0535	-1.0560	-0.0025
106	3-Methyl- <i>cis</i> -2-hexene	0.5039	-0.9618	-0.9446	0.0172
107	2-Ethyl-1-pentene	0.4195	-1.0501	-1.0425	0.0077
108	3-Methyl- <i>trans</i> -2-hexene	0.4859	-0.9996	-0.9655	0.0340
109	3-Methyl- <i>cis</i> -3-hexene	0.4559	-1.0052	-1.0002	0.0050
110	2-Methyl-2-hexene	0.4625	-1.0039	-0.9925	0.0114
111	<i>cis</i> -3-Heptene	0.4366	-1.0119	-1.0226	-0.0107
112	<i>trans</i> -3-Heptene	0.3941	-1.0218	-1.0718	-0.0500
113	3-Ethyl-2-pentene	0.5072	-0.9784	-0.9408	0.0376
114	2,3-Dimethyl-2-pentene	0.5564	-0.9531	-0.8838	0.0693
115	Cyclopentene	0.0149	-1.7069	-1.6934	0.0135
116	1-Methylcyclopentene	0.4011	-1.2426	-1.2461	-0.0035
117	Cyclohexene	0.5415	-1.1244	-1.0834	0.0410
118	3-Ethylcyclopentene	0.6203	-0.9190	-0.9920	-0.0730
119	1-Methylcyclohexene	0.9016	-0.6882	-0.6662	0.0221
120	1,3-Butadiene	-0.4724	-2.5176	-2.4845	0.0331
121	2-Methyl-1,3-butadiene	-0.0070	-1.9539	-1.9453	0.0086
122	3-Methyl-1,2-butadiene	0.0095	-1.8788	-1.9262	-0.0475
123	<i>trans</i> -1,3-Pentadiene	0.1252	-1.8620	-1.7922	0.0698
124	<i>cis</i> -1,3-Pentadiene	0.1752	-1.8187	-1.7342	0.0845
125	1,2-Pentadiene	0.0966	-1.8091	-1.8253	-0.1062
126	2,3-Pentadiene	0.1072	-1.7788	-1.8130	-0.0341
127	1,5-Hexadiene	0.1838	-1.6260	-1.7242	-0.0982

*Computation of the class-distinction equation*

For relative retention times on  $m = 2$  stationary phases, eqn. 5 can be formulated as:

$$\log t_1 = a \cdot \log t_2 + a_f \quad (13)$$

where  $a$  is constant for all solutes, and  $a_f$  is a term characteristic of the particular association of functional groups. The relative retention times are  $t_1$  and  $t_2$  for any substance on the two selected stationary phases. Eqn. 13 describes a series of parallel lines, each one corresponding to a particular class of compounds on a logarithmic diagram.

Retention data for five classes of hydrocarbons have been subjected to a regression procedure for eqn. 13 and the corresponding regression parameters are reported in Table I. The specific terms are approximately additive with respect to carbon-to-carbon double bond as well as the cyclization although the latter contribution is smaller. The detailed list of retention data, measured and calculated, is given in Table II.

*Computation of the carbon number equation*

An equation with relative retention times, analogous to eqn. 12, can be written for  $m = 2$  stationary phases in the form:

$$n = b_1 \cdot \log t_1 + b_2 \cdot \log t_2 + b_f \quad (14)$$

where  $b_1$  and  $b_2$  are a set of coefficients, the same for all solutes, and  $b_f$  is a term characteristic of a particular association of functional groups and thus the same for compounds only within a given class. The symbols  $t_1$  and  $t_2$  represent the relative retention times on two chosen stationary phases. On logarithmic diagrams, eqn. 14 describes a system of parallel lines, divided into series, corresponding to each class of compounds. These lines connect points for compounds with a particular carbon number. The distance between any two neighbouring carbon number lines within a class is the same throughout the whole system.

Retention data for five classes of hydrocarbons have been used in a regression procedure to fit eqn. 14. Relative retention times on DMS and SQ are respectively  $t_1$  and  $t_2$ . The corresponding regression parameters are reported in Table III. Similar to

TABLE III

PARAMETERS OF THE CARBON NUMBER EQUATION OBTAINED THROUGH A REGRESSION PROCEDURE, USING DMS AND SQ

Class of hydrocarbons	Parameters		
	$b_1$ (DMS)	$b_2$ (SQ)	$b_f$
Alkanes	-5.851	7.121	15.52
Cycloalkanes	-5.851	7.121	15.82
Alkenes	-5.851	7.121	17.00
Cycloalkenes	-5.851	7.121	17.19
Alkadienes	-5.851	7.121	18.73
Common residual standard deviation			0.21



TABLE IV

COMPARISON OF ACTUAL CARBON NUMBERS WITH THOSE CALCULATED FROM THE CARBON NUMBER EQUATION FOR ALKANES, CYCLOALKANES, ALKENES, CYCLOALKENES AND ALKADIENES USING RETENTION DATA ON DMS AND SQ  
Numbers refer to the compounds listed in Table II.

<i>No. of compound</i>	<i>Carbon number</i>		<i>Difference</i>
	<i>Real</i>	<i>Calculated</i>	
1	2	2.32	0.32
2	3	3.25	0.25
3	4	4.08	0.08
4	4	4.30	0.30
5	5	4.78	-0.22
6	5	5.13	0.13
7	5	5.34	0.34
8	6	5.84	-0.16
9	6	6.01	0.01
10	6	6.14	0.14
11	6	6.15	0.15
12	6	6.38	0.38
13	7	6.77	-0.23
14	7	6.85	-0.15
15	7	6.70	-0.30
16	7	6.89	-0.11
17	7	6.99	-0.01
18	7	7.14	0.14
19	7	7.16	0.16
20	7	7.11	0.11
21	7	7.41	0.41
22	8	7.53	-0.47
23	8	7.77	-0.23
24	8	7.86	-0.14
25	8	7.83	-0.17
26	8	7.62	-0.38
27	8	7.78	-0.22
28	8	7.74	-0.26
29	8	7.69	-0.31
30	8	7.94	-0.06
31	8	7.85	-0.15
32	8	8.15	0.15
33	8	8.10	0.10
34	8	7.92	-0.08
35	8	7.85	-0.15
36	8	8.03	0.03
37	8	8.17	0.17
38	8	8.41	0.41
39	5	5.09	0.09
40	6	6.03	0.03
41	6	6.29	0.29
42	7	6.69	-0.31
43	7	6.99	-0.01
44	7	6.97	-0.03
45	7	6.94	-0.06
46	7	6.97	-0.03

(Continued on p. 212)

TABLE IV (continued)

No. of compound	Carbon number		Difference
	Real	Calculated	
47	7	7.17	0.17
48	7	7.11	0.11
49	8	7.64	-0.36
50	8	7.86	-0.14
51	8	7.80	-0.20
52	8	7.92	-0.08
53	7	7.23	0.23
54	8	8.00	0.00
55	8	8.13	0.13
56	8	8.19	0.19
57	3	2.68	-0.32
58	4	3.67	-0.33
59	4	3.94	-0.06
60	4	4.07	0.07
61	4	4.00	0.00
62	5	4.85	-0.15
63	5	5.01	0.01
64	5	4.83	-0.17
65	5	5.17	0.17
66	5	5.09	0.09
67	5	5.00	0.00
68	6	5.67	-0.33
69	6	5.93	-0.07
70	6	5.74	-0.26
71	6	5.96	-0.04
72	6	6.10	0.10
73	6	5.90	-0.10
74	6	6.07	0.07
75	6	5.99	-0.01
76	6	6.18	0.18
77	6	6.34	0.34
78	6	6.05	0.05
79	6	6.13	0.13
80	6	6.29	0.29
81	6	6.18	0.18
82	6	5.98	-0.02
83	7	6.71	-0.29
84	6	6.03	0.03
85	7	6.90	-0.10
86	7	6.83	-0.17
87	7	6.55	-0.45
88	7	6.81	-0.19
89	7	6.92	-0.08
90	7	6.72	-0.28
91	7	6.90	-0.10
92	7	6.97	-0.03
93	7	7.12	0.12
94	7	6.84	-0.16
95	7	6.84	-0.16
96	7	7.25	0.25

TABLE IV (continued)

No. of compound	Carbon number		Difference
	Real	Calculated	
97	7	6.92	-0.08
98	7	7.09	0.09
99	7	6.97	-0.03
100	7	7.21	0.21
101	7	7.18	0.18
102	7	7.07	0.07
103	7	6.91	-0.09
104	7	7.01	0.01
105	7	7.11	0.11
106	7	7.20	0.20
107	7	7.07	0.07
108	7	7.04	0.04
109	7	7.17	0.17
110	7	7.14	0.14
111	7	7.24	0.24
112	7	7.42	0.42
113	7	7.06	0.06
114	7	6.96	-0.04
115	5	4.95	-0.05
116	6	6.00	0.00
117	6	6.02	0.02
118	7	7.02	0.02
119	7	7.02	0.02
120	4	3.57	-0.43
121	5	4.86	-0.14
122	5	5.30	0.30
123	5	4.74	-0.26
124	5	4.75	-0.25
125	5	5.28	0.28
126	5	5.43	0.43
127	6	6.08	0.08

the specific term  $a_f$  of the class-distinction equation, the specific term  $b_f$  exhibits an additive character with respect to functional groups, which explains the observed order of specific terms. The detailed list of carbon numbers and calculated carbon numbers is given in Table IV.

## DISCUSSION

Organic substances may be systematically identified using a set of linear equations. These equations employ a number of retention parameters (for two stationary phases in the present work) and an increasing number of terms related to structure. The term  $a_f$  can be determined from eqn. 13 and points to the particular combination of functional groups. The latter information can be used in eqn. 14 to select the right term  $b_f$  and calculate the carbon number. Of course, more equations can be added to these two and the structural analyses expanded. Branching, for example, can be established by equations analogous to the one reported by Ladon and Walraven<sup>5</sup>.

A very important problem is the exactness required in determining the terms related to structure. The range of calculated  $a_f$  terms should be narrow enough so that no overlap between classes of compounds with close  $a_f$  values occurs. Such an overlap obscures the identification. The best solution to this problem is to have a safe, unoccupied distance between zones occupied by different classes on the structural dimension. This can be achieved by experimenting with new or greater combinations of stationary phases. The same approach can be used to determine carbon numbers. If the difference between the calculated and real carbon number tends to exceed 0.5, an error can arise.

With DMS and squalane some slight overlap exists between classes of compounds, in the class-distinction equation. However, if the values obtained from the carbon number equation are rounded to the nearest integer, they are equal to the actual carbon numbers.

The exact agreement between calculated and real values will never be possible, which is the essence of the "non-calculability" principle<sup>10</sup>. This should not obscure the fact that some correlations are better than others, and the differences between calculation and reality can always be minimized to such an extent that the applicability of an identification scheme becomes safe. Some obvious illustrations of this possibility have already been reported in previous work<sup>6</sup> concerning carbon number in correlations on three stationary phases as well as with the retentions at three different temperatures on one stationary phase.

## CONCLUSION

The correlation between retention and molecular class and the correlation between retention and carbon number can be expressed by two mathematical formulae which, if used in succession, initiate a systematic procedure for the total structural elucidation of an unknown substance.

## ACKNOWLEDGEMENT

The author thanks J. A. Lowden for help in the preparation of the manuscript.

## REFERENCES

- 1 A. T. James and A. J. P. Martin, *Biochem. J.*, 50 (1952) 679.
- 2 B. Smith, R. Ohlson and G. Larson, *Acta Chem. Scand.*, 17 (1963) 436.
- 3 J. Bricteux and G. Duyckaerts, *J. Chromatogr.*, 22 (1966) 221.
- 4 J. J. Walraven, A. W. Ladon and A. I. M. Keulemans, *Chromatographia*, 1 (1968) 195.
- 5 A. W. Ladon and J. J. Walraven, in E. Sz. Kováts (Editor), *5th International Symposium on Separation Methods: Column Chromatography, Lausanne, 1969*, Sauerlander A.G., Aarau, 1970, p. 167.
- 6 A. W. Ladon, *Ph.D. Thesis*, University of Toronto, Toronto, 1973.
- 7 A. W. Ladon and S. Sandler, *Anal. Chem.*, 45 (1973) 921.
- 8 C. A. Cramers, *Ph.D. Thesis*, Technical University Eindhoven, Eindhoven, 1967.
- 9 R. A. Hively and R. E. Hinton, *J. Gas Chromatogr.*, 6 (1968) 203.
- 10 A. W. Ladon, *Chromatographia*, 4 (1971) 171.