Journal of Chromatography, 158 (1978) 43–56
© Elsevier Scientific Publishing Company, Amsterdam

CHROM. 11,203

TOPOLOGICAL ANALYSIS: A TECHNIQUE FOR THE PHYSICO-CHEMICAL EXPLOITATION OF RETENTION DATA IN GAS–LIQUID CHROMATOGRAPHY

JACQUES R. CHRÉTIEN and JACQUES-EMILE DUBOIS

Institut de Topologie et de Dynamique des Systèmes, Université Paris VII, associé au C.N.R.S., 1 Rue Guy-de-la-Brosse, 75005 Paris (France)

SUMMARY

For saturated aliphatic esters, topological analysis has been applied to the chromatographic data (Kováts retention indices and retention index increments ΔI) and to solute factors (a, c) resulting from previous factorial analysis. These different sets of data can be used to obtain a quantitative evaluation of the specific solute–stationary phase interactions localized at the level of the skeletal carbon atoms of the solute molecules.

INTRODUCTION

We recently presented topological analysis as a means for the physico-chemical exploitation of chromatographic retention data[1]. Starting from molecular information, interest in topological analysis lies in obtaining physico-chemical information localized, for example, at the level of the carbon atoms of the carbon skeleton of the solute molecules. Our purpose here is to show that it is possible in this way to achieve a quantitative evaluation of specific solute–stationary phase interactions.

In chromatography, there are often several series of Kováts retention indices relating to a simple set of compounds or to isotopologous sets of compounds. Their dependence can be expressed in thermodynamic terms by retention index increments (ΔI) or by isotopology factors (τI) (ref. 2), and their applications are either of a predictive or of a physico-chemical nature. When we presented the first type of application for improving the prediction of Kováts indices, we showed, for example, how part of the information associated with new compounds (dibromoalkanes, isomers and diastereoisomers) can be estimated from that for their parent isotopologous alkenes[3]. Therefore, we shall limit our discussion here to physico-chemical applications.

The physico-chemical applications of topological analysis were considered in a previous paper dealing with the topological analysis of the gas–liquid behaviour of alkenes[1] based on the use of Kováts indices for a set of about 50 alkenes analysed on five different stationary phases. In that paper, we showed that the contributions associated with the topological sites in the topology–information diagrams have a

clear chromatographic meaning that can be used for a physico-chemical purpose. However, stress was laid on the underlying ideas and principles of the method. In this paper, stress will be laid on results.

Thus, we have chosen a set of saturated esters from a major experimental study and data treatment by Ashes and Haken[4,5] in which factorial analysis was used. So as to show the different levels of data treatment where topological analysis can come into play, we shall successively use the Kováts indices for 99 saturated esters measured with SE-30 as stationary phase (weakly polar), those measured with Silar 5 cp as stationary phase (strongly polar), the corresponding Kováts index increments (ΔI) and substance polarity factors or solute factors (a, c). We shall then show that factorial analysis and topological analysis are complementary, and shall consider the evolution of statistical tests, in accordance with the criteria defined by Souter[6], depending on the quality of the information treated: raw experimental data (I_A, I_P), derived data (ΔI, ...) or calculated data (solute factor, ...). Thus, on a more general level, we shall show how, and up to which limits, it is possible to use these different sets of data in order to arrive at a quantitative evaluation of specific solute–stationary phase interactions.

METHOD

Calculations are based on the establishment of topology–information correlations, and require the organization of a set of compounds and the topological description of their molecular environment starting from an origin taken as the focus, in accordance with the concepts of the DARC (*D*escription, *A*cquisition, *R*etrieval and *C*omputer-aided design) topological system[7]. The DARC/PELCO (*P*erturbation of an *E*nvironment that is *L*imited, *C*oncentric and *O*rdered) procedure has been used for the establishment of the correlations[8].

As the principles of the DARC/PELCO method have already been discussed with respect to their applications in chromatography[9], we shall not dwell on them here. However, certain terms and points should be recalled. Fig. 1 shows that the set of ten esters R–COO–R′ are derived from methyl formate by progressive substitution. The pattern (COOC) is taken as the focus. A graph is associated with each of these esters. The topological sites correspond to the nodes of the graph. Here, these nodes correspond to the skeletal carbon atoms of the ester molecules. Superposition of these elementary graphs gives the characteristic imprint or trace of this set of compounds. The two development directions, DD_1 and DD_2, are non-equivalent and correspond to the description of the alkanol and acid chain, respectively (*i.e.*, R′ and R).

Fig. 2 gives the imprint of Ashes and Haken's[4,5] population of 99 esters (Table I). The generation order of all sites whose appropriate information contribution must be taken into account is expressed by the concept of an *E*nvironment that is *L*imited *C*oncentric and *O*rdered (ELCO). Each site is localized in the ELCO by a linear order labelling A_i or B_{ij} (Fig. 2b) and its influence is interpreted as a perturbation term. For example, the evaluation of term B_{12} of the first environment E_B^1 of the second development direction DD_2 (Fig. 2b) corresponds to the difference between information I_9 and information I_8 in Fig. 1.

These perturbation terms are the components of vector $\vec{I}(m)$ that characterizes

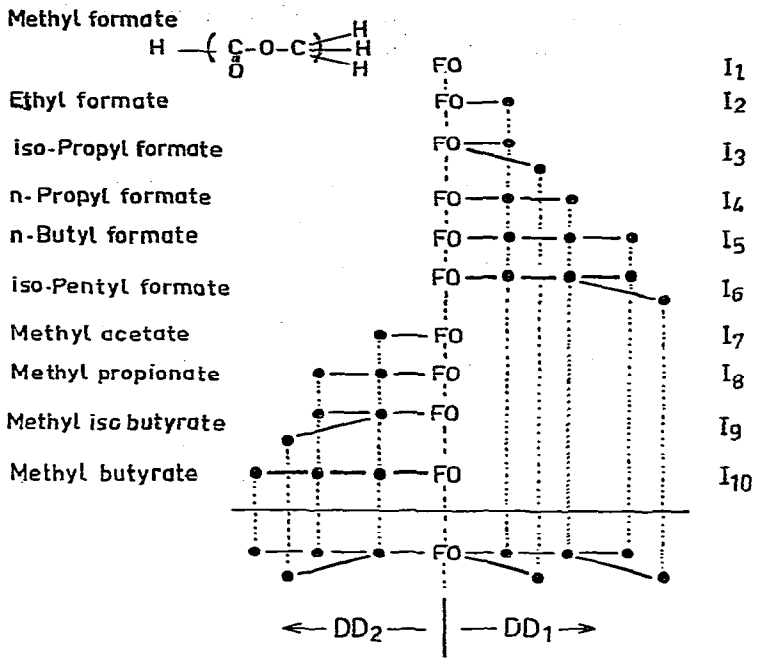

Fig. 1. Superposition of elementary ester graphs giving an imprint characteristic of the compound studied. Data $I_1, \ldots, I_{10}$ are used to calculate average perturbation terms associated with the sites in the topology–information correlation.

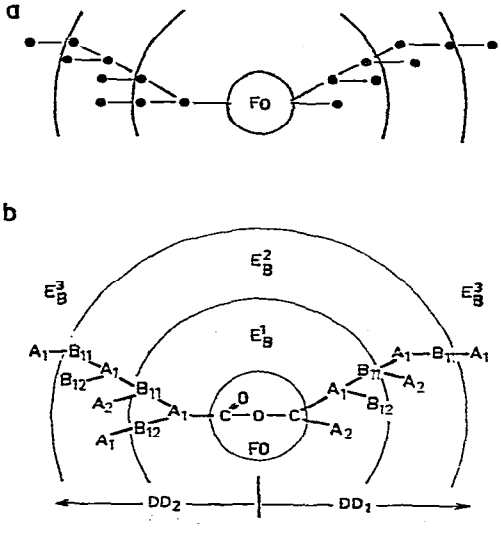

Fig. 2. (a) Imprint of the population of 99 esters having 17 topological sites giving rise to an 18-parameter correlation. (b) Organization of molecular environment by the ELCO concept.

the information for an experimental population made up of m compounds. This vector is defined by the basic topology–information relationship:

$$I(\mathscr{E}) = < \vec{T}(\mathscr{E}) \mid \vec{I}(m) >$$

where $I(\mathscr{E})$ is the contribution from the environment and $\vec{T}(\mathscr{E})$ is the topological vector of the environment. $\vec{T}(\mathscr{E}) = (x_1, \ldots, x_n)$; thus $x_n = 1$ when the nth site of the environment is occupied, and $x_n = 0$ when it is not. The PELCO method consists of calculating the $\vec{I}(m)$ vector and in defining thereby a topology–information correlation which can be presented as a topology–information diagram such as Figs. 4–6. Actually, the perturbation term of a site is an average value, and is optimized by using a multiple regression program that takes into account all the members of a given population containing this site. Information $I(X)$ concerning compound X can be calculated from information $I(X_0)$ concerning reference compound X_0 (methyl formate in the present case) and can be expressed by $I(X) = I(X_0) + I(\mathscr{E})$.

RESULTS AND DISCUSSION

By using various factorial analysis methods, Rohrschneider's[10] and McReynolds'[11] sets of data have led to fundamental work on the characterization and selection of stationary phases[12,13]. In order to characterize better the diversity of behaviour of the stationary phases, Rohrschneider and McReynolds sought the widest possible diversity of solutes on the level of the chemical function as well as on that of the structure. The absence of structural homology between most of these compounds explains why attempts to characterize solutes have attracted the attention of chromatographers to a lesser extent[14] because the establishment of structure–retention correlations was impossible.

All that is required in order to verify the structural homology between the compounds in any set is to ascertain whether every compound is derived from the reference compound by progressive substitution according to the principles of the DARC topological system. Thus, the saturated esters that we have chosen to study are structurally homologous. However, the series of data corresponding to a homogeneous set of compounds can themselves be more or less precise, depending on their nature.

Fig. 3 shows the four types of chromatographic data that can be subjected to topological analysis. These are:

(1) *Raw retention data*: Kováts indices (I_A, ..., I_n, ..., I_p) measured on stationary phases with various polarities, and relating to a single set of compounds (S_f); Kováts indices (I_n, $I_{n,t}$) measured on a single stationary phase, but relative to two isotopologous sets of compounds (S_f and $S_{f,t}$).

(2) *Derived retention data*: Kováts index increments, $\Delta I = I_p - I_A$, or isotopology factors, $\tau I = I_{n,t} - I_n$.

(3) *Unstandardized derived retention data*: $\Delta I'$ or $\tau I'$ calculated from two sets of data determined under different chromatographic conditions (temperature, reference series, ...).

(4) *Solute factors* resulting from previous treatment by factorial analysis.

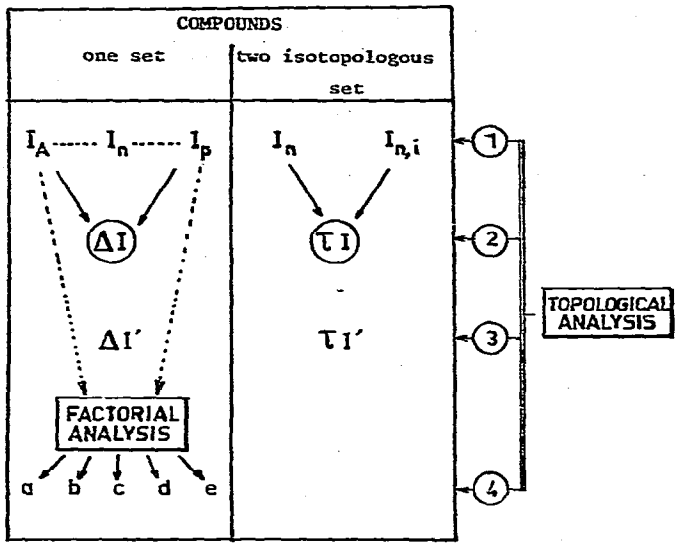

Fig. 3. Four types of topological analysis of chromatographic data: (1) raw chromatographic data, Kováts indices; (2) derived data, Kováts index increment, ΔI, or isotopology factor, τI; (3) unstandardized derived data, $\Delta I'$ or τI; (4) solute factors or substance polarity factors established with Rohrschneider's equation $\Delta I = ax + by + cz + du + es$, where $x, y, \ldots, s$ are the stationary phase polarity factors.

Levels of data treatment and statistical correlation tests

Of the aliphatic esters analyzed chromatographically by Ashes and Haken[4,5], we chose the 99 saturated esters shown in Table I and their associated data. Of the four types of data mentioned above, we have used some of those in categories 1, 2 and 3, namely Kováts indices obtained with SE-30 as the stationary phase (weakly polar) and with Silar 5 cp as the stationary phase (strongly polar), the corresponding Kováts index increments ($\Delta I = I_{\text{Silar 5 cp}} - I_{\text{SE-30}}$), and solute factors a and c determined by Ashes and Haken[5] according to the principles defined by Rohrschneider and corresponding respectively to benzene and 2-pentanone as test compound.

The results of the statistical tests for the 18-parameter (17 topological sites, plus the focus) topology–information correlation used for each of the five series of data on the 99 esters are shown in Table II. The quality of these tests decreases from left to right, *i.e.*, starting from the analysis using the Kováts indices corresponding to the SE-30 stationary phase, up to the analysis using solute factor c. The correlation coefficient, R, is excellent for indices obtained on SE-30 ($R = 0.9991$), very good for those obtained on Silar 5 cp ($R = 0.9985$), average for the ΔI increments ($R = 0.984$), poor for solute factor a ($R = 0.948$) and very poor for solute factor c ($R = 0.798$). The root mean square of the deviations and the standard deviation, between experimental and calculated values, reflects the quality of the correlations. However, the standard deviation takes into account the number of parameters used in these correlations.

SE-30, which is a methylsilicone, is weakly polar and, according to the sum of index increments for five reference compounds, is characterized by $\sum_{1}^{5} = 217$ (ref.

TABLE I

TOPOLOGICAL STUDY OF THREE TYPES OF CHROMATOGRAPHIC DATA FOR A SET OF 99 ESTERS[4,5]

Parameters of the corresponding topology–information correlations are given in Figs. 4 and 5; values in parentheses indicate compounds that have not been used to establish the correlation.

Types of chromatographic data	*Kováts index on SE-30*	*Kováts index on Silar 5 cp*	*Index increments:* $\Delta I = I_{SE\text{-}30} - I_{Silar}$	*Solute factor a*
Correlation				
Number of compounds used	99	92	97	90
Number of parameters	18	18	18	19
Correlation coefficient, R	0.9991	0.9990	0.9947	0.981
Standard deviation, σ	0.907	0.863	3.73	6.28
Root mean square of the deviation	0.82	0.77	3.37	5.58
F-test	2710	2258	432	102

Comparison between experimental and calculated values

No.	*Ester*	*Exp. (a)*	*Calc. (b)*	*Difference (a) − (b)*	*Exp. (c)*	*Calc. (d)*	*Difference (c) − (d)*	*Exp. (e)*	*Calc. (f)*	*Difference (e) − (f)*	*Exp. (g)*	*Calc. (h)*	*Difference (g) − (h)*
1	Methyl formate	386	434.7	−48.7	762	805.6	−43.6	376	376.5	− 0.5	47.88	47.88	0.0
2	Ethyl formate	495	510.1	−15.1	860	863.1	− 3.1	365	357.7	7.3	38.99	44.15	− 5.16
3	Propyl formate	602	606.1	− 4.1	955	955.0	0.0	353	353.8	− 0.8	32.95	37.74	− 4.79
4	Butyl formate	707	699.9	7.1	1059	1048.0	11.0	352	352.8	− 0.8	(8.98)	(22.81)	(−13.82)
5	Pentyl formate	810	792.9	17.1	1156	1139.3	16.7	346	351.1	− 5.1	− 1.23	5.45	− 6.68
6	Hexyl formate	907	885.8	21.2	(1252)	(1225.9)	(26.1)	345	347.2	− 2.2	−13.41	−11.46	− 1.95
7	Isopropyl formate	552	551.6	0.4	869	868.5	0.5	317	321.4	− 4.4	(81.87)	(44.95)	(36.92)
8	Isobutyl formate	670	663.8	6.2	(1015)	(993.6)	(21.4)	345	336.5	8.5	41.31	22.82	18.48
9	Isopentyl formate	777	761.1	15.9	1113	1094.4	18.6	336	338.0	− 2.0	9.33	9.24	0.09
10	Methyl acetate	509	525.2	−16.2	(850)	(874.6)	(−24.6)	341	346.6	− 5.6	(95.49)	(73.56)	(21.93)
11	Ethyl acetate	592	600.7	− 8.7	923	932.0	9.0	331	327.7	3.3	58.45	51.18	7.27
12	Propyl acetate	695	696.7	− 1.7	1013	1024.0	−11.0	318	323.8	− 5.8	46.47	44.77	1.70
13	Butyl acetate	794	790.5	3.5	1116	1116.9	− 0.9	322	322.8	− 0.8	33.68	29.84	3.84
14	Pentyl acetate	891	883.5	7.5	1212	1208.3	3.7	321	321.2	− 0.2	14.22	12.48	1.73
15	Hexyl acetate	988	976.3	11.7	1303	1294.8	8.2	315	317.3	− 2.3	− 4.27	− 4.43	0.16
16	Isopropyl acetate	643	642.1	0.9	(914)	(937.4)	(−23.4)	(271)	(291.4)	(−20.4)	56.66	51.98	4.68
17	Isobutyl acetate	750	754.3	− 4.3	1063	1062.5	0.5	313	306.6	6.4	19.72	29.85	−10.13
18	Isopentyl acetate	859	851.7	7.3	1172	1163.3	8.7	313	308.0	5.0	7.03	16.27	9.24
19	Methyl propionate	617	621.9	− 4.9	947	953.5	− 6.5	330	332.0	− 2.0	(91.35)	(70.42)	(20.93)
20	Ethyl propionate	692	697.3	− 5.3	998	1010.9	−12.9	306	313.2	− 7.2	(76.78	(48.03)	(28.75)
21	Propyl propionate	789	793.3	− 4.3	1098	1102.9	− 4.9	309	309.3	− 0.3	47.56	41.63	5.93
22	Butyl propionate	886	887.1	− 1.1	1195	1195.9	− 0.9	309	308.3	0.7	31.29	26.69	4.60

23	Pentyl propionate	980	980.1	− 0.1	1289	1287.2	1.8	309	306.6	2.4	5.10	9.34	− 4.24
24	Hexyl propionate	1074	1073.0	1.0	1378	1373.8	4.2	304	302.7	1.3	−13.37	− 7.57	− 5.80
25	Isopropyl propionate	733	738.8	− 5.8	1001	1016.3	−15.3	268	276.8	− 8.8	51.48	48.84	2.64
26	Isobutyl propionate	848	851.0	− 3.0	1141	1141.5	− 0.5	293	292.0	1.0	33.28	26.71	6.57
27	Isopentyl propionate	948	948.3	− 0.3	1243	1242.2	0.8	295	293.4	1.6	17.75	13.13	4.62
28	Methyl butyrate	702	703.3	− 1.3	1032	1029.1	2.9	330	326.1	3.9	54.11	48.71	5.40
29	Ethyl butyrate	778	778.8	− 0.8	1085	1086.6	− 1.6	307	307.2	− 0.2	(45.41)	(26.33)	(19.08)
30	Propyl butyrate	875	874.8	0.2	1179	1178.5	0.5	304	303.4	0.6	22.00	19.92	2.08
31	Butyl butyrate	969	968.6	0.4	1272	1271.5	0.5	303	302.3	0.7	4.35	4.99	− 0.64
32	Pentyl butyrate	1062	1061.6	0.4	1362	1362.8	− 0.8	300	300.7	− 0.7	−12.93	−12.37	− 0.56
33	Hexyl butyrate	1156	1154.4	1.6	1451	1449.4	1.6	295	296.8	− 1.8	−29.11	−29.28	0.17
34	Isopropyl butyrate	820	820.2	− 0.2	1091	1092.0	− 1.0	271	270.9	0.8	26.95	27.13	− 0.18
35	Isobutyl butyrate	933	932.4	0.6	1217	1217.1	− 0.1	284	286.1	− 2.1	2.34	5.00	− 2.66
36	Isopentyl butyrate	1029	1029.8	− 0.8	1316	1317.8	− 1.8	287	287.5	− 0.5	−12.19	− 8.58	− 3.61
37	Methyl pentanoate	807	797.3	9.7	1131	1115.7	15.3	324	318.9	5.1	45.32	41.82	3.51
38	Ethyl pentanoate	876	872.7	3.3	1180	1173.2	6.8	304	300.1	3.9	22.64	19.43	3.21
39	Propyl pentanoate	971	968.7	2.3	(1221)	(1265.1)	(−44.1)	(250)	(296.2)	(−46.2)	13.32	13.03	0.29
40	Butyl pentanoate	1063	1062.5	0.5	1361	1358.1	2.9	298	295.2	2.8	− 6.37	− 1.91	− 4.46
41	Pentyl pentanoate	1155	1155.5	− 0.5	1451	1449.5	1.6	296	293.6	2.4	−22.65	−19.26	− 3.39
42	Hexyl pentanoate	1247	1248.3	− 1.3	1539	1536.0	3.0	292	289.7	2.3	−47.51	−36.17	−11.34
43	Isopropyl pentanoate	915	914.2	0.8	1180	1178.6	1.4	265	263.8	1.2	12.89	20.24	− 7.35
44	Isobutyl pentanoate	1027	1026.3	0.7	1307	1303.7	3.3	280	278.9	1.1	− 8.08	−1.89	− 6.19
45	Isopentyl pentanoate	1132	1123.7	8.3	1405	1404.5	0.5	273	280.4	− 7.4	− 4.07	−15.47	11.41
46	Methyl hexanoate	902	891.8	10.2	(1225)	(1209.6)	(15.4)	323	319.8	3.2	14.08	19.91	− 5.83
47	Ethyl hexanoate	976	967.2	8.8	1275	1267.1	7.9	299	301.0	− 2.0	− 0.11	− 2.47	2.36
48	Propyl hexanoate	1064	1063.2	0.8	1363	1359	4.0	299	297.1	1.9	−13.02	− 8.88	− 4.14
49	Butyl hexanoate	1156	1157.1	− 1.1	1451	1452.0	− 1.0	295	296.1	− 1.1	−24.83	−23.81	− 1.02
50	Pentyl hexanoate	1246	1250.0	− 4.0	1539	1543.3	− 4.3	293	294.5	− 1.5	−44.28	−41.16	− 3.12
51	Hexyl hexanoate	1337	1342.9	− 5.9	1626	1630.1	− 3.9	289	290.6	− 1.6	−61.65	−58.08	− 3.57
52	Isopropyl hexanoate	1008	1008.7	− 0.7	1275	1272.5	2.5	267	264.7	2.3	−10.27	− 1.67	− 8.60
53	Isobutyl hexanoate	1119	1120.9	− 1.9	1397	1397.6	− 0.6	278	279.8	− 1.8	−11.68	−23.80	12.12
54	Isopentyl hexanoate	1212	1218.2	− 6.2	1494	1498.4	− 4.4	282	281.3	0.7	−25.58	−37.38	11.80
55	Methyl isobutyrate	665	662.3	2.7	957	954.7	2.3	292	292.7	− 0.7	46.65	49.88	− 3.23
56	Ethyl isobutyrate	732	737.7	− 5.7	1016	1012.2	3.8	284	273.9	10.1	(−18.58)	(27.50)	(−46.08)
57	Propyl isobutyrate	836	833.7	2.3	1107	1104.1	2.9	271	270.0	1.0	17.54	21.09	− 3.55
58	Butyl isobutyrate	931	927.5	3.5	1201	1197.1	3.9	270	269.0	1.0	7.14	6.16	0.98
59	Pentyl isobutyrate	1024	1020.5	3.5	1294	1288.4	5.6	270	267.3	2.7	− 1.95	−11.20	9.25
60	Hexyl isobutyrate	1117	1113.3	3.7	1383	1375.0	8.0	266	263.4	2.6	−22.77	−28.11	5.34
61	Isopropyl isobutyrate	780	779.2	0.8	1014	1017.6	− 3.6	234	237.6	− 3.6	21.61	28.30	− 6.70
62	Isobutyl isobutyrate	899	891.3	7.7	1149	1142.7	6.3	250	252.7	− 2.7	− 3.18	6.17	− 9.35

(continued on p. 50)

TABLE I *(continued)*

No.	Ester	Exp. (a)	Calc. (b)	Difference (a) − (b)	Exp. (c)	Calc. (d)	Difference (c) − (d)	Exp. (e)	Calc. (f)	Difference (e) − (f)	Exp. (g)	Calc. (h)	Difference (g) − (h)
63	Isopentyl isobutyrate	994	988.7	5.3	1249	1243.4	5.6	255	254.2	0.8	−14.48	− 7.41	− 7.07
64	Methyl isopentanoate	763	757.1	5.9	1063	1056.9	6.1	300	300.1	− 0.1	35.59	40.01	− 4.42
65	Ethyl isopentanoate	839	832.6	6.4	1114	1114.3	− 0.3	275	281.2	− 6.2	24.54	17.63	6.91
66	Propyl isopentanoate	929	928.6	0.4	1206	1206.3	− 0.3	277	277.4	− 0.4	8.10	11.23	− 3.13
67	Butyl isopentanoate	1021	1022.4	− 1.4	1296	1299.2	− 3.2	275	276.3	− 1.3	− 6.19	− 3.71	− 2.48
68	Pentyl isopentanoate	1112	1115.4	− 3.4	1386	1390.6	− 4.6	274	274.7	− 0.7	−21.99	−21.06	− 0.93
69	Hexyl isopentanoate	1204	1208.2	− 4.2	1474	1477.1	− 3.1	270	270.8	− 0.8	−37.43	−37.98	0.55
70	Isopropyl isopentanoate	874	874.0	0.0	1127	1119.7	7.3	253	244.9	8.1	20.73	18.43	2.30
71	Isobutyl isopentanoate	985	986.2	− 1.2	1246	1244.8	1.2	261	260.1	0.9	− 5.18	− 3.70	− 1.48
72	Isopentyl isopentanoate	1081	1083.6	− 2.6	1343	1345.6	− 2.6	262	261.5	0.5	−14.61	−17.28	2.67
73	Methyl isohexanoate	875	861.6	13.4	(1177)	(1158.8)	(18.3)	302	299.5	2.5	24.23	19.45	4.8
74	Ethyl isohexanoate	943	937.0	6.0	1219	1216.2	2.8	276	280.7	− 4.7	− 7.89	− 2.93	− 4.9
75	Propyl isohexanoate	1035	1033.0	2.0	1313	1308.1	4.9	278	276.8	1.2	−10.66	− 9.3	− 1.3
76	Butyl isohexanoate	1125	1126.8	− 1.8	1402	1401.1	0.9	277	275.8	1.2	(24.86)	(−24.97)	(49.13)
77	Pentyl isohexanoate	1215	1219.8	− 4.8	1490	1492.5	− 2.5	275	274.1	0.8	−39.91	−41.62	1.71
78	Hexyl isohexanoate	1306	1312.7	− 6.7	1578	1579.0	− 1.0	272	270.2	1.7	−54.49	−58.53	4.04
79	Isopropyl isohexanoate	979	978.5	0.5	1229	1221.6	7.4	250	244.4	5.6	7.94	− 2.12	10.06
80	Isobutyl isohexanoate	1089	1090.7	− 1.7	1340	1346.7	− 6.7	251	259.5	− 8.5	−34.97	−24.25	−10.72
81	Isopentyl isohexanoate	1181	1188.0	− 7.0	1442	1447.5	− 5.5	261	261.0	0.0	−41.42	−37.84	− 3.58
82	Methyl 2-methylpentanoate	853	837.6	15.4	1131	1117.0	14.0	278	279.6	− 1.6	21.94	21.28	0.66
83	Ethyl 2-methylpentanoate	917	913.1	3.9	1173	1174.4	− 1.4	256	260.8	− 4.8	− 1.06	− 1.10	0.04
84	Propyl 2-methylpentanoate	1009	1009.1	− 0.1	1266	1266.3	− 0.3	257	257.0	0.0	− 4.73	− 7.51	2.78
85	Butyl 2-methylpentanoate	1097	1102.9	− 5.9	1352	1359.3	− 7.3	255	255.9	− 0.9	−23.7	−22.45	− 1.23
86	Pentyl 2-methylpentanoate	1187	1195.9	− 8.9	1440	1450.7	−10.7	253	254.3	− 1.3	−36.32	−39.80	3.48
87	Hexyl 2-methylpentanoate	1277	1288.7	−11.7	1527	1537.2	−10.2	250	250.4	− 0.4	−52.11	−56.71	4.60
88	Isopropyl 2-methylpentanoate	952	954.6	− 2.6	1176	1179.8	− 3.8	224	224.5	− 0.5	2.84	− 0.30	3.14
89	Isobutyl 2-methylpentanoate	1064	1066.7	− 2.7	1302	1304.9	− 2.9	238	239.7	− 1.7	−17.37	−22.43	5.06
90	Isopentyl 2-methylpentanoate	1153	1164.1	−11.1	1394	1405.7	−11.7	241	241.1	− 0.1	−40.22	−36.01	− 4.21
91	Methyl 2-ethylbutyrate	845	831.3	13.7	1116	1106.2	9.8	271	275.2	− 4.2	16.98	17.85	− 0.87
92	Ethyl 2-ethylbutyrate	914	906.8	7.2	1171	1163.7	7.3	257	256.4	0.6	−14.21	− 4.54	− 9.67
93	Propyl 2-ethylbutyrate	1005	1002.8	2.2	1260	1255.6	4.4	255	252.5	2.5	− 6.76	−10.94	4.18
94	Butyl 2-ethylbutyrate	1093	1096.6	− 3.6	1343	1348.6	− 5.6	250	251.5	− 1.5	−25.48	−25.88	0.40
95	Pentyl 2-ethylbutyrate	1183	1189.6	− 6.6	1434	1440.0	− 6.0	251	249.8	1.2	−40.47	−43.33	2.76
96	Hexyl 2-ethylbutyrate	1273	1282.4	− 9.4	1520	1526.5	− 6.5	247	245.9	1.1	−52.34	−60.14	7.80
97	Isopropyl 2-ethylbutyrate	954	948.2	5.8	1174	1169.1	4.9	220	220.0	0.0	(21.34)	(− 3.73)	(25.07)
98	Isobutyl 2-ethylbutyrate	1060	1060.4	− 0.4	1294	1294.2	− 0.2	234	235.2	− 1.2	−27.58	−25.86	− 1.72
99	Isopentyl 2-ethylbutyrate	1149	1157.8	− 8.8	1387	1394.9	− 7.9	238	236.6	1.4	−42.32	−39.44	− 2.88

15). It is with this stationary phase, used by Ashes and Haken[4] as a reference, that the behaviour of the solutes under study is most homogeneous. In contrast to this stationary phase, Silar 5 cp, which is a phenylcyanopropylsilicone, is strongly polar and is characterized by $\sum_{1}^{5} = 2428$.

For experimental reasons related either to adsorption problems on the gas–liquid interface or to the choice of the *n*-alkane reference series, it is well known that Kováts indices are determined with lower precision with relatively polar stationary phases than with those whose polarity is clearly weaker. Further, using SE-30 first and then going to the Silar 5 cp stationary phase increases the contribution of the non-ideal solute behaviour. These two reasons explain the slightly weaker coherence in behaviour of the esters studied with Silar 5 cp than with SE-30.

The precision of the derived ΔI data is simultaneously related to those of the data for I_p and I_A. The statistical tests remain fairly satisfactory, and show that topological analysis of these derived data is possible.

For the correlations with solute factors *a* and *c*, successively corresponding to benzene and 2-pentanone as test compounds, the quality of the statistical tests

TABLE II

STATISTICAL TEST OF THREE TYPES OF CHROMATOGRAPHIC DATA CORRELATIONS

Raw data, Kováts indices; derived data, ΔI; and solute factors arising from previous factorial analysis. The correlations are established for the same parameters and compounds.

Parameter	*Kováts indices*		*Index increments:* $\Delta I = I_{Silar} - I_{SE\text{-}30}$	*Solute factors*	
	SE-30	*Silar 5 cp*		*a*	*c*
Number of compounds used in the correlation	99	99	99	99	99
Number of parameters	18	18	18	18	18
Correlation coefficient, *R*	0.9991	0.9985	0.984	0.948	0.798
Standard deviation, σ	0.907	1.07	6.33	11.6	6.75
Root mean square of standard deviation	0.821	0.966	5.72	10.5	6.11
F-test	2710	1625	149	42.0	8.35

decreases noticeably, and is most striking for solute factor *c* (Table II). By subjecting the set of Kováts indices for saturated esters determined on 14 stationary phases to Benzecri's "analyse factorielle des correspondances"[16], in the same way that Chastrette[17] applied it to Rohrschneider's and McReynolds' sets of data, it can be seen that the test compounds, benzene, 1-butanol, 2-pentanone, nitropropane and pyridine, are not in the cluster of esters. None of these test compounds particularly reflects the ester character. This results from the fact that these test compounds have been chosen to reflect the diversity of different types of interactions encountered in every possible kind of functional group.

The determination of these solute factors by factorial analysis is conventionally

based on two considerations: the choice of test compounds, and the value arbitrarily set for them. It is therefore normal for fairly strong distortions in the various series of solute factors to be introduced when these are calculated using factorial analysis. These distortions strongly weaken the behavioural coherence found by topological analysis.

The results in Table I are relative to the correlations carried out using the indices obtained with SE-30 and Silar 5 cp, the ΔI increments and solute factors a. As indicated by the statistical tests shown in Table I, the results are better than the corresponding results in Table II. Eighteen parameters were used for the correlations carried out with the indices obtained with Silar 5 cp and for those carried out with the ΔI increments, but in the first instance, the number of data correlated was reduced to 92 and in the second to 97. For example, both of these correlations were established without using the value for propyl pentanoate (compound No. 39), as the retention index proposed for this compound (1221) with Silar 5 cp has probably been erroneously miscopied. Linear interpolation yields a retention index of 1268 for this compound on Silar 5, thereby indicating a corresponding ΔI of 297; these two values are in excellent agreement with those calculated from the correlation, *i.e.*, 1265.1 and 296.2, respectively. The other six compounds that were eliminated for establishment of the correlation of the Silar 5 retention indices correspond to a behaviour which, statistically, is not good, mainly because of the distortions introduced by the first terms of the homologous series, notably by the formates.

In order to take into account the distortions introduced by the first terms of the homologous series, when these distortions are coherent between the various sites, additional parameters of interaction between unrelated sites can be introduced. This explains the nineteenth parameter in the correlation of the solute factors.

Specific interactions between esters and the Silar 5 cp polar stationary phase

A thermodynamic study of Kováts indices[18–20] and other derived data makes it possible to attach a thermodynamic meaning to the perturbation terms associated with the topological sites in the topology–information correlations[2]. In general, the Kováts index obtained on a polar phase, I_p, can be expressed as a function of the vapour pressures, $P^0_{x,r}$, and of the activity coefficients, $\gamma^\infty_{x,r}$, of solute x and the reference alkanes r. Formally, this is expressed by $I_p = f(P^0_{x,r}, \gamma^\infty_{x,r})$.

The difference in behaviour between the esters studied on Silar 5 cp and those studied on SE-30 is expressed by the retention index increments, and gives the specific solute–stationary phase interactions on Silar 5 cp. Thus, $\Delta I = I'_p - I_A \approx f(\gamma^\infty_{x,r})$.

Topological analysis of the behaviour of esters on Silar 5 cp (Fig. 4a) can be extended by dissociating, at the level of each topological site, the relatively ideal behaviour of the solutes (given by their reference behaviour on SE-30) from the contribution of the specific solute–stationary phase interactions (given by ΔI) (Fig. 4b). An examination of these contributions follows below.

The behaviour of the esters studied on Silar 5 cp is shown in the topology–information diagram in Fig. 4a. These diagrams have a clear meaning for chromatographers[1]. If the choice of focus is known to correspond here to the pattern (COOC), then pentyl acetate has an ordered graph (A_1–FO–A_1–B_{11}–A_1–B_1) with a corresponding calculated index, *i.e.*, $I_{calc} = 68.9 + 805.6 + 57.5 + 91.9 + 93.0 + 91.4 = 1208.3$, and an experimental value of 1212 Kováts indices (K.I.). Likewise,

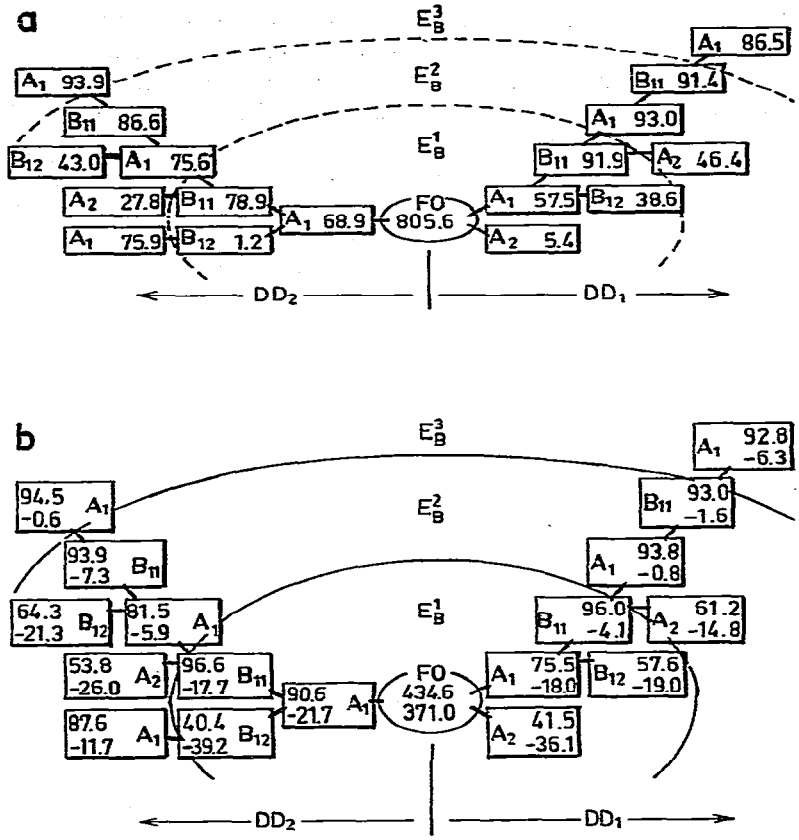

Fig. 4. Specific interactions between esters and the polar Silar 5 cp stationary phase determined by topological analysis of Kováts indices. (a) Behavior of saturated aliphatic esters on *Silar 5 cp* for every topological site corresponding here to every skeletal carbon atom. (b) Dissociation of the ideal behaviour contribution (SE-30, values in upper part of boxes) from specific interactions with Silar 5 cp (values in lower part of boxes).

the ordered graph for ethyl isobutyrate, $\begin{matrix} & B_{11} & \\ & | & \\ A_1 & -FO- & A_1 \\ & | & \\ & B_{12} & \end{matrix}$, has a corresponding calculated index, *i.e.*, $I_{calc} = 78.9 + 1.2 + 68.9 + 805.6 + 57.5 = 1012.1$, and an experimental value of 1016 K.I. The strong contribution of the focus, 805.6 K.I., should be noted, as should the influence of the chain lengthening of the *n*-alkanols for which, in agreement with Ashes and Haken[4], the weight of the CH_2 groups does not tend towards 100 K.I.; on the contrary, it decreases for those sites situated in the second and third environments, E_B^2 and E_B^3. The chain lengthening of the *n*-acids indicated in the second development direction, DD_2, is not symmetrical to that of the *n*-alkanols. In the first development direction, DD_1, site A_2 of E_B^1 corresponding to the substitution of the focus has a very weak contribution, *i.e.*, 5.4 K.I.; this indicates that the index of the isopropyl esters is very close to that of the ethyl esters. In con-

trast, the branching of the alkanol chains moves away from the focus, and the contribution of the corresponding B_{12} and A_2 sites increases to 38.6 and 46.4, respectively.

Similar behaviour is observed for the branchings of the acid chain in DD_2. The formal contribution from the isobutyrates of site B_{12} in E_B^1 is very weak (1.2 K.I.), and indicates that the isobutyrate indices are very near those of the propionates; indeed, the deviation between the comparable elements of these two homologous series is less than 1%. Sites A_2 and B_{12} in E_B^1 express the branchings introduced by the isopentanoates and the isohexanoates, respectively. Their contribution increases as their distance from the focus becomes greater, *i.e.*, 27.8 and 43.0 K.I.

The interpretation of the observed behaviour of the esters studied on Silar 5 cp can be developed by consulting Fig. 4b. The values in the upper part of boxes representing the sites in Fig. 4b express the relatively ideal behaviour of these esters studied on SE-30, used as the reference behaviour; the values are mainly a function of the solute vapour pressures. The contribution of the focus is 434.6 K.I. The presence of site A_1 related to the focus in DD_1 introduces a perturbation of only 75.5 K.I.; this perturbation is related to the non-linearity of the variation in retention index in the homologous series. In contrast, site A_1, related to the focus in DD_2, has a higher value (90.6 K.I.). The branchings expressed by sites A_2, B_{12} and A_2 in DD_1 increase in value (41.5, 57.6 and 61.2) as their distance from the focus becomes greater. Similar variations are observed for sites B_{12}, A_2 and B_2 in DD_2, which correspond to branchings in the *n*-acid chain.

The difference in the behaviour between the esters studied on Silar 5 cp and those studied on SE-30 give the additional specific solute–stationary phase interactions introduced by the change from the weakly polar SE-30 to the strongly polar Silar 5 cp. These specific interactions are indicated by the values in the lower part of the boxes representing the sites in Fig. 4b, which are virtually identical with those obtained by the direct topological analysis of ΔI (*cf.*, results in Table I).

It should be noted that the strong contribution of the focus (371 K.I.) is positive, whereas all the other contributions of the topological sites are negative. This indicates that the strong interaction of the esters studied on Silar 5 cp is limited to the ester functional group. All of the other sites contribute more or less to the weakening of this specific interaction. This weakening is fairly strong for most of the sites of the first environment, being *ca.* —20 K.I. on average. This weakening decreases fairly regularly when the *n*-acid or *n*-alkanol chain is lengthened, and results in an increase in the *n*-alkyl character of the esters whose behaviour is similar to that of the reference *n*-alkanes. Sites A_2 and B_{12}, which are the ones nearest the focus in DD_1 and DD_2 of environment E_B^1, correspond to branchings of the alkanol and acid chains that most weaken the contribution of the focus: —36.1 and —39.2 K.I., respectively. This might be explained by steric hindrances on the solutes which interact with the stationary phases. It should be recalled that, with alkenes analysed on β,β'-oxydipropionitrile, tri- or a tetra-substitution at the ethylenic focus results in an electronic effect that causes a positive contribution of the activity coefficient on the corresponding topological sites related to the focus[1].

Topological analysis of solute factors a and c

The parameters of the topology–information correlation with solute factors *a* are shown in Fig. 5. The corresponding statistical tests and the re-calculated values

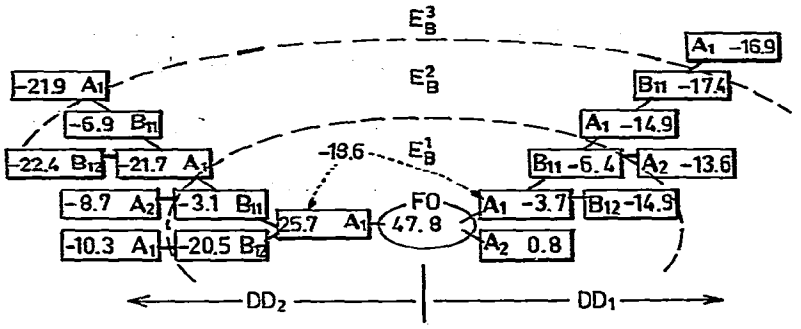

Fig. 5. Topological analysis of solute factors a for esters: comparison of esters and benzene (test compound).

are given in Table I. Solute factor a corresponds to the benzene test compound. Fig. 6 shows the treatment of solute factors c, established using 2-pentanone as the test compound.

Comparison of Figs. 5 and 6 shows that, in the topological analysis of solute factors c, the value assigned to the focus is stronger, and that the contribution of the sites corresponding to a chain lengthening has a negative value that very rapidly nears zero. In contrast, in the topological analysis of solute factors a, the sites corresponding to chain lengthenings have a strong and constant negative contribution. This indicates that the similarity in behaviour between the benzene test compound and the esters remains localized at the level of the focus of the esters, and that the similarity of the esters is greater with 2-pentanone than with benzene. The interest of this topological analysis thus lies in specifying the quantitative contribution of the different structural parameters.

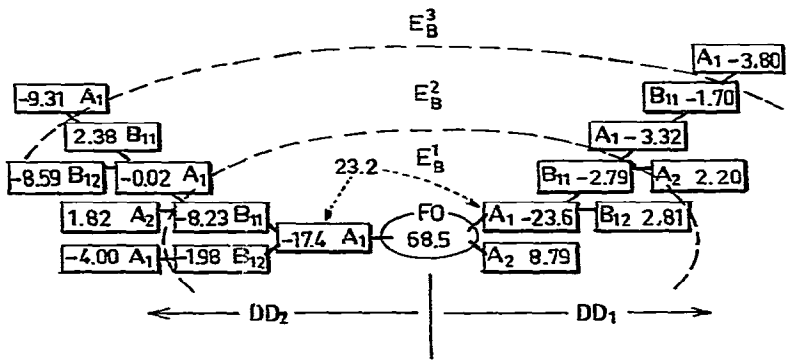

Fig. 6. Topological analysis of solute factors, c for esters: comparison of esters and 2-pentanone (test compound).

CONCLUSION

Topological analysis can be used to treat different types of chromatographic data, such as raw data, derived ΔI and τI data and solute factors. Although we have not dealt here with any treatment of τI, we have used it previously[3] for prediction

purposes, and know that it can be used fruitfully in physico-chemical applications. The possibility of subjecting solute factors resulting from a preliminary factorial analysis to topological analysis shows that both methods are complementary. Currently, however, the main interest in using topological analysis to treat chromatographic data resides in the possibility of calculating specific solute–stationary phase interactions, at the level of topological sites, studied by gas–liquid, gas–solid or liquid chromatography, in order to investigate, for example, electrochemical, catalytic or biological problems.

REFERENCES

1 J. R. Chrétien and J.-E. Dubois, *Anal. Chem.*, 49 (1977) 756.
2 J. R. Chrétien, *C.R. Acad. Sci., Ser. C*, 281 (1975) 151.
3 J. R. Chrétien and J.-E. Dubois, *J. Chromatogr.*, 126 (1976) 171.
4 J. R. Ashes and J. K. Haken, *J. Chromatogr.*, 101 (1974) 103.
5 J. R. Ashes and J. K. Haken, *J. Chromatogr.*, 118 (1976) 183.
6 P. Souter, *J. Chromatogr. Sci.*, 12 (1974) 418 and 424.
7 J.-E. Dubois in W. T. Wipke, S. Heller, R. Feldmann and E. Hyde (Editors), *Computer Representation and Manipulation of Chemical Information*, Wiley, New York, 1974, p. 239.
8 J.-E. Dubois, D. Laurent and A. Aranda, *J. Chim. Phys.*, 70 (1973) 1608 and 1716.
9 J.-E. Dubois and J. R. Chrétien, *J. Chromatogr. Sci.*, 12 (1974) 811.
10 L. Rohrschneider, *J. Chromatogr.*, 22 (1966) 6.
11 W. O. McReynolds, *J. Chromatogr. Sci.*, 8 (1970) 685.
12 A. Eskes, F. Dupuis, A. Dijkstra, H. de Clercq and D. L. Massart, *Anal. Chem.*, 47 (1975) 2168.
13 J. K. Haken, M. S. Wainwright and N. Do Phuong, *J. Chromatogr.*, 117 (1976) 23.
14 P. H. Weiner and J. F. Parcher, *Anal. Chem.*, 45 (1973) 302.
15 Applied Science Labs., State College, Pa., Catalog No. 20 (1977) 19.
16 J. P. Benzecri, *L'Analyse des Données*, Vol. 2, Dunod, Paris, 1973.
17 M. Chastrette, *J. Chromatogr. Sci.*, 14 (1976) 357.
18 D. A. Tourres, *J. Chromatogr.*, 30 (1967) 357.
19 J. Krupcik, O. Liska and L. Sojak, *J. Chromatogr.*, 51 (1970) 119.
20 L. Rohrschneider, *Advan. Chromatogr.*, 4 (1974) 239.