Classification of Mesogenic Benzalazines by Multivariate Data Analysis

JOSE LUIS SERRANO,[a] MERCEDES MARCOS,[a] ENRIQUE MELENDEZ,[a] CHRISTER ALBANO,[b] SVANTE WOLD[b] and JOSE ELGUERO[c]

[a] Departamento de Química Orgánica, Facultad de Ciencias, Universidad de Zaragoza, 50009-Zaragoza, Spain, [b] Research Group for Chemometrics, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden and [c] Instituto de Química Médica, Juan de la Cierva 3, Madrid-6, Spain

A method of pattern recognition, SIMCA, was applied to a collection of 170 mesogenic and non-mesogenic compounds, belonging to a family of benzalazines. Four types of liquid crystals were considered: monotropic nematics, enantiotropic nematics, smectics and nematic-smectics. The method leads to a classification that follows the increase of molecular order. Geometric and electronic variables, which can be calculated from the structural formulae, have been used. The best set of variables for a study of the mesogenic properties of benzalazines turned out to be the one formed by a combination of geometrical variables, dipole moment and Hammett parameters.

At present there exist a sufficiently large number of data about different kinds of substituent constants to make it possible to describe, in an approximate manner, some of the properties of a compound, taking into account the nature and the position of its substituents. Particularly complete is the collection of aromatic substituent constants. Since most of the known liquid crystals have at least one aromatic ring, it was tempting to use the above-mentioned constants to explain the mesogenic properties of an aromatic compound.

If a relationship can be found it would be possible to predict the mesogenic properties of a new compound from its structural formula and a series of tabulated values. However, the thermal and thermodynamic properties depend on a large number of factors, which make "intuitive" classifications almost impossible. On the other hand, the foreseeable existence of interactions between the "independent" variables and the qualitative nature of the response excludes the formulation of an empirical model, simple enough to apply the modern methods of building optimum experimental matrices.[1]

The best approach to a problem of this complexity is to use pattern recognition methods[2] among which SIMCA[3] was selected. The theory and practical utilisation of this method has already been described (Ref. 4 and references cited therein).

EXPERIMENTAL

In this first attempt at the classification of mesogenic compounds by means of multifactorial analysis, a restricted experimental domain was selected, namely benzalazine derivatives

Table 1. Experimental Space; *cf.* Scheme 1.

	R_1	R_1'	R_4	R_4'	Class
$R_2=R_2'=CH_3$ $R_3=R_3'=H$					
1	$O-CH_3$	$O-CH_3$	H	H	B(M)[a]
2	$O-C_2H_5$	$O-C_2H_5$	H	H	B(M)
3	$O-C_3H_7$	$O-C_3H_7$	H	H	B
4	$O-n-C_4H_9$	$O-n-C_4H_9$	H	H	A
5	$O-n-C_5H_{11}$	$O-n-C_5H_{11}$	H	H	A
6	$O-n-C_6H_{13}$	$O-n-C_6H_{13}$	H	H	A
7	$O-n-C_7H_{15}$	$O-n-C_7H_{15}$	H	H	A
8	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	H	A
9	$O-n-C_9H_{19}$	$O-n-C_9H_{19}$	H	H	A
10	$O-n-C_{10}H_{21}$	$O-n-C_{10}H_{21}$	H	H	A
11	$O-n-C_{12}H_{25}$	$O-n-C_{12}H_{25}$	H	H	A
12	$O-n-C_{14}H_{29}$	$O-n-C_{14}H_{29}$	H	H	A
13	$O-n-C_{16}H_{33}$	$O-n-C_{16}H_{33}$	H	H	A
14	$O-n-C_{18}H_{37}$	$O-n-C_{18}H_{37}$	H	H	A
15	H	H	H	H	A
16	H	H	OH	OH	A
17	OH	OH	H	H	A
18	$O-CO-CH_3$	$O-CO-CH_3$	H	H	B(M)
19	$O-CO-C_6H_5$	$O-CO-C_6H_5$	H	H	A
20	H	H	CH_3	CH_3	A
21	CH_3	CH_3	H	H	A
22	OH	OH	CH_3	CH_3	A
23	$O-C_2H_5$	$O-C_2H_5$	CH_3	CH_3	A
24	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	CH_3	CH_3	A
25	H	H	Cl	Cl	A
26	Cl	Cl	H	H	A
27	OH	OH	Cl	Cl	A
28	$O-C_2H_5$	$O-C_2H_5$	Cl	Cl	A
29	$O-n-C_8H_{17}$	Cl	Cl	Cl	A
30	H	H	NO_2	NO_2	A
31	NO_2	NO_2	H	H	A
32	OH	OH	NO_2	NO_2	A
33	$O-C_2H_5$	$O-C_2H_5$	NO_2	NO_2	A
34	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	NO_2	NO_2	A
35	NH_2	NH_2	H	H	A
$R_2=R_2'=CH_3$ $R_3=R_3'=OH$					
36	$O-CH_3$	$O-CH_3$	H	H	C(M)
37	$O-C_2H_5$	$O-C_2H_5$	H	H	C
38	$O-n-C_3H_7$	$O-n-C_3H_7$	H	H	C
39	$O-n-C_4H_9$	$O-n-C_4H_9$	H	H	C(M)
40	$O-n-C_5H_{11}$	$O-n-C_5H_{11}$	H	H	E(M)
41	$O-n-C_6H_{13}$	$O-n-C_6H_{13}$	H	H	E(M)
42	$O-n-C_7H_{15}$	$O-n-C_7H_{15}$	H	H	D(M)
43	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	H	D(M)
44	$O-n-C_9H_{19}$	$O-n-C_9H_{19}$	H	H	D
45	$O-n-C_{10}H_{21}$	$O-n-C_{10}H_{21}$	H	H	D(M)
46	$O-n-C_{12}H_{25}$	$O-n-C_{12}H_{25}$	H	H	D(M)
47	$O-n-C_{14}H_{29}$	$O-n-C_{14}H_{29}$	H	H	D(M)
48	$O-n-C_{16}H_{33}$	$O-n-C_{16}H_{33}$	H	H	D
49	$O-n-C_{18}H_{37}$	$O-n-C_{18}H_{37}$	H	H	D(M)
50	H	H	H	H	A

	R_1	R_1'	R_4	R_4'	Class
51	OH	OH	H	H	A
52	$O-CH_2-CH=CH_2$	$O-CH_2-CH=CH_2$	H	H	C(M)
53	$O-iso-C_3H_7$	$O-iso-C_3H_7$	H	H	A
54	$O-iso-C_4H_9$	$O-iso-C_4H_9$	H	H	A
55	$O-iso-C_5H_{11}$	$O-iso-C_5H_{11}$	H	H	E
56	$O-sec-C_4H_9$	$O-sec-C_4H_9$	H	H	A
57	$O-C_2H_5$	$O-C_2H_5$	H	H	B(M)
58	$O-ter-C_4H_9$	$O-ter-C_4H_9$	H	H	A
59	$O-CH_2-CH_2-OH$	$O-CH_2-CH_2-OH$	H	H	C(M)
60	$O-C_2H_5$	$O-CH_2-CH_2-OH$	H	H	C(M)
61	$O-CH_2-COO-C_2H_5$	$O-CH_2-COO-C_2H_5$	H	H	A
62	$O-C_2H_5$	$O-CH_2-COO-C_2H_5$	H	H	C
63	$O-CH_2-COO-CH_3$	$O-CH_2-COO-CH_3$	H	H	A
64	$O-C_2H_5$	$O-CH_2-COO-CH_3$	H	H	C(M)
65	$O-C_2H_5$	$O-cyclo-C_6H_{11}$	H	H	C(M)
66	$O-CH_2-C_6H_5$	$O-CH_2-C_6H_5$	H	H	A
67	$O-C_2H_5$	$O-CH_2-C_6H_5$	H	H	C(M)
68	$O-CO-CH_3$	$O-CO-CH_3$	H	H	A
69	$O-C_2H_5$	$O-CO-CH_3$	H	H	C(M)
70	$O-CO-C_6H_5$	$O-CO-C_6H_5$	H	H	C(M) [b]
71	$O-C_2H_5$	$O-CO-C_6H_5$	H	H	C[b]
72	$O-C_2H_5$	$O-CH_3$	H	H	C(M) [b]
73	$O-C_2H_5$	$O-n-C_8H_{17}$	H	H	C(M)
74	$O-n-C_3H_7$	$O-n-C_5H_{11}$	H	H	C
75	$O-n-C_5H_{11}$	$O-n-C_7H_{15}$	H	H	E(M)
76	$O-CH_3$	$O-n-C_5H_{11}$	H	H	C
77	$O-C_2H_5$	$O-n-C_4H_9$	H	H	C(M)
78	$O-CH_3$	$O-n-C_8H_{17}$	H	H	C
79	$O-n-C_8H_{17}$	$O-n-C_{10}H_{21}$	H	H	D(M)

$R_2=R_2'=CH_3$ $R_3=OH$ $R_3'=H$

80	$O-CH_3$	$O-CH_3$	H	H	B(M)[c]
81	$O-C_2H_5$	$O-C_2H_5$	H	H	C(M)
82	$O-n-C_3H_7$	$O-n-C_3H_7$	H	H	C
83	$O-n-C_4H_9$	$O-n-C_4H_9$	H	H	C
84	$O-n-C_5H_{11}$	$O-n-C_5H_{11}$	H	H	C(M)
85	$O-n-C_6H_{13}$	$O-n-C_6H_{13}$	H	H	C(M)
86	$O-n-C_7H_{15}$	$O-n-C_7H_{15}$	H	H	E(M)
87	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	H	E(M)
88	$O-n-C_9H_{19}$	$O-n-C_9H_{19}$	H	H	D(M)
89	$O-n-C_{10}H_{21}$	$O-n-C_{10}H_{21}$	H	H	D(M)
90	$O-n-C_{12}H_{25}$	$O-n-C_{12}H_{25}$	H	H	D(M)
91	$O-n-C_{14}H_{29}$	$O-n-C_{14}H_{29}$	H	H	D
92	$O-n-C_{16}H_{33}$	$O-n-C_{16}H_{33}$	H	H	D(M)
93	$O-n-C_{18}H_{37}$	$O-n-C_{18}H_{37}$	H	H	D(M)
94	$O-n-C_5H_{11}$	$O-CH_3$	H	H	C
95	$O-n-C_5H_{11}$	$O-C_2H_5$	H	H	C(M)
96	$O-C_2H_5$	$O-n-C_5H_{11}$	H	H	C
97	OH	H	H	H	A
98	$O-C_2H_5$	H	H	H	A
99	$O-n-C_8H_{17}$	H	H	H	A
100	$O-n-C_4H_9$	OH	H	H	A
101	$O-CH_3$	$O-C_2H_5$	H	H	C(M)
102	$O-CH_3$	$O-n-C_5H_{11}$	H	H	C

	R_1	R_1'	R_4	R_4'	Class
103	$O-n-C_4H_9$	$O-C_2H_5$	H	H	C
104	$O-CH_3$	$O-n-C_8H_{17}$	H	H	C(M)
105	$O-n-C_8H_{17}$	$O-C_2H_5$	H	H	C
106	$O-n-C_5H_{11}$	$O-n-C_7H_{15}$	H	H	E(M)
107	$O-n-C_7H_{15}$	$O-n-C_9H_{19}$	H	H	E
108	$O-n-C_{10}H_{21}$	$O-n-C_8H_{17}$	H	H	E(M)
109	OH	CH_3	H	H	A
110	$O-C_2H_5$	CH_3	H	H	C
111	$O-n-C_8H_{17}$	CH_3	H	H	C(M)
112	OH	H	H	CH_3	A
113	$O-C_2H_5$	H	H	CH_3	A
114	$O-n-C_8H_{17}$	H	H	CH_3	A
115	$O-C_2H_5$	$O-C_2H_5$	H	CH_3	C(M)
116	$N-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	CH_3	E(M)
117	OH	Cl	H	H	A
118	$O-C_2H_5$	Cl	H	H	C(M)
119	$O-n-C_8H_{17}$	Cl	H	H	C
120	OH	H	H	Cl	A
121	$O-C_2H_5$	H	H	Cl	A^e
122	$O-n-C_8H_{17}$	H	H	Cl	A
123	$O-C_2H_5$	$O-C_2H_5$	H	Cl	C(M)
124	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	Cl	?^{d}
125	OH	NO_2	H	H	A
126	$O-C_2H_5$	NO_2	H	H	C
127	$O-n-C_8H_{17}$	NO_2	H	H	C(M)
128	OH	H	H	NO_2	A
129	$O-C_2H_5$	H	H	NO_2	A
130	$O-n-C_8H_{17}$	H	H	NO_2	A
131	OH	OH	H	NO_2	A
132	$O-C_2H_2$	$O-C_2H_5$	H	NO_2	A
133	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	NO_2	D(M)
134	$O-C_2H_5$	$O-CH_3$	H	H	C
135	$O-n-C_8H_{17}$	$O-CH_3$	H	H	C(M)
136	$O-C_2H_5$	H	H	$O-CH_3$	A
137	$O-n-C_8H_{17}$	H	H	$O-CH_3$	A
138	$O-C_2H_5$	$O-C_2H_5$	H	$O-CH_3$	B(M)
139	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	$O-CH_3$	E(M)
140	$O-n-C_4H_9$	C_2H_5	H	H	C(M)
141	$O-n-C_6H_{13}$	$n-C_4H_9$	H	H	C(M)
$R_4=R_4'=H$ $R_3=R_3'=H$					
142	$O-CH_3$	$O-CH_3$	H	H	C(M)
143	$O-C_2H_5$	$O-C_2H_5$	H	H	C
144	$O-n-C_3H_7$	$O-n-C_3H_7$	H	H	C(M)
145	$O-n-C_4H_9$	$O-n-C_4H_9$	H	H	C
146	$O-n-C_5H_{11}$	$O-n-C_5H_{11}$	H	H	C
147	$O-n-C_6H_{13}$	$O-n-C_6H_{13}$	H	H	C(M)
148	$O-n-C_7H_{14}$	$O-n-C_7H_{14}$	H	H	E(M)
149	$O-n-C_8H_{17}$	$O-n-C_8H_{17}$	H	H	E
150	$O-n-C_9H_{19}$	$O-n-C_9H_{19}$	H	H	E(M)
151	$O-n-C_{10}H_{21}$	$O-n-C_{10}H_{21}$	H	H	E(M)
152	$O-CO-CH_3$	$O-CO-CH_3$	H	H	C(M)
153	$O-CO-C_2H_5$	$O-CO-C_2H_5$	H	H	C
154	$O-CO-(n)C_3H_7$	$O-CO-(n)C_3H_7$	H	H	C(M)
155	$O-CO-(n)C_4H_9$	$O-CO-(n)C_4H_9$	H	H	C(M)

	R_1	R_1'	R_4	R_4'	Class
156	$O-CO-CH_2-CH(CH_3)_2$	$O-CO-CH_2-CH(CH_3)_2$	H	H	C
157	$O-CO-C_6H_5$	$O-CO-C_6H_5$	H	H	C(M)
158	$CH_2-(CH_2)_2-CN$	$CH_2-(CH_2)_2-CN$	H	H	E(M)
159	$O-(CH_2)_2-OH$	$O-(CH_2)_2-OH$	H	H	C
160	CH_3	CH_3	H	H	B(M)
161	$n-C_3H_7$	$n-C_3H_7$	H	H	C(M)
162	$n-C_5H_{11}$	$n-C_5H_{11}$	H	H	C(M)
163	$n-C_5H_{11}$	$n-C_5H_{11}$	CH_3	CH_3	C
164	$n-C_4H_9$	$n-C_5H_{11}$	CH_3	CH_3	C(M)
165	$n-C_5H_{11}$	$O-n-C_3H_7$	CH_3	CH_3	A^e
166	$n-C_4H_9$	$n-C_4H_9$	CH_3	CH_3	C
167	$n-C_5H_{11}$	$n-C_4H_9$	CH_3	H	C(M)
168	$n-C_4H_9$	$O-n-C_5H_{11}$	CH_3	H	C(M)
169	$n-C_4H_9$	$O-n-C_3H_7$	CH_3	H	C
170	$n-C_4H_9$	$n-C_4H_9$	CH_3	H	C(M)

^{a} (M) object for the training set; ^{b} These objects were included erroneously in class E for classification using the I, II and IIIrd set of variables. ^{c} Compounds for the training set; ^{d} This compound was included in class B for classification using the I, II and IIIrd set of variables; ^{e} These compounds were included erroneously in class C for classification using the I, II and IIIrd set of variables.

(Scheme 1). Most of the compounds were synthetized and studied in our laboratory (Zaragoza),[5-9] the remaining being selected from the literature.[10,11]

The collection includes 170 compounds, 140 being synthesized by us. They were divided into five different groups according to their mesogenic properties:

A, 62 compounds that melt normally (no liquid crystals).

B, 8 compounds that present nematic monotropism, N_m.

C, 67 enantiotropic nematic compound, N.

D, 16 compounds showing smectic mesophases, S_A and/or S_C.

E, 16 compounds showing nematic and smectic mesophases, N and S_C or S_A.

(one compound, the 124th, does not belong to any of the above groups, as will be explained later on).

The collection was divided into two parts, the first one (73 compounds) being used as a training set and the second (97 compounds) as a test set.[4] The division of the compounds into two groups was carried out in order to include representatives of all possible geometrical structures in the training set. In a second analysis, all the compounds were included.

The compounds are gathered in Table 1 with the mesogenic group they belong to and the indication of the set (training or test).

The choice of the variables used to characterize a compound is rather subjective, but takes into account the factors believed to be responsible for mesomorphism. Classically, three factors are regarded as important:

The molecular shape.

The existence of a permanent dipole moment in the molecule.

The anisotropy of the electronic polarizability.
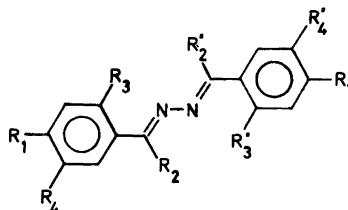
Scheme 1.

Table 2. Variables used to characterize the structures. The symbols used refer to Fig. 1.

Geometrics				
1. lM^b	(I,II,III,IV,V)[a]		40. $\Sigma|\mathscr{F}|R_1$	(I)
2. wM^c	(I,II,III,IV,V)		41. $\Sigma|\mathscr{F}|R_1'$	(I)
3. lR_1	(I,II,III,IV,V)		42. $\Sigma\mathscr{F}R_1+\Sigma\mathscr{F}R_1'$	(I,IV)
4. lR_1'	(I,II,III,IV,V)		43. $\Sigma|\mathscr{F}|R_1+\Sigma|\mathscr{F}|R_1'$	(I)
5. wR_1	(I,II,III)			
6. wR_1'	(I,II,III)		**Related to dipolar moments**	
7. $w\alpha$	(I,II,III,IV,V)		μ (vectorials)	
8. $w\alpha'$	(I,II,III,IV,V)		44. $\mu_x M$	(II)
9. lR_1+lR_1'	(IV,V)		45. $\mu_y M$	(II)
10. wR_1/lR_1	(IV,V)		46. $\mu_x R_1$	(II,IV,V)
11. wR_1'/lR_1'	(IV,V)		47. $\mu_Y R_1$	(II,IV,V)
12. Molecular weight	(I,II,III,IV,V)		48. $\mu_x R_1'$	(II,IV,V)
13. Number of intra-			49. $\mu_y R_1'$	(II,IV,V)
molecular H-bonding			50. $\mu_x \alpha$	(II,IV,V)
in the central core	(I,II,III)		51. $\mu_y \alpha$	(II)
			52. $\mu_x \alpha'$	(II,IV,V)
			53. $\mu_y \alpha'$	(II)
Related to Swain and Lupton's constants				
			μ (scalars)	
$\mathscr{R}$ (vectorials)			54. $\Sigma\mu M$	(II,IV,V)
14. $\Sigma\mathscr{R}_x\alpha^d$	(I,IV)		55. $\Sigma\mu|M|$	(II,IV,V)
15. $\Sigma\mathscr{R}_y\alpha^e$	(I)		56. $\Sigma\mu R$	(II,IV,V)
16. $\Sigma\mathscr{R}_x\alpha'$	(I,IV)		57. $\Sigma|\mu|R_1$	(II)
17. $\Sigma\mathscr{R}_y\alpha'$	(I)		58. $\Sigma\mu R_1'$	(II,IV,V)
18. $\Sigma\mathscr{R}_x M$	(I)		59. $\Sigma|\mu|R_1'$	(II)
19. $\Sigma\mathscr{R}_y M$	(I)		60. $\Sigma\mu R_1+\Sigma\mu R_1'$	(II,IV,V)
			61. $\Sigma|\mu|R_1'+\Sigma|\mu|R_1'$	(II)
$\mathscr{R}$ (scalars)			62. $\Sigma\mu\alpha$	(II,IV,V)
20. $\Sigma\mathscr{R}\alpha^f$	(I,IV)		63. $\Sigma|\mu|\alpha$	(II)
21. $\Sigma\mathscr{R}\alpha'$	(I,IV)		64. $\Sigma\mu\alpha'$	(II,IV,V)
22. $\Sigma|\mathscr{R}|\alpha^g$	(I)		65. $\Sigma|\mu|\alpha'$	(II)
23. $\Sigma|\mathscr{R}|\alpha'$	(I)		66. $\Sigma\mu$ bond of M	(II)
24. $\Sigma\mathscr{R}M$	(I,IV)		67. Number of lone pairs in	
25. $\Sigma|\mathscr{R}|M$	(I,IV)		the whole molecule, M	(II,IV,V)
$\mathscr{F}$ (vectorials)			**Related to Hammett's constants**	
26. $\Sigma\mathscr{F}_x\alpha$	(I,IV)		σ (vectorials)	
27. $\Sigma\mathscr{F}_y\alpha$	(I)		68. $\Sigma\sigma_x\alpha$	(III,V)
28. $\Sigma\mathscr{F}_x\alpha'$	(I,IV)		69. $\Sigma\sigma_y\alpha$	(III)
29. $\Sigma\mathscr{F}_y\alpha'$	(I)		70. $\Sigma\sigma_x\alpha'$	(III,V)
30. $\Sigma\mathscr{F}_x M$	(I)		71. $\Sigma\sigma_y\alpha'$	(III)
31. $\Sigma\mathscr{F}_y M$	(I)		72. $\Sigma\sigma_x M$	(III)
			73. $\Sigma\sigma_y M$	(III)
$\mathscr{F}$ (scalars)				
32. $\Sigma\mathscr{F}\alpha$	(I,IV)		σ (scalars)	
33. $\Sigma\mathscr{F}\alpha'$	(I,IV)		74. $\Sigma\sigma\alpha$	(III,V)
34. $\Sigma|\mathscr{F}|\alpha$	(I)		75. $\Sigma\sigma\alpha'$	(III)
35. $\Sigma|\mathscr{F}|\alpha'$	(I)		76. $\Sigma|\sigma|\alpha$	(III)
36. $\Sigma\mathscr{F}M$	(I,IV)		77. $\Sigma|\sigma|\alpha'$	(III)
37. $\Sigma|\mathscr{F}|M$	(I,IV)		78. $\Sigma\sigma M$	(III,V)
38. $\Sigma\mathscr{F}R_1$	(I,IV)		79. $\Sigma|\sigma|M$	(III,V)
39. $\Sigma\mathscr{F}R_1'$	(I,IV)			

[a] Classification groups including this variable; [b] l=length; [c] w=width; [d] $\mathscr{R}_x\alpha=\mathscr{R}$ component along the X-axis (α group); [e] $\mathscr{R}_y\alpha=\mathscr{R}$ component along the Y-axis (α group); [f] $\mathscr{R}\alpha$=Scalar algebraic sum; [g] $|\mathscr{R}|\alpha$=Scalar absolute sum.

According to that, the following independent variable were selected:

Intramolecular geometrical distances in Å, obtained from molecular models.[12] The bond lengths used are the ones determined by X-ray for structurally-similar compounds.[13-16]

Dipole moment and polarizability were calculated from the values of group and bond dipole moments.[17,18]

The Hammett [19] and Swain and Lupton [19] aromatic substituent constants.

For the polysubstituted compounds the independent variables were considered collectively. Three different summations were used (Table 2):

Vectorial: Take into account the origin and the direction of the substituents in relation to the principal axis of the molecule (Figure 1). $\mathscr{F}$, $\mathscr{R}$, μ and σ were treated in this way.

Scalar: Do not take into account the direction of the substituent. *Algebraic:* The constants were added up taking into consideration their sign (ΣCte). *Absolute:* The sum of the absolute values of the constants ($\Sigma|$Cte$|$). The following variables were treated in this way: $\mathscr{F}$, $\mathscr{R}$ and μ.

Due to the relationship existing between Hammett's σ and the Swain and Lupton's $\mathscr{F}$ and $\mathscr{R}$ constants, no model including both types of variables was tested.

The following sets of variables were used in the classification trials:

I, Geometric data and Swain and Lupton's constants (40 variables).

II, Geometric data and constants obtained from dipole moments (34 variables).

III, Geometric data and Hammett's constants (22 variables).

The first calculations were carried out with these sets (I, II and III). For each group of liquid crystals, the corresponding mathematical model was obtained and analyzed. The importance of the variables was determined for the different models.

Once this had been carried out, two supplementary sets of variables (IV and V) were constructed. They are combinations of the most representative variables of the first three sets (Table 2).

IV, Geometric data and a selection of the Swain and Lupton's constants plus a selection of variables data related to the dipole moments (39 variables).

V, Geometric data, a selection of the Hammett's constants and a selection of variables related to the dipole moments (30 variables).

Sets IV and V were similarly used to obtain the mathematical models for each group of liquid crystals (B, C, D, E).

In Table 2 the 79 different variables, their definition and the set (I to V) they belong to are gathered.

These five sets of variables were used to obtain the mathematical model from the training set of compounds. The validity of the model was tested against the test set.

RESULTS AND DISCUSSION

Classification using the I, II and IIIrd sets of variables. Class A: Non-mesogenic compounds were impossible to classify, even the changing the collection of compounds making up the training set. There is no spatial structure in this group, the compounds appearing randomly distributed all over the experimental domain. However, as the class A compounds do not belong to any of the four mesogenic group (B, C, D and E), they can be assigned by exclusion. Hence this is a case of asymmetric data structure.[3,4]

The most interesting results of the mathematical models for the mesogenic groups B, C, D and E are gathered in Table 3.

The performances of the models are also reported. The columns in Table 3 are as follows:

a: Mesogenic properties of the group (B to E).

b: Sets of variables used in the classification (I to III).

c: Number of compounds used in the training set, the total number of compounds in the group appears in brackets.

d: Minimum number of principal components (p.c.) of the model.

e: Total residual standard deviation (RSD) of the objects in a given p.c. model.

f: Percentage of the explained model variance.

g: Value of the "tolerance distance". A compound belongs to a class model if its RSD is not significantly larger than this value.

h: Number of well classified compounds in the group (percentage).

i: Number of wrongly classified compounds in the group (liquid crystals classified in group A) (percentage).

j: Number (J) of classified compounds from other groups of liquid crystals within the group. Percentage calculated in relation to the sum of compounds in the remaining groups: $100J/(\Sigma N)$. For instance, 23 compounds belonging to C, D and E groups were classified erroneously in group B (Table 3): $100 \times 23/(67+16+16)=23$ %, since there are 67, 16 and 16 compounds belonging to class C, D and E, respectively.

k: Number (K) of class A compounds classified within the group. Percentage calculated in relation to the total number of the group A compounds: $100K/62$.

A close examination of the compounds abnormally far from their class model, leads to the discovery of some experimental errors which were subsequently corrected.

Compounds *121* and *165* belonging to class A were included in class C.

Compound *124* shows nematic and smectic monotropism. Nematic mesophase has a very small temperature range. It was classified in class B, but in fact it did not belong to any of the four groups.

Compounds *70, 71* and *72* belonging to class C were included erroneously in class E. The mathematical treatment was repeated with a set which did not contain these compounds (see bottom of table 3) with a considerable improvement in the classification results (from 56 % to 94 %, column h of Table 3).

When the compounds that deviated most were excluded, the fitting of the models improved (the variance decreased) for each class of liquid crystals. On the other hand their discriminating power[4] worsened and a large number of false positives appear from the other classes, class A included.

For compound *94*, the SIMCA method pointed to an anomaly. An examination of the data matrix showed an error in the variables, an error that was subsequently corrected.

In other wrongly classified compounds no error was found either in the calculated data or in the mesogenic classification.

The results using the first three sets of variables (I, II, III) are good enough to detect big anomalies, but the discriminating power of the mathematical models is too poor to be useful. To go further, it would be necessary to analyse the role that the independent variables play in the classification.

Analysis of the variables used in sets I, II and III and selection of the new variables for sets IV and V. Among the geometrical variables, the most important are those related to molecular width (2, 7 and 8); the importance of the remaining ones depends on the class of mesogenic compounds under consideration. Whereas the molecule length (1) and the terminal chains (3 and 4) are fundamental for the classification of groups D and E, they are of little importance for groups B and C. On the other hand, variable 12, the molecular weight, is of fundamental importance for the classification of the last two groups. The variables 5 and 6 corresponding to the width of the terminal chains do not play any significant part in the classification. In sets IV and V they were replaced by the more interesting variables 10 and 11.

Variable 13 is of special interest. In the case of benzalazines, when R_3 (Scheme 1) is an
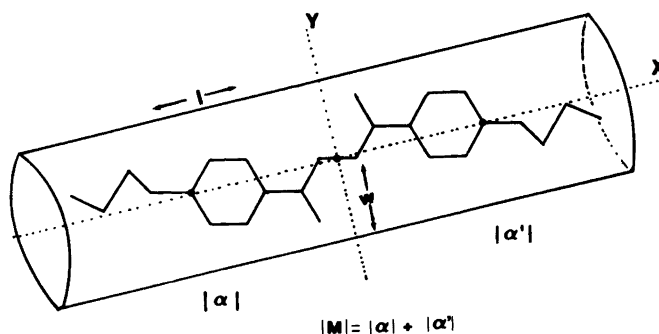
Fig. 1. Definition of the independent variables.

hydroxyl group, a six-membered hydrogen-bonded ring is formed with the nearest nitrogen atom of the azine bridge. The coplanarity that the chelate ring imposes on the phenyl and the central bridge has important consequences both for the structure and the mesogenic behaviour of benzalazines.[16] As expected, variable 13 (intramolecular hydrogen bond) has a considerable influence on the mathematical models. However, its own specificity limits its interest and in sets IV and V it was decided to omit this variable for the sake of generality.

For the other variables it was apparent that the "vectorial" variables corresponding to the whole molecule are on the whole of little importance whereas the ones corresponding to the aromatic rings, α and α' (Fig. 1) are important. With regard to the latter the most significant are the ones defined in relation to the principal axis of the molecule (X, Fig. 1). For this reason, in sets IV and V only the variables corresponding to the aromatic rings α and α' have been kept.

Other variables of moderate importance are those belonging to the group of "vectorially" added up variables related to the dipole moments of the terminal chains R_1 and R_1' (46, 47, 48 and 49). They were kept in sets IV and V because they are the only "vectorial" variables related to the terminal chains.

In the group of scalar variables, the "algebraic" ones are those which most influenced the classification. For this reason, they were kept in sets IV and V. Among the "absolute" variables, only those in rapport with the whole molecule $|M|$ were kept, since they have the largest weight in the classification.

Variable 67 represents the number of lone pairs in the molecule. Since it is of considerable importance in the classification, it was retained in sets IV and V. Variable 66 was excluded for the opposite reason.

Classification using the IV and Vth sets of variables. According to the previous considerations, the IVth set was formed by the following 39 variables:

Geometrical variables (1−4, 7−12).

Swain and Lupton's constants (14, 16, 20, 21, 24−26, 28, 32, 33, 36−39, 42).

Dipole moment (46−50, 52, 54−56, 58, 60, 62, 64, 67).

The last set differs from the preceding one in that the Swain and Lupton's constants were replaced by Hammett's.

V (30 variables):

Geometrical variables (1−4, 7−12).

Dipole moment (46−50, 52, 54−56, 58, 60, 62, 64, 67).

Hammett's constants (68, 70, 74, 75, 78, 79).

Table 3. Results of the mathematical models for the mesogenic groups B, C, D and E using the I, II and IIIrd sets of variables.

a	b	c	d	e	f	g	h	i	j	k
B	I	8[a]	2	0.636	60	0.636	7(87.5 %)	1(12.5 %)	23(23/99=23 %)	3(5 %)
	II	8[a]	1	0.768	42	0.768	7(87.5 %)	1(12.5 %)	37(37/99=37 %)	6(10 %)
	III	8[a]	1	0.791	37	0.791	6(75 %)	2(12.5 %)	40(40/99=40 %)	5(8 %)
C	I	40(67)	1	0.701	51	0.701	49(74 %)	18(27 %)	6(6/40=15 %)	4(6.5 %)
	II	40(67)	1	0.699	51	0.699	52(79 %)	15(22 %)	5(5/40=12.5 %)	7(11 %)
	III	40(67)	1	0.803	36	0.803	45(68 %)	22(33 %)	5(5/40=12.5 %)	7(11 %)
D	I	13(16)	2	0.613	62	0.613	14(87.5 %)	2(12.5 %)	13(13/91=14.3 %)	1(1.5 %)
	II	13(16)	2	0.565	68	0.565	15(94 %)	1(6 %)	8(8/91=8.8 %)	0(0 %)
	III	13(16)	3	0.391	85	0.391	15(94 %)	1(6 %)	20(20/91=22 %)	3(5 %)
E	I	16[b](16)	2	0.635	60	0.635	12(75 %)	4(25 %)	26(26/91=28.6 %)	1(1.5 %)
	II	16[b](16)	2	0.567	68	0.567	13(81 %)	3(19 %)	16(16/91=17.6 %)	3(5 %)
	III	16[b](16)	3	0.486	76	0.486	9(56 %)	7(44 %)	16(16/91=17.6 %)	3(5 %)
	III	13[c](16)	3	0.447	80	0.447	15(94 %)	1(6 %)	25(25/91=27.5 %)	3(5 %)

[a] Compound 124 included; [b] Compounds 70, 71 and 72 were included in class E by mistake; [c] Compounds 70, 71 and 72 deleted.

Table 4. Results of the mathematical models for the mesogenic groups B, C, D and E using the IV and Vth sets of variables.

a	b	c	d	e	f	g	h	i	j	k
B	IV	7(8)	2	0.580	66.5	0.140	8(100 %)	0	7[a](7/99=7 %)	0
	V	7(8)	2	0.577	67.0	0.120	8(100 %)	0	7[a](7/99=7 %)	1(1.5 %)
C	IV	40(67)	2	0.677	54.3	0.170	62(92.5 %)	5[b](7.5 %)	32[c](32/40=80 %)	20(32 %)
	V	40(67)	2	0.725	47.5	0.115	62(92.5 %)	5[b](7.5 %)	23[d](23/40=57.5 %)	12(19 %)
D	IV	13(16)	2	0.353	87.5	0.100	15(94 %)	1[e](6 %)	4[f](4/91=4.5 %)	0
	V	13(16)	3	0.274	92.5	0.040	15(94 %)	1[e](6 %)	2[f](2/91=2 %)	0
E	IV	13(16)	2	0.496	75.5	0.200	16(100 %)	0	10[g](10/91=11 %)	5(8 %)
	V	13(16)	3	0.447	80.0	0.145	16(100 %)	0	8[h](8/91=9 %)	1(1.5 %)

[a] From class C; [b] Compounds 59, 70, 126, 127, 159 classified in class A; [c] 7 from class B, 15 from class E and 10 from class D; [d] 5 from class B, 12 from class E and 6 from class D; [e] Compound 133; [f] From class E; [g] 5 from class C and 5 from class D; [h] 4 from class C and 4 from class D.
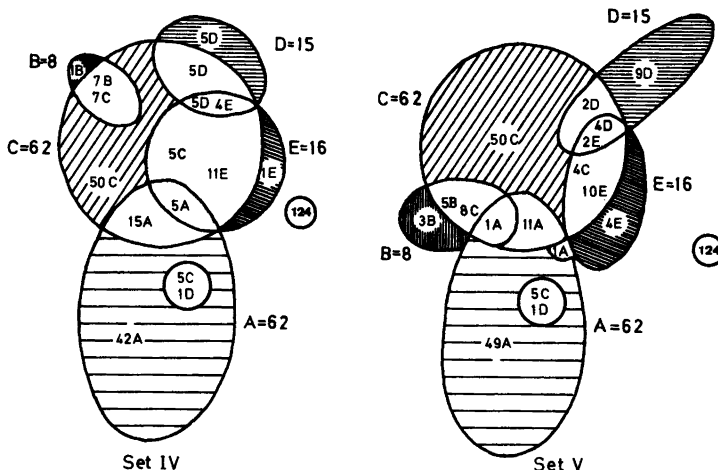
Fig. 2. Topological representation of the relationships between the different groups of liquid crystals.

The most interesting properties of the mathematical models corresponding to the mesogenic compounds B, C, D and E and the IV and Vth sets of variables are gathered in Table 4. The meaning of the columns is identical to that of the columns in Table 3.

The classification power of the new sets is considerably better than that of sets I, II and III. Remarkable is the decrease of false positives (columns j and k, Tables 3 and 4).

The mathematical models for classes B, D and E are clearly defined, but the model obtained for class C is somewhat vague, having a very low discriminating power. Its domain intersects with the domains (hyperboxes[4]) of the other classes. This corresponds to an inhomogeneity of the compounds making up class C with regard to the sets of variables used. For this group it would be necessary to include other structural properties in the sets of variables.

Fig. 2 represents graphically the distribution of the compounds in different classes using the sets of variables IV and V.

These representations deserve a few comments. To begin with, they only have a topological meaning. Secondly, the white areas correspond to wrongly classified compounds. For instance, in set IV the upper left white area contains seven compounds belonging to group B, classified in both B and C groups and seven other compounds belonging to group C also classified in both groups B and C. The central white area of set IV marked 5D+4E corresponds to five D class compounds classified in groups D, C and E and four E class compounds classified in the same groups. Apart from the fact that set V of variables is a better classifier than set IV, the main difference between the two sets is that there is a non-liquid crystalline compound A classified in B and C groups obliging group B to shift.

Ignoring the compounds that melt normally (class A) the relationships between the mesogenic groups in both diagrams (Fig. 2) can be summarized as follows (Scheme 2):

$$B \rightleftarrows C \rightleftarrows E \rightleftarrows D$$
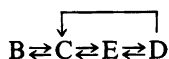
Scheme 2.

The arrows indicate that compounds of one class are included in another class. This phenomenon is generally reciprocal, save for the fact that no class C compound was included in class D.

Structurally, the classification of liquid crystals according to the increase of molecular order follows the sequence $N_m < N < S_A < S_C$, *i.e.* $B < C < E < D$, which corresponds quite well to Scheme 2 which shows no relationship between class B and classes E and D and a weaker relationship between classes C and D. Taking into account the fact, previously stated, that class C is ill-defined, the agreement between the pictures of figure 2 (see also scheme 2) and the molecular reality is satisfactory. It is also worth pointing out that the transition from one class of liquid crystals to another is quite smooth. For instance, a closely related series of compounds could even belong to four different classes of liquid crystals, see compounds *36−49* or *80−96* in Table 1.

Both sets of variables lead to the same six false negative compounds (Table 4, column i). Two of them (*59* and *159*) share in common an hydroxyethoxy ($HOCH_2CH_2O-$) substituent on the terminal chain (R_1, Scheme 1). This substituent can form intermolecular hydrogen bonds which modify the Van der Waals interactions responsible for the mesophase molecular order. None of the variables used takes into account this possibility. An examination of other compounds with the same terminal substituent shows that all of them are at the periphery of experimental region, confirming the necessity of introducing new variables in order to define the hydroxyethoxy substituent correctly.

Compounds *126, 127* and *133* are nitro derivatives substituted. We have already observed that this substituent shows a special behaviour compared with other substituents.[20]

All the variables of sets IV and V have an appreciable weight in the classification, with the exception of those related with the dipole moments of the terminal chains, 40−49, which may be omitted. Considering that set V has fewer variables than set IV and that it is a better classifier, (the number of compounds belonging to group A that can be defined by exclusion is superior) we recommend set V for the study of the mesogenic properties of benzalazines.

Using this last set, four still unsynthesized compounds have been classified with the following result:

2,2'-Dihydroxy-4,4'-dipentyloxybenzalazine: C group 2,2'-Dihydroxy-4,4'-dioctyloxy-benzalazine: C+D groups 3'-Aza-2-hydroxy-4,4'-dioctyloxy-α-methylbenzalazine: C+E groups.

4,4'-Acetylamino-α,α'-dimethylbenzalazine: A group (no liquid crystal).

We intend to synthesize these compounds in order to verify the predictive character of the SIMCA model obtained with set V of 30 variables.

REFERENCES

1. Albano, C., Blomqvist, G., Dunn, W.J., III, Edlund, V., Eliason, B., Johansson, E., Nordén, B., Sjöström, M., Söderstrom, B. and Wold, S. *Characterization and Classification Based on Multivariate Data Analyses,* In Varmavuori, A., Ed., *IUPAC 27th Intern. Congr. of Pure and Applied Chemistry,* Pergamon, Oxford 1976, p. 377.
2. Wold, S. *Pattern Recognition. 8* (1976) 127.
3. Albano, C., Dunn W.J., III, Edlund, V., Johansson, E., Nordén, B., Sjöström, M. and Wold, S. *Anal. Chim. Acta Comp. Techn. Optim. 2* (1978) 429.
4. Wold, S. and Sjöström, M. *Chemometrics: Theory and Application, Am. Chem. Soc. Symp. Ser. No. 52* (1977).
5. Marcos, M., Meléndez, E. and Serrano, J.L. *Mol. Cryst. Liq. Cryst. 91* (1983) 157.

6. Barberá, J., Marcos, M., Meléndez, E. and Serrano, J.L. *Mol. Cryst. Liq. Cryst. 94* (1983) 367.
7. Marcos, M., Meléndez, E. and Serrano, J.L. *An. Quim. C 79* (1983) 422.
8. Marcos, M., Meléndez, E., Romero, P. and Serrano, J.L. *An. Quim. Ser. C. In press.*
9. Serrano, J.L. *Tesis,* Universidad de Zaragoza, Zaragoza 1980. *Personal communication.*
10. Demus, D. and Zaschke, H. *Flüssige Kristalle in Tabellen,* 2nd Ed., Deut. Verlag Grundstoffind: Leipzig 1976.
11. Fergason, J.L. and Marshall, K.L. *U.K. Pat. Application GB* 2 027 026A (1980).
12. Framework Molecular models from Practice-Hall Inc. were used.
13. Galigne, S.L. and Falgueirettes, *J. Acta Cryst. 824* (1968) 1523.
14. Arcovito, L., Bonamico, M., Domenicano, A. and Vaciego, A. *J. Chem. Soc. B* (1969) 733.
15. Fayos, J., Martínez-Ripoll, M., Carcía-Mina, M.C., Conzalez-Martínez, J. and Arrese, F. *Acta Cryst. B 36* (1980) 1952.
16. Carcía-Mina, M.C., Arrese, F., Martínez-Ripoll, M., García-Blanco, S. and Serrano, J.L. *Acta Cryst. B 38* (1982) 2726.
17. Hill, N., Price, A.M., Vaughan, W.E. and Davies, M. *Dielectric Properties and Molecular Behaviour,* Van Nostrand, New York 1969, p. 246.
18. Diaz, M. and Roig, A. *Química Física VI,* Alhambra, Madrid 1972, p. 289.
19. Hansch, C. and Leo, A.L. *Substituent Constant for Correlation Analysis in Chemistry and Biology,* Wiley, New York 1979, p. 48.
20. Meléndez, E. and Serrano, J.L. *Mol. Cryst. Liq. Cryst. 91* (1983) 173.

-