JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 3, 651–658 (1990)

EFFECTS OF THE HETEROAROMATIC MOIETY ON SPECTROSCOPIC PROPERTIES, pK_a AND REACTIVITY OF AZOLES: A CHEMOMETRIC STUDY

CYNTHIA EBERT
Istituto di Chimica Farmaceutica, Università di Trieste, Piazzale Europa 1, 34128 Trieste, Italy
JOSÉ ELGUERO*
Instituto de Quimica Medica, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain
GIUSEPPE MUSUMARRA*
Dipartimento di Scienze Chimiche, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy

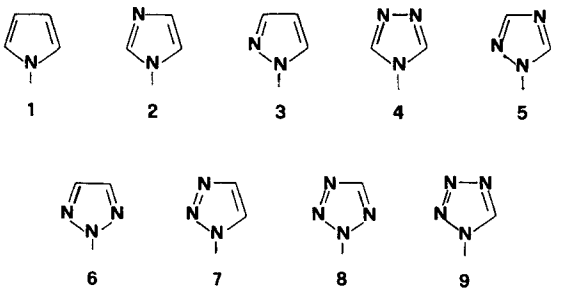
The spectroscopic and physicochemical properties of parent azoles substituted on the nitrogen by a proton, a methyl group, a *p*-aminophenyl group and an acetyl group have been gathered. This information on azol-1-yl substituents has been treated by principal component analysis and partial least-squares analysis. The first result of these analyses is that it has been possible to complete some missing properties of azol-1-yl substituents, such as pK_a for proton loss of some azoles and hydrolysis rates of certain azolides. Another significant result of the chemometric approach is the assignment of the 1-acetyltetrazole IR and NMR data to one of the two possible isomers.

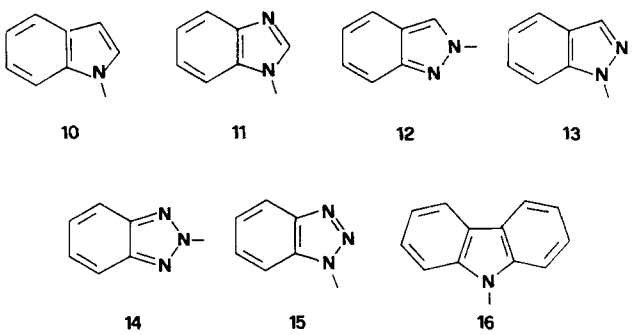
INTRODUCTION

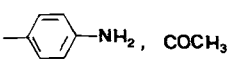
Among the classical heteroaromatic compounds, azines and azoles, the latter are characterized by the existence of a bond between the pyrrole-like sp^2 nitrogen and the substituent X. This situation is unique in neutral aromatic compounds, not including functional derivatives such as pyridones.

Ar-X

Ar:





X: H, CH_3,  , $COCH_3$

a b c d

The effects of *C*-substituents, common to all aromatic and heteroaromatic compounds, have been widely studied. In contrast, there are very few reports dealing with the effects of *N*-substituents. To study these effects, there are several possibilities, e.g. the use of pyridinium salts, pyridones and azoles. Regarding

* Authors for correspondence.

0894–3230/90/100651–08$05.00

Received 28 November 1989
Revised 24 April 1990

azoles, only the effect of substituents X on ^{13}C NMR parameters has been carefully examined.[1]

The effect of *N*-substituents and of the heteroaromatic moiety on the spectroscopic properties, equilibria and reactivity of azoles have also been studied. We have already discussed the influence of the azole on the conformation of *N*-arylazoles (dihedral angle),[2-5] on the isomerism, conformation, rotational barrier and spectroscopic proprieties of *N*-acylazoles[6-12] and on the structure of *N,N*-linked biazoles.[13] Moreover, the azoles have often been used as test compounds for theoretical calculations owing to their homogeneity and uniqueness.[14]

METHODS

Multivariate statistical methods. A data set suitable for a multivariate analysis consists of a table (matrix) where a number (M) of experimental values (variables) are collected for each of N compounds (objects). The geometrical interpretation of each object is a point in the M-dimensional space, where each variable defines an orthogonal axis. Accordingly, the data set has the form of N points in an M-space. Multivariate methods search for the structure of the data, i.e. they are aimed at recognizing systematic patterns, if present.[15-18]

Principal component analysis (PCA), the statistical method adopted here as an alternative to multiple regression analysis (MRA),[19] looks for systematic variations in the data matrix to elucidate the structure of the objects in the M-space.[20] PCA requires no *a priori* assumptions and selects the best model, with the minimum number of dimensions, which explains the data structure.

The SIMCA method carries out PCA on a data matrix containing elements x_{ik}, where k represents the experimental measurements (variables) and i the chemical compounds (objects). Each element is described by the equation

$$x_{ik} = \bar{x}_k + \sum_{a=1}^{A} t_{ia} p_{ak} + e_{ik} \qquad (1)$$

where the number A of significant cross-terms (components) and the parameters p_{ak} and t_{ia} are calculated by minimizing the squared residuals e_{ik}, after subtracting $\bar{x}_k$ (the mean value of the i experimental quantities x_k).

In this model, parameters $\bar{x}_k$ and p_{ak} depend only on the variables and t_{ia} only on the compounds. The deviations from the model are expressed by the residuals e_{ik}. The number of significant components (A) is determined using the cross-validation technique.[21]

The relevance of each variable in describing the mathematical model is given by its modelling power, $MPOW_k$:

$$MPOW_k = 1 - s_k^{\{A=A\}} / s_k^{\{A=0\}} \qquad (2)$$

where s_k is the residual standard deviation for each variable after A dimensions and after dimension zero.

The recently developed algorithm called partial least-squares (PLS)[22-24] is aimed at finding the relationship existing between one or more 'dependent' variables and a group of explanatory variables. This method gives a description of the X matrix by the one principal component-like model [equation (1)] and a description of the y vector as a predictive relationship with the latent variable t [equation (3)], where b_a is a proportionality coefficient for each dimension.

$$y_{ia} = \Sigma\, b_a t_{ia} + h_{ia} \qquad (3)$$

The algorithm used in the SIMCA/MACUP method, presented in detail elsewhere,[22] is iterative for each dimension as in PCA. It consists in finding the latent variables of the X matrix t_{ia} in such a way that the relationship between y_i and t_i is maximized.

PCA and PLS have already been applied successfully to investigate the applicability of the Hammett equation to five-membered heterocycles,[25] to study the simultaneous dependence of nucleophilic displacement rates on alkyl group structure and leaving group nucleofugacity in *N*-alkyl pyridinium and related cations[26] and to demonstrate the orthogonality of classic and magnetic aromaticity scales in five- and six-membered heterocycles.[27]

Objectives of the study. In a systematic exploration of a family of compounds, there are always some data missing owing to the instability of some of them (minor tautomers) or to the difficulty in measuring certain properties (very slow or very rapid kinetics). A possibility for solving this difficulty is through theoretical calculations ('compounds not easily amenable to experiment');[14] another method for obtaining a complete picture of spectroscopic, equilibrium and reactivity data is to use a chemometric approach. Thus, one of the objectives of this study was to fill the empty values in Table 1.

Further, the parameters can be used to assign spectroscopic data (measured for one of two possible isomers) to the isomer which fits the model better, i.e. to identify the isomeric structure of a heterocycle.

RESULTS AND DISCUSSION

Principal component analysis

Class A

[60 elements; 12 objects (1–3, 5–7, 10–13, 15 and 16); 5 variables (1–5)]

IR and ^{13}C NMR data (variables 2, 3 and 4) reported in Table 1 for 1-(*N*-acetyl)tetrazole (**9d**), in the reference paper[1] were not assigned to one of the two poss-

Table 1. Available characteristics for azoles **1–16**

Compound	Ar	Var. No: / *X*:	1 *p*-NH_2Ph σ_p [a]	2 $COCH_3$ ν(CO) [b]	3 $COCH_3$ δ(CO) [c]	4 $COCH_3$ $\delta(CH_3)$ [c]	5 CH_3 $\delta(^{15}N)$ [d]	6 H pK_a [e]	7 $COCH_3$ ln $t_{1/2}$ [f]
1	Pyrr-1-yl		0·10	1732	167·5	21·9	230·2	17·51	—
2	Imidazol-1-yl		0·24	1747	166·3	22·3	218·1	14·40	3·71
3	Pyrazol-1-yl		0·19	1746	169·2	21·3	179·3	14·21	6·81
4	1,2,4-Triazol-4-yl		0·33	[g]	[g]	[g]	217·1	[h]	[g]
5	1,2,4-Triazol-1-yl		0·37	1765	167·8	21·8	171·3	[h]	1·86
6	1,2,3-Triazol-2-yl		0·35	1780	165·7	21·8	131·4	[i]	—
7	1,2,3-Triazol-1-yl		0·40	1762	168·3	22·1	143·0	[i]	3·28
8	Tetrazol-2-yl		0·59	[j]	[j]	[j]	101·8	[k]	—
9	Tetrazol-1-yl		0·52	1779	164·7	23·1	151·4	[k]	—
10	Indol-1-yl		—	1710	168·2	23·5	253·6	16·97	—
11	Benzimidazol-1-yl		0·38	1730	166·9	23·3	236·4	12·86	6·63
12	Indazol-2-yl		—	1752	170·8	21·8	162·1	[l]	—
13	Indazol-1-yl		—	1720	170·7	22·7	203·8	[l]	—
14	Benzotriazol-2-yl [m]		—	—	—	—	117·0	[n]	—
15	Benzotriazol-1-yl		—	1735	169·3	22·8	161·8	[n]	4·74
16	Carbazol-1-yl		0·38	1692	169·3	27·2	272·7	16·70	—

[a] Hammett σ_p constants for *N*-(*p*-aminophenyl) derivatives (**c**), calculated from NMR chemical shifts, from Ref. 29.
[b] Stretching carbonyl frequencies, for *N*-acetyl derivatives (**d**), from Refs 31 and 32.
[c] ^{13}C NMR chemical shifts for *N*-acetyl derivatives (**d**), from Ref. 1.
[d] ^{15}N NMR chemical shifts for *N*-methyl derivatives (**b**), from Ref. 33.
[e] pK_a for proton loss [(**a**) derivatives, X = H], from Ref. 30.
[f] ln $t_{1/2}$ for the hydrolysis of *N*-acetylazoles (**d**), from Ref. 31.
[g] *N*-Acetyl derivative too unstable to be isolated.
[h] A pK_a of 10·04 was obtained for a possible mixture of tautomers.
[i] A pK_a of 9·26 was obtained for a possible mixture of tautomers.
[j] Available data were assigned to the 1-yl derivative **9**; see discussion of PCA, class A.
[k] A pK_a of 4·90 was obtained for a possible mixture of tautomers.
[l] A pK_a of 13·86 was obtained for a possible mixture of tautomers.
[m] Object not included in any calculation.
[n] A pK_a of 8·38 was obtained for a possible mixture of tautomers.

ible isomers **8d** or **9d**. PCA of a data matrix (class A) without tetrazoles **8** and **9** provided a three significant principal component model, which explains 77% of the total variance. In this PCA (class A), the values of the residuals for the variables within each object (tetrazole) will be adopted as a guide for the assignment of variables 2, 3 and 4 to a specific isomeric structure. This procedure has already proved to be successful for the assignment of ^{13}C NMR shifts.[28]

Both tetrazoles **8** and **9** (assigning the same values for variables 2, 3 and 4) were then fitted, as test objects, into the above PCA model with the aim of excluding one of the two possible isomeric structures on the basis of the 'object' standard deviations (s_i) and of the 'variable' residuals (e_{ik}). In other words, if the s_i and e_{ik} values are significantly lower for one isomer, this isomer fits the PCA model better (i.e. its values of variables 1–5 are in better agreement with those of other azoles) and then it is likely that the IR and ^{13}C NMR values for variables 2, 3 and 4 refer to this isomer.

The residual standard deviation and the residuals reported in Table 2 are significantly lower for 1-tetrazole (**9**), showing that this isomer fits the PCA model better than 2-tetrazole (**8**). On the basis of the above considerations, the values 1779, 164·7 and 23·1 are reported in Table 1 as variables 2, 3 and 4, respectively, for 1-(*N*-acetyl)tetrazole (**9d**), whereas for 2-(*N*-acetyl) tetrazole (**8d**) the above variables are considered as unknown values.

Table 2. Residual standard deviation for each tetrazole (s_i) and residuals for variables 1–5 (e_{ik}) in class A

Compound (*i*)	s_i	e_{i1}	e_{i2}	e_{i3}	e_{i4}	e_{i5}
8	0·52	−0·09	−0·40	−0·06	0·18	−0·58
9	0·06	0·02	−0·01	−0·02	−0·06	0·05

Class B

[70 elements; 14 objects (1–3, 5–13, 15 and 16); 5 variables (1–5)].

The second PCA was carried on a data matrix where the values of variables 2, 3 and 4 for 1-(*N*-acetyl)tetrazole (**8d**) were now given as unknown

Table 3. PC 'scores' for azoles (class B)

Compound	Ar	t_1	t_2	t_3
1	Pyrr-1-yl	−0·69	1·09	1·61
2	Imidazol-1-yl	0·09	0·03	1·39
3	Pyrazol-1-yl	0·15	1·57	0·30
5	1,2,4-Triazol-1-yl	1·05	0·25	0·06
6	1,2,3-Triazol-2-yl	2·15	−0·18	0·47
7	1,2,3-Triazol-1-yl	1·12	0·23	−0·57
8	Tetrazol-2-yl	4·26	−1·06	0·20
9	Tetrazol-1-yl	1·97	−1·72	0·25
10	Indol-1-yl	−1·77	−0·35	0·76
11	Benzimidazol-1-yl	−0·69	−0·82	0·72
12	Indazol-2-yl	0·28	2·07	−0·64
13	Indazol-1-yl	−1·22	1·60	−0·21
15	Benzotriazol-1-yl	−0·19	0·80	−0·59
16	Carbazol-1-yl	−3·59	−1·60	−0·50

values, available data being already assigned to compound (**9**) (cf. class A). The aim of this analysis was to evaluate the information content of the variables (characteristics) from the 'loadings' plot and to look for systematic patterns exhibited by the azoles (objects) in the 'scores' plot.

The number of PC existing in characteristic vector space is less than the number of variables in the data set. For class B, a three significant PC model, which accounts for 86% of the total variance (42% first PC + 24% second PC + 20% third PC) was found. The first PC is the best summary of the linear relationship exhibited in the data. The second component (orthogonal to the first) accounts for the most residual variance when the effect of the first component is removed from the data. Subsequent components are defined similarly until all the systematic variance in the data is exhausted. The PC 'scores' plots for azoles (compounds) are reported in Table 3 and represented graphically on the 'scores' plots of t_2 vs t_1 and t_3 vs t_1 (Figure 1). These plots show that t_1 values generally increase on replacement of a ring carbon atom by a 'pyridine-like' nitrogen, particularly when the latter is adjacent to the pyrrole-like nitrogen (see azoles **1–3–6**, **2–5–8** or **2–7–8**, **10–13** and **11–15**), and decrease with benzo substitution (see **1–10–16**, **2–11**, **3–13** and **7–15**). Also, the t_2 values for pyrrole and imidazole decrease significantly with benzo substitution.

The PC loadings for characteristics **1–5**, which will be used as explanatory variables in the PLS analysis, are given in Figure 2. For the interpretation of the loadings plots it is worth mentioning that variables which lie in the same direction with respect to the origin (0,0 point) have similar information content. Figure 2 shows clearly that the variables under examination exhibit different information contents.

PLS analysis

In the PLS treatment, we select, for a set of compounds (reference set), a number of available characteristics (independent variables, X-block) as descriptors of the property taken in turn as the dependent variable (y). The Hammett σ_p for N-azoles,[29] the pK_a for the acidity

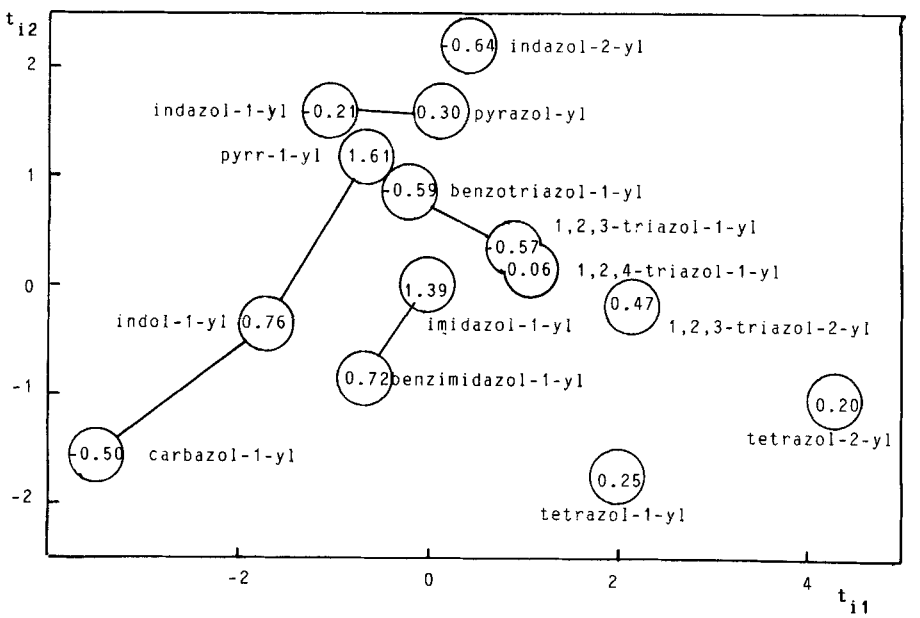


Figure 1. Scores plot of t_2 vs t_1 with t_3 indicated inside circles for class B

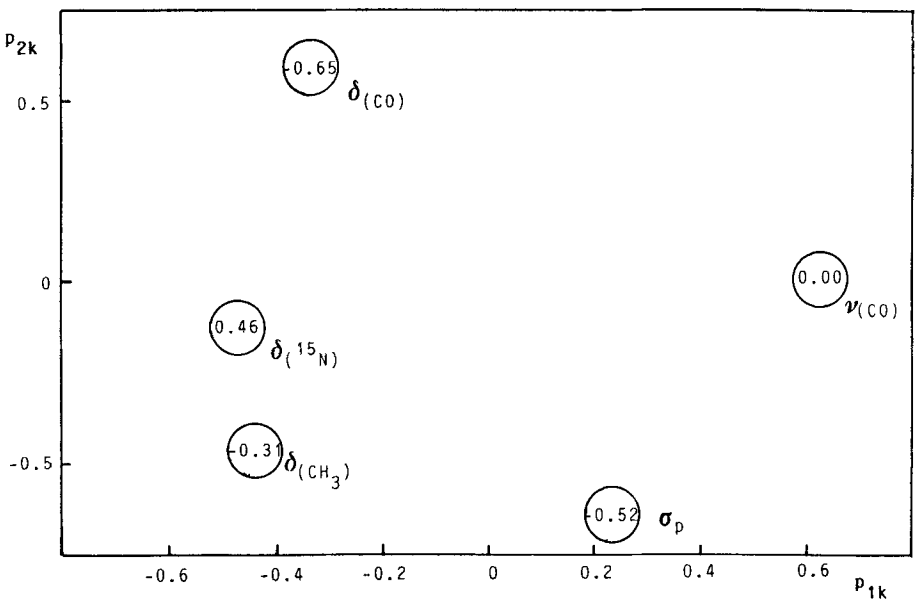


Figure 2. Loadings plot of p_2 vs p_1 with p_3 indicated inside circles for class B

of azoles[30] and the logarithm of half-lives (min) for the hydrolysis of *N*-acetylazoles[31] were taken, in turn, as the dependent variable in three PLS treatments (classes C–E) using variables 2–5 (class C) or variables 1–5 (classes D and E) as explanatory variables (*X*-block). In order to carry out the PLS analysis, the objects (azoles) were divided into a reference or learning set (in which all values for both explanatory and dependent variables are needed) used to build the PLS model, and a test set including objects for which data for one or more of the above dependent variables were missing. A typical data matrix for a PLS analysis is reported in Figure 3. The objects and the variables included in the reference set of each class (according to data availability) are recorded in Table 4, together with the percentage of explained variance and the number of significant PLS components (PLSC) obtained.

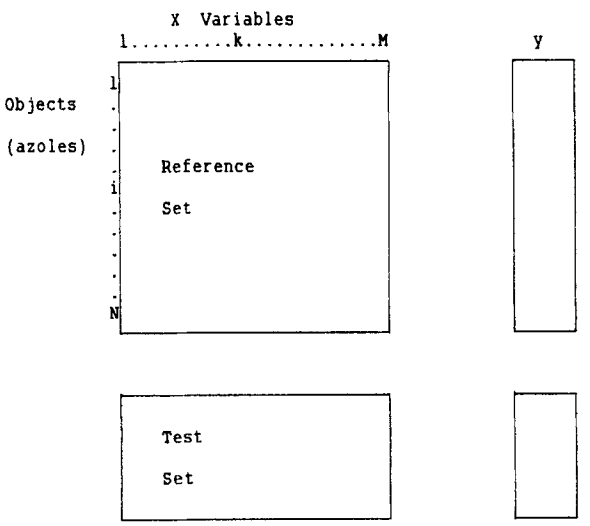


Figure 3. Typical data matrix for PLS analysis. *X* variables are the explanatory variables or descriptors (Hammett σ_p, IR and NMR data), *y* is, in turn, the dependent variable (Hammett σ_p, equilibrium or reactivity data). The reference set is used to build the PLS model whose parameters allow the PLS predictions for compounds in the test set

Class C

PLS analysis using variables 2–5 as descriptors and Hammett σ_p as the dependent variable, provided a two PLSC model which explains only 62% of total variance. Variables 2–5 turned out to be not particularly appropriate descriptors to describe the variation of σ_p.

The modelling powers after two PLSC for the explanatory variables, reported in Table 5, are significantly higher for the infrared carbonyl stretching frequencies (variable 2) and for the methyl ^{13}C NMR shifts (variable 4) of *N*-acetylazoles, indicating that the above variables are better descriptors for the dependent variable (σ_p). This finding can be rationalized taking into account the relative position of the azole and of the 'probe' used for the evaluation of its 'effects' (Figure 4). Hammett σ_p values were calculated from proton chemical shifts of the amino group in 1-(*p*-aminophenyl)azoles where the 'probe' (the amino

Table 4. Objects and variables used for each PLS analysis

Class[a]	Objects in the reference set	Explanatory variables	y[b]	PLSC[c]	%V[d]
C	**1, 2, 3, 5, 6, 7, 9, 11, 16**	2, 3, 4, 5	1	2	62 (36 + 26)
D	**1, 2, 3, 11, 16**	1, 2, 3, 4, 5	6	3	94 (31 + 55 + 8)
E	**2, 3, 5, 7, 11**	1, 2, 3, 4, 5	7	3	96 (53 + 27 + 16)

[a] Each class made selecting the appropriate values from Table 1.
[b] Dependent variable.
[c] Number of PLS components.
[d] Percentage of total variances explained; % variance for first, second and third PLSC given in parentheses.

Table 5. Modelling powers ($MPOW_k$) for explanatory variables in the PLS analysis

	Class							
	C		D			E		
Var. (k)	$MPOW_{1k}$	$MPOW_{2k}$	$MPOW_{1k}$	$MPOW_{2k}$	$MPOW_{3k}$	$MPOW_{1k}$	$MPOW_{2k}$	$MPOW_{3k}$
1	—	—	0	0·93	0·93	0	0	0·82
2	0·36	0·80	0·67	0·89	0·89	0·70	0·70	0·92
3	0·42	0·38	0	0	0·99	0	0·69	0·88
4	0	0·98	0·41	0·77	0·77	0	0·30	0·84
5	0·26	0·47	0·29	0·29	0·90	0·48	0·75	0·69

group) is connected to the substituent (the azole) by transmitting system (the benzene ring). Carbonyl stretching frequencies and methyl carbon shifts of *N*-acetylazoles (variables 2 and 4) are 'probes' not directly linked to the azole and then somewhat similar to the Hammett σ_p. Carbonyl shifts of *N*-acetylazoles and nitrogen-15 shifts of *N*-methylazoles (variables 3 and 5) are 'probes' located at the N—C bond and then influenced by specific proximity effects of the azole, which are relevant to the prediction of other properties such as the hydrolysis rates (see discussion of class E).

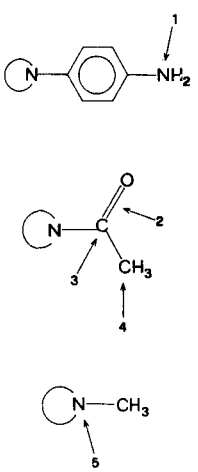


Figure 4. Positions of the 'probe' (indicated by an arrow) used in variables 1–5 for the evaluation of the 'effect' of azoles

The above considerations and the low percentage of variance explained by the PLS model prevent its use for precise predictive purposes. However, PLS predictions reported in Table 6 provide an approximate estimation for unavailable σ_p values of azoles (test set).

Class D

PLS analysis using variables 1–5 as descriptors and the pK_a for the acidity of azoles as the dependent variable provided a three significant PLSC model which accounts for 94% of the total variance. The modelling powers recorded in Table 5 show that the carbonyl stretching (variable 2) and, to a lesser extent, methyl and nitrogen NMR chemical shifts (variables 4 and 5) are relevant in determining the first PLSC (31% variance), whereas the second PLSC (55% variance) is mainly determined by σ_p (variable 1); after three PLSC all explanatory variables exhibit approximately the same modelling power, i.e. all of them contribute to the PLS model.

The predictions reported in Table 6 are in excellent agreement with the experimental values not only for compounds in the reference set (**1, 2, 3, 11** and **16**) but

Table 6. PLS predictions for dependent variables in classes C, D and E

		Class:	C		D		E	
		Y:	σ_p		pK_a		log $t_{1/2}$	
Compound	Ar		Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
1	Pyrr-1-yl		0·10	0·20	17·51	17·66	∞	7·41
2	Imidazol-1-yl		0·24	0·30	14·40	14·09	3·71	3·56
3	Pyrazol-1-yl		0·19	0·18	14·21	14·20	6·81	6·83
4	1,2,4-Triazol-4-yl		0·33	[a]	10·04[b]	[a]	—	[a]
5	1,2,4-Triazol-1-yl		0·37	0·29		9·97	1·86	2·17
6	1,2,3-Triazol-2-yl		0·35	0·42	9·26[b]	7·90	—	−2·76
7	1,2,3-Triazol-1-yl		0·40	0·32		8·72	3·28	3·03
8	Tetrazol-2-yl		0·59	[a]	4·90[b]	[a]	< −0·70	[a]
9	Tetrazol-1-yl		0·52	0·52		5·68	—	−3·67
10	Indol-1-yl		—	0·23	16·97	17·43	∞	11·35
11	Benzimidazol-1-yl		0·38	0·31	12·86	13·14	6·63	6·70
12	Indazol-2-yl		—	0·18	13·86[b]	12·51	—	7·84
13	Indazol-1-yl		—	0·16		15·27	—	12·62
15	Benzotriazol-1-yl		—	0·27	8·38	12·90	4·74	8·62
16	Carbazol-1-yl		0·38	0·40	16·70	16·60	—	16·98

[a] PLS predictions not available owing to the lack of explanatory variables.
[b] pK_a obtained for a possible mixture of tautomers.

also for indole, which belongs to the test set. The experimental acidities for 1,2,4-triazole (10·04) and for 1,2,3-triazole (9·26) are consistent with a great excess of the 1H-tautomer **5a** and **7a**, respectively (for a detailed discussion for tautomeric equilibria see Ref. 34); the less satisfactory PLS predictions for indazoles **12a** and **13a** also suggest **13a** as the possible prevailing tautomer.

No information about the tautomeric equilibria of tetrazole can be obtained by the present model owing to the lack of sufficient explanatory variables, which does not allow reliable PLS predictions.

Class E

The hydrolysis rates of *N*-acetylazoles, selected as the dependent variable for this class, appear to be described fairly well by the selected explanatory variables (1–5). A three significant component model explaining 96% of total variance is found. The modelling powers in Table 5 show that the first PLSC (i.e. the greatest systematic descriptor variation which can be related to the dependent variable) accounts for 53% of the variance and is determined by the IR carbonyl stretching frequencies of *N*-acetylazoles (variable 2) and by the nitrogen shifts of *N*-methylazoles (variable 5) with no contribution from other descriptors. This trend is again consistent with the overall picture depicted in Figure 4. For the description of the rates of hydrolysis of *N*-acetylazoles, implying C—N bond breaking, descriptors whose 'probes' are either atoms forming this bond or properties sensitive (related) to its strength are needed. Both the IR carbonyl stretching frequencies of *N*-acetylazoles (variable 2) and the nitrogen shifts of *N*-methylazoles (a property describing the azole nitrogen environment when attached to a carbon atom) appear to be very appropriate explanatory variables.

The second PLSC (accounting for 27% residual variance when the effect of the first PLSC is removed from the data) is determined by the carbonyl carbon (the carbon forming the C—N bond) shifts of *N*-acetylazoles (variable 3, $MPOW_2 - MPOW_1 = 0{\cdot}69$), by the methyl carbon shifts of *N*-acetylazoles (variable 4, $MPOW_2 - MPOW_1 = 0{\cdot}30$) and again by the nitrogen shifts of *N*-methylazoles (variable 5, $MPOW_2 - MPOW_1 = 0{\cdot}27$).

The third significant PLSC (accounting for 16% of the variance) is mainly required to describe the explanatory capability of the azole electronic effects, expressed by σ_p values (variable 1), which are also important in determining the hydrolysis rates. After three PLSC all variables contribute, approximately to the same extent, to the PLS model and no systematic variance is left in the data.

The PLS predictions in Table 6, satisfactory for compounds in the reference set, provide the order of magnitude for the half-lives of *N*-acetylpyrrole (**1d**) and *N*-acetylindole (**10d**), for which limit infinite values were available, whereas *N*-acetylbenzotriazole (**15d**) appears to be an outlier of the PLS model.

CONCLUSIONS

The two main objectives of this study have been achieved. First, the PLS predictions in Table 6 complete fairly well the experimental results in Table 1,

making possible the use of azoles in QSAR studies when the corresponding log *Ps* are known. It is interesting to compare such a chemometric approach with the theoretical approach. For azoles, we can examine if there is a relationship between INDO-calculated parameters[14] and calculated values in Table 6. For instance, there exists a linear equation which relates the calculated pK_a values in Table 6 to the charge of the NH hydrogen, $pK_a = -163{\cdot}6 - 233{\cdot}1q_H$, $r^2 = 0{\cdot}96$ (compounds **1**, **2**, **3**, **5**, **6**, **7** and **9**).

The second objective, to use the comparison between the observed value and those calculated for a pair of isomers for deciding the structure of the most stable isomer, has also been achieved. Thus, the deviations from the PCA model (expressed as residual standard deviations and as residuals) allow the assignment of the isomeric structure of 1-(*N*-acetyl)tetrazole (**9d**) on the basis of spectroscopic data.

ACKNOWLEDGEMENTS

We thank the Consiglio Nazionale delle Ricerche (CNR, Rome) and the Consejo Superior de Investigaciones Cientificas (CSIC, Madrid) for a travel grant to G.M. and the Ministero della Ricerca Scientific e Tecnologica (MURST) for financial support to C.E. and G.M. for research on chemometrics.

REFERENCES

1. M. Begtrup, J. Elguero, R. Faure, P. Camps, C. Estopà, D. Ilavsky, A. Fruchier, C. Marzin and J. de Mendoza, *Magn. Reson. Chem.* **26**, 134–151 (1988).
2. J. Elguero, R. Jacquier and S. Mondon, *Bull. Soc. Chim. Fr.* 1346–1351 (1970).
3. J. P. Rioux, C. Clément, R. Jacquier, G. Tarrago and J. Elguero, *Bull. Soc. Chim. Fr.* 2144–2149 (1970).
4. J. Elguero, A. Fruchier and R. Jacquier, *An. Asoc. Quim. Argent.* **58**, 307–312 (1970).
5. T. Avignon, L. Bouscasse and J. Elguero, *Tetrahedron*, **34**, 1139–1145 (1978).
6. J. Elguero, A. Fruchier and R. Jacquier, *Bull. Soc. Chim. Fr.* 3041–3042 (1966).
7. S. Mignonac-Mondon, J. Elguero and R. Lazaro, *C. R. Acad. Sci.* **276**, 1533–1535 (1973).
8. J. Elguero, C. Marzin and L. Pappalardo, *Bull. Soc. Chim. Fr.* 1137–1141 (1974).
9. J. Elguero, A. Fruchier, L. Knutsson and R. Lazaro, *Can. J. Chem.* **52**, 2744–2750 (1974).
10. H. Sauvaitre, J. Teysseyre and J. Elguero, *Bull. Soc. Chim. Fr.* 635–641 (1976).
11. J. Elguero, A. Fruchier, L. Pappalardo and M. C. Pardo, *An. Quim.* **74**, 1529–1535 (1978).
12. M. Begtrup, R. M. Claramunt and J. Elguero, *J. Chem. Soc., Perkin Trans. 2* 99–104 (1978).
13. M. L. Castellanos, S. Olivella, N. Roca, J. De Mendoza and J. Elguero, *Can. J. Chem.* **62**, 687–695 (1984).
14. J. Catalán, J. L. G. de Paz, M. Yañez and J. Elguero, *Chem. Scr.* **24**, 84–91 (1984), and references therein.
15. B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.* **94**, 5632–5639 (1972).
16. M. Sjöström and S. Wold, *Acta Chem. Scand., Ser. B* **35**, 537–554 (1981).
17. S. Wold, M. Sjöström, in *Chemometrics: Theory and Application*, edited by B. R. Kowalski ACS Symposium Series, No. 52, pp. 243–282. American Chemical Society, Washington, DC (1977).
18. C. Albano, W. J. Dunn, III, U. Edlund, E. Johansson, B. Norden, M. Sjöström and S. Wold, *Anal. Chim. Acta* **103**, 429–443 (1978).
19. N. Draper and M. Smith, *Applied Regression Analysis*, 2nd ed. Wiley–Interscience, New York (1981).
20. (a) K. V. Mardia, J. T. Kent and J. M. Biddy, *Multivariate Analysis*. Academic Press, London (1979); (b) S. Wold, K. Esbensen and P. Geladi, *Chemometr. Intell. Lab. Syst.* **2**, 37–52 (1987).
21. S. Wold, *Technometrics* **20**, 397–405 (1978).
22. S. Wold, C. Albano, W. J. Dunn, III, K. Esbensen, S. Hellberg, E. Johansson and M. Sjöström, *Food Research and Data Analysis*, edited by H. Martens and H. Russwurm, pp. 147–162. Applied Science, Barking (1983).
23. H. Wold, in *Systems Under Indirect Observation, Causality–Structure Prediction*, edited by K. G. Joreskog and H. Wold, Vol. 2. North-Holland, Amsterdam (1981).
24. S. Clementi, G. Cruciani and G. Curti, *Anal. Chim. Acta* **191**, 149–160 (1986).
25. (a) S. Alunni, S. Clementi, C. Ebert, P. Linda, G. Musumarra, M. Sjöström and S. Wold, *J. Chem. Soc., Perkin Trans. 2* 485–490 (1985); (b) G. Musumarra, in *Thiophene and its Derivatives*, edited by S. Gronowitz, Vol. 44, Part 3, Chapt. V, pp. 975–1153. Wiley, New York (1986).
26. G. Musumarra, M. Bruno, A. R. Katritzky, K. Sakizadeh, S. Alunni and S. Clementi, *J. Chem. Soc., Perkin Trans. 2* 1887–1894 (1985).
27. A. R. Katritzky, P. Barczynski, G. Musumarra, D. Pisano and M. Szafran, *J. Am. Chem. Soc.* **111**, 7–15 (1989).
28. G. Musumarra, S. Wold and S. Gronowitz, *Org. Magn. Reson.* **17**, 118–123 (1981).
29. P. Bouchet, C. Coquelet, J. Elguero, *J. Chem. Soc., Perkin Trans. 2* 449–451 (1974).
30. J. Catalán, J. L. M. Abboud and J. Elguero, *Adv. Heterocycl. Chem.* **41**, 187–274 (1987).
31. H. A. Staab, *Angew. Chem., Int. Ed. Engl.* **1**, 351–367 (1962).
32. (a) W. Otting, *Chem. Ber.* **89**, 1940–1945 (1956); (b) H. W. Thompson and R. J. L. Popplewell, *Z. Elektrochem.* **64**, 746–747 (1960); (c) L. Birkofer and P. Wegner, *Chem. Ber.* **100**, 3485–3494 (1967); (d) J. Elguero and A. Fruchier, unpublished results.
33. (a) G. J. Martin, M. L. Martin and J. P. Gouesnard, *^{15}N-NMR Spectroscopy*. Springer, New York (1981); (b) M. Witanowski, L. Stefaniak and G. A. Webb, *Annu. Rep. NMR Spectrosc.* **18**, 459–469 (1986).
34. J. Elguero, C. Marzin, A. R. Katritzky and P. Linda, *Adv. Heterocycl. Chem.* Suppl. 1, 27 (1976).