

APPLICATION OF CHARACTERISTIC VECTOR ANALYSIS FOR CALCULATIONS OF CONSTANT OF DISSOCIATION FOR SOME INDOLES

Blagoja S. Andonovski*, Bogdan T. Bogdanov, Goran M. Stojković

*Institute of Chemistry, Faculty of Natural Sciences and Mathematics, The "Sv. Kiril & Metodij" University,
P. O. Box 162, MK-1001 Skopje, Republic of Macedonia*

*E-mail: blagand@iunona.pmf.ukim.edu.mk

In the present work we give the results obtained from the method Characteristic Vector Analysis (CVA is a version of the Principal Component Analysis (PCA)) to overcome difficulties in calculating the constant of dissociation for indole, 3-methylindole, 3-formylaldehyde, 3-acetylindole, 2-(3-indolyl) acetic acid, 3-(3-indolyl) propionic acid, 4-(3-indolyl)butyric acid, D-tryptophan and indolyl-2-carboxylic acid. Constant of dissociations have been extensively investigated by ultraviolet spectrophotometry in sulfuric acid solutions. The CVA method was also used to separate the effect of protonation from the generalized medium effect.

Key words: indoles; dissociation constant; Characteristic Vector Analysis; UV spectra

INTRODUCTION

It has been reported that a number of reactions, electrophilic substitution, oxidative rearrangements, hydrations, carbonyl reaction are initiated or influenced by protonation of the indole nucleus [1, 2]. pK_a values are also very useful in the optimization of the spectrophotometric determination of the nitrite in water [3] or can readily be applied to reaction kinetics [4]. Protonation of the some indoles [indole(I), 3-methylindole (3-MI), 3-formylaldehyde (3-FI), 3-acetylindole (3-AI), 2-(3-indolyl) acetic acid (3-IAAH), 3-(3-indolyl) propionic acid (3-IPAH), 4-(3-indolyl)butyric acid (3-IBAH), D-tryptophan (D-TrpH) and indolyl-2-carboxylic acid (2-ICAH)] in different solvents (perchloric acid [5, 6], sulfuric acid [7–16] or phosphoric and hydrochloric acid [17]) was studied by UV spectroscopy in the 400 – 190 nm region. Protonation of all investigated indoles gives the usual large bathochromic shift of the long wavelength band [13–16]. Protonation of all indoles causes a more complex change in the UV spectra. The effect of the medium is not removed by a simple lateral shift. We were able to separate the medium effect from the larger change due to the pro-

tonation, by the analysis of the curve using characteristic vectors. This paper deals with Characteristic Vector Analysis (CVA), which has been described in detail by J. L. Simonds [18] and used by R. L. Reeves [19] and S. C. Wong [20, 21] to investigate problems dealing with protonation. However, superimposed on the large effects of protonation have smaller "medium effect" which prevented the family of any spectral curves from passing through isobestic points. In connection with our pK_a investigation [5–8, 17] we examined the utility of CVA for the treatment of UV spectra.

Method

1) The characteristic vector analysis (sometimes referred to as a "the principal component analysis" or an eigenvector analysis) is a method of separating independent factors for sets of multivariate response data. The method can be used empirically for estimating the number of independent factors contributing to the total variation observed in a family of UV spectra. If p independent factors are involved in generating the absorbance curve,

the sample responses at each wavelength for a given concentration will be given by

$$\begin{aligned} A_1 &= A_1 + c_1 v_{11} + c_2 v_{21} + \dots + c_p v_{p1} \\ A_2 &= A_2 + c_1 v_{12} + c_2 v_{22} + \dots + c_p v_{p2} \\ A_r &= A_r + c_1 v_{1r} + c_2 v_{2r} + \dots + c_p v_{pr} \end{aligned} \quad (1)$$

Where the choice of A is arbitrary and the mean values of the absorbance seem to be a convenient choice. The v are characteristic vectors, and the c are weighting coefficients.

2) The values of $\log c(\text{BH}^+)/c(\text{B})$ for all indole derivatives are plotted against H_i , or H_a and X . Good linear plots were obtained for all indole de-

rivatives. From such slopes m , m' and $\text{p}K_a$ values were obtained fitting the equation of Yates and McClelland [30–35]

$$\log c(\text{BH}^+)/c(\text{B}) = mH_i + \text{p}K_a \quad (2)$$

Also, in Table 4 are listed m and $\text{p}K_a$ values obtained by applying the similar equation

$$\log c(\text{BH}^+)/c(\text{B}) = m' H_A + \text{p}K_a \quad (3)$$

and m^* and $\text{p}K_a$ obtained by applying Cox-equation

$$\log c(\text{BH}^+)/c(\text{B}) - \log c(\text{H}^+) = m^* X + \text{p}K_a \quad (4)$$

EXPERIMENTAL

The UV spectra were subjected to CVA, following a computer program described in Ref. 22, 23 the dissociation ratios and constant were then calculated from reconstituted spectra (Fig. 2 and

3). The spectrophotometric procedure used to determine the dissociation ratio was basically the same as that employed previously [5–8].

RESULTS AND DISCUSSION

Ultraviolet spectra of indoles (I, 3-MI, 3-FI, 3-AI, 3-IAAH, 3-IPAH, 3-IBAH, D-TrpH and 2-ICAH) in solutions of increasing acidity did not show isobestic points, indicating lateral solvent shifts within the range of acid concentration used.

Fig. 1 shows an experimental family of six UV spectra for D-TryH in sulfuric acid solutions. The UV spectrum of D-TryH exhibits four bands which result from $\pi \rightarrow \pi^*$ transitions. The UV spectrum of D-TryH was assigned and interpreted using the FREE ELECTRON MODEL [24–28]. The bands in the wavelength range (260 to 350 nm) are due to the intramolecular charge transfer transition. These transitions are denoted $H_B \rightarrow V_S$ [24]. Fig. 2 shows a family of UV spectra calculated by CVA, the first vector 99.05 % and the second vector 0.29 % is accounting for the total variability. The calculation is in such an order that the first vector accounts for the largest amount of the total response variability, the second vector accounts for the second largest amounts of the variability and so forth. The mathematical analysis shows that any of the experimental UV spectrum can be reconstituted by addition to the mean UV spectrum of Fig. 3, a contribution at each wavelength from the first characteristic vector given in Fig. 3, weighted by the factor c_1 from Table 1, and another contribution of this wavelength from the

second characteristic vector given in Fig. 3, weighted by the factor c_2 from Table 1.

Table 1

Multiples to characteristic vectors v_1 and v_2 for D-TryH in sulfuric acid solutions

RC spectrum	Multiples c_1 to v_1	Multiples c_2 to v_2
1	0.3842824	0.4582980
2	0.3319540	0.0909669
3	0.3090683	-0.3081981
4	0.2583358	-0.1685126
5	0.2233384	0.1512564
6	-0.1644543	-0.5556445
7	-0.1476863	-0.3174946
8	-0.3527012	-0.4453678
9	-0.3883627	0.1545335
10	-0.453769	0.0494277

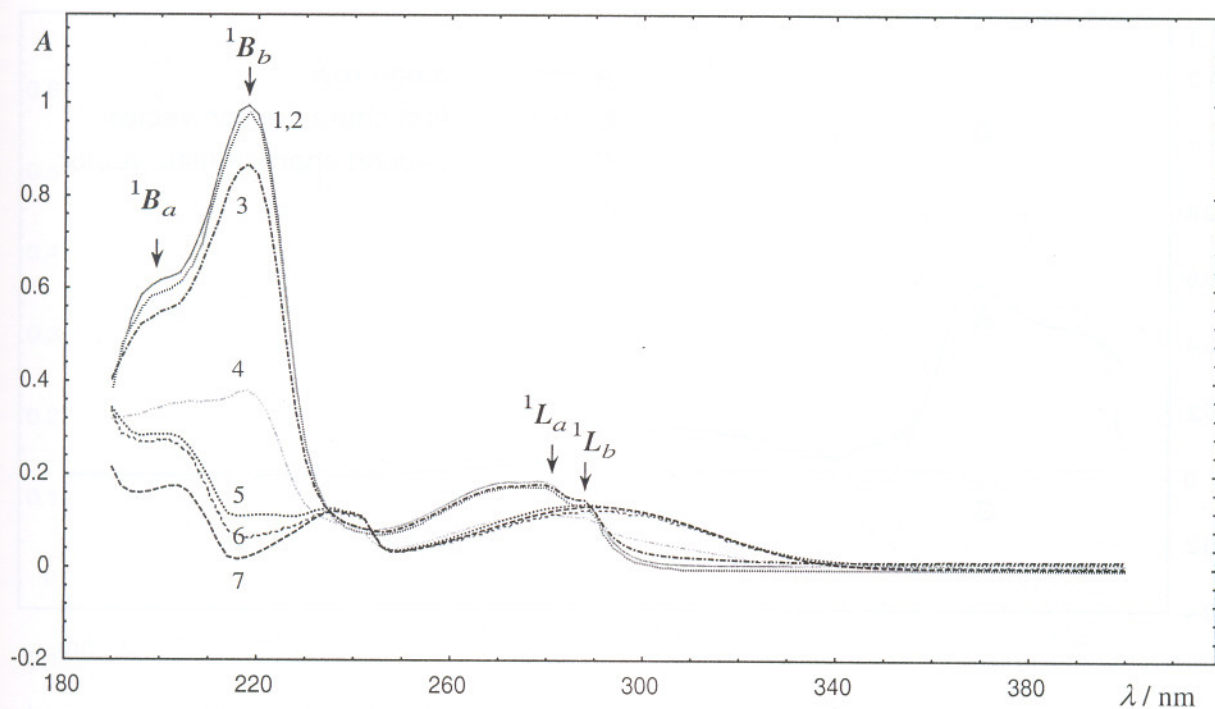


Fig. 1. UV spectra of D-tryptophan ($c = 4.0 \cdot 10^{-5} \text{ mol/dm}^3$) as a function of sulfuric acid concentration (mol/dm^3): 6.0 (1); 7.0 (2); 8.0 (3); 8.5 (4); 9.5 (5); 10.5 (6) and 11.0 (7)

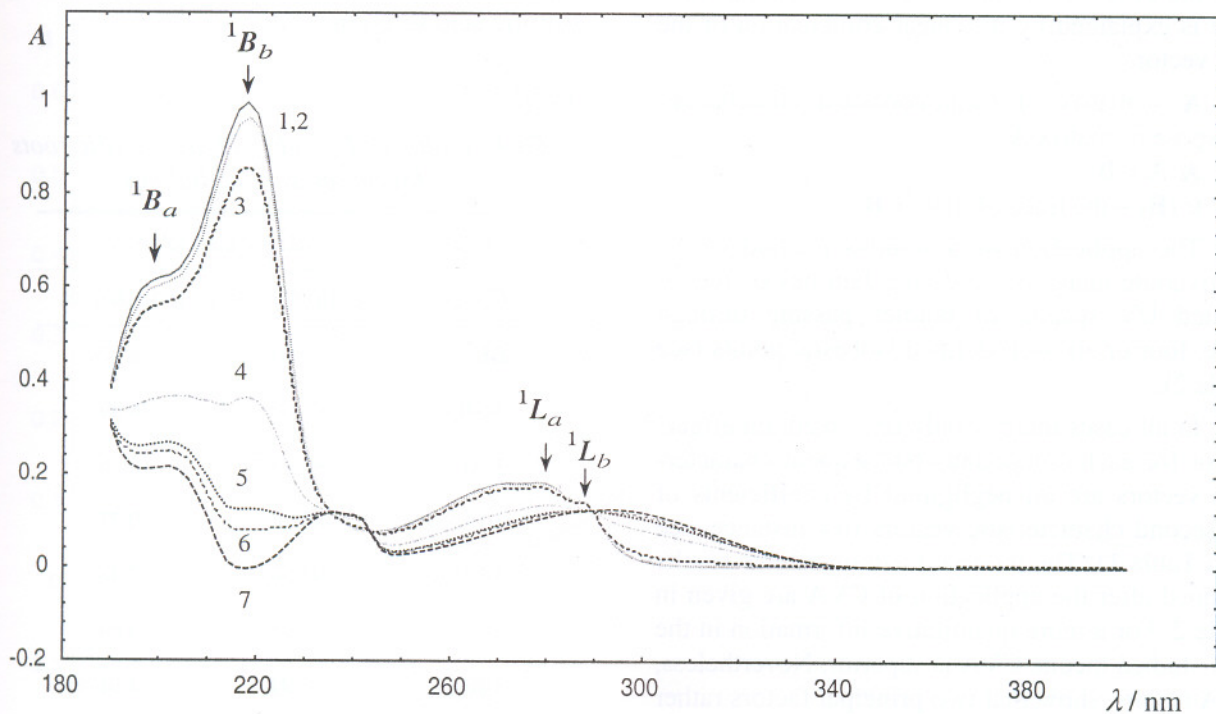


Fig. 2. UV spectra reconstituted from the mean curve and the first characteristic vector of D-tryptophan ($4.0 \cdot 10^{-5} \text{ mol/dm}^3$) as a function of sulfuric acid concentration mol/dm^3 : 6.0 (1); 7.0 (2); 8.0 (3); 8.5 (4); 9.5 (5); 10.5 (6) and 11.0 (7) mol/dm^3

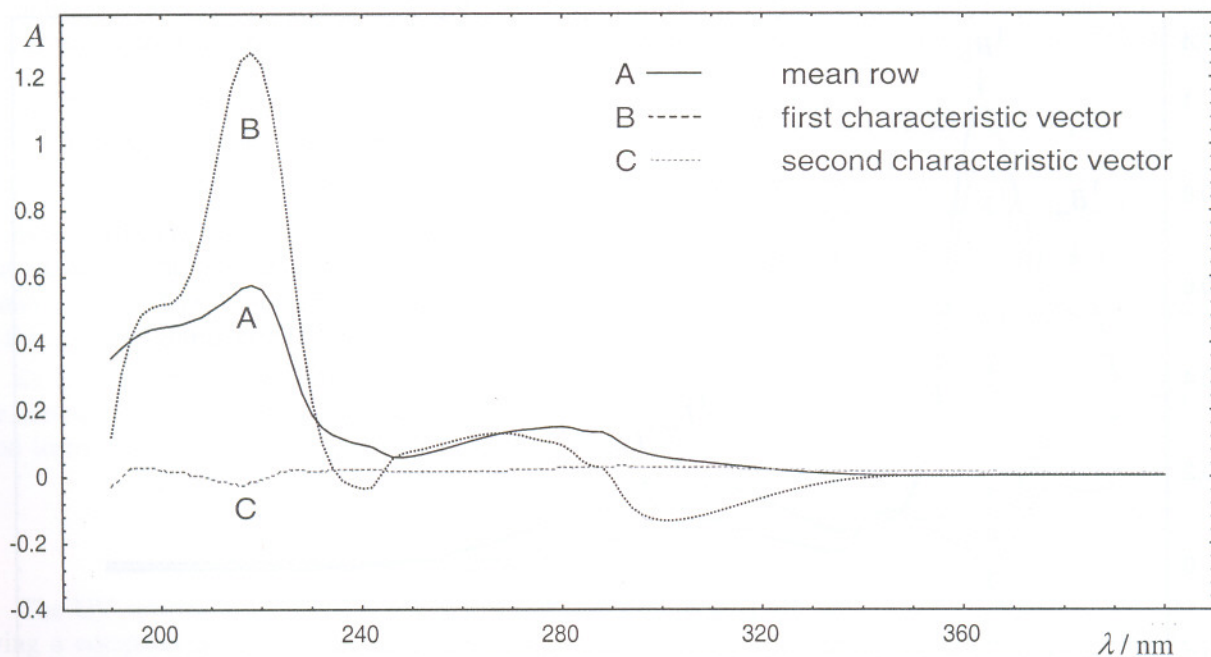


Fig. 3. Mean curve (A) and curves for the first characteristic vector (B) and the second characteristic vector (C) of D-tryptophan

The sum of the diagonal elements of **B** is called the trace of **B**. The trace of **B** is equal to the sum of all the characteristic roots of **B**. The ratio of the sum of the first roots where the trace of **B** is a measure of the percentage of the original data variance is explained by the linear combination of the first vector.

A – matrix of mean-corrected absorbances
transpose of matrix **A**

$$A^t \cdot A = B$$

tr (**B**) – the trace of matrix **B**

The application of Simonds's method of the multivariate analysis yields the families of reconstituted UV spectra of indoles, passing through three, four or six well-defined isobestic points (see Table 2).

In all cases there is only one "medium effect" vector for each compounds (subsequent characteristic vectors are all negligible) by coefficients of the second characteristic vectors (for instance, see c_2 in Table 1). The modern medium effects results obtained after the application of CVA are given in Table 2. For a more quantitative information in the well-studied media, CVA is superior. Nevertheless, CVA results, show that two principal factors rather than one are required to explain variabilities in λ_{\max} or λ_{is} of UV spectra of all investigated indoles (Table 3). The first vector always accounted for 90 % or more of the total variability. The second vector

was usually weaker in the case of the protonation of the O-atom. Small differences in the λ_{\max} or λ_{is} in mineral acid solutions appear at a very high acidity, see Fig. 4 and 5. Fig. 4 shows an experimental family of six UV spectra for 2-ICAH in a sulfuric acid solution.

Table 2

CVA results for L_1 and L_2 characteristic roots for all the studied indoles

Media	Sulfuric acid solutions	
	L_1 [tr(B)] / %	L_2 [tr(B)] / %
IH	89.80	6.47
3-MI	89.05	6.89
3-AI	90.15	8.16
3-IAAH	89.96	9.77
3-IPAH	95.50	2.48
2-ICAH	93.98	4.64
3-IBAH	96.91	1.91
D-TrpH	99.05	0.29
3-FI	88.80	10.27

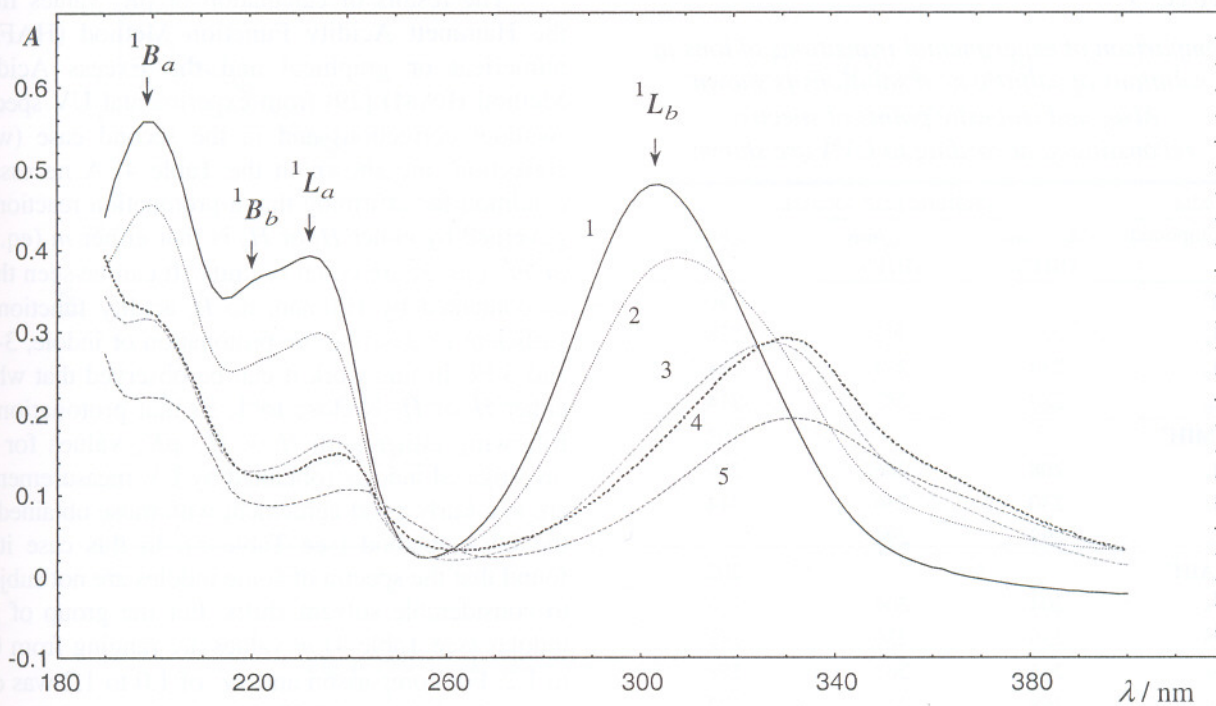


Fig. 4. UV absorption spectra of indolyl-2-carboxylic acid ($4.0 \cdot 10^{-5}$ mol/dm³) as a function of sulfuric acid concentration (mol/dm³): 12.0 (1); 12.5 (2); 14.0 (3); 14.5 (4) and 17.0 (5) mol/dm³

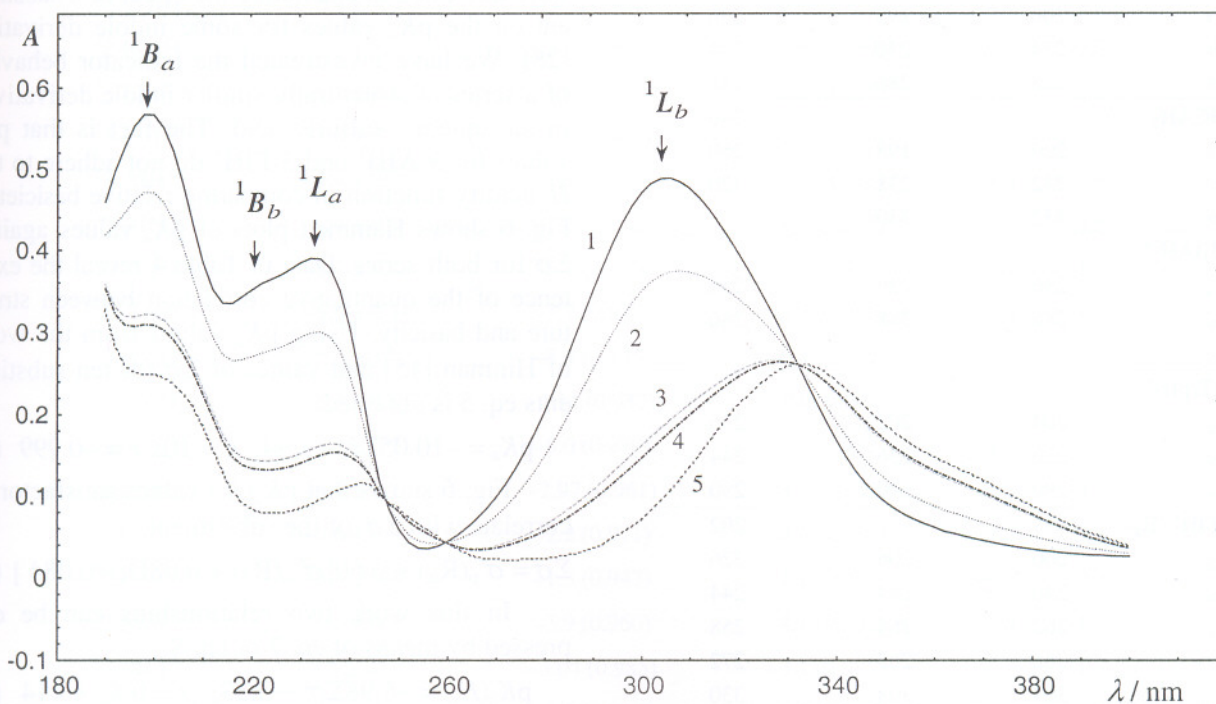


Fig. 5. UV spectra reconstituted from the mean curve and the first characteristic vector of indolyl-2-carboxylic acid ($4.0 \cdot 10^{-5}$ mol/dm³) as a function of sulfuric acid concentration (mol/dm³): 12.0 (1); 12.5 (2); 14.0 (3); 14.5 (4) and 17.0 (5) mol/dm³

Table 3

Comparison of experimental transitions of ions in solutions of sulfuric acid solutions is shown. Also, and isobestic points of spectra reconstituted according to CVA are shown

Media	Sulfuric acid solutions		
Compound	$^1\lambda_{\max}/\text{nm}$ (BH ⁺)	$^2\lambda_{\max}/\text{nm}$ (BH ⁺)	$\lambda_{\text{is}}/\text{nm}$
IH ⁺			196
¹ B _a	202	202	238
¹ B _b	234	234	288
¹ L _b	280	280	314
3-MIH ⁺			192
¹ B _a	208	206	286
¹ B _b	230	236	314
¹ L _b	292	294	
3-AIH ⁺			202
¹ B _a	204	204	226
¹ B _b	236	236	242
¹ L _a	262	262	258
¹ L _b	268	268	272
	344	340	316
3-IAAH ₂ ⁺			196
¹ B _a	236	236	222
¹ B _b	244(sh)	248(sh)	304
¹ L _a	288;	288	
¹ L _b	312	320	
3-IPAH ₂ ⁺			194
¹ B _a	200	198	236
¹ B _b	234	230	244
¹ L _b	288	286	288
2-ICAH ₂ ⁺			246
¹ B _a	200	198	260
¹ B _b	242	238	320
¹ L _b	332	330	
3-IBAH ₂ ⁺			236
¹ B _a	236	236	236
¹ B _b	286	288	286
¹ L _b			
D-TrpH ₂ ⁺			236
¹ B _a	204	202	236
¹ B _b	236	236	244
¹ L _b	290	290	290
3-FIH ⁺ ¹ B _a			202
¹ B _b	206	206	226
¹ L _b	246	244	244
¹ L _a	262	264	258
	266	268	272
	304	308	330

¹Uncorrected for the medium shift

²Absorption maximum at the wavelength of curves reconstituted from the mean curve and the first vector

The results of calculation of pK_a values from the Hammett Acidity Function Method (HAFM) numerical or graphical and the Excess Acidity Method (EAM) [29] from experimental UV spectra (without correction) and in the second case (with correction) are shown in the Table 4. A necessary condition for affirming that a protonation reaction is governed by either H_i or H_a is that either m (eq. 2) or m' (eq. 3) are equal to unity. It can be seen that, as contented by Hinman, no H_i acidity function is sufficient to describe the protonation of indole, 3-AI and 3-FI. In this work it can be observed that when either H_i or H_a is close to 1, so that protonation is following either with H_i or H_a , pK_a values for all investigated indoles (obtained by UV measurements) are in a fairly good agreement with those obtained in the CVA method (see Table 3). In this case it is found that the spectra of some indoles are not subject to considerable solvent shifts. For the group of six indoles (see Table 4) m values are ranging from 0.8 to 1.2. For comparison and m' of 1.0 to 1.2 was obtained for 3-AI and 3-FI. The deviation of these m values might be the result of different indicated structures. The effects of the medium on the UV spectra of these indoles appear to be small. A comparison of the pK_a 's obtained from experimental UV spectra and with the correction of all UV spectra reveals the satisfactory validity of the CVA method. We were interested in examining the effect of a substituent on the pK_a values for some indole derivative [28]. We have investigated the indicator behavior of a series of structurally similar indole derivatives in an aqueous sulfuric acid. The fact is that pK_a values for 3-AIH⁺ and 3-FIH⁺ do not adhere to the H_i acidity function in comparing relative basicities. Fig. 6 shows Hammett plots of pK_a values against $\Sigma\sigma$ for both series. Data in Table 4 reveal the existence of the quantitative correlation between structure and basicity. Using pK_a values from the work of Hinman [15] and values of $\Sigma\sigma$ for ten substituents eq. 5 is obtained:

$$pK_a = -10.05\Sigma\sigma - 3.94; \quad n = 10; \quad r = -0.999 \quad (5)$$

Fig. 6 shows that $pK_a(H_i)$ values satisfactorily correlate with $\Sigma\sigma$ of the substituent.

$$\Sigma\sigma = \sigma_p^+(R_\alpha) + 0.60[\sigma_m^+(R_\beta) + 0.08D_s - 0.084] \quad (6)$$

In this work two relationships can be expressed by means of eq. 7 or eq. 8.

$$pK_a(H_i) = -6.98\Sigma\sigma - 3.39; \quad r = 0.8; \quad n = 14 \quad (7)$$

From the above correlations an interesting pointing may be brought. By excluding pK_a values for 3-AIH⁺ and 3-FIH⁺ eq. 8 is obtained.

$$pK_a(H_i) = -9.37\Sigma\sigma - 3.86; \quad r = 0.999; \quad n = 12 \quad (8)$$

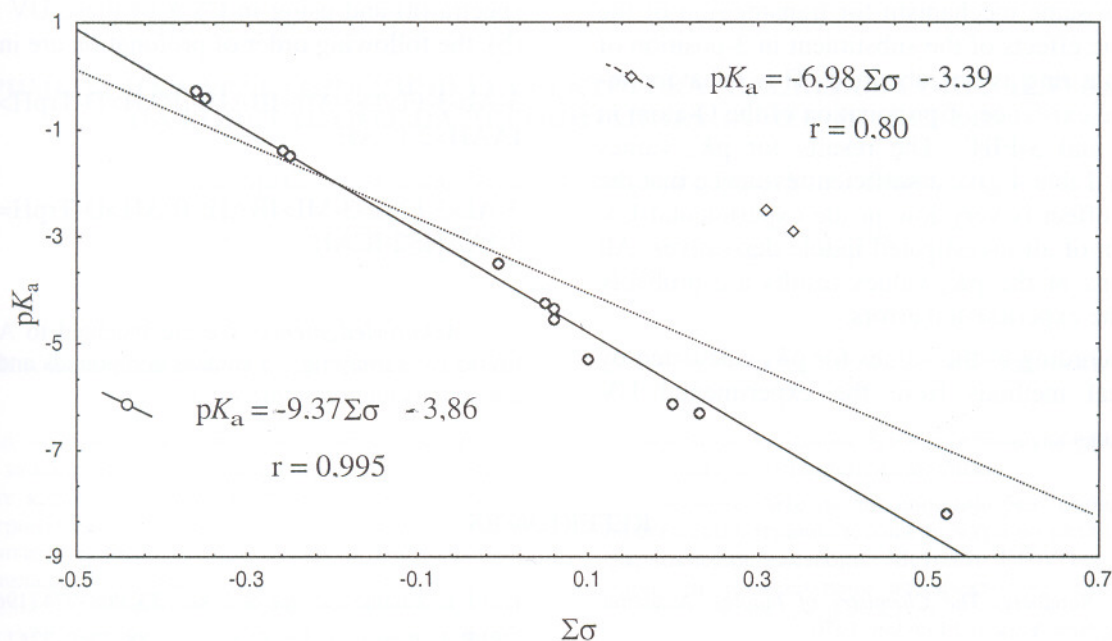


Fig. 6. Plots of pK_a values for all indole derivatives against $\Sigma\sigma$

Table 4

pK_a values for some indoles (in sulfuric acid solutions) obtained by EAM and HAFM (found from experimental UV spectra and using in CVA of these UV spectra)

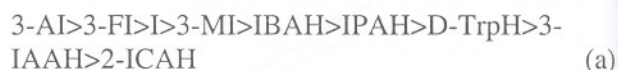
Ions	Numerically (HAFM)	Graphically (EAM)	Graphically (HAFM)	m	m^*
¹ 2-ICAH	-8.4 ± 0.2	-5.5 (0.997)	-6.8 (0.995)	-0.8	0.8
² 2-ICAH	-8.8 ± 0.2	-5.1 (0.994)	-6.9 (0.999)	-0.8	0.7
¹ 3-FIH ⁺ a	-1.7 ± 0.1	-1.0 (0.986)	-2.2 (0.985)	-1.2	1.2
² 3-FIH ⁺ a	-1.8 ± 0.1	-1.6 (0.988)	-2.3 (0.989)	-1.3	0.7
¹ 3-AIH ⁺ a	-1.4 ± 0.2	-1.3 (0.997)	-2.1 (0.996)	-1.4	0.8
² 3-AIH ⁺ a	-1.4 ± 0.1	-1.2 (0.972)	-1.4 (0.982)	-1.0	0.5
¹ D-TrpH ₂ ⁺	-5.4 ± 0.1	-5.3 (0.971)	-6.1 (0.971)	-1.1	1.5
² D-TrpH ₂ ⁺	-5.6 ± 0.1	-3.7 (0.977)	-3.3 (0.975)	-0.6	1.0
¹ 3-IPAH ₂ ⁺	-5.3 ± 0.2	-5.4 (0.995)	-6.4 (0.972)	-1.2	1.7
² 3-IPAH ₂ ⁺	-5.5 ± 0.1	-4.14 (0.932)	-4.45 (0.963)	-0.8	1.0
¹ IBAH ₂ ⁺	-5.1 ± 0.2	-5.0 (0.979)	-5.4 (0.972)	-1.0	1.5
² IBAH ₂ ⁺	-4.7 ± 0.1	-2.97 (0.961)	-3.9 (0.907)	-0.8	0.9
¹ IH ⁺ b	-3.6 ± 0.2	-2.2 (0.971)	-4.4 (0.981)	-1.2	0.8
² IH ⁺ b	-3.5 ± 0.4	-1.3 (0.947)	-1.1 (0.958)	-0.3	0.3
¹ 3-MIH ⁺	-4.6 ± 0.2	-3.9 (0.960)	-4.1 (0.962)	-0.9	1.3
² 3-MIH ⁺	-3.6 ± 0.05	-3.04 (0.983)	-3.1 (0.987)	-0.9	1.3
¹ 3-IAAH ₂ ⁺	-5.6 ± 0.2	-5.3 (0.979)	-6.5 (0.992)	-1.1	1.5
² 3-IAAH ₂ ⁺	-5.5 ± 0.3	-3.1 (0.916)	-3.1 (0.949)	-0.6	0.7

^a pK_a values obtained by the H_a acidity function. ^b pK_a values obtained by the H_i acidity function

The same mechanism for transmission of the electronic effects of the substituent in 3-position of the indole ring is not obtained. This behavior reflects the existence of protonation of the O-atom in 3-AIH⁺ and 3-FIH⁺. The results for pK_a values listed in Table 4 give a sufficient evidence that the solvent effect is very low in the experimental UV spectrum of all investigated indole derivatives. All deviations of the pK_a values results are probably due to the experimental errors.

According to the values for pK_a calculated by numerical methods from the experimental UV

spectra (a) and using in CVA of these UV spectra (b), the following order of protonation are inferred.



Acknowledgments. We are indebted to A. Nikolvski for supplying the studied compounds and for his continuous interest in the work.

REFERENCES

- [1] R. J. Sundberg, *The Chemistry of Indoles*, Academic Press, New York and London, 1970.
- [2] T. H. Lowry and K. S. Richards, *Mechanism and Theory in Organic Chemistry*, Harper and Row, New York, 1981, pp. 232–242.
- [3] A. Chauba, A. N. Baveja and V. K. Gupta, *Anal. Chim. Acta*, **143**, 273–276 (1982).
- [4] R. A. Cox, K. Yates, *Can. J. Chem.*, **57**, 2944–2951 (1979).
- [5] B. Andonovski, I. Spirevska, A. Nikolovski, *Croat. Chem. Acta*, **96**, 1201–1213 (1996).
- [6] B. Andonovski, *Croat. Chem. Acta*, **72**, 4, 711–726 (1999).
- [7] B. Andonovski, L. Šoptrajanova, I. Spirevska, *Bull. Chem. Technol. Macedonia* **17**, 2, 97–104 (1998).
- [8] B. Andonovski, G. Stojković, *J. Serb. Chem. Soc.*, **64**, 12, 775–787 (1999).
- [9] H. Podkovinska, G. Novotarska, B. Barcizevska, *Zesz. Nauk. Acad. Ekon. Poznań, Ser-I*, **58**, 119–125 (1974); in *CA* **83**, 192107y (1975).
- [10] G. Novotarska and H. Podkovinska, *Rocz. Chem.*, **50**, 789–793 (1976).
- [11] G. Novotarska, *Symp. Pap.-IUPAC Int. Symp. Chem. Nat. Prod.*, **2**, 293–295 (1978); in *CA* **91**, 192402f (1979).
- [12] G. Novotarska, *Zesz. Nauk. Acad. Ekon. Poznań, Ser-I*, **80**, 128–132 (1978); in *CA* **94**, 64903v (1981).
- [13] R. L. Hinman, E. B. Whipple, *J. Am. Chem. Soc.*, **84**, 2534–2539 (1962).
- [14] G. Berti, A. da Settimo, D. Segnini, *Gazz. Chim. Ital.*, **91**, 571–586 (1961).
- [15] R. L. Hinman, J. Lang, *J. Am. Chem. Soc.*, **86**, 3796–3806 (1964).
- [16] R. L. Hinman, J. Lang, *Tetrahedron Lett.*, **21**, 12–15 (1960).
- [17] B. Andonovski, G. Stojković, *Acta Chem. Slov.*, **47**, 349–362 (2000).
- [18] J. L. Simonds, *J. Opt. Soc. Am.*, **53**, 968–974 (1963).
- [19] R. L. Reeves, *J. Am. Chem. Soc.*, **88**, 2240–2247 (1966).
- [20] J. T. Edward, S. C. Wong, *J. Am. Chem. Soc.*, **99**, 4229–4232 (1977).
- [21] S. C. Wong, *Ph.D. Thesis*, McGill University, Montreal, Canada, 1974.
- [22] J. T. Edward, *Ph. D. Thesis*, McGill University, Montreal, Canada 1976.
- [23] B. Andonovski, *Ph. D. Thesis*, The “Sv. Kiril & Metodij” University, Skopje, Macedonia, 1999.
- [24] S. Nagakura and J. Tanaka, *J. Chem. Phys.*, **22**, 236–240 (1954).
- [25] H. H. Jafe, M. Orchin, *Theory and Applications of Ultraviolet Spectroscopy*, J. Wiley and Sons, Inc. New York, 1962, p. 294.
- [26] P. S. Song, W. E. Kurtin, *J. Am. Chem. Soc.*, **91**, 4892–4906 (1969).
- [27] H. E. Auer, *J. Am. Chem. Soc.*, **95**, 3003–3011 (1973).
- [28] F. Tomas-Vert, C. A. Ponce, M. R. Estrada, J. Silber, J. Singh, J. Anunciatta, *J. Mol. Struct.*, **246**, 203–215 (1991).
- [29] R. A. Cox, *Advances Physical Organic Chemistry*, **35**, 1–66 (2000).
- [30] K. Yates, J. B. Stevens and A. R. Katritzky, *Can. J. Chem.*, **42**, 1957–1970 (1964).
- [31] R. A. Cox, M. F. Goldman, K. Yates, *Can. J. Chem.*, **57**, 2961–2966 (1979).
- [32] R. A. Cox, K. Yates, *J. Am. Chem. Soc.*, **100**, 3861–3867 (1978).
- [33] N. C. Marziano, G. M. Cimino, R. C. Passerini, *J. Chem. Soc. Perkin Trans.*, **2**, 1915–1920 (1973).
- [34] L. P. Hammett, A. J. Deyrup, *J. Am. Chem. Soc.*, **54**, 2721–2734 (1932).
- [35] P. A. Cohen and L. A. Cohen, *Can. J. Chem.*, **70**, 282–284 (1992).

Резиме

ПРИМЕНА НА КАРАКТЕРИСТИЧНАТА ВЕКТОРСКА АНАЛИЗА ЗА ОПРЕДЕЛУВАЊЕ НА КОНСТАНТИТЕ НА ДИСОЦИЈАЦИЈА ЗА НЕКОИ ИНДОЛИ

Благоја С. Андоновски*, Богдан Т. Богданов, Горан М. Стојковиќ

*Институт за хемија, ПМФ, Универзитет „Св. Кирил и Методиј“,
б. бр. 162, МК-1001 Скопје, Република Македонија*

Клучни зборови: индоли; константи на дисоцијација; карактеристична векторска анализа; УВ-спектри

Во презентираната работа се изложени резултатите за константите на дисоцијација определени со методот карактеристична векторска анализа (КВА) за отстранување на грешките во пресметувањето на константите на дисоцијација за индол, 3-метилиндол, 3-формилалдехид, 3-ацетилиндол, 2-(3-индолил)оцетна киселина, 3-(3-индолил)пропионска киселина, 4-(3-индолил)бутерна киселина, D-триптофан и индолил-2-

карбоксилна киселина. КВА претставува една верзија на основната компонентна анализа.

Константите на дисоцијација беа определувани од снимени ултравиолетови спектри во различни сулфурнокисели раствори. Методот КВА исто така е користен за одделување на ефектот на протонирање од ефектите на растворувачот.