VERSITA

Cent. Eur. J. Chem. • 8(3) • 2010 • 494–507
DOI: 10.2478/s11532-010-0023-1

Central European Journal of **Chemistry**

The thermodynamic dissociation constants of methotrexate by the nonlinear regression and factor analysis of multiwavelength spectrophotometric pH-titration data

Invited Paper

Milan Meloun[1]*, Zuzana Ferenčíková[1] and Aleš Vrána[2]

[1]*Department of Analytical Chemistry, University of Pardubice CZ 532 10 Pardubice, Czech Republic*

[2]*IVAX Pharmaceuticals, s.r.o., CZ 747 70 Opava, Czech Republic*

Received 30 December 2009; Accepted 17 February 2010

Abstract: The mixed dissociation constants of *methotrexate* – chemically (2S)-2-[(4-{[(2,4-diamino-7,8-dihydropteridin-6-yl)methyl](methyl)amino}phenyl)formamido]pentanedioic acid (the cas number 59-05-2) at various ionic strengths I of range 0.01 - 0.4, and at temperatures of 25ºC and 37ºC, were determined with the use of two different multiwavelength and multivariate treatments of spectral data, SPECFIT32 and SQUAD(84) nonlinear regression analyses and INDICES factor analysis according to a general rule of first,determining the number of components, and then calculating the spectral responses and concentrations of the components. Concurrently, the experimental determination of the thermodynamic dissociation constants was in agreement with its computational prediction of the PALLAS programme based on knowledge of the chemical structures of the drug. The factor analysis in the INDICES programme predicts the correct number of light-absorbing components when the data quality is high and the instrumental error is known. Three thermodynamic dissociation constants were estimated by nonlinear regression of $\{pK_a, I\}$ data: for methotrexate $pK_{a1}^T = 2.895(13)$, $pK_{a2}^T = 4.410(14)$, $pK_{a3}^T = 5.726(15)$ at 25ºC and $pK_{a1}^T = 3.089(15)$, $pK_{a2}^T = 4.392(12)$, $pK_{a3}^T = 5.585(11)$ at 37ºC, where the figure in brackets is the standard deviation in last significant digits. The reliability of the dissociation constants of the drug were proven by conducting goodness-of-fit tests of the multiwavelength spectrophotometric pH-titration data.

Keywords: *Spectrophotometric titration • Dissociation constant • Methotrexate • Specfit • Squad*

1. Introduction

Methotrexate (MTX, *4-amino-N10-methylpteroyl glutamic acid, Amethopterin*) (Fig. 1) is a folic acid antagonist, and acts as an inhibitor of dihydrofolate reductase, leading to DNA damage and cell death [1]. Intravenous administration of MTX at a high dose is widely used as chemotherapy in humans for various neoplastic diseases such as acute leukemia, osteogenic sarcoma [2,3], non-Hodgkin lymphoma and breast carcinoma [4]. Recently, however, lower doses of MTX have been used in humans to treat rheumatoid arthritis [5-7]. MTX usage in racehorses has been reported most likely due to the reported effectiveness of treating rheumatoid arthritis in humans. Most equine lameness conditions are not due to rheumatoid arthritis, thus with the existence of regulatory limits on select non-steroidal anti-inflammatory medications, the use of MTX to manage lameness in race horses is without merit.

Methotrexate (MTX) is practically insoluble in all water immiscible solvents; therefore solid phase extraction is a more efficient choice for MTX enrichment

* *E-mail: milan.meloun@upce.cz*

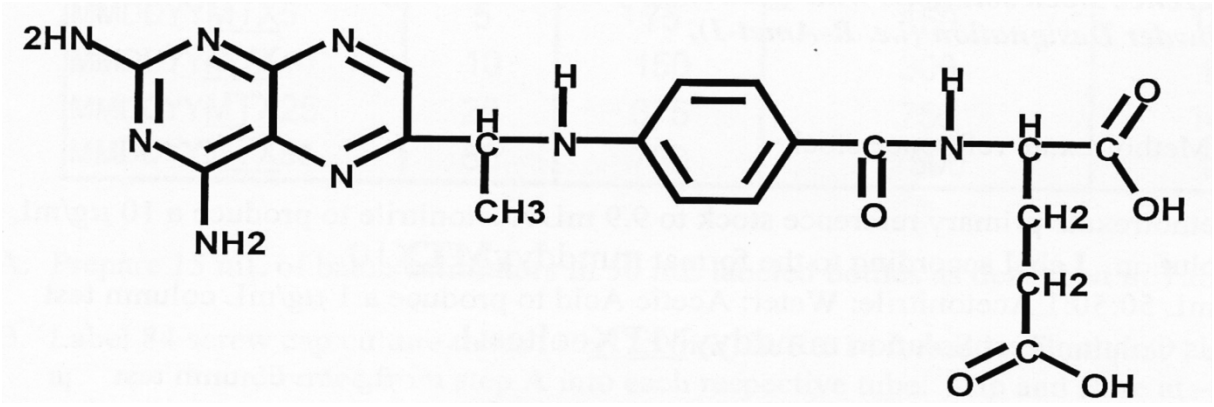

Figure 1. Methotrexate.

from aqueous biofluids [6]. Polyglutamates and 7-hydroxy-methotrexate have been reported [1,2,8-10] as detectable metabolites of MTX; the method focuses on the parent methotrexate in equine urine as the target analyte.

Methotrexate is a very polar, poly-functional compound, closely related to folic acid as it is a methylated derivative of the folic acid. This polar nature precludes extraction by conventional liquid-liquid extraction procedures. Methotrexate is a weak dicarboxylic acid with pK_a approx. 4.8 and 5.5, and thus it is expected to mostly ionize at physiologic pH [6]. However, ionized forms of drugs - in general - can barely cross the (lipophilic) cell membrane bilayer.

In previous work [11-19] the authors have shown that the spectrophotometric method in combination with suitable chemometric tools can be used for the determination of protonation constants β_r or acid dissociation constants K_a even for barely soluble drugs. Spectrophotometry is a convenient method for K_a determination in very diluted aqueous solutions (about 10^{-5} to 10^{-6} M), provided that the compound possesses pH-dependent light absorption due to the presence of a chromophore in proximity to the ionisation centre *cf.* [20-34]. There are many cases where the spectral responses of two and sometimes even more components overlap considerably, and the analysis is not straightforward. Problems arise because of strong overlapping chemical components involved in the equilibrium, and uncertainties which arise from the mathematical algorithms used to solve these problems. In such cases, much more information can be extracted if multivariate and multiwavelength spectrophotometric data are analyzed by means of an appropriate multivariate data analysis method. The most relevant algorithms are the hard-modeling programme SQUAD [23-28] and the soft-modeling programme SPECFIT [31-33, 35].

In this study, we have attempted to complete the information on the protonation/dissociation constants for methotrexate. Concurrently, the experimental determination of protonation constants was combined with its computational prediction based on a knowledge of chemical structures [36].

2. Theory

2.1 Procedure for the determination of the mixed protonation/dissociation constants

An acid-base equilibrium of the drug studied is described in terms of the thermodynamic protonation of the Brönstedt base L^{z-1} according to the equation $L^{z-1} + H^+ \rightleftharpoons HL^z$ characterized by the protonation constant

$$K_H = \frac{a_{HL^z}}{a_L^{z-1} a_{H^+}} = \frac{[HL^z]}{[L^{z-1}][H^+]} \frac{y_{HL^z}}{y_{L^{z-1}} y_{H^+}} \qquad (1)$$

The protonation equilibria between the anion L (the charges are omitted for the sake of simplicity) of a drug and a proton H are considered to form a set of variously protonated species L, HL, H_2L, and H_3L ...*etc.*, which have the general formula H_rL in a particular chemical model and which are represented by n_c the number of species, r_i, i = 1, ..., n_c where index i labels their particular stoicheiometry; the overall protonation (stability) constant of the protonated species, β_r, may then be expressed as

$$\beta_r = [H_rL] / ([L]\ [H]^r) = c / (l\ h^r) \qquad (2)$$

where the free concentration [L] = l, [H] = h and [H_rL] = c. For dissociation reactions realized at constant ionic strength the so-called „mixed dissociation constants“ are defined as

$$K_{a,j} = \frac{[\mathrm{H_{j-1}L}]\, a_{\mathrm{H^+}}}{[\mathrm{H_jL}]} \quad (3)$$

Each aqueous species is characterized by its own spectrum. UV/VIS experiments and the *i*th solution are measured at the *j*th wavelength. The Lambert-Beer law relates the absorbance, $A_{i,j}$, which isg defined as

$$A_{i,j} = \sum_{n=1}^{n_c} \varepsilon_{j,n}\, c_n = \sum_{n=1}^{p} (\varepsilon_{r,j}\, b_r\, l\, h^r)_n \quad (4)$$

where $\varepsilon_{r,j}$ is the molar absorptivity of the H_rL species with the stoichiometric coefficient *r* measured at the *j*th wavelength. The absorbance $A_{i,j}$ is an element of the absorbance matrix ***A*** of size (n_s *H* n_w) which is measured for n_s solutions with known total concentrations of n_z = 2 basic components, c_L and c_H, at n_w wavelengths. Calculations related to the determination of protonation constants may be performed by a regression analysis of spectra using versions of the SQUAD(84) programme family [23,28] and SPECFIT/32 [31–33,35] and which have been described previously [37-39].

2.2 Determination of the thermodynamic protonation/dissociation constant

Let us consider the dependence of the mixed dissociation constant $K_a = a_{H+}\ [L^{z-1}]/[HL^z]$ on an ionic strength, in which both ions HL^z and L^{z-1} have roughly the same ion-size parameter *a* in the dissociation equilibrium $HL^z \rightleftarrows L^{z-1} + H^+$ with the thermodynamic dissociation constant $K_a^T = a_{H+}\ a_L/a_{HL}$,; and suppose the overall salting-out coefficients is given by $C = C_{HL} - C_L$. This dependence is expressed by the Debye-Hückel equation as

$$\mathrm{p}K_a = \mathrm{p}K_a^T - \frac{A\,(1-2\,z)\,\sqrt{I}}{1 + B\,\mathring{a}\,\sqrt{I}} + C\,I \quad (5)$$

where A = 0.5112 $mole^{-1/2}\ L^{1/2}\ K^{3/2}$ and B = 0.3291 $mole^{-1/2}\ m^{-1}\ L^{1/2}\ K^{1/2}10^{10}$ for aqueous solutions at 25°C. The mixed dissociation constant pK_a represents a dependent variable while the ionic strength *I* stands for the independent variable. Three unknown parameters ***b*** = {pK_a, *å*, *C*} are to be estimated by a minimization of the sum of the squared residuals [40]

$$U(\boldsymbol{b}) = \sum_{i=1}^{n} w_i \left[\mathrm{p}K_{a,\mathrm{exp},i} - \mathrm{p}K_{a,\mathrm{calc},i}\right]^2 = \sum_{i=1}^{n} w_i \left[\mathrm{p}K_{a,\mathrm{exp},i} - f(I; \mathrm{p}K_a^T, \mathring{a}, C)\right]^2 = \mathrm{minimum} \quad (6)$$

The nonlinear estimation problem is simply a problem of optimization in the parameter space, in which the pK_a and *I* are known and given values. The parameters $\mathrm{p}K_a^T$, *å*, and *C* are unknown variables that need to be estimated [37-43].

2.3 Reliability of the estimated dissociation constants

The adequacy of a proposed regression model with experimental data and the reliability of a parameter estimates $pK_{a,i}$ found, being denoted for the sake of simplicity as unknown parameters b_j, and ε_{ij}, j = 1, ..., m, may be examined by the goodness-of-fit test, *cf.* a previous tutorial [37,39].

2.4 Determination of the number of light-absorbing species

A qualitative interpretation of the spectra aims (a) to evaluate the quality of the dataset and remove spurious data, and (b) to estimate the minimum number of factors, *i.e.*, contributing aqueous species, which are necessary to describe the experimental data. The INDICES [44] determine the number of dominant light-absorbing species present in the equilibrium mixture. The various indicator function *PC(k)* techniques in the INDICES programme were developed to deduce the exact size of the true component space; these can be classified into two general categories and were described in detail previously [11,15]: (a) precise methods based upon knowledge of the experimental error of the absorbance data, $s_{inst}(A)$, and (b) approximate methods requiring no knowledge of the experimental error. In general, most precise and approximate methods are based on the procedure to find the point where the slope of the indicator function *PC(k) = f(k)* changes.

2.4.1. Precise indices

The determination of a number of light-absorbing components in a mixture is based on a comparison of an actual index of a method used with the experimental error of the instrument used, $s_{inst}(A)$, [29,45]:

(a) *Kankare's residual standard deviation,* $s_k(A)$. The $s_k(A)$ values for different numbers of components *k* are plotted against an index *k*, $s_k(A) = f(k)$, and the number of significant components is an integer $n_c = k$ for which $s_k(A)$ is close to the instrumental error of absorbance $s_{inst}(A)$ [11,29].

(b) *Residual standard deviation, RSD(k),* is used analogously as in the previous method $s_k(A)$.

(c) *Root mean square error, RMS:* analogically as in the previous method is based on finding the point where the slope of the indicator function changes.

(d) *Average error criterion, AE(k),* is used analogously as in the preceding method $s_k(A)$.

(e) *Bartlett* χ^2 *criterion,* $\chi^2(k)$ is used when the true

number of significant components corresponds to the first *k* value for which $\chi^2(k)$ is less than critical $\chi^2(k)_{expected} = (n - k)(m - k)$.

(f) *Standard deviation of eigenvalues s(g)* is based on finding the point where the slope of the indicator function changes.

(g) *Eigenvalues:* the first *k* eigenvalues, called a set of *primary eigenvalues,* contribute from the real components and should be considerably larger than those containing only noise.

2.4.2. Approximate methods

Most of the techniques presented here are empirical functions:

(h) *Exner function ψ(k):* The $\psi(k) = (k)$ function can vary from zero to infinity, with the best fit approaching zero.

(i) *Scree test, RPV(k):* When the residual percent variance is plotted against the number of *k* PC dimensions used in the data reproduction, $RPV(k) = f(k)$, the curve should drop rapidly and level off at some point.

(j) *Imbedded error function, IE(k):* The imbedded error function *IE*(k) is an empirical function developed to identify those *k* latent variables which contain error without relying upon an estimate of the error associated with the absorbance data matrix.

(k) *Factor indicator function, IND(k):* The factor indicator function *IND(k)* is an empirical function which reaches a minimum when the correct number of latent variables or *k* PC dimensions is employed in the data reproduction.

(l) *Ratio of eigenvalues calculated by smoothed PCA and those by ordinary PCA, RESO(k):* The index $RESO_i^k$ or the ratios between $\frac{\lambda^s_{x,i}}{\lambda^o_i}$ for different *k* and plot log ($RESO_i^k$) versus component number is calculated. The number of components is estimated by examining the log ($RESO_i^k$) versus component number plots. The number of log($RESO_i^k$)s are located which are very close to each other and do not change substantially with the variation of *k* compared to the remaining log($RESO_i^k$)s . This is the number of components which exist in the examined mixture.

2.5 Signal-to-noise ratio SER

The level of "experimental noise" should be used as a critical factor in the experiment. Therefore, it is necessary to have a consistent definition of the *signal-to-noise ratio SNR* so that the impact of this parameter can be critically assessed. Traditional approaches to *SNR* are typically based on the ratio of the maximum signal to the maximum noise value. As an alternative, the concept of instrumental error was employed and the *signal-to-error ratio SER* was defined using the instrumental standard deviation of absorbance, $s_{inst}(A)$for an error . The plot of small absorbance changes in the spectrum of the drug studied means that the value of the absorbance difference for the *j*th-wavelength of the *i*th-spectrum $\Delta_{ij} = A_{ij} - A_{i,acid}$ is divided by the instrumental standard deviation $s_{inst}(A)$, and the resulting ratios $SER = \Delta/s_{inst}(A)$; these are plotted in dependence of wavelength λ for all absorbance matrix elements, where $A_{i,acid}$ is the initial spectrum of the acid form of the drug and measured for the starting pH value of the pH range which is being studied. This *SER* ratio is then compared with the limiting *SER* value to determine if the absorbance changes are significantly larger than the instrumental noise.

The plot of the ratio $e/s_{inst}(A)$, *i.e.*, the ratio of the residuals divided by the instrumental standard deviation $s_{inst}(A)$, is dependent on wavelength λ for all the residual matrix elements for tests if the residuals are of the same or similar magnitude as is the instrumental noise to prove the best curve fitting achieved.

3. Experimental Procedure

3.1 Chemicals and solutions

Hydrochlorid acid, 1M, was prepared from concentrated HCl (*p.a.*, Lachema Brno) using redistilled water and standardized against HgO and KI with reproducibility of less than 0.20%. *Potassium hydroxide*, 1 M, was prepared from pellets (*p.a.*, Aldrich Chemical Company) with carbondioxide-free redistilled water and standardized against standardized HCl with a reproducibility of 0.1%. The preparation of other solutions from analytical reagent-grade chemicals has been described previously [37-39, 41-43]. *Methotrexate* 5×10^{-5} M, was prepared from solid samples (IVAX *a.s.*, Opava) using redistilled water. The supplier guaranteed high purity of the substances (over 98%).

3.2 Apparatus and pH-spectrophotometric titration procedure

The apparatus used, and the pH-spectrophotometric titration procedure have been described in detail previously [37-39]. The experimental and computation scheme for the determination of the protonation constants of the multicomponent system is taken from Meloun *et al., cf.* page 226 in [20] and the five steps are described in detail elsewhere [37]. (1) instrumental error of absorbance measurements, $s_{inst}(A)$, (2) experimental design, (3) number of light-absorbing species, (4) choice

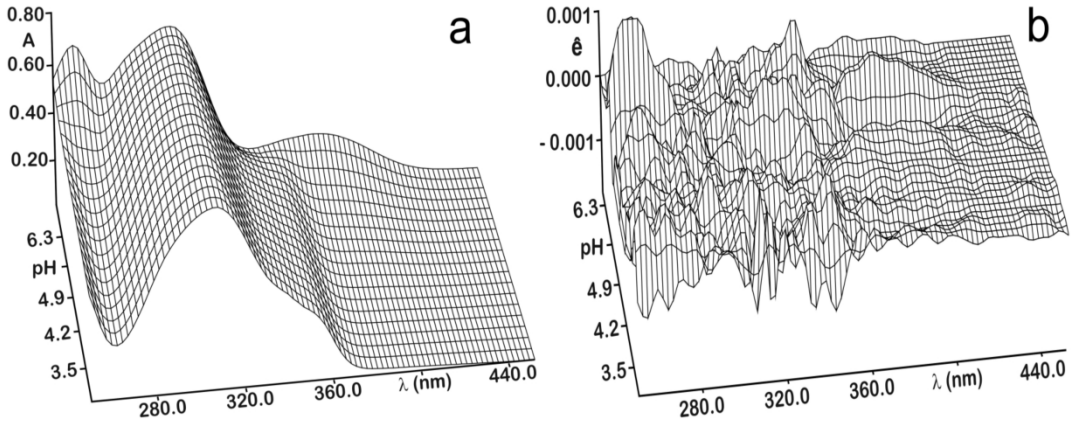

Figure 2. (a) The 3D-absorbance-response-surface representing the measured multiwavelength absorption spectra for methotrexate dependent on a pH at 25°C, (b) the 3D-residuals map after non-linear regression performed with SPECFIT and SQUAD, (S-Plus).

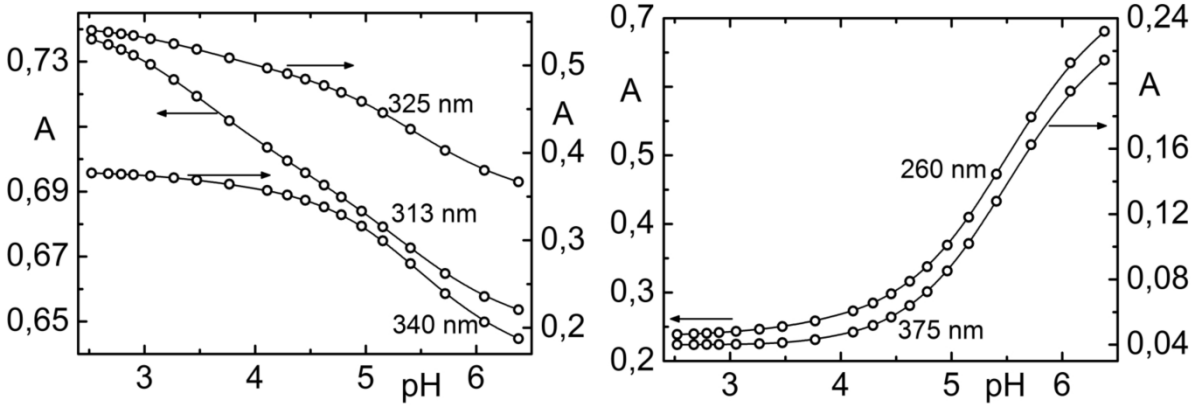

Figure 3. The absorbance vs.pH curves for 325 nm, 313 nm, 340 nm, 260 nm and 375 nm at 25°C.

of computational strategy, (5) diagnostics indicating a correct chemical model: When a minimization process terminates, selected curve-fitting diagnostics are examined to determine whether the results should be accepted: the physical meaning of parametric estimates, the physical meaning of the species concentrations, the goodness-of-fit test and the deconvolution of spectra.

3.3 Software

Computation relating to the determination of dissociation constants was performed by a regression analysis of the UV/VIS spectra using the SQUAD(84) [25] and SPECFIT/32 [35] programmes. Most of the graphs were plotted using ORIGIN 7.5 [46] and S-Plus [44]. The thermodynamic dissociation constant pK_a^T was estimated using the MINOPT nonlinear regression programme in the ADSTAT statistical system (TriloByte Statistical Software, Ltd., Czech Republic) [47]. A

Table 1. The search for a protonation model of methatrexate with the SQUAD(84) analysis of A-pH spectra at 25°C and an ionic strength I = 0.0012. The standard deviations of the parameters estimated in the last valid digits are in brackets.

Model hypothesis	1st hypothesis: L, HL, H_2L	2nd hypothesis: L, HL, H_2L, H_3L
k, s_k(A) [mAU]	3, 0.52	4, 0.46 mAU
pK_{a1}	3.813(18)	3.086(167)
pK_{a2}	5.639(2)	4.403(63)
pK_{a3}	---	5.675(6)
s(A) [mAU]	0.87	0.65
s(e) [mAU]	0.87	0.65
| ē | [mAU]	0.58	0.42
g_1	-0.07	-0.28
g_2	3.21	3.32
R [%]	0.18	0.13
Model is	rejected	accepted

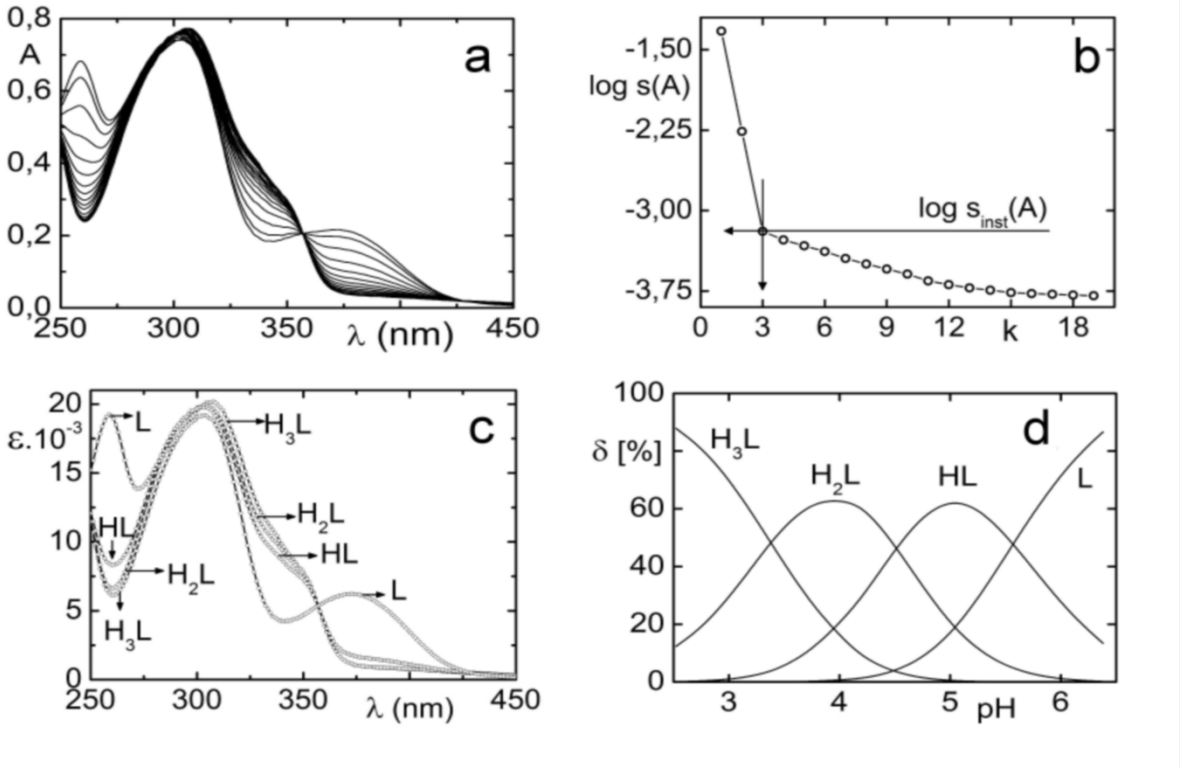

Figure 4. The non-linear regression analysis of the protonation equilibria model and factor analysis of methotrexate: (a) Absorption spectra of 4×10^{-5} M MTX dependent on pH at 25°C; (b) Cattel's scree plot of the Wernimont-Kankare procedure for the determination of the number of light-absorbing species in the mixture $k^*=3$ leads to $n_c=3$ and the actual instrumental error of the spectrophotometer $s_{inst}(A) = 0.65$ mAU (INDICES in S-Plus); (c) pure spectra profiles of molar absorptivities vs. wavelengths for the variously protonated species L, HL, H_2L and H_3L; (d) distribution diagram of the relative concentrations for both variously protonated species L, HL, H_2L and H_3L of rasagiline base dependent on pH at 25°C. The charges of species are omitted for the sake of simplicity (SPECFIT, ORIGIN).

qualitative interpretation of the spectra with the use of the INDICES programme [45] aims to evaluate the quality of the dataset and remove spurious data, and to estimate the minimum number of *factors*, *i.e.,* contributing aqueous species necessary to describe the experimental data and to determine the number of dominant species present in the equilibrium mixture. PALLAS [36] is the programme for making predictions based on the structural formulae of drug compounds. Entering the compound topological structure descriptors graphically, pK_a values of organic compounds are predicted using approximately hundreds of Hammett and Taft equations and quantum chemistry calculus.

3.4 Supporting information available

Complete experimental and computational procedures, input data specimens and corresponding output in numerical and graphical form for the programmes, INDICES, SQUAD(84) and SPECFIT/32 are available free of charge on line at *http://meloun.upce.cz* and in the block DOWNLOAD and block DATA.

4. Results and Discussion

Recently, methotrexate studied in our laboratory represents a drug acid which exhibits small changes in spectra. Other instrumental methods could not be used for limited solubility in water. It is wise before starting a regression to analyze actual experimental data, to search for scientific library sources, to obtain a good default for the number of ionizing groups, and numerical values for the initial guess as to relevant stability (protonation) constants and the probable spectral traces of all the expected components [48].The program PALLAS [36] provides a collection of powerful tools for making a prediction of the pK_a values of any organic compound on the basis of the structural formulae of the compounds, using approximately 300 Hammett and Taft equations. Depending on the nature of the chemical structure and hypothesis that the ionization state of a particular group is dependent upon its subenvironments constituted by its neighboring atoms and bonds, a hierarchical tree is constructed outward from the ionizing atom. Predicted

Table 2. The search for a protonation model of methatrexate with the ADSTAT analysis of A-pH curves measured at 260, 313, 325, 240, 375 nm. The standard deviations of the parameters estimated in the last valid digits are in brackets.

	A-pH curve at 260 nm		A-pH curve at 313 nm		A-pH curve at 325 nm	
Model	L, HL, H_2L	L, HL, H_2L, H_3L	L, HL, H_2L	L, HL, H_2L, H_3L	L, HL, H_2L	L, HL, H_2L, H_3L
pK_{a1}	3.919(94)	3.376(0)	3.538(45)	3.377(0)	3.549(45)	3.377(0)
pK_{a2}	5.521(09)	4.514(1)	5.363(45)	4.515(1)	5.469(20)	4.515(0)
pK_{a3}		5.572(0)		5.573(0)		5.572(0)
ε_L	0.736(2)	0.738(0)	0.650(1)	0.649(0)	0.3523(0)	0.351(2)
ε_{LH}	0.279(5)	0.320(0)	0.696(1)	0.685(0)	0.489(2)	0.470(2)
ε_{LH2}	0.238(1)	0.255(0)	0.740(1)	0.705(0)	0.5444(1)	0.501(1)
ε_{LH3}	-	0.236(0)	-	0.742(0)	-	0.546(0)
s(A)	4.91E-04	3.16E-07	3.66E-06	3.09E-07	4.75E-04	3.30E-07
s(e)	4.21E-04	2.51E-07	3.14E-04	2.46E-07	4.08E-04	2.58E-07
| ē |	3.73E-04	2.22E-07	2.76E-04	2.08E-07	3.65E-04	2.12E-07
R [%]	11.3	0.01	4.48	0.00	8.45	0.01
Model is	rejected	accepted	rejected	accepted	rejected	accepted

	A-pH curve at 340 nm		A-pH curve at 375 nm	
Model	L, HL, H_2L	L, HL, H_2L, H_3L	L, HL, H_2L	L, HL, H_2L, H_3L
pK_{a1}	3.676(72)	3.376(1)	4.927(282)	3.345(18)
pK_{a2}	5.510(09)	4.514(1)	5.670(123)	4.528(7)
pK_{a3}		5.572(0)		5.574(10)
ε_L	0.165(1)	0.164(0)	0.239(1)	0.238(1)
ε_{LH}	0.356(1)	0.336(0)	0.097(35)	0.065(3)
ε_{LH2}	0.378(1)	0.364(0)	0.034(0)	0.040(4)
ε_{LH3}	-	0.379(0)	-	0.040(0)
s(A)	2.70E-04	3.23E-07	1.75E-04	2.33E-05
s(e)	2.32E-04	2.57E-07	1.50E-04	1.85E-05
| ē |	2.08E-04	2.27E-07	1.17E-04	1.30E-05
R [%]	6.96	0.01	15.6	1.93
Model is	rejected	accepted	rejected	accepted

values of three dissociation constants are $pK_{a1,pred}$ = 3.27, $pK_{a2,pred}$ = 4.72 and $pK_{a3,pred}$ = 5.36.

The pH-spectrophotometric titration enables absorbance-response data (Fig. 2a) to be obtained for analysis by the least-squares nonlinear regression. The reliability of parameter estimates (p*K*´s and ε´s) can be evaluated on the basis of the goodness-of-fit test of residuals (Fig. 2b). Moreover, the *A*-pH curves at 325, 323, 340, 260 and 375 nm (Fig. 3) also show that the dissociation constants of methotrexate may be indicated spectrophotometrically. The changes in spectra (Fig. 4a) are small within deprotonation. In fact, both of the variously protonated species H_2L and H_3L exhibit quite similar absorption bands. The adjustment of pH value from 2 to 7 causes the absorbance to change by only 0.150 of the A-pH curve, thus making the monitoring of the four components L, HL, H_2L, and H_3L of the protonation equilibrium rather unclear. As the changes in spectra are small, a very precise measurement of absorbance is neccessary for a reliable detection of the deprotonation equilibrium.

The first step of the regression spectra analysis estimated the number of light-absorbing species using the INDICES algorithm [45] (Figs. 5 and 4b). The number of light-absorbing species *p* can be predicted from the index function values by finding the point $p = k$ where the slope of the index function $PC(k) = f(k)$ changes, or by comparing $PC(k)$ values to the instrumental error $s_{inst}(A)$. This is the common criterion for determining *p* (Fig. 5). The very low value of $s_{inst}(A)$ in Fig. 4b proves that a sufficiently precise spectrophotometer and efficient experimental technique were used. The position of the break point on the $s_k(A) = f(k)$ curve in the factor analysis scree plot is calculated and gives $k = 3$ with the corresponding co-ordinate $s_3(A)$ = 0.43 mAU (*i.e.,* log $s_3(A)$ = -0.36) or $k = 4$ with the corresponding co-ordinate $s_4(A)$ = 0.36 m AU(*i.e.*, log $s_4(A)$ = -0.44),

Table 3. Dependence of the estimated mixed dissociation constants pK_a of methotrexate on ionic strength using regression analysis of pH–spectrophotometric data with SPECFIT and SQUAD at 25°C and at 37°C. The standard deviations of the pK_a in the last valid digits are in brackets.

		25°C					
Ionic strength		0.0012	0.0218	0.0630	0.1246	0.2274	0.3510
SPECFIT	pK_{a1}	2.887(126)	2.871(161)	2.907(150)	3.004(150)	3.197(145)	3.393(76)
	pK_{a2}	4.372(19)	4.308(20)	4.332(25)	4.424(22)	4.603(26)	4.777(22)
	pK_{a3}	5.673(8)	5.601(12)	5.581(13)	5.635(20)	5.857(14)	6.063(13)
	s(A) [mAU]	0.36	0.35	0.27	0.25	0.37	0.35
SQUAD	pK_{a1}	3.085(167)	2.864(132)	2.909(176)	3.038(307)	3.194(155)	3.289(122)
	pK_{a2}	4.403(62)	4.286(61)	4.326(73)	4.426(75)	4.606(106)	4.789(94)
	pK_{a3}	5.675(5)	5.605(9)	5.583(7)	5.637(8)	5.845(22)	6.061(21)
	s(A) [mAU]	0.65	0.48	0.47	0.48	0.56	0.52

		37°C					
Ionic strength		0.0012	0.0218	0.0632	0.1246	0.2274	0.3510
SPECFIT	pK_{a1}	3.066(110)	3.080(180)	3.096(130)	3.213(66)	3.351(89)	3.588(76)
	pK_{a2}	4.362(94)	4.211(89)	4.298(51)	4.482(46)	4.776(65)	5.182(64)
	pK_{a3}	5.545(12)	5.469(21)	5.451(16)	5.627(16)	5.783(16)	6.363(15)
	s(A) [mAU]	0.26	0.33	0.96	0.28	0.34	0.45
SQUAD	pK_{a1}	3.097(148)	3.071(200)	3.109(151)	3.198(80)	3.359(120)	3.584(172)
	pK_{a2}	4.347(50)	4.248(60)	4.240(72)	4.486(70)	4.765(90)	5.157(76)
	pK_{a3}	5.532(5)	5.453(10)	5.477(14)	5.638(15)	5.795(15)	6.369(20)
	s(A) [mAU]	0.42	0.54	1.10	0.52	0.54	0.59

Table 4. Thermodynamic dissociation constants pK_a^T at 25°C and 37°C and apparent dissociation constants pK_a at 25°C from literature [48] and predicted values from the PALLAS programme for methotrexate. Th e standard deviations in the last valid digits are in brackets. (X) indicates UV-inactive dissociation steps and pK_a values were taken from pressure-assisted capillary electrophoresis (PACE) and fixed during least-squares evaluation.

Estimated with SPECFIT			Apparent pK_a values at 25°C at ionic strength equal to				Predicted with
	Value at 25°C	Value at 37°C	0.05 M with PACE	0.05 M with PACE	0.15 M with UV-VIS	0.15 M with pot. titr. aq.	with PALLAS
pK_{a1}^T	2.895(4)	3.089(8)	3.22(4)	3.22(4)	3.22(X)	3.5(1)	3.27
pK_{a2}^T	4.410(6)	4.392(10)	4.53(6)	4.45(6)	4.45(X)	4.4(1)	4.72
pK_{a3}^T	5.726(8)	5.585(14)	5.62(6)	5.46(6)	5.52(3)	5.5(1)	5.36

which may also be taken as the actual instrumental error $s_{inst}(A)$ of the spectrophotometer. All 12 selected methods of modified factor analysis in Fig. 5 estimate the three or four light-absorbing components L, HL, H_2L, and H_3L of the protonation equilibrium. The two or three dissociation constants and three or four molar absorptivities of methotrexate calculated for 29 wavelengths constitute 2 + (3×29) = 89 or 3 + (4×29) = 119 unknown parameters, which are estimated and refined by SQUAD(84) or SPECFIT32 in the first run. The reliability of the parameter estimates may be tested using the following five diagnostics:

The *1st diagnostic* indicates whether all of the parametric estimates $pK_{a,r}$ and ε_r have a physical meaning and reach realistic values for the 20 values of pH-spectra of methotrexate measured at 29 wavelengths. The SQUAD(84) programme gives the parameter estimates pK_{a1} = 3.086(0.167), pK_{a2} = 4.403(0.063) and pK_{a3} = 5.675(0.006) at 25°C with

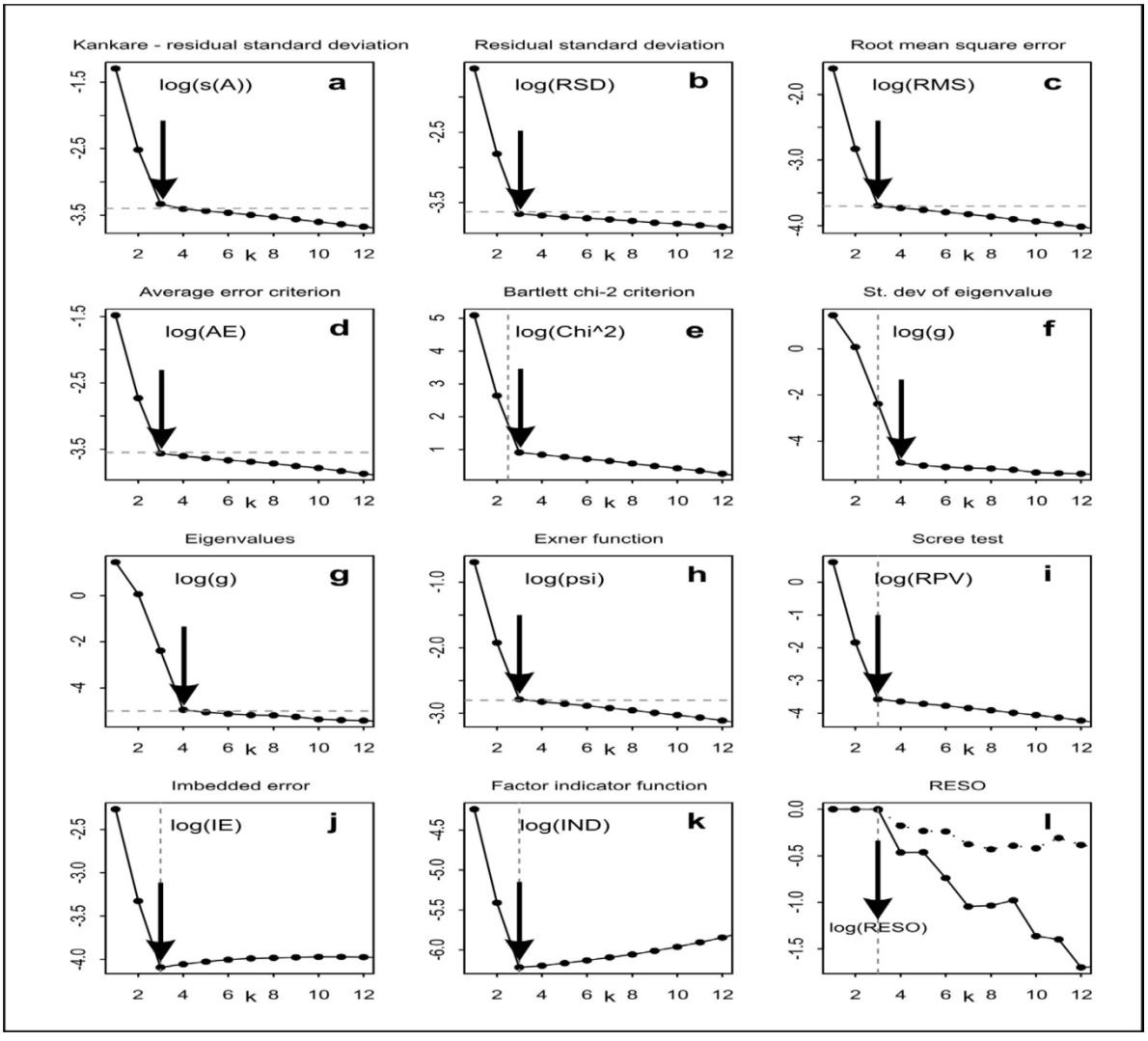

Figure 5. The logarithm dependence of the cattels index plot of eigenvalues in form of 12 indices modifying methods as a function of the number of principal components k for the pH-absorbance matrix: (a) Kankare's residual standard deviation, $s_k(A)$, (b) Residual standard deviation, RSD(k), (c) Root mean square error, RMS, (d) Average error criterion, AE(k), (e) Bartlett χ^2 criterion, $\chi^2(k)$, (f) Standard deviation of eigenvalues s(g), (g) Eigenvalues, (h) Exner function ψ(k), (i) Scree test, RPV(k), (j) Imbedded error function, IE(k), (k) Factor indicator function, IND(k), (l) Ratio of eigenvalues calculated by smoothed PCA and those by ordinary PCA, RESO(k). The arrows indicate that most of the methods lead to 3 light-absorbing species in a pH-equilibrium mixture (S-Plus).

$s(A)$ = 0.65 mAU, while the SPECFIT programme leads to estimates pK_{a1} = 2.887(0.126), pK_{a2} = 4.372(0.019) and pK_{a3} = 5.673(0.008) at 25ºC with $s(A)$ = 0.36 mAU. As the standard deviations $s(pK_a)$ of parameters pK_a and $s(\varepsilon_r)$ of parameters ε_r are significantly smaller than their corresponding parameter estimates (Table 1), all of the variously protonated species are statistically significant at a significance level of α = 0.05. The physical meaning of the dissociation constant $pK_{a,r}$, molar absorptivities ε_r, and stoichiometric indices r are examined in search of the protonation equilibria model in Table 1. The 1st hypothesis of the protonation model L, HL, H_2L is rejected, as poor fitness was achieved. The absolute values of $s(pK_{a,j})$, $s(\varepsilon_j)$ give information about the last U-contour of the hyperparaboloid in the proximity of the pit, U_{min}. For well-conditioned parameters, the last U-contour is a regular ellipsoid, and the standard deviations are reasonably low. High s values are found with ill-conditioned parameters and a "saucer"-shaped pit. The relationship of $s(\beta_j) \times F_\sigma < \beta_j$ should be met where F_σ is equal to 3. Fig. 4c shows that the estimated molar absorptivities of all of the variously protonated species ε_L, ε_{HL}, ε_{H2L} and ε_{H3L} of methotrexate dependinge on wavelength are realistic. Some spectra overlap and may cause some resolution difficulties in the regression analysis. As the two protonation models in the model

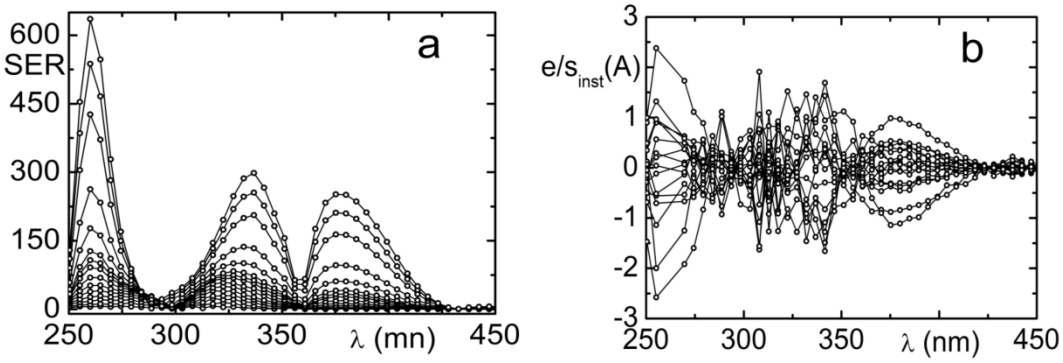

Figure 6. (a) The plot of small absorbance changes in the spectrum of methotrexate means that the value of the absorbance difference for the jth-wavelength of the ith-spectrum $\Delta_{ij} = A_{ij} - A_{i,acid}$ is divided by the instrumental standard deviation $s_{inst}(A)$, and the resulting ratios $SER = \Delta/s_{inst}(A)$ are plotted in dependence of wavelength λ for all absorbance matrix elements, where $A_{i,acid}$ is the limiting spectrum of the acid form of the drug measured. This ratio is compared with the limiting SER value for the methotrexate to test if the absorbance changes are significantly larger than the instrumental noise. (b) The plot of the ratio $e/s_{inst}(A)$, i.e., the ratio of the residuals divided by the instrumental standard deviation $s_{inst}(A)$ dependent on wavelength λ for all the residual matrix elements for methotrexate tests if the residuals are of the same magnitude as the instrumental noise.

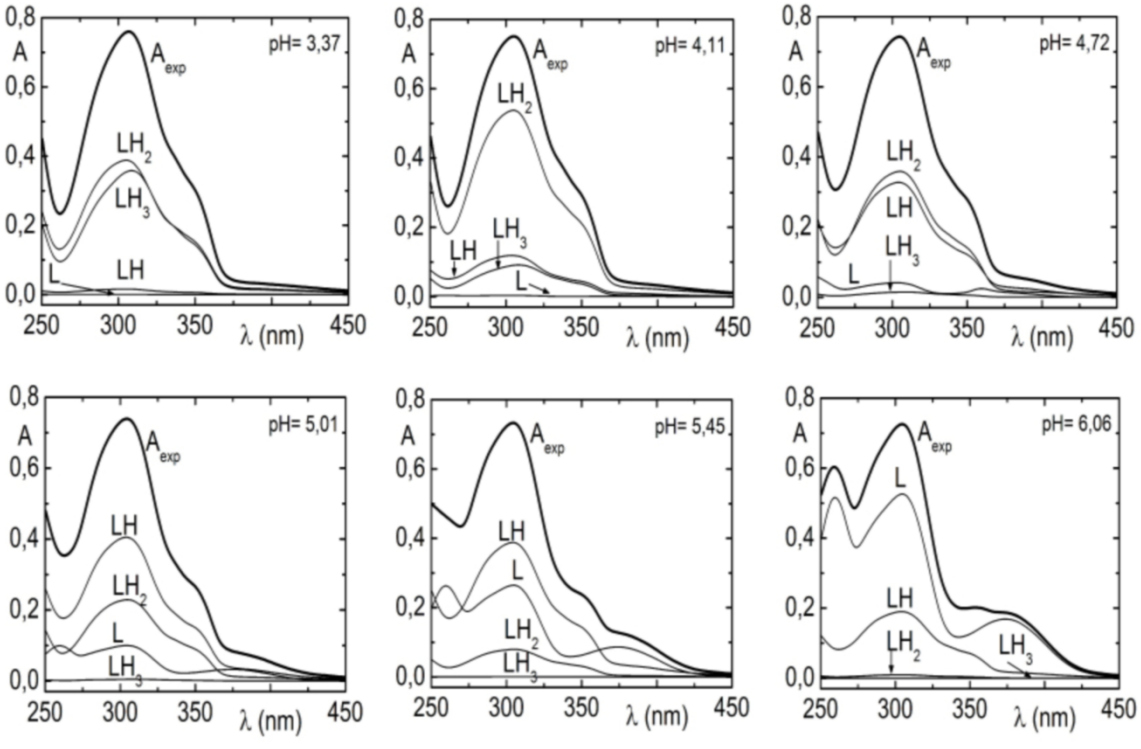

Figure 7. Deconvolution of the experimental absorption spectrum of methotrexate for 39 wavelengths into spectra of the individual variously protonated species L, HL, H_2L and H_3L, in solution of each particular absorption spectrum for a selected value of pH equal to 3.37, 4.11, 4.72, 5.01, 5.45, and 6.06. The charges of species are omitted for the sake of simplicity (SQUAD, ORIGIN).

search depicted in Table 1 (1st model: L, HL, H_2L, 2nd model: L, HL, H_2L, and H_3L) were tested, it may be concluded that the regression spectra analysis can distinguish between these two models; and on the basis of a very good spectra fitting, the model L, HL, H_2L, and H_3L was proven.

The *2nd diagnostic* tests whether all of the calculated free concentrations of the three variously protonated species on the distribution diagram (the relative concentration expressed as a percentage) have

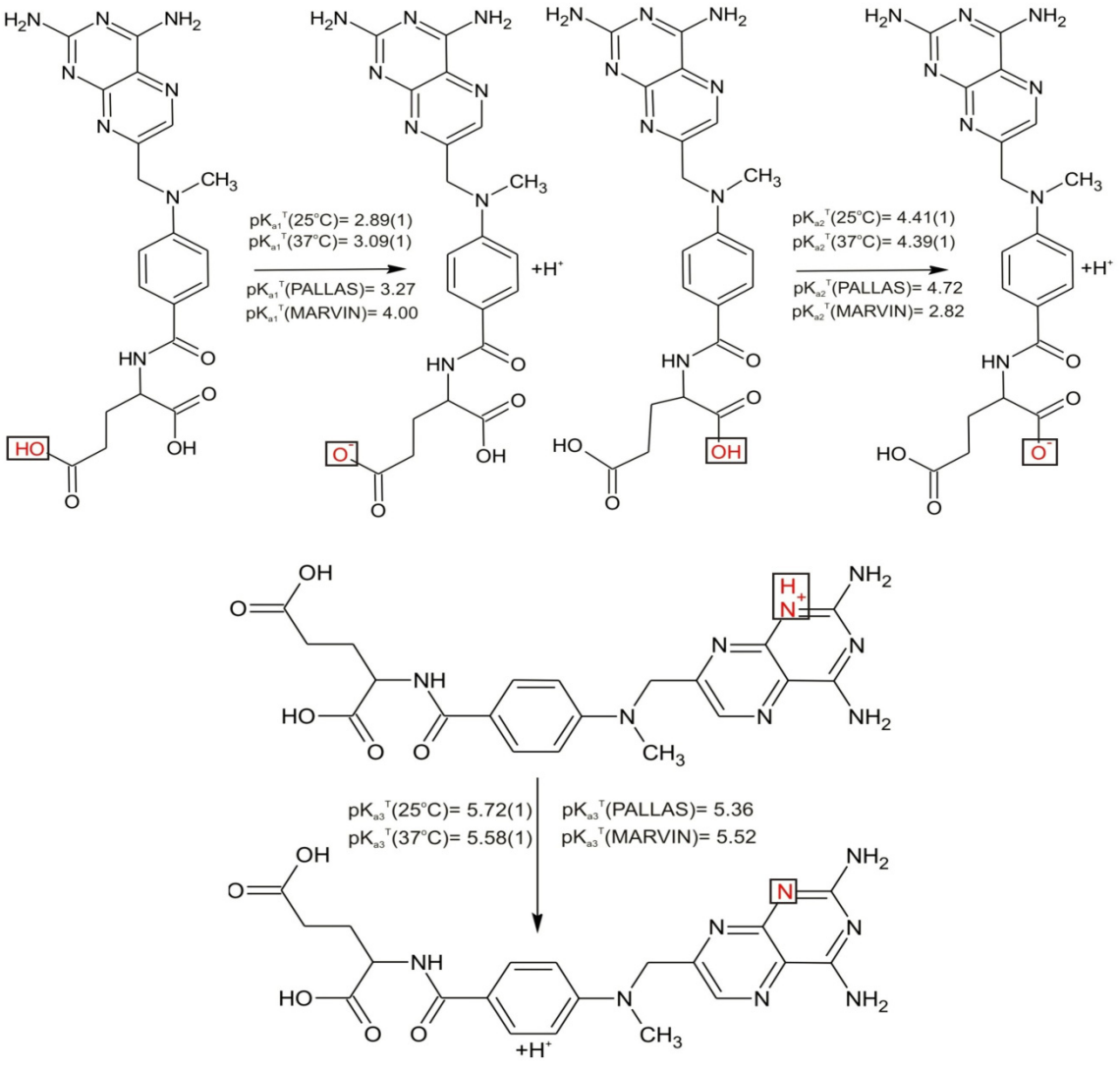

Figure 8. Protonation schema of methotrexate.

physical meaning. This proved to be the case. The calculated free concentration of the basic components and variously protonated species of the protonation equilibria model should show molarities to about 10^{-8} M. Species at about 1% relative concentration or less in an equilibrium behave as numerical noise in a regression analysis. A distribution diagram of relative concentration shows the protonation equilibria of L, HL, H_2L, and H_3L depending on the pH as illustrated in Fig. 4d and makes it easier to determine more efficiently, the contributions of the individual species to the total concentration. Since the molar absorptivities will generally be in the range of 10^3 - 10^5 L mol^{-1} cm^{-1}, species present at a relative concentration of less than *ca.* 0.1% which will affect the absorbance significantly if their ε is extremely high.

The *3rd diagnostic* concerning the matrix of correlation coefficients in output of SQUAD(84) proves that there is an interdependence of two pairs of protonation methotrexate $r(\log \beta_{11}$ *vs.* $\log \beta_{12}) = 0.9086$ and $r(\log \beta_{12}$ *vs.* $\log \beta_{13}) = 0.8823$ constants which indicate that their estimation will be more difficult.

The *4th diagnostic* addresses the goodness-of-fit. The goodness-of-fit achieved is easily examined using the difference between the experimental and calculated values of absorbance, $e_i = A_{exp,\ i,\ j} - A_{calc,\ i,\ j}$. The examination of the spectra and the graph of the predicted absorbance response-surface, as seen in all of the experimental points (Fig. 2b), should reveal whether the calculated results are consistent and whether any severe experimental errors have been made in the spectra measurement (Fig. 6b). One of the most important statistics calculated is the standard

deviation of absorbance, *s(A)*. This is calculated from a set of refined parameters at the termination of the minimization process. This is usually compared to the standard deviation of absorbance calculated by the INDICES programme [45], $s_k(A)$, and if $s(A) \le s_k(A)$, or $s(A) \le s_{inst}(A)$, the instrumental error of the applied spectrophotometer, the fit is considered to be statistically acceptable (Table 1). This proves that the $s_3(A)$ value is equal to 0.43 mAU or $s_4(A)$ value is equal to 0.36 mAU and is close to the standard deviation of absorbance upon the termination of the the SPECFIT minimization process, $s(A)$ = 0.35 mAU (or 0.65 mAU SQUAD(84)). Although this statistical analysis of residuals gives the most rigorous test for the degree-of-fit, realistic empirical limits must be used. The statistical measures of all residuals *e* prove that the minimum of the eliptic hyperparaboloid *U* is reached with SQUAD(84): the residual standard deviation *s(e)* = 0.52 mAU always has sufficiently low values. Graphical presentation of the residuals on Fig. 2b and 6b assists the detection of an outlier spectrum point, a trend in the spectrum residuals, or an abrupt level shift in the spectra. The statistical measures of all the residuals depicted in Fig. 2b prove that the minimum of the eliptic hyperparaboloid is reached: the residual mean $\overline{e}$ = 1.65×10^{-16} proves that there is no bias or systematic error in the spectra fitting. The mean residual $|\overline{e}|$ = 0.42 mAU and the residual standard deviation *s(e)* = 0.65 mAU (or 0.36 mAU SPECFIT) have sufficiently low values. The skewness $g_1(e)$ = -0.28 is close to zero and proves a symmetric distribution of the residuals set, while the kurtosis $g_2(e)$ = 3.32 is close to 3 proving a Gaussian distribution. The Hamilton *R-factor* of relative fitness is 0.13% calculated with SQUAD(84) only, and thus proving an excellent achieved fitness; the parameter estimates may therefore be considered reliable. If the Hamilton *R-factor* of relative fit, expressed as a percentage, is < 0.5%, the fit is excellent; If it is > 2%, the fit is poor. The criteria of resolution used for the hypotheses were: (1) a failure of the minimization process in a divergency or a cyclization; (2) an examination of the physical meaning of the estimated parameters to ensure that they were both realistic and positive; and (3) the residuals should be randomly distributed around the predicted regression spectrum. Systematic departures from randomness were taken to indicate that either the chemical model or the parameter estimates were unsatisfactory.

Using the experimental and evaluation strategy of the A-pH curves measured at dominant analytical wavelengths (Fig. 3) also allows to find a true hypothesis for the protonation equilibria model of methotrexate. To test the reliability of a model and the estimated protonation/dissociation constants, the goodness-of-fit test using residuals statistical analysis was applied; results appear in Table 2. For methotrexate equilibria studied, the most efficient tools, such as the Hamilton *R*-factor, the mean residual and the standard deviation of residuals were applied: as the *R*-factor in all cases reaches a value of less then 0.02% indicates an excellent fitness and reliable parameter estimates. The standard deviation of absorbance $s(A)$ after the termination of the minimization process is always better than 10^{-2} and the proposal of a good chemical model and reliable parameter estimates are proven.

To express and analyze small absorbance changesin the spectral set, the absorbance differences for the *j*-th wavelength of the *i*-th spectrum $\Delta_i = A_{ij} - A_{i,acid}$ were calculated to subtract the actual pH absorbance spectrum value from the acidic form. The absorbance difference Δ_i was then divided by the actual instrumental standard deviation $s_{inst}(A)$ of the spectrophotometer, and the resulting value represents the *signal-to-error* value *SER*. Fig. 6a shows a graph of the *SER* depending on the wavelength in the measured range for the drug. When the *SER* is larger than 10, a factor analysis sufficiently predicts the correct number of light-absorbing components in the equilibrium mixture. To prove that a non-linear regression also can analyze such data of small absorbance changes, the residuals set was compared with the instrumental noise $s_{inst}(A)$. If the ratio $e/s_{inst}(A)$ is of similar magnitude, *i.e.*, nearly equal to one, it means that sufficient curve fitting is achieved by the non-linear regression of the spectra set and that the minimization process determined the minimum of the residual-square-sum function U_{min}. Fig. 6b displays a comparison of the ratio $e/s_{inst}(A)$ which is dependent on the wavelength for the measured methotrexate. From the figure it is obvious that most of the residuals are of the same magnitude as the instrumental noise, thus proving a sufficient reliability of the performed regression process.

The *5th diagnostic*, the spectra deconvolution in Fig. 7, shows the deconvolution of the experimental spectrum into spectra of the individual variously protonated species and examines whether the experimental design is efficient. If, for a particular pH range, the spectrum consists of just a single component, further spectra for that range would be redundant, although they could improve the rate of precision. Several spectra should be measured in pH ranges where more components contribute significantly to the spectrum. Such a spectrum provides sufficient information for a regression analysis which monitors at least two species in equilibrium none of which is a minor species. The minor species has a relative concentration in a distribution diagram of less

than 5% of the total concentration of the basic component c_L. When, on the other hand, only one species prevails in solution, the spectrum yields poor information for a regression analysis, and the parameter estimate is vague and definitely not sufficiently reliable. Spectrum deconvolution proves to be quite usefulin proposing an efficient experimentation strategy.

The unknown parameter pK^T_a has been estimated by applying a Debye-Hückel equation to the data in Table 3 according to the regression criterion (6). Table 4 defines point estimates of the thermodynamic methotrexate dissociation constants of studied at two temperatures. Because of the small range of ionic strength, the ion-size parameter *å* and the salting-out coefficient *C* could not be estimated .

5. Conclusions

When drugs are poorly soluble, pH-spectrophotometric titration may be used with the non-linear regression of the absorbance-response-surface data instead of performing a potentiometric determination of the dissociation constants. The reliability of the dissociation constants of the drug methotrexate studied may be proven with goodness-of-fit tests of the absorption spectra measured at various pH levels. Three thermodynamic dissociation constants were estimated by nonlinear regression of {pK_a, *I*} data: for methotrexate pK_{a1}^T= 2.895(4), pK_{a2}^T= 4.410(6), pK_{a3}^T= 5.726(8) at 25°C and pK_{a1}^T= 3.089(8), pK_{a2}^T= 4.392(10), pK_{a3}^T= 5.585(14) at 37ºC, where the figure in brackets is the standard deviation in last significant digits. The agreement with the literature data is shown in Table 4. Most indices always predict the correct number of components when the *signal-to-error ratio SER* is far exceeds 10. The Wernimont-Kankare procedure in INDICES performs a reliable determination of the instrumental standard deviation of the spectrophotometer used $s_{inst}(A)$ which predicts the number of light-absorbing components present n_c and can also solve an ill-defined problem with severe collinearity in the spectra or very small changes in spectra.

Methotrexate is a weak dicarboxylic acid with dissociation constants approximately 4.8 and 5.5, and thus it was expected to mostly ionize at physiologic pH. However, ionized forms of drugs in general can only barely cross the (lipophilic) cell membrane bilayer. Whereas protonated species predominate at pH of the gastric juice, then, at pH > 6, the non-protonated form of the drug is dominant. The distribution of protonated/non-protonated forms has turned cross-over towards unprotonated species (Fig. 8).

Acknowledgments

The financial support from the Grant Agency IGA MZ ČR (Grant No NR-9055-4/2006 and No NS9831-4/2008) and the Czech Ministry of Education (Grant No MSM0021627502) is gratefully acknowledged.

References

[1] A. Albertioni et al., J. Chrom. B 665, 163 (1995)
[2] F. Albertioni et al., Clinical Chemistry 42, 39 (1966)
[3] J.L. Wisnicki, W.P. Tong, D.B. Ludlum, Cancer Treatment Reports 62:4, 529 (1978)
[4] R. Turci, G. Micoli, C. Minoia, Rapid Commun. Mass Spectrom. 14, 685 (2000)
[5] B. Nuernberg, M. Kohlbrenner, R. Faulkner, D.E. Furst, J. Chromatogr., Biomed. Applic. 487, 476 (1989)
[6] B.N. Cronstein, J.R. Bertino, Milestones in Drug Therapy; Methotrexate (Birkhauser Verlag AG, Basel-Berlin-Boston, 2000)
[7] B.N. Cronstein, M.A. Eberle, H.E. Gruber, R. I. Levin, Proc. Nail. Acad. Sci. USA. 88, 2441 (1991)
[8] R. Turci, M.L. Fiorentino, C. Stanni, C. Minoia, Rapid Commun. Mass Spectrom. 14, 173 (2000)
[9] O. Beck et al., Ther. Drug Monit. 13, 528 (1991)
[10] S. Steinborner, J. Henion, Anal. Chem 71, 2340 (1999)
[11] M. Meloun, J. Čapek, P. Mikšík, R.G. Brereton, Anal. Chim. Acta. 423, 51 (2000)
[12] M. Meloun, M. Pluhařová, Anal. Chim. Acta, 416, 55 (2000)
[13] M. Meloun, P. Černohorský, Talanta 52, 931 (2000)
[14] M. Meloun, D. Burkoňová, T. Syrový, A. Vrána, Anal. Chim. Acta, 486, 125 (2003)
[15] M. Meloun, T. Syrový, A. Vrána, Anal. Chim. Acta 489, 137 (2003)
[16] M. Meloun, T. Syrový, A. Vrána, Talanta 62, 511 (2004)
[17] M. Meloun, T. Syrový, A. Vrána, Anal. Chim. Acta 533, 97 (2005)
[18] M. Meloun, T. Syrový, A. Vrána, Anal. Chim. Acta 543, 254 (2005)
[19] M. Meloun, M. Javůrek, J. Militký, Microchimica Acta 109, 221 (1992)

[20] M. Meloun, J. Havel, E. Högfeldt, Computation of Solution Equilibria (Ellis Horwood, Chichester, 1988)
[21] M. Meloun, J. Havel, Computation of Solution equilibria, 1. Spectrophotometry, Folia Fac. Sci. at. Univ. Purkyn. Brunensis (Chemia), XXV,Brno 1984, 2. Potentiometry, XXVI, Brno 1985.
[22] L. G. Sillén, B. Warnqvist, Acta Chem. Scand. 22, 3032 (1968)
[23] SQUAD: D.J. Leggett (Ed.), Computational Methods for the Determination of Formation Constants (Plenum Press, New York, 1985), (a) pp. 99-157; (b) pp. 291-353
[24] J. Havel, M. Meloun, In: D.J. Leggett (Ed.), Computational Methods for the Determination of Formation Constants (Plenum Press, New York, 1985), (a) pp. 19; (b) pp. 221
[25] SQUAD(84): M. Meloun, M. Javůrek, J. Havel, Talanta 33, 513 (1986)
[26] D.J. Leggett, W.A.E. McBryde, Anal. Chem. 47, 1065 (1975)
[27] D.J. Leggett, Anal. Chem. 49, 276 (1977)
[28] D.J. Leggett, S.L. Kelly, L.R. Shiue, Y.T. Wu, D. Chang, K.M. Kadish, Talanta 30, 579 (1983)
[29] J.J. Kankare, Anal. Chem. 42, 1322 (1970)
[30] P. Gans, A. Sabatini, A. Vacca, Talanta 43, 1739 (1996)
[31] H. Gampp, M. Maeder, C.J. Mayer, A. Zuberbuhler, Talanta 32, 95 (1985)
[32] H. Gampp, M. Maeder, C.J. Meyer, A. Zuberbühler, Talanta 32, 251 (1985)
[33] (a) H. Gampp, M. Maeder, C.J. Meyer, A. Zuberbühler, Talanta 32, 1133 (1985); (b) H. Gampp, M. Maeder, C.J. Meyer, A. Zuberbühler, Talanta 33, 943 (1986)
[34] Y. Kin Tam, K. Takács-Novák, Anal. Chim. Acta 434, 157 (2001)
[35] SPECFIT/32: Spectrum Software Associates, 197M Boston Post Road West, Marlborough, MA, 01752 U.S.A. (2004), (http://www.bio-logic.info/rapid-kinetics/specfit.html).
[36] Pallas: Compudrug International Inc., Morgan Drive 115, Sedona, AZ86351, U.S.A.
[37] M. Meloun, S. Bordovská, T. Syrový, A. Vrána, Anal. Chim. Acta 580, 107 (2006)
[38] M. Meloun, S. Bordovská, A. Vrána, Anal. Chim. Acta 584, 419 (2007)
[39] M. Meloun, T. Syrový, S. Bordovská, A. Vrána, Anal. Bioanal. Chem. 387, 941 (2007)
[40] (a) M. Meloun, J. Militký, M. Forina, Chemometrics for Analytical Chemistry, PC-Aided Regression and Related Methods (Ellis Horwood, Chichester, 1994) Vol. 2; (b) M. Meloun, J. Militký, M. Forina, PC-Aided Statistical Data Analysis (Ellis Horwood, Chichester, 1992) Vol. 1
[41] M. Meloun, S. Bordovská, T. Syrový, J. Phys. Org. Chem. 20, 690 (2007)
[42] M. Meloun, S. Bordovská, L. Galla, J. Pharm. Biomed. Anal. 45, 552 (2007)
[43] M. Meloun, S. Bordovská, Anal. Bioanal. Chem. 389, 1267 (2007)
[44] S-PLUS: TIBCO SOFTWARE INC., Hillview Avenue 3303, Palo Alto, CA 94304, U.S.A.
[45] M. Meloun, T. Syrový, A. Vrána, Anal. Chim. Acta 489, 137 (2003)
[46] ORIGIN: OriginLab Corporation, One Roundhouse Plaza, Suite 303, Northampton, MA 01060, USA.
ADSTAT: ADSTAT 1.25, 2.0, 3.0 (Windows 95),
[47] TriloByte Statistical Software Ltd., Pardubice, Czech Republic
[48] L. Xing, R.C. Glen, J. Chem. Inf. Comput. Sci. 42, 796 (2002)
[49] Z. Szakács, B. Noszál, Electrophoresis 27, 3399 (2006)