Anal Bioanal Chem (2007) 387:941–955
DOI 10.1007/s00216-006-0993-1

ORIGINAL PAPER

Reliability and uncertainty in the estimation of pK_a by least squares nonlinear regression analysis of multiwavelength spectrophotometric pH titration data

Milan Meloun · Tomáš Syrový · Sylva Bordovská ·
Aleš Vrána

Received: 23 August 2006 / Revised: 30 October 2006 / Accepted: 6 November 2006 / Published online: 10 January 2007

Abstract When drugs are poorly soluble then, instead of the potentiometric determination of dissociation constants, pH-spectrophotometric titration can be used along with nonlinear regression of the absorbance response surface data. Generally, regression models are extremely useful for extracting the essential features from a multiwavelength set of data. Regression diagnostics represent procedures for examining the *regression triplet* (*data, model, method*) in order to check (a) the data quality for a proposed model; (b) the model quality for a given set of data; and (c) that all of the assumptions used for least squares hold. In the interactive, PC-assisted diagnosis of data, models and estimation methods, the examination of data quality involves the detection of *influential points*, outliers and high leverages, that cause many problems when regression fitting the absorbance response hyperplane. All graphically oriented techniques are suitable for the rapid estimation of influential points. The reliability of the dissociation constants for the acid drug silybin may be proven with goodness-of-fit tests of the multiwavelength spectrophotometric pH-titration data. The uncertainty in the measurement of the pK_a of a weak acid obtained by the least squares nonlinear regression analysis of absorption spectra is calculated. The procedure takes into account the drift in pH measurement, the drift in spectral measurement, and all of the drifts in analytical operations, as well as the relative importance of each source of uncertainty. The most important source of uncertainty in the experimental set-up for the example is the uncertainty in the pH measurement. The influences of various sources of uncertainty on the accuracy and precision are discussed using the example of the mixed dissociation constants of silybin, obtained using the SQUAD(84) and SPECFIT/32 regression programs.

Keywords Measurement uncertainty · Spectrophotometric titration · Dissociation constant · Protonation · pK_a reliability · Regression triplet · Residuals · Outliers · Influential points · Goodness-of-fit test · Silybin · SPECFIT · SQUAD · INDICES

Introduction

Proton transfer is a vital part of many chemical and biochemical processes and is determined by the acid dissociation constants (pK_a) of the chemicals involved. The acid–base character of a xenobiotic is an important property in the study of drug action, and in the development of new human and veterinary drugs, crop protecting agents, anticancer drugs, etc. Moreover, the pK_a values of ionizable drugs also affect their lipophilicity and permeability, which are important physicochemical considerations when predicting bioavailability. Most of the aspects of computational procedures used to estimate pK_a values have been described previously, such as software packages, descriptors, etc. (see [1] and references therein). Well-defined experimental methods used to estimate pK_a values, such as potentiometric titration (the standard approach) and spectrophotometric titrations (the alternative approach), as well as new approaches such as capillary electrophoresis, are described

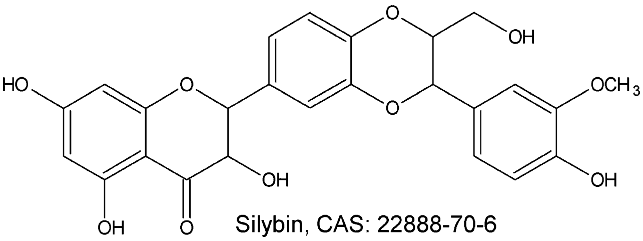
M. Meloun (✉) · T. Syrový · S. Bordovská
Department of Analytical Chemistry, University of Pardubice,
532 10 Pardubice, Czech Republic
e-mail: milan.meloun@upce.cz

A. Vrána
IVAX Pharmaceuticals, s.r.o.,
747 70 Opava, Czech Republic

in [2] and references therein. In previous work [1], the authors have shown that the spectrophotometric method in combination with suitable chemometric tools can be used to determine protonation constants β_{qr} or acid dissociation constants pK_a of even barely soluble drugs.

This paper describes a series of powerful general diagnostics for detecting observations that differ from the bulk of the data. These may be individual observations that do not belong to the general model, i.e., influential points or outliers. The identification of influential points and regression diagnostics is a relatively new topic in the chemometric literature, but it is rapidly gaining recognition and acceptance by practitioners as a supplement to the traditional analysis of residuals [3]. A single case approach to the detection of outliers can, however, fail because of the masking effect, in which outliers go undetected because of the presence of another, usually adjacent, observation. Regression diagnostics represent procedures for examining the *regression triplet* (*data, model, method*) in order to identify (a) the data quality for a proposed model; (b) the model quality for a given set of data; and (c) whether all the assumptions of least squares are fulfilled. The main difference between the use of regression diagnostics and that of classical statistical tests is that there is no need for an alternative hypothesis; all kinds of deviations from the ideal state are discovered. Our concept of exploratory regression analysis is based on the fact that "the user knows more about the data than the computer" [4].

In this paper, an estimation of the uncertainty in the measurement of the pK_a of a weak acid (obtained by multiwavelength spectrophotometric pH-titration) is presented. The procedure of pK_a uncertainty estimation takes into account various adjustable parameters such as the drift in pH measurement, the drift in spectral measurement, and all the drifts in analytical operations. The relative importances of various sources of uncertainty in terms of the whole experimental strategy is investigated, and their effects on the accuracy and precision of the dissociation constants pK_a are elucidated. By way of example, the various dissociation constants of the acid drug silybin at ionic strength I=0.30 and at a temperature of 25 °C are estimated using two nonlinear regression programs, SQUAD(84) [5–8] and SPECFIT/32 [9–12] (Scheme 1).



Scheme 1 Chemical structure of silybin

Theoretical

Nonlinear absorbance response hyperplane fitting is an important tool for multiwavelength spectrophotometric pH-titration data analysis. Before the advent of the nonlinear regression program approach to pK_a determination, linear relationships were "graphically fitted," with ruler and graph paper. Nonlinear relations had to be linearized in some appropriate way (a survey of this approach is provided by [13–16]). Subsequent analysis involved a manual straight-line fit, and slope and intercept were interpreted according to the linearization used. While it is possible to computerize this approach, it is inadequate to do so. Error analysis is seriously hampered by the distortions imposed by the linearization function used. Nonlinear least squares fitting is superior, as there are no distortions of the noise structure of the data.

Estimation of protonation constants by nonlinear least squares regression

Computations related to the determination of protonation constants β_{qr} (or dissociation constants pK_a) may be performed by least squares regression analysis of multiwavelength spectra using versions of the SQUAD family of programs [5–8] and SPECFIT/32 [9–12]; as has been described in the tutorial [1]. The experimental and computational schemes used to determine the protonation constants of a multicomponent system are taken from Meloun et al. [14–16]. Numerical details of the computer data treatment, data inputs and corresponding outputs are listed in the "Supporting information."

In order to briefly to explain the methodology for the analysis of sets of spectra of the complexity described above, it is necessary to review the principles involved in performing nonlinear least squares fitting of the absorbance response hyperplane. The task is to determine the best set of parameters β_{qr} (or pK_a) and molar absorptivities ε_{qr} for a given sets of spectra, and a predefined protonation equilibria model hypothesis.

If 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, LH, LH_2, LH_3, ...etc., which have the general formula L_qH_r in a particular chemical model and are represented by n_c, the number of species, $(q, r)_i$, i=1, ..., n_c, where index i labels their particular stoichiometry, then the overall protonation (stability) constant of the protonated species, β_{qr}, may be expressed as

$$\beta_{qr} = \left[L_qH_r\right]/([L]^q[H]^r) = c/(l^q h^r)$$

where the free concentration [L]=l, [H]=h and $[L_qH_r]=c$. An acid–base equilibrium of the drug studied is described in terms of the protonation of the Brönstedt base L^{z-1} according to the equation $L^{z-1}+H^+ \approx HL^z$, characterized by the protonation constant

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

and in the case of a polyprotic species this is protonated to yield a polyprotic acid H_jL:

$$L^{z-} + H^+ \approx HL^{1-z};\ K_{H1}$$
$$HL^{1-z} + H^+ \approx H_2L^{2-z};\ K_{H2}$$

The subscript to K_H indicates the ordinal number of the protonation step. The direct formation of each protonated species from the base L^{z-} can be expressed by the overall reaction L^{z-1}+j $H^+ \approx H_jL^z$ and by the overall constant $\beta_{Hj}=K_{H1}K_{H2}...\ K_{Hj}$, where j denotes the number of protons involved in the overall protonation. For dissociation reactions realized at constant ionic strength, so-called "mixed dissociation constants" are defined as

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

Since each aqueous species is characterized by its own spectrum, for UV/VIS experiments and the ith solution measured at the jth wavelength, the Lambert–Beer law relates to the absorbance, $A_{i,j}$, which is defined as

$$A_{i,j} = \sum_{n=1}^{n_c} \varepsilon_{j,n}\ c_n = \sum_{n=1}^{n_c}\left(\varepsilon_{qr,j}\ \beta_{qr}\ l^q\ h^r\right)_n$$

where $\varepsilon_{qr,\ j}$ is the molar absorptivity of the L_qH_r species with the stoichiometric coefficients q, r measured at the jth wavelength and an optical pathlength equal to unity. The absorbance $A_{i,\ j}$ is an element of the absorbance matrix $\boldsymbol{A}$ of size (n_s×n_w) that is measured for n_s solutions with known total concentrations of two (i.e., n_z=2) basic components, c_L and c_H, at n_w wavelengths. A multicomponent spectra analyzing program can adjust β_{qr} and ε_{qr} for absorption spectra by minimizing the residual square sum function *RSS*, denoted here as $U(\boldsymbol{b})$,

$$U(\boldsymbol{b}) = \sum_{i=1}^{n_s}\sum_{j=1}^{n_w}\left(A_{\exp,i,j} - A_{\mathrm{calc},i,j}\right)^2$$
$$= \sum_{i=1}^{n_s}\sum_{j=1}^{n_w}\left(A_{\exp,i,j} - \sum_{k=1}^{n_c}\varepsilon_{j,k}c_k\right)^2 = \text{minimum}$$

where $A_{i,\ j}$ represents an element of the experimental absorbance response surface of size n_s×n_w, and the independent variables c_k are the total concentrations of the basic components c_L and c_H that are adjusted in n_s solutions. The unknown parameters are the best estimates for the protonation constants, $\beta_{qr,i}$, i=1, ..., n_c, which are adjusted by the regression algorithm. At the same time, a matrix of molar absorptivities ($\varepsilon_{qr,\ j}$, j=1, ..., $n_w)_k$, k=1, ..., n_c is estimated as non-negative reals, based on the current values of the protonation constants. For a set of current values of $\beta_{qr,i}$, the free concentrations of ligand l for each solution are calculated, as h is known from pH measurements. Then, the concentrations of all the species in the equilibrium mixture $[L_qH_r]_j$, j=1, ..., n_c are obtained; they represent n_s solutions of the matrix $\boldsymbol{C}$.

The least squares (LS) method does not ensure that the model is fully acceptable from the statistical and physicochemical points of view. One source of problems may be found in the components of a *regression triplet*:

$$\text{quality of fit} = f(\mathit{Data},\ \mathit{Model},\ \mathit{Method\ of\ estimation})$$

The LS method provides accurate estimates only when all assumptions about the data and about the regression model are fulfilled [3]. When some assumptions are not fulfilled, the LS method is inconvenient. The quality of the fit is usually defined using regression diagnostics, as the sum of squared residuals between the measured data and their computationally modeled representation. The least squares estimates $\boldsymbol{b}$ for the regression parameters $\boldsymbol{\beta}$ are obtained by finding the minimal length of the *residual vector* $\widehat{\boldsymbol{e}} = \boldsymbol{A} - \widehat{\boldsymbol{A}}_{\mathbf{p}}$, where $\widehat{\boldsymbol{A}}_{\mathbf{p}}$ is the *predictor vector*. When determining the statistical properties of random vectors $\widehat{\boldsymbol{A}}_{\mathbf{p}}$, $\widehat{\boldsymbol{e}}$, and $\boldsymbol{b}$, some basic assumptions are necessary for the least squares method to be valid [17]. (1) The regression parameters $\boldsymbol{\beta}$ are not bounded, although in chemometric practice, there are some restrictions on the parameters, based on their physical meaning. (2) The regression model is linear or nonlinear in its parameters, and an additive model for the measurement of errors is valid. (3) The matrix of nonrandom controllable values of the explanatory variable $\boldsymbol{X}$ has a column rank equal to m. (4) The mean value of the random errors ε_i is zero; $E(\varepsilon_i)$=0. (5) The random errors ε_i have constant and finite variance, $E\left(\varepsilon_i^2\right) = \sigma^2$, and the data are therefore said to be homoscedastic. (6) The random errors ε_i are uncorrelated and therefore cov(ε_i, ε_i)=E(ε_i, ε_i)=0. This corresponds to independence of the measured absorbances A_i. (7) The random errors ε_i have a normal distribution $N(0,\sigma^2)$.

Reliability of β_{qr} or pK_a estimates obtained by the goodness-of-fit test

Regression diagnostics detect and assess the quality and reliability of a regression model. The goal of diagnostics is twofold: to recognize important phenomena resulting from outliers rather than the bulk of the data, and to suggest appropriate remedies in order to find a better regression model. Regression diagnostics are performed to narrow the gap between theoretical assumptions and observed data. In contrast to robust regression, which solves this problem by dampening the effect of outliers, regression diagnostics identify the outliers and deal with them directly. They look for model misspecification, departure from the normality assumption and from homoscedasticity of the residuals, collinearity in the predictor variables and influential observations. Residual analysis, comprising numerical and graphical analysis of the ordinary and various derived residuals, is one of the most important parts of regression diagnostics [3]. A collection of statistics known as influence analysis measures how well a protonation model fits the multiwavelength and multivariate data, e.g., how well a regression model accounts for the variance of the response variable. Examination of data quality involves the detection of the *influential points*, which cause many problems in regression analysis by shifting the parameter estimates or increasing the variance of the parameters. The influential points may instead be classified according to data location as follows. (i) *Outliers*, which differ from the other points in value on the absorbance axis and are separated from the bulk of the data. These may distort statistics calculated from such a sample. Outliers must be detected and tested to determine whether they should be discarded before modeling. (ii) *High-leverage points*, also called *extremes*, which differ from the other points in value on the pH axis. (iii) *Both outliers and leverages*. Outlier identification by examination of the residuals is relatively simple, and can be done once the regression model has been constructed.

Graphical analysis of residuals

Residual analysis is based on examining residuals from a regression model via graphical and/or numerical diagnostics in order to check the quality of nonlinear models. A variety of residual plots, such as *the bar plot, box-and-whisker plot, dot plot, midsum plot, symmetry plot, kurtosis plot, differential quantile plot, quantile-box plot, frequency polygon, histogram, quantile plot, quantile-quantile plot, rankit plot, scatter plot*, and *autocorrelation plot*, have been widely used by Cook and Weisberg [18], Atkinson [19], Chatterjee and Hadi [20], Anscombe [21], Draper and Smith [22], Carrol and Ruppert [23] and others. The resulting graphs are used for goodness-of-fit tests and the identification of influential points, cf. page 289 in [4]. Systematic departures of residuals from randomness also indicate that the model is not satisfactory. The following plots seem to be the most important:

(1) The residual index plot provides an initial impression of the absorbance residuals using interactive computer graphics, enabling detection of outliers, detection of a trend in the residuals, detection of a sign change, and detection of an abrupt shift of level in the experiment. This scatter plot is also used to verify the normality and homoscedasticity assumptions for the residual. The ideal plot shows a horizontal band of points with constant vertical scatter from left to right. A similar analysis is performed by a scatter plot of residuals vs. independent variables and a scatter plot of the residual vs. the prediction, which indicate suspicious points that could be influential.

(2) The kernel estimation of the probability density plot and histogram detect an actual sample distribution.

(3) The rankit Q–Q plot has the quantile of the standardized normal distribution $u_{\mathrm{P}_\mathrm{i}}$ for $P_\mathrm{i}=i/(n+1)$ on the x-axis and the ordered residuals on the y-axis, i.e., increasingly ordered values of various types, but mostly classical residuals. To examine the normality of a residual distribution, the rankit plot, also called the normal probability plot, may be applied. Data points lying along a straight line indicate distributions of similar shape. The intercept of the line indicates a difference in location, while a slope shows a difference in scale. This plot enables the classification of a sample distribution according to its skewness, kurtosis and tail length. A convex or concave shape indicates a skewed sample distribution. A sigmoidal shape indicates that the tail lengths of the sample distribution differ from those of a normal one.

(4) The halfsum plot gives information about the symmetry of a distribution. For a symmetric distribution, the halfsum plot forms a horizontal line $y=M$ (median).

(5) The quantile-box plot is a universal tool for examining the statistical features of data: for symmetrical distributions, the sample quantile function has a sigmoid shape, whereas for an asymmetrical one the quantile function is convex or concave-increasing. A symmetric unimodal distribution contains individual boxes arranged symmetrically inside one another, and the value of relative skewness is close to zero. Outliers are indicated by a sudden increase in the quantile function outside the quartile F box, and the slope may approach infinity. Departure from the normal straight line indicates non-normality or model misspecification; an opposite curvature at the ends indicates long or short tails, while a convex or concave curvature is related to asymmetry.

(6) The autocorrelation trend plot detects one important violation of basic assumptions for least squares and checks for evidence of any serial process fluctuation dependence or trend in an observed time series. If the process is stationary, the trend in residuals does not depend on time.

Statistical analysis of residuals

The plots recommended are visual techniques for easy checking of some of the basic assumptions of the least squares method and the proposed model. Certain statistics provide a numerical measure of some of the discrepancies previously described. If the proposed model represents the data adequately, the residuals should form a random pattern that has a normal distribution $N(0,\ s^2)$ with the residual mean equal to zero, $E(\widehat{e}) = 0$, and the standard deviation of residuals $s(\widehat{e})$ being near to the noise $\boldsymbol{\varepsilon}$, i.e., near to the experimental error $s_{\text{inst}}(A)$. Systematic departures from randomness indicate that the model and parameter estimates are not satisfactory. Statistical analysis of residuals is the main diagnostic tool used to search for the "best" model when more than one is possible or proposed. The goodness-of-fit test analyzes the set of residuals, and examines the following criteria [4]. (1) The *residual bias* is the arithmetic mean of residuals $E(\widehat{e})$ and should be equal to zero $E(\widehat{e}) = 0$; all residual values lying outside the modified Hoaglin's inner bounds B_L* and B_U* (cf. page 81 in [17]) are considered to be outliers. (2) The *mean of absolute values of residuals* $E|\widehat{e}|$, and the square root of the residual variance $s^2(\widehat{e}) = U(\boldsymbol{b})/(n - m)$, known as the estimate of the *residual standard deviation*, $s(\widehat{e})$, should both be of the same magnitude as the instrumental error of the regressed variable absorbance A, i.e., $s_{inst}(A)$. Obviously it is also valid that $s(\widehat{e}) \approx s_{\text{inst}}(A)$. (3) The *residual skewness*, $g_1(\widehat{e})$, for a symmetric distribution of residuals should be equal to zero. (4) The *residual kurtosis*, $g_2(\widehat{e})$, for a normal distribution should be equal to 3. (5) The *determination coefficient D* calculated from the relationship $D = 1 - U(\boldsymbol{b}) / \sum_{i=1}^{n} \left(A_{\exp,i} - \overline{A}_{\exp}\right)^2$ multiplied by 100% is called the *regression rabat*, and is equal to the percentage of points which correspond to the proposed regression model. (6) The *Hamilton R-factor of relative fitness* is often used in the chemical laboratory, and is expressed by the relationship $R - factor = \sqrt{\mathrm{U(b)} \Big/ \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{A_i^2}}$. There is an empirical rule of a fitness classification with the use of the *Hamilton R-factor*: for a good fitness, the *Hamilton R-factor* reaches a value ≤1%, and for excellent fitness it is lower than 0.5%. (7) The *Akaike information criterion AIC* is appropriate for distinguishing between various models. It is defined by the relationship $AIC = -2L(b) + 2m$, or $AIC = n\,pt \ln\left[\frac{U(b)}{n}\right] + 2m$, where n is the number of data points and m is the number of estimated parameters. The best regression model is considered to be that for which this criterion reaches a minimal value. The most suitable model is the one which gives the lowest values for the mean quadratic error of prediction *MEP* and Akaike information criterion *AIC* and the highest value of the regression rabat *D*, but not all software is able to provide these as outputs.

Procedure used to build and test the protonation model

The adequacy of a proposed regression model with experimental data, and the reliability of the parameter estimates $\beta_{qr,j}$ or p$K_{a,j}$ found (denoted for the sake of simplicity as b_j, j=1, ..., m and $\boldsymbol{\varepsilon}_{ij}$, j=1, ..., n_{w}), may be examined by the goodness-of-fit test.

(1) The quality of the parameter estimates b_j, j=1, ..., m found is depends on their variances $D(b_j)$. An empirical rule is often used: parameter b_j differs significantly from zero when its estimate is greater than three standard deviations, $3\sqrt{D(b_j)} < |b_j|$, j=1, ..., m. Higher parameter variances can be caused by termination of a minimization process before reaching a minimum.

(2) The quality of the experimental data is examined by identifying influential points through the use of regression diagnostics.

(3) The quality of curve fit achieved, or the adequacy of the proposed model and m parameter estimates found with n values of experimental data, is examined by a goodness-of-fit test based on the statistical analysis of classic residuals. If the proposed model adequately represents the data, the residuals should form a random pattern with a normal distribution $N(0,\ s^2)$, with a residual mean equal to zero, $\overline{e} = 0$, and with the standard deviation of residuals $s(e)$ being near to noise, i.e., the experimental error of absorbance measured, $s(\widehat{e}) \approx s_{\text{inst}}(A)$. Systematic departures from randomness indicate that the model and parameter estimates are not satisfactory. Examinations of residual plots may be assisted by graphical analysis of the residuals.

Uncertainty in the estimated dissociation constants

The adequacy of a proposed regression model with experimental data and the reliability of the parameter estimates $\beta_{qr,j}$ or p$K_{a,j}$ found can be examined by the goodness-of-fit test. Direct results x_j from experimental and instrumental operations in a laboratory are always approximate, mainly because of the limited accuracy of measuring instruments. Results from a chemical analysis or physicochemical constants (e.g., the protonation or dissociation constant y) are calculated from several measured quantities x_1, ..., x_n by the function y=G(x_1, ..., x_n). The resulting approximate relationship for the variance $s^2(y)$ is formed from m sources of uncertainties, and each source has its own variance $\sigma^2(x_i)$. The following expression is termed the rule of propagation of absolute uncertainties:

$$s^2(y) = \sum_{i=1}^{m} s^2(x_i) + 2\sum_{i=1}^{m-1}\sum_{j=i}^{m} cov\ \left(x_i,\ x_j\right),$$

where cov(x_i, x_j) is a measure of the linear dependence between the two variables x_i and x_j.

Experimental

Chemicals and solutions

The drug silybin and other chemicals and solutions have been described previously [24].

Apparatus and pH-spectrophotometric titration procedure

The apparatus used and the pH-spectrophotometric titration procedure have been described elsewhere [1, 24].

Software used

Computation relating to the determination of dissociation constants was performed by regression analysis of the UV/VIS spectra using the SQUAD(84) [6], SPECFIT/32 [12] and INDICES [25] programs. Most of the graphs were plotted using ORIGIN 7.5 [26]. In order to create regression diagnostic graphs and compute regression-based characteristics, an algorithm was written in S-PLUS [27], and the Linear Regression module of the ADSTAT package [28] was used.

Supporting information available

Complete experimental and computational procedures, input data specimens and corresponding outputs (in numerical and graphical form) from both SQUAD(84) and SPECFIT/32 are available free of charge via the Internet at http://meloun.upce.cz in the block *DATA*.

Results and discussion

Silybin was chosen as a typical example of the acid drugs analyzed in our laboratory to demonstrate the reliability of the protonation model and the estimation of protonation constants because of two issues. The first is evaluating the protonation equilibria for the drug silybin in cases of strongly overlapping equilibria, as the difference between two consecutive dissociation constants is less than 3 (about 1.2 here). Such close equilibria are always difficult to evaluate and therefore the user needs to prove the reliability of each dissociation constant estimation. A distribution diagram of the relative concentrations of all of the variously protonated species demonstrates the overlapping protonation equilibria for three consecutive dissociation constants. The second issue concerns the small differences between the molar absorptivities of the variously protonated species within a spectrum. It may happen that nonlinear regression fails when the small differences in absorbance are of the same magnitude as the instrumental noise, $s_{\mathrm{inst}}(A)$.

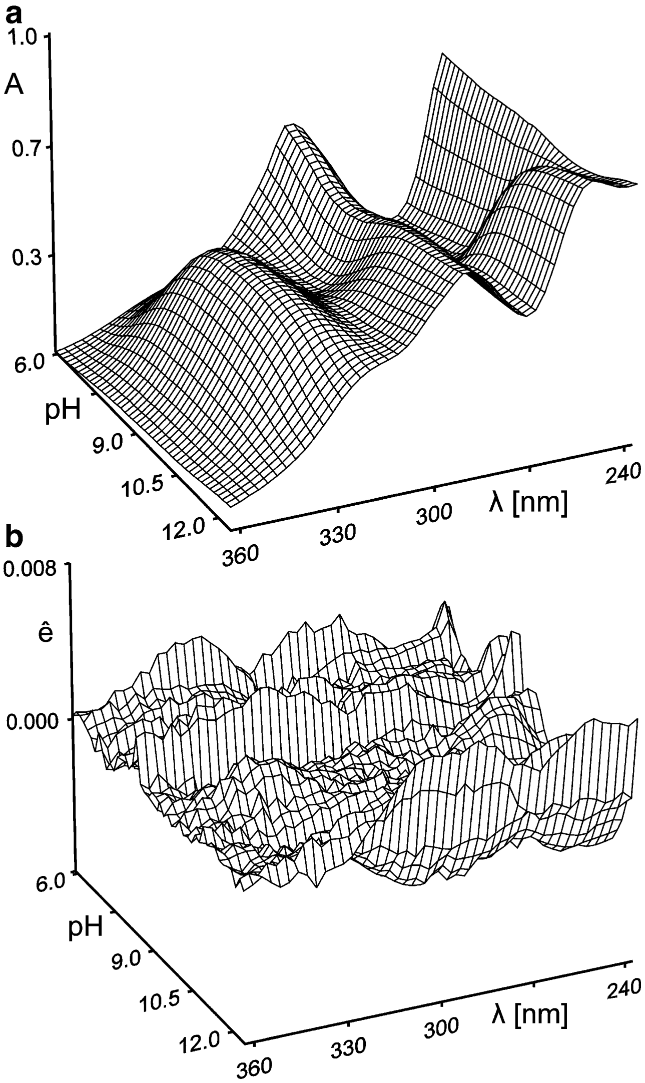


Fig. 1 **a** The 3-D absorbance response surface representing the dependence on pH at 25 °C of 33 absorption spectra for the protonation equilibria of silybine after removal of influential outlying spectra (S-PLUS). **b** The 3-D overall diagram for the residuals, representing the response surface showing the quality of the goodness-of-fit after the removal of influential outlying spectra (S-PLUS)

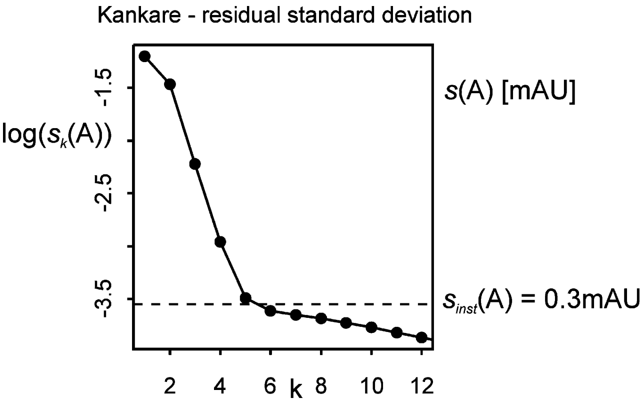


Fig. 2 Cattel's scree plot of the Kankare criterion $s(A)$ for the determination of the number of light-absorbing species in the mixture k^*=5 and the actual instrumental error of the spectrophotometer used $s_5{}^*(A)$=0.3 mAU (INDICES in S-PLUS)

Reliability of protonation model estimation

The number of light-absorbing species in the protonation equilibria

The proposed strategy for efficient protonation constant determination followed by spectral data treatment is presented for the protonation equilibria of the drug acid silybin [24]. pH-spectrophotometric titration enables absorbance response surface data (Fig. 1a) to be obtained for nonlinear regression analysis. The reliability of parameter estimates (for pK_a and ε) may be evaluated on the basis of a goodness-of-fit test of the set of residuals (Fig. 1b). The SQUAD(84) program [6] analytical process starts with data smoothing followed by a factor analysis based on the Kankare method using the INDICES procedure [25], as described in ref. [1]. The position of a breakpoint on the $s_k(A)$=f(k) curve in the scree plot is calculated and gives k*=5 with the corresponding co-ordinate $s_5{}^*(A)$=0.3 mAU, which also represents the instrumental error $s_{\mathrm{inst}}(A)$ of the spectrophotometer used (Fig. 2).

Least squares nonlinear regression of the absorbance response hyperplane

Four protonation constants and five molar absorptivities of silybin for 39 wavelengths constitute $4 + (5 \times 39) = 199$ unknown parameters, which are refined by the MR algorithm in the first run of the SQUAD(84) program on pH spectra (Fig. 3). In the second run, the NNLS algorithm makes a final refinement of all previously found parameter estimates, with all molar absorptivities kept non-negative. The reliability of the parameter estimates may be tested using SQUAD(84) diagnostics.

The first diagnostic indicates whether all parametric estimates β_{qr} and ε_{qr} have physical meanings and reach realistic values. As the standard deviations $s(\log\ \beta_{qr})$ of parameters $\log\ \beta_{qr}$ and $s(\varepsilon_{qr})$ of parameters ε_{qr} are significantly smaller than their corresponding parameter estimates (Table 1), all the variously protonated species are statistically significant. Figure 4 shows how the estimated molar absorptivities of all of the variously protonated species (ε_{L}, $\varepsilon_{\mathrm{LH}}$, $\varepsilon_{\mathrm{LH2}}$, $\varepsilon_{\mathrm{LH3}}$ and $\varepsilon_{\mathrm{LH4}}$) of silybin depend on wavelength. Some spectra overlap, and this may cause some resolution difficulties in a nonlinear regression approach.

The second diagnostic tests whether all of the calculated free concentrations of the variously protonated species on the distribution diagram have physical meaning, which proved to be the case (Fig. 4). The diagram shows that overlapping protonation equilibria exist here.

The third diagnostic, concerning the matrix of correlation coefficients in Table 1, proves that there is an absence of an interdependence between any pair of protonation constants of silybin except for species LH_1 vs. LH_2, and LH_3 vs. LH_4. The significant correlations of these two pairs may be explained by the protonation constants being too close, which is related to overlapping equilibria.

The fourth diagnostic, concerning the goodness-of-fit (Fig. 5), indicates influential points and outliers in the

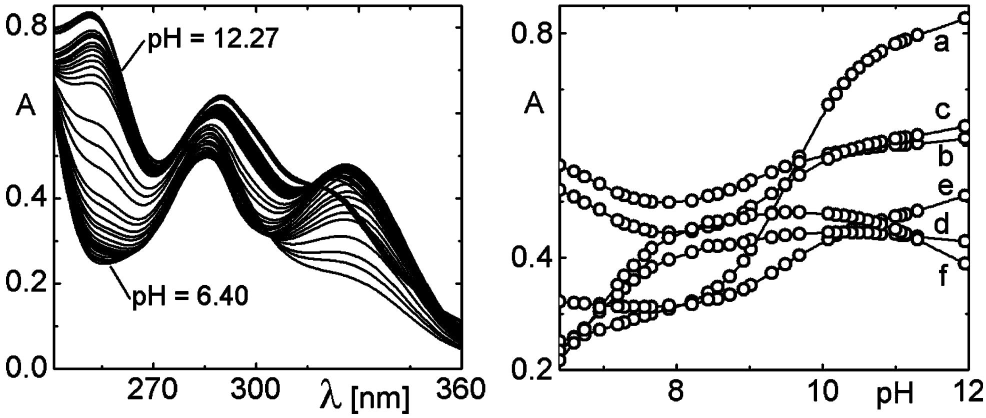


Fig. 3 The absorption spectra of silybin (*left*), the *A*–pH curve at selected wavelengths in dependence on pH (*right*) for dominant analytical wavelengths [nm]: *a*, 252.4; *b*, 285.3; *c*, 291.3; *d*, 318.2; *e*, 303.21; *f*, 327.1 (SPECFIT, ORIGIN)

Table 1 The best chemical model found for the protonation equilibria of silybin using double-checked nonlinear least squares regression analysis of multiwavelength and multivariate pH spectra with SQUAD (84) and SPECFIT/32 (**bold**) for n_s=20 (and 33) spectra measured at n_w=39 (and 43) wavelengths for n_z=2 basic components L and H forming n_c=5 variously protonated species

Estimated protonation constants				Partial correlation coefficients			
L_qH_r	log β_{qr}	pK_a	s (log β_{qr})	L_1H_1	L_1H_2	L_1H_3	L_1H_4
L_1H_1	11.501, **11.485**	11.501, **11.485**	0.008, **0.004**	1	–	–	–
L_1H_2	21.112, **21.108**	9.611, **9.623**	0.010, **0.002**	0.9105	1	–	–
L_1H_3	29.778, **29.776**	8.666, **8.668**	0.021, **0.008**	0.5191	0.7709	1	–
L_1H_4	36.676, **36.659**	6.898, **6.883**	0.022, **0.002**	0.4927	0.7473	0.9902	1
Determination of the number of light-absorbing species by factor analysis							
			SQUAD		SPECFIT		
Number of spectra measured n_s			20		33		
Number of wavelengths n_w			39		43		
Number of light-absorbing species k*			5		5		
Residual standard deviation $s_k*(A)$, [mAU]			0.3		Not estimated		
Goodness-of-fit test via statistical analysis of residuals							
Residual mean $\bar{e}$ [mAU]			3.50×10^{-17}		-1.83×10^{-8}		
Mean residual $\bar{e}$ [mAU]			0.67		0.52		
Standard deviation of residuals $s(e)$ [mAU]			1.01		0.65		
Residual skewness $\widehat{g}_1(e)$			0.29		−0.04		
Residual kurtosis $\widehat{g}_2(e)$			2.43		3.56		
Hamilton R-factor [%]			0.2		Not estimated		
ε (all species) vs. λ are			Realistic		Realistic		

The charges of the ions are omitted for the sake of simplicity. The resolution criterion and the reliability of the parameter estimates found is proven with goodness-of-fit statistics such as the residual square sum *RSS*, the standard deviation of absorbance after termination of the regression process, $s(A)$ [mAU], the residual standard deviation by factor analysis $s_k(A)$ [mAU], the mean residual $|\bar{e}|$, the residual standard deviation $s(e)$, the residual skewness $\widehat{g}_1(e)$ and the residual kurtosis $\widehat{g}_2(e)$, which proves that a Gaussian distribution applies, the Hamilton R-factor [%], and the presence of non-negative and realistic estimates for the calculated molar absorptivities of all of the variously protonated species ε vs. λ.

spectra set. The basic features and statistical properties of the residual set of each experimental spectrum are described by the symmetry and kurtosis of the residual distribution, their dispersion, and the presence or absence of outliers. The various *exploratory diagnostic plots* (EDPs) offer information about these statistical data features, some of which are shown in Fig. 5. When a sufficient number of points is available, estimating the probability density function and histogram can help to elucidate the structure of the sample. While the left part of Fig. 5 exhibits a symmetrical normal distribution, the right part shows a skewed non-normal asymmetric distribution, proving it to be an outlying spectrum.

Although the left Q–Q graph in Fig. 5 indicates rather longer tails than an ideal normal distribution, all points are well fitted with the straight line and therefore the residuals exhibit agreement between the residual distribution and the normal distribution. The right graph in Fig. 5 does not fit the straight line well, and these residuals therefore do not exhibit normality.

The halfsum plot indicates that most points in the left graph are in the confidence band around a median line

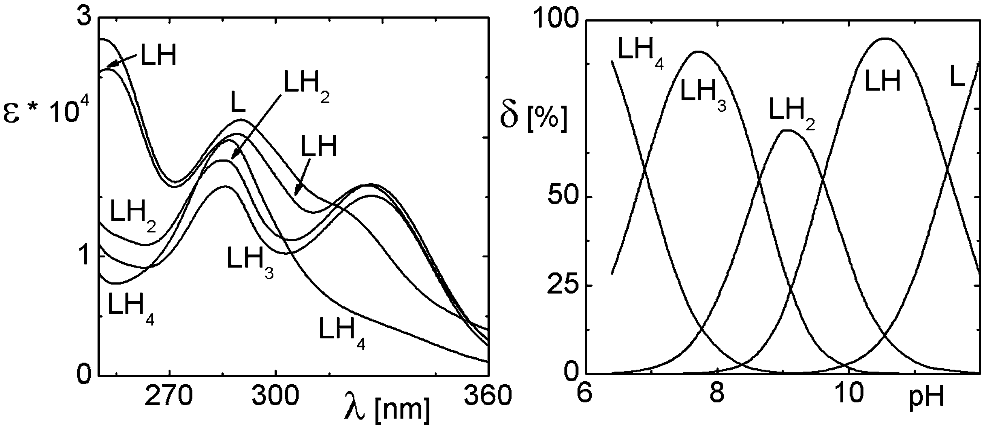
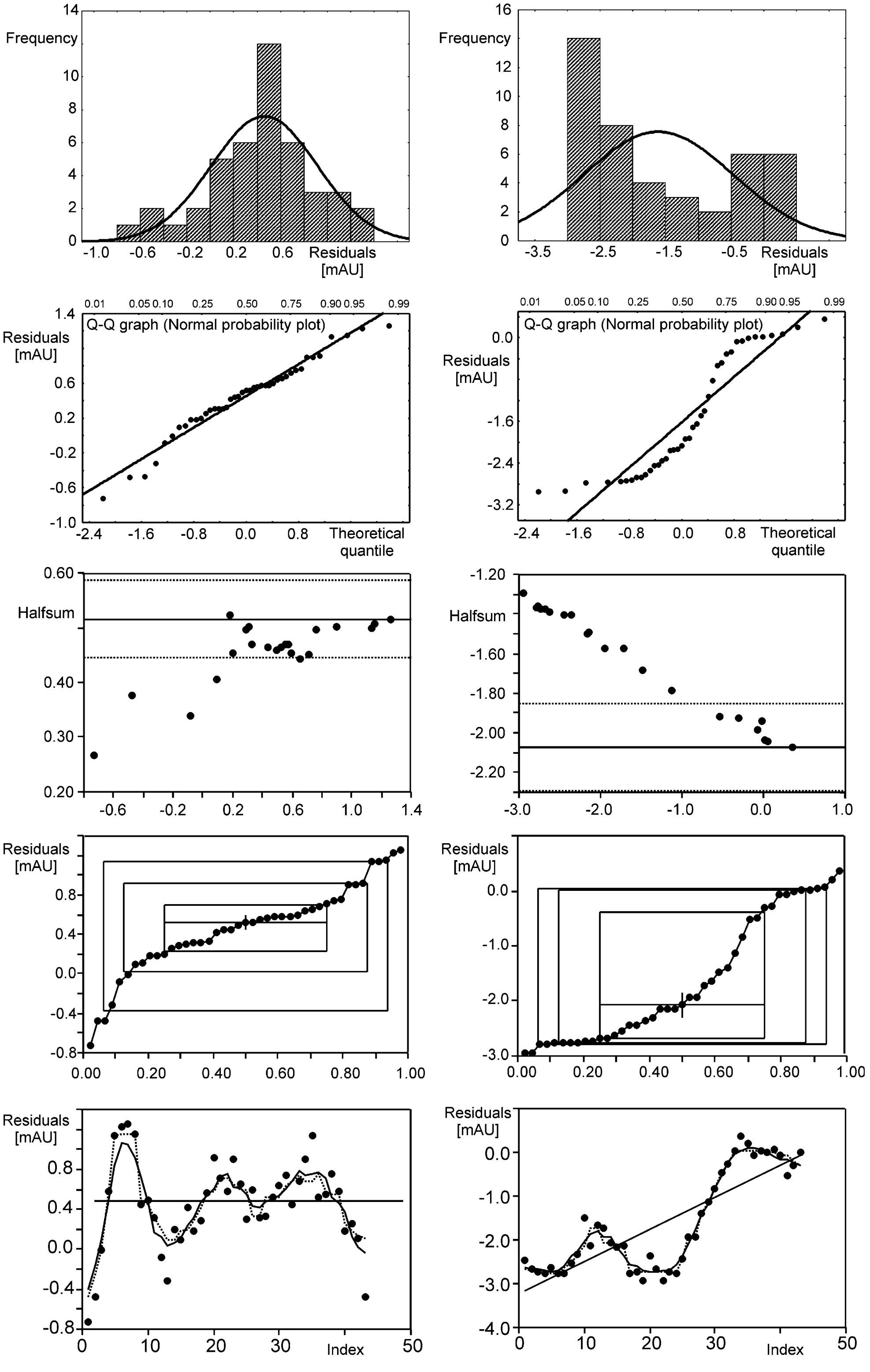


Fig. 4 The pure spectral profiles for the molar absorptivity vs. wavelength for the variously protonated species L, LH, LH_2, LH_3, LH_4 of silybin, and the distribution diagram for the relative concentrations of all of the variously protonated species L, LH, LH_2, LH_3, LH_4 of silybin in relation to pH; the charges of species are omitted for the sake of simplicity (SPECFIT, ORIGIN)

Fig. 5 Detecting and removing influential outlying spectra with the use of graphical exploratory data analysis involving goodness-of-fit test of residuals. From the set of 33 spectra, the two examples show either a good spectral fit (*left*) or a poor spectral fit (*right*): *1st row:* histogram and kernel estimation of the probability density; *2nd row:* the quantile-quantile (Q–Q) plot; *3rd row:* the halfsum plot; *4th row:* the quantile-box plot; *5th row:* the plot of a trend analysis, (QCEXPERT)



y=*M*, and that the residuals therefore exhibit a symmetric distribution. As most points in the right graph are not in the confidence band of a median line, this distribution deviates from a symmetrical one.

The quantile-box plots in both parts of Fig. 5 exhibit significant differences in terms of the shapes and symmetries of the boxes. While the left graph suggests a symmetric distribution of a Gaussian nature, the right graph shows an asymmetric distribution with some outliers.

The residual-index scatter plot is very informative, as it is able to indicate an autocorrelated trend in the residuals. The left graph does not prove any trend but a horizontal band of points with constant vertical scatter from left to right. No trend is proven with the statistical test. The right graph exhibits an obviously increasing trend in the residuals and this proves that this spectrum is strongly outlying among the set of spectra. This trend is proven to be statistically significant.

Statistics from the goodness-of-fit test prove that the $s_5(A)$ value of 0.3 mAU is closer to the standard deviation of absorbance when the minimization process terminates; $s(A)$=1.01 mAU (SPECFIT 0.65 mAU) when the outlying spectra have been removed (Fig. 6). Outlying spectra in the original set of 45 spectra are detected with rectangles in the left graph of Fig. 6, and the right graph shows the fitness when the outlying spectra are removed. Numerical values of statistical measures of the residuals now indicate very good fitness, and also prove that the minimum of the elliptic hyperparaboloid U was reached: the residual mean $\bar{e} = 3.50 \times 10^{-17}$ (SPECFIT -1.83×10^{-8}) proves that there is no bias or systematic error in the spectra fitting. The mean residual $|\bar{e}| = 0.67$ (SPECFIT 0.52) mAU and the residual standard deviation $s(e)$=1.01 (SPECFIT 0.65) mAU have sufficiently low values. The standard deviation of absorbance $s(A)$ after termination of the minimization process is always better than 2 mAU, and the proposal of a good chemical model and reliable parameter estimates are thus proven. The skewness $\hat{g}_1(e) = 0.29$ (SPECFIT −0.04) is quite close to zero and proves the symmetric distribution of the set of residuals, while the kurtosis $\hat{g}_2(e) = 2.43$ (SPECFIT 3.56) is close to 3, proving that a Gaussian distribution applies.

The fifth diagnostic, the spectral deconvolution in Fig. 7, shows the deconvolution of the experimental spectrum into spectra for the individual variously protonated species, to examine whether the experimental design is efficient.

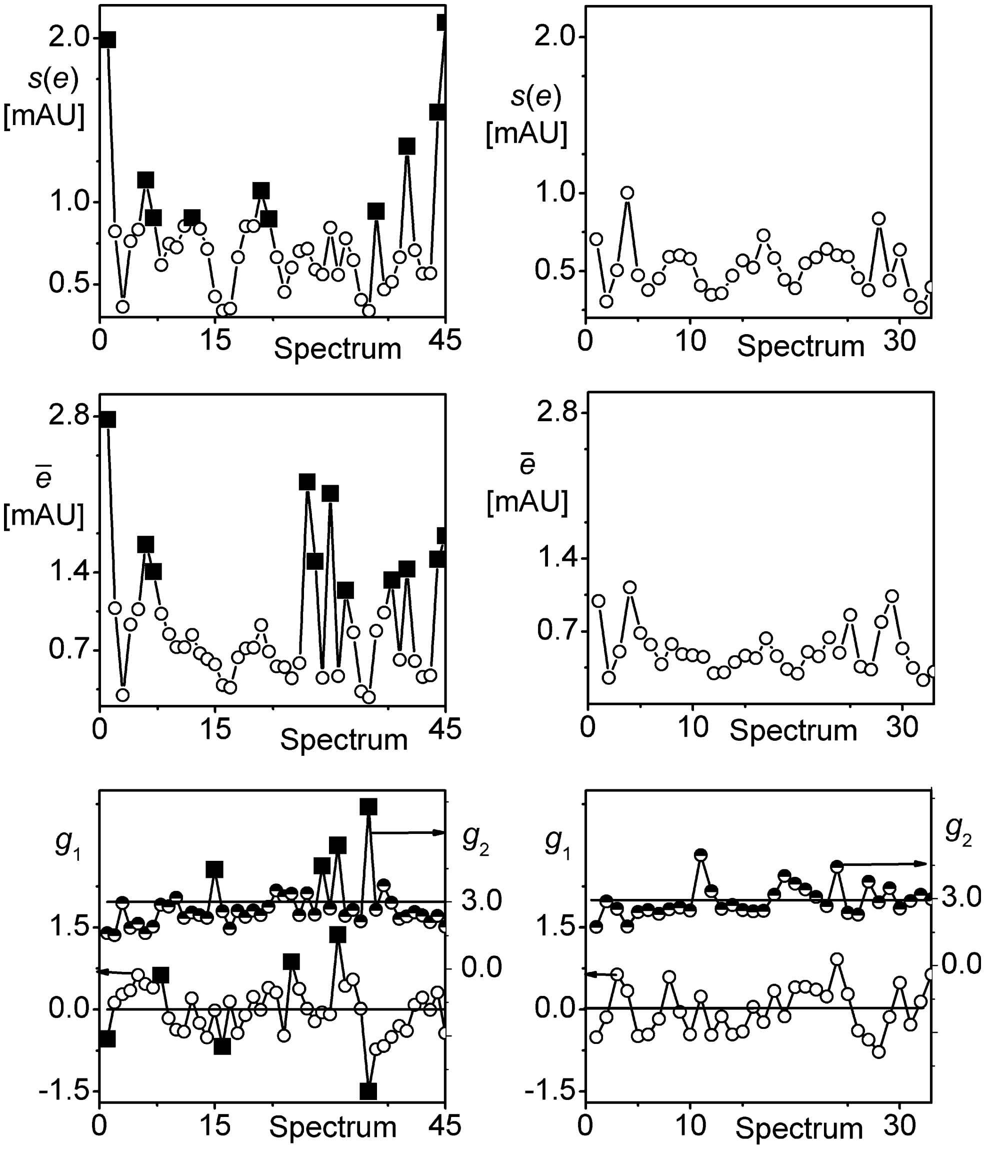


Fig. 6 Detecting and removing influential outlying spectra using a goodness-of-fit test. Spectral fitness achieved before (*left*) and after (*right*) removing outliers. *Rectangles* indicate outliers: *1st row:* the plot of the residual standard deviation $s(e)$; *2nd row:* the mean residual $|\bar{e}|$; *3rd row:* test of residual distribution symmetry using skewness g_1 and kurtosis g_2; (SPECFIT, QCEXPERT, ORIGIN)

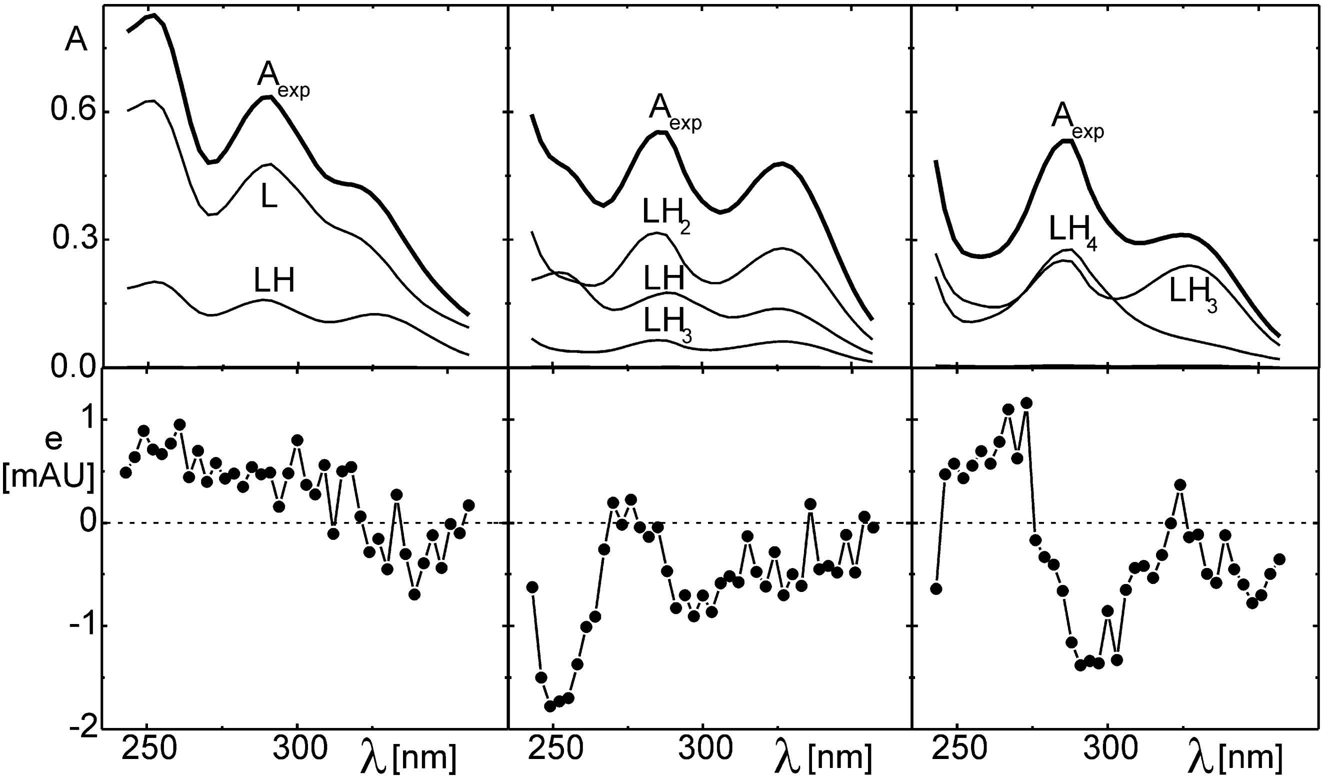


Fig. 7 Deconvolution of the experimental absorption spectrum of silybin for 39 wavelengths into spectra for the individual variously protonated species L, LH, LH_2, LH_3, LH_4 in solution (*above*), and the statistical analysis of the residuals (*below*) from each particular absorption spectrum for a selected value of pH equal to: (*a*) 11.96, (*b*) 9.31 and (*c*) 6.95. The charges of the species are omitted for the sake of simplicity. (SQUAD, ORIGIN)

Spectral deconvolution seems to be quite a useful tool when proposing a strategy for efficient experimentation. Such a spectrum provides sufficient information for a regression analysis that monitors at least two species in equilibrium, where none of them is a minor species. A 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 is prevalent in solution, the spectrum yields quite poor information for a regression analysis, and the parameter estimate is rather unsure and definitely not reliable enough. The upper part of Fig. 7 shows a spectral deconvolution and the lower part a plot of the residual scatter vs. wavelength. This graph detects the quality of curve fitting and also proves the reliability of fitting the actual spectrum in question.

Uncertainty in the estimated dissociation constants

Uncertainty in pK$_a$ caused by drifts in pH measurement

To gauge the uncertainty arising from pH measurement, the pH-meter was adjusted via standard buffers at 4.006±0.005 and 9.180±0.005, and then a third buffer of value 6.865± 0.005 was measured for six hours and the time drift in the pH values monitored (Fig. 8). A decrease in pH value by 0.002 pH units every six hours does not appear to be significant during a 90-minute measurement of the spectra set. It was found that for a given experiment the uncertainty in the pH meter adjustment is ±0.004, the uncertainty due to time drift is ±0.007, and the uncertainty in the pH value of the pH standard is ±0.005. Based on the propagation of errors law, the uncertainty caused by the pH measurements is equal to ±0.009.

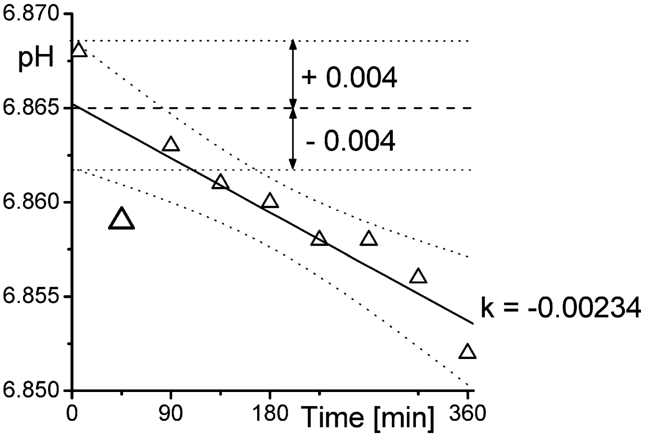


Fig. 8 Uncertainty in pK_a caused by drifts in pH measurement over time

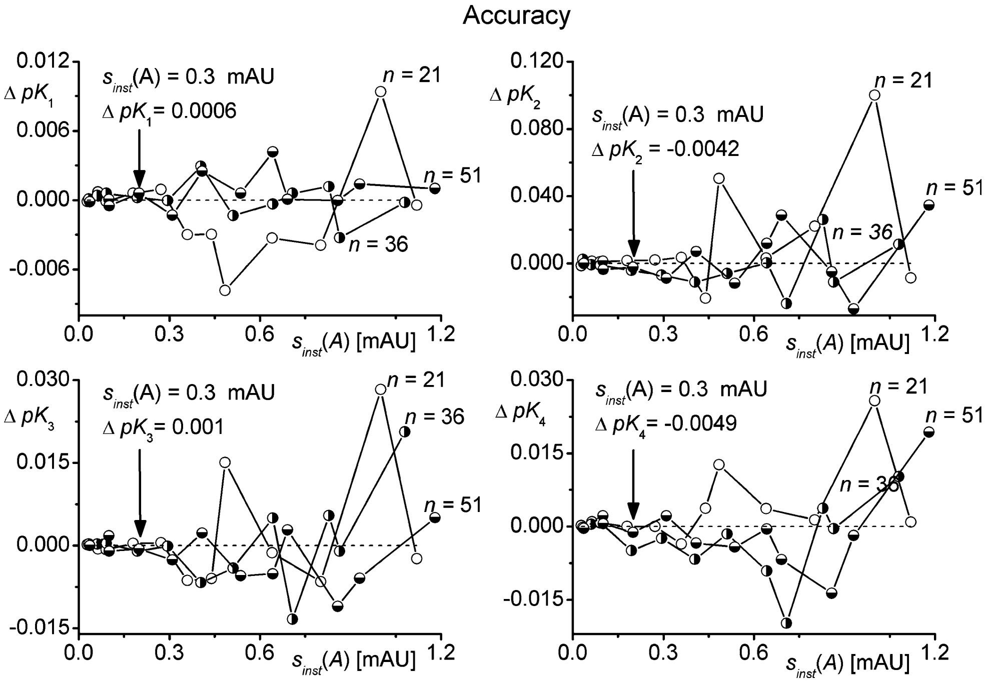


Fig. 9 Dependence of the accuracy of the pK_a estimates on the instrumental error of the spectrophotometer used $s_{\text{inst}}(A)$

Uncertainty in pK_a caused by drifts in spectral measurement

The uncertainty estimation procedure is applied to dissociation constant determination for silybin, which has three close dissociation constants due to overlapping protonation equilibria $|\mathrm{p}K_i - \mathrm{p}K_{i+1}| < 3$ and a distant protonation equilibrium, i.e., for an ionic strength I=0.03 the dissociation constants were found to be p$K_{a,1}$=6.898, p$K_{a,2}$=8.666, p$K_{a,3}$=9.611, p$K_{a,4}$=11.501 (Fig. 4). Moreover, two differently protonated species have very similar absorption bands. The known values for the molar absorption coefficients of all the variously protonated species and the values of four dissociation constants were used to generate the absorption spectra. A set of precise values for the absorbance at 39 wavelengths was then loaded with random

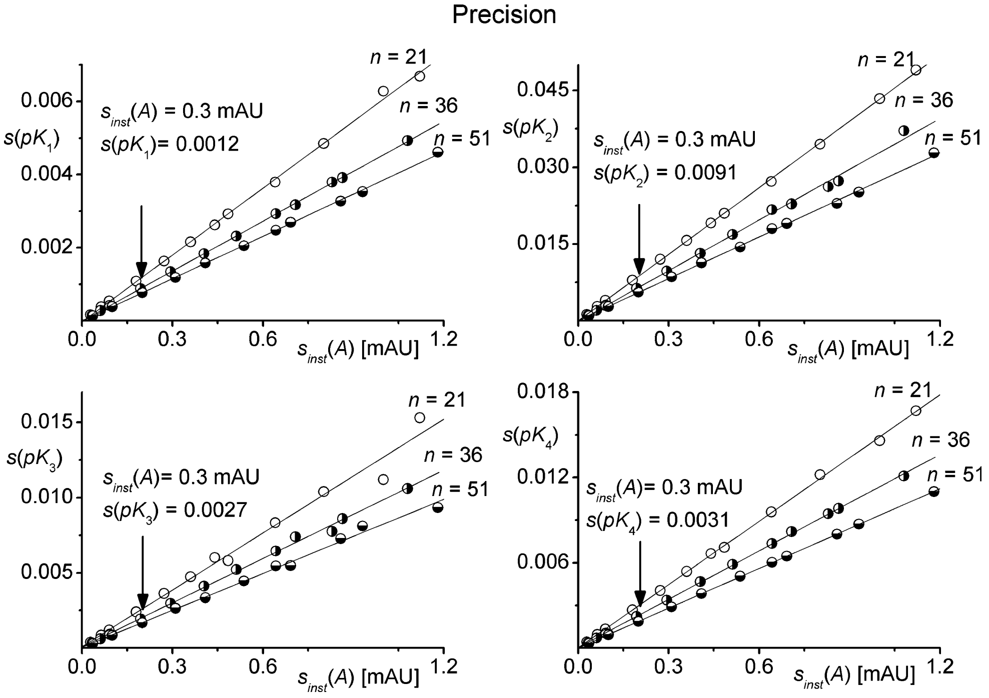


Fig. 10 Dependence of the precision of the pK_a estimates on the instrumental error of the spectrophotometer used $s_{\text{inst}}(A)$

Table 2 The accuracy of the pK_i estimates, investigated via the bias ΔpK_i, i=1, ..., 4, as expressed in the linear regression model $\Delta pK_i = \beta_0 + \beta_1 s_{inst}(A)$ as a function of the instrumental standard deviation $s_{inst}(A)$

	Intercept $b_0(s)$	t_{exp} and H_0: b_0=0 is	Slope $b_1(s)$	t_{exp} and H_0: b_1=0 is	Regression model is
21 spectra, pK_{a1}	-1.6×10^{-3} (2.0×10^{-3})	−0.78, accepted	1.65 (3.52)	0.47, accepted	not significant
pK_{a2}	-4.8×10^{-3} (1.5×10^{-2})	−0.33, accepted	38.70 (25.17)	1.54, accepted	not significant
pK_{a3}	-2.2×10^{-3} (4.8×10^{-3})	−0.46, accepted	8.54 (8.25)	1.04, accepted	not significant
pK_{a4}	-9.5×10^{-4} (3.4×10^{-3})	−0.28, accepted	10.56 (5.94)	1.78, accepted	not significant
36 spectra, pK_{a1}	6.4×10^{-4} (7.3×10^{-4})	0.88, accepted	−1.2 (1.25)	−0.96, accepted	not significant
pK_{a2}	-5.3×10^{-3} (6.4×10^{-3})	−0.82, accepted	6.67 (10.94)	0.61, accepted	not significant
pK_{a3}	-3.7×10^{-3} (3.9×10^{-3})	−0.95, accepted	8.83 (6.73)	1.31, accepted	not significant
pK_{a4}	-3.6×10^{-3} (3.9×10^{-3})	−0.92, accepted	2.25 (6.61)	0.34, accepted	not significant
51 spectra, pK_{a1}	4.2×10^{-5} (7.2×10^{-4})	0.06, accepted	1.32 (1.18)	1.13, accepted	not significant
pK_{a2}	-4.4×10^{-3} (8.5×10^{-3})	−0.52, accepted	12.96 (13.90)	0.93, accepted	not significant
pK_{a3}	-5.5×10^{-4} (2.4×10^{-3})	−0.23, accepted	−2.16 (3.85)	−0.56, accepted	not significant
pK_{a4}	-2.4×10^{-3} (4.0×10^{-3})	−0.60, accepted	3.5 (6.46)	0.54, accepted	not significant

In the interval of $s_{inst}(A)$ investigated, from 0.1 to 1.0 mAU, both of the parameter estimates, the intercept β_0 and the slope β_1, are statistically tested using a Student t-test of the null hypotheses H_0: b_0=0 and H_0: b_1=0. The significance of the proposed linear regression model is then proven with the Fisher-–Snedecor F-test.

errors generated for the preselected standard deviation of absorbance $s(A)$ and that were equal to the instrumental noise of the spectrophotometer used, $s_{inst}(A)$.

The first and most important experimental parameter affecting the accuracy and precision of the estimated dissociation constants is the value of the instrumental noise of the absorbance measurement, $s_{inst}(A)$. The value of this noise was therefore generated and varied within an interval of 0.1–1.0 mAU (Figs. 9 and 10). The second important parameter was the spectral sample size n, which was varied here, taking values of n=21, 36 and 51 using a pH interval of 5–13. The noise generated was added to a precisely calculated matrix of the spectra for 39 wavelengths that has a Gaussian distribution of the random error of a zero mean (actually 10^{-20}), and a standard deviation equal to the preselected value for the noise. When testing the statistical significance of the estimated parameters, the critical value of the Student t-test was t_{crit}=2.23 (Table 2).

Simulated spectra were treated using two regression programs, SQUAD(84) and SPECFIT/32. Both programs yielded similar pK_i estimates and similar curves for the molar absorptivities vs. wavelength. The accuracy of the pK_i estimates was investigated via the bias ΔpK_i, i=1, ..., 4, as expressed in the linear regression model $\Delta pK_i = \beta_0 +$

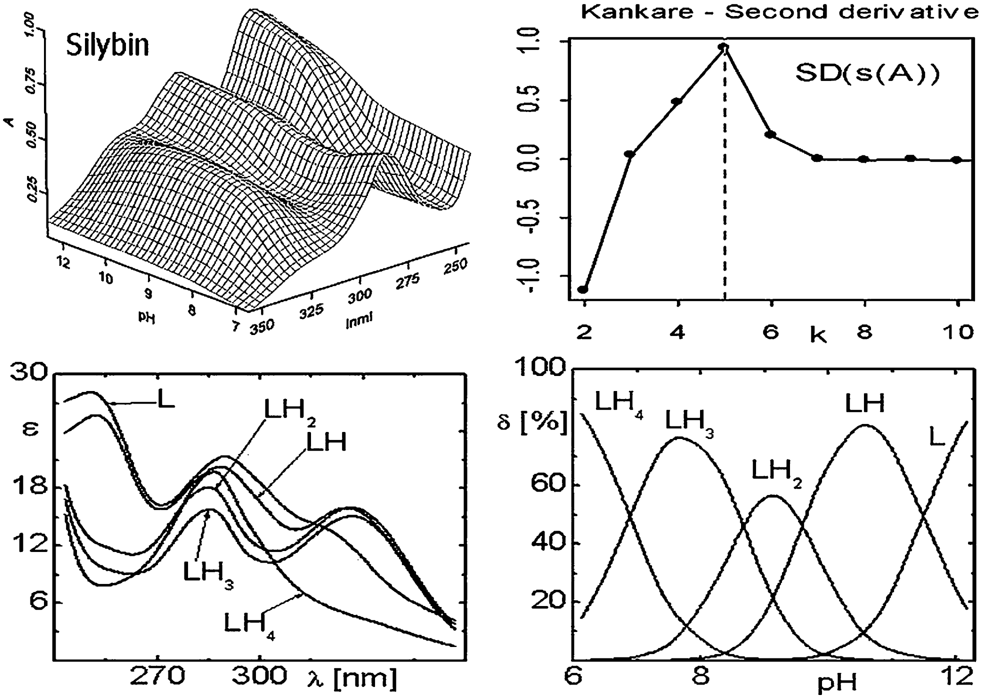


Fig. 11 Spectrophotometric determination of pK_a

Table 3 Estimated accuracies and precisions (uncertainties) of individual dissociation constants pK_i, i=1, ..., 4, for the instrumental standard deviation $s_{\mathrm{inst}}(A)$=0.3 mAU

Accuracy				
Bias in pK_i	p$K_{\mathrm{a}1}$	p$K_{\mathrm{a}2}$	p$K_{\mathrm{a}3}$	p$K_{\mathrm{a}4}$
	±0.002	±0.014	±0.004	±0.010
Precision				
Uncertainty in pK_i	p$K_{\mathrm{a}1}$	p$K_{\mathrm{a}2}$	p$K_{\mathrm{a}3}$	p$K_{\mathrm{a}4}$
	±0.001	±0.010	±0.003	±0.003

Table 4 The drift during analytical operations exerts no significant influence

Analytical operation	Value and uncertainty
Weighting	50.0±0.1 mg
Pipet 1	25.00±0.03 ml
Pipet 2	10.00±0.01 ml
Volumetric flask	250.0±0.3 ml
Microburette	1.250±0.001 ml
Purity of drug	97.5±0.1 %

Analytical operations add uncertainty to the concentration of a drug acid and so they will only affect the uncertainty in the estimated molar absorptivities of the variously protonated species. The law of uncertainty propagation leads us to conclude that the uncertainty in the molar absorptivity is about ±0.29%.

$\beta_1 s_{\mathrm{inst}}(A)$ as a function of the instrumental standard deviation $s_{\mathrm{inst}}(A)$. In the investigated interval of $s_{\mathrm{inst}}(A)$, from 0.1 to 1.0 mAU, neither of the parameter estimates b_0 and b_1 are statistically significant, and so the dissociation constants pK_i are accurate for all values of $s_{\mathrm{inst}}(A)$.

Precision was examined based on the estimated standard deviation of the dissociation constant s(pK) of the calculated ith dissociation constant pK_i in relation to the noise level $s_{\mathrm{inst}}(A)$, expressed in the linear regression model $s(\mathrm{p}K)=\beta_0+\beta_1 s_{\mathrm{inst}}(A)$ (Fig. 10). In all cases the intercept was statistically insignificant, and it was found that increasing the sample size decreased the value of the slope (Fig. 10). While parameter pK_2 is well-conditioned in the regression model, as the pK_2 is sensitive enough to the absorbance noise, the other three parameters (pK_1, pK_3 and pK_4) are less sensitive to the magnitudes of random errors, and are therefore badly conditioned in the regression model (Table 3).

The noise level $s_{\mathrm{inst}}(A)$ has an influence on the precision of the estimated parameters pK_i when closely overlapping equilibria exist. This is the case with the LH_2 and LH_3 species, which exhibit overlapping spectra with LH, and therefore the estimation of pK_2 is more difficult and the precision of the estimation depends on the noise level of the spectral data. Table 3 shows the bias in the accuracies and uncertainties of the dissociation constants for the noise level $s_{\mathrm{inst}}(A)$=0.3 mAU, which corresponds to common experimental data.

Uncertainties in pK$_a$ caused by drifts during analytical operations

The uncertainties in pK_i arising from drifts during analytical operations have no significant influence, and therefore were not propagated in the total uncertainty of pK_i. Analytical operations add uncertainty to the concentration of the acid drug and will thus only affect the uncertainties in the estimated molar absorptivities of the variously protonated species. The law of uncertainty propagation leads us to conclude that the uncertainty in the molar absorptivity is about ±0.29% (Table 4).

Uncertainties in the overall pK$_a$ values

Based on the aforementioned uncertainties and experimental noise level $s_{\mathrm{inst}}(A)$=0.3 mAU, the resulting uncertainties

Table 5 The estimated dissociation constants in the protonation model (L, LH, LH_2, LH_3, LH_4) of silybin for various ionic strengths at 25 °C are proven with the goodness-of-fit test

	Ionic strength							
	0.011		0.032		0.089		0.128	
	SQUAD	SPECFIT	SQUAD	SPECFIT	SQUAD	SPECFIT	SQUAD	SPECFIT
p$K_{\mathrm{a},1}$	6.871(44)	6.871(4)	6.898(22)	6.897(2)	6.858(40)	6.857(3)	6.841(43)	6.839(5)
p$K_{\mathrm{a},2}$	8.938(43)	8.919(24)	8.666(21)	8.668(12)	8.549(38)	8.533(22)	8.574(41)	8.551(24)
p$K_{a,3}$	9.721(18)	9.714(9)	9.611(10)	9.612(4)	9.579(19)	9.577(6)	9.556(19)	9.551(7)
p$K_{\mathrm{a},4}$	11.644(9)	11.640(8)	11.501(8)	11.501(7)	11.666(15)	11.661(12)	11.622(14)	11.621(12)
Goodness-of-fit test, $s_k(A)$ [mAU]=0.25								
RSS [mAU]	1.29	1.28	0.6	0.58	1.51	1.48	1.97	2.01
$\bar{e}$ [mAU]	1.05	1.04	0.67	0.66	0.99	0.98	1.11	1.13
$s(A)$ [mAU]	1.6	1.35	1.01	0.86	1.61	1.38	1.84	1.6

in the estimated dissociation constants pK_i were calculated (Table 5). This also means that when the standard deviation for the dissociation constant s(pK), calculated using nonlinear regression, is larger than these values, the noise level $s_{\mathrm{inst}}(A)$ in the experimental data is larger than the supposed value 0.3 mAU.

Conclusions

The reliability of the dissociation constants for the acid drug silybin can be proven be performing goodness-of-fit tests on the absorption spectra measured at various pH values (Fig. 11). Goodness-of-fit tests for various regression diagnostics enabled the reliability of the parameter estimates to be determined. When drugs are poorly soluble, pH-spectrophotometric titration may be used along with nonlinear regression of the absorbance response surface data instead of potentiometry to determine the dissociation constants. Regression diagnostics represent procedures for examining the *regression triplet* (*data, model, method*) in order to check (a) the data quality for a proposed model, (b) the model quality for a given set of data, and (c) whether all of the assumptions of least squares are fulfilled.

Acknowledgments The financial support of the Grant Agency IGA (Grant No NR9055-4/2006) and of the Czech Ministry of Education (Grant No MSM253100002) is gratefully acknowledged.

References

1. Meloun M, Bordovská S, Syrový T, Vrána A (2006) Anal Chim Acta 580:107–121
2. Maeder M, Neuhold Y-M, Puxty G, Gemperline P (2006) Chemometr Intell Lab Syst 82:75–82
3. Meloun M, Militký J, Hill M, Brereton RG (2002) Analyst 127:433–450
4. Meloun M, Militký J, Forina M (1994) Chemometrics for analytical chemistry, vol 2. PC-aided regression and related methods. Ellis Horwood, Chichester, UK
5. Leggett DJ (ed)(1985) SQUAD. In: Computational methods for the determination of formation constants. Plenum, New York, pp 99–157, 291–353
6. Meloun M, Javůrek M, Havel J (1986) Talanta 33:513–524
7. Leggett DJ, McBryde WAE (1975) Anal Chem 47:1065–1070
8. Leggett DJ (1977) Anal Chem 49:276–281
9. Gampp H, Maeder M, Mayer Ch J, Zuberbühler A (1985) Talanta 32:95–101
10. Gampp H, Maeder M, Meyer Ch J, Zuberbühler A (1985) Talanta 32:251–264
11. Gampp H, Maeder M, Meyer Ch J, Zuberbühler A (1985) Talanta 33:943–951
12. Spectrum Software Associates (2004) SPECFIT/32. Spectrum Software Associates, Marlborough, MA (see http://www.bio-logic.info/rapid-kinetics/specfit.html, last accessed 16th November 2006)
13. Meloun M, Javůrek M, Högfeldt E (1988) Chem Scripta 28:323–329
14. Meloun M, Havel J, Högfeldt E (1988) Computation of solution equilibria. Ellis Horwood, Chichester, UK
15. Meloun M, Havel J (1984) Computation of solution equilibria, 1. Spectrophotometry, Folia Fac. Sci. Nat. Univ. Purkyn. Brunensis (Chemia), Brno, XXV
16. Meloun M, Havel J (1985) Computation of solution equilibria, 2. Potentiometry, Folia Fac. Sci. Nat. Univ. Purkyn. Brunensis (Chemia), Brno, XXVI
17. Meloun M, Militký J, Forina M (1992) Chemometrics for analytical chemistry, vol 1. PC-aided statistical data analysis. Ellis Horwood, Chichester, UK
18. Cook RD, Weisberg S (1982) Residuals and influence in regression. Chapman & Hall, London
19. Atkinson AC (1985) Plots, transformations and regression: an introduction to graphical methods of diagnostic regression analysis. Clarendon, Oxford
20. Chatterjee S, Hadi AS (1988) Sensitivity analysis in linear regression. Wiley, New York
21. Anscombe FJ (1961) Proc Fourth Berkeley Symp Math Statist Prob I:1–36
22. Draper NR, Smith H (1966) Applied regression analysis, 1st edn. Wiley, New York
23. Carrol RJ, Ruppert D (1988) Transformation and weighting in regression. Chapman and Hall, New York
24. Meloun M, Burkoňová D, Syrový T, Vrána A (2003) Anal Chim Acta 486:125–141
25. Meloun M, Syrový T, Vrána A (2003) Anal Chim Acta 489:137–151
26. OriginLab Corporation (2006) ORIGIN. OriginLab Corporation, Northampton, MA
27. Insightful Corp. (2006) S-PLUS. Insightful Corp., Seattle, WA, (see http://www.insightful.com/products/splus, last accessed 16th November 2006)
28. TriloByte Statistical Software Ltd. (2006) ADSTAT 1.25, 2.0, 3.0 (Windows 95). TriloByte Statistical Software Ltd., Pardubice, Czech Republic