

METHODOLOGY

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Using a Matlab Implemented Algorithm for UV-vis Spectral Resolution for pK_a Determination and Multicomponent Analysis

Yotam Gonen^{1,2} and Giora Rytwo^{2,3}

¹The Faculty of Agriculture, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem, Rehovot 76100, Israel. ²Tel Hai Academic College, Environmental Sciences Department, Upper Galilee 12210, Israel.

³MIGAL, Galilee Technological Center, Kiryat Shmona, Israel. Email: yotamg@telhai.ac.il

Abstract: A Matlab implemented computer code for spectral resolution is presented. The code enables the user to resolve the UV-visible absorption spectrum of a mixture of up to 3 previously known components, to the individual components, thus, evaluating their quantities. The resolving procedure is based on searching the combination of the components which yields the spectrum which is the most similar (minimal RMSE) to the measured spectrum of the mixture. Examples of using the software for pK_a value estimation and multicomponent analysis are presented and other implementations are suggested.

Keywords: UV-vis, spectral resolution, matlab, computer code

Analytical Chemistry Insights 2009:4 21–27

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

UV-vis spectroscopy is a cheap and fast technique for quantification of solutes, mainly organic molecules. According to the assumptions of Beer-Lambert equation, if there are several solutes with overlapping spectra, and no mutual interaction affects the spectrum profile; the result would be a superposition of the individual spectra of the solutes.¹ Therefore, when dealing with a sample that may contain two or more solutes with overlapping spectra, there is a need to deconvolute the superimposed spectra.² Such spectral resolution may be essential when actual separation of the absorbing species is impossible or may affect the result, as in the case of tautomeric equilibria, short-lived intermediates and low stability complexes.² Spectral deconvolution was also suggested in several other applications as in identification and enumeration of the water born parasite *Cryptosporidium Parvum* Oocysts,³ analysis of blood platelets size distribution and chemical composition⁴ and pK_a value determination.⁵

The problem of overlapping spectral bands may be addressed in two main ways:

- a. Mathematical deconvolution of the superimposed spectra to single bands represented as mathematical function (usually Gaussian function) and assigning the bands to the analytes.² This method is very useful when the components of the mixture are unknown but may generate artifacts by adding inexistent components while lowering the root mean square error (RMSE). In addition it should be kept in mind that since electronic absorption spectra usually appears as a series of broad and smooth bands rather than as sharp peaks,⁶ the representation of the single band as a mathematical function is just an approximation and therefore presents another cause for errors.² Several commercial software products are available that provide such spectral deconvolution as “Specfit” (<http://www.pha.jhu.edu/~gak/specfit.html>) and “Peakfit” Sysat softwer Inc. (<http://www.systat.com/products/PeakFit/>).
- b. The second approach, which is useful when having previous knowledge of the identity of the absorbing components, and seeking for their concentration, is selection of the best combination of the several previously known solutes which yields the best fit to the measured spectrum. The drawback

of the method is that it is essential to know the spectra of the absorbing components. However, if such knowledge is available this method is less susceptible for artifacts.

In previous studies a preliminary version of a similar program was prepared using electronic worksheet program.⁷ However, the program was slow and only two components were considered. In this study we present a program based on a Matlab code that performs such deconvolution for up to three base spectra. The main principle is systematically searching the combination of the three spectra of the components that yield the spectrum which is most similar to that of the mixture.

Assumptions

The assumptions required for such spectral resolution are:

1. Individual spectra are not affected by interactions with other sample ingredients (in other words—superposition can be assumed).¹
2. Overlapping spectra are not completely identical, at least in part of the measured range.
3. Detection limits allow measuring the two or three solutes in the same dilution and still obtain acceptable O.D. values. Thus, optical densities of all components in the measured sample are of the same order of magnitude.

The Matlab programming environment is specifically designed for manipulating vectors and matrices. Since spectra can be easily displayed as a row vector in which the position of each number is related to the wavelength (or wave number) and value is related to the measured optical density (O.D.) in each wavelength, the program yields accurate and very fast deconvolutions.

Materials and Methods

Calculation scheme

The program accepts 4 row vectors, each representing a spectrum: the mixture spectra, with the unknown amounts of up to three components (*mix*), and spectra of the individual components (denoted as *X*, *Y* and *Z*). The spectra are introduced as row vectors in which the place in the vector represents the wave length and the value represents the optical density.



Assuming $X(\lambda)$, $Y(\lambda)$ and $Z(\lambda)$ are the spectrum vectors of solutes X , Y and Z respectively, then the overall absorbance vector would be:

$$\text{Mix}(\lambda) = AX(\lambda) + BY(\lambda) + CZ(\lambda) \quad (1)$$

where $\text{Mix}(\lambda)$ is the vector representing the measured mixture, and A , B and C are unknown scalars representing the ratio between each component in the mixture, and the relevant known-sample spectrum. By dividing each spectra by its maximum (i.e. maximum optical density value), we would get three normalized spectra (e.g. X_N , Y_N and Z_N) and we can then write:

$$\text{Calculated} = X_N(\lambda) + K_1 Y_N(\lambda) + K_2 Z_N(\lambda) \quad (2)$$

where K_1 and K_2 are the ratios between Y_N and Z_N to X_N respectively. Next step is normalizing the calculated vector (i.e. calculated_N). The program systematically checks all possible pairs of K_1 and K_2 within a range and interval defined by the user. Finally, the program locates the K_1 and K_2 pair which yielded the lowest RMSE between the calculated_N vector and the normalized mixture vector and then calculates the scalars A , B and C . The program returns as output the maximum O.D. and λ_{max} for each of the solutes X , Y and Z , the root mean square error and a graphical representation of the calculated and original vector along with the three base vectors. From the results, the concentration of each of the solutes in the mixture is calculated. The whole process is very fast: in computers with a CPU of 2 GHz a three-solution spectra (each contains 316 absorption values) calculation takes less than a second.

The problem of resolving a vector to a number of its components with a tolerance level for deviations has typically more than one solution. Although, this problem may be addressed by matrix calculation algorithms available in Matlab as well as other linear algebra softwares, these algorithms uses iterative procedures that returns the first solution encountered by the calculation procedure that conforms to the defined tolerance. Thus, using such approach yields a solution that is based on a local RMSE minimum which is not necessarily the best solution. In the code presented here, the program searches all the possible solutions

(within the defined range and using a defined interval) and chooses the best one, thus, the value of RMSE is an absolute minimum and therefore provides a higher reliability of the calculation procedure.

The user interface window

The graphical user interface (Fig. 1) enables the user to redefine the spectra and the calculation parameters while viewing the graphical representation of the calculated and the measured spectra.

Examples:

a. Method verification: In order to verify the efficiency and accuracy of the spectral resolution code, 14 test tubes containing 10 ml mixtures of 2,4,5-trichlorophenol (TCP), 2-chlorophenol (2CP) and 4-chlorophenol (4CP) were prepared. All of the tubes were buffered to pH 10 to ensure that the three solutes existed only as phenolate anions. These solutes were selected due to the high similarity of their UV-vis spectra (Fig. 2), in order to emphasize the abilities of the deconvolution process. All test tubes were measured in an HP 8452A diode-array spectrophotometer in the range of 230–350 nm and the measured spectra were resolved to find TCP, 2CP and 4CP concentrations. In order to make the test as realistic as possible the preparation of the mixtures and the spectral resolution process were done by two different people and the researcher conducting the spectral resolution was unaware of the real composition of the mixture until the end of the process.

b. pK_a measurement: The pK_a values of TCP and pentachlorophenol (PCP) were estimated by the following procedure: 200 ml of analyte solution (0.2 mM and 0.05 mM for TCP and PCP respectively) were placed on a magnetic stirrer and pH was changed by adding concentrated HCl or NaOH. Added volumes of NaOH and HCl were very small (<50 μl) thus had negligible effect on analyte concentrations. Samples in the pH range of 4.45–9.06 and 3.53–4.94 for TCP and PCP respectively, were taken out and measured by an HP 8452A diode-array spectrophotometer. All of the measured spectra were resolved to the spectra representing the completely protonated (represented by samples with pH 9.06 and 8.3 for TCP and PCP respectively) and completely deprotonated

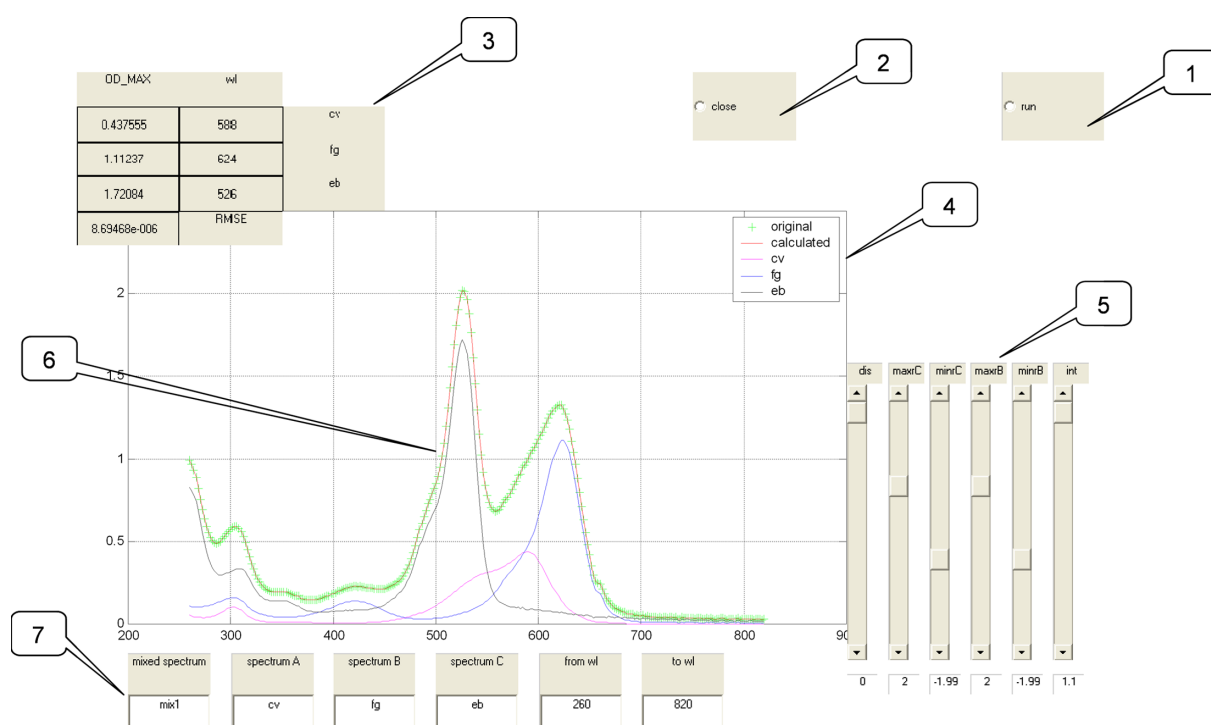


Figure 1. The user interface window.

The user interface window is presented in Figure 1. The user interface includes: 1) Run button—pressing the run button starts the calculation procedure. 2) Close button—Closes the user interface window and saves the results as a matrices named “results”. 3) Results table—displays the results (λ_{\max} and O.D. maximum for each of the components and the RMSE between the calculated and original spectra) of the latest run. 4) Figure legend. 5) Sliders for determination of the following calculation parameters: “int”—defines the primary geometric interval for K and K’. “minrB”/“minrC”—defines the minimal ratio between the maximal value of the normalized vectors of A and B or C respectively. “maxrB”/“maxrC”—defines the maximal ratio between the maximal value of the normalized vectors of A and B or C respectively. 6) Graphical representation of the mixed and calculated vectors with the three base vectors. 7) Text editing windows for the vectors and for the desired wavelength range to be considered.

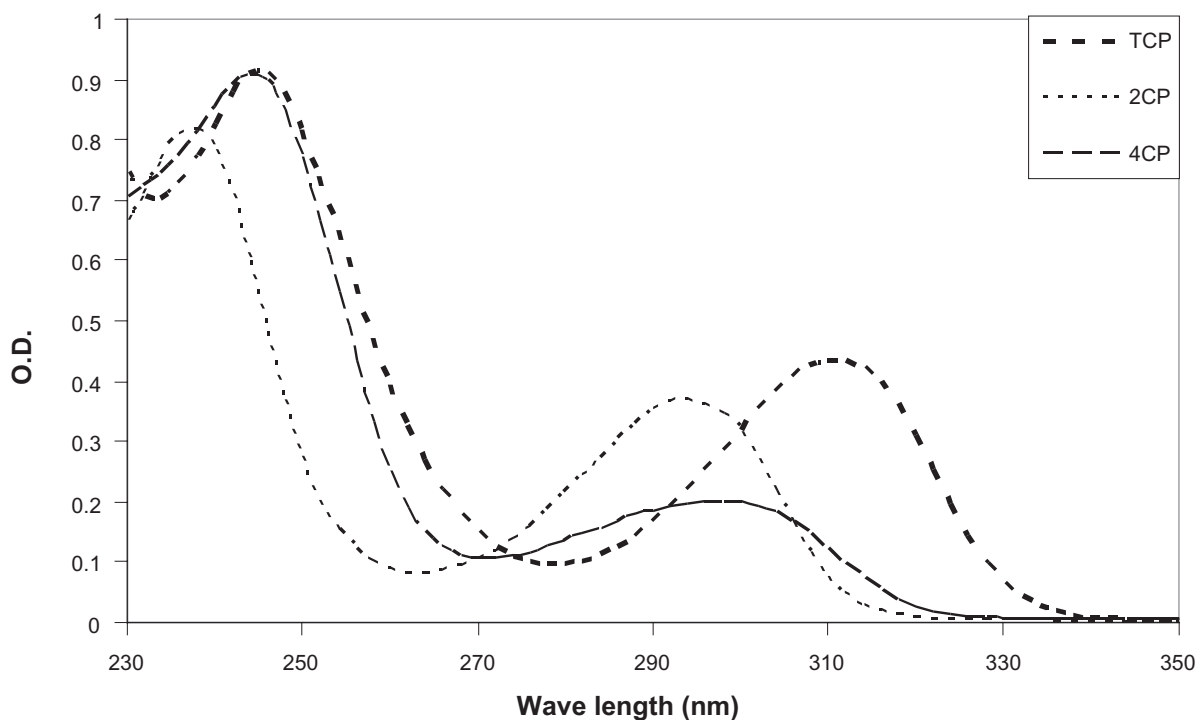


Figure 2. UV-vis spectra of TCP, 2CP and 4CP.

UV-vis spectra of TCP, 2CP and 4CP. All solutes are at a concentration of 0.1 mM and were measured at pH = 10.

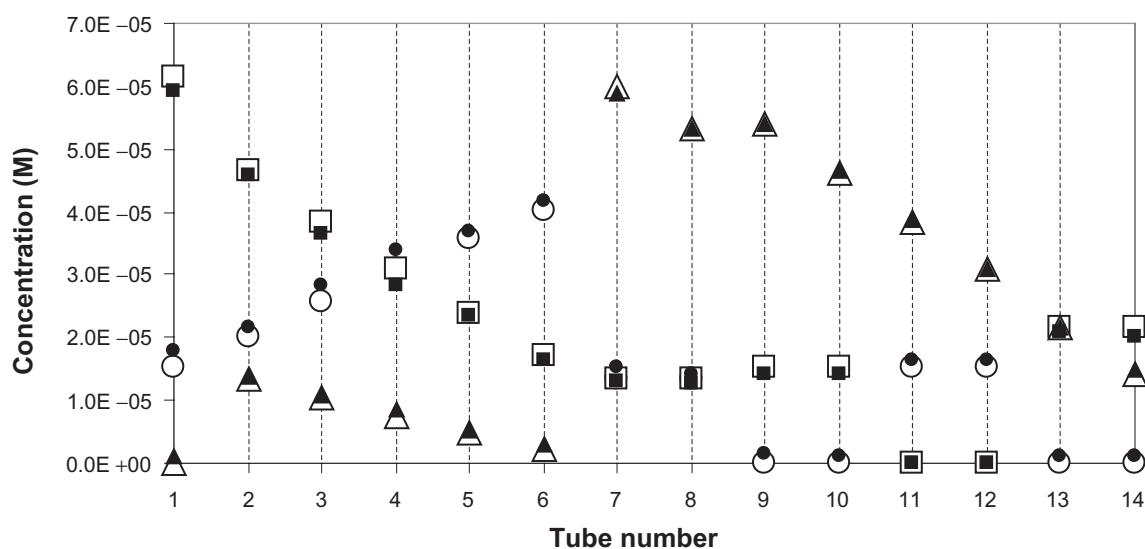


Figure 3. Added and calculated concentrations of TCP, 2CP and 4CP. Added (open symbols) and calculated (black symbols) of: TCP (circles), 2CP (triangles) and 4CP (squares).

(represented by samples with pH 4.45 and 2.74 for TCP and PCP respectively).

Results

a. Method verification: the results of the verification of the spectral separation procedure are presented in Figure 3. The results show a very good fit between the added concentration and the concentration deduced from the spectral resolution process. The root mean square error between the added and calculated concentrations was 1.29 μM which is very reasonable for UV-vis measurements of such solutes.

b. pK_a measurement: The spectra of TCP and PCP in different pH values are presented in Figures 4a and 4b respectively. As may be observed in Figure 4, the spectra in the intermediate pH value

appear to be a superposition of the spectra at a high pH and the spectra at a low pH (representing the deprotonated phenolate specie and the protonated specie respectively).

The protonation reaction of TCP/PCP can be written as

$$\frac{(A^-)(H^+)}{(AH)} = k_a \quad (3)$$

where (A^-) is the activity of the deprotonated species, (AH) is the activity of the protonated species and (H^+) is the proton activity. Thus after resolving each of the measured spectra to the spectra of the protonated and deprotonated species, and assuming activity coefficients are close to 1 (thus activities equals

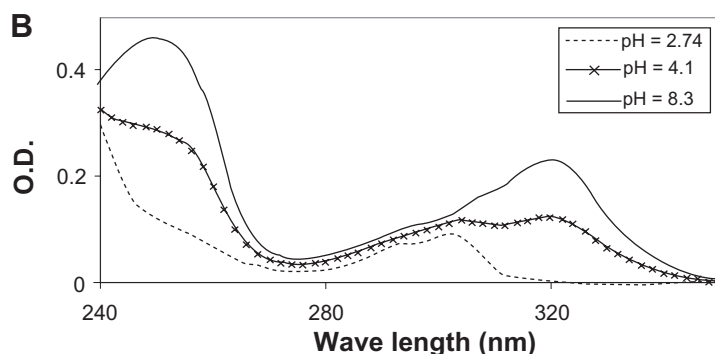
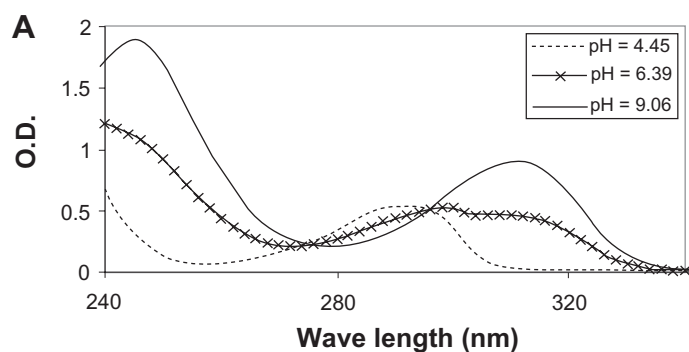


Figure 4. The UV-vis spectra of TCP and PCP at different pH values. UV-vis spectra of: **A)** TCP at pH values of 4.45, 6.39 and 9.06 and **B)** PCP at pH values of 2.74, 4.1 and 8.3.



concentrations) the pK_a value can be estimated by plotting the product of (A^-) and (H^+) against (AH) , the slope of the line would be the deprotonation constant (Figs. 5a and 5b).

Calculating the $-\text{Log}_{10}$ of the slope of the TCP line in Figure 5a yields $pK_a = 6.41$. This value is lower than reported in the literature 6.7–6.94.^{8,9} In order to estimate the reliability of the resolution process, mass conservation of TCP was calculated by summing concentrations of protonated and deprotonated species and dividing the result by the total TCP concentration (i.e. 0.1 mM). The average mass conservation was $100.4\% \pm 0.458\%$. Such accuracy in the mass balance may indicate the reliability of the process even though the slight discrepancies with literature values.

Calculating the $-\text{Log}_{10}$ of the slope of the PCP line in Figure 5b yields $pK_a = 4.14$. This value is much lower than reported in the literature for experimental values 4.4–5.26.^{10,11} The mass conservation of PCP was calculated the same way mentioned for TCP and was $99.3\% \pm 6.9\%$.

The very wide range of literature values for PCP may indicate a problem of pK_a determination using the conventional titration method (probably because of its limited water solubility). On the other hand, the method of deducing pK_a from the slope of such plot appears very sensitive: For example, a change of 0.1 pH units in the pK_a value calculated for PCP, would result in a 20.6% decrease or 26% increase in the slope for downward and upward changes respectively.

Even though example [b] focuses on mixtures of two components, example [a] demonstrate that the program can accurately deal with three different components in a mixture. Additional possible mixtures in environmental conditions can be three different dyes

dissolved in a textile factory effluents, or two species of the same chemical (like in example [b]), in which the solution includes also unknown concentration of a known impurity, with a well defined UV-visible spectrum, as humic acid- for example.

Discussion

The software presented here, may be useful whenever two or three known solutes with overlapping spectra need to be quantified. In addition, whereas most diode-array operating softwares are able to subtract linear baselines from the spectra, the software presented here may be used for subtraction of any background spectra needed.

The software enables the user to adjust the number of wavelengths (vector length), and the wavelength interval, thus it can be used for different types of diode array spectrophotometers.

Although, the method presented here works even when overlapping between the component spectra are very high (see example a), the main drawback of the method is that it relies on perfect superposition of the component spectra. Thus, whenever there is a deviation from Beer-Lamberts law (for examples the formation of dimers and aggregates), a mutual influence of solutes (for example—complexation) or even influence of pH or ionic strength on the spectra, this method should be avoided.

An important advantage of the proposed software arises from its being an “open code” computing solution, this enables any user with elementary knowledge in programming to understand the calculation procedure and add or change features to suite any specific requirements. As long as the assumptions mentioned earlier are met, the code may be adjusted

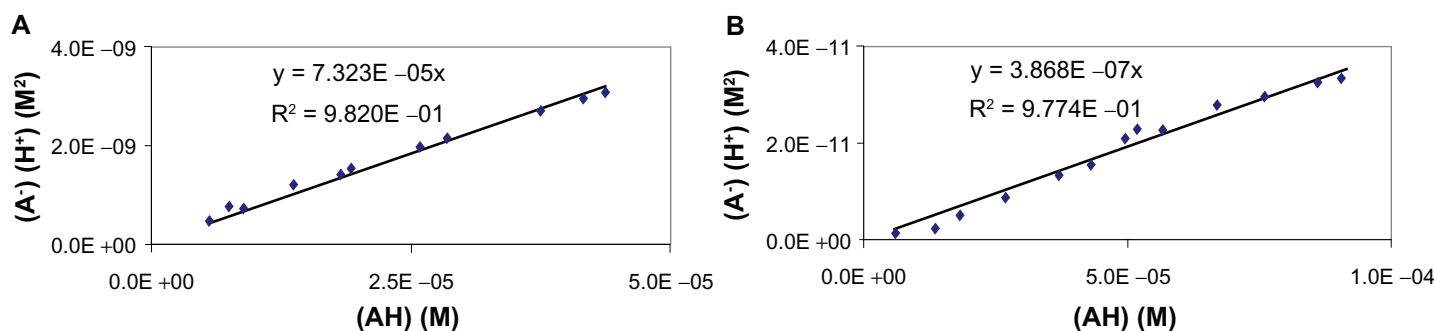


Figure 5. $(A^-)(H^+)$ versus (AH) plots for TCP and PCP. A plot of $(A^-)(H^+)$ versus (AH) ; for **A**) TCP samples in the pH range of 5.46–7.89 and **B**) PCP samples in the pH range of 3.43–4.94.



for use in different analytical methods as fluorescence spectroscopy, chromatography (while time is used instead of wavelength), and with some adjustment possibly even ATR infra red spectrophotometry. The full code was not presented here for brevity; and it may be received by E-mail (free of charge) along with a short user manual following a request to the authors.

Acknowledgment

This study was partly supported by grant 849/05 by the Israeli Science Foundation, and by an Infrastructure grant of the Israeli Ministry of Science and Technology, on Novel Tailored Water Treatment Solutions Based on Clay Nano-Technologies.

References

1. Skoog DA, Leary JJ. *Principles of instrumental analysis* (Saunders College Publishing, USA, 1992) fourth ed., p. 161–4.
2. Antonov L, Nedeltcheva D. *Chem Soc Rev Environ Sci Technol*. 2003; 37:5254.
3. Callahan MR, Rose JB, Garcia-Rubio L. *Environ Sci Technol*. 2003; 37:5254.
4. Mattley Y, Leparé G, Potter R, Garcia-Rubio L. *Photochemistry and photobiology*. 2000;71:610.
5. Palter KP, Molloy JL, Niemeyer ED. *Journal of Chemical Education*. 2005;82:1682.
6. Brown DW, Floyd AJ, Sainsbury M. *Ultraviolet and Visible Spectroscopy, in "Organic Spectroscopy"* (Wiley J & Sons Ltd., New York. 1988).
7. Rytwo G, Nir S, Margulies L. *Soil Sci Soc Am J*. 1995;59:554.
8. Dentel K, Bottero JY, Khatib K, Demougeot H, Duguet JP, Anselme C. *Wat Res*. 1995;29:1273.
9. Schellenberg K, Leuenberger C, Schwarzenbach RP. *Environ Sci Technol*. 1984;18:652.
10. Myers AG, Goldberg SD. *Angew Chem Int Ed*. 2000;39:2732.
11. Ez'erskis Z, Jusys Z. *Pure Appl Chem*. 2001;73:1929.

Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>