

MULTIPARAMETRIC CURVE FITTING—IV

COMPUTER-ASSISTED ESTIMATION OF SUCCESSIVE DISSOCIATION CONSTANTS AND OF MOLAR ABSORPTIVITIES FROM ABSORBANCE-pH CURVES BY THE DCLET PROGRAM

MILAN MELOUN

Department of Analytical Chemistry, College of Chemical Technology,
53210 Pardubice, Czechoslovakia

and

JOSEF ČERMÁK

Department of Physics, College of Chemical Technology, 53210 Pardubice, Czechoslovakia

(Received 12 July 1978. Accepted 28 November 1978)

Summary—The program DCLET evaluates the dissociation (or protonation) constants and molar absorptivities ϵ_{H_jL} of all light-absorbing species H_jL ($j = 0, \dots, J$) of a polyprotic acid H_JL by a non linear regression of the function $A = f(\text{pH}; \text{p}K_a, \epsilon_{H_jL})$, i.e., by multiparametric curve fitting. The three specific subroutines DATADC, UBBEDC, SKRDC described form part of the general program ABLET. The goodness of fit is tested by statistical examination of the residuals. Heuristic or algorithmic strategies of minimization may be used, and experimental or synthetic data may be processed. By examining the effects on the results of varying the density and distribution of points of a synthetic data set, it is possible to improve the planning of experiments, and by introducing parametric weight, to improve the sensitivity of a particular parameter in a model.

Existing methods for determination of dissociation constants by various approaches have already been discussed comprehensively.¹⁻⁵ For a protolytic system in which some species have characteristic absorption spectra in the ultraviolet and/or the visible region, a spectrophotometric method can be used with advantage. A sufficiently large set of experimental data can readily be obtained by microtitration combined with simultaneous monitoring of pH.⁶⁻¹⁰ For monoprotic systems, the normal graphical procedure involves transformation of the experimental absorbance-pH function.² A straightforward numerical version based on this approach makes use of simple linear regression.^{11,12}

The methods proposed for determining values of two successive dissociation constants usually require knowledge of the absorptivities of the two boundary species of the protonation system. The best values of the other parameters are then calculated by successive approximation. Irving, Rossotti and Harris¹³ have presented a method which may be used even if the molar absorptivities of individual species cannot be determined directly. Thamer and Voigt¹⁴ have developed a method which is applicable where the ratio of the dissociation constants is less than 10^3 . In a subsequent paper, Thamer¹⁵ outlined two general methods of calculation for two close successive dis-

sociation constants, which do not require the determination of absorptivities. Both the direct approach and the calculation by successive approximation are designed for a desk calculator; however, they are rather tedious and susceptible to arithmetical errors. Roth and Bunnett⁶ have expressed the absorbance-pH function as a linear equation in five unknowns. A suitable selection of five points is made from the experimental A -pH graph to obtain a set of five simultaneous equations which is solved by a computer. This method, like that of Thamer,¹⁵ has the disadvantage that all the other experimental points are ignored.

The Taylor's-series approximation in non-linear least-squares analysis has been adapted by Auld and French,⁸ with the use of an electronic desk calculator, to allow the dissociation constants for a monoprotic and a diprotic acid to be calculated even when the constants are similar. Heys, Kinns and Perrin¹⁷ have described an approach in which all the experimental data are used to find the best fit for the molar absorptivity of the intermediate species and for K_{a1} and K_{a2} by a least-square minimization method. The A -pH function is rewritten to obtain an expression in three unknowns. For each of n experimental points an equation is obtained and the set of n simultaneous equations is solved by a matrix transposition technique using the IBM library subroutine¹⁸ based on the method of Golub.¹⁹ Multiparametric curve-fitting can also be used in combination with deviation-pat-

tern recognition to distinguish the titration curve of a diprotic species with close dissociation constants from that of a monoprotic system.²⁰

During the sixties various non-linear estimation methods were developed and applied to the calculation of equilibrium constants.^{2-5,21} Multiparametric least-squares curve-fitting programs for laboratory minicomputers,^{22,23} recently described, have been used for various analytical purposes, including the evaluation of rate and equilibrium constants.²⁴ The pit-mapping program LETAGROP VRID by Sillén *et al.*^{25,26} offers perhaps the most comprehensive approach to the computation of the stability constants. On the other hand, the transfer of a large program of this type, such as SPEFO,²⁶ to a large computer requires much effort. Two programs, FA608 and EY608, designed by Kankare²⁷ for small computers (16 kbyte) can evaluate equilibrium constants and their errors for multicomponent systems, from spectrophotometric data. Factor analysis is used to estimate the number of independently varying components, and for data reduction. Molar absorptivities of individual components are also calculated and spectra interpolated from these values can be plotted. Kaden and Zuberbühler²⁸ have published a general program capable of processing either potentiometric or absorbance data by changing one or two subroutines. By systematic computer analysis of spectral data, the number of absorbing species, the stoichiometry, and the equilibrium constants or acid dissociation constants of Pyrocatechol Violet, and of the system tin(IV)-Pyrocatechol Violet were determined by means of a program written by Wakley and Varga.²⁹ The program SQUAD, written by Leggett and McBryde,³⁰ is capable of calculating, simultaneously or individually, acid association constants (and hence pK_a values) and overall stability constants (of the concentration type) for any species formed in systems containing up to two metals and two ligands.

This paper deals with computation of both distinct and similar dissociation constants from absorbance-pH data. The program DCLET is based on the subroutine LETAG,³¹⁻³³ which is a version of Sillén's LETAGROP VRID, adapted to be used with the subroutine LETAG.

THEORY

Equation of an absorbance-pH curve

The spectrophotometric determination of dissociation (or protonation) constants of a polyprotic acid H_jL is based on the interpretation of the absorbance-pH curve measured over the pH range covering the main regions of existence of all the species of interest. The equation for the A -pH curve is written in the following manner

$$A = d(\epsilon_L[L] + \epsilon_{HL}[HL] + \epsilon_{H_2L}[H_2L] + \dots) \quad (1)$$

where d is the path-length and ϵ_L , ϵ_{HL} , ϵ_{H_2L} , ... etc. are the absorptivities of those species of the protona-

tion system which absorb light at the chosen wavelength. Each equilibrium concentration of a species which occurs to a measurable extent within the pH range of interest is a function of the hydrogen-ion activity.

Since

$$[H_jL] = c_{H_jL} \cdot \beta_{H_jL} \cdot a_H^j / (1 + \beta_{HL} \cdot a_H + \dots + \beta_{H_jL} \cdot a_H^j)$$

and

$$\beta_{H_jL} = [H_jL]/[L] \cdot a_H^j,$$

where c_{H_jL} is the analytical concentration of acid H_jL , equation (1) may be written as

$$A = dc_{H_jL} \frac{\epsilon_L + \epsilon_{HL}\beta_{HL}a_H + \epsilon_{H_2L}\beta_{H_2L}a_H^2 + \dots}{1 + \beta_{HL}a_H + \beta_{H_2L}a_H^2 + \dots}$$

$$= dc_{H_jL} \frac{\epsilon_L + \sum_{j=1}^J \epsilon_{H_jL} 10^{(j \cdot \log a_H + \log \beta_{H_jL})}}{1 + \sum_{j=1}^J 10^{(j \cdot \log a_H + \log \beta_{H_jL})}} \quad (2)$$

In the expression $j \cdot \log a_H + \log \beta_{H_jL}$, the conventional activity pH scale ($pH = -\log a_H$) may be used and the protonation constant β_{H_jL} may be expressed as a function of the mixed stepwise dissociation constant $K_{ai} = a_H \cdot [H_{i-1}L]/[H_iL]$:

$$j \cdot \log a_H + \log \beta_{H_jL} = \sum_{i=1}^j pK_{ai} - j \cdot pH \quad (2a)$$

In many practical cases the protonation equilibria are sufficiently separated, ($pK_{a(j-1)} - 3 \leq pK_{aj} \leq pK_{a(j+1)} + 3$) for each of them to be treated as a monoprotic system, in which case equation (2) can be written as

$$A = dc_{HL} \frac{\epsilon_L + \epsilon_{HL} 10^{(pK_{a1} - pH)}}{1 + 10^{(pK_{a1} - pH)}} \quad (3)$$

If a graphical approach is used,¹ equation (3) is usually put in the linearized logarithmic form

$$\log \frac{A - dc_{HL}\epsilon_L}{dc_{HL}\epsilon_{HL} - A} = pK_{a1} - pH \quad (4)$$

and the value of pK_{a1} is found as the point where the straight line intersects the pH-axis.

Regression analysis

In the analysis of an absorbance-pH curve in terms of equation (2), it must be remembered that the values of the parameters ϵ_{H_jL} and pK_a are known only approximately from the graph, but by means of the following refinement process their exact values can be determined.

For fitting the non-linear function [equation (2)], a least-squares curve-fitting method seems to be the

Table 1. Example of the use of program DCLET. Experimental data set for Pyrocatechol Violet; absorbances measured at 590 nm, $c_1 = 3.0 \times 10^{-5} M$ (output shortened and simplified)

<i>Data</i>			
Number of points	23		
Total concentration of dye (<i>M</i>)	1.00000		
Coefficients Q_1, Q_2	1.0 1.0		
WT (Temperature, <i>K</i>)	298.16		
WW [Liquid-junction potential correction (mV)]	0.0		
WZ (pa_H of standard buffer solution)	7.413		
WK [Nernstian slope of glass electrode (mV)]	57.59		
	<i>i</i>	pH_{read}	pa_{H+i}
	1	4.220	4.133
	2	5.300	5.242
	3	6.490	6.465
	4	6.800	6.783
	5	7.020	7.009
	6	7.200	7.194
	7	7.480	7.482
	8	7.600	7.605
	9	7.800	7.811
	10	8.050	8.067
	11	8.200	8.221
	12	8.350	8.376
	13	8.740	8.776
	14	8.960	9.002
	15	9.160	9.208
	16	9.370	9.423
	17	9.600	9.660
	18	9.790	9.855
	19	10.040	10.112
	20	10.250	10.327
	21	10.500	10.584
	22	10.790	10.882
	23	11.220	11.324
			$A_{exp,i}$
			0.008
			0.019
			0.101
			0.169
			0.246
			0.326
			0.479
			0.546
			0.665
			0.793
			0.862
			0.907
			0.957
			0.952
			0.916
			0.858
			0.783
			0.710
			0.606
			0.535
			0.472
			0.425
			0.389

Leta

Values of parameters during minimization process

Initial guess	ϵ_L	ϵ_{HL}	pK_{11}	ϵ_{H2L}	pK_{21}	<i>U</i>
After 1st iteration	0.389	1.000	9.800	0.038	7.600	3.7954E-02
After 2nd iteration	0.363	1.064	9.854	0.018	7.581	7.2909E-04
After 3rd iteration	0.366	1.067	9.835	0.020	7.589	6.3589E-04
After 4th iteration	0.366	1.067	9.835	0.020	7.589	6.3588E-04
After termination	0.36633	1.06713	9.83496	0.01969	7.58880	6.3587E-04
	+-	+-	+-	+-	+-	
	0.00461	0.00519	0.01480	0.00343	0.00871	

SIGY (standard deviation of absorbance) = 0.0059

Skrik

Curve fitting analysis by statistical test of residuals, $\delta_i = (A_{exp,i} - A_{calc,i}) \times 10^4$

<i>i</i>	δ_i	<i>i</i>	δ_i
1	-121	21	-2
2	-54	22	10
3	81	23	6
4	75		
5	79	<i>Statistical parameter</i>	
6	54	Arithmetic mean 1.7E-07	
7	0	Mean deviation 4.1E-03	
8	-67	Standard deviation σ 0.0053	
9	-75	Variance σ^2 2.8E-05	
10	-74	Moment coefficient of skewness -0.369	
11	25	Moment coefficient of kurtosis 2.51	
12	30	Observed χ^2 2.39	
13	20	[χ^2 (6; 0.95) should be less than 12.6]	
14	59		
15	1		
16	-45		
17	4		
18	31		
19	-21		
20	-17		

most convenient; thus the sum of the squares of the residuals

$$U = \sum_{i=1}^N w_i (A_{\text{exp},i} - A_{\text{calc},i})^2 \quad (5)$$

for N experimental points $\{A_{\text{exp}}, \text{pH}\}$ is minimized. The statistical weight, w_i , is usually set at unity for experimental data sets, but for synthetic data sets the weight is calculated, and is different from unity. Non-linear estimation is a problem of optimization in multiparametric space, in which the values of A and pH are the given numbers and the parameters $\epsilon_{\text{H},i}$, $\text{p}K_{a,i}$ are the variables. Starting from an initial guess of the values of the parameters, the "best" values are reached by a few successive approximations. If it is assumed that the error-square sum U is a second-degree function of m parameters, then $0.5(m+1)(m+2)$ points in parametric space suffice for calculation of the position of the minimum.

Successful application of this computational method depends on the use of the heuristic (trial-and-error) minimization process and also on the number of data available. An accurate, well-planned A - pH data set should be obtained, and the measurement should be properly conducted (e.g., a combined pH -photometric microtitration technique seems to be best¹⁰). The number of protons taking part in the protonation equilibria within the relevant pH range should be known so that a suitable mathematical model may be chosen. The data should first be treated graphically to detect and eliminate possible systematic errors and to obtain initial guess for the parameters to be determined.

COMPUTER PROGRAM

The program DCLET is written in Fortran and has been run on an EC 1040 (Robotron, GDR) which is compatible with the IBM 360 computer. Some of the basic subroutines of the minimization process will be published in another paper from this series.³³ An older version of DCLET (1972) for a smaller store-capacity computer (Hewlett-Packard 2116, Tesla 200), called JDC-LETAG, is available on request.

The program will accept up to 50 points from the A - pH curve, and 7 unknown parameters (i.e. $J \leq 3$ in H_2L). The program has a flexible structure allowing a heuristic (trial-and-error) strategy to be used in the minimization process.

Program DCLET starts with initial guesses of the parameters from the data, i.e., experimental values of the independent variable pH , and of the dependent variable A . A "best" approximation to the unknown parameters ($\text{p}K_a$, $\epsilon_{\text{H},i}$), and their standard deviations ($\sigma_{\text{p}K_a}$, $\sigma_{\epsilon_{\text{H},i}}$) are calculated. The agreement between the calculated A - pH curve and the experimental points can be seen in a table. The residuals (i.e., the differences between the experimental and calculated absorbances) are analysed statistically.

If the model is satisfactory, and there are no systematic errors in the data, the residuals should have a Gaussian distribution with the following properties: arithmetic mean zero, standard deviation σ , mean deviation $(\sqrt{2}\pi)\sigma \approx 0.8\sigma$, variance σ^2 , moment coefficient of skewness zero, and moment coefficient of kurtosis 3.³⁴ A value for χ^2 is then derived from the difference between the observed and calculated probability. A fit can be accepted at the appropriate confidence level if the observed value of χ^2 is less than the theoretical value.

The program may be used to study minimization strategy, and to plan experimental work. For this, a synthetic data set simulating an A - pH curve in which the points are loaded by calculated errors, is generated for each point of an A - pH curve. Thus, it is possible to increase the influence of each parameter values of the independent variable, pH , it is possible to find an optimum plan for an experiment. The structure of the program allows a weight to be calculated for each point of an A - pH curve. Thus, it is possible to increase the influence of each parameter on the value of the error-square sum.

The program consists of a main program and several subroutines. The subroutine DCLET allows the user to read the data, to call another subroutine for a simulation of synthetic data, if desired, for the minimization process, and for output. The subroutine DATADC does all the calculations necessary to transform the primary experimental data. The measured pH values are corrected for any deviation from Nernstian slope of the glass electrode, for any change in the temperature from 298.16 K, and for the liquid-junction potential (for details see previous paper³³ in this series^{35,36}).

The subroutine UBEDC calculates the absorbance for each pH value [equation (2)] and a set of values of the parameters resulting at the particular stage of the minimization process. Subsequently, the sum U of the squares of the residuals is calculated [equation (5)]. The subroutine SKRDC gives as output the values of the parameters and their standard deviations, the statistical analysis of residuals and a graph of experimental and calculated points.

The other subroutines are basic in character and have been described in a previous contribution.³³ They are READI, READR, STATS, LETAG with internal subroutines MULLE, INVER, PINUS, and WEIGHT, PLOTT, SIMUL, NORAND and RANDOM.

A listing of the program DCLET, and instructions for preparation of input data and interpretation of output are available on request.

REFERENCES

1. A. Albert and E. P. Serjeant, *Ionization Constants of Acids and Bases*, Methuen, London, 1962.
2. I. Ya. Bershtein and Yu. L. Kaminskii, *Spektrofotometricheskii analiz v organicheskoi khimii*, Izdatelstvo Khimiya, Leningrad, 1975.

3. W. A. E. McBryde, *Talanta*, 1974, **21**, 979.
4. C. W. Childs, P. S. Hallman and D. D. Perrin, *ibid.*, 1969, **16**, 1119.
5. F. J. C. Rossotti, H. S. Rossotti and R. J. Whewell, *J. Inorg. Nucl. Chem.*, 1971, **33**, 2051.
6. H. E. Boaz and J. W. Forbes, *Anal. Chem.*, 1958, **30**, 456.
7. R. W. Rinehart and J. E. Stafford, *Microchem. J.*, 1962, **6**, 567.
8. D. S. Auld and T. C. French, *Anal. Biochem.*, 1970, **34**, 262.
9. T. C. Bruice and J. R. Maley, *ibid.*, 1970, **34**, 275.
10. M. Meloun, V. Riha, S. Kotrlý and J. Pánel, *Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Pardubice* 1975, **33**, 161.
11. J. Havel and V. Kubáň, *Scripta Fac. Sci. Nat. Univ. Purkynianae Brun., Chemia*, 1971, **2**, 87.
12. V. Kubáň, *ibid.*, 1972, **2**, 81.
13. H. Irving, H. S. Rossotti and G. Harris, *Analyst*, 1955, **80**, 83.
14. B. J. Thamer and A. F. Voigt, *J. Phys. Chem.*, 1952, **56**, 225.
15. B. J. Thamer, *ibid.*, 1955, **59**, 450.
16. B. Roth and J. F. Bunnett, *J. Am. Chem. Soc.*, 1965, **87**, 334.
17. G. Heys, H. Kinns and D. D. Perrin, *Analyst*, 1972, **97**, 52.
18. *IBM Scientific Subroutine Package*, Subroutine LLSQ, IBM Corporation, New York, 1968.
19. G. Golub, *Num. Math.*, 1965, **7**, 206.
20. L. Meites and D. M. Barry, *Talanta*, 1973, **20**, 1173.
21. K. Nagano and D. E. Metzler, *J. Am. Chem. Soc.* 1967, **89**, 2891.
22. L. Meites and L. Meites, *Talanta*, 1972, **19**, 1131.
23. D. M. Barry and L. Meites, *Anal. Chim. Acta*, 1974, **68**, 435; D. M. Barry, L. Meites and B. H. Campbell, *ibid.*, 1974, **69**, 143.
24. W. A. de Olivera, *M. S. Thesis*, Clarkson College of Technology, 1972.
25. N. Ingri and L. G. Sillén, *Ark. Kemi*, 1964, **23**, 97.
26. L. G. Sillén and B. Warnqvist, *ibid.*, 1969, **31**, 377.
27. J. J. Kankare, *Anal. Chem.*, 1970, **42**, 1322.
28. T. Kaden and A. Zuberbühler, *Talanta*, 1971, **18**, 61.
29. W. D. Wakley and L. P. Varga, *Anal. Chem.*, 1972, **44**, 169.
30. D. J. Leggett and W. A. E. McBryde, *Anal. Chem.* 1975, **47**, 1065.
31. Version of LETAG for ODRA 1013 in auto-code MOST F13: J. Čermák and M. Meloun, *Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Pardubice*, 1974, **32**, 37.
32. Version of LETAG for Hewlett-Packard 2116 in Fortran: J. Čermák and M. Meloun, *2nd Symposium on Algorithms in Computation Technique, High Tatras 1974, Knížnica algoritmov, Vol. 2*, SVTS, Bratislava, 1974.
33. Version of LETAG for IBM 360 in Fortran: J. Čermák and M. Meloun, *Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Pardubice*, 1978, **39**, 41.
34. M. R. Spiegel, *Theory and Problems of Statistics, SI (metric) Ed.*, Chap. 7, McGraw-Hill, New-York, 1972.
35. M. Meloun and J. Čermák, *Talanta*, 1976, **23**, 15.
36. M. Meloun and S. Kotrlý, *Collection Czech. Chem. Commun.*, 1977, **42**, 2115.