DETERMINATION OF EQUILIBRIUM CONSTANTS FROM SPECTROPHOMETRIC DATA OBTAINED FROM SOLUTIONS OF KNOWN pH: THE PROGRAM pHab

Peter GANS[a](°), Antonio SABATINI[b] and Alberto VACCA[b]

a, School of Chemistry, The University of Leeds, Leeds LS2 9JT, England

b, Dipartimento di Chimica, Università degli Studi di Firenze, Via Maragliano 77, 50144 Firenze, and I.S.S.E.C.C., C.N.R., Via Nardi 39, 50132 Firenze, Italy.

Summary - A new program suite, pHab, has been written for the determination of equilibrium constants from spectrophotometric data obtained from solutions at known pH. The suite consists of a custom data-editor, a refinement program and an output viewer.

Riassunto - Un nuovo insieme di programmi, pHab, è stato scritto per la determinazione di costanti di equilibrio da dati spettrofotometrici ottenuti da soluzioni a pH noto. Il pacchetto consiste di un editore di dati dell'utente, del programma di affinamento delle costanti di equilibrio e di un visualizzatore dei risultati.

INTRODUCTION

We have recently given a summary of the most important programs used for the determination of equilibrium constants.[1] Most of these program handle data from one type of experiment, the majority dealing with potentiometric data. HYPERQUAD,[1] DALSFEC,[2] NONLIN15,[3] and PSEQUAD[4] are general programs designed to handle both potentiometric and spectrophotometric data simultaneously. Of general programs designed to handle solely spectrophotometric data, LETAGROP-SPEFO[5] was the first and most subsequent programs share a similar approach to the mathematical problems. SQUAD[6] seems to be the most widely used program, but it uses numerical approximations to the elements of the Jacobian (as do all the programs in the LETAGROP family) which mitigates against robustness. Other notable programs include SPECA[7] and STAR.[8]

(°), corresponding author; fax: +44 113 233 6565; e-mail: peterg@chem.leeds.ac.uk

SPECFIT,[9] has some mathematical features which distinguish it from all other programs. The most important of these are the use of principal component analysis to compress the experimental data and a refinement algorithm which "eliminates" the so-called linear parameters from the equilibrium constant refinement. Both of the features depend on the fact that calculated absorbance is a linear function of the concentrations of the species present, as exemplified by the Beer-Lambert law.

$$A_{s\lambda}^{calc} = l_s \sum_{i=1-na} c_{si}\varepsilon_{i\lambda} \qquad [1]$$

Eqn. 1 is to be read as follows. The calculated absorbance of solution *s* at wavelength λ is the sum of *na* terms, each term being the product of c_{si}, the concentration of the *i*th species in the solution *s* and $\varepsilon_{i\lambda}$, the molar absorbance of the *ith* species at the wavelength λ. The sum is scaled by l_s, the optical path-length used for the solution *s*. In the absence of noise in the experimental data the matrix $\boldsymbol{A}^{obs}$ of measured absorbances should have rank *na*, the number of species which absorb light. By performing a singular value decomposition **A=USV** a compressed data matrix can be constructed by using the first *na+2* rows of **V** as a projection operator: $\mathbf{A \leftarrow AV^T}$. This compression has the merit of reducing the number of columns in the absorbance matrix from the number of wavelengths to a few more than the number of light-absorbing species and allows data from all wavelengths to be included in the calculation.

The second feature is that the Jacobian, $\frac{\partial \mathbf{A}}{\partial \beta}$, is calculated by a procedure in which the molar absorbances, **ε**, become implicit quantities in the analytical expressions.[9] The method of calculation of the Jacobian has recently been revised for speed of execution.[10]

By contrast, the traditional approach[5] has been to alternate explicit calculation of the molar absorbances by a linear least-squares process with a cycle of non-linear refinement of the equilibrium constants. This was the procedure adopted for Hyperquad.[1] However Hyperquad requires that the analytical concentrations of all reagents be known. This is because the free concentrations are found by solving the complete set of mass-balance equations, eqn. 2:

$$T_k = [A_k] + \sum_{i=1,nk} p_{ij}\beta_i \prod [A_j]^{p_{ij}} \qquad [2]$$

The index *k* runs over all the reagents, *[A]* represents a free concentration, p_{ij} is the stoichiometric coefficient of the *j*th reagent in the *i*th complex species and β_i is the equilibrium constant for the *i*th species. The total concentration of the *k*th reagent, T_k, is given by eqn. 3,

$$T_k = \frac{n_k + va_k}{v_0 + v} \qquad [3]$$

which shows the total concentration of a reagent during a titration in which n_k moles of reagent *k* are present initially in the titration vessel in a volume v_0, a_k is the concentration of the reagent in the burette, and *v* is the volume added at a titration point.

If pH is known rather than the analytical concentration of the hydrogen ion then $[A_H]$ rather than T_H is known. This requires that the method of calculation of the free concentrations be modified. pHab was therefore derived from Hyperquad by adapting for this eventuality, by removing those parts

relevant to potentiometric data only and by creating a new type of data file for the spectrophotometric data, whilst retaining the standard form for the model file.

METHOD AND IMPLEMENTATION

The free concentrations $[A_k]$ are obtained by solving the reduced set of equations of mass balance (no equation for the proton) and using $[H^+]=10^{-pH}$ for the free hydrogen ion concentration. Inside the program the proton is placed last in the list of reagents, but users may place it anywhere in the list. The Jacobian for the system is obtained by partial differentiation of eqn. 1 and consists of blocks with the following structure:

$$\begin{array}{cccccccccccccc}
\frac{\partial A_1}{\partial \varepsilon_{1,1}} & \cdots & \frac{\partial A_1}{\partial \varepsilon_{na,1}} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \frac{\partial A_1}{\partial \beta_1} & \cdots & \frac{\partial A_1}{\partial \beta_{nk}} & \frac{\partial A_1}{\partial \alpha_1} & \cdots & \frac{\partial A_1}{\partial \alpha_{nd}} \\
0 & \cdots & 0 & \frac{\partial A_2}{\partial \varepsilon_{1,2}} & \cdots & \frac{\partial A_2}{\partial \varepsilon_{na,2}} & \cdots & 0 & \cdots & 0 & \frac{\partial A_2}{\partial \beta_1} & \cdots & \frac{\partial A_2}{\partial \beta_{nk}} & \frac{\partial A_2}{\partial \alpha_1} & \cdots & \frac{\partial A_2}{\partial \alpha_{nd}} \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & 0 & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \frac{\partial A_{nl}}{\partial \varepsilon_{nl,1}} & \cdots & \frac{\partial A_{nl}}{\partial \varepsilon_{na,nl}} & \frac{\partial A_{nl}}{\partial \beta_1} & \cdots & \frac{\partial A_{nl}}{\partial \beta_{nk}} & \frac{\partial A_{nl}}{\partial \alpha_1} & \cdots & \frac{\partial A_{nl}}{\partial \alpha_{nd}}
\end{array}$$

Here, the first *nl*×*na* columns constitute a partition, $\mathbf{J}_\varepsilon$, which relates to the unknown molar absorbances, of which there is one at each wavelength for each light-absorbing species, the next *nk* columns constitute the partition $\mathbf{J}_\beta$ and the last *nd* columns constitute the partition, $\mathbf{J}_\alpha$, which relates to the so-called dangerous parameters (total amount, burette concentration). There is a block such as this for each solution used. Expressions for the partial derivatives have been given previously.[1]
The full Jacobian, $\mathbf{J}=(\mathbf{J}_\varepsilon|\mathbf{J}_\beta|\mathbf{J}_\alpha)$ is used to calculate the adjustments to α and β, by partially solving eqn. 4:

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\begin{pmatrix}\Delta\varepsilon\\ \Delta\beta\\ \Delta\alpha\end{pmatrix}=\mathbf{J}^{\mathrm{T}}\left(\mathbf{A}^{\mathbf{obs}}-\mathbf{A}^{\mathbf{calc}}\right) \qquad [4]$$

After these adjustments have been applied the concentrations are recalculated and the molar absorbances are obtained by fitting eqn. 1 by the method of linear least-squares, i.e. by solving eqn. 5:

$$\left(\mathbf{J}_\varepsilon{}^{\mathrm{T}}\mathbf{J}_\varepsilon\right)\varepsilon=\mathbf{J}_\varepsilon{}^{\mathrm{T}}\left(\mathbf{A}^{\mathbf{obs}}-\mathbf{A}^{\mathbf{calc}}\right) \qquad [5]$$

In order to make the program pHab easier to use we have written a custom data editor, pHabEDIT, which is similar in operation to HEDIT[1] and an output interpreter, pHabOUT which is similar to HANDOUT.[1] These programs were written with the aid of the program library that was developed for Hyperquad and offer many of the same facilities. The spectrophotometric data for pHab is held in a file with the extension PAV (Program Absorbance Values), which differs from the .PAD files required by Hyperquad principally in that it holds the observed pH values and does not hold any information regarding total hydrogen ion concentration. The equilibrium model is held in a standard Hyperquad.PAR file. Ancillary files that are included in the pHab package include a set of sample data files, template files for new data and a manual written in WinWord.[11]

DISCUSSION

It appears that pHab and SPECFIT give identical results when applied to the same data (protonation of chromate).[12] This is reassuring for the two programs differ mainly in their methods of successive approximation. An important difference is that SPECFIT uses factor analysis to enable data compression. The advantage of this procedure is that data from all wavelengths is utilised. The disadvantage is that the compression process must introduce some distortion into the data. This arises because singular value decomposition is itself a least-squares process,[13] in which noise cannot be removed from the data without some concomitant removal of signal.[14] This result therefore confirms the fact that the compression process used in SPECFIT introduces negligible signal distortion.

When using pHab it is advisable to select the wavelengths for the calculation of the equilibrium constants in order to avoid generating a very large Jacobian. The traditional argument for wavelength selection is that choosing the wavelength corresponding to an absorbance maximum in a species' spectrum gives the most information concerning the concentration of that species. Other wavelengths give less information and each implies the existence of another set of unknown molar absorbances. Therefore the purpose of doing multi-wavelength calculations is to ameliorate a little the effects of experimental noise, but its value is not as great as might be expected since the ratio of the number of observations to the number of unknown quantities does not increase by a large amount. Furthermore, the truth is that molar absorbance and equilibrium constant are multiplied together in Beer's law (eqn. 1) so that there is a correlation between these quantities that is masked by the method of calculation. Indeed these methods of calculation were developed precisely because the direct method which treats both ε and β as parameters at the same level often fails because of that correlation. For this reason also an error in equilibrium constant may lead, through correlation of errors, to a calculated molar absorbance having a negative value. In pHab we do not apply non-negativity constraints, being of the opinion that a negative molar absorbance indicates something wrong with the model/data that should not be ignored. However, whilst the calculation may use selected wavelengths one should, as part of the analysis of results, calculate the spectra at all wavelengths. It is also important that the final estimates of errors on the equilibrium constants be calculated using the full Jacobian.

We believe that the last major problem with determining equilibrium constants from spectrometric titration data concerns baseline errors. It seems to be essential that the cuvette is moved as little as

possible during a titration as minute cuvette movements can introduce baseline errors above the level of the noise in modern instruments. Baseline drift remains a serious problem. Otherwise programs like pHab offer significant possibilities for the study of equilibria in solution, especially when combined advanced techniques of factor analysis.[15, 16]

Received June 30th, 1998

Acknowledgements. We gratefully acknowledge the EC grant "Metals and Environmental Problems" (Human Capital and Mobility) which has facilitated communications between Leeds and Florence.

REFERENCES

1. P. GANS, A. SABATINI and A. VACCA, *Talanta*, **43**, 1739 (1996).
2. R.A. ALCOCK, F.R. HARTLEY and D.E. ROGERS, *J. Chem. Soc. Dalton Trans.*, **1978**, 115.
3. P.D. TAYLOR, I.E.G. MORRISON and R.C. HIDER, *Talanta*, **35**, 507 (1988).
4. L. ZEKANY and I. NAGYPAL, in D.J. Leggett (Ed), *Computational Methods For The Determination of Formation Constants*, Plenum Press, New York, 1985, p 291.
5. L.G. SILLÉN and B. WARNQVIST, *Arkiv. Kemi*, **31**, 377 (1968).
6. D.J. LEGGETT in D.J. Leggett (Ed), *Computational Methods For The Determination of Formation Constants*, Plenum Press, New York, 1985, p 159.
7. R. CAZALLAS, M.J. CITORES, N. ETXEBARRIA, L.A. FERNÁNDEZ AND J.M. MADARIAGA, *Talanta*, **41**, 1637 (1994).
8. J.L. BELTRÁN, R. CODONY and M.D. PRAT, *Anal. Chim. Acta*, **276**, 441 (1993).
9. H. GAMPP, M. MAEDER, C.J. MEYER and A. ZUBERBÜHLER, *Talanta*, **32**, 95, 257 (1985).
10. M. MAEDER and A. ZUBERBÜHLER, *Anal. Chem.*, **62**, 2220 (1990).
11. http://www.chim1.unifi.it/group/vacsab/phab.htm
12. J.J. CRUYWAGEN, J.B.B. HEYNS and E.A. ROHWER, *Polyhedron*, **17**, 1741 (1998).
13. E.R. MALINOWSKI and D.G. HOWERY, *Factor Analysis in Chemistry*, Robert E. Krieger Publishing Company, Malabar, Florida, 1989.
14. P. GANS, *Data fitting in the Chemical Sciences*, Wiley, Chichester, 1992.
15. R. TAULER, A. IZQUERDO-RIDORSA and E. CASASSAS, *Chemometrics and Intelligent Lab. Systems*, **18**, 293 (1993).
16. R.M. DYSON, S. KADERLI, G.A. LAWRANCE, M. MAEDER and A.D. ZUBERBÜHLER, *Anal. Chim. Acta*, **353**, 381 (1997).