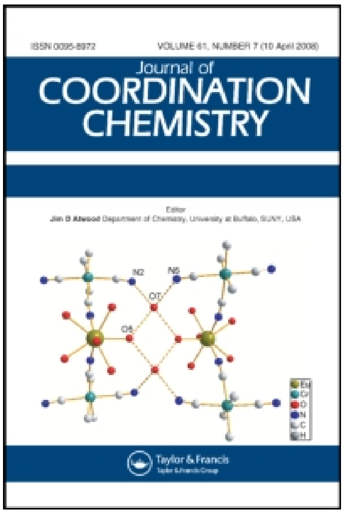

Journal of Coordination Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713455674

THE SIGNIFICANCE AND RELATIONSHIP OF CORRELATION COEFFICIENTS FOR STEPWISE FORMATION CONSTANTS (K) AND CUMULATIVE FORMATION CONSTANTS (β)

Kenneth N. Raymond[a]; James M. Mccormick[a]
[a] Department of Chemistry, University of California, Berkeley, CA, USA

To cite this Article Raymond, Kenneth N. and Mccormick, James M.(1998) 'THE SIGNIFICANCE AND RELATIONSHIP OF CORRELATION COEFFICIENTS FOR STEPWISE FORMATION CONSTANTS (K) AND CUMULATIVE FORMATION CONSTANTS (β)', Journal of Coordination Chemistry, 46: 1, 51 — 57

To link to this Article: DOI: 10.1080/00958979808047194

URL: http://dx.doi.org/10.1080/00958979808047194

J. Coord. Chem., 1998, Vol. 46, pp. 51–57
Reprints available directly from the publisher
Photocopying permitted by license only

Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in Malaysia.

THE SIGNIFICANCE AND RELATIONSHIP OF CORRELATION COEFFICIENTS FOR STEPWISE FORMATION CONSTANTS (K) AND CUMULATIVE FORMATION CONSTANTS (β)

KENNETH N. RAYMOND* and JAMES M. McCORMICK

Department of Chemistry, University of California, Berkeley, CA 94720-1460, USA

(Received 19 November 1996; Revised 27 May 1997; In final form 11 September 1997)

The relationship between the correlation coefficients for stepwise formation constants (K) and cumulative formation constants (β) is derived and discussed. The correlation between pairs of cumulative formation constants in a non-linear least-squares fit does not generally reflect the correlation between the underlying stepwise formation constants. The consequences of these results are discussed.

Keywords: Formation constants; correlation; refinement; solution thermodynamics

INTRODUCTION

Increasingly the quantitative description of metal complex stability and equilibria are of concern to chemists in fields as varied as environmental monitoring of toxic metals[1] and medical diagnostic agents based on metal ions.[2,3,4] Consequently much more information on stability constants for metal complex formation is appearing in the literature. While the use of such data is generally well understood, the limitations due to correlations between the parameters is not generally appreciated. It is the purpose of this article to illustrate the marked differences in the apparent correlation of chemical equilibria as described by cumulative *vs.* overall stability constants.

---

* Corresponding author. E-mail: raymond@garnet.berkeley.edu.

In general the theory of solution thermodynamics writes equilibrium reactions as cumulative formation constants ($\beta_{m\ell h}$) corresponding to the formation of a complex of composition $M_mL_\ell H_h$. The $M_mL_\ell H_h$ are regarded as having been formed by an equilibrium reaction involving the aqueous metal ion (M), free deprotonated ligand (L) and aqueous protons (H), respectively (**1**).

$$\beta_{m\ell h} = \frac{[\mathrm{M_mL_\ell H_h}]}{[\mathrm{M}]^{\mathrm{m}}[\mathrm{L}]^{\ell}[\mathrm{H}]^{\mathrm{h}}} \tag{1}$$

Many least-squares refinement programs carry out their computations using these overall formation constants,[5,6] although others also refine on the stepwise formation constants, K_j.[6] The $\beta_{m\ell h}$ and the K_j form two different basis sets for the mathematical description of the same physical process. Since these are product relationships the logarithmic values are linearly related. What may not be obvious is that as a consequence of this relationship there is a great difference between correlations involving one set of variables *vs.* the other set. It will be demonstrated that the $\beta_{m\ell h}$ constants tend to be highly correlated even if there is little or no correlation between the underlying K_j constants, and a method to convert the correlation coefficients of the $\beta_{m\ell h}$ to those of the K_j will be presented.

RESULTS AND ANALYSIS

For the set of individual observations y_i, a non-linear least-squares fit seeks to minimize the difference between the observed and calculated values of the observations (**2**).

$$R = \sum_{i=1}^{m} [(\mathrm{y_{obs}} - \mathrm{y_{calc}})/\sigma_{\mathrm{obs}}]_i^2 \tag{2}$$

For potentiometric titrations the y_i are pH values, while for spectrophotometric titrations they are absorbances. In a non-linear least-squares refinement the derivatives of the functional parameter with respect to the variable parameter, D_{ij}, are a key part of the calculation. Suppose for the moment that the $\log K_j$ are being refined, then the D_{ij} are given by (**3**).

$$D_{ij} = (\delta y_i/\delta \log K_j) \tag{3}$$

If m and n are, respectively, the total number of observations and variables, then **D** is an m by n matrix. The variance–covariance matrix S for the non-linear least squares fit is given by (**4**),[7]

$$\mathbf{S} = \left(\frac{\mathrm{R}}{\mathrm{m} - \mathrm{n}}\right)(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1} \tag{4}$$

where the superscript T denotes the transpose of the matrix. The significance of the variance–covariance matrix is that it summarizes the errors in the refined parameters. Defining the individual standard deviations of the *i*th refined parameter as σ_i, and the correlation of the *i*th refined parameter with the *j*th refined parameter as c_{ij} allows us to write the elements of the variance–covariance matrix as (**5**).

$$S_{ij} = \sigma_i \sigma_j c_{ij} \tag{5}$$

The definition of the cumulative stability constants are such that they are products of the underlying stepwise formation constants as formally defined below:

$$\beta_\ell = \prod_{j=1}^{\ell} K_j \tag{6}$$

When this is converted into a logarithmic form (the logarithms of the formation constants are usually the variables being reported and refined), the equation becomes:

$$\log \beta_\ell = \sum_{j=1}^{\ell} \log K_j \tag{7}$$

From this we see that the linear relationship between β's and K's is given by a lower triangular matrix **L** which relates the column vector of log K values to a column vector of logβ values (**8**).[8]

$$\begin{pmatrix} \log \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \log \beta_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \end{pmatrix} \begin{pmatrix} \log K_1 \\ \log K_2 \\ \cdot \\ \cdot \\ \cdot \\ \log K_n \end{pmatrix}$$

$$\frac{\delta \log \beta_\ell}{\delta \log K_j} = \mathbf{L}_{\ell j} = 1 \text{ or } 0 \text{ for } j < \ell \text{ and } = 0 \text{ for } j > \ell \tag{8}$$

The derivative chain rule gives:

$$(\delta y_i/\delta \log K_j) = \sum_{\ell=1}^{n} (\delta y_i/\delta \log \beta_\ell)(\delta \log \beta_\ell/\delta \log K_j) \tag{9}$$

From the original derivative equation (**3**) it is clear how the transformation from K's to β's changes the derivative equation. Thus we obtain (**10**), where $\mathbf{P}_{i\ell} = (\delta y_i/\delta \log \beta_\ell)$ and **L** is from (**7**).

$$\mathbf{D} = \mathbf{P}\mathbf{L}^T \tag{10}$$

The determinent of **L** is 1, hence the inverse of **L** exists so that we can write:

$$\mathbf{P} = \mathbf{D}(\mathbf{L}^{-1})^T \tag{11}$$

Now let **Q** be the variance–covariance matrix for the $\log \beta$ values (**12**),

$$\mathbf{Q} = \frac{R}{m-n}(\mathbf{P}^T\mathbf{P})^{-1} = \left(\frac{R}{m-n}\right)\left[(\mathbf{L})^{-1}\mathbf{D}^T\mathbf{D}(\mathbf{L}^{-1})^T\right]^{-1} \tag{12}$$

which leads to (**13**)

$$\mathbf{Q} = \mathbf{L}\mathbf{S}\mathbf{L}^T \tag{13}$$

by matrix inversion rules, and implies (**14**).

$$\mathbf{S} = \mathbf{L}^{-1}\mathbf{Q}(\mathbf{L}^{-1})^T \tag{14}$$

The transformation equations (**13**) and (**14**) describe how the variance–covariance matrices (and hence the standard deviations and correlation coefficients) are related. As specific examples consider the following cases:

Example 1

Assume that there are three formation constants ($\log K_1$, $\log K_2$, $\log K_3$) which have zero correlation and individual standard deviations of σ_1, σ_2, σ_3. The corresponding $\log \beta_\ell$ are given by the following equations:

$$\begin{aligned}\log \beta_1 &= \log K_1\\ \log \beta_2 &= \log K_1 + \log K_2\\ \log \beta_3 &= \log K_1 + \log K_2 + \log K_3\end{aligned}$$

The **L** matrix that relates the $\log\beta_\ell$ to the $\log K_i$ is thus:

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and the correlation matrix **S** of the $\log K_i$ is:

$$\mathbf{S} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

The variance–covariance matrix for the $\log\beta_\ell$ is then:

$$\mathbf{Q} = \mathbf{LSL}^T = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2+\sigma_3^2 \end{pmatrix}$$

This gives new σ values for $\log\beta_1$ through $\log\beta_3$ of σ_1, $(\sigma_1^2+\sigma_2^2)^{1/2}$, $(\sigma_1^2+\sigma_2^2+\sigma_3^2)^{1/2}$ respectively. The correlation matrix is then given by $c_{ij} = Q_{ij}/\sigma_i\sigma_j$. To see how this changes the numerical values, suppose that the estimated standard deviations in the $\log K_i$ are all about the same value, so $\sigma_1 = \sigma_2 = \sigma_3 = \sigma(\log K)$. Then $\sigma(\log\beta_1) = \sigma(\log K)$, $\sigma(\log\beta_2) = \sqrt{2}\sigma(\log K)$, $\sigma(\log\beta_3) = \sqrt{3}\sigma(\log K)$ and $c_{12} = 1/\sqrt{2}$, $c_{13} = 1/\sqrt{3}$, $c_{23} = \sqrt{2/3}$. The original zero correlation between the $\log K_i$ now gives an 81.8% correlation between $\log\beta_2$ and $\log\beta_3$!

Example 2

In a determination of the protonation constants of TRENCAM,[9] the values for the sequential protonation constants shown in Table I were obtained. The estimated standard deviations ($\sigma(\log\beta_\ell)$) shown for the cumulative protonation constants were obtained from the refinement program.[10] The upper half of the symmetric variance–covariance matrix for the $\log\beta_\ell$ values obtained in the same fit is shown in Table II. The parameters are highly correlated, with all of the correlation coefficients nearly 90% or larger. It is somewhat difficult to make sense out of the correlations with respect to the values of the observations, but using the protocol presented here the corresponding correlation coefficients for the stepwise formation constants can be calculated and are shown in Table III. These calculations are conveniently performed using the matrix manipulation functions of a commercial spreadsheet (a Microsoft Excel spreadsheet of this calculation is included as Supplemental Material).

TABLE I Cumulative and stepwise stability constants for TRENCAM with their estimated standard deviations as determined from the variance–covariance matrix of the non-linear least-squares refinement and the procedure described in the text

#	$\log\beta$	$\sigma(\log\beta)$	$\log K$	$\sigma(\log K)$
1	11.3456	0.0078	11.3456	0.0078
2	20.1257	0.0076	8.7801	0.00326
3	28.7425	0.0078	8.6168	0.00329
4	35.4688	0.0081	6.7263	0.00275
5	41.3288	0.0081	5.8600	0.00280

TABLE II Correlation coefficient matrix of the $\log\beta_\ell$ for TRENCAM

#	1	2	3	4	5
1	1.0000	0.9105	0.9754	0.9253	0.9346
2		1.0000	0.9093	0.8857	0.8814
3			1.0000	0.9407	0.9550
4				1.0000	0.9401
5					1.0000

TABLE III Correlation coefficient matrix of the $\log K_i$ for TRENCAM

#	1	2	3	4	5
1	1.0000	−0.2697	0.2095	−0.0411	0.0269
2		1.0000	−0.8604	0.1673	−0.0932
3			1.0000	−0.2241	0.1268
4				1.0000	−0.6261
5					1.0000

It is clear from Table III that there is only one large correlation, between parameters 2 and 3, which is not surprising given that both these protonations are occurring in the same pH range. This result is not obvious if one were to only consider the correlations of the log(β) values given in Table II, but would be if the refinement had been on the log(K) values. Since statistically independent protonation sites with equal acidities will give apparent pK_a's that are separated by 2log(2),[11] the 0.1 pK_a difference in the protonation constants is at or below the true resolution limit. These two protonation constants are highly correlated; their sum is known very accurately, but their individual values are not.

CONCLUSION

Analysis of the relationship between the correlation coefficients found in non-linear least squares for stepwise formation constants and the cumulative formation constants demonstrates that care must be taken in interpreting the correlation coefficients between these thermodynamic parameters. The cumulative formation constants ($\beta_{m\ell h}$) may be highly correlated even when the underlying stepwise formation constants (K_i) are uncorrelated. It is, therefore, prudent to check the correlation coefficients between both the $\log\beta_{m\ell h}$ and the K_i before interpreting potentiometric or spectrophotometric titration data.

Acknowledgements

This research was supported by NIH grant AI 11744 and by a NIH postdoctoral fellowship to J.M.M.

References

[1] F.R. Hartley, C. Burgess and R. Alcock, "Solution Equilibria", Ellis Horwood, Ltd.; Chichester, England, 1980.

[2] S.J. Franklin and K.N. Raymond, *Inorg. Chem.*, **33**, 5794, (1994).

[3] C. Paul-Roth and K.N. Raymond, *Inorg. Chem.*, **34**, 1408, (1995).

[4] J. Xu, S.J. Franklin, D.W. Whisenhunt Jr. and K.N. Raymond, *J. Am. Chem. Soc.*, **117**, 7245, (1995).

[5] "Computational Methods for the Determination of Formation Constants", Leggett, D.J., Ed.; Plenum Press, NY, 1985.

[6] A.E. Martell and R.J. Motekaitis, "Determination and Use of Stability Constants", VCH Publishers, Inc; New York, 1988.

[7] For a general discussion of solving least squares problems see: C.L. Lawson and R.H. Hanson, "Solving Least Squares Problems", Prentice-Hall; Englewood, New Jersey, 1974.

[8] Note that for hydrolysis reactions some of the elements of the lower triangle of **L** may be 0 because all of the protonation reactions will be independent of the hydrolysis reactions.

[9] S.J. Rogers, C.-W. Lee, C.Y. Ng and K.N. Raymond, *Inorg. Chem.*, **26**, 1622, (1987).

[10] The $\sigma(\log\beta_\ell)$ given in Table I should not be regarded as a true estimate of the accuracy of $\log\beta_\ell$. The standard deviations for a least squares refinement reflect only the internal agreement between the model and the data. Limits on experimental accuracy typically put an absolute limit on the reported standard deviations for $\log\beta_\ell$ as not being much less than 0.01 log units.

[11] E.Q. Adams, *J. Am. Chem. Soc.*, **38**, 1503, (1916).