

ANALISIS DE LAS CURVAS SIGMOIDES. APLICACION A LAS CURVAS DE TITULACION DE OXIDO-REDUCCION

Claude Marmasse

Departamento de Bioquímica

Universidad de Tennessee

(Recibido: octubre 27, 1969)

ABSTRACT:

The logistic method of Reed and Berkson essentially assumes the knowledge of a "perfect" point of the curve; in addition, it introduces an unjustified ponderation of the data. A new method is proposed in which: 1) corrections to estimated values of the parameters are computed iteratively by means of a statistical fitting of the experimental data at an expansion (limited to its linear terms) of the equation of the curve; 2) each point is weighed by a factor inversely proportional to the variance of the experimental potential. This variance is estimated by means of the formula of the propagation of errors; 3) in these conditions, a fitting by the method of least squares is efficient and unbiased and fiducial limits for the calculated values of the parameters are easily obtained; 4) the possible existence of systematic deviations is analyzed by use of the theory of runs.

I. INTRODUCCION

Todavía no ha sido solucionado correctamente el problema del análisis de las curvas de titulación (curvas sigmoides); en efecto, además del procedimiento gráfico clásico de estimación del potencial normal y del número de electrones, el único método conocido de explotación numérica de este tipo de datos es el de Reed y Berkson, o método logístico. Este procedimiento supone esencialmente que se conoce un punto "perfecto"; entonces se analiza el resto de la curva con referencia a este punto. Aunque las medidas en la proximidad del punto de inflexión de la curva son a menudo precisas y confiables, es claro que tal procedimiento tiene un vicio de arbitrariedad. Además, ha sido notado por Clark¹ que este método² presupone implícitamente que la variante de la variable dependiente es constante sobre el dominio experimental de los valores investigados. Pero, y especialmente en el caso de las titulaciones de óxido-reducción, esta hipótesis no se justifica: es muy raro conseguir potenciales estables y bien definidos en la vecindad de 100% de oxidación y de 100% de reducción.

El objeto de este artículo es describir un método que elimina la primera debilidad del método logístico y que remedia la segunda. Esta nueva técnica necesita de cálculos iterativos que se ejecutan fácilmente en una computadora digital.

II. TEORIA FUNDAMENTAL

Se reduce el problema¹ al ajuste de un conjunto de datos $\{v_i, E_i\}$ ($i = 1, 2, \dots, n$) a una ecuación del tipo

$$E = E_0 - K \text{ Log } \frac{v - \epsilon}{A - (v - \epsilon)} \quad (1)$$

donde K es un parámetro dado y donde E_0 , ϵ y A son parámetros que deben ser estimados.

En el caso de una titulación de óxido-reducción, por ejemplo, E es el potencial de un electrodo inerte, E_0 es el potencial normal (standard) de la molécula investigada, v es la cantidad de agente reductor u oxidante añadido durante la titulación, ϵ es la cantidad* de agente reductor u oxi-

* Se puede considerar que ϵ define el origen de la titulación del componente principal.

dante que sería necesario añadir para corregir la cantidad v y A es la cantidad de agente reductor u oxidante necesaria para convertir el sistema principal (si ϵ hubiera sido igual a cero). El parámetro K se define como $K = \frac{RT}{n\mathfrak{F}}$ donde R es la constante de los gases, T es la temperatura absoluta, \mathfrak{F} es el faraday y n es el número de electrones que son intercambiados en la reacción. Este número, que es entero, exceptuando el caso de reacciones muy complejas en las cuales se puede hallar a veces un número aparente de electrones fraccionario, se consigue generalmente por medio de procedimientos gráficos y se puede suponer conocido.

Consideremos el conjunto de los mejores valores, (o valores verdaderos) $\{E_{0v}, A_v, \epsilon_v\}$ y que se puede obtener un conjunto de valores aproximados $\{E_0 = E_{0v} + \Delta E_0, A = A_v + \Delta A, \epsilon = \epsilon_v + \Delta \epsilon\}$, por ejemplo, por medio de un procedimiento gráfico. Los valores verdaderos son una solución estática de la ecuación

$$E = E_{0v} - K \text{Log} \frac{v - \epsilon_v}{A - (v - \epsilon_v)} \quad (2)$$

que se puede transformar en

$$E = E_0 + \Delta E_0 - K \text{Log} \frac{v - \epsilon}{A - (v - \epsilon)} \frac{1 - \frac{\Delta \epsilon}{v - \epsilon}}{1 + \frac{\Delta A + \Delta \epsilon}{A - (v - \epsilon)}} \quad (3)$$

la cual se puede desarrollar, dentro de la región \mathcal{D} definida por

$$\left. \begin{array}{l} \left| \frac{\Delta \epsilon}{v - \epsilon} \right| < 1 \\ \left| \frac{\Delta A + \Delta \epsilon}{A - (v - \epsilon)} \right| < 1, \end{array} \right\} \quad (4)$$

para obtener la relación,

$$E = E_0 + \Delta E_0 - K \operatorname{Log} \frac{v - \epsilon}{A - (v - \epsilon)} + \frac{K \Delta \epsilon}{v - \epsilon} \quad (5)$$

$$+ \frac{K \Delta A}{A - (v - \epsilon)} + \mathcal{O} [(\Delta \epsilon)^2] + \mathcal{O} [(\Delta A)^2] .$$

Si ahora se escribe la ecuación (5) como

$$E = E_0 + \Delta E_0 - K \operatorname{Log} \frac{v - \epsilon}{A - (v - \epsilon)} + \frac{K \Delta \epsilon}{v - \epsilon} + \frac{K \Delta A}{A - (v - \epsilon)} \quad (6)$$

la podemos interpretar como la definición de una regresión multi-lineal de la variable dependiente E sobre las dos variables independientes $K/(v - \epsilon)$ y $K/[A - (v - \epsilon)]$. En estas condiciones, un ajuste estadístico de los datos experimentales a la relación (5) proporcionará las correcciones ΔE_0 , $\Delta \epsilon$ y ΔA , correspondientes a los tres valores aproximados E_0 , ϵ y A . Entonces se consiguen las mejores aproximaciones de los parámetros por medio de las relaciones

$$E_{0,m} = E_0 + \Delta E_0$$

$$\epsilon_m = \epsilon + \Delta \epsilon$$

$$A_m = A + \Delta A .$$

Dentro de la región \mathcal{D} , la iteración es convergente; es decir es posible reducir los incrementos hasta que sean despreciables.

Un ajuste estadístico de los datos experimentales a la regresión (6) por medio del método de mínimos cuadrados no es eficiente, porque, como ha sido notado previamente, la varianza de E no es constante. Aunque es desconocida la variación exacta de esta varianza, se puede estimar por medio de

$$\operatorname{Var} (E) = \left(\frac{\partial E}{\partial v} \right)^2 \operatorname{Var} (v) , \quad (7)$$

y se obtiene,

$$\text{Var}(E) = \left\{ \frac{KA}{(v - \epsilon)[A - (v - \epsilon)]} \right\}^2 \text{Var}(v)$$

La varianza de v puede ser considerada constante, ya que en la mayoría de los casos v es un volumen medido con una bureta o una microbureta. Por eso,

$$\text{Var}(E) \propto \{(v - \epsilon)[A - (v - \epsilon)]\}^{-2} \quad (8)$$

La cantidad $E/\text{Var}(E)$ tiene una varianza constante³. En estas condiciones se pueden conseguir los tres incrementos ΔE_0 , $\Delta \epsilon$ y ΔA con exactitud por medio de un ajuste estadístico de los datos experimentales a la regresión trilineal, por el método de cuadrados mínimos:

$$z = \rho_1 \Delta E + \rho_2 \Delta \epsilon + \rho_3 \Delta A \quad (9)$$

con

$$z \equiv \left[E - E_0 + K \text{Log} \frac{v - \epsilon}{A - (v - \epsilon)} \right] \rho_1$$

$$\rho_1 \equiv (v - \epsilon)^2 [A - (v - \epsilon)]^2$$

$$\rho_2 \equiv K\rho_1 / (v - \epsilon)$$

$$\rho_3 \equiv K\rho_1 / [A - (v - \epsilon)]$$

III. TECNICA DE CALCULO. DETERMINACION DE LIMITES FIDUCIARIOS PARA LOS PARAMETROS.

La ecuación (1) describe el curso de la titulación del componente

principal; es decir que esta relación tiene un sentido físico solamente* cuando

$$v - \epsilon_v \geq 0 \quad (10)$$

$$A_v - (v - \epsilon_v) \geq 0.$$

Sean v_{n_1} y v_{n_2} los valores experimentales extremos de v que satisfacen las relaciones¹⁰. Para el primer ciclo del cálculo, es conveniente tomar $n_1 = 2$, $n_2 = n - 1$, $\epsilon = 0$ y $A = v_n$ pero, para cada ciclo siguiente de la iteración es necesario determinar nuevos valores de n_1 y n_2 con referencia a las relaciones (10).

De una manera general, se definen elementos t_{kl} y p_k ($k, l = 1, 2, 3$) por medio de

$$t_{kl} = \sum_{i=n_1}^{n_2} \rho_K(v_i) \rho_l(v_i), \quad k, l = 1, 2, 3$$

y

$$p_k = \sum_{i=n_1}^{n_2} z(v_i) \rho_k(v_i), \quad k = 1, 2, 3.$$

Sea T la matriz cuadrada simétrica de elemento general t_{jk} ($j, k = 1, 2, 3$). T tiene un inverso T^{-1} de elemento general t^{jk} , entonces se consiguen los tres incrementos por medio de las relaciones

$$\Delta E_0 = \sum_{k=1}^3 t^{1k} p_k$$

$$\Delta \epsilon = \sum_{k=1}^3 t^{2k} p_k$$

$$\Delta A = \sum_{k=1}^3 t^{3k} p_k.$$

* Sin embargo, desde un punto de vista matemático, las igualdades no son aceptables porque el término logarítmico es infinito.

Ya que se ha impuesto que la regresión (9) pasa por el origen, no hay ningún grado de libertad para la desviación de cero de esta regresión; por eso, se calcula la varianza s^2 por medio de la relación

$$(n_2 - n_1 + 1) s^2 = \sum_{i=n_1}^{n_2} [z(v_i)]^2 - (p_1 \Delta E_0 + p_2 \Delta \epsilon + p_3 \Delta A).$$

Entonces, la varianza de los tres incrementos es

$$s^2 t^{11} \quad \text{para } \Delta E_0$$

$$s^2 t^{22} \quad \text{para } \Delta \epsilon$$

$$s^2 t^{33} \quad \text{para } \Delta A$$

y se puede probar independientemente cada incremento versus cero por medio de una comparación de las razones $s \sqrt{t^{11}} / \Delta E_0$, $s \sqrt{t^{22}} / \Delta \epsilon$ y $s \sqrt{t^{33}} / \Delta A$ con la distribución t . Sin embargo, es posible probar globalmente si un conjunto de valores $\{\Delta E_0, \Delta \epsilon, \Delta A\}$ es significativamente diferente de cero; esto es equivalente a probar la consistencia del conjunto de los parámetros calculados con el conjunto de los valores experimentales. Para hacerlo, se puede considerar, según Williams³, la forma cuadrática

$$Q = \sum_{k=1}^3 \sum_{j=1}^3 b_k b_j t^{kj},$$

con $b_1 \equiv \Delta E_0$, $b_2 \equiv \Delta \epsilon$ y $b_3 \equiv \Delta A$; Q es una suma de cuadrados con 3 grados de libertad y por lo tanto, la expresión $Q/(3s^2)$ está distribuida como F , con 3 y $n_1 - n_2 - 2$ grados de libertad. Es decir, un conjunto de valores calculados $\{\Delta E_0, \Delta \epsilon, \Delta A\}$ no es significativamente diferente del conjunto $\{0, 0, 0\}$ si $Q \leq 3s^2 F$, a cierto nivel de probabilidad.

IV. EJEMPLO NUMERICO Y DISCUSION

A título de ejemplo, se ha aplicado este nuevo método al análisis de la titulación de un indofenol, realizada por Clark y Cohen (Tabla I). Con

v (ml)	E (mV)
0	287.0
2	158.1
4	147.9
6	141.5
8	136.8
10	132.7
12	129.2
14	125.6
16	122.4
18	119.2
20	115.9
22	112.4
24	108.5
26	104.1
28	98.5
30	90.5

Tabla I. Titulación de un indofenol. Datos de Clark y Cohen⁴ citados por Clark¹.

el método logístico, se consigue:

$$E_0 = 0.1224 \text{ V}$$

$$\epsilon = 0 \text{ ml}$$

$$A = 32.68 \text{ ml.}$$

Además, Clark, notó la presencia de un pequeño error sistemático, o sea una pequeña desviación de la fórmula (1); la incertidumbre del potencial normal fue estimada como 0.1 o quizás 0.2 mV.

Por medio del método iterativo se consigue

$$E_0 = 0.1217_5 \text{ V}$$

$$\epsilon = 0.1_9 \text{ ml}$$

$$A = 32.4_7 \text{ ml.}$$

Se nota, en primer lugar, que los dos grupos de resultados no son muy diferentes. Eso proviene de que la curva experimental seleccionada es especialmente buena; de hecho, no sería posible emplear el método logístico con otro tipo de datos a causa de la necesidad de un punto "perfecto", mientras que en el método iterativo no se presupone nada con referencia a cualquier punto.

Por otra parte, el método logístico no permite practicar un análisis estadístico y particularmente, no permite cuantificar la incertidumbre sobre el valor del potencial normal; lo único que es posible, es una comparación cruda entre los valores calculados y los valores observados, aunque el método iterativo permite calcular la varianza del potencial normal. Y se puede interpretar fácilmente esta varianza porque (a causa de la ponderación introducida y de la manera de realizar el cálculo) ΔE_0 tiene una distribución normal (o casi normal) con 0 como medida y $s^2 t^{11}$ como varianza). Es decir, en el caso presente, el potencial normal E_0 está distribuido normalmente (o casi normalmente) con respecto a 121.75 mV con una varianza de $1.08 \times 10^{-2} \text{ mV}^2$. Los límites fiduciaros correspondientes se dan en la Tabla II; estos límites precisan la estimación de Clark, pero son menos optimistas.

2P	Límites fiduciaros (mV)
0.050	0.22
0.010	0.31
0.005	0.34
0.001	0.43

Tabla II Límites fiduciaros del potencial normal correspondiente a los datos de la Tabla I.

Es claro que tanto el éxito como el valor del método están íntimamente ligados a la estimación correcta de la varianza de E_0 . Por otra parte, se sabe que la estimación de la varianza por medio de la fórmula (7) de propagación de los errores no es necesariamente eficaz y que a veces, se pueden introducir errores substanciales. Sin embargo, en el caso presente, no es necesario conocer $Var(E)$ con una gran precisión, ya que esta varianza se emplea sólo como un término de ponderación; además, y porque la función E es simétrica, los momentos de orden impar respecto a la medida son iguales a cero y eso disminuye automáticamente el número de términos que sería preciso añadir a

$$\left(\frac{\partial E}{\partial v}\right)^2 Var(v)$$

para conseguir la expresión completa de $Var(E)$. Todo esto conduce a conjeturar que la estimación (8) es una forma límite, que es satisfactoria para ponderar la regresión (9). Sin embargo, ya que el término ponderador se vuelve infinito para 0% y 100% de reducción se puede producir una pérdida ligera de información. Desde un punto de vista físico, sería más satisfactorio un término de la forma

$$\{(v - \epsilon) [A - (v - \epsilon)] + B\}^{-2}$$

donde B es una constante que puede ser pequeña y que prácticamente no introduciría ningún cambio numérico sobre la influencia de los términos ponderadores, pero permitiría eventualmente utilizar los dos puntos extremos. Serían deseables experimentos para estimar el orden de magnitud de B .

Finalmente, se puede estudiar la existencia de una desviación sistemática por medio del análisis estadístico de la secuencia de los signos de los residuales. En el caso de los datos de Clark y Cohen, la secuencia de estos signos es como sigue:

- - - + + + - - - - - + + + .

Si cada arreglo es igualmente probable, la probabilidad de que se arreglen los seis signos positivos en dos series de tres es⁵:

$$\frac{(0+0+2)!}{0!0!2!} \binom{9+1}{0+0+2} \Bigg/ \binom{6+9}{6} = 0.009 .$$

Esta pequeña probabilidad es indicativa de una desviación sistemática que, por ejemplo, podría reflejar la existencia de un ligero fenómeno perturbador que no se toma en cuenta en la ecuación (1).

BIBLIOGRAFIA

1. W.M. Clark, "Oxidation-reduction potentials of organic systems", The Williams and Wikins Co., Baltimore, 1960.
2. L.J. Reed and J.J. Berkson, J. Chem. Phys., 33 (1929) 760.
3. E.J. Williams, "Regression analysis", John Wiley and Sons, New York, 2nd print. 1964.
4. Clark and Cohen, Public Health Reports (U.S.), 39, No. 904, 381, quoted by Clark (ref. 1).
5. W. Feller, "An introduction to probability theory and its applications", John Wiley and Sons, New York, 2nd ed., 1960, p. 60.

RESUMEN

El método logístico de Reed y Berkson supone, esencialmente, el conocimiento de un punto "perfecto" de la curva; además, se introduce una ponderación injustificada de los datos. Se propone un nuevo método con las siguientes ventajas: 1) Las correcciones a los valores estimados de los parámetros se calculan iterativamente, por medio de un ajuste estadístico de los datos experimentales, usando un desarrollo, restringido hasta términos lineales de la ecuación de la curva; 2) cada punto se pesa por un factor inversamente proporcional a la varianza del potencial experimental. Esta varianza se estima usando la fórmula de propagación de errores; 3) en estas condiciones, el ajuste por el método de mínimos cuadrados es eficiente y los límites de confianza para los valores calculados de los parámetros se pueden obtener fácilmente.