

What Is the Best Way to Fit Equilibrium Data?

Introduction — The Question

In Chapter 9 I discuss the variation of the equilibrium constant with temperature. I mentioned there (p. 260) that

Because $\ln K$ is proportional to $\Delta_r G^\circ$ (equation 9.11) and $\ln \rho$ is specified at specific values of T and P , it follows that an expression relating $\ln K$ and $\ln \rho$ is logically equivalent to one giving $\Delta_r G^\circ$ as a function of T and P . In Chapter 4 we saw that an equation giving $\Delta_r G^\circ$ as a function of T and P is called a *fundamental equation*, and that it implicitly contains information on the variation of *all* thermodynamic parameters with T and P .

Similarly, knowing $\log K$ as a function of T is equivalent to knowing $\Delta_r G^\circ$ as a function of T . Therefore whatever equation is used to represent the variation of $\log K$ with T contains implicit functional relationships between the fit coefficients and the various temperature derivatives of $\Delta_r G^\circ$, such as $\Delta_r H^\circ$ and $\Delta_r C_P^\circ$. The simplest example is the equation

$$\log K = a/T + b$$

which implies that $\Delta_r H^\circ$ and $\Delta_r S^\circ$ are constants and that $\Delta_r C_P^\circ$ is zero. A more complex example is the density model. Many equations have been suggested and used, containing various combinations of reciprocal, log and exponential terms. This raises the question, what in fact is the theoretically “best” form for such an equation? Or is there one?

The Clarke and Glew Answer

Clarke and Glew (1966) address this question by deriving an equation in which provides an unbiased, and therefore “best” estimate of the variation of $\Delta_r H^\circ$ with T . In their own words (and equation numbers),

Since ΔH_T° is a well-behaved function of T , its value can be properly expressed as a perturbation on the value of ΔH_θ° at reference θ using the Taylor’s series expansion,

$$\Delta H_T^\circ = \Delta H_\theta^\circ + \Delta C_{P\theta}^\circ(T - \theta) + \frac{1}{2}(d\Delta C_{P\theta}^\circ/dT)_\theta(T - \theta)^2 + \frac{1}{6}(d^2\Delta C_{P\theta}^\circ/dT^2)_\theta(T - \theta)^3 + \frac{1}{24}(d^3\Delta C_{P\theta}^\circ/dT^3)_\theta(T - \theta)^4 \quad (3)$$

Similarly for ΔS_T° the relationship to ΔS_θ° is given by

$$\Delta S_T^\circ = \Delta S_\theta^\circ + \int_\theta^T (\Delta C_{PT}^\circ/T) dT \quad (4)$$

in which ΔC_{PT}° is defined by the Taylor’s series expansion,

$$\Delta C_{PT}^\circ = \Delta C_{P\theta}^\circ + (d\Delta C_{P\theta}^\circ/dT)_\theta(T - \theta) + \frac{1}{2}(d^2\Delta C_{P\theta}^\circ/dT^2)_\theta(T - \theta)^2 + \frac{1}{6}(d^3\Delta C_{P\theta}^\circ/dT^3)_\theta(T - \theta)^3 \quad (5)$$

The relationship between $R \ln K$ and T is then obtained by substituting equation (3) and the integrated equation (4) into

$$\begin{aligned} R \ln K &= -\Delta G_T^\circ \\ &= \Delta S_T^\circ - (\Delta H_T^\circ / T) \end{aligned} \quad (2)$$

resulting in their equation (6) (not shown here). They then point out that as $(T - \theta)$ is usually much smaller than θ , it is advantageous to rescale T in equation (6) by using the new variable $x = (T - \theta)/\theta$ to replace T . Making this substitution and simplifying results in the final equation,

$$\begin{aligned} R \ln K &= -\frac{\Delta G_\theta^\circ}{\theta} + \frac{\Delta H_\theta^\circ}{\theta} \left[\frac{x}{1+x} \right] + \Delta C_{P\theta}^\circ \left[x^2 \sum_{n=1}^{\infty} \frac{n}{n+1} (-x)^{n-1} \right] \\ &\quad + \frac{\theta}{2} (d\Delta C_P^\circ/dT)_\theta \left[x^3 \sum_{n=1}^{\infty} \frac{n}{n+2} (-x)^{n-1} \right] \\ &\quad + \frac{\theta^2}{6} (d^2\Delta C_P^\circ/dT^2)_\theta \left[x^4 \sum_{n=1}^{\infty} \frac{n}{n+3} (-x)^{n-1} \right] \\ &\quad + \frac{\theta^3}{24} (d^3\Delta C_P^\circ/dT^3)_\theta \left[x^5 \sum_{n=1}^{\infty} \frac{n}{n+4} (-x)^{n-1} \right] \end{aligned} \quad (7)$$

where the Taylor's series is terminated at $d^3\Delta C_P^\circ/dT^3)_\theta$, giving a five-variable equation, which is considered sufficient to properly represent any accurate set of measurements.

The Working Equation

Equation (7) is then rewritten as

$$R \ln K = b_0 + b_1 u_1 + b_2 u_2 + b_3 u_3 + b_4 u_4 + b_5 u_5 \quad (10)$$

where $u_1 \cdots u_5$ are the temperature functions

$$\begin{aligned} u_1 &= \left[\frac{x}{1+x} \right] \\ u_2 &= \left[x^2 \sum_{n=1}^{\infty} \frac{n}{n+1} (-x)^{n-1} \right] \\ &\vdots \\ u_5 &= \left[x^5 \sum_{n=1}^{\infty} \frac{n}{n+4} (-x)^{n-1} \right] \end{aligned}$$

$T^{\circ}\text{C}$	K
5	359.20×10^{-5}
10	357.45×10^{-5}
15	352.93×10^{-5}
20	346.94×10^{-5}
25	338.76×10^{-5}
30	329.78×10^{-5}
35	319.52×10^{-5}
40	308.51×10^{-5}
45	296.40×10^{-5}

Table 1: The equilibrium constant for the ionization of cyanoacetic acid at various temperatures, from Feates and Ives (1956).

and $b_0 \dots b_5$ are least squares fit coefficients. Comparing equations (10) and (7) shows that the coefficients are related to thermodynamic parameters as follows,

$$\begin{aligned}
b_0 &= -\frac{\Delta G_{\theta}^{\circ}}{\theta} & b_1 &= \frac{\Delta H_{\theta}^{\circ}}{\theta} \\
b_2 &= \Delta C_{P\theta}^{\circ} & b_3 &= \frac{\theta}{2}(d\Delta C_P^{\circ}/dT)_{\theta} \\
b_4 &= \frac{\theta^2}{6}(d^2\Delta C_P^{\circ}/dT^2)_{\theta} & b_5 &= \frac{\theta^3}{24}(d^3\Delta C_P^{\circ}/dT^3)_{\theta}
\end{aligned}$$

from which the parameters are derived as

$$\begin{aligned}
\Delta G_{\theta}^{\circ} &= -b_0\theta \\
\Delta H_{\theta}^{\circ} &= b_1\theta \\
\Delta S_{\theta}^{\circ} &= b_0 + b_1 \\
\Delta C_{P\theta}^{\circ} &= b_2 \\
(d\Delta C_P^{\circ}/dT)_{\theta} &= 2b_3/\theta
\end{aligned}$$

and so on.

Cyanoacetic Acid Example

This method has been coded in the Matlab program CGfit.m, included in the Matlab folder. Clarke and Glew (1966) use the data of Feates and Ives (1956) on the ionization of cyanoacetic acid to demonstrate the usefulness of this method. Their data are shown in Table 1, and the results obtained by Clarke and Glew are shown in Table 2. These can be used to verify the results of the Matlab program. The Feates and Ives data as output from the program CGfit.m are shown in Figure 1

Parameter ($\theta = 298.15$ K)	Value
$\Delta G_{\theta}^{\circ}$ cal mol $^{-1}$	3369.59
$\Delta H_{\theta}^{\circ}$ cal mol $^{-1}$	-894.1
$\Delta C_{p\theta}^{\circ}$ cal mol $^{-1}$ K $^{-1}$	-36.1
$(d\Delta C_{p}^{\circ}/dT)_{\theta}$ cal mol $^{-1}$ K $^{-2}$	0.39
$(d^2\Delta C_{p}^{\circ}/dT^2)_{\theta}$ cal mol $^{-1}$ K $^{-3}$	-0.086
$(d^3\Delta C_{p}^{\circ}/dT^3)_{\theta}$ cal mol $^{-1}$ K $^{-4}$	-0.009

Table 2: Thermodynamic parameters derived from the data of Feates and Ives by Clarke and Glew (1966), using all five terms.

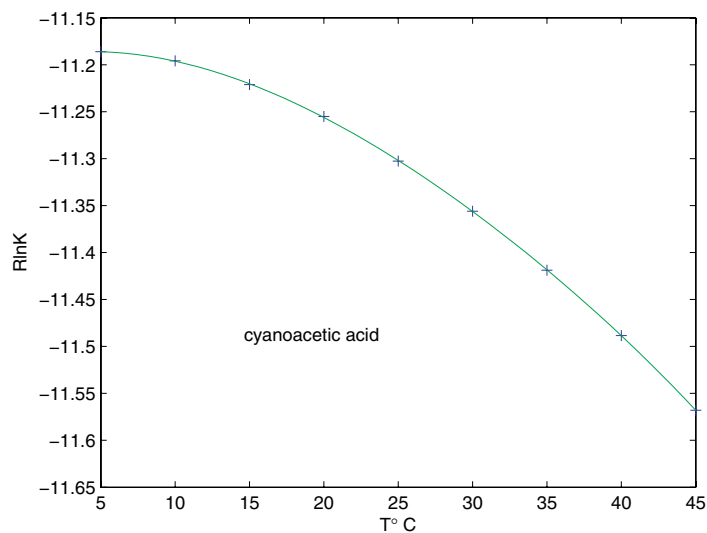


Figure 1: The data of Feates and Ives (+) and the line fitted by the Clarke and Glew method.

Statistical Significance

Clarke and Glew present an extensive statistical study of their results, and conclude that the data of Feates and Ives are not sufficiently precise to justify use of five constants, and that therefore the value of the third derivative of C_p° is not statistically significant. Bolton (1970) also discusses the statistics of the fitting process, and says that in the study of some 125 sets of acid ionization constants in the temperature range 0 to 70°C, none required more than the three temperature-variable form of the equation to give the statistically best representation of the data, i.e.,

$$R \ln K = b_0 + b_1 u_1 + b_2 u_2 + b_3 u_3$$

Even the use of the fourth variable in the Feates and Ives data is borderline.

It appears then that even for data of the highest quality, only three temperature variables need be used, and perhaps for high temperature data where experimental difficulties are greater, only two temperature variables would be sufficient. This would still result in values for $\Delta_r G^\circ$, $\Delta_r H^\circ$, $\Delta_r S^\circ$, and $\Delta_r C_p^\circ$.

HCl Example

Just to see what would happen, I fit data for the ionization of $\text{HCl}(aq)$ from SUPCRT92 using CGfit.m, by changing the value of θ and rerunning the program several times. The data and fitted line are shown in Figure 2, and the calculated parameters are shown in Figure 3. These results might be compared with those in Figure 10.12 in the text.

Conclusions

Bolton (1970) says

When used correctly the C and G procedure gives a thermodynamically meaningful and non-empirical representation of the data, and the values of the thermodynamic functions obtained are the best unbiased estimates calculable from the data.

He adds in a footnote that the use of the term “non-empirical” can be fairly questioned, but that the C and G analysis is considerably more general than any other, and that the assumptions involved could be said “to reduce to the single assumption that the thermodynamic functions involved behave in a rational and predictable manner.”

References

- Bolton, P.D., 1970, Calculation of thermodynamic functions from equilibrium data. Jour. Chem. Education, v. 47, pp.638–641.
- Clarke, E.C.W., and Glew, D.N., 1966, Evaluation of thermodynamic functions from equilibrium constants. Trans. Faraday Soc., v. 62, pp. 539–547.

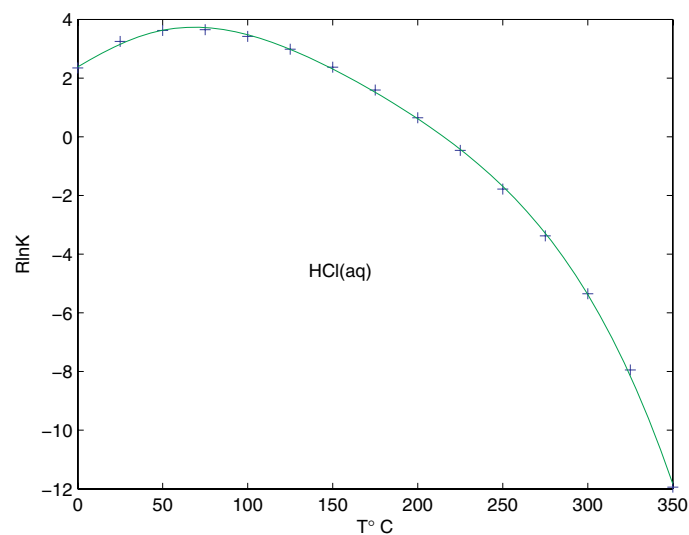


Figure 2: Ionization constants for $\text{HCl}(aq)$ from SUPCRT92. The line is calculated using the Clarke and Glew method.

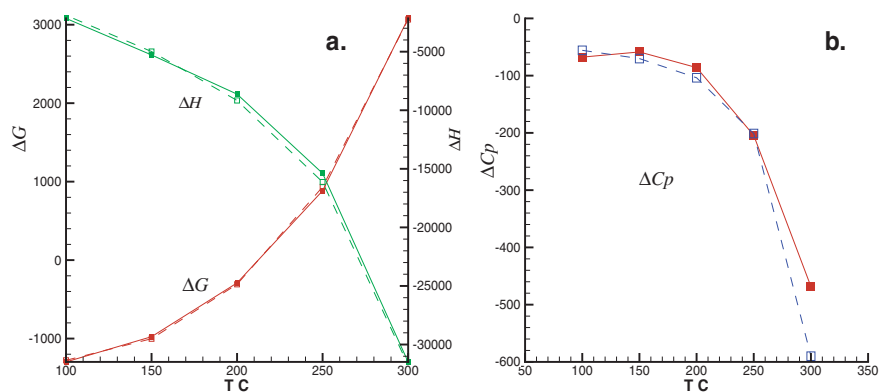


Figure 3: Thermodynamic parameters for the ionization of $\text{HCl}(aq)$. a. In red— $\Delta_r G^\circ$. In green— $\Delta_r H^\circ$. Solid line—C&G. Dashed line—SUPCRT92. b. $\Delta_r C_p^\circ$. Solid line—C&G. Dashed line—SUPCRT92.

Feates, F.S., and Ives, D.J.G., 1956, The ionization functions of cyanoacetic acid in relation to the structure of water and the hydration of ions and molecules. Jour. Chem. Soc., pp. 2798–2812.