File: /General/MLAB-Text/Papers/msb/msb.tex

Step-Wise Multiple-Site Binding

Gary D. Knott, Ph.D.
Civilized Software, Inc.
12109 Heritage Park Circle
Silver Spring, MD 20906 USA
Tel. (301) 962-3711
Email: csi@civilized.com

Cooperative effects often arise when there are multiple binding sites for a ligand that are located spatially close to one another. Some of the issues that arise are discussed in "Mixed and uniform cooperativity of ligand binding to multisite proteins: The cooperativity types allowed by the Adair equation and conditions for them" by Edward P. Whitehead, in the Journal of Theoretical Biology, pp. 153:170, vol. 87, 1980.

A particular situation of interest is that where many F molecules bind step-wise through a series of reactions to a G molecule with distinct affinities. For example, oxygen binds to hemaglobin in this manner. This is one explanation for apparent cooperative binding. Thus we may consider the situation:

$$\begin{aligned}
F + G_0 &\underset{D_1}{\overset{A_1}{\rightleftharpoons}} G_1, \\
F + G_1 &\underset{D_2}{\overset{A_2}{\rightleftharpoons}} G_2, \\
&\vdots \\
F + G_{i-1} &\underset{D_i}{\overset{A_i}{\rightleftharpoons}} G_i, \\
&\vdots \\
F + G_{N-1} &\underset{D_N}{\overset{A_N}{\rightleftharpoons}} G_N,
\end{aligned}$$

where G_0 is defined as free G.

When N (the number of F-binding sites on each G molecule) is not large, the kinetic model differential equations can be used in curve-fitting; however, in general, we deal with the equilibrium model instead.

Define the molar equilibrium constants $K_i = A_i/D_i$. Let $F(t)$ be the concentration of (unbound) F at time t, let $G_i(t)$ be the concentration of G_i at time t, and let F_e and G_{ie} be these concentrations at $t = t_e$, the time when our system approaches equilibrium. Then: $K_i = G_{ie}/(G_{i-1,e}F_e)$, so $K_1K_2 \ldots K_i = G_{ie}/(G_{0e}F_e^i)$. Now we may define $B_i = K_1K_2 \ldots K_i$.

Now, let F_b be the concentration of bound F molecules at equilibrium, so $F_b = G_{1e} + 2G_{2e} + \ldots + NG_{Ne}$. Note that $F_b + F_e$ is the total concentration of F present, *i.e.* $F_b + F_e = F(0)$. Also, let H be the concentration of G molecules in either a bound or free state, so $H = G_{0e} + G_{1e} + \ldots + G_{Ne}$.

Now, define v as the mean number of F molecules bound to each G molecule. Then $v = F_b/H$, or

$$v = (G_{1e} + 2G_{2e} + \ldots + NG_{Ne})/(G_{0e} + G_{1e} + \ldots + G_{Ne}).$$

But, $G_{ie} = B_iG_{0e}F_e^i$, so

$$v = (B_1G_{0e}F_e + 2B_2G_{0e}F_e^2 + \ldots + NB_NG_{0e}F_e^N)/(G_{0e} + B_1G_{0e}F_e + B_2G_{0e}F_e^2 + \ldots + B_NG_{0e}F_e^N),$$

or

$$v = (B_1F_e + 2B_2F_e^2 + \ldots + NB_NF_e^N)/(1 + B_1F_e + B_2F_e^2 + \ldots + B_NF_e^N).$$

This equation is known as the Adair-Klotz stepwise equilibrium model.

Now given data points (F_e, F_b), each based on different initial values of H and $F_e + F_b$, corresponding data points of the form $(F_e, F_b/H)$ can be constructed, and v can be treated as a function of F_e and fit to such data thus allowing the parameters B_1, B_2, ..., B_N to be estimated (which hence provides estimates of the equilibrium constants K_1, K_2, ..., K_N). If the number of sites, N, is not known, N can be set to 1, 2, 3, etc., and that value of N which yields the best fit can be taken as the estimate of the true N-value. Note the model for data points of the form (F_e, G_{0e}) can be expressed in terms of a `ROOT` expression in $G_0(0)$ and K_1, ..., K_N; but no nice general form is apparent. Note also that data of the form $(F_e, F_b/H)$ has error in both the first and second components. This means that, at a minimum, correct weights should be used in fitting.

The following is a *MLAB* dialog that demonstrates the above mentioned curve fitting for $N = 4$. We first define the model function $v(F_e)$, read-in the data, set the initial guesses for B_1, B_2, B_3 and B_4, and then fit the data to the model.

```
fct p(x) = 1+sum(I,1,N, B[I]*x^I)
```

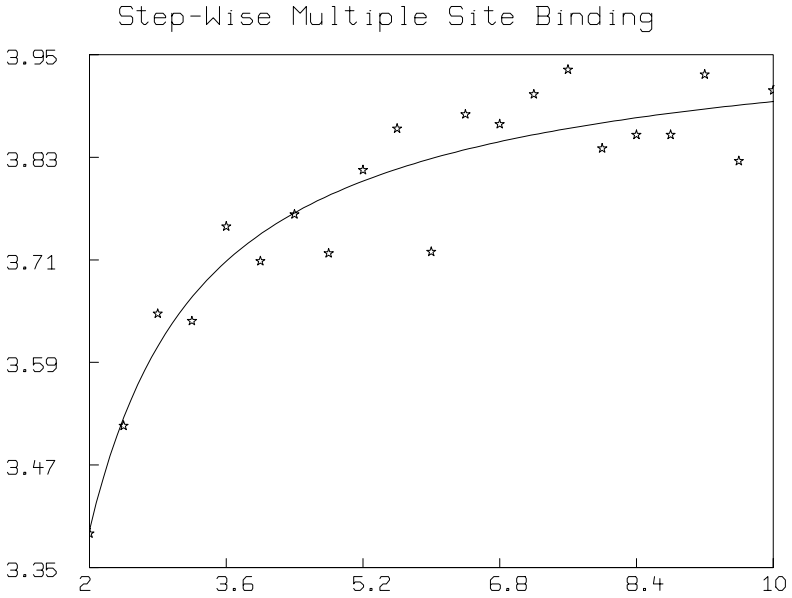
```
fct v(fe) = fe*p'x(fe)/p(fe) /* model function */
N = 4 /* number of reaction steps */
data = read(msb, 21, 2)
t = minv((data col 1)):maxv(data col 1)!100

/* initial guesses */
B = 2:5

fit(B), v to data with weight ewt(data)

final parameter values
      value               error                dependency    parameter
   2.7807664141      164.6641428061          0.9999816051   B[1]
   0.0702266931       16.0436904339           0.999253623   B[2]
   3.1455369331      133.7340925626          0.9999952965   B[3]
   2.7828358122      116.4894630769          0.9999972398   B[4]
3 iterations
CONVERGED
best weighted sum of squares = 3.047449e+01
weighted root mean square error = 1.338886e+00
weighted deviation fraction = 9.451356e-03
R squared = 8.959082e-01

draw data lt none pt star ptsize 0.01
draw points(v,t)
top title "Step-Wise Multiple Site Binding"
view
```



In order to curve-fit for $B_1, \ldots, B_N$, this model often requires accurate guesses, and these may be difficult to obtain. One method is to obtain such guesses following section 10 of "Models in Regression", by Peter Sprent. The idea is to convert our model to a related linear model which admits a unique solution with arbitrary guesses.

We have $v = B_1F_e(1-v) + B_2F_e^2(2-v) + \ldots + B_NF_e^N(N-v)$, so we can establish the data matrix whose columns are values of $F_e(1-v), F_e^2(2-v)$, $\ldots$, $F_e^N(N-v)$, v, and fit the above model to obtain $B_1, \ldots, B_N$, using as weights the values $1 + B_1F_e + \ldots + B_NF_e^N$, based on the initial guesses for $B_1, \ldots, B_N$. The resulting B-values should then be used to recompute the weights $1 + B_1F_e + \ldots + B_NF_e^N$, and then this fitting process should be reiterated in this manner. *MLAB* can provide such iterative reweighting by specifying a weight function. In general constraints must be used, or special deviations from this process must be introduced to cope with negative B-values; for example each value B_i may be replaced with $\max(B_i, 0)$. Finally, the results obtained may be used as initial guesses for a non-linear regression analysis using the original model for v.

Consider the polynomial $p(x) = 1 + B_1x + B_2x^2 + \ldots + B_Nx^N$. Note $v(F_e) = F_e\frac{dp}{dx}(F_e)/p(F_e)$. Fletcher, Spector, and Ashbrook have shown in their paper, "Analysis of macromolecule-ligand binding by determination of

stepwise equilibrium constants", Biochemistry, Vol. 9, pp. 4580:4587, 1970, that if $p(x)$ has N real roots, R_1, R_2, ..., R_M of multiplicities N_1, N_2, ..., N_M respectively, then by partial-fraction decomposition, v can be written as:

$$v(F_e) = N_1 r_1 F_e/(1 + r_1 F_e) + \ldots + N_M r_M F_e/(1 + r_M F_e),$$

where $r_i = -1/R_i$, the negative reciprocals of the roots of p.

This form was suggested by Scatchard as a generalization of the form $rF_e/(1 + rF_e)$ which arises for a single-site simple binding reaction. Its interpretation is that there are M classes of binding sites, with N_i sites in the ith class, having an average equilibrium constant r_i. If the number of classes is not too large, this model can be used to describe $(F_e, v(F_e))$ data by curve-fitting to find the parameters N_1, N_2, ..., N_M, and r_1, r_2, ..., r_M, under the constraints $N_1 + N_2 + \ldots + N_M = N$, and $N_i > 0$, and $r_i > 0$. (Then the values of N_i can be judiciously adjusted to integer-values so as to preserve $N_1 + N_2 + \ldots + N_M = N$; alternatively, M can be taken to be N, the N_i's can be dispensed with (*i.e.* each $N_i = 1$), and the number of r_i's increased accordingly.) Constraints are usually needed to get a satisfactory fit.

Fletcher, et.al., have shown that, when $v(F_e)$ can be expressed in Scatchard's form, the following relations hold. Let L_1, L_2, ..., L_N be defined by: $L_j = r_i$ for $N_{i-1} + 1 \le j \le N_i$, with $N_0 = 0$.

Then: $v(F_e) = \sum_{i=1}^{n} L_i F_e/(1+L_i F_e)$ and $B_j = \sum_{1 \le i_1 < i_2 < \ldots i_j \le N} L_{i_1} L_{i_2} \ldots L_{i_j}$.

Note, B_j is the jth elementary symmetric function on L_1, ..., L_N. It is the coefficient of z^{N-j} in the polynomial $(z + L_1)(z + L_2) \cdots (z + L_N)$. The symmetric functions can be defined in MLAB by the following recursive function definitions.

```
*FUNCTION S(A,Z) = IF Z=N THEN SUM(I,A,Z,L[I]) ELSE \
       SUM(I,A,Z,L[I]*S(I+1,Z+1))
*FUNCTION B(J) = IF J<1 OR J>N THEN 1 ELSE S[1,N-J+1]
```

Another relationship which could be used is:

$B_j(i) = B_j(i-1) + L_i B_{j-1}(i-1)$, where $B_j(i)$ is the jth symmetric function on L_1, L_2, ..., L_i. Then $B_j = B_j(N)$.

Now, the original equilibrium constants may be obtained as: $K_i = B_i/B_{i-1}$ where $B_0 = 1$.

When our polynomial, $p(x)$, does not have N real roots, the Scatchard form is not equivalent to the Adair form. Rather a partial-fraction expansion yields a sum of rational terms with quadratic denominators. Sometimes, there is a "near-by" polynomial with N real roots, and if a Scatchard form is

fit to data, the resulting estimates of $K_1, \ldots, K_N$ can be useful, even though we are using an inexact model. One usually would check the obtained K_i-values in the Adair model in any event. Usually useful results can be had only when $N < 6$, unless highly-accurate data over a wide range of F_e-values is available.

The general analysis of multiple-site binding is quite complex. An excellent survey, including many practical tips, is: "The analysis of equlibrium binding data by the fitting of models" by John E. Fletcher, DCRT, N.I.H., May 1982.