

Ultracentrifuge Data Analysis

Gary D. Knott, Ph.D.
Civilized Software, Inc.
12109 Heritage Park Circle
Silver Spring, MD 20906
Tel. (301) 762-1785
email: csi@civilized.com
URL: <http://www.civilized.com>

One common application of curve-fitting occurs in the analysis of ultracentrifuge data. One of the most common uses of an ultracentrifuge is for molecular weight determination. Suppose we load the cell (a wedge-like container with circular-profile surfaces at each end) of an ultracentrifuge with a solution of a substance with an unknown molecular weight, and then spin the cell until the molecules in solution distribute themselves in the cell according to the forces that act upon them. This equilibrium arrangement will have the solute distributed in the cell in an exponential fashion with relatively more of the solute at the bottom (which is the outer end) of the cell. Therefore the concentration of the solute varies at different positions along the cell (*i.e.*, at different radial positions or distances from the center of the ultracentrifuge.)

Let us denote a given radial position by r , measured in centimeters from the center of rotation, where the cell ranges from the top position at r_m to the bottom position at the outer end of the cell at r_b . Then the concentration of the solute, c , measured in grams per liter, at the radial position r is a function of r given as:

$$c(r) = c_h e^{AN(r^2 - r_h^2)},$$

where

c_h is the concentration in grams per liter of the solute at a given reference radial position, r_h . The value c_h can be computed from N and c_0 , the total amount of solute in the cell measured in grams per liter, however, in practice, c_h can be determined by curve-fitting.

r_h is the given radial reference position (it is not necessary that r_h be r_b or r_m , indeed r_b is generally a poor choice).

N is the apparent molecular weight of the solute.

A is a certain constant, defined as: $(1 - \bar{v}\rho)w^2/(2RT)$, where \bar{v} is the partial specific volume of the solute ($1/\bar{v}$ is the density of the anhydrous solute, so \bar{v} is measured in cm^3/gram), ρ is the solution density in grams/cm^3 , w is the angular velocity of the rotor in radians per second, R is the gas constant ($\approx 8.314 \times 10^7$ ergs/degree/mole), and T is the absolute temperature.

Due to the non-ideal diffusion behavior of most materials, the observable or apparent molecular weight N , at a given radial position r , is approximately the following non-linear function of the actual molecular weight, M : $N \approx M/(1 + BMc(r))$, where the parameter B is the so-called virial coefficient, with $B \geq 0$.

Thus N is a function of r , and we have:

$$\begin{aligned} c(r) &= c_h \exp(AM(r^2 - r_h^2)/(1 + BMc(r))), \quad \text{or} \\ c(r) &= \text{root}(x, 0, c_{\max}, (c_h \exp(AM(r^2 - r_h^2)/(1 + MBx)) - x)). \end{aligned}$$

The notation $\text{root}(x, a, b, f(x))$ denotes a value, x_0 , in the interval $[a, b]$ which is a root of the expression $f(x)$, *i.e.*, x_0 is a value such that $f(x_0) = 0$. `MLAB` provides the root operator as a built-in function `ROOT`. The value c_{\max} is any value greater than $c(r_b)$, which is an upper bound for $c(r)$.

Now, if we measure the concentration distribution with a UV scanner or with a Rayleigh interferometer attached to the ultracentrifuge and express the results as a table of points (r_i, c_i) , which gives the observed concentration c_i at the radial position r_i in the cell, then the root-form equation for $c(r)$ given above can be used as a model to fit these points. In this way, M can be determined by curve-fitting. The virial coefficient B is also generally treated as a fitting parameter. Often c_h is not exactly known, and can also be a fitting parameter. Concentration can be measured in any of a number of different units. Fundamentally, we have fringe numbers or optical densities, and such observations are readily convertible by scaling to grams-per-liter concentration units.

Here is an example showing how the fit is done in `MLAB`. We first define the model, and then read-in some data, set initial-guess values of the parameters to be fitted, and fit the model to the given data. Finally, we draw pictures of the data and the fitted model.

```
fct c(r)=root(x, 0.5,20,(ch*exp(A*M*(r^2-rh^2)/(1+M*B*x))-x))
```

```

data = read(centri,51,2) /* read in the data from file centri.dat */

q = 0.26 /* this term is 1-zbar*rho */
omega = 1047.2 /* corresponding to 10000 rpm machine rotation */
R = 8.314*10^7 /* gas concentration */
t = 298.15 /* absolute temprature */
A = (q*omega^2)/(2*R*t); /* A = 5.75119093E-6 */
rh = 6.7 /* reference radius in cm */

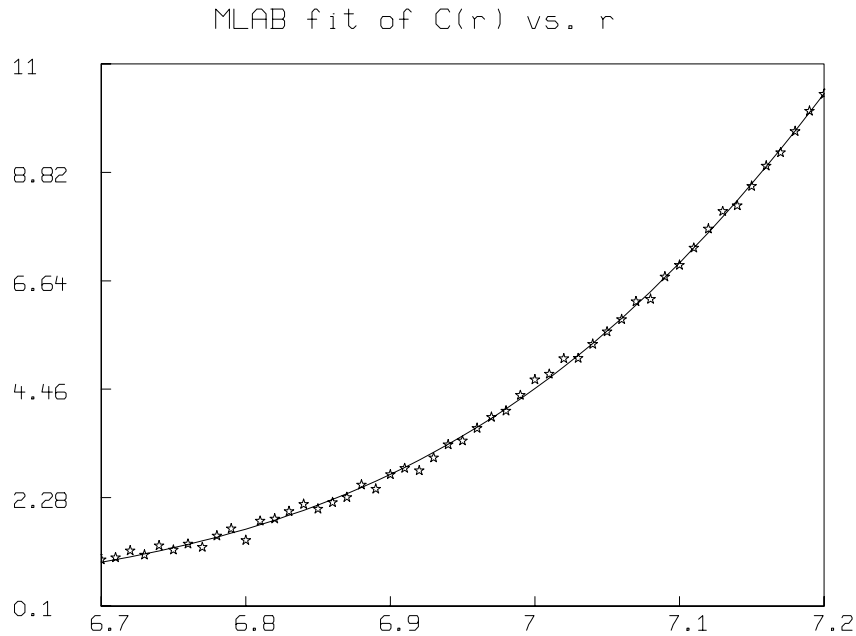
/* set initial guesses for B, ch, M */
B = 10^(-7)
ch = 1.25;
M = 60000

fit(M,B,ch), c to data

final parameter values
      value          error          dependency    parameter
68743.0466061385    1414.898785691    0.9928704243    M
 2.283270139e-07    1.901124729e-08    0.9729368851    B
 0.9799669867      0.0201841151      0.9789628624    ch
4 iterations
CONVERGED
best weighted sum of squares = 3.946440e-01
weighted root mean square error = 9.067387e-02
weighted deviation fraction = 9.884748e-03
R squared = 9.989925e-01

draw data, lt none, pt star, ptsize 0.01
draw points(c,(data col 1))
top title "MLAB fit of C(r) vs. r"
view

```



Note that our model $c(r)$ is an *implicit* function. One of the features of *MLAB* is the ability to handle such models. The *MLAB* curve-fitting process involves computing the derivatives of c with respect to B, M and c_h . Here are the derivative functions generated by *MLAB*. When differentiating an implicit function, *MLAB* makes use of the *eval* operator defined so that $eval(x, E, F)$ denotes the value of the expression F with the value E substituted for x within F .

```
* type c'B
FUNCTION c'B(r) = EVAL(x,ROOT(x,0.5,20,ch*EXP((A*M*(r^2-rh^2))/(1+M*B*x))-x),
-(((M*x*A*M*(r^2-rh^2))/((1+M*B*x)*(1+M*B*x)))*EXP((A*M*(r^2-rh^2))
/(1+M*B*x))*ch)/(((M*B*A*M*(r^2-rh^2))/((1+M*B*x)*(1+M*B*x)))
*EXP((A*M*(r^2-rh^2))/(1+M*B*x))*ch+1))
*
* type c'M
FUNCTION c'M(r) = EVAL(x,ROOT(x,0.5,20,ch*EXP((A*M*(r^2-rh^2))/(1+M*B*x))-x),
(((A*(r^2-rh^2)*(1+M*B*x)-B*x*A*M*(r^2-rh^2))/((1+M*B*x)*(1+M*B*x)))
*EXP((A*M*(r^2-rh^2))/(1+M*B*x))*ch)/(((M*B*A*M*(r^2-rh^2))
/((1+M*B*x)*(1+M*B*x)))*EXP((A*M*(r^2-rh^2))/(1+M*B*x))*ch+1))
```

```

*
* type c'ch
FUNCTION c'ch(r) = EVAL(x,ROOT(x,0.5,20,ch*EXP((A*M*(r^2-rh^2))/(1+M*B*x))-x),
EXP((A*M*(r^2-rh^2))/(1+M*B*x))/(((M*B*A*M*(r^2-rh^2))/((1+M*B*x)*(1+M*B*x))))
*EXP((A*M*(r^2-rh^2))/(1+M*B*x))*ch+1))

```

If we have several solutes with unknown molecular weights, then the equation for the combined concentration distribution is merely the sum of the separate concentration distributions. Thus, curve-fitting can be used to determine several molecular weights simultaneously.

In general, we have: $c(r) = c_1(r) + c_2(r) + \dots + c_n(r)$, where the concentration of the i th solute is $c_i(r) = c_{hi} \exp(A_i N_i (r^2 - r_h^2))$, N_i is the apparent molecular weight of the i th solute, and c_{hi} is the grams per liter concentration of the i th solute at the reference position r_h , and A_i is the usual constant $(1 - \bar{v}\rho)w^2/(2RT)$ described above, particularized for the i th solute.

As before, N_i is a function of r and the true molecular weight, M_i , namely: $N_i \approx M_i/(1 + B_i M_i c(r))$. In fact this is an approximation to the relationship:

$$N_i = M_i / (1 + \sum_{k=1}^n \sum_{j=1}^{\infty} B_{ikj} (M_i c_k(r))^j)$$

Usually assuming each value B_{ik1} is the same, namely B_i/n , for $1 \leq k \leq n$ and $B_{ikj} = 0$ for $1 < j$ is adequate, and this results in the simple form used above.

Suppose, for example, we have two solutes with molecular weights M_1 and M_2 . Then

$$c(r) = \text{root}(x, 0, c_{\max}, c_{h1} \exp(A_1 M_1 (r^2 - r_h^2)/(1 + B_1 M_1 x)) + c_{h2} \exp(A_2 M_2 (r^2 - r_h^2)/(1 + B_2 M_2 x)) - x)$$

It is often of interest to compute the amounts of each component present in the ultracentrifuge cell directly from the gram-per-liter concentration profile function. This can be done by using the equation:

$$q_i = \int_{r_m}^{r_b} c_i(r) v dr,$$

where q_i is the total mass in grams of the given solute i , v is a constant based on the geometry of the cell, r_b is the radial position of the bottom of the cell, and r_m is the radial position of the top, or meniscus, of the solution in the cell. For a cell 1.2 cm deep with a chord length of .33 cm at the radial position $r_b = 7.2$ cm while spinning, we have $v = 1.2 \cdot 2 \cdot \arcsin(.33 \cdot 7.2/2)$. It is important to use the most accurate possible geometry constant v and bottom

radial position r_b ; when spinning, r_b can be approximately 1% greater than the value measured at rest. Finally note that this computation can be confounded when the solute sticks to the the interior walls of the cell or builds up excessively at the bottom of the cell.

MLAB can be used to compute the equilibrium constant K for the simple binding reaction $X + Y \rightleftharpoons Z$ occurring within the ultracentrifuge cell in many cases.

Let $c_1(r)$ denote the grams-per-liter concentration of the first solute substance X at radial position r , let $c_2(r)$ denote the grams-per-liter concentration of the second solute substance Y at radial position r , and let $c_3(r)$ denote the grams-per-liter concentration of the composite third solute substance Z at radial position r . Then the combined concentration distribution function is $c(r) = c_1(r) + c_2(r) + c_3(r)$, where the grams-per-liter concentration distribution of the i th solute is $c_i(r) = c_{hi} \exp(A_i N_i (r^2 - r_h^2))$ for $i = 1, 2, 3$.

Also, at any fixed radial position r , the stoichiometric relationship $K = (\frac{c_3(r)}{M_3}) / ((\frac{c_1(r)}{M_1})(\frac{c_2(r)}{M_2}))$ holds, where M_i denotes the molecular weight of solute i for $i = 1, 2, 3$. Note that $M_3 = M_1 + M_2$. Thus,

$$c_3(r) = \frac{M_1 + M_2}{M_1 M_2} K c_1(r) c_2(r).$$

Let $g_i(r, x) := c_{hi} \exp(A_i M_i (r^2 - r_h^2) / (1 + B_i M_i x))$ for $i = 1, 2, 3$. Note $g_i(r, c(r)) = c_i(r)$. Now define $f(c_1, c_2) := c_1 + c_2 + K c_1 c_2 (M_1 + M_2) / (M_1 M_2)$. Then, $c(r) = \text{root}(x, 0, c_{\max}, f(g_1(r, x), g_2(r, x)) - x)$ where c_{\max} satisfies $c_{\max} > \max_r c(r)$. We thus have an implicit model for the c vs. r data corresponding to the simple binding reaction $X + Y \rightleftharpoons Z$ occurring in the ultracentrifuge cell. The potentially-unknown parameters of this model are c_{1h} , c_{2h} , M_1 , M_2 , B_1 , B_2 , A_1 , A_2 and K . MLAB curve-fitting may now in principle be employed to estimate these parameters. In the case of a self-associative dimerization reaction, this model simplifies with $c_{1h} = c_{2h}$, $M_1 = M_2$, $A_1 = A_2$ and $B_1 = B_2$.

Generally we cannot successfully estimate the equilibrium constant K when K is small (say $< .5 \cdot 10^4$), because our model tends to become increasingly ill-conditioned as K decreases. The exact limit depends on the molecular weights, the initially-loaded amounts, and the virial coefficients of the reactants.

There are several more “rules of thumb” that we should remember when estimating the equilibrium constant K .

(1) The molecular weights M_1 and M_2 should be predetermined and should not be fitting parameters. Any of a number of methods can be used

to determine M_1 and M_2 ; if sequence composition information is known, of course, it should be used to compute the corresponding molecular weight exactly. Similarly, the virial coefficients B_1 and B_2 should likewise be pre-determined, if possible. (Note that once we have obtained M_1 , M_2 , B_1 , B_2 , c_{1h} , and c_{2h} , we can subtract $c_1(r) + c_2(r)$ from our data, and then fit g_3 to the result in order to estimate B_3 and c_{3h} if this is desired.)

(2) The value c_{\max} should be as small as possible so as to minimize the occurrence of overflows in the derivatives of c . This implies that relatively dilute solutions of the reactants should be used. On the other hand, we need to produce an adequate amount of Z material to be represented in the $c(r)$ -data. The speed of the ultracentrifuge should be chosen to obtain a distribution of mass which is neither too “flat” nor too “sharp”, *i.e.*, material should not “pile-up” at the bottom of the cell. It may sometimes be convenient to discard the data for radial positions near the bottom of the cell in order to be able to reduce the value of c_{\max} . It is generally wise to discard the data near the bottom of cell in any event because the error in measuring mass concentration is large near the boundary of the cell.

(3) It is generally useful to replace K by $\exp(L)$, where $L = \log(K)$, and then fit to estimate L . This does not generally provide a more exact estimate of K , but the normal theory standard errors computed for L are more likely to be meaningful than are the same quantities for K when estimated directly.

Here is an example of using MLAB to estimate the equilibrium constant for a simple binding reaction. Note we have also used the conservation-of-mass relations for substance 1 (X) and for substance 2 (Y). Choosing parameters to fit the computed masses of X and Y to the known initially-loaded masses is a valuable device to improve our parameter estimates. This can only be done, however, when accurate values of v and r_b are available. (Another device that can help in obtaining more accurate estimates is to simultaneously fit gram-per-liter concentration-profile data obtained at equilibrium for several different rotor speeds.)

```
* rm = 6.7 /* minimum radius in cm */
* rb = 7.2 /* maximum radius in cm */
* rh = rm /* reference radius in cm */
* vc = 2*asin(.33/(2*rb)) /*angle of the cell wedge */
* vc=vc*1.2 /*volume constant, total solution vol.= (vc/2)*(rb^2-rm^2) cm^3 */

* r1 = rm:rb:.01 /* radial positions for curve-plotting */

* qv = 0.26 /* this term is 1 - zbar*rho */
```

```

* omega = 1047.2 /* corresponding to 10000 rpm machine rotation */
* R = 8.314*10^7 /* gas concentration */
* t = 298.15 /* absolute temperature */
* A1 = (qv*omega^2)/(2*R*t); A2 = A1 /* Both A1 and A2 = 5.75119093E-6 */

* /* predetermined values for known parameters */
* B1 = 2*10^(-8); B2=10^(-9);
* M1 = 67000; M2=6800

* constraints q={9<lk, lk<40, ch1>0, ch2>0, ch1+ch2<1.4}

* fct g1(r,x)=ch1*exp(A1*M1*(r^2-rh^2)/(1+M1*B1*x))
* fct g2(r,x)=ch2*exp(A2*M2*(r^2-rh^2)/(1+M2*B2*x))
* fct f(c1,c2)=c1+c2+exp(lk)*c1*c2*(M1+M2)/(M1*M2)
* fct c(r) = root(x,(ch1+ch2)/3,100000, f(g1(r,x),g2(r,x))-x)

* fct c1(r)=g1(r,c(r))
* fct c2(r)=g2(r,c(r))
* fct c3(r)=c3h(r,c(r))
* fct c3h(r,x)=x-g1(r,x)-g2(r,x)

* /*tm1, tm2 = total mass in grams of substances 1 and 2 */
* fct tm1()=integral(r,rm,rb,tm1z(r,c(r))*vc*r)
* fct tm1f(c1,c2,c)=c1+(M1/(M1+M2))*(c-c1-c2)
* fct tm1z(r,x) = tm1f(g1(r,x),g2(r,x),x)

* fct tm2()=integral(r,rm,rb,tm2z(r,c(r))*vc*r)
* fct tm2f(c1,c2,c)=c2+(M2/(M1+M2))*(c-c1-c2)
* fct tm2z(r,x) = tm2f(g1(r,x),g2(r,x),x)

* /*specify the total mass in grams of substances 1 and 2 */
* mass1[1]= .113689839;
* mass2[1]= .00613855051

* /* read-in the r vs. c(r) data to be modeled. */
* data = read(data.dat,1000,2)
* n=nrows(data);
*

```



```

* /* define the weights for fitting and normalize them to sum to 1 */
* zw=ewt(data); zw=zw/rowsum(zw)

* /*set initial guesses for k, ch1,and ch2 */
* k=10000; lk=log(k)
* ch1=.15;
* ch2=.035

* fit(lk,ch1,ch2), c to data with weight zw, tm1 to mass1, tm2 to mass2,\
: constraints q

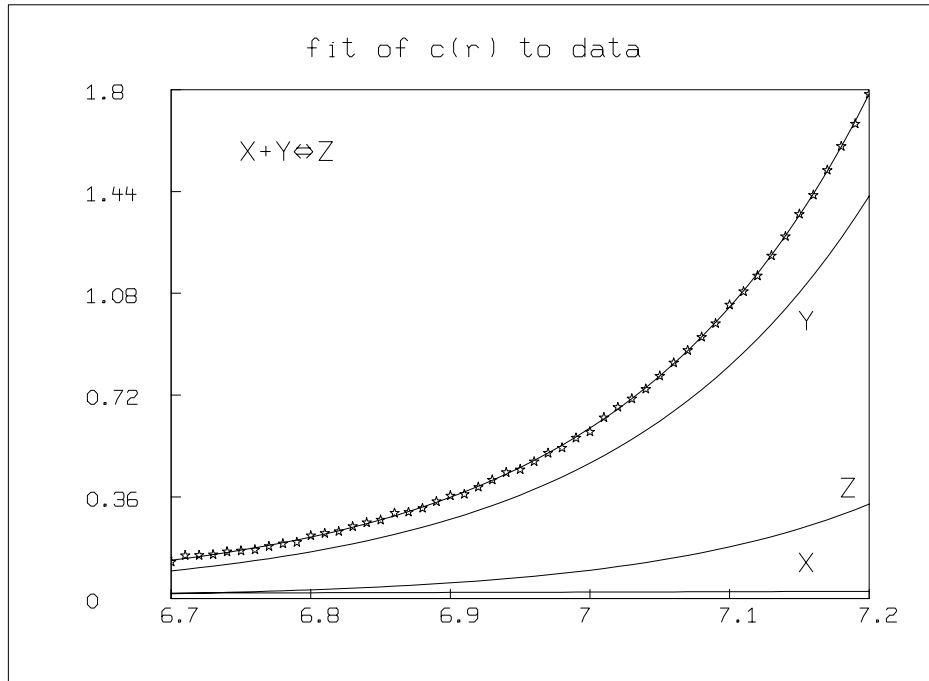
final parameter values
      value          error          dependency    parameter
10.92859862      0.1533653124      0.9980490348      LK
  0.09847696449      0.003231742687      0.9986427864      CH1
  0.01982622037      0.001226384415      0.9914474897      CH2
4 iterations
CONVERGED
best weighted sum of squares = 2.386287e-05
weighted root mean square error = 6.908382e-04
weighted deviation fraction = 5.217909e-03
R squared = 9.998792e-01
no active constraints

* exp(lk) /* k=exp(lk) */
  = 55748.1007
* tm1() /* estimated total mass in grams of X */
  = .113740813
* tm2() /* estimated total mass in grams of Y */
  = 6.25476406E-3

* draw data, lt none, pt star, ptsize 0.01
* draw points(c,r1)
* top title "fit of c(r) to data"
* draw points(c1,r1), color red
* draw points(c2,r1), color green
* draw points(c3,r1), color yellow
* title "Y" at (7.15,.95) world
* title "Z" at (7.18,.36) world
* title "X" at (7.15,.095) world

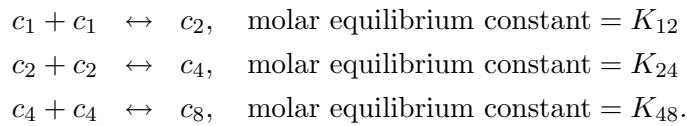
```

```
* title "X+Y'2T='RZ" at (6.75,1.55) world
* view
```



When we have self-associating molecules in an ultracentrifuge that self-associate in a more complex manner than simple dimer formation, we can still determine the monomer molecular weight, and the equilibrium constants of the known association reactions by curve-fitting in MLAB.

For example, suppose we have the self-associating reactions:



If we let these reactions take place in an ultracentrifuge cell, and measure the weight-average molecular weight, M_w , of the resulting mixture in grams for different values of c , the total concentration in grams per liter of all the species c_1 , c_2 , c_4 , and c_8 , then we can simultaneously determine the monomer molecular weight of the species c_1 : M_1 , and the equilibrium constants K_{12} , K_{24} , K_{48} , by curve-fitting.

We have $M_w = (c_1M_1 + c_2M_2 + c_4M_4 + c_8M_8)/c$, where M_i is the molecular weight of c_i , and the symbol c_i , as a function, denotes the gram-per-liter equilibrium concentration of the species c_i . c is the total gram-per-liter concentration, so $c = c_1 + c_2 + c_4 + c_8$.

Note that $M_2 = 2M_1$, $M_4 = 4M_1$, and $M_8 = 8M_1$. Also,

$$\begin{aligned} K_{12} &= (c_2/M_2)/(c_1/M_1)^2, \\ K_{24} &= (c_4/M_4)/(c_2/M_2)^2, \quad \text{and} \\ K_{48} &= (c_8/M_8)/(c_4/M_4)^2. \end{aligned}$$

Thus, the corresponding equilibrium constants for gram-per-liter concentrations can be defined as:

$$\begin{aligned} E_{12} &= 2K_{12}/M_1 \\ E_{24} &= K_{24}/M_1, \quad \text{and} \\ E_{48} &= K_{48}/(2M_1), \end{aligned}$$

so that,

$$\begin{aligned} c_2 &= E_{12}c_1^2, \\ c_4 &= E_{24}c_2^2 = E_{24}E_{12}^2c_1^4, \quad \text{and} \\ c_8 &= E_{48}c_4^2 = E_{48}E_{24}^2E_{12}^4c_1^8. \end{aligned}$$

Define $E_{14} = E_{24}E_{12}^2$, and $E_{18} = E_{48}E_{24}^2E_{12}^4$.

Now, $c = c_1 + c_2 + c_4 + c_8$, so $c = c_1 + E_{12}c_1^2 + E_{14}c_1^4 + E_{18}c_1^8$, and thus, $c_1 = \text{root}(x, 0, c, (x + E_{12}x^2 + E_{14}x^4 + E_{18}x^8 - c))$.

Also, $M_w = M_1(c_1 + 2E_{12}c_1^2 + 4E_{14}c_1^4 + 8E_{18}c_1^8)/c$.

Due to the non-ideal gas behavior of most materials, the observed weight-average molecular weight N is, in fact, more nearly the following non-linear function of the actual weight-average molecular weight: $N = M_w/(1 + BM_wc)$, where the parameter B is the so-called virial coefficient.

Thus, given data: c vs. N , we can fit for the parameters E_{12} , E_{24} , E_{48} , B , and M_1 in the following model, as defined in MLAB.

```
*FUNCTION MA(C) = NID(C,M1*MWX(C,E12,E14,E18),B)
*FUNCTION NID(C,MW,B) = MW/(1+B*C*MW)
*FUNCTION MWX(C,A1,A2,A3) = P(C1(C,A1,A2,A3),A1,A2,A3)/C
*FUNCTION P(C1,A1,A2,A3) = C1+2*A1*C1^2+4*A2*C1^4+8*A3*C1^8
*FUNCTION C1(C,A1,A2,A3) = ROOT(X,0,C,X+A1*X^2+A2*X^4+A3*X^8-C)
```

In order to obtain the data c vs. N , preliminary analysis of the actual observations is required. What is, in fact, observed is r vs. c , where r is a radial position in the cell and where the total gram-per-liter concentration c is observed as a function of r . We have the fundamental relation:

$$c(r) = c_h \exp(AN(r^2 - r_h^2)),$$

where A , c_h , and r_h are defined above.

Now one can take a few neighboring observations, say five points $(r_1, c(r_1))$, $(r_2, c(r_2))$, \dots , $(r_5, c(r_5))$, and fit the above equation to these points with $c_h = c(r_3)$, and $r_h = r_3$, and thus determine the value of N which corresponds to $c(r_3)$. Sliding this operation over all the observations transforms the data so that the model above for c vs. N data is now applicable.

Alternatively, one may use the r vs. c observations directly by using an implicit model based on the following equations.

$$\begin{aligned} c_1(r) &= c_{1h} \exp(AN_1(r^2 - r_h^2)), \\ N_1 &= M_1/(1 + BM_1c(r)), \quad \text{and} \\ c(r) &= c_1(r) + E_{12}c_1(r)^2 + E_{14}c_1(r)^4 + E_{18}c_1(r)^8. \end{aligned}$$

In MLAB we have:

```
*FUNCTION C(R) = P(C1(R,B,M1,C1H),E12,E14,E18)
*FUNCTION P(X,A1,A2,A3) = X+A1*X^2+A2*X^4+A3*X^8
*FCT C1(R,B,M1,CH1) = ROOT(Y,0,MAXC1,Y-CH1*EXP(A*(R*R-RH*RH)*M1/(1+B*M1*Y)))
```

Although using this model to fit r vs. c data to obtain the parameters B , M_1 , E_{12} , E_{14} , E_{18} , and possibly c_{h1} , is probably more accurate than the indirect method above, it has the disadvantage that when, as is common, different experiments with different c_{h1} values (and perhaps different A values) are used to cover a wider range of c values, then a curve-fit involving several models being fit simultaneously to adjust shared parameters must be employed, and this is impractical when too many experiments are involved.

In either case, K_{12} , K_{24} , and K_{48} can then be determined from E_{12} , E_{14} , E_{18} , and M_1 . Similar models can be obtained for many other specific forms of self-association. The special case where all possible species can arise due to sequential binding, with each binding requiring the same change of free energy as any other is discussed below.

Suppose we have a non-specific isodesmic “stacking” reaction involving a monomer c_1 .

$$c_1 + c_1 \leftrightarrow c_2, \quad \text{molar equilibrium constant} = K \quad (1)$$

$$c_1 + c_2 \leftrightarrow c_3, \quad \text{molar equilibrium constant} = K \quad (2)$$

$$\vdots \quad \quad \quad \vdots \quad (3)$$

An isodesmic association is one where the free energy for adding one more monomer, c_1 , to an oligomer, c_n , is the same for all n .

As before, let c_i also denote the number of grams-per-liter of the species c_i in the cell at equilibrium, thus c_i is the grams-per-liter concentration of the identically-named species. Also let $c = c_1 + c_2 + \dots$, and $M_w = (c_1M_1 + c_2M_2 + \dots)/c$, where M_i is the molecular weight of the species c_i ; thus c is the total grams-per-liter concentration of all species present, and M_w is the weight-average molecular weight of all species.

Now, $K = (c_{i+1}/M_{i+1})/((c_1/M_1)(c_i/M_i))$ so $c_{i+1}/M_{i+1} = K(c_1/M_1)(c_i/M_i)$, and hence by successive substitution $c_{i+1}/M_{i+1} = K_i(c_1/M_1)^{i+1}$.

Thus $c_i = i(K/M_1)^{i-1}c_1^i$, and hence $c = c_1(1 + 2Kc_1/M_1 + 3(Kc_1/M_1)^2 + \dots)$. But $1 + 2x + 3x^2 + \dots = 1/(1-x)^2$, so $c = c_1/(1 - Kc_1/M_1)^2$, where K is the molar equilibrium constant, subject to the constraint that $Kc_1/M_1 < 1$.

Also, $M_w = (M_1c_1 + 2M_1(2(Kc_1/M_1)c_1) + 3M_1(3(Kc_1/M_1)2c_1) + \dots)/c$, so $M_w = c_1M_1(1 + 2^2(Kc_1/M_1) + 3^2(Kc_1/M_1)2 + \dots)/c$; and $1^2 + 2^2x + 3^2x^2 + \dots = (1+x)/(1-x)^3$, so $M_w = c_1M_1(1 + Kc_1/M_1)/(1 - Kc_1/M_1)^3/c$. Substituting $c_1/(1 - Kc_1/M_1)^2$ for c , we have $M_w = M_1(M_1 + Kc_1)/(M_1 - Kc_1)$.

Now, c_1 is the function of c given by $c_1(c) = \text{root}(x, 0, c, x/(1 - Kx/M_1)^2 - c)$, and in terms of this function $M_w = M_1(M_1 + Kc_1(c))/(M_1 - Kc_1(c))$. As before the apparent weight-average molecular weight N_w is approximately given by $M_w/(1 + BM_wc)$, so we may use the model

$$N_w(c) = M_1/((M_1 - Kc_1(c))/(M_1 + Kc_1(c)) + BM_1c)$$

to fit c vs. N_w data.

Also, since $c(r) = c_h \exp(AN_w(r_2 - r_h^2))$, we can use the model

$$c(r) = \text{root}(x, 0, c_{max}, (x - c_h \exp(AN_w(x)(r_2 - r_h^2)))$$

to fit r vs. c data.

The equations above are examples of the type of models involved in the analysis of data obtained from the use of an ultracentrifuge. There are many other useful experiments, with other equations (or differential equations) describing the phenomena involved. A good reference on the use of

the ultracentrifuge and the associated mathematical modeling is the chapter by K.E. Van Holde in the book *The Proteins*, 3rd edition, Vol. 1, 1975. Another source is *Foundations of Ultracentrifugical Analysis*, by Hiroshi Fujita, Wiley, 1975.