Biophysical Chemistry, 32 (1988) 131–146
Elsevier

BPC 01319

The determination of equilibrium constants for heterogeneous macromolecular interactions

Systems forming 2:1 complexes

Mary Lou Crowl-Powers and James K. Zimmerman

Department of Biological Sciences, 132 Long Hall, Clemson University, Clemson, SC 29634-1903, U.S.A.

Received 1 July 1988
Revised manuscript received 12 August 1988
Accepted 16 August 1988

Association constant; Macromolecular interaction; Light scattering

In developing a method for analyzing the heterogeneous association $nA + mB \rightleftharpoons A_nB_m$, we have specifically investigated the case of $n = 2$, $m = 1$ for both the specific case of no appreciable intermediates and the more general case allowing intermediates. Computer-simulated three-dimensional surfaces of the 2:1 model generated from total concentrations of species A and B and the resulting weight-average molecular weights were analyzed with a Gauss-Newton nonlinear least-squares minimization routine. The surfaces generated included normalized random error of varying standard deviations imposed upon both the concentrations and weight-average molecular weights. For comparison purposes, these surfaces were analyzed not only by using the correct 2:1 model, but also by an incorrect (1:1) model and by the other (incorrect) 2:1 model. Except for those situations where the 'experimental' noise was consistently higher than the concentration of one of the species, correct K values were obtained and the correct model was easily distinguished from the incorrect model. The computer routine similarly distinguished between data correctly described as 1:1 and the same data incorrectly analyzed as either 2:1 model. For those cases in which a microscopic K_i value predicts an association such that all species involved for that particular K_i are in appreciable amounts, the K_i value is returned correctly. Correct overall equilibrium constants are also converged upon as long as adequate amounts of A_2B, B and A are present.

1. Introduction

Associations between unlike macromolecules are of great importance in biological systems. Methods for analyzing these associations, and especially heterogeneous associations, are often specifically designed for the experimental procedure being used. The most common methods are based on sedimentation equilibrium experiments [1–9], although other procedures such as osmotic pressure [10,11], analytical gel chromatography [12–14] and light scattering [15–17] have also been used. Each procedure has its own advantages and disadvantages depending on the macromolecular system being studied [11]. Usually, this entails developing specialized mathematical analyses which are not easily interchanged between experimental systems. A mathematical method would be more useful if the analysis could be used for several types of experimental procedures.

We have previously reported the development of a general analytical method for the interaction of two macromolecules forming a 1:1 complex [18], requiring only measurements of the total concentrations of each of the two macromolecular species and the determination of the weight-aver-

Correspondence address: J.K. Zimmerman, Department of Biological Sciences, 132 Long Hall, Clemson University, Clemson, SC 29634-1903, U.S.A.

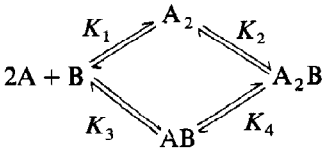
age molecular weight of the complex. This method has been expanded to the formation of a 2:1 complex either with or without appreciable intermediate species.

The case with no intermediates can be described by:

$2A + B \rightleftharpoons A_2B$ (model I)

This model is of course mechanistically unlikely because the probability of a three-body collision would be very small.

A more realistic model for producing A_2B would be to add B to an A_2 dimer or to add a second molecule of A to an AB complex as described below:



(model II)

Of course, both arms of model II could also be operating simultaneously. If the equilibrium concentrations of A_2 and AB are both small, models II and I will be essentially identical.

Our computer simulations were based on analyses of more than 80 data points per surface. The minimum number of experimental points required for the equations to be pertinent are of course equal to the number of dependent variables to be determined (as few as one for model I or as many as three for model II), however, the precision will increase as the square root of the number of observations. One system that is capable of generating a large number of A_0, B_0, M_W data points is sedimentation equilibrium when the data are obtained through a scanner system and when the point-by-point total concentrations of each species can be determined because of spectral differences. Another system for generating a large number of data points is currently under development by us and uses a light scattering system with a flow cell and with proportional pumps delivering varying mixtures of species A and B. We repeat, however, that the analytical model presented here is derived independently of the data collection mode.

2. Method

2.1. Fitting method

The general procedure used to analyze a three-dimensional surface has been described previously [18] for the case of a 1:1 complex. This procedure requires a model-specific subroutine which will describe the fitting function solely in terms of the desired equilibrium constant(s) and the relationships between the input parameters, A_0, B_0, and M_W from each data point. This requires calculation of the values for free A and free B for each data point at a given K value. In our previous work [18], this information could be obtained exactly through a quadratic fit, however, for both models of the 2:1 system it was necessary to calculate the free A concentration (and thus the free B concentration) for each A_0, B_0 combination at each estimate of K (or set of K values) using a Newton-Raphson root-finding routine. After the concentration of each species in each concentration pair was established for a given estimate of K (or set of K values), a Gauss-Newton fit was employed to generate the next estimate of K (or set of K values). The cycle was repeated until a value of K (or set of K values) was reached that met the conditions of convergence. The appropriate mathematics are given in appendix A for model I and in appendix B for model II.

2.2. Simulated data

The limits of the system were tested using computer-simulated data. The molecular weights of species A and B used in generating the data were 90000 and 180000, respectively. A range of concentration pairs of both species, from 0.1- up to 0.9-times that of the indicated maximum concentrations, was generated. The method of data generation and of applying random error was described earlier [18]. For model I maximum concentrations of 0.1, 1.0 and 10 mg/ml and association constants of 0.1, 10 and 1000 $(ml/mg)^2$ were used; the corresponding molar association constants thus ranged from 4×10^7 to 4×10^{15} M^{-2}. For model II, $K_1K_2 = K_3K_4 = K'$. An overall association constant, K', of 10 $(ml/mg)^2$ (4×10^{11}

M^{-2}) and a maximum concentration of 1.0 mg/ml were used for all cases. 25 data sets were made with five ratios each for $K_1 : K_2$ and $K_3 : K_1$ ranging from 1 : 100 to 100 : 1. The 'experimental' error imposed on the weight-average molecular weight, M_W, and total concentrations A_0 and B_0 were between 0 and 3%. The algorithm for all 1 : 1 data sets was described previously [18].

3. Results

In order for the fitting procedure to be useful, it must not only obtain the correct answers for K, but also be able to distinguish between correct and incorrect models. Therefore, models generated by each of the 2 : 1 models and data generated for a 1 : 1 model [18] were tested.

3.1. Comparison of 1 : 1 model and 2 : 1 model I

Table 1A shows results for data generated by a 1 : 1 model, then analyzed by both the 1 : 1 model and a 2 : 1 model (model I). Table 1B lists results for data generated by model I then analyzed by both model I and the 1 : 1 model. In each section the computed values for K and the variance of the fit, s^2, are listed. By comparing ratios of the variances for each fit, one can determine which model better fits the data [19–21]. The last column of each table lists F ratios. Each of the data sets in table 1 have 100 degrees of freedom for each of the two models (101 data points and one fitting constant). Under these conditions, in order to eliminate, with 95% confidence, the model with the larger variance, the F ratio must be greater than 1.39 (this cut-off value was calculated using the FPROB function of SPEAKEZ). For 99% confidence, the cut-off value is 1.6.

Table 1A shows fits for four 1 : 1 data sets chosen to cover the extremes of the 1 : 1 data analysis. Both cases A and D are considered outside the range of acceptable analysis. This designation is made if more than one-third of the data points contain a particular species at a level less than the 'experimental' noise. Even with these extreme values, three of the four cases were still able to distinguish clearly between models; in case A a 71% confidence level was still achieved.

Table 1B shows a very wide range of data generated by model I. Case A is such a weak association that in 100% of the data points the concentration of A_2B is less than the 3% error imposed on the system. In this case, as expected, one can neither calculate the correct K value nor distinguish between models. In the other eight cases the correct model is easily distinguished. It should be noted that the only other incorrect K values returned, cases H and I, both fall outside the defined range of analysis. In case I more than half of the data points contain free A or free B below the experimental noise level; in case H approximately one-third of the data points had free A or free B below the noise level. In both cases the program still returned a K within 17% of the correct value.

3.2. Comparison of 2 : 1 models with and without intermediates

3.2.1. With 0% random error

As discussed earlier, a 2 : 1 model can also include intermediate species (model II). In this model three independent parameters are required to fit the data. We have chosen K', K_1 and K_3. The remaining parameters can be calculated from these. Because of the number of parameters to be fitted, it was important to test the sensitivity of the analysis with data containing zero error. This allows us to distinguish between difficulties inherent in the fitting method and those arising from experimental noise. Table 2 is arranged in a similar manner to table 1, with table 2A analyzing data generated from model I by both models I and II, while table 2B shows the same analysis of data generated by model II. All of these data have 0% experimental error of M_W, A_0 and B_0. In all cases 91 data points were generated resulting in 90 degrees of freedom for model I and 88 degrees of freedom for model II.

In table 2A all of the K values were returned correctly when analyzed by model I (the model generating the data) but in five of the nine cases model II also returned the correct K' value and,

Table 1

Data fit by 1:1 model and 2:1 model (without intermediates)

Indicated random Gaussian error was imposed on molecular weights and concentrations. Concentrations listed are the maximum values for the system.

A Data generated as 1:1	Results if data analyzed as				F ratios $(s_B^2 : s_A^2)$
	1:1		2:1		
	K	s_A^2	K	s_B^2	
(A) $K = 0.1$ (ml/mg), $C = 1$ (mg/ml), STD $M_W = 3\%$ STD concentration = 3%	9.50E−02	1.44E+07	6.40E−02	1.61E+07	1.12E+00
(B) $K = 1.0$ (ml/mg), $C = 1$ (mg/ml), STD $M_W = 1\%$, STD concentration = 1%	9.98E−01	2.66E+06	6.49E−01	3.70E+07	1.39E+01
(C) $K = 1.0$ (ml/mg), $C = 1$ (mg/ml), STD $M_W = 3\%$, STD concentration = 3%	1.02E+00	1.70E+07	6.63E−01	5.76E+07	3.39E+00
(D) $K = 100$ (ml/mg), $C = 1$ (mg/ml), STD $M_W = 3\%$, STD concentration = 3%	7.46E+01	3.20E+07	7.69E+00	8.14E+08	2.54E+01

B Data generated as 2:1	Results if data analyzed as				F ratios $(s_A^2 : s_B^2)$
	1:1		2:1		
	K	s_A^2	K	s_B^2	
$K = 0.1$ (ml/mg)^2, STD M_W, concentration = 3%					
(A) $C = 0.1$ (mg/ml)	2.80E−02	2.31E+05	2.60E−26	2.31E+05	1.00E+00
(B) $C = 1.0$ (mg/ml)	9.05E−01	6.34E+08	1.17E−01	2.36E+07	2.69E+01
(C) $C = 10.0$ (mg/ml)	2.58E+00	1.44E+13	9.80E−02	4.11E+09	3.50E+03
$K = 10$ (ml/mg)^2, STD M_W, concentration = 3%					
(D) $C = 0.1$ (mg/ml)	1.02E+00	5.44E+05	1.02E+01	1.69E+05	3.22E+00
(E) $C = 1.0$ (mg/ml)	4.09E+00	1.67E+10	1.04E+01	4.98E+07	3.35E+02
(F) $C = 10.0$ (mg/ml)	2.95E−01	3.45E+12	9.32E+00	9.04E+09	3.82E+02
$K = 1000$ (ml/mg)^2, STD M_W, concentration = 3%					
(G) $C = 0.1$ (mg/ml)	1.35E−03	8.37E+07	1.05E+03	4.51E+05	1.86E+02
(H) $C = 1.0$ (mg/ml)	2.46E+01	1.39E+11	8.33E+02	8.50E+07	1.64E+03
(I) $C = 10.0$ (mg/ml)	5.14E−02	3.40E+12	8.17E+02	1.00E+10	3.40E+02

in those five cases, the F ratio test could not distinguish between the two models. (F ratio 1.44 for elimination at the 95% level, 1.7 at the 99% level). Closer inspection of the model II analyses, however, showed that, in those five cases, equilibrium constants K_1 and K_3 are returned as quite low values while the constants relating the intermediates to the A_2B complex are very high,

thus predicting no appreciable intermediates. Since essentially only A, B and A_2B are present, one would expect the return of a good overall equilibrium constant. For the remaining four cases, cases H and I allow the two models to be clearly distinguished, case A distinguishes between the models at the 86% confidence level and case F resulted in program failure when attempting to fit the data with the incorrect model.

On the other hand, when appreciable intermediates are present (table 2B), the analysis easily distinguishes between models and gives a correct overall K' value. Data set I returned a value for K' furthest from the true value due to the fact that a majority of data points had very little, if any, A_2B complex formed. In general, this analysis returns correct K_i values, especially for the K_i values of the more predominant pathway. Case A, where only a very small amount of AB is formed while larger amounts of A_2 are present, shows this quite nicely.

3.2.2. With 3% random error

In table 3 the same systems are shown as in table 2 except that a random 3% error was imposed on 'measured' values of M_W, A_0 and B_0. In table 3A the analysis using the generating model, model I, has already been discussed (table 1B, section 3.1). The K' values returned using the incorrect model (model II) in all but four cases (B, C, E, and G) returned values similar to those for the correct model, and two more, D and H, are incorrect by approx. 20%. With the addition of 3% error on the data, no distinction can be made between the two models in any of the cases. In two cases (F and I), the program was unable to determine any acceptable values when using model II for analysis.

In table 3B only four of the nine cases returned values for K' within 10% of the correct value when analyzed by the generating model (model II). For the five cases where the correct values were not returned, the majority of the data showed the concentration of at least one species to be less than the 3% experimental noise in the total concentration. Seven of the nine cases have greater than 95% confidence of elimination of the incorrect model (without intermediates), while in the remaining two cases (A and B), there is 92 and 80% confidence, respectively, in eliminating the incorrect model. In fact, all five of the cases returning incorrect K' values still were able to identify the correct model with 99% confidence. As one would expect with experimental error introduced into the data, the returned K_i values have larger errors associated with them.

As in the simpler 2:1 system (model I), the microscopic association constants must be such that both reactants and products are in appreciable amounts in order for the fitting routine to converge to the proper values. For those data sets which have such a tight association that the concentration of one species is lower than the 'experimental' error, as is the K_3 value in data sets F and I, the proper value for that microscopic K_i is not returned. The same is true for very weak associations where very little product is formed, such as for K_3 values in data sets A, B, D and G and for K_1 values in sets A–C.

3.3. Surface maps

Figs. 1 and 2 represent surfaces from data sets generated without and with intermediate species, models I and II, respectively. The total concentrations for species A and B are along the x- and y-axis and the weight-average molecular weight is along the z-axis. Shown in fig. 1A–D are the surfaces as the association constant, initial concentrations, and 'experimental' error are varied. The surface for an association of 10 $(ml/mg)^2$, maximum concentrations for both species of 1.0 mg/ml and 0% error is shown in fig. 1A. Without error imposed on any of the three variables (M_W, A_0 and B_0), the surface is smooth. The data used to generate this surface are easily fitted by the routine (data set E, table 2A). As the error is increased to 3%, the surface becomes irregular (fig. 1B), reflecting this increase. Decreasing the association constant to a very weak association ($K = 0.1$ $(ml/mg)^2$) and the concentrations to low maximum concentrations of both species ($C = 0.1$ mg/ml), as in fig. 1C, the surface becomes almost flat. As expected, the fit of these data is very poor (data set A, table 3A). If the association constant and maximum concentrations of species A and B

Table 2

Data fit by 2:1 model with and without intermediates; 0% error on molecular weights and concentrations

Concentrations listed are maximum values for the system.

A

Data generated as 2:1 (without intermediates)	Results if data analyzed as								F ratios $(s_B^2 : s_A^2)$
	2:1 (without intermediates)		2:1 (with intermediates)						
	K	s_A^2	K_1	K_2	K_3	K_4	K'	s_B^2	
$K = 0.1$ (ml/mg)^2, STD M_W and concentration = 0%									
(A) $C = 0.1$ (mg/ml)	9.60E−02	9.87E+00	4.83E−03	1.69E+01	7.17E−05	1.14E+03	8.16E−02	1.26E+01	1.28E+00
(B) $C = 1.0$ (mg/ml)	1.00E−0.1	8.95E+02	8.76E−06	1.14E+04	2.29E−06	4.36E+04	1.00E−01	9.34E+02	1.04E+00
(C) $C = 10.0$ (mg/ml)	9.90E−02	9.12E+04	1.34E−18	7.47E+06	1.44E−04	6.92E+02	9.99E−02	9.23E+04	1.01E+00
$K = 10$ (ml/mg)^2, STD M_W and concentration = 0%									
(D) $C = 0.1$ (mg/ml)	1.00E+01	8.95E+00	1.75E−06	5.73E+06	5.85E−08	1.71E+08	1.00E+01	9.18E+00	1.03E+00
(E) $C = 1.0$ (mg/ml)	9.99E+00	9.12E+02	2.25E−07	4.45E+07	1.32E−03	7.57E+03	9.99E+00	9.24E+02	1.01E+00
(F) $C = 10.0$ (mg/ml)	1.00E+01	9.41E+04	–	–	–	–	–	–	– [a]
$K = 1000$ (ml/mg)^2, STD M_W and concentration = 0%									
(G) $C = 0.1$ (mg/ml)	1.00E+03	9.12E+00	1.02E−07	9.82E+09	1.29E−02	7.73E+04	9.99E+02	9.23E+01	1.01E+00
(H) $C = 1.0$ (mg/ml)	1.00E+03	9.42E+02	7.20E+10	8.76E−27	7.20E+10	8.76E−27	6.31E−16	3.68E+09	3.91E+06
(I) $C = 10.0$ (mg/ml)	1.00E+03	9.38E+04	2.73E−05	1.27E+21	2.63E+13	1.32E+03	3.47E+16	1.95E+10	2.08E+05

B

Data generated as 2:1 (with intermediates) ($K = 10$ (ml/mg)^2, $C = 1.0$ (mg/ml))					Results if data analyzed as								F ratios $(s_A^2 : s_B^2)$
					2:1 (without intermediates)		2:1 (with intermediates)						
	K_1	K_2	K_3	K_4	K	s_A^2	K_1	K_2	K_3	K_4	K'	s_B^2	
(A)	0.316	31.62	0.003	3162.0	9.09E+00	5.59E+06	3.16E−01	3.16E+01	4.80E−03	2.08E+03	1.00E+01	9.05E+02	6.18E+03
(B)	0.316	31.62	0.316	31.62	8.99E+00	5.56E+06	3.16E−01	3.16E+01	3.16E−01	3.16E+01	9.99E+00	8.86E+02	6.28E+03
(C)	0.316	31.62	31.62	0.316	9.13E+00	3.67E+08	3.16E−01	3.16E+01	3.16E+01	3.16E−01	9.98E+00	7.67E+02	4.78E+05
(D)	3.162	3.162	0.032	316.2	5.50E+00	6.54E+07	3.17E+00	3.16E+00	3.08E−02	3.25E+02	1.00E+01	8.55E+02	7.65E+04
(E)	3.162	3.162	3.162	3.162	6.42E+00	9.60E+07	3.16E+00	3.17E+00	3.16E+00	3.17E+00	1.00E+01	8.75E+02	1.10E+05
(F)	3.162	3.162	316.2	0.032	1.45E+01	5.81E+08	3.16E+00	3.15E+00	3.16E+02	3.15E−02	9.96E+00	8.60E+02	6.76E+05
(G)	31.62	0.316	0.316	31.62	2.60E+00	2.24E+08	3.15E+01	3.17E−01	3.16E−01	3.16E+1	9.98E+00	9.26E+02	2.42E+05
(H)	31.62	0.316	31.62	0.316	9.47E+00	3.36E+08	3.16E+01	3.16E−01	3.16E+01	3.16E−01	9.98E+00	8.90E+02	3.78E+05
(I)	31.62	0.316	3162.0	0.003	2.16E+01	5.89E+08	3.17E+01	3.02E−01	3.18E+03	3.01E−03	9.58E+00	9.11E+02	6.47E+05

[a] Program fails when incorrect model is used for analysis.

Table 3

Data fit by 2:1 model with and without intermediates; 3% error on molecular weights and concentrations

Concentrations listed are maximum values for the system.

A

Data generated as 2:1 (without intermediates)	Results if data analyzed as								F ratios $(s_B^2 : s_A^2)$
	2:1 (without intermediates)		2:1 (with intermediates)						
	K	s_A^2	K_1	K_2	K_3	K_4	K'	s_B^2	
$K = 0.1$ (ml/mg)^2, STD M_W and concentration = 3%									
(A) $C = 0.1$ (mg/ml)	2.60E−26	2.31E+05	1.05E−02	3.70E−01	5.06E−03	7.65E−01	3.87E−03	2.38E+05	1.03E+00
(B) $C = 1.0$ (mg/ml)	1.17E−01	2.36E+07	1.72E−02	6.34E+00	4.40E−08	2.47E+06	1.09E−01	2.42E+07	1.03E+00
(C) $C = 10.0$ (mg/ml)	9.80E−02	4.11E+09	9.56E−10	1.02E+08	8.42E−11	1.16E+09	9.78E−02	4.21E+09	1.02E+00
$K = 10$ (ml/mg)^2, STD M_W and concentration = 3%									
(D) $C = 0.1$ (mg/ml)	1.02E+01	1.69E+05	1.31E−01	5.98E+01	2.29E−01	3.41E+01	7.84E+00	1.72E+05	1.02E+00
(E) $C = 1.0$ (mg/ml)	1.04E+01	4.98E+07	5.89E−03	1.76E+03	8.88E−08	1.17E+08	1.04E+01	5.11E+07	1.03E+00
(F) $C = 10.0$ (mg/ml)	9.32E+00	9.04E+09	–	–	–	–	–	–	– [a]
$K = 1000$ (ml/mg)^2, STD M_W and concentration = 3%									
(G) $C = 0.1$ (mg/ml)	1.05E+03	4.51E+05	7.96E−04	1.31E+06	1.26E+00	8.27E+02	1.04E+03	4.68E+05	1.04E+00
(H) $C = 1.0$ (mg/ml)	8.33E+02	8.50E+07	1.21E−25	6.69E+27	1.31E−01	7.18E+03	8.12E+02	8.73E+07	1.03E+00
(I) $C = 10.0$ (mg/ml)	8.17E+02	1.00E+10	–	–	–	–	–	–	– [a]

B

Data generated as 2:1 (with intermediates) ($K = 10$ (ml/mg)^2, $C = 1.0$ (mg/ml))					Results if data analyzed as								F ratios $(s_A^2 : s_B^2)$
					2:1 (without intermediates)		2:1 (with intermediates)						
	K_1	K_2	K_3	K_4	K	s_A^2	K_1	K_2	K_3	K_4	K'	s_B^2	
(A)	0.316	31.62	0.003	3162.0	9.25E+00	4.85E+07	5.59E−01	1.49E+01	1.20E−01	9.04E+01	1.09E+01	3.51E+07	1.38E+00
(B)	0.316	31.62	0.316	31.62	8.91E+00	6.10E+07	3.56E−01	2.79E+01	1.86E−01	5.35E+01	9.95E+00	5.03E+07	1.21E+00
(C)	0.316	31.62	31.62	0.316	8.81E+00	4.45E+08	4.15E−01	1.89E+01	3.03E+01	2.60E−01	7.87E+00	2.42E+07	1.84E+01
(D)	3.162	3.162	0.032	316.2	5.21E+00	1.25E+08	3.37E+00	2.88E+00	1.69E−08	5.73E+08	9.71E+00	4.04E+07	3.09E+00
(E)	3.162	3.162	3.162	3.162	6.48E+00	1.67E+08	3.25E+00	2.89E+00	3.89E+00	2.49E+00	9.69E+00	4.21E+07	3.97E+00
(F)	3.162	3.162	316.2	0.032	1.49E+01	6.76E+08	3.34E+00	6.10E+00	2.71E+02	7.53E−02	2.04E+01	4.47E+07	1.51E+01
(G)	31.62	0.316	0.316	31.62	2.41E+00	3.06E+08	2.79E+01	1.23E−01	1.74E+00	1.97E+00	3.40E+00	2.91E+07	1.05E+01
(H)	31.62	0.316	31.62	0.316	1.01E+01	4.15E+08	3.07E+01	8.59E−01	2.80E+01	9.39E−01	2.63E+01	5.64E+07	7.36E+00
(I)	31.62	0.316	3162.0	0.003	2.01E+01	7.16E+08	3.41E+01	7.46E−10	1.01E+03	2.53E−11	2.55E−08	5.92E+07	1.21E+01

[a] Program fails when incorrect model is used for analysis.

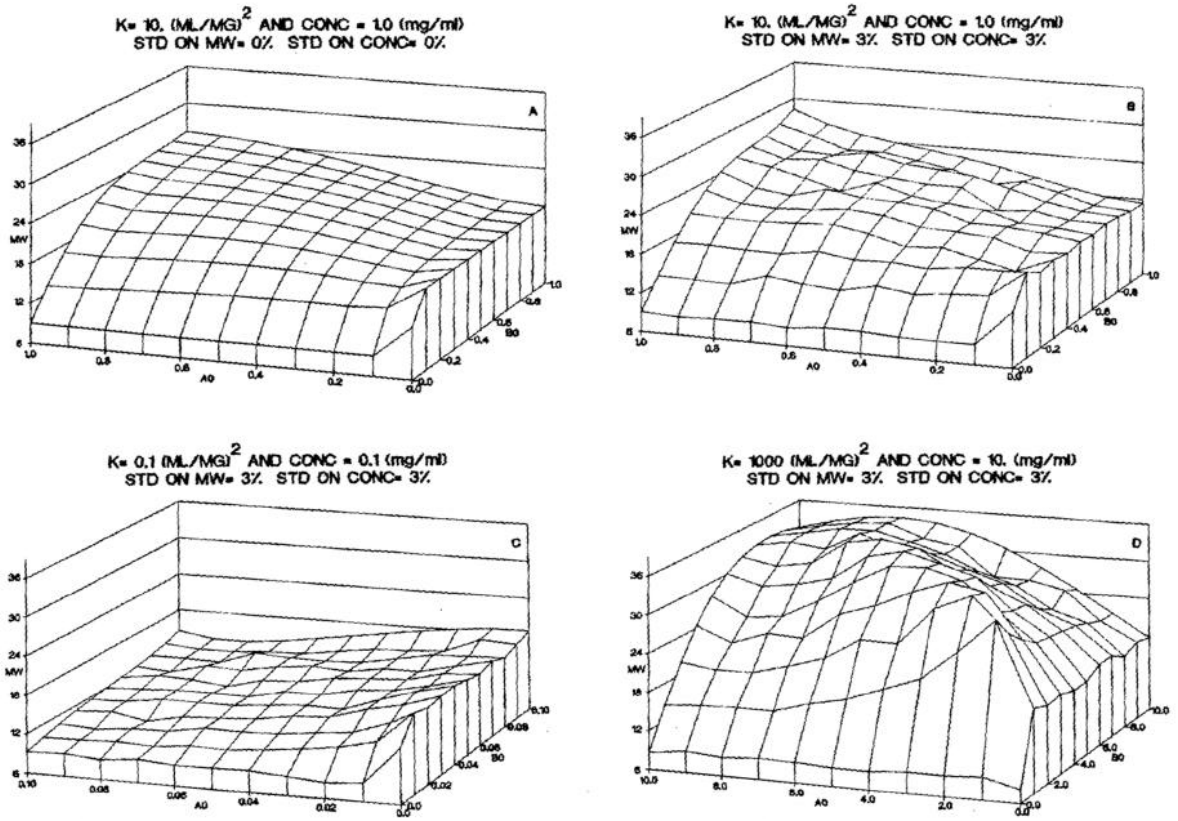


Fig. 1. Direct surface maps for system without intermediates present. Surface generated by plotting initial concentrations of species A and B vs. resulting weight-average molecular weight for four cases. Equilibrium (association) constants, maximum concentrations and magnitude of random errors are given in each panel.

are increased, as in fig. 1D ($K = 1000$ (ml/mg)^2 and $C = 10$ mg/ml), the surface again has the characteristic shape seen in fig. 1A and B. With 3% error imposed on M_W, A_0 and B_0, the fit of these data is still relatively good (data set I, table 3A).

The surface maps in fig. 2A–D depict the changes which occur due to variation in the $K_1 : K_2$ and $K_3 : K_1$ ratios as well as to increasing experimental error. Fig. 2A represents a surface with 0% error and ratios such that K_1, K_2, K_3 and K_4 are all equal. The surface is smooth and results in convergence upon the correct macroscopic and microscopic association constants (data set E, table 2B). As the error is increased to 3%, the surface becomes irregular (fig. 2B). The characteristic shape of the association surface is still present and results in proper convergence upon K', K_1 and K_3 (data set E, table 3B). In fig. 2C both K_1 and K_3 values are decreased by a factor of 10 and 1000, respectively. Enough A_2B complex is still formed that the overall K' is easily converged upon (data set A, table 3B); however, both K_1 and K_3 values are sufficiently low that not enough intermediates are formed to be able to return the proper microscopic constants. In fig. 2d both K_1 and K_3 values are increased (from those values for fig. 2B) by a factor of 10 and 1000, respectively. In this case very little A_2B complex is formed; as a result the correct K' value is not

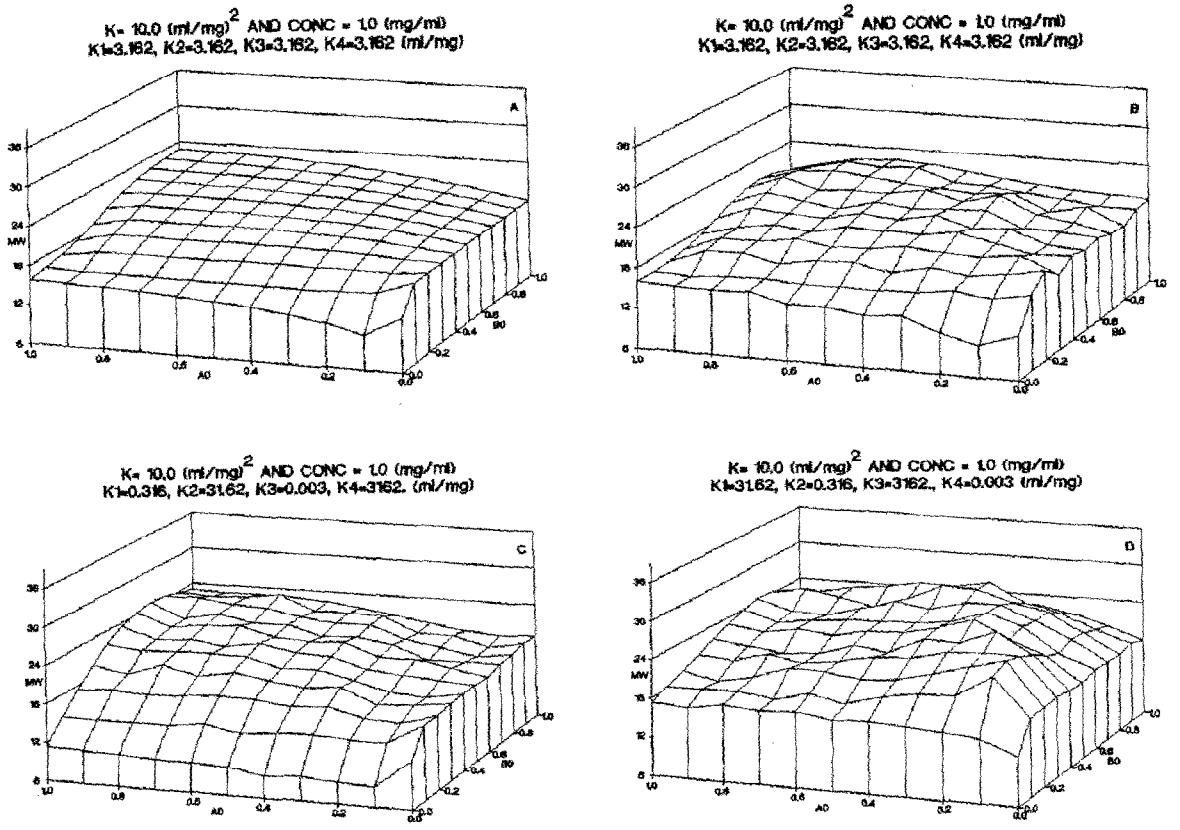


Fig. 2. Direct surface maps for system with intermediates present. Surface generated as in fig. 1. Macroscopic and microscopic equilibrium (association) constants and maximum concentrations are given in each panel. The magnitude of random error is 0% for panel A and 3% for panels B–D.

obtained (data set I, table 3B) although K_1 and K_3 are converged upon properly. Note the slight peak or ridge along the axis where B_0 is twice A_0, indicating a 1:1 molar ratio of the complexed species. In this case the surface map is more characteristic of the surface for an AB complex rather than for an A_2B complex.

3.4. Contour maps

Figs. 3 and 4 are two-dimensional contour plots of the data shown in figs. 1 and 2B–D. The initial concentrations for species A and B are along the x- and y-axes, respectively, and the weight-average molecular weight is represented by contour levels. For each case contour levels start at 90000 and are located near the x-axis. The contour level lines increase every 10000. For fig. 3A the lines of stoichiometry for a 1:1 and a 2:1 system are indicated. The thick dotted line is drawn for a 1:1 molar complex (2 g B per g A; $M_A = 90000$, $M_B = 180000$). A 2:1 molar ratio is indicated by the thin dotted line.

For a system without intermediates (model I), the stoichiometry (in weight ratios) can easily be determined from these contour plots. This is true even in some cases where the correct K value cannot be determined (fig. 3C). Under conditions of low concentrations and association constants, it is more difficult to determine the stoichiometry (fig. 3B). When intermediates (AB and/or A_2) are present in appreciable levels in addition to the

terminal A_2B complex, the stoichiometry is more difficult to determine. In fig. 4A all three species are present and the stoichiometry taken from the contour map is intermediate to that of either 1 : 1 or 2 : 1. Fig. 4B shows a case with low association constants and low total concentrations forming very little AB or A_2. Sufficient A_2B is present so that the stoichiometry taken from the map shows a 2 : 1 molar ratio. For the case represented in fig. 4C, little A_2 and virtually no A_2B complex is formed. This data set has exclusively AB present (in addition to free A and B) and thus the stoichoimetry estimated from this figure is very close to a 1 : 1 molar ratio. Overall, fig. 4 shows that contour maps are not as useful if intermediates are present and can easily lead to misinterpretation of the system.

4. Conclusion

The analysis procedure described here for a heterogeneous interacting macromolecular system which form 2 : 1 complexes can be used with any experimental procedure which determine the weight-average molecular weight of the complexes and the total concentrations of each of the two molecular species. This method can easily distinguish between a 1 : 1 and 2 : 1 complex as well as being able to distinguish between 2 : 1 systems with and without appreciable intermediate species. Even with a 3% random error imposed on measured A_0, B_0, and M_w values a high majority of the nine cases examined for model I (no intermediates) and the 25 cases examined (not all shown) for model II returned the correct overall

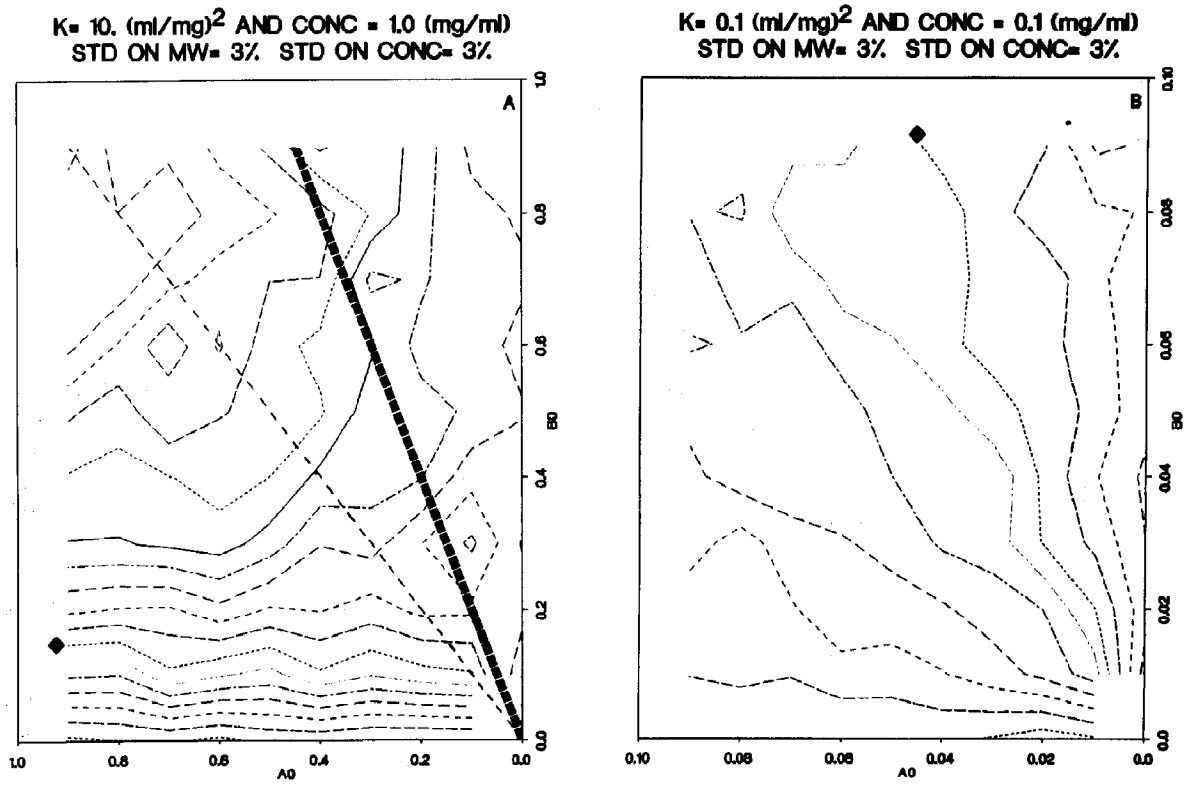


Fig. 3. Contour maps for system without intermediates present. Same data as in fig. 1B–D but represented as a contour. The contour levels differ in molecular weight by 10000. For ease of interpretation the 150000 level is marked with a filled diamond. The two lines in A represent the stoichiometry for 1 : 1 (thick dotted line) and 2 : 1 (thin dotted line) molar ratios.

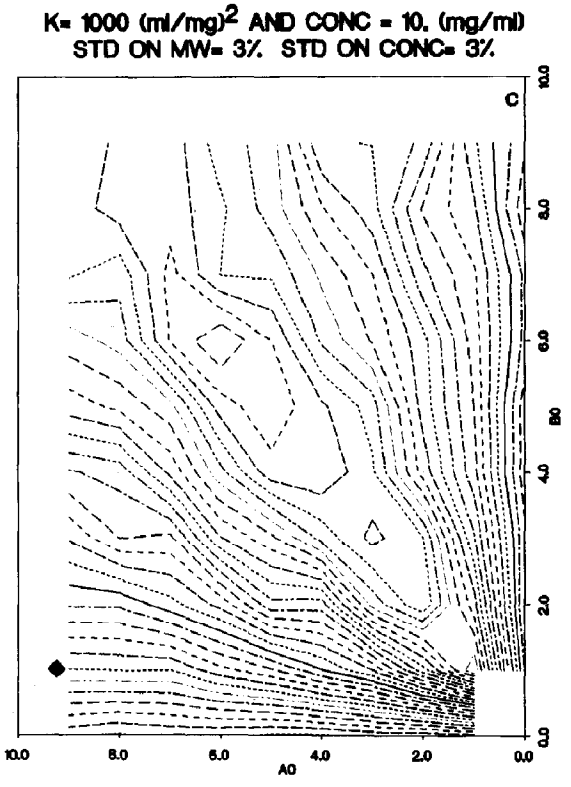


Fig. 3 (continued).

equilibrium constant and the correct microconstants for the predominate pathway in model II. The major limitation of this method is that for a given constant the procedure requires all reactants and products to have concentrations above the 'experimental' error in a substantial fraction of the data points included in the analysis.

Appendix A: System without intermediates – model I

For the system

$$2A + B \rightleftharpoons A_2B = C$$

the association constant can be expressed, in units of $(ml/mg)^2$, as

$$K = c/(a^2 b)$$

Let M_r of species A $= M_1$, of B $= M_2$ and of C $= M_3$, where $M_3 = 2M_1 + M_2$. Let $r_1 = 2M_1/M_3$ and $r_2 = M_2/M_3$. The total concentration of species A (a_0) and species B (b_0) can be expressed as

$$a_0 = a + cr_1 \tag{A1}$$

$$b_0 = b + cr_2. \tag{A2}$$

By definition,

$$M_W(a_0 + b_0) = aM_1 + bM_2 + a^2 bM_3 K. \tag{A3}$$

A1. Newton-Raphson procedure for solving b

Rearranging eqs. A1 and A2 and setting them equal to each other,

$$(a_0 - a)/r_1 = c = (b_0 - b)/r_2.$$

Solving for a

$$a = a_0 + (r_1/r_2)(b - b_0). \tag{A4}$$

Rearranging eq. A2 and substituting $c = a^2 bK$

$$b + a^2 br_2 K - b_0 = 0 \tag{A5}$$

Substituting eq. A4 into eq. A5,

$$\begin{aligned} 0 &= b + \{a_0 + (r_1/r_2)(b - b_0)\}^2 br_2 K - b_0 \\ &= b^3\{(r_1^2/r_2)K\} + b^2\{2r_1 K(a_0 - b_0(r_1/r_2))\} \\ &\quad + b\{r_2 K(a_0 - b_0(r_1/r_2))^2 + 1\} - b_0. \end{aligned} \tag{A6}$$

Let $X = a_0 - b_0(r_1/r_2)$, $Q_A = (r_1^2/r_2)K$, $Q_B = 2r_1 KX$ and $Q_C = r_2 KX^2 + 1$, then eq. A6 becomes

$$G = b^3 Q_A + b^2 Q_B + bQ_C - b_0 = 0. \tag{A7}$$

$$dG/db = 3b^2 Q_A + 2bQ_B + Q_C$$

$$b_{new} = b_{old} - \{G/(dG/db)\}.$$

Using eq. A4 to evaluate corresponding a values and recycle until G is within acceptable limits.

A2. Gauss-Newton procedure

Define $Y = M_w(a_0 + b_0)$, $F = aM_1 + bM_2 + a^2bM_3K$ and $R = \sum(Y - F)^2$.

$$dR/dK = -2\left\{\sum(Y - F)(dF/dK)\right\} = 0 = G$$

$$G = \sum Y(dF/dK) - \sum F(dF/dK) \qquad \text{(A8)}$$

$$dF/dK = (da/dK)\{M_1 + 2abM_3K\} + (db/dK)\{M_2 + a^2M_3K\} + a^2bM_3. \qquad \text{(A9)}$$

Substituting a^2bK for c and rearranging eq. A1 to solve for a

$$a = \left\{-1 \pm (1 + 4ba_0r_1K)^{1/2}\right\}/2br_1K \qquad \text{(A10)}$$

Let $Q = \{1 + 4ba_0r_1K\}^{1/2}$, then eq. A10 becomes

$$a = (Q - 1)/2br_1K. \qquad \text{(A11)}$$

For da/dK:

$$da/dK = d\{(Q - 1)/2br_1K\}/dK = \{bK(dQ/dK) - (Q - 1) \times (b + K(db/dK))\}/2r_1(bK)^2$$

Substituting for dQ/dK, where $dQ/dK = (2a_0r_1/Q)\{b + K(db/dK)\}$

$$da/dK = (b + K(db/dK))\left\{((2ba_0r_1K/Q) - (Q - 1))/2r_1(bK)^2\right\}$$

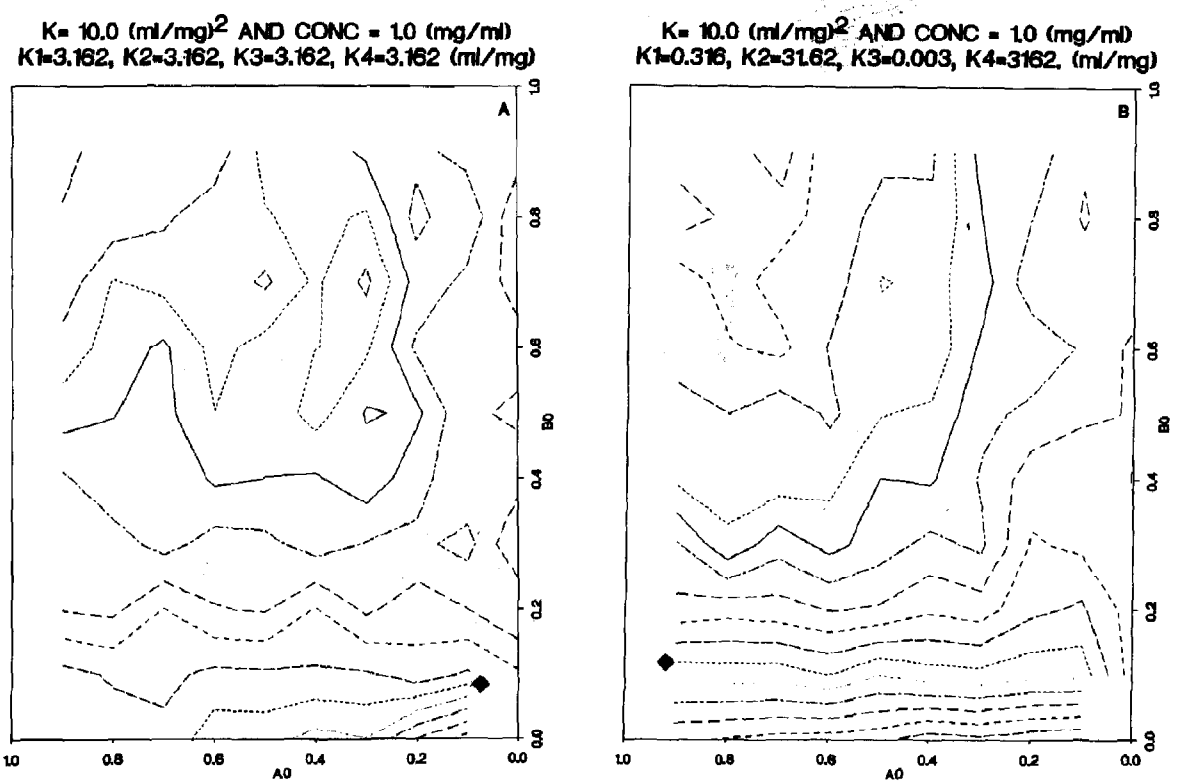


Fig. 4. Contour maps for system with intermediates present. Same data as in fig. 2B–D but represented as a contour. The contour levels differ in molecular weight by 10000. For ease of interpretation, the 150000 level is marked with a filled diamond.

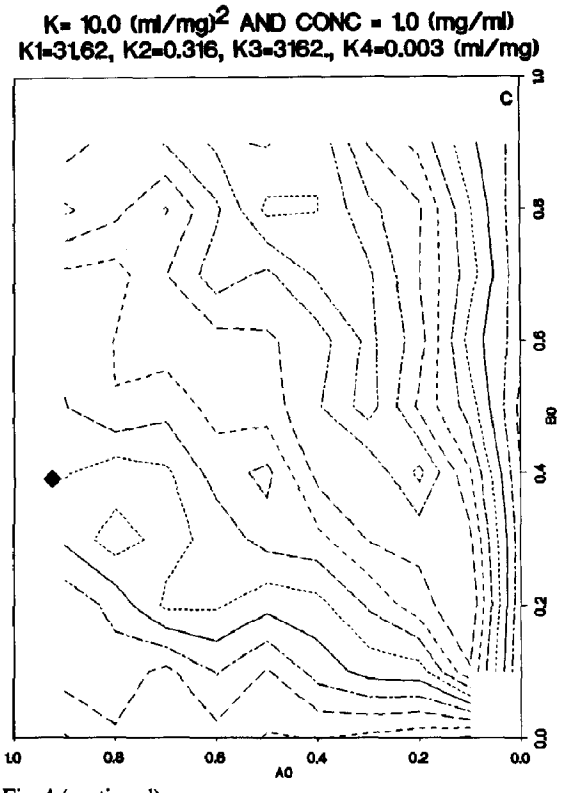


Fig. 4 (continued).

Let $Z = \{(2ba_0r_1K/Q) - (Q-1)\}/2r_1(bK)^2$, then

$$da/dK = \{b + K(db/dK)\}Z. \tag{A12}$$

Rearranging eq. A5

$$b = b_0/(1 + a^2r_2K). \tag{A13}$$

Let $W = 1 + a^2r_2K$, then eq. A13 becomes

$$b = b_0/W. \tag{A14}$$

For db/dK:

$$db/dK = d\{b_0/W\}/dK = -b_0(dW/dK)/W^2$$

Substituting for dW/dK, where $dW/dK = r_2\{a^2 + 2aK(da/dK)\}$,

$$db/dK = -(b_0/W^2)\{r_2(a^2 + 2aK(da/dK))\}$$
$$= -(b_0/W^2)r_2a^2 - (da/dK) \times \{2aK(b_0/W^2)\} \tag{A15}$$

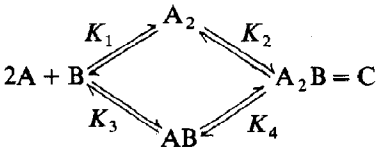
Then substituting eq. A12 for da/dK into eq. A15

$$db/dK = -(b_0/W^2)\{a^2r_2 + 2abKZ\} / \{1 + 2aK^2Z(b_0/W^2)\}. \tag{A16}$$

Eqs. A12 and A16 provide the expressions necessary to complete eq. A9. This provides the necessary terms for eq. A8 which is the equation one is minimizing.

Appendix B: System with intermediates – model II

For the system



the overall association constant, K', can be expressed as

$$K' = K_1K_2 = K_3K_4$$

where, microscopic association constants, K_1, K_2, K_3, and K_4 can be expressed as

$$K_1 = [A_2]/[A]^2 \qquad K_2 = [C]/[A_2][B]$$
$$K_3 = [AB]/[A][B] \qquad K_4 = [C]/[AB][B]$$
$$[A_2] = K_1[A]^2 \qquad [C] = K'[A]^2[B]$$
$$[AB] = K_3[A][B].$$

Let M_r of species A = M_1, of B = M_2, of AB = M_3 and of C = M_4, where $M_3 = M_1 + M_2$ and $M_4 = 2M_1 + M_2$. Since the concentration units in the paper are expressed in mg/ml, then let $r_1 = M_1/M_3$, $r_2 = 2M_1/M_4$, $r_3 = M_2/M_3$ and $r_4 = M_2/M_4$. The total concentration of species A(a_0) and species B(b_0) can be expressed as

$$a_0 = [A] + [A_2] + r_1[AB] + r_2[C]$$
$$= a + a^2K_1 + abr_1K_3 + a^2br_2K' \tag{B1}$$
$$b_0 = [B] + r_3[AB] + r_4[C]$$
$$= b + abr_3K_3 + a^2br_4K' \tag{B2}$$

where the concentrations are designated by lower-case letters, without brackets. By definition,

$$M_w(a_0 + b_0) = aM_1 + bM_2 + 2a^2K_1M_1 + abK_3M_3 + a^2bK'M_4. \quad \text{(B3)}$$

B1. Newton-Raphson procedure for solving a

Rearranging eq. B2 and solving for b

$$b = b_0/(1 + ar_3K_3 + a^2r_4K').$$

Let $D = 1 + ar_3K_3 + a^2r_4K'$, then

$$b = b_0/D. \quad \text{(B4)}$$

Rearranging eq. B1 and setting it equal to zero

$$a^2(K_1 + br_2K') + a(1 + br_1K_3) - a_0 = 0. \quad \text{(B5)}$$

Substituting for b, from eq. B4

$$a^2(K_1 + (b_0/D)r_2K') + a(1 + (b_0/D)r_1K_3) - a_0 = 0.$$

Rearranging and substituting for D,

$$0 = a^4(r_4K_1K') + a^3(r_3K_1K_3 + r_4K') + a^2(K_1 + b_0r_2K' + r_3K_3 - a_0r_4K') + a(1 + b_0r_1K_3 - a_0r_3K_3) - a_0. \quad \text{(B6)}$$

Let $Q_1 = r_4K_1K'$, $Q_2 = r_3K_1K_3 + r_4K'$, $Q_3 = K_1 + b_0r_2K' + r_3K_3 - a_0r_4K'$ and $Q_4 = 1 + b_0r_1K_3 - a_0r_3K_3$, then eq. B6 becomes

$$a^4Q_1 + a^3Q_2 + a^2Q_3 + aQ_4 - a_0 = 0 = G \quad \text{(B7)}$$

$$dG/da = 4a^3Q_1 + 3a^2Q_2 + 2aQ_3 + Q_4$$

$$a_{new} = a_{old} - \{G/(dG/da)\}.$$

Use eq. B4 to evaluate corresponding b values and recycle until G is within acceptable limits.

B2. Gauss-Newton procedure

Define $Y = M_w(a_0 + b_0)$, $F = aM_1 + bM_2 + 2a^2K_1M_1 + abK_3M_3 + a^2bK'M_4$ and $R = \Sigma(Y - F)^2$.

F is a function of three independent variables K', K_1, K_3.

$$dR = 0 = G = \Sigma Y(dF) - \Sigma F(dF), \quad \text{(B8)}$$

where

$$dF = (\delta F/\delta K')dK' + (\delta F/\delta K_1)\,dK_1 + (\delta F/\delta K_3)dK_3 \quad \text{(B9)}$$

$$\delta F/\delta K' = (\delta F/\delta a)(\delta a/\delta K') + (\delta F/\delta b)(\delta b/\delta K') + a^2bM_4 \quad \text{(B10)}$$

$$\delta F/\delta K_1 = (\delta F/\delta a)(\delta a/\delta K_1) + (\delta F/\delta b)(\delta b/\delta K_1) + 2a^2M_1 \quad \text{(B11)}$$

$$\delta F/\delta K_3 = (\delta F/\delta a)(\delta a/\delta K_3) + (\delta F/\delta b)(\delta b/\delta K_3) + abM_3 \quad \text{(B12)}$$

$$\delta F/\delta a = M_1 + 4aK_1M_1 + bK_3M_3 + 2abK'M_4$$

$$\delta F/\delta b = M_2 + aK_3M_3 + a^2K'M_4.$$

Let $Q_A = K_1 + br_2K'$ and $Q_B = 1 + br_1K_3$. Substituting Q_A and Q_B into eq. B5 and solving for a,

$$a = \{-Q_B \pm (Q_B^2 + 4a_0Q_A)^{1/2}\}/2Q_A.$$

Let $T = (Q_B^2 + 4a_0Q_A)^{1/2}$, then

$$a = \{T - Q_B\}/2Q_A. \quad \text{(B13)}$$

Using eq. B13 to find $\delta a/\delta K_i$, where i indicates any of the three K parameters,

$$\delta a/\delta K_i = \{Q_A\{(\delta T/\delta K_i) - (\delta Q_B/\delta K_i)\} + (Q_B - T)(\delta Q_A/\delta K_i)\}/2Q_A^2 \quad \text{(B14)}$$

$$\delta Q_A/\delta K' = br_2 + (\delta b/\delta K')r_2K'$$
$$\delta Q_B/\delta K' = (\delta b/\delta K')r_1K_3$$
$$\delta Q_A/\delta K_1 = 1 + (\delta b/\delta K_1)r_2K'$$
$$\delta Q_B/\delta K_1 = (\delta b/\delta K_1)r_1K_3$$
$$\delta Q_A/\delta K_3 = (\delta b/\delta K_3)r_2K'$$
$$\delta Q_B/\delta K_3 = br_1 + (\delta b/\delta K_3)r_1K_3$$

$$\delta T/\delta K_i = \{Q_B(\delta Q_B/\delta K_i) + 2a_0(\delta Q_A/\delta K_i)\}/T,$$

then

$$\delta T/\delta K' = \{(\delta b/\delta K')(r_1K_3Q_B + 2a_0r_2K') + 2a_0br_2\}/T$$

$$\delta T/\delta K_1 = \{(\delta b/\delta K_1)(r_1K_3Q_B + 2a_0r_2K') + 2a_0\}/T$$

$$\delta T/\delta K_3 = \{(\delta b/\delta K_3)(r_1K_3Q_B + 2a_0r_2K') + br_1Q_B\}/T.$$

Let $Q_C = r_1K_3Q_B + 2a_0r_2K'$, then

$$\delta T/\delta K' = \{(\delta b/\delta K')Q_C + 2a_0br_2\}/T$$

$$\delta T/\delta K_1 = \{(\delta b/\delta K_1)Q_C + 2a_0\}/T$$

$$\delta T/\delta K_3 = \{(\delta b/\delta K_3)Q_C + br_1Q_B\}/T.$$

Substituting for $\delta T/\delta K_i$, $\delta Q_A/\delta K_i$ and $\delta Q_B/\delta K_i$ into eq. B14,

$$\delta a/\delta K' = \{(\delta b/\delta K')\{(Q_AQ_C/T) + r_2K'(Q_B - T) - r_1K_3Q_A\} + br_2\{(Q_B - T) + (2a_0Q_A/T)\}\}/2Q_A^2$$

$$\delta a/\delta K_1 = \{(\delta b/\delta K_1)\{(Q_AQ_C/T) + r_2K'(Q_B - T) - r_1K_3Q_A\} + \{(Q_B - T) + (2a_0Q_A/T)\}\}/2Q_A^2$$

$$\delta a/\delta K_3 = \{(\delta b/\delta K_3)\{(Q_AQ_C/T) + r_2K'(Q_B - T) - r_1K_3Q_A\} + br_1Q_A\{(Q_B/T) - 1\}\}/2Q_A^2.$$

Let $Q_F = (Q_B - T) + 2a_0Q_A/T$, $XXX = \{(Q_AQ_C/T) + r_2K'(Q_B - T) - r_1K_3Q_A\}/2Q_A^2$,

$YYK = br_2Q_F/2Q_A^2$, $YY_1 = Q_F/2Q_A^2$ and $YY_3 = \{br_1Q_A\{(Q_B/T) - 1\}/2Q_A^2$, then

$$\delta a/\delta K' = (\delta b/\delta K')XXX + YYK \quad \text{(B15)}$$

$$\delta a/\delta K_1 = (\delta b/\delta K_1)XXX + YY_1 \quad \text{(B16)}$$

$$\delta a/\delta K_3 = (\delta b/\delta K_3)XXX + YY_3. \quad \text{(B17)}$$

From eq. B4

$\delta b/\delta K_i = -b_0(\delta D/\delta K_i)/D^2$, where

$$\delta D/\delta K' = (\delta a/\delta K')\{r_3K_3 + 2ar_4K'\} + a^2r_4$$

$$\delta D/\delta K_1 = (\delta a/\delta K_1)\{r_3K_3 + 2ar_4K'\}$$

$$\delta D/\delta K_3 = (\delta a/\delta K_3)\{r_3K_3 + 2ar_4K'\} + ar_3.$$

Let $Q_G = r_3K_3 + 2ar_4K'$, then

$$\delta b/\delta K' = (-b_0/D^2)\{(\delta a/\delta K')Q_G + a^2r_4\} \quad \text{(B18)}$$

$$\delta b/\delta K_1 = (-b_0/D^2)\{(\delta a/\delta K_1)Q_G\} \quad \text{(B19)}$$

$$\delta b/\delta K_3 = (-b_0/D^2)\{(\delta a/\delta K_3)Q_G + ar_3\}. \quad \text{(B20)}$$

Substituting $\delta b/\delta K'$ into eq. B15 and solving for $\delta a/\delta K'$,

$$\delta a/\delta K' = (-b_0/D^2)\{(\delta a/\delta K')Q_G + a^2r_4\} \times XXX + YYK = \{YYK - (a^2b_0r_4XXX/D^2)\}/\{1 + (b_0/D^2)Q_GXXX\}.$$

Substituting $\delta b/\delta K_1$ into eq. B16 and solving for $\delta a/\delta K_1$,

$$\delta a/\delta K_1 = (-b_0/D^2)\{(\delta a/\delta K_1)Q_G\}XXX + YY_1 = YY_1/\{1 + (b_0/D^2)Q_GXXX\}.$$

Substituting $\delta b/\delta K_3$ into eq. B17 and solving for $\delta a/\delta K_3$,

$$\delta a/\delta K_3 = (-b_0/D^2)\{(\delta a/\delta K_3)Q_G + ar_3\}XXX + YY_3 = \{YY_3 - XXX(b_0/D^2)ar_3\}/\{1 + (b_0/D^2)Q_GXXX\}.$$

Let $Q_H = 1 + (b_0/D^2)Q_GXXX$, then

$$\delta a/\delta K' = \{YYK - a^2b_0r_4XXX/D^2\}/Q_H \quad \text{(B21)}$$

$$\delta a/\delta K_1 = YY_1/Q_H \quad \text{(B22)}$$

$$\delta a/\delta K_3 = \{YY_3 - XXX(b_0/D^2)ar_3\}/Q_H \quad \text{(B23)}$$

Substituting equations B18–B23 into B10–B12 to

evaluate $\delta F/\delta K_i$ provides all necessary terms for the system to be evaluated.

Acknowledgement

This work was carried out under PHS Grant AI23099.

References

1 S.H. Tindall and K.C. Aune, Biochemistry 20 (1981) 4861.
2 S.H. Tindall and K.C. Aune, Anal. Biochem. 120 (1982) 71.
3 M.S. Lewis and R.J. Youle, J. Biol. Chem. 261 (1986) 11571.
4 P. Hensley, M.C. O'Keefe, C.J. Spangler, J.C. Osborne, Jr and C.W. Vogel, J. Biol. Chem. 261 (1986) 11038.
5 E. Eienstein, L.T. Duong, R.L. Ornberg, J.C. Osborne, Jr and P. Hensley, J. Biol. Chem. 261 (1986) 12814.
6 R.C. Chatelier and A.B. Minton, Biophys. J. 51 (1987) 94a.
7 M.S. Lewis, Biophys. J. 51 (1987) 441a.
8 P. Lollar, Biophys. Chem. 28 (1987) 245.
9 M.S. Lewis, S.W. Luborsky and K.M. Yamada, Biophys. J. 53 (1988) 75a.
10 E.T. Adams, Jr, A.H. Pekar, D.A. Soucek, L.H. Tang, G. Barlow and J.L. Armstrong, Biopolymers 7 (1969) 5.
11 B.W. McCarty and E.T. Adams, Jr, Biophys. Chem. 28 (1987) 149.
12 G.K. Ackers, in: Methods of protein separation, vol. 2, ed. N. Catsimpoolas (Plenum, New York, 1976) p. 1.
13 L.W. Nichol, R.J. Siezen and D.J. Winzor, Biophys. Chem. 9 (1978) 47.
14 J.R. Cann and D.J. Winzor, Arch. Biochem. Biophys. 256 (1987) 78.
15 R.F. Steiner, Arch. Biochem. Biophys. 49 (1954) 71.
16 A.H. Pekar, P.J. Wan and E.T. Adams, Jr, Adv. Chem. 9 (1978) 47.
17 P. Lollar, C.G. Parker and S. Krishnaswamy, Biophys. J. 53 (1988) 76a.
18 J.K. Zimmerman and M.L. Crowl-Powers, Biophys. Chem. 29 (1988) 231.
19 E.D. Sprague, C.E. Larrabee, Jr and H.B. Halsall, Anal. Biochem. 101 (1980) 175.
20 H.J. Motulsky and L.A. Ransnas, FASEB J. 1 (1987) 365.
21 W.G. Bardsley, P.B. McGinlay and M.G. Roig, Biophys. Chem. 26 (1987) 1.