Biophysical Chemistry 28 (1987) 245–251
Elsevier

BPC 01203

Heterogeneous, ideal associations at sedimentation equilibrium

Pete Lollar

Departments of Medicine and Biochemistry, University of Vermont, Burlington, VT 05405, U.S.A.

Received 7 July 1987
Accepted 30 September 1987

Binding constant; Sedimentation equilibrium; Heterogeneous association; Self-association; Stoichiometry; Nonlinear regression

A general approach to analyze associations of the type $mA + nB = A_mB_n$ at sedimentation equilibrium under ideal conditions was derived. A simple transform to generate a function Q was defined in which the optical measurement (absorbance or refractive index increment) at a given radial distance is divided by the value that would result if no association occurred. Explicit calculations for the cases of $m = 1$, $n = 1$ and $m = 2$, $n = 1$ were used to simulate centrifuge experiments. Nonlinear least-squares regression of Q vs. the optical measurement showed that accurate estimation of parameters could be obtained in selected instances for both types of association only if the appropriate model was used. Additionally, the effect of random experimental and systematic error (e.g., the presence of contaminants) was evaluated.

1. Introduction

The analysis of the interaction of unlike molecules is of fundamental importance in biology. Analytical ultracentrifugation is potentially a useful method in this regard, since a concentration gradient is generated at sedimentation equilibrium that depends on stoichiometry and binding energy. Initially, the study of interacting systems by equilibrium sedimentation involved the calculation of average molecular weight moments as a function of concentration [1]. Nonlinear regression fitting to models of association is potentially a more accurate method, since it avoids numerical differentiation inherent in the calculation of molecular weight moments. As a result, analysis of centrifuge data, particularly from self-associating systems, by nonlinear regression methods has become more widely used [2,3]. In this study, a general approach to the calculation of the concentration distribution of all species at sedimentation equilibrium for ideal associations of the type $mA + nB = A_mB_n$ is given. Additionally, a practical method for evaluating models of association and estimating equilibrium constants, based on simulations for the cases $m = 1$, $n = 1$ (type 1:1 association) and $m = 2$, $n = 1$ (type 2:1 association), is suggested.

Correspondence address: P. Lollar, Departments of Medicine and Biochemistry, University of Vermont, Burlington, VT 05405, U.S.A.

2. Theory

2.1. General

Consider an ideal association of the type $mA + nB = A_mB_n$ in which B can self-associate to form an n-mer and there are m/n binding sites for A per monomer of B. A set of $[m(1 + 2 + \ldots n)/n] + n - 1$ equilibrium reactions can be written:

$$B + B_{s-1} \leftrightarrows B_s \qquad s = 2, 3, \ldots, n \qquad (1a)$$

$$A + A_{r-1}B_s \leftrightarrows A_rB_s \qquad r = 1, 2, \ldots, (ms/n) \quad (A_0B_s = B_s)$$

$$s = 1, 2, \ldots, n \qquad (1b)$$

These reactions can be described by the macroscopic equilibrium constants

$$\frac{c_{A_rB_s}}{c_A c_{A_{r-1}B_s}} = \hat{k}_{rs} \tag{2a}$$

$$\frac{c_{B_s}}{c_B c_{B_{s-1}}} = \hat{k}_{0s} \tag{2b}$$

where $c_i(x^2)$ represents the weight concentration (g/l) of i as a function of radial position and $\hat{k}_{rs}$ the equilibrium constant (l/g). It is assumed that the equilibrium constants are independent of radial position. Self-association of A can also be included in the model but is omitted here for the sake of simplicity.

Given its initial concentration, the concentration of any species at equilibrium is given by the Rinde equation (ref. 4, eq. II-83)

$$c_i = \frac{c_i'(\sigma_{i/2})(x_b^2 - x_m^2)\exp\left[(\sigma_{i/2})(x^2 - x_m^2)\right]}{(E_i - 1)} \tag{3}$$

where c_i' is the whole-cell concentration of i at sedimentation equilibrium, σ_i the effective reduced molecular weight, x_b^2 and x_m^2 the squared radial position at cell bottom and meniscus, respectively, x the radial position, and

$$E_i = \exp\left[\sigma_i/2(x_b^2 - x_m^2)\right] \tag{4}$$

The effective reduced molecular weight equals

$$M_i(1 - \bar{v}_i\rho)\omega^2/RT \tag{5}$$

where M_i and $\bar{v}_i$ are the molecular weight and partial specific volume of i, respectively, ρ the solution density, ω the angular velocity, R the ideal gas constant, and T the absolute temperature. Substitution of eq. 3 into eq. 2a yields

$$\frac{c'_{A_rB_s}}{c'_A c'_{A_{r-1}B_s}\alpha_{rs}} = \hat{k}_{rs} \tag{6}$$

where

$$\alpha_{rs} = \frac{\sigma_A\left[(r-1)\sigma_A + s\sigma_B\right]}{2(r\sigma_A + s\sigma_B)} \cdot \frac{(E_{A_rB_s} - 1)}{(E_A - 1)(E_{A_{r-1}B_s} - 1)} \tag{7}$$

To obtain eq. 4, we note that effective reduced molecular weights are additive (e.g., $\sigma_A + \sigma_B = \sigma_{AB}$) (ref. 1, eq. 6.105). Additionally, the exponentials involving x^2 in eq. 3 cancel, since the equilibrium constant is independent of x^2. A similar expression can be written for the self-association process. Now it is necessary to convert to molar concentrations because the stoichiometry numbers m and n are defined in molar terms. Since $C_i' = c_i'/M_i$ represents the whole-cell molar concentration of i,

$$\frac{C'_{A_rB_s}}{C'_A C'_{A_{r-1}B_s}} = K'_{rs} \tag{8}$$

where

$$K'_{rs} = \frac{\hat{k}_{rs}\alpha_{rs}M_A\left[(r-1)M_A + sM_B\right]}{rM_A + sM_B} \tag{9}$$

can be viewed as an apparent whole-cell equilibrium constant (M^{-1}). Including A and B, $[m(1 + 2 + \ldots n)/n] + n + 1$ species are defined by eq. 1. Together with the conservation of mass equations that exist for both A and B, the number of equations involving K'_{rs} equals the number of unknown concentrations at equilibrium.

Generally, a set of simultaneous nonlinear equations describing chemical equilibria can be solved for the unknown concentrations of all species by numerical methods [5]. For the simplest associations, specific solutions may be used as outlined below.

2.2. Type 1 : 1 associations

This process is defined by $m = n = 1$. Explicit solutions for the concentrations of the three species, A, B and AB, can be obtained in this case. Using the conservation equations

$$C'_{A,o} = C'_A + C'_{AB} \text{ and } C'_{B,o} = C'_B + C'_{AB} \tag{10}$$

where $C'_{i,o}$ is the initial whole-cell molar concentration of i, eq. 8 becomes

$$\frac{C'_{AB}}{(C'_{A,o} - C'_{AB})(C'_{B,o} - C'_{AB})} = K'_{11}, \tag{11}$$

Given K'_{11} and the known starting concentration

of A and B, the whole-cell molar concentrations of all species at sedimentation equilibrium can be determined using the quadratic formula:

$$C'_{AB} = \frac{(C'_{A,o} + C'_{B,o} + 1/K'_{11})}{2} - \frac{\sqrt{\left[(C'_{A,o} + C'_{B,o} + 1/K'_{11})^2 - 4C'_{A,o}C'_{B,o}\right]}}{2} \quad (12)$$

Since optical systems in the ultracentrifuge measure quantities that are proportional to weight concentrations, the molecular weights of the various species are used to convert back to weight concentrations. The weight concentration of each species as a function of radial position, c_i, is then obtained by substitution into eq. 3. The optical signal is given by

$$S(x^2) = P_A c_A + P_B c_B + P_{AB} c_{AB} \quad (13)$$

where P_A, etc., are optical constants (extinction coefficients or refractive index increments) for each species.

Although the optical constants of A and B can be determined, that of AB is usually inaccessible. It can be estimated (ref 1, eq. 6.114) from

$$P_{AB} = \frac{M_A P_A + M_B P_B}{M_{AB}} \quad (14)$$

Therefore, given the initial loading concentrations, molecular weights, partial specific volumes, and optical constants of A and B, as well as the equilibrium constant and centrifuge conditions, it is possible to determine the optical signal at any position.

2.3. Type 2:1 associations

If B has two binding sites for A and does not self-associate, then the equilibria are described by eq. 1b with $m = 2$ and $n = 1$. The two equilibrium constants are not necessarily equal. Converting to molar concentration terms as before,

$$C'_{AB} = K'_{11} C'_A C'_B \quad (15)$$

$$C'_{A_2B} = K'_{21} C'_A C'_{AB} \quad (16)$$

$$C'_{A,o} = C'_A + C'_{AB} + 2C'_{A_2B} \quad (17)$$

$$C'_{B,o} = C'_B + C'_{AB} + C'_{A_2B} \quad (18)$$

Substituting eq. 17 into eq. 16 and rearranging yields:

$$C'_{A_2B} = \frac{K'_{21} C'_{AB}(C'_{A,o} - C'_{AB})}{(1 + 2K'_{21} C'_{AB})} \quad (19)$$

Substituting eq. 17 and 18 into eq. 15 and using eq. 19 to eliminate C'_{A_2B},

$$C'_{AB} = K'_{11}\left[C'_{A,o} - C'_{AB} - 2jK'_{21}C'_{AB}(C'_{A,o} - C'_{AB})\right] \times \left[C'_{B,o} - C'_{AB} - jK'_{21}C'_{AB}(C'_{A,o} - C'_{AB})\right] \quad (20)$$

where

$$j = (1 + 2K'_{21}C'_{AB})^{-1}.$$

This cubic equation can be solved numerically for C'_{AB} using the Newton-Raphson iteration. Eqs. 17–19 are then solved for C'_{A_2B}, C'_A and C'_B. The analysis then proceeds as in the case of A + B = AB.

2.4. Least-squares analysis

Experimental curves result from the superposition of several exponential curves, one for each species in the cell. When applied to sedimentation equilibrium data, nonlinear least-squares fitting to multiple exponentials can be complicated by highly correlated parameters [6]. An empirical approach is to reduce the data by dividing all values of the dependent variable (e.g., absorbance) by some function to obtain a new equation that is not a sum of exponentials. Therefore, we define the function

$$Q(x^2) = \frac{S(x^2)}{\left[U_A(x^2) + U_B(x^2)\right]} \quad (21)$$

where U_A and U_B are the optical signals that would result from A and B if no association occurred. These values are determined using eq. 3 into which the starting concentrations of A and B are substituted into the right-hand side for c'_i. The values of U are then obtained using the optical

constants for A and B, respectively. This conversion adds more information to the analysis in that each data pair represents a comparison between an observed value and a calculated value that is based on the initial concentrations and effective reduced molecular weights of the reacting species. These values must be obtained from independent measurements. Data sets of Q vs. c are fitted using the Marquardt algorithm [7]. For example, for type 1:1 interactions, there is a single meaningful value of x^2 that satisfies eq. 13 for a given parameter estimate (in this case $\hat{k}_{11}$). It can be found using the Newton-Raphson iteration. Then the denominator in eq. 21 is calculated using this value of x^2. This determines a calculated value of Q for any given value of $S(x^2)$. The squared differences between the calculated Q values and the experimental Q values are then summed. The algorithm attempts to minimize the differences by changing the parameter estimate(s). Least-squares analysis for type 2:1 associations proceeds similarly.

If the error in the measurement of the optical signal is normally distributed with constant standard deviation, then the Q function will also be normally distributed with a standard deviation that is proportional to $[U_A(x^2) + U_B(x^2)]^{-1}$. Therefore, the data are weighted by this factor in the analysis. When viewed graphically, there is greater scatter in the Q function at low concentrations, but these points have less weight.

The estimated standard error in the parameters is accurate only in the limit that the analytical function approximates a linear function. Johnson et al. [2] have shown that the confidence interval of the estimate can be more accurately obtained using the F statistic. Therefore, the error reported from the standard regression analysis only approximately reflects the true error. Moreover, this error is due solely to the random error resulting from instrumental uncertainty. A form of systematic error also occurs because of error in the constant values of effective reduced molecular weights, initial concentrations, and optical constants that are used in the calculation of all values of the Q function. The magnitude of this error can be estimated by propagation of error analysis [7]. This requires a calculation of the change in the dependent variable, $\hat{k}_{rs}$, as a function of σ_i, etc. This cannot be obtained analytically, but can be estimated numerically by observing the change in $\hat{k}_{rs}$ due to small fluctuations in σ_i, etc., in simulated experiments.

3. Results

3.1. Type 1:1 association

An example of a data fitting to a simulated type 1:1 association between two unlike molecules is shown in fig. 1. The simulated data have been transformed using the definition of the Q function and fitted to a type 1:1 model using weighted nonlinear least-squares regression. The data are typical of those that might be obtained using absorption optics. The simulated experimental error, $0.005A$, probably slightly overestimates the error resulting from available optical systems [8]. The estimated binding constant, 19.2 ± 0.71 l/g is close to the 'true' equilibrium constant of 20 l/g and corresponds to less than 5% error. Analy-

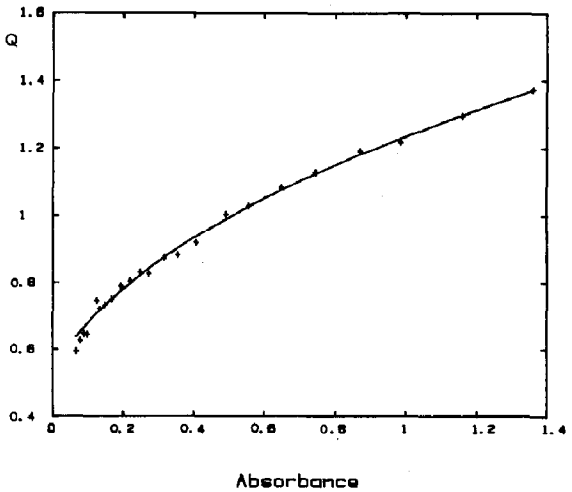

Fig. 1. Least-squares fit of Q vs. absorbance for a type 1:1 association. Input conditions: $M_A = 65\,000$; $M_B = 45\,000$; $\bar{v}_A = 0.72$ ml/g; $\bar{v}_B = 0.71$ ml/g; $P_A = 1$ l/g per cm; $P_B = 1.5$ l/g per cm; initial concentrations were 0.15 and 0.1 g/l for A and B, respectively; rpm = 16 000; $\rho = 1$ g/ml; $T = 20\,^{\circ}C$; $x_b = 7.176$ cm; $x_m = 6.964$ cm; $\hat{k}_{11} = 20$ l/g; error = $0.005A$. A parameter estimate of 19.2 ± 0.71 l/g was obtained. The curve connects the calculated values.

sis of repetitive simulations revealed no bias in the estimation of the parameter. Attempts to fit a type 2:1 model to the data resulted in a progressive decrease in the estimate of $\hat{k}_{21}$ and no convergence.

In the experiment described in fig. 1, it was assumed that there was no error in the fixed parameters. If the same data are used and instead an incorrect value for M_B of 48 000 is used (7% error) then the estimated value of $\hat{k}_{11}$ becomes 15.6 ± 0.56 l/g resulting in a systematic error that is larger than the random error. The fact that the binding constant is underestimated is not apparent from inspection of a graph of the fitted data, nor does it become apparent from changing the initial concentrations or rotor conditions (data not shown). The error used in the simulation is for illustrative purposes and is larger than that found under actual conditions using absorption optics ($\approx 3\%$) [9] if the effective reduced molecular weights of A and B are determined using the same preparations as used in the association experiment.

3.2. Effect of inactive molecules

As a second example, an experiment was simulated for a type 1:1 association between two macromolecules with very different physical properties including no absorbance of B (table 1). Additionally, in one experiment 20% of the A molecules were 'dead'. This results in a systematic error due to error in the initial concentration of A along with superimposed heterogeneity due to another species that is not accounted for in the model. The inactive molecules lead to an underestimation of the equilibrium constant. The estimated errors are not increased and no trend in the incorrect estimate is evident when the concentration is doubled.

Table 1

Effect of inactive molecules on $\hat{k}_{11}$ for a type 1:1 association

Conditions: $M_A = 50\,000$; $M_B = 10\,000$: $\bar{v}_A = 0.72$ ml/g; $\bar{v}_B = 0.65$ ml/g; $P_A = 1$; $P_B = 0$; rpm = 24 000; $\hat{k}_{11} = 25$ l/g. The mass ratio of A to B (whole cell) was 5. Other conditions were as in fig. 1.

% inactive A	Nominal concentration of A plus B (g/l)	$\hat{k}_{11}$ (l/g)
0	0.09	31.7 ± 4.9
20		16.9 ± 2.8
0	0.18	25.8 ± 1.9
20		16.7 ± 1.2

3.3. Similar sized visible species

The example in table 1 also illustrates the fact that the Q function can provide good fits to the data under difficult circumstances. The B molecule is invisible so that the two visible species differ in molecular weight by only about 20%. The estimated standard error in the simulation is fairly small. When the simulation was repeated ten times, $\hat{k}_{11} = 25.8 \pm 1.8$ l/g (mean $\pm$ S.D.). Because of the small difference in molecular weight between the visible species, systematic error due to error in molecular weight determination led to larger errors than in the previous example. For instance, when an incorrect molecular weight of 51 000 for A was used (2% error), $\hat{k}_{11}$ decreased from 25 to 16 l/g.

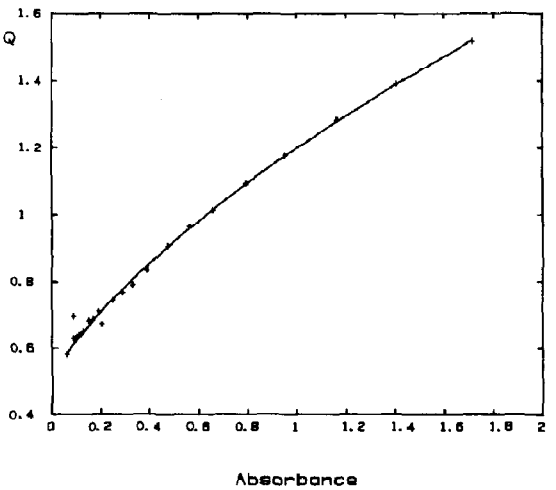

Fig. 2. Least-squares fit of Q vs. absorbance for a type 2:1 association. Conditions as in fig. 1 except $\hat{k}_{11} = 10$ l/g and $\hat{k}_{21} = 2$ l/g were used as input parameters. The best fit to the data gave values of 10.9 ± 1.4 l/g and 1.6 ± 0.32 l/g for $\hat{k}_{11}$ and $\hat{k}_{21}$, respectively.

Table 2

Type 2:1 association to a smaller invisible molecule

Conditions were the same as for table 1. The nominal concentration of A plus B was 0.18 g/l.

	Equilibrium constant (l/g)		Correlation coefficient
	$\hat{k}_{11}$	$\hat{k}_{21}$	
Actual	25	10	−0.987
Calculated	35.2 ± 10.8	8.9 ± 0.83	

3.3. Type 2:1 association

A sample simulation is shown in fig. 2. Reasonably good estimates of the two equilibrium constants were obtained. Convergence was not reached when a type 1:1 association was used as the model. When more than one parameter is estimated, the correlation coefficient between any two parameters can be determined [7]. As the absolute value of a correlation coefficient increases it becomes more difficult to find unique parameter values that give a best fit to the data. Although the cutoff limit is somewhat arbitrary, Johnson and Frasier [10] have suggested 0.96 as a reasonable value. For the simulation in fig. 1, the two parameters were not highly correlated ($r = -0.94$). The experiment indicates that under selected circumstances higher-order associations may be studied using this technique.

An example where the Q function fails occurs when the simulation conditions are identical to those given in table 1 except that a type 2:1 association is present with $\hat{k}_{11} = 25$ and $\hat{k}_{21} = 10$ (table 2). There is a large error in $\hat{k}_{11}$ and the parameters are highly correlated.

4. Discussion

In comparison to the analysis of self-associating molecules, heterogeneous associations at sedimentation equilibrium have been less extensively studied. The purpose of this investigation was to simulate heterogeneous associations between molecules that may or may not have different effective reduced molecular weights, optical constants, or starting concentrations and to develop a method to fit the data. The key to the analysis is the assumption of conservation of mass in the centrifuge cell and the fact that each species has a concentration distribution at equilibrium given by the Rinde equation (eq. 3). Additionally, when substituting conservation equations into expressions involving an equilibrium constant, particular attention must be paid to the units of concentration. Stoichiometry is defined in this study in molar terms (i.e., m and n represent moles of A and B, respectively). Therefore, the equilibrium constant expressions must be temporarily converted to molar concentrations (eq. 8). The units ultimately are converted back to weight concentration terms, since measurements in the ultracentrifuge are usually proportional to this quantity.

The problem of determining the concentration of all species at sedimentation equilibrium becomes one of finding the solution of a set of simultaneous nonlinear equations. This is straightforward for type 1:1 and 2:1 associations. For higher-order associations, a general algorithm for numerically solving simultaneous nonlinear equations becomes more useful.

Experiments have been simulated for type 1:1 and 2:1 associations. It is convenient to transform the data to generate a Q function as defined by eq. 21. The Marquardt algorithm for nonlinear least-squares regression can then be used to obtain accurate estimates of the binding constant(s) if the correct model is used to fit Q as a function of the optical measurement. Use of the incorrect model leads to lack of convergence indicating that the method can be used to exclude models. This algorithm is the most widely employed method to fit data to arbitrary functions. This indicates that the Q function may have broad applicability to fit models of heterogeneous association.

Thus, the Q function is an empirical transformation of the data set to a form that is more amenable to least-squares regression. The transformation was suggested by the work of Morris and Ralston [11] who used a nonlinear regression method to fit reduced sedimentation equilibrium data to models of self-associating systems. They reduced data using the Ω function defined by Milthorpe et al. [3]. The experimental determina-

tion of the Q function is very similar to that of the Ω function. The Q function differs in that a reference concentration is not used and the denominator has two exponential terms with different amplitudes since two nonassociated species are possible.

The most general analysis of heterogeneous associations at sedimentation equilibrium has been performed by Tindall and Aune [12]. Their iterative procedure for parameter estimation is more complicated than the method used in this study, which may limit its usefulness. Recently, Lewis and Youle [8] described a least-squares regression method for reducing sedimentation equilibrium data to analyze type 1 : 1 associations restricted to equal starting concentrations of A and B. They were able to determine successfully the single equilibrium constant for the subunit interaction of ricin by nonlinear regression. Hensley et al. [6] also analyzed a type 1 : 1 association between cobra venom factor and human complement factor B using the modified Gauss-Newton nonlinear regression algorithm described by Johnson and Frasier [10]. They were able to obtain a satisfactory estimate for the dissociation constant of the interaction by simultaneously analyzing data collected at two rotor speeds. Attempts to fit data from a single rotor speed resulted in a high degree of correlation between the parameters which were the meniscus concentrations of the three possible species.

Systematic error in the determination of binding constants results from error in the fixed Q function parameters such as effective reduced molecular weight, initial concentration, or the optical constant. This error is not apparent from the visual inspection of fitted Q function curves. It also does not result in an identifiable trend in the estimated parameters as a function of loading concentration (table 1) or angular velocity. Since this error can be greater than the error due to instrumental noise, it must be taken into consideration.

The simulations described here proceed quickly using a microcomputer. Transformation to the Q function appears useful for data-fitting purposes at least up to certain type 2 : 1 associations. For higher-order associations, this becomes progressively more difficult due to increasing variation of σ_i and because of the possibility of large differences in the equilibrium constants. This problem conceivably could be overcome by simultaneously analyzing several starting concentrations and treating the starting concentrations as independent variables in the nonlinear regression analysis. This approach is currently being investigated.

Acknowledgements

This work was supported by a Grant-in-Aid from the American Heart Association (85-119) with funds contributed in part by the Vermont Affiliate, NIH grant HL01538, and the Vermont SCOR in Thrombosis (HL35058)

References

1 H. Fugita, Foundations of ultracentrifugal analysis (Wiley, New York, 1975).
2 M.L. Johnson, J.J. Correia, D.A. Yphantis and H.A. Halvorson, Biophys. J. 36 (1981) 575.
3 B.K. Milthorpe, P.D. Jeffrey and L.W. Nichol, Biophys. Chem. 3 (1975) 169.
4 H. Kim, R.C. Deonier and J.W. Williams, Chem. Rev. 77 (1977) 659.
5 E.E. Tucker, CABIOS 2 (1986) 159.
6 P. Hensley, M.C. O'Keefe, C.J. Spangler, J.C. Osborne and C. Vogel, J. Biol. Chem. 261 (1986) 11038.
7 P.R. Bevington, Data reduction and error analysis for the physical sciences (McGraw-Hill, New York, 1969).
8 M.S. Lewis and R.J. Youle, J. Biol. Chem. 261 (1986) 11571.
9 T.F. Holtzman, S.P. Leytus, J.O. Baldwin and W.F. Mangel, Anal. Biochem. 119 (1982) 62.
10 M.L. Johnson and S.G. Frasier, Methods Enzymol. 117 (1985) 301.
11 M. Morris and G.B. Ralston, Biophys. Chem. 23 (1985) 49.
12 S.H. Tindall and K.C. Aune, Anal. Biochem. 120 (1982) 71.