Biophysical Chemistry 23 (1985) 49–61
Elsevier

BPC 01024

DETERMINATION OF THE PARAMETERS OF SELF-ASSOCIATION BY DIRECT FITTING OF THE OMEGA FUNCTION

Michael MORRIS and G.B. RALSTON

Department of Biochemistry, University of Sydney, Sydney, NSW 2006, Australia

Received 14th April 1985
Revised manuscript received 1st July 1985
Accepted 15th July 1985

Key words: Sedimentation equilibrium; Omega function; Spectrin; Self-association; Nonlinear regression

Nonlinear regression is used to fit the omega function vs. protein concentration curves (first described by B.K. Milthorpe, P.D. Jeffrey and L.W. Nichol, Biophys. Chem. 3 (1975) 169) obtained from sedimentation equilibrium experiments on self-associating macromolecules. Nonlinear regression allows the direct fit of these curves with discrete or indefinite self-association reaction models in order to obtain values for the equilibrium constants and second virial coefficient. The method is independent of the choice of reference concentration and avoids the original method of extrapolating an omega function curve to zero concentration and then using the extrapolated value to construct a monomer activity curve used for analysis. This extrapolation can become very difficult for mild to strong self-associations where incorrectly extrapolated values lead to systematic error in the monomer activity curves. The method is applied to results from a mild, indefinite self-association, exemplified by the self-association of human spectrin, and to computer-simulated data of weak, mild and strong, indefinite self-associations.

1. Introduction

The omega function was introduced by Milthorpe et al. [1] as a means of fitting reaction models to sedimentation equilibrium data obtained with self-associating solutes. By extrapolating the omega function to zero concentration, the thermodynamic activity of the self-associating monomer can be determined at a reference point in the centrifuge cell. Milthorpe et al. [1] showed that the omega function was devoid of turning points, and that the required extrapolation should be, in many cases, straightforward. Once estimated, the activity of the monomer at the reference point can be used to calculate the activity of the monomer as a function of the total solute concentration. The activity of the monomer is a particularly useful quantity since, for a particular set of reaction conditions, it is a function only of the total solute concentration and can be written to include the parameters of self-association pertaining to a particular model [1]. The values of the parameters can be relatively easily determined by simultaneous solution of reaction model equations using the data from a number of arbitrarily chosen points [1].

One of the major features of this method is that the successive differentiation and integration of the data required by many other methods [2–4] are avoided, and that, as a result, less error is introduced into the calculation.

We have recently used the omega function analysis in a study of the self-association of the red cell cytoskeletal protein spectrin [5]. Although this approach appeared to be satisfactory, difficulties in extrapolating the omega function to zero concentration were encountered. It occurred to us that, since the omega function itself was dependent only on the total protein concentration, the parameters of the self-association model could, in

principle, be determined by directly fitting this function. This would avoid extrapolating the omega function curve and then fitting the activity curve derived from the extrapolated value.

This paper describes the direct fitting of the omega function by nonlinear regression to obtain the parameters of self-association models. Nonlinear regression uses all the data to determine the parameters, rather than selected sets of arbitrarily chosen points. The method is tested by fitting indefinite self-association models to computer simulated examples and to results from the self-association of spectrin.

2. Materials and methods

2.1. Preparation of spectrin

Spectrin dimer was extracted and purified from human erythrocytes as previously described [6] but with the inclusion of 0.3 mM sodium azide in all buffers. After repeated chromatography on a column of Sepharose 4B (3.0 × 50 cm) in a buffer comprising 0.1 M NaCl, 0.01 M phosphate, pH 7.5, 5 mM EDTA and 0.1 mM dithiothreitol, the dimer fraction was centrifuged for 40 min at 30000 × g to remove any insoluble material and was then dialysed four times (3 × 1 h, 1 × 6 h) against the same buffer plus 0.3 mM phenylmethylsulfonyl fluoride. The purified spectrin was used immediately for sedimentation equilibrium experiments to minimize the proteolytic damage that may occur during storage.

2.2. Meniscus depletion sedimentation equilibrium

Three different initial loading concentrations of spectrin were centrifuged at 25°C and 9038 rpm for 29.5 h in a Beckman-Spinco analytical ultracentrifuge fitted with electronic speed control and an RTIC unit. An An-D rotor and Yphantis 12 mm 6-channel centerpiece were used. The use of silicone layering oil was avoided in sedimentation equilibrium experiments [7]. At 29.5 h the Rayleigh interference pattern was photographically recorded on Kodak metallographic plates. The plates were measured on a Nikon comparator at 20 × magnification and the displacements were corrected for baseline error [8]. A concentration conversion factor of 4.04 fringes per 1 g/l was used [9]. The effective-monomer molecular weight, M_1, of spectrin was taken as 480000 [10], a partial specific volume, $\bar{v}_1$, of 0.733 ml/g was used [11] and the solution density, ρ, was 1.003 g/ml.

$\Omega(r)$ vs. $c(r)$ curves were calculated using eq. 1a (section 3). To determine if all three samples were homogeneous and had reached chemical equilibrium during the time course of the experiment, a reference concentration common to all three channels was chosen and the $\Omega(r)$ vs. $c(r)$ curves were examined for coincidence over the common concentration range [1]. The square of the radial position of the reference concentration in each channel was estimated by interpolation using a six-point quadratic or cubic regression.

The Adams-Fujita approximation [12] for the activity of the monomer, $a_1(r)$, was used when fitting the $\Omega(r)$ vs. $c(r)$ and the $a_1(r)$ vs. $c(r)$ curves:

$$a_1(r) = c_1(r)\exp[BM_1c(r)]$$

where $c_1(r)$ is the monomer concentration at radial distance r, and B the second virial coefficient. The equations describing $c_1(r)$ in the g/l scale for the SEK I and SEK III reaction models [3] were as follows: For the SEK I model (where the effective monomer adds to any sized polymer with the same change in free energy), $c_1(r)$ is given by [13]:

$$c_1(r) = \frac{1}{k} + \frac{1 - \sqrt{[1 + 4c(r)k]}}{2c(r)k^2} \qquad (kc_1(r) < 1)$$

where k is the intrinsic equilibrium constant. The molar intrinsic equilibrium constant, K, is given by:

$$K = kM_1$$

For the SEK III model, $c_1(r)$ is an implicit function of $c(r)$:

$$c(r) = c_1(r)\left\{1 + \frac{k_{12}c_1(r)[2 - kc_1(r)]}{[1 - kc_1(r)]^2}\right\} \qquad (kc_1(r) < 1)$$

where k_{12} is the dimerization constant and k the

intrinsic equilibrium constant describing all steps except dimerization. The corresponding molar equilibrium constants are given by:

$K_{12} = k_{12} M_1$

and

$K = kM_1$

Simulated curves of $\Omega(r)$ vs. $c(r)$ for the SEK I and SEK III reaction models were calculated using a combination of eq. 1b (section 3), the Adams-Fujita approximation and the appropriate equation for $c_1(r)$. For all simulations $M_1 = 480\,000$, $\bar{v}_1 = 0.73$ ml/g, $\rho = 1.0$ g/ml and $T = 298$ K. In certain cases the $c(r)$ values, in the raw data set of $c(r)$ vs. r, were perturbed by random, normally distributed errors of 0.0027 g/l standard deviation (corresponding to a standard deviation of approx. 3 μm in the baseline-corrected plate reading).

2.3. Nonlinear regression

Experimental and simulated $\Omega(r)$ vs. $c(r)$ and $a_1(r)$ vs. $c(r)$ curves were fitted by nonlinear regression based on the Gauss-Newton algorithm. Parameters were re-estimated for each iteration using the approximation by central differences of the partial first derivatives of the fitting function. Convergence to a final set of parameters was achieved from initial estimates on either side of the final values, with a tolerance of less than one part in 10^6.

For the SEK III model, the implicit function for $c_1(r)$ was solved by means of the Newton-Raphson iterative procedure to obtain $c_1(r)$ for any particular $c(r)$.

2.4. Electrophoresis

The purity of spectrin samples was examined using acrylamide gel electrophoresis in the presence of SDS according to the method of Fairbanks et al. [14]. Samples were found to be greater than 98% pure even at the completion of centrifugation. No traces of actin or band 4.1 could be detected in any of the samples.

3. Results

3.1. The omega function

The omega function, $\Omega(r)$, is defined as [1]:

$$\Omega(r) = \frac{c(r)\exp\left[\phi_1 M_1 (r_F^2 - r^2)\right]}{c(r_F)} \quad (1a)$$

$$= \frac{a_1(r_F)c(r)}{a_1(r)c(r_F)} \quad (1b)$$

where $\phi_1 = (1 - \bar{v}_1\rho)\omega^2/2RT$ with $\bar{v}_1$ the partial specific volume of the effective monomer, ρ the solution density, ω the angular velocity, R the universal gas constant and T the absolute temperature of the sedimentation equilibrium experiment. M_1 is the effective-monomer molecular weight, $c(r)$ and $c(r_F)$ the total protein concentrations at radial positions r and r_F, respectively, and a_1 the thermodynamic activity of the effective monomer. From eq. 1a [1], if M_1 and $\bar{v}_1$ are known and a reference concentration, $c(r_F)$, is chosen, a plot of $\Omega(r)$ vs. $c(r)$ can be drawn from measurements obtained with Rayleigh interference or absorption optics. By definition $\Omega(r) = 1$ at $c(r_F)$.

Importantly, plots of $\Omega(r)$ vs. $c(r)$ provide a very sensitive test for determining if the self-association reactions from a series of ultracentrifugal experiments have reached chemical equilibrium and if the samples are homogeneous. If chemical equilibrium has been achieved and the samples are homogeneous the plots will be coincident over their common concentration range when a common $c(r_F)$ is used (provided no volume changes accompany the formation of polymers) [1].

In order to fit reaction models to experimental results, Milthorpe et al. [1] suggested extrapolating the plot of $\Omega(r)$ vs. $c(r)$ to infinite dilution. From eq. 1b it can be shown that [1]:

$$\lim_{c(r)\to 0} \Omega(r) \equiv \Omega_0 = a_1(r_F)/c(r_F) \quad (2)$$

Since $c(r_F)$ is known, the monomer activity, $a_1(r_F)$, can be calculated. The value of $a_1(r_F)$ is then used to determine the monomer activity, $a_1(r)$, at any radial position and hence at any measured concentration, $c(r)$, using the following

equation [1]:

$$a_1(r) = a_1(r_F)\exp\left[\phi_1 M_1\left(r^2 - r_F^2\right)\right] \quad (3)$$

A plot of $a_1(r)$ vs. $c(r)$ can now be constructed and fitted with various reaction models. The values of the reaction model parameters can be determined by simultaneous solution of the reaction model equations using the data from a number of arbitrarily chosen points [1]. Alternatively, the entire data set could be used for the fitting procedure using nonlinear regression.

3.2. Extrapolation of the omega function

The accuracy of the analysis just described is highly dependent on the accuracy with which the $\Omega(r)$ vs. $c(r)$ curve can be extrapolated. For example, let us take an arbitrary set of data for which the reference concentration, $c(r_F)$, was measured without error, but for which the extrapolation to infinite dilution was in error by 2%. From eqs. 2 and 3, a 2% error in the extrapolated value, Ω_0, would lead to a 2% error in every calculated monomer activity, $a_1(r)$.

Milthorpe et al. [1] demonstrated that the extrapolation of an omega function curve to infinite dilution for a nearly ideal, weakly self-associating solute is relatively unambiguous. There are two reasons for this: Firstly, there is little or no curvature in the omega function near zero concentration. Secondly, if a small reference concentration is chosen, the intercept will have a value close to its maximum possible value of 1.0. In this way, if an error is made in the extrapolation, the percentage error is likely to be small and, therefore, the error in $a_1(r_F)$ and the other monomer activities will be small.

Fig. 1 shows a simulated omega function curve for a weakly self-associating solute. A good extrapolation can be made to within about 2% of the true value in this example. Thus, for a weakly self-associating solute, an accurate monomer activity curve can be drawn because the intercept can be determined accurately. Analysis of the monomer activity curve with any particular reaction model will give good estimates of the parameters.

As the strength of the self-association increases,

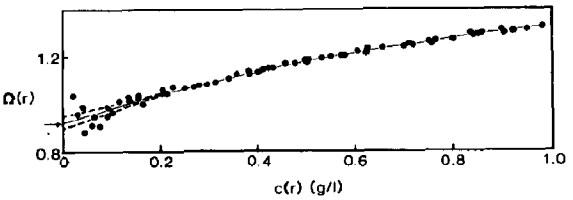

Fig. 1. Simulated omega function curve for a weakly self-associating solute. The solute is associating in accordance with the SEK III model ($k_{12} = 0.5$ l/g, $k = 0.25$ l/g, $B = 3.7\times10^{-7}$ l mol g^{-2}, $c(r_F) = 0.16557$ g/l). The solid line represents our best attempt to average the data by eye. The extrapolated value for this line lies very close to the correct extrapolated value (arrow). The dashed lines enclose the feasible range of extrapolations. The range is approx. $\pm 2\%$ of the correct value. Normalized, random error has been placed on the data.

the problems of extrapolating become more apparent. For mild self-associations (fig. 2) the omega function curve becomes steeper at low concentration and the curvature in this region becomes more uncertain. In addition, even if a small reference concentration is chosen, the intercept value will be substantially smaller than 1.0. In these cases, a small error in the extrapolation will result in a large percentage error in the intercept value and hence in the monomer activity values. This percentage error will tend to increase as the strength of the self-association increases.

Milthorpe et al. [1] suggest two means by which a poor extrapolation can be rectified: Firstly, an incorrect extrapolation can be detected because the monomer activity curve derived from the extrapolation will not pass through the origin. However, rearranging eq. 1b:

$$a_1(r) = \frac{a_1(r_F)c(r)}{c(r_F)\Omega(r)} \quad (4)$$

For a self-associating system (provided that the chosen monomer species exists), as $c(r)$ approaches zero, $\Omega(r)$ approaches some nonzero value so that:

$$\lim_{c(r)\to 0} a_1(r) = 0 \quad (5)$$

Therefore, a plot of $a_1(r)$ vs. $c(r)$ will pass through the origin irrespective of the error in $a_1(r_F)$ pro-

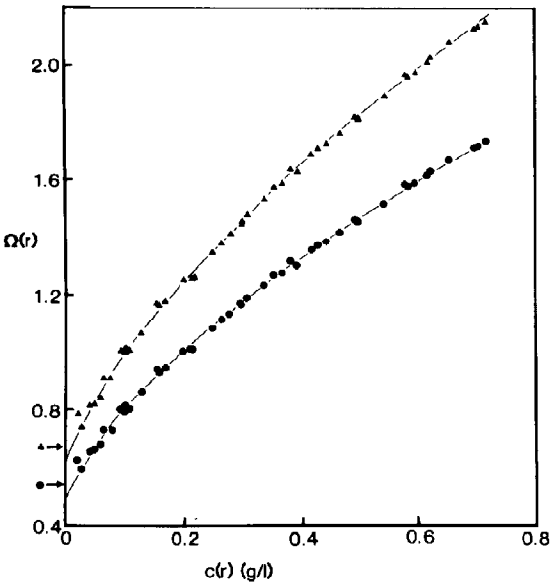

Fig. 2. Simulated omega function curves based on the SEK III model. Mild self-association ($k_{12} = 3$ l/g, $k = 2$ l/g, $B = 3.7 \times 10^{-7}$ l mol g^{-2}). The solid lines represent attempts to average the data in order to obtain the extrapolated values at zero concentration. A reasonable extrapolation for one curve (●, $c(r_F) = 0.2005$ g/l) has underestimated the correct extrapolation (●→) by about 9%. When the data are averaged in a similar manner for a curve based on a new $c(r_F)$ (▲, $c(r_F) = 0.10783$ g/l) a similar percentage error in the extrapolated value is obtained. Normalized, random error has been placed on the data.

duced by a poor extrapolation. Secondly, Milthorpe et al. suggest that an improved estimate of the reference monomer activity can be determined by averaging the values of $a_1(r_F)$ obtained from several curves based on different reference concentrations. However, this procedure does not necessarily lead to a better estimate of $a_1(r_F)$. For a particular set of data, the relative positions of the points in the omega function curves are similar irrespective of the reference concentration used. So, if a reasonable, but incorrect, extrapolation is made for one curve (fig. 2) then it is not unreasonable that the points used to guide the extrapolation will be treated in a similar manner for a curve based on a different reference concentration. The result is a tendency to propagate a systematic error in the evaluation of the intercepts and hence in the value of $a_1(r_F)$. Instead of drawing several omega function curves it would be sufficient to draw a single curve, using the lowest reference concentration available, and use a range of feasible extrapolations as a guide to the errors involved.

Fig. 3 shows the effect an incorrect extrapola-

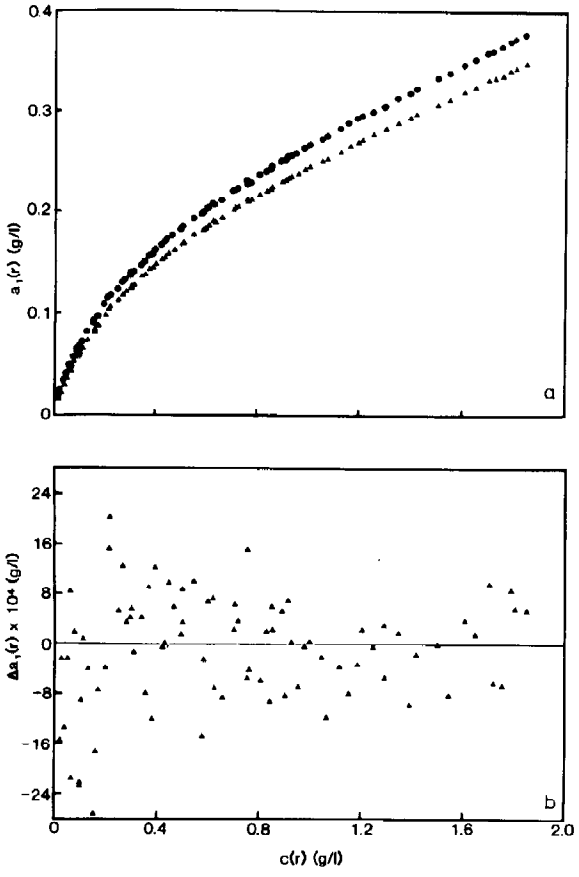

Fig. 3. (a) Monomer activity curves for a mild self-association based on the correct and incorrect extrapolated values in fig. 2 ($c(r_F) = 0.10783$ g/l). The solid circles are generated from the correct extrapolated value of 0.670 and the solid triangles from the incorrect value of 0.62. In (b) a nonlinear regression fit to the incorrect activity curve gave a nonrandom distribution of residuals and returned incorrect parameter values: $k_{12} = 4.00 \pm 0.05$ l/g, $k = 1.90 \pm 0.05$ l/g, $B = (3.28 \pm 0.11) \times 10^{-7}$ l mol g^{-2}. When the correct activity curve was fitted the distribution of residuals was random and the correct parameter values were returned: $k_{12} = 3.05 \pm 0.03$ l/g, $k = 2.03 \pm 0.04$ l/g, $B = (3.75 \pm 0.08) \times 10^{-7}$ l mol g^{-2}.

tion has on the monomer activity curve and the values of the self-association parameters. The correct and incorrect extrapolations for one of the simulated omega function curves in fig. 2 were used to construct the corresponding monomer activity curves (fig. 3a). The monomer activity curves then were fitted with the correct model, the SEK III model, using nonlinear regression.

The values of the returned parameters are inaccurate for the incorrect curve and the residuals are nonrandomly distributed (fig. 3b). If the correct curve is fitted, the correct parameter values are returned and the residuals are randomly distributed. For strong self-associations (data not shown), where reasonable extrapolations of the omega function curves can give very large percentage errors in the extrapolated values, the problems of nonrandom residuals distributions and inaccurate parameter estimates are accentuated.

Therefore, for mild and strong self-associations there is a possibility of rejecting the correct self-association model because of an error made in extrapolating the omega function curve to zero concentration.

3.3. *Direct fitting of the omega function*

In principle, the extrapolation and its associated problems can be avoided entirely by fitting the omega function directly, since the omega function, like the activity function, is dependent only on the total solute concentration and can be written in terms of the self-association parameters.

Using the Adams-Fujita approximation (see section 2) eq. 1b can be rewritten:

$$\Omega(r) = \frac{c(r)c_1(r_F)\exp\{BM_1[c(r_F)-c(r)]\}}{c(r_F)c_1(r)} \quad (6)$$

Various self-association models can be used to evaluate the right-hand side of eq. 6 for comparison with the experimental $\Omega(r)$ values, provided an implicit or explicit function of $c_1(r)$ in terms of $c(r)$ and the equilibrium constant(s) is available. The values of the reaction model parameters can be determined by nonlinear regression analysis of the omega function data as a function of $c(r)$.

However, with the equilibrium constants and the second virial coefficient as parameters the fitting function is forced to pass through the reference point, since from eq. 1a, $\Omega(r)=1$ at the reference point irrespective of the values of the parameters or the error made in measuring the reference concentration. This means that the values of the returned parameters and the distribution of the residuals of the fit are greatly dependent on the choice of the reference concentration.

This problem can be overcome by using several measured concentrations covering the concentration range (or, more appropriately, all of the measured concentrations) as reference concentrations. In this way a set of omega function curves is produced from the experimental data; each curve can then be fitted with the reaction model and the parameters calculated. A good fit will be indicated if the values for each parameter from the set of

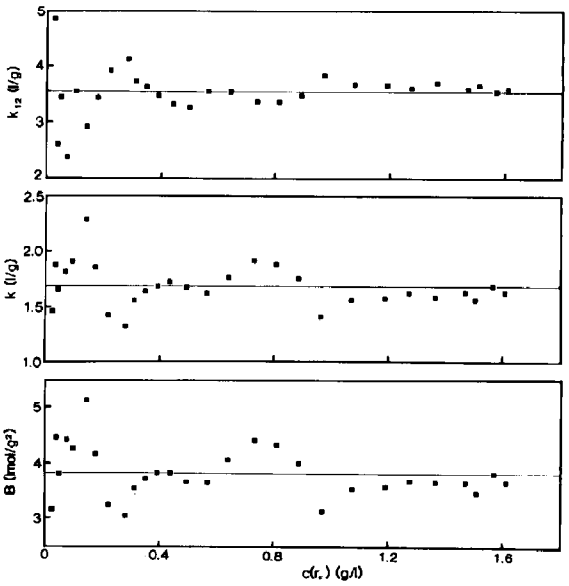

Fig. 4. The dependence of the fitting parameters (k_{12}, k and B) on the reference concentration, $c(r_F)$, for SEK III model fits to a set of omega function curves. The set of curves was obtained from a single channel of a sedimentation equilibrium experiment on spectrin. Each measured concentration was used, in turn, to construct an omega function curve which was then fitted with the model using nonlinear regression. The mean value for each parameter is given by the solid line in each panel. A two-sided t-test for each parameter shows no significant trend in the data at a probability level of 0.05.

curves are randomly distributed about a central value. That is, the correlation between the reference concentrations and the values of the parameter will be at, or close to, zero. Fig. 4 shows the results of such an analysis from a single channel of an experiment on spectrin. Each measured concentration has been used as a reference concentration to produce an omega function curve and each omega function curve has been fitted with the SEK III model. For each parameter, the correlation between the reference concentration and the values of the parameter is zero, within limits of error, indicating that the SEK III model is a good one.

The analysis becomes unwieldy, however, when one wishes to consider the data from several channels either separately or simultaneously. If the channels are considered separately then the analysis must be repeated for each channel. If the channels are considered simultaneously then there is the added problem that the same reference concentration must be used from each channel to produce each omega function curve in the set. In most cases a particular reference concentration will not be explicitly measured in each channel and the position of the reference concentration in the solution column must be determined by interpolation.

To avoid the laborious procedure of constructing a set of omega function curves and fitting each curve in the set, it would be useful if the nonlinear regression fit was not forced to pass through the reference point but, instead, the reference point was given the same weight as any other point.

3.4. The reference point

We shall first consider the effect that an error in the reference concentration has on the omega function, since, from eqs. 1, the reference concentration is used to calculate every omega function value, and our function fitting procedure is forced to pass through the reference point. For a set of data in the absence of errors:

$$\Omega(r) = \frac{c(r)\exp\left[\phi_1 M_1\left(r_F^2 - r^2\right)\right]}{c(r_F)} \tag{1a}$$

and the omega function curve intercept, Ω_0, is given by:

$$\Omega_0 \equiv \lim_{c(r)\to 0} \Omega(r) = a_1(r_F)/c(r_F) \tag{2}$$

Consider the case where the measurement of $c(r_F)$ is in error by an amount δ, so that:

$$c'(r_F) = c(r_F) + \delta \tag{7}$$

where the superscript indicates an apparent value. In this case all omega function values will be incorrect, being affected by the error in the reference value:

$$\Omega'(r) = \frac{c(r)\exp\left[\phi_1 M_1\left(r_F^2 - r^2\right)\right]}{c'(r_F)} \tag{8}$$

By combining eqs. 1a and 8 the apparent value of Ω at any radial position can be related to the true value:

$$\Omega'(r) = \frac{\Omega(r)c(r_F)}{c'(r_F)} \tag{9}$$

If, for example, δ is positive, the result is systematically to lower the entire set of Ω values except for the reference value, which, from eq. 8, is still fixed at 1.0.

Fig. 5 illustrates simulated omega function curves based on the SEK III reaction model. One curve is calculated from a set of perfect data and the other is flawed by a positive error in the reference concentration. An SEK III fit to the true curve passes through all the points and returns the correct parameter values and the correct intercept, Ω_0. The correct reference monomer activity, $a_1(r_F)$, can be obtained in the usual way by rearranging eq. 2:

$$a_1(r_F) = \Omega_0 c(r_F) \tag{10}$$

An SEK III fit to the flawed curve gives a poor fit and an intercept value very different from that obtained from the true curve. Even if the fit were not forced to pass through the reference point, the intercept for the flawed curve, Ω_0', would still be different from the intercept for the true curve, since, from eqs. 2 and 9:

$$\Omega_0' \equiv \lim_{c(r)\to 0} \Omega'(r) = \frac{\Omega_0 c(r_F)}{c'(r_F)} \tag{11}$$

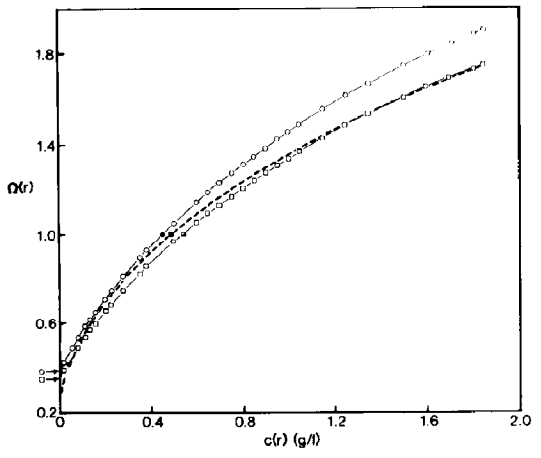

Fig. 5. Simulated omega function curves based on the SEK III model ($k_{12} = 3$ l/g, $k = 2$ l/g, $B = 3.7 \times 10^{-7}$ l mol g^{-2}). One curve (○) has been calculated in the absence of errors and the other curve (□) has been flawed by an error in the reference concentration. When the equilibrium constants and the second virial coefficient are used as fitting parameters, the SEK III fit to the unflawed data (solid line) passes through the center of all the points including the reference point (●, $c(r_F) = 0.45$ g/l), gives the correct intercept (○→) and returns the correct parameter values. The SEK III fit to the flawed curve (dashed line) passes through the center of the reference point (■, $c(r_F) = 0.49$ g/l) only and returns incorrect values: $k_{12} = 9.4 \pm 4.4$ l/g, $k = 0.6 \pm 1.1$ l/g, $B = (1.5 \pm 1.9) \times 10^{-7}$ l mol g^{-2}. The intercept for the fit to the flawed curve falls below the intercept value calculated from eq. 11 (□→). When $\hat{c}(r_F)$ is included as an additional fitting parameter (see section 3.5 for details) the fit (solid line) is no longer forced to pass through the reference point but passes through the more appropriate point determined by the value of $\hat{c}(r_F)$ (◪, $\hat{c}(r_F) = 0.542$ g/l). The correct parameter values are returned now and the intercept value is equal to that calculated from eq. 11. The reference points have not been included in the fits.

However, in spite of the fact that Ω'_0 and $c'(r_F)$ are in error in the flawed curve, the correct reference monomer activity can be obtained from eqs. 10 and 11:

$$\Omega'_0 c'(r_F) = \Omega_0 c(r_F) = a_1(r_F) \tag{12}$$

Note that the value of $a_1(r_F)$ is not appropriate to the measured reference concentration, $c'(r_F)$, but to the true reference concentration, $c(r_F)$. This means that, irrespective of the error in $c'(r_F)$, a good extrapolation of the flawed $\Omega'(r)$ vs. $c(r)$ curve to obtain the intercept will yield the correct value of the reference monomer activity at the appropriate radial position, r_F. (Once the value of $a_1(r_F)$ has been obtained, monomer activities at any other radial position can be calculated using eq. 3 without further reference to $c'(r_F)$. Therefore, in a plot of $a_1(r)$ vs. $c(r)$ there will be no propagation of systematic error caused by the use of $c'(r_F)$.) Although in any real experiment all data points will contain unknown errors, the above discussion shows that the determination of the activity of the monomer at the reference point need not be prejudiced by errors in the reference concentration.

3.5. Removing the privilege of the reference point

In most cases the appropriate intercept, Ω'_0, can only be obtained if the fitting procedure is not forced to pass through the reference point. This can be achieved by defining an additional parameter, $\hat{c}(r_F)$, such that $\Omega[\hat{c}(r_F)] = 1.0$.

Although $\hat{c}(r_F)$ has no physical meaning, when $c(r) = \hat{c}(r_F)$, $\Omega'(r) = 1.0$, so that, using eq. 1b:

$$\frac{a_1(r_F)\hat{c}(r_F)}{c'(r_F)\hat{a}_1(r_F)} = 1.0 \tag{13}$$

where $\hat{a}_1(r_F)$ is the monomer activity corresponding to $\hat{c}(r_F)$. Rearranging eq. 13 gives:

$$\frac{a_1(r_F)}{c'(r_F)} = \frac{\hat{a}_1(r_F)}{\hat{c}(r_F)} \tag{14}$$

Combining eqs. 12 and 14, the ratio $\hat{a}_1(r_F)/\hat{c}(r_F)$ gives us the appropriate intercept value, Ω'_0, and the true reference monomer activity, $a_1(r_F)$, for the flawed curve $\Omega'(r)$ vs. $c(r)$:

$$\frac{\hat{a}_1(r_F)}{\hat{c}(r_F)} = \frac{a_1(r_F)}{c'(r_F)} = \Omega'_0 \tag{15}$$

Therefore, with $\hat{c}(r_F)$ as an additional parameter, a nonlinear regression fit of the flawed $\Omega'(r)$ vs. $c(r)$ curve is no longer forced to pass through the reference point $[c'(r_F), 1.0]$. Instead, the fit will pass now through the more appropriate point $[\hat{c}(r_F), 1.0]$, and the appropriate intercept, Ω'_0, will be returned.

It remains to be determined if the same vector of reaction model parameters (the equilibrium constant(s) and the second virial coefficient) will be returned for both the true omega function curve and the flawed curve when the fit passes through the more appropriate point $[\hat{c}(r_F), 1.0]$. For a given vector of parameters and the true reference concentration, $c(r_F)$, the ratio $a_1(r_F)/c(r_F)$ is a constant, independent of $c(r)$:

$$\frac{a_1(r_F)}{c(r_F)} = \Omega_0 \tag{2}$$

Similarly, for a given vector of parameters and a fixed value of $\hat{c}(r_F)$, the ratio $\hat{a}_1(r_F)/\hat{c}(r_F)$ is also a constant, independent of $c(r)$ (eq. 15). We can compute expressions for $\Omega(r)$ and $\Omega'(r)$ based on the same vector of parameters by substituting the appropriate intercept into eq. 1b:

$$\Omega(r) = \frac{\Omega_0 c(r)}{a_1(r)} \tag{16}$$

$$\Omega'(r) = \frac{\Omega'_0 c(r)}{a_1(r)} \tag{17}$$

Since the concentration dependence of $\Omega(r)$ and $\Omega'(r)$ resides in the same term, $c(r)/a_1(r)$, and $c(r)$ and $a_1(r)$ are dependent on the reaction model parameters, the same vector of parameters can be used to describe both the $\Omega(r)$ vs. $c(r)$ curve and the $\Omega'(r)$ vs. $c(r)$ curve. Thus, we have a means of fitting omega function curves using nonlinear regression which is not forced to pass through the reference point and which is independent of any error in the reference concentration.

This can be illustrated (fig. 5) by fitting the SEK III omega function curve in which the reference concentration is in error. The fit is no longer forced to pass through the reference point but, instead, passes through the more appropriate point determined by the additional parameter, $\hat{c}(r_F)$. The correct values of the reaction model parameters are returned now. The intercept differs from the true intercept in accordance with eq. 11 and when the reference monomer activity is calculated, it equals the true value.

Importantly, function fitting by nonlinear regression to omega function curves can differentiate between reaction models. For example, an omega function curve drawn from SEK III data is fitted properly by the SEK III model but is not fitted well by an SEK I model (fig. 6a). On the other hand, an omega function curve drawn from SEK I data is fitted perfectly by an SEK III model, since the SEK I model is simply a limited SEK III case with $k_{12} = k$ (fig. 6b).

The advantage of the additional parameter, $\hat{c}(r_F)$, is exemplified by experimental results for the self-association of human spectrin. With only the equilibrium constants and second virial coeffi-

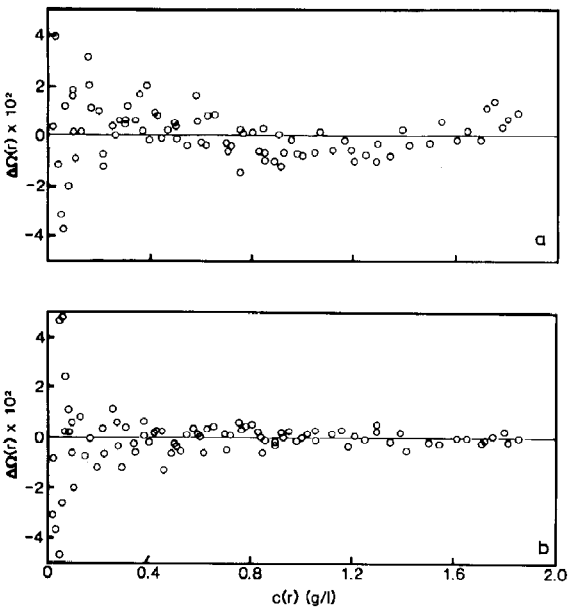

Fig. 6. (a) Residuals plot for an SEK I model fit to an omega function curve drawn from SEK III data ($k_{12} = 3$ l/g, $k = 2$ l/g, $B = 3.7\times10^{-7}$ l mol g^{-2}, $c(r_F) = 0.2005$ g/l). $\hat{c}(r_F)$ has been included as a fitting parameter. The returned parameter values for the SEK I model were: $k = 2.32\pm0.03$ l/g, $B = (4.56\pm0.06)\times10^{-7}$ l mol g^{-2}, $\hat{c}(r_F) = 0.2061\pm0.0010$ g/l. An SEK III model fit to the curve gave a random distribution of residuals and returned the correct values: $k_{12} = 3.05\pm0.14$ l/g, $k = 2.02\pm0.06$ l/g, $B = (3.75\pm0.15)\times10^{-7}$ l mol g^{-2}, $\hat{c}(r_F) = 0.2020\pm0.0010$ g/l. (b) Residuals plot for an SEK III model fit to an omega function curve drawn from SEK I data ($k = 2.3$ l/g, $B = 4.6\times10^{-7}$ l mol g^{-2}). $\hat{c}(r_F)$ has been included as a fitting parameter. The returned parameter values for the SEK III fit were $k_{12} = 2.32\pm0.17$ l/g, $k = 2.28\pm0.08$ l/g, $B = (4.55\pm0.23)\times10^{-7}$ l mol g^{-2}, $\hat{c}(r_F) = 0.219\pm0.002$ g/l. Normalized, random error has been placed on the data.

cient as parameters, many, and preferably all, measured concentrations should be used as reference concentrations in order to determine whether

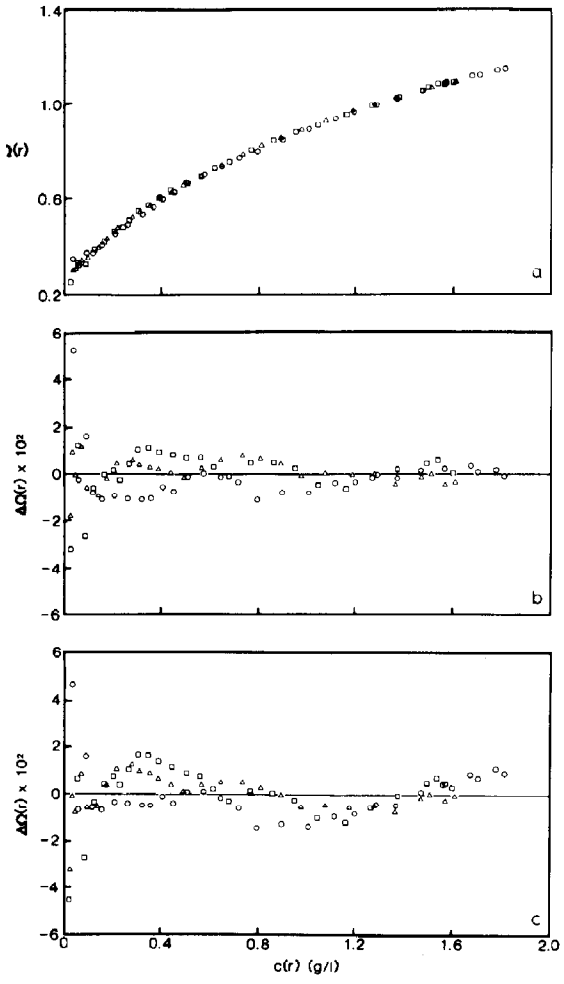

Fig. 7. (a) Omega function curves for a sedimentation equilibrium experiment performed at 25°C on three different initial loading concentrations of spectrin (1.0 g/l (○), 0.75 g/l (△) and 0.5 g/l (□)). $c(r_F)$ for each channel has been chosen at 1.28347 g/l. The three curves have been fitted simultaneously with (b) the SEK III model and (c) the SEK I model using $\hat{c}(r_F)$ as an additional parameter. The results in (b) and (c) have been plotted as the residuals to the fits. The returned parameter values for the SEK III model were: $k_{12} = 3.60 \pm 0.41$ l/g, $k = 1.64 \pm 0.19$ l/g, $B = (3.39 \pm 0.47) \times 10^{-7}$ l mol g^{-2}, $\hat{c}(r_F) = 1.278 \pm 0.006$ g/l. The returned values for the SEK I model were: $k = 2.27 \pm 0.07$ l/g, $B = (5.09 \pm 0.15) \times 10^{-7}$ l mol g^{-2}, $\hat{c}(r_F) = 1.265 \pm 0.006$ g/l.

the model is a good fit and, if so, to obtain good estimates of the parameters (fig. 4). With the additional parameter, $\hat{c}(r_F)$, only one reference concentration need be chosen in order to determine if the model is a good choice and to obtain good estimates of the parameters. Furthermore, the choice of reference concentration is arbitrary, since any selected concentration will yield the same residuals distribution and parameter estimates.

The method is easily extended to simultaneous analysis of results from several experiments using different operating conditions (different loading concentrations, angular velocities and radial positions) provided a common reference concentration is used. Fig. 7a displays omega function curves calculated from the results of three different initial loading concentrations of spectrin in a meniscus depletion experiment. The curves overlap closely indicating that all three solutions are homogeneous and have reached chemical equilibrium during the time course of the experiment. The curves have been fitted with the SEK I and SEK III reaction models. The values of the returned parameters are reasonable for both models in the light of published results on the self-association of spectrin [15] and its size and shape [16]. The residuals plot for the SEK III model (fig. 7b) shows a more random distribution of residuals than for the SEK I model (fig. 7c) indicating a better fit by the SEK III model to the data. Rejection of the SEK I model can be justified objectively by a criterion such as a runs test. While the residuals of fig. 7b appear to be distributed randomly by this criterion, those of fig. 7c show significant departure from a random distribution at the 5% level. Data from individual channels also show significant departure from random distribution with the SEK I model.

4. Discussion

The method we have outlined allows the determination of the parameters of self-association using the omega function without reliance on visual extrapolation, and without the use of an extrapolated value to construct an activity curve as a function of concentration. Although extrapolation of the omega function may be straightforward for

weakly self-associating solutes (fig. 1), the chance of introducing errors, and the relative size of such errors, increases with the strength of the association (fig. 2). Any error introduced in this way is propagated throughout the monomer activity curve, leading to inaccurate estimates of the reaction model parameters (fig. 3). In addition, the distortion of the distribution of residuals introduced by errors in extrapolation may lead to the rejection of the correct model for association.

In the method we have described, all of the data points are used to make maximum use of the information available. The difficulties of extrapolation and the construction of a monomer activity curve are avoided by fitting the omega function directly using nonlinear regression. This can be achieved in two ways: One method uses only the reaction model parameters as fitting parameters. For this method to be successful most, and preferably all, the measured concentrations should be used sequentially as reference concentrations to produce a set of omega function curves (fig. 4). However, this procedure is laborious, especially when data from several channels and several reaction models are to be tested. In addition, it is uncertain whether the values obtained for each parameter are normally distributed with respect to the error in the reference concentration. A much better way is to include the additional fitting parameter, $\hat{c}(r_F)$, to avoid forcing the fitted curve through a fixed point. Instead, the curve will pass through a more appropriate point determined by $\hat{c}(r_F)$ (fig. 5). With this method only one omega function curve needs to be fitted, since the fitting procedure is independent of the value of the reference concentration and independent of any measured error in the reference concentration.

The errors calculated for the parameters when fitting the omega function curves are more realistic than the errors calculated when fitting the monomer activity curves since the monomer activity function (eq. 3) is a smoothed function and gives unrealistically small errors. We have estimated our errors for fits to the omega function curves using standard statistical techniques [17]. These techniques do not strictly apply to nonideal self-association models since the parameters are not independent of each other and the fitting functions are not linear in form [18]. However, the errors calculated using the standard techniques should at least be of the right magnitude [18,19] and it should be possible to estimate the errors using more sophisticated statistical analysis if necessary.

It is advisable with mild and strong self-associations to collect extra data in the low concentration region since it is in this region that the proportion of effective monomer is highest and the first equilibrium constant can be calculated most accurately. If too few points are measured at low concentration the random error may appear to be systematic over a small concentration range and this will adversely affect the estimation of the first equilibrium constant. Since there is correlation between the parameters the values of all parameters will be affected. Possibly, this problem can be overcome by independently determining the value of the second virial coefficient [20]. In this way the second virial coefficient will be treated as a constant instead of a parameter whose value is estimated in the nonlinear regression calculation. This may force any other parameters to assume more realistic values.

Direct fitting of the omega function was used to test the self-association of human spectrin (fig. 7). Of the two models tested the SEK III model (fig. 7b) appeared to give a better fit to the data than the SEK I model (fig. 7c). It is possible that the slightly nonrandom distribution of residuals for the SEK I model results from correlation of the residuals due to the curve-fitting process and that the model may still be valid [21]. However, simulated data based on SEK III parameters (whose values were very close to those obtained for the experiments on spectrin) showed the same distribution of residuals as the experimental data when fitted with the two models (fig. 6). The close agreement between simulated and experimental residuals plots strongly suggests that the SEK I model does not describe the self-association of spectrin as well as the SEK III model over the concentration range 0–1.8 g/l.

Although we have presented the method for direct fitting of the omega function using indefinite self-associating macromolecules the principle can be applied to discrete self-associating systems

and self-associations of small molecules, including micelle formations.

Johnson et al. [18] have developed a nonlinear least-squares method that determines virial coefficients and self-association constants by fitting concentration as a function of radial position. This is the most direct method for determining parameter values for specific association schemes. In many cases, however, this method can present difficulties: The basic test for heterogeneity for samples centrifuged under different operating conditions (different initial loading concentrations, radial positions and angular velocities) is the coincidence of plots of apparent weight average molecular weight vs. concentration. The calculation of apparent weight average molecular weights is a procedure prone to systematic error due to a differentiation step [1,13] and therefore this test may indicate heterogeneity where none exists. Since the precision of the test is not as high as the precision of the concentration data used directly to calculate the parameters, it should be used only as a guide to the presence of heterogeneity.

The omega function analysis has a much more sensitive test for heterogeneity based on the overlap of the plots of the omega function vs. concentration [1]. This method avoids any differentiation of the data. In our experience, data from three separate channels plotted as apparent weight average molecular weight against concentration can sometimes indicate heterogeneity in samples due to the slight, apparent lack of overlap of the plots. Often, when the data have been replotted in terms of the omega function vs. concentration, the overlap of the three plots over the common concentration range has indicated that the samples are, in fact, homogeneous.

Johnson and co-workers [22] have also devised a more quantitative test for heterogeneity in self-associating systems based on the overlap of the confidence regions of the parameters of self-association. A separate regression analysis is performed on the data from each of a number of channels. If the confidence regions do not overlap the sample is taken to be heterogeneous. In general, since the correct mode of self-association is rarely known, an assumed (and possibly incorrect) model must be used to perform the test. In these cases the data sets from the separate channels must be truncated to a common concentration range before each channel is regressed separately [22]. The choice of an incorrect model and the subsequent truncation procedure present no problem provided the density of the data points in the truncated region for each channel is sufficiently high. A large number of data points can be achieved readily with automatic plate reading techniques [22,23]. When automatic plate reading is not available the number, and the density, of points measured is generally greatly reduced and the number of points in each truncated set may be small. If this is the case, then the requirement that the distribution of the data in each of the truncated sets be identical may no longer apply and the test cannot be used sensibly. In addition, with a small number of points, the random error may appear to be systematic over a small concentration range and this may give rise to large changes in the parameter estimates without greatly increasing the confidence regions [21].

The truncation test for heterogeneity is at present restricted to heterogeneity produced by non-associating, monomer-like species present in solution with solutes undergoing discrete self-associations [22]. The applicability of the test with respect to indefinite self-associations has yet to be determined.

When any one of the above limitations is not upheld the truncation method cannot be employed with confidence and users of the method of Johnson et al. [18] must then resort to plots of weight average molecular weight vs. concentration as a test for heterogeneity.

The method of Johnson et al. is restricted to the simultaneous analysis of results from each channel of meniscus depletion experiments or intermediate-speed experiments [8] performed at a single speed. Thus, data from low-speed experiments cannot be analysed. With direct fitting of the omega function, data from low-speed experiments can be incorporated into the analysis provided the concentration range for the low-speed experiments overlaps the concentration range for the meniscus depletion experiments and a common reference concentration is used. We are also preparing a paper in which a procedure is detailed for incorporating into the analysis the results of

low-speed experiments whose concentration range does not overlap that of the meniscus depletion (or intermediate-speed) experiments.

Acknowledgements

We wish to thank Dr. Simon Easterbrook-Smith for allowing us to use his nonlinear regression program. This work was supported by the Australian Research Grants Scheme.

References

1 B.K. Milthorpe, P.D. Jeffrey and L.W. Nichol, Biophys. Chem. 3 (1975) 169.
2 R.F. Steiner, Arch. Biochem. Biophys. 39 (1952) 333.
3 E.T. Adams, L.-H. Tang, J.L. Sarquis, G.H. Barlow and W.M. Norman, in: Physical aspects of protein interactions, ed. N. Catsimpoolas (Elsevier/North-Holland, Amsterdam, 1978) p. 1.
4 D.E. Roark and D.A. Yphantis, Ann. N.Y. Acad. Sci. 164 (1969) 245.
5 M. Morris and G.B. Ralston, Proc. Aust. Biochem. Soc. 16 (1984) 17.
6 G.B. Ralston and J.C. Dunbar, Biochim. Biophys. Acta 579 (1979) 20.
7 M. Morris and G.B. Ralston, Biochim. Biophys. Acta 788 (1984) 132.
8 D.C. Teller, Methods Enzymol. 27 (1973) 346.
9 J. Babul and E. Stellwagen, Anal. Biochem. 28 (1969) 216.
10 G. Ralston, J. Dunbar and M. White, Biochim. Biophys. Acta 491 (1977) 345.
11 Z. Kam, R. Josephs, H. Eisenberg and W.B. Gratzer, Biochemistry 16 (1977) 5568.
12 E.T. Adams and H. Fujita, in: Ultracentrifugal analysis in theory and experiment, ed. J.W. Williams (Academic, London, 1963) p. 119.
13 H. Kim, R.C. Deonier and J.W. Williams, Chem. Rev. 77 (1977) 659.
14 G. Fairbanks, T.L. Steck and D.F.H. Wallach, Biochemistry 10 (1971) 2606.
15 E. Ungewickell and W. Gratzer, Eur. J. Biochem. 88 (1978) 379.
16 J.C. Dunbar and G.B. Ralston, Biochim. Biophys. Acta 667 (1981) 177.
17 W.W. Cleland, Adv. Enzymol. 21 (1967) 1.
18 M.L. Johnson, J.J. Correia, D.A. Yphantis and H.R. Halvorson, Biophys. J. 36 (1981) 575.
19 M.S. Lewis and G.D. Knott, Biophys. Chem. 5 (1976) 171.
20 P.R. Wills, L.W. Nichol and R.J. Siezen, Biophys. Chem. 11 (1980) 71.
21 J.G. Reich, J. Winkler and I. Zinke, Stud. Biophys. 43 (1974) 77.
22 D.A. Yphantis, J.J. Correia, M.L. Johnson and G.-M. Wu, in: Physical aspects of protein interactions, ed. N. Catsimpoolas (Elsevier/North-Holland, Amsterdam, 1978) p. 275.
23 D.J. DeRosier, P. Munk and D.J. Cox, Anal. Biochem. 50 (1972) 139.