ARTICLE

Helmut Cölfen · Stephen E. Harding

MSTARA and *MSTARI*: interactive PC algorithms for simple, model independent evaluation of sedimentation equilibrium data

Abstract This paper describes a program available for PC's for the evaluation of molecular weights from sedimentation equilibrium. This program, in its two forms – MSTARA for absorption optical records and MSTARI for interference optical records – requires no prior assumption of the nature of the system (ideal, non-ideal, monodisperse, polydisperse, self-associating etc.) and takes into consideration the whole solute distribution (i.e. from solution meniscus to cell base) in the ultracentrifuge cell rather than just a selected data-set. MSTARA or MSTARI are therefore recommended as a first analysis programme of sedimentation equilibrium data coming off an absorption or interference based analytical ultracentrifuge. These programmes are therefore particularly well suited if heterogeneity (polydispersity or interaction phenomena) or non-ideality is suspected. Their use is demonstrated for a series of data-set types (ideal, non-ideal, polydisperse and self-associating). Although MSTARA and MSTARI are model independent, they provide the basis for more detailed analysis of interactions, polydisperse distributions or non-ideality via easy export of ASCII datafiles to model dependent routines.

Key words Analytical ultracentrifugation · Sedimentation equilibrium · MSTAR function · Molar mass determination · Polydispersity

Introduction

A user of sedimentation equilibrium analysis in the ultracentrifuge is usually seeking for himself the following types of basic or more advanced information. The types of *basic information* would be one or more of: 1. What is the molecular weight of my system?; 2. Does my sample contain species of different molecular weight?; 3. If my sample contains species of different molecular weight what is the average molecular weight of the system? After this, the user may be seeking answers to the following more *advanced questions:* 4. If there are different species present are they in chemical equilibrium (i.e. a reversible association) or is my system a polydisperse non-interacting mixture, or are both phenomena present?; 5. If my system is interacting, what is the interaction stoichiometry and strength (as measured e.g. by association constants, (K_a) or dissociation constants, (K_d) etc)?; 6. If my system is polydisperse, what is the molecular weight distribution?; 7. Am I safe to neglect the effects of non-ideality or should I take it into consideration?

With the renaissance of ultracentrifugation following H. Schachman's announcement in *Nature* (Schachman 1989) and R. Giebeler's description (Giebeler 1992) of a new commercially available ultracentrifuge with on-line uv-absorption optics {the "XL-A", Beckman Instruments, Palo Alto, USA} there has been considerable interest in software for both sedimentation velocity and equilibrium analysis. With the launch of a further Beckman ultracentrifuge {the "XL-I"} with integrated on-line absorption and Rayleigh Interference optics, (Furst 1997) interest has now been extended further from the analysis of absorption optical records of velocity and equilibrium distributions to those recorded by Interference optics.

Whereas advances in data capture and analysis of sedimentation velocity data – such as the Bridgman "g(s)" procedure – have been covered elsewhere (Stafford unpublished i.e. 1992, 1997) we examine here software for *sedimentation equilibrium* analysis, and in particular, the "basic information" points 1–3 above, viz.: *measuring the (weight average) molecular weight M_w and answering the question as to whether or not the system is homogeneous or a heterogeneous mixture of interacting or non-interacting components.* There are several algorithms, both commercially available (from Beckman Instruments) and from ultracentrifuge workers (made available through the

H. Cölfen (✉)
Max-Planck Institute for Colloid and Interface Research, Colloid Chemistry, Kantstrasse 55, D-14513 Teltow, Germany

S. E. Harding
National Centre for Macromolecular Hydrodynamics, University of Nottingham, Sutton Bonington, LE12 5RD, UK

RASMB database) for the analysis of sedimentation equilibrium optical records based on models for a macromolecular system. Examples include (i) IDEAL1 (McRorie and Voelker 1993): works out M_w for a thermodynamically ideal single solute system (ii) IDEAL2 (McRorie and Voelker 1993): works out the molecular weights for 2 ideal non-interacting solutes in a mixture (iii) NONIDEAL (McRorie and Voelker 1993): allows the determination of the molecular weight as well as of the second osmotic virial coefficient for a nonideal single solute system (iv) ASSOC4 (McRorie and Voelker 1993): works out association constants for a thermodynamically ideal associating system of user specified stoichiometry (v) NONLIN (Johnson et al. 1981): works out association constants for a nonideal associating system (vi) RINDE and RINDENON works out molecular weight distributions of a discrete type (non-interacting monomers, dimers etc.) for both ideal and non-ideal systems respectively; (vii) XLASE (Lechner 1992) works out weight (as well as estimates for z- and number) average molecular weights and the molecular weight distribution for a polydisperse system for a user specified type (Schulz-Zimm, Gaussian, Log-normal etc.).

All these routines (apart from the first part of XLASE) assume a model for the system: single solute, ideal, non-ideal, monomer-dimer, Gaussian distribution etc. and hence the results of such fitting procedures can only be meaningful *if the model has been chosen correctly:* the need for a reliable, user-friendly universal evaluation procedure for sedimentation equilibrium experiments is thus clearly indicated.

Molecular weights for unknown systems

Beckman Instruments own IDEAL1 (McRorie and Voelker 1993) is popularly used for giving the molecular weight of an "unknown" system, by fitting a chosen data-set in the concentration distribution according to:

$$c(r) = c(r_F) \cdot \exp\{k \cdot M_{w,app}(r^2 - r_F^2)\} \tag{1}$$

where $k = \dfrac{(1 - \bar{v}\rho)\omega^2}{2RT}$, $\bar{v}$ = partial specific volume of the solute, ρ = solvent density, ω = angular velocity, R = gas constant, T = thermodynamic temperature, c = polymer concentration, r = radial distance from the center of rotation, and r_F = reference radius (commonly taken as the radial position of the meniscus, "r_a"). The "app" signifies apparent weight average molecular weight (i.e. prior to any correction for thermodynamic non-ideality, if necessary). $M_{w,app} = M_w$, the true molecular weight only in the case of a thermodynamically ideal system, i.e. where exclusion volume and charge effects = 0. The main disadvantage of IDEAL1 is however that it records the average over the species present in only a chosen part of the ultracentrifuge cell, *not over the whole contents of the cell* which may not reflect the initial loading distribution. This approximation is usually sufficient for fairly simple systems (one compo-

nent, or two or more components, provided they do not differ too much in mass) but it can lead to serious underestimates for heterogeneous systems in which the molecular weights of the various components differ considerably (although a useful check for this is to examine any dependence of M_w measured in this way with rotor speed (Laue 1992a)): $\ln c(r)$ vs r^2 plots can be strongly curved which makes it difficult to determine the whole cell weight average molecular weight, M_w, particularly if the regions near the cell base are not well defined.

XLASE (1st part) considers the whole of the (macromolecular) solute distribution in the ultracentrifuge cell and uses the classical alternative form of Eq. (1) for M_w and a related form for estimating M_z (Creeth and Pain 1967):

$$M_{w,app.} = \frac{(c_b - c_a)}{k\, c_0\, (r_b^2 - r_a^2)} \tag{2}$$

$$M_{z,app.} = \frac{1}{k(c_b - c_a)}\left[\frac{1}{r_b}\left(\frac{dc(r)}{dr}\right)_b - \frac{1}{r_a}\left(\frac{dc(r)}{dr}\right)_a\right] \tag{3}$$

where c_0 = initial polymer concentration and r = radial distance from the center of rotation with the indices b = cell bottom and a = meniscus. The concentrations c(r) can be represented in terms of equivalent absorbance units, A(r) or interference fringe displacement units, J(r).

Use of Eqs. (2) {or 3} again can be problematical for heterogeneous systems where difficult extrapolations to get c_b and c_a may be required: determination of $M_{w,app.}$ and $M_{z,app.}$ using the concentration difference between meniscus c_a and cell bottom c_b (see Eqs. (2) and (3)) can be wrong if c_a and especially c_b can only be estimated with poor precision.

Especially for complicated heterogeneous systems such as polysaccharides and glycoconjugates this can be a serious problem: this inspired the derivation of the M^* procedure (Creeth and Harding 1982) which was shown to have particularly advantageous features. The M^*, an operational *point average* (i.e. a function of radial distance at a point in the ultracentrifuge cell from the axis of rotation) molecular weight is defined by

$$M^*(r) = \frac{c(r) - c_a}{k\, c_a(r^2 - r_a^2) + 2k \int_{r_a}^{r} r[c(r) - c_a]\, dr} \tag{4}$$

The only parameter in Eq. (4) which requires extrapolation is c_a: this extrapolation can be achieved without difficulty with absorption optical records by a simple linear extrapolation to the meniscus because usually the concentration gradient is flat in the meniscus region. Since Rayleigh interference optical records give directly only the solute *concentration relative to the meniscus*, a more subtle extraction procedure for c_a is required, unless the so-called meniscus depletion method is followed (Wales et al. 1951; Yphantis 1964). Although interference records therefore require more care in extraction of c_a, such optical records usually give considerably smoother concentration distributions.

In Eq. (4), no differentiation process – which is sensitive to noise in the data – is involved. The M* function was shown in the original Creeth and Harding (1982) paper to have a number of useful identities, the most important of which is that at the cell bottom (b):

$$M^*(r=b) = M_{w,app.} \quad (5)$$

i.e. the apparent weight average molecular weight over the whole distribution of solute in the cell, from the solution meniscus (r = a) to the cell base (r = b). Other useful properties of M* are (Creeth and Harding 1982):

$$M^*(r=a) = M_{w,app.}(r=a) \quad (6)$$

i.e. the point weight average molecular weight at the meniscus, and

$$M^*(c \to 0) = M_{n,app.}(r=a) \quad (7)$$

i.e. the point number average molecular weight at the meniscus. The most useful identity is Eq. (5), because it provides a basis for obtaining $M_{w,app.}$ without involving a concentration extrapolation to the cell base (but rather a less severe extrapolation of M* to the cell base, which is less sensitive to error). The extrapolation of M*(r) to the cell base is usually simple and can be done by applying an extrapolation line in most cases. This extrapolation is possible even if the raw data set is of poor quality, a feature not uncommon with absorption optical data records.

The earliest precursor of the current interactive MSTARA and MSTARI programs for PC was an algorithm set up for a Wang desk top calculator by J. M. Creeth; this was subsequently translated and extended into a FORTRAN version for a mainframe computer written by one of us (Harding et al. 1992). Not only did these algorithms provide for the evaluation of $M_{w,app.}$ for the whole solute distribution in the cell using the M* procedure (Eq. [5]) but they also provided for the evaluation of point weight average molecular weights $M_{w,app.}(r)$ as a function of local radial position r in the ultracentrifuge cell via the relation

$$M_{w,app.}(r) = \frac{1}{k}\left[\frac{d \ln c(r)}{d(r^2)}\right] \quad (8)$$

and also point z-average molecular weights (Creeth and Pain 1967) $M_{z,app.}(r)$:

$$M_{z,app.}(r) = M_{w,app.}(r) + \frac{1}{k}\left[\frac{d \ln M_{w,app.}(r)}{d(r^2)}\right]$$
$$= \frac{1}{k}\frac{d}{d(r^2)}\ln\left(\frac{1}{r}\frac{dc(r)}{dr}\right) \quad (9)$$

Further, the evaluation of an "ideal" compound average, $M_{y2}(r)$, was provided for, whereby first order effects of thermodynamic non-ideality (and associative phenomena) are eliminated by combining $M_{w,app}(r)$ with $M_{z,app}(r)$:

$$M_{y2}(r) = M_{w,app}^2(r)/M_{z,app}(r) \quad (10)$$

$M_{w,app}(r)$, $M_{z,app}(r)$ and $M_{y2}(r)$ could be plotted either as a function of r, or an equivalent normalised parameter in r^2 (normalised so that it has a value of 0 at the solution meniscus and a value of 1 at the cell base):

$$\xi = (r^2 - r_a^2)/(r_b^2 - r_a^2) \quad (11)$$

or as a function of the corresponding local concentrations c(r).

This FORTRAN version still however had some undesirable features:

– extrapolation of M*(r) to the cell bottom was performed "graphically" by eye
– there was no possible axis scaling by the user permitting to zoom into interesting plot regions and thus performing more accurate extrapolations,
– no possibility to interactively re-set the sliding strip length which sets the smoothing factor for the determination of $M_{w,app.}$ via the differentiation of the ln c vs. r^2 plot (Eq. (8)), a plot which is very sensitive to noise. The same of course applies to the second derivative necessary for the determination of $M_{z,app.}(r)$ via Eq. (9),
– in the particular case of MSTARI, which requires evaluation of the meniscus concentration c_a (or in fringe unit terms, J_a) sometimes very noisy local $J_{a,app.}$ values were accrued, depending on the strip width of the sliding strip procedures used. On several occasions it proved to be useful to input an independently determined J_a value, which can be determined by many alternative procedures as summarized in Creeth and Pain (1967),
– speed of output and thus evaluation was not sufficient for the large amounts of generated data from the short column equilibrium technique in multichannel centerpieces (see, e.g., Laue 1992a),
– the ability to export data to other model dependent algorithms for further analysis was very cumbersome.

For those reasons and our general experiences within the National Centre for Macromolecular Hydrodynamics, it was decided to upgrade the MSTAR programs to provide an interactive base for general ultracentrifuge users on a widespread PC basis rather than for the dwindling community with mainframe computer experience and to facilitate rapid export and import of data: the aim has been to set up both MSTARA and MSTARI so that they are interactive in all basic steps to allow as much user input as possible. We now describe the essentials of the new versions of MSTARA and MSTARI which are available as public domain software (see end of this text).

The MSTAR programs

There exist two different versions of the MSTAR program due to the differences in the optical detection of the concentration gradient. Both versions have in common that they were designed as an open system, meaning that the data pairs of each plot set up during the evaluation are saved as an ASCII file. These files can then be imported into any kind of further evaluation software, spreadsheets or plot routines or whatever the user desires. Hence, the MSTAR

programs can serve as a platform for more sophisticated analysis. For further details on the original versions the reader is referred to Harding et al. (1992).

MSTARA

This is the first program version and is designed for the evaluation of the sedimentation equilibrium concentration distribution recorded using uv-absorption optics according to the classical Lambert-Beer law.

As there might still be absorbing impurities (even after thorough dialysis) which are not able to sediment, it is always desirable to determine the baseline absorption by overspeeding the centrifuge (if possible) to a speed where all solute has sedimented towards the cell bottom (this is the same as the "baseline offset" E in IDEAL1, IDEAL2, NONIDEAL or ASSOC4). In the earliest stage of the program the user is asked to enter this value which is then subtracted from the solute equilibrium gradient to give the true absorbance gradient. In some cases, the baseline absorbance might even be as high as 0.3–0.4. Normally it is reasonable to assume that this is constant throughout the solution, but occasionally care has to be expressed if a gradient of absorbing solvent components is expected. If the equilibrium speed is high anyway, the overspeed method will not yield a baseline value. In this case, thorough dialysis or an equivalent procedure using a chromatography column is an absolute pre-requisite to ensure the non-macromolecular absorbing components in solvent and solution channels in an ultracentrifuge cell are exactly matched.

The data input stage therefore requires the baseline absorbance (if any) as well as the standard parameters of rotor speed, temperature, solvent density, partial specific volume and the radial position at the base of the cell.

After the data input stage which consists of the user input of the necessary experimental parameters and system constants appearing in Eq. (4), the data file from the Beckman Optima XL-A or any other absorption A(r) vs. r data file in such ASCII format is read in; this will include the radial position of the meniscus. In the case of noisy data, it may sometimes happen that local absorbance values are beyond the baseline and would therefore give physically meaningless negative absorbances after the baseline correction. To avoid this, the first user interaction is required. The MSTARA program displays the absorbance vs. radial distance plot with the selected baseline and hence gives the user the chance to check, if the determined baseline is realistic. If not, the user can input a more realistic baseline. If the baseline is set as correct by the user and there still should be some data points which should give negative absorbance values after the baseline correction, these individual corrected absorbances are set to 0. Practice has shown that such procedure does not influence the results of the molecular weight determinations as long as the baseline absorbance is set correctly. We would like to stress the importance of a baseline correction since wrong absorbance and thus concentration gradients will lead to

erroneous results, whatever the evaluation technique for sedimentation equilibrium will be. Hence it is important to maintain a user input here (see part a of Fig. 1). The next section of the programme ("b" in Fig. 1) predicts the cell loading concentration A_0 in absorption units using the equation (see, e.g. Fujita 1975):

$$A_0 = \frac{2}{r_b^2 - r_a^2} \int_{r_a}^{r_b} A(r)\, r\, dr \tag{12}$$

A_0 is not needed in the $M^*(r)$ evaluation but is useful as a check on the macromolecular component of the injected concentration (supramolecular aggregates would be lost from optical registration). The concentration at the meniscus A_a is evaluated (step c) in Fig. 1)) using freely defineable data ranges in the absorbance vs. r plot and performing an extrapolation to the meniscus using regression polynomials up to degree 15 (usually, degree 1 is sufficient). This subroutine, common to other sections of the pro-
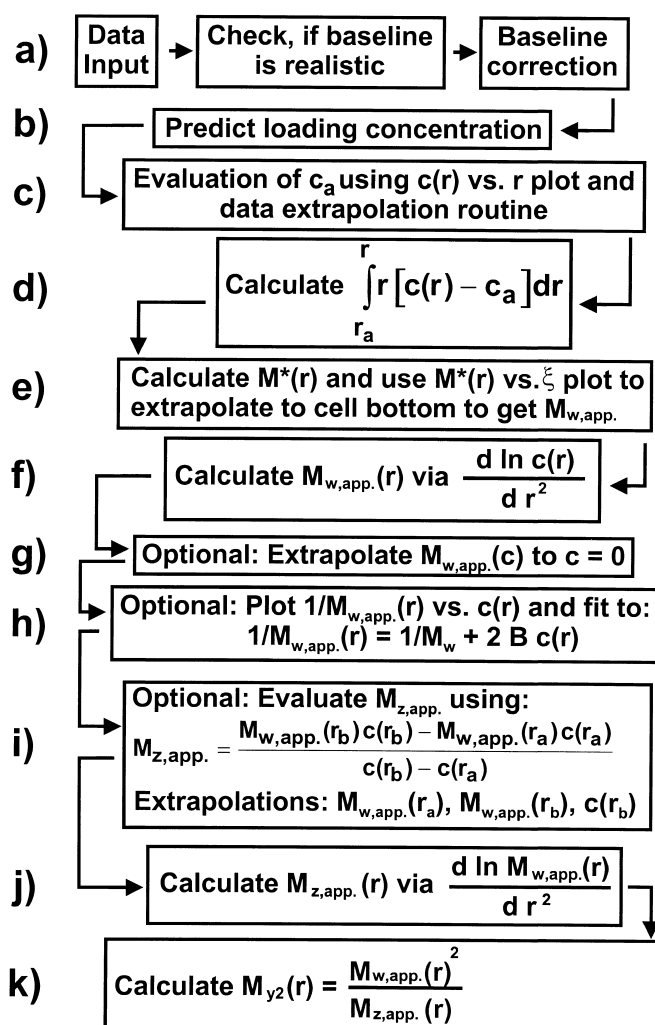

Fig. 1 Block diagram of the basic modules used in the MSTARA program. The indices (a–k) indicate the program steps referred to in the text. Note that the concentration in MSTARA is expressed in absorption units

gramme, is called the *"extrapolation routine"* in all that follows. Although it is known that the concentration gradient is exponential, no exponential regression was used because it is *not* the aim of the extrapolation routine to fit the whole concentration gradient but just a part thereof which is selected by the user. This makes the extrapolation module more universal for all extrapolations in MSTAR. Experience has shown, that in most cases, a linear regression line is sufficient to perform this and the other extrapolations in MSTAR. Polynomials with degrees higher than 3 are usually not necessary and can mislead the unwary.

The benefits of the user interaction are illustrated in Fig. 2. Here a dataset was evaluated which has a minor peak in the concentration gradient which is clearly an artefact. Two ranges are selected by the user which are then used for the data extrapolation as can be seen in Fig. 2. Curve fitting of the entire concentration gradient would probably lead to erroneous results due to the peak.

After evaluation of A_a the integral in Eq. (4) (see step d in Fig. 1) is evaluated using the trapezoid rule. The original FORTRAN mainframe routine used a more advanced numerical quadrature package but comparison of the performance of this with the simpler and computationally more economical trapezoid procedure showed no significant loss in precision or accuracy. After the evaluation of the local values of these integrals mentioned above, $M^*(r)$ is calculated using Eq. (4) (see step e) in Fig. 1).

Plotting $M^*(r)$ vs. ξ permits the determination of $M_{w,app.}$ using the data extrapolation routine for the extrapolation to the cell bottom. It should be noted here that one characteristic of the M^* function is that it starts very noisy near the meniscus (because of small concentration increments) but then gets more precise towards the cell bottom and thus allowing good determinations of $M_{w,app.}$. The use of M^* in this way can be thought of as an accumulator starting at the cell meniscus steadily homing in on the true $M_{w,app}$ as all the information about the solute distribution is gathered by the time the cell base is reached. Performing several extrapolations – easy with the interactive routine – the user gets a quite realistic impression of the error range of the extrapolations performed.

$M_{w,app.}(r)$ values are calculated using the classical method (step f of Fig. 1) via the local slopes in the $\ln A(r)$ vs. ξ plot according to Eq. (8). To perform the differentiation step, a so called sliding strip procedure is applied (Teller 1973). This is a regression of a data range left and right to the data point under consideration where the data point number at each side is defined as the sliding strip. Usually sliding strips of at least 10 are applied which yield reasonable but not excessive data smoothing. The regression function can be chosen as a straight line or a parabola and the sliding strip length is freely definable by the user. MSTARA uses a default value sliding strip length based on the data point number (data point number/10) which the user has an opportunity to change.

As a better diagnostic of the existence of self-association, polydispersity or non-ideality, the $M_{w,app.}(r)$ values are not only plotted vs. ξ but also vs. $A(r)$. As optional modules which are especially useful for nonideal, non-associating systems, MSTAR offers the possibility to extrapolate the $M_{w,app.}(r)$ {and/or $1/M_{w,app.}(r)$} vs. $c(r)$ (here $A(r)$) to zero concentration which can give an estimate of the true weight average molar mass M_w and the 2nd thermodynamic virial coefficient B via the relation:

$$1/M_{w,app.}(r) = (1/M_w) + 2\,B\,c(r) \qquad (13)$$

For analyses of more severe types of non-ideality the data have to be exported to a more specialised routine. At this stage also the user can examine his plots of $\ln c(r)$ versus r^2 (or ξ), and $M_{w,app.}(r)$ versus ξ or $c(r)$ and decide if the data is worth exporting to model dependent packages such as, for the analysis of self-associations, ASSOC4, NONLIN or, as a bolt-on end module to MSTAR, routines called OMEGA & PSI (which will be described elsewhere) or, for the analysis of polydisperse distributions, XLASE, RINDE or RINDENON. Prior export to a simple plotting package such as ORIGIN of several data-sets of $M_{w,app.}$ versus $c(r)$ for different loading concentrations c_a should also be useful for assaying for distinguishing self-association from non-interacting polydispersity phenomena (McRorie and Voelker 1993).

In MSTARA, the user also has the option of evaluating $M_{z,app.}(r)$ and $M_{y2}(r)$ as a function of ξ or $c(r)$ via a further sliding strip differentiation process (Eqs. (9), (10)): this however is even more sensitive to experimental noise than $M_{w,app.}(r)$ (see steps j, k) in Fig. 1). In general, the quality of absorbance data from the XL-A ultracentrifuge is not high enough to yield realistic estimates of either $M_{z,app.}(r)$ or $M_{y2}(r)$: the option of evaluating these functions is more appropriate for MSTARI.

$M_{z,app.}$ *for the whole distribution* (as opposed to local point averages) can however be realistically achieved using the following equation:

Fig. 2 Demonstration of the user interaction selecting appropriate data ranges for extrapolation processes

$$M_{z,app.} = \frac{M_{w,app.}(b)c_b - M_{w,app.}(a)c_a}{c_b - c_a} \qquad (14)$$

and the user is given this option. This still requires however a difficult extrapolation of the radial concentration gradient to the cell bottom and the extrapolation of $M_{w, app.}(r)$ to meniscus and cell bottom: in cases when these extrapolations are too difficult, the user is recommended resorting to a Schlieren optics based detection system which yields $M_{z, app.}$ data directly (see Creeth and Pain 1967).

MSTARI

MSTARI is similar in principle to MSTARA with the exception of step a in Fig. 1. This is caused by the difference between absorption and Rayleigh interference patterns, the first being directly proportional to the solute concentration (between certain limits), the latter reflecting *local* concentrations relative to that at the meniscus c_a, or, in terms of fringe displacement units, J_a. Thus it is of fundamental importance for the correct evaluation of Rayleigh interference patterns to determine J_a correctly. We would advise users to be very careful before considering using a method which avoids the need for measurement of J_a. This is the "meniscus depletion" or "high-speed" method (Wales et al. 1951; Yphantis 1964) in which the equilibrium distribution is run at a speed high enough that $J_a \to 0$: this method however can only be applied if a series of conditions are met (Yphantis 1964) and is generally unsuitable for heterogeneous systems because of the risk of losing a proportionally higher part of the higher molecular weight species from optical registration at the cell base.

MSTARI takes an input ASCII file of a data-set of relative fringe displacement $j(r)$ (relative to the meniscus, $r = a$) as a function of radial displacement r: this data file can either be obtained off-line via a photographic stage (Harding and Rowe 1989; Rowe et al. 1992) or directly online from CCD-array captured data (Laue 1992b; Furst 1997). {Any physically meaningless negative fringe shifts $j(r)$ deriving from noise near the meniscus region are set to zero}.

Blank correction. A "blank" correction (somewhat analagous to the baseline correction for MSTARA) of the ASCII $j(r)$ versus r data is sometimes necessary – particularly at higher equilibrium speeds – because of possible window distortions affecting the interference fringes (Teller 1973). This correction is probably best made by recording the corresponding interference pattern with just water in both solvent and solution channels (making sure the cell is torqued to the same stress) at the same equilibrium speed. The resulting $j(blank(r))$ versus r data set is then subtracted from the original $j(r)$ versus r to give a corrected data-set. This of course is not only for MSTARI but any interference based sedimentation equilibrium analysis algorithm. In practice we find, along with Teller (1973) that these corrections are only very small, especially at low rotor speed.

The next stage is to determine J_a from which the absolute fringe displacements $J(r)$ (directly proportional to c(r)) can be obtained from

$$J(r) = j(r) + J_a \qquad (15)$$

There exist a whole variety of methods to deterine J_a as reviewed by Creeth and Pain (1967). The user has the option of determining J_a using a separate method (Creeth and Pain 1967) or he/she can use a mathematical extraction procedure based on a plot of

$$\frac{j(r)}{(r^2 - r_a^2)} \quad \text{vs.} \quad \frac{I(r)}{(r^2 - r_a^2)} \qquad (16)$$

where $I(r)$ is the integral defined in Fig. 3. $J_a = 2*$ intercept (at the meniscus)/limiting slope at the meniscus (Creeth and Harding 1982). This can be obtained either by directly obtaining a linear fit from user specified limits in this plot or, since the limiting regions near the meniscus are also more noisy due to the lower fringe increments, by evaluating local $J_{a, app.}$ from local regions at points away from the meniscus (using a sliding strip procedure) to give an apparent $J_{a, app}$ and extrapolating this back to the meniscus (as described in Harding et al. 1992). MSTARI sets the sliding strip lengths according to the data point number dividing it through 20, 10, 6 and 4 and rounding to the nearest integer value.

At this stage MSTARI requires the users chosen value of J_a; the $j(r)$ versus r dataset is then converted to one of $J(r)$ versus r (or ξ) via Eq. (15), and the rest of the evaluation is essentially the same as MSTARA (points b–k) in Fig. 1, with the exception that $c(r)$ is now represented by $J(r)$ values instead of $A(r)$.

The other difference with MSTARA is that, since Rayleigh interference optical records give generally more accurate records of $c(r)$ versus r, the extraction of $M_{z, app.}(r)$,
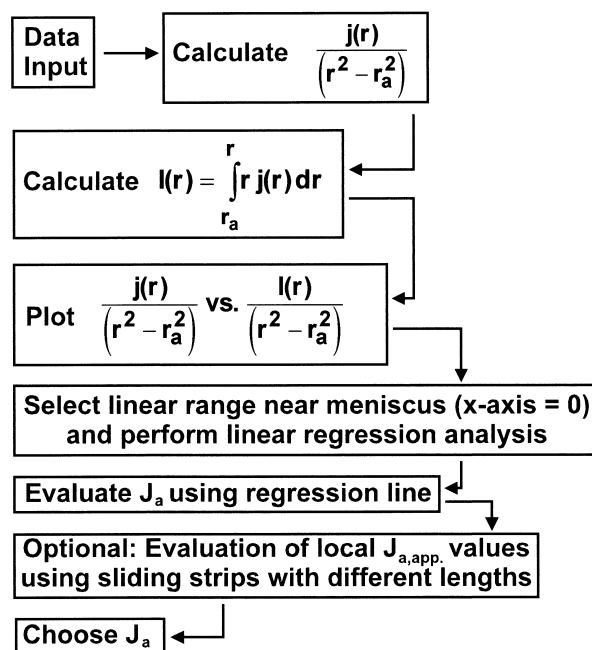

Fig. 3 Block diagram of the modules used in the MSTARI program which are different to those in MSTARA (see text)

$M_{z,app.}$ and $M_{y2}(r)$ becomes realistic, and also the data-sets of $M_{w,app.}(r)$ versus $\xi(r)$ or $J(r)$ becomes much sharper.

Conversion of A(r) or J(r) to weight concentration units

This is normally trivial. To convert $A(r)$ to $c(r)$ the extinction coefficient, ε_λ at a wavelength λ must be known:

$$c(r) = A(r)/\{\varepsilon_\lambda \cdot 1\} \tag{17}$$

where l is the optical path length of the cell (1.2 cm for a standard Beckman cell). As an example, the value for ε_λ for ovalbumin is $660\ ml \cdot g^{-1} cm^{-1}$ at a wavelength λ of 280 nm: ε_λ is a strong function of the wavelength and the amino acid content of a protein (Perkins 1986). To convert $J(r)$ to $c(r)$ the specific refractive increment $(dn/dc)_\lambda$ is needed (cf. Eqs. (2.4) and (2.5) of Creeth and Pain 1967):

$$c(r) = J(r) \cdot \lambda/\{(dn/dc) \cdot 1\} \tag{18}$$

Typically, for visible light $(dn/dc) \sim 0.18\text{–}0.19\ ml/g$ for proteins and $\sim 0.15\ ml/g$ for polysaccharides, and shows some variation with λ. Clearly, a comprehensive tabulation of ε_λ and $(dn/dc)_\lambda$ values for macromolecules would be useful for ultracentrifuge users and work is currently in progress at Nottingham.

Sample output

To give the user an impression of what to expect we have for Figs. 4 and 5 given a sample of some of the output data using MSTARA and MSTARI respectively. For the MSTARA sample output (Fig. 4) from an experiment performed on the cytoplasmic domain of the membrane protein "band 3" (a dimeric protein, of monomer molecular weight $M_1 = 40\,000$ g/mol and molar dissociation constant $K_d \sim 3\ \mu M$) the printout consists of (i) a reflection of user entered data and A_0 evaluation (not shown here) (ii) $A(r)$ versus r trace and linear extrapolation to give the absorption at the meniscus (Fig. 4a) (iii) $LnA(r)$ versus $\xi(r)$ plot (Fig. 4b) (iv) $M^*(r)$ versus $\xi(r)$ extrapolation $(M_{w,app.} = M^*(\xi \to 1))$ (Fig. 4c) (v) $M_{w,app.}(r)$ versus $\xi(r)$ (not shown) or $A(r)$ (Fig. 4d); (vi) $M_{z,app.}$ (obtained interactively from the $M_{w,app.}(r)$ versus $\xi(r)$ plot) and optional plots (not shown) of $1/M_{w,app.}(r)$ versus $A(r)$, $M_{z,app.}(r)$ versus $\xi(r)$ or $A(r)$ and a plot of $M_{y2}(r)$ versus $\xi(r)$ or $A(r)$. For MSTARI the same output protocol is followed with the exception that the $A(r)$ versus r plot (Fig. 4a) is replaced by $j(r)$ versus r (Fig. 5a) together with two further plots for extracting J_a (Fig. 5b, c), shown here from an experiment performed on a polydisperse and highly non-ideal polycationic chitosan molecule. Figures 5b and c illustrate some of the features a user can encounter when extracting the meniscus concentration J_a. The virtue of varying the sliding strip width (done automatically in MSTARI) in Fig. 5c is seen with the longer strip lengths giving the smoother form: The x-axis notation "Point number" refers to the

Table 1 ASCII data-files created for export from MSTAR

	MSTARA file attachment	MSTARI file attachment
A(r) or j(r) versus r	.AA	.IDJ
J_a extraction plot	–	.IJA
ln A(r) or ln J(r) vs ξ	.A	.IJ
M*(r) vs ξ	.AM	.IM
$M_{w,app}$ versus ξ	.AMW	.IMW
$M_{w,app}$ versus A(r) or J(r)	.AWJ	.IWJ
$M_{z,app}$ versus ξ	.AMZ	.IMJ
$M_{z,app}$ versus A(r) or J(r)	.AZJ	.IZJ

point number in the data-set (1 would be the meniscus). As explained in the Harding et al. (1992) paper, the limiting $J_{a,app.}$ at the meniscus ($\xi = 0$) equals the true meniscus concentration J_a, and can normally be found to with a few hundredth's of a fringe. In the case of Fig. 5c, a value of J_a of ~ 0.1 is returned, confirmed by Fig. 5b. Errors in J_a extraction only become serious if the total fringe increment across the cell is small ($< 2\text{–}3$ fringes). No other plots are shown here for MSTARI (the rest of the output has the identical form of MSTARA, except that $A(r)$ is substituted by $J(r)$) apart from the $M_{w,app.}(r)$ versus $J(r)$ plot (Fig. 5d) shown for two reasons (i) greater smoothness compared to the absorption data records (ii) the effect of thermodynamic non-ideality on macromolecular systems (sharp decrease in $M_{w,app.}(r)$ with increase in $J(r)$).

Finally, Figs. 4 and 5 are the raw output from the printer. The user can export any of the plots to a more advanced plotting package, as well as to more advanced routines for further analysis, as noted above. The codes are given in Table 1.

MSTAR output for ideal, heterogeneous, non-ideal and self-associating systems

To give the user an idea or further insight as to what the various types of system he may encounter can produce in terms of the form of the concentration distribution and molecular weight plots, we consider here for systems of simulated data (200 data points per system): (a) *a simple ideal one component system* {M = 20,000 g/mol, speed = 15,000 RPM, temperature = 20 °C, baseline absorbance 0.01, $r_a = 6.0$ cm, $r_b = 7.2$ cm, $\bar{v} = 0.73$ ml/g, $\rho = 1.001$ g/ml, loading concentration, $A_0 = 0.060$} (b) *an ideal polydisperse system of two non-interacting components* {$M_1 = 10,000$ g/mol + $M_2 = 30,000$ g/mol, speed = 15,000 RPM, temperature = 20 °C baseline absorbance 0.01, $r_a = 6.0$ cm, $r_b = 7.2$ cm, $\bar{v} = 0.73$ ml/g for both components, $\rho = 1.001$ g/ml, $A_{0,1} = 0.048$, $A_{0,2} = 0.050$, $M_w = 20,167$ g/mol, $M_z = 25,124$ g/mol}; (c) *a non-ideal one component system* {M = 20,000 g/mol, speed = 15,000 RPM, temperature = 20 °C, baseline absorbance 0.01, $r_a = 6.0$ cm, $r_b = 7.2$ cm, $\bar{v} = 0.73$ ml/g, $\rho = 1.001$ g/ml, BM (absorbance units^{-1}) = 0.4, $A_0 = 0.058$, $M_{w,app.} = 19,119$ g/mol}; (d) *an ideal dimerising system* {$M_1 = 20,000$ g/mol, $k_2 = 9.98$

Plot of A (Y-axis) versus r (X-Axis) for a:\535c3m.ra3

A (Meniscus) = .1140752

Plot of ln(A) (Y-axis) versus Xsi (X-Axis) for a:\535c3m.ra3

Plot of Mstar * 10^5 (Y-axis) versus Xsi (X-Axis) for a:\535c3m.ra3

Mw,app = 77228 g/mol

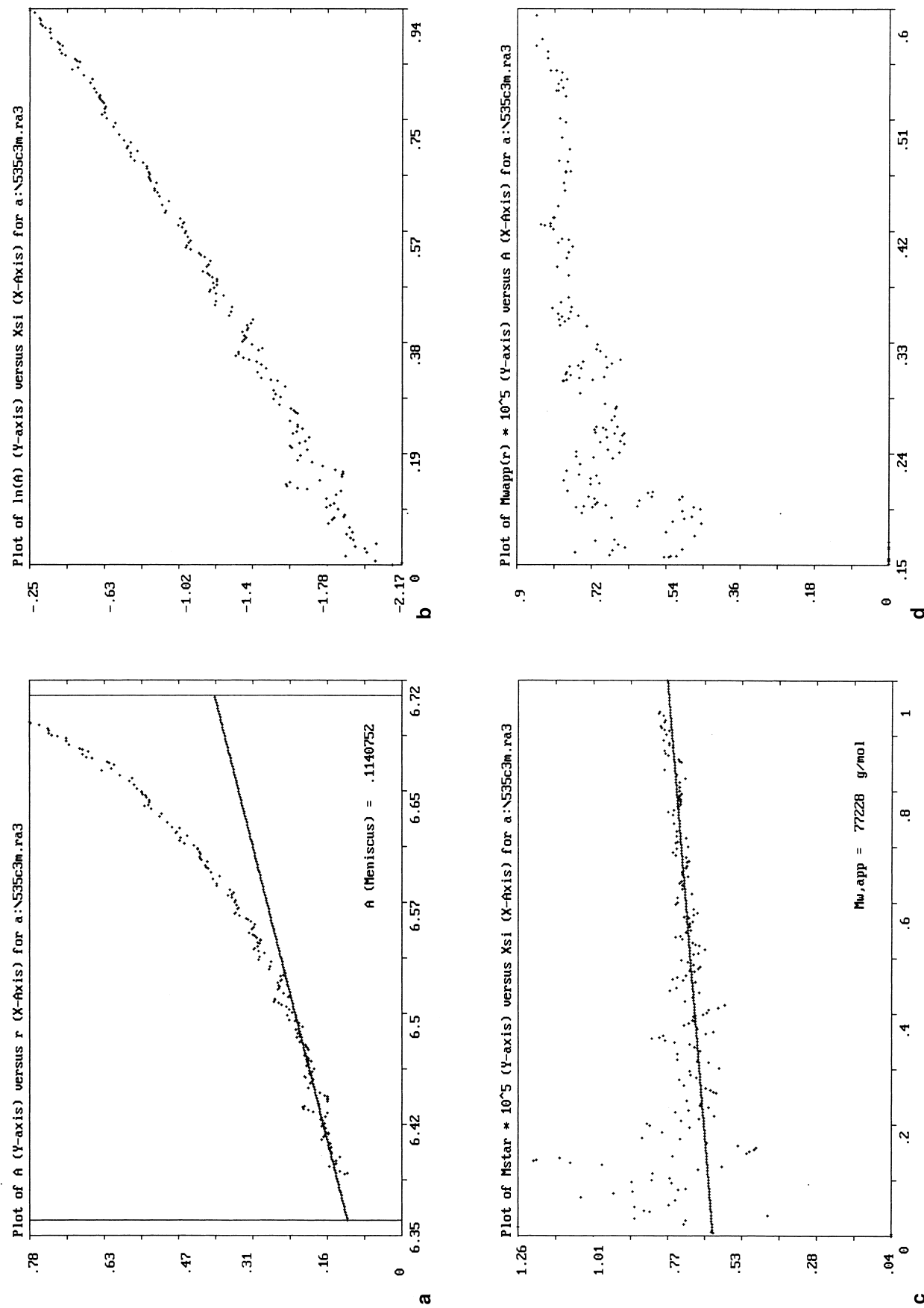**Plot of Mwapp(r) * 10^5 (Y-axis) versus A (X-Axis) for a:\535c3m.ra3**

Fig. 4 Sample output (raw plots) from part of MSTARA. Data from an XL-A run on the cytoplasmic domain of the band 3 membrane protein. $c_0 = 0.175$ mg/ml, rotor speed = 10.000 rev/min, temperature = 20.0 °C, wavelength = 280 nm, 12 mm path length cell. Baseline absorbance = 0.045. (Cölfen et al. 1996)
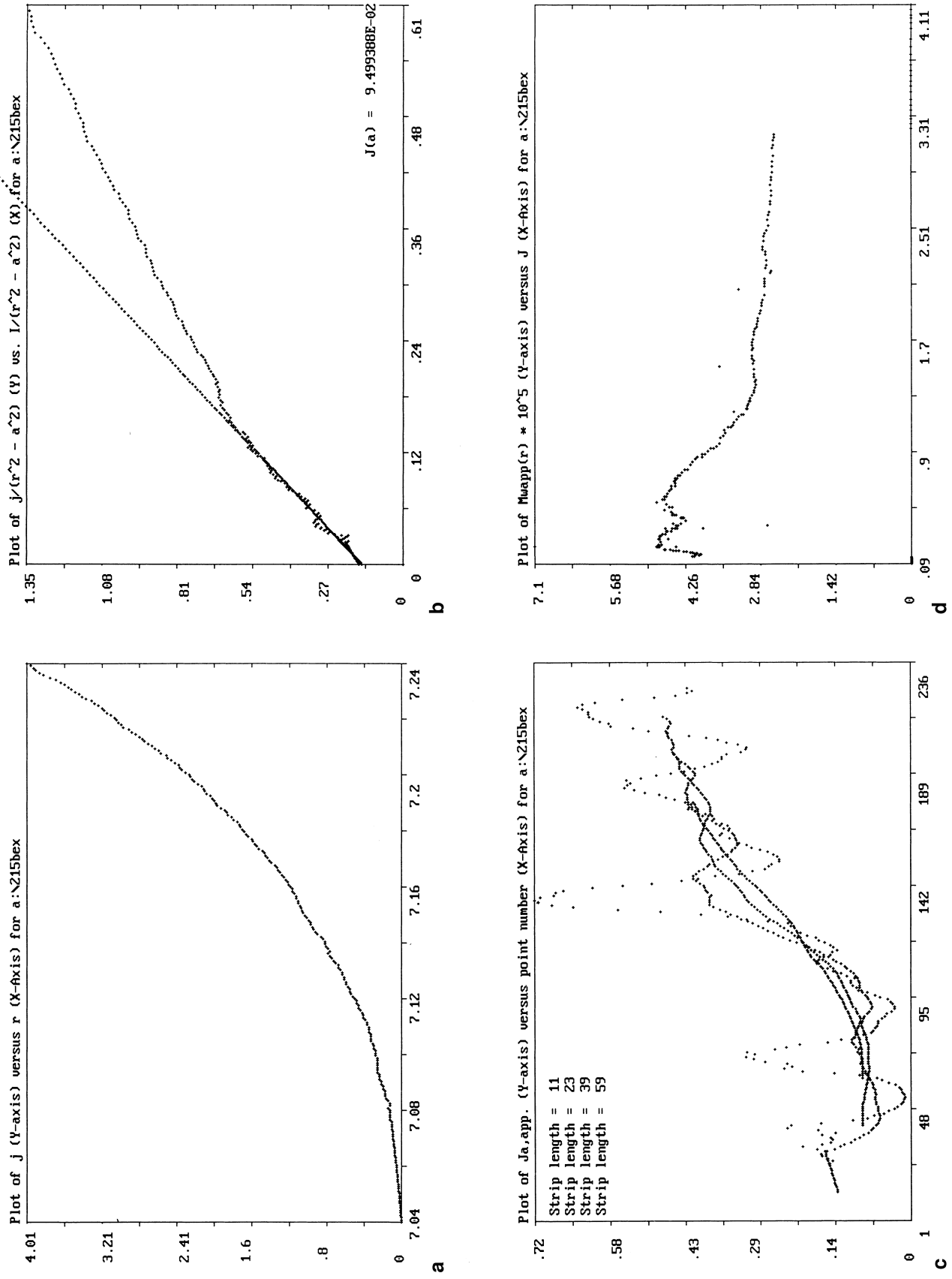
Plot of J (Y-axis) versus r (X-Axis) for a:\215bex

4.01

3.21

2.41

1.6

.8

0

7.04   7.08   7.12   7.16   7.2   7.24

a

Plot of J/(r^2 - a^2) (Y) vs. 1/(r^2 - a^2) (X),for a:\215bex

1.35

1.08

.81

.54

.27

0

0   .12   .24   .36   .48   .61

J(a) = 9.49938BE-02

b

Plot of Ja,app. (Y-axis) versus point number (X-Axis) for a:\215bex

.72

.58

.43

.29

.14

0

1   48   95   142   189   236

Strip length = 11
Strip length = 23
Strip length = 39
Strip length = 59

c

Plot of Mwapp(r) * 10^5 (Y-axis) versus J (X-Axis) for a:\215bex

7.1

5.68

4.26

2.84

1.42

0

.09   .9   1.7   2.51   3.31   4.11

d

Fig. 5 Sample output (raw plots) from part of MSTARI. Data from a Model-E run on a chitosan "FR2". $c_0 = 0.6$ mg/ml, rotor speed = 9000 rev/min, temperature = 23.5 °C, 30 mm path length cell

(absorbance units^{-1}), speed = 15,000 RPM, temperature = 20 °C, baseline absorbance 0.01, r_a = 6.0 cm, r_b = 7.2 cm, $\bar{v}$ = 0.73 ml/g, ρ = 1.001 g/ml, A_0 = 0.095, $A_{0,1}$ = 0.060, $A_{0,2}$ = 0.035, M_w = 27,461 g/mol, M_z = 30,870 g/mol). For convenience, the output is given in MSTARA form, with concentrations given in absorbance units. Fig. 6 shows the concentration distribution for each system.

The ln A vs. ξ plot (Fig. 7)

Note the linear plot for an ideal single solute system (Fig. 7a), upward curvature for a heterogeneous mixture of either non-interacting (Fig. 7b) or interacting (Fig. 7d) components, and downward curvature for the non-ideal system (Fig. 7c).

The M vs. ξ plot*

By definition, M*(r) at the cell base (r = b or ξ = 1) equals $M_{w,app.}$ over the whole distribution (i.e. *not* just at the cell bottom). In practice, the data cannot always be observed completely to the cell bottom which makes extrapolations necessary, and Fig. 8a–d show the form these may take depending on the system. Figure 8a shows that even for an ideal single solute system, the greatest accuracy in the
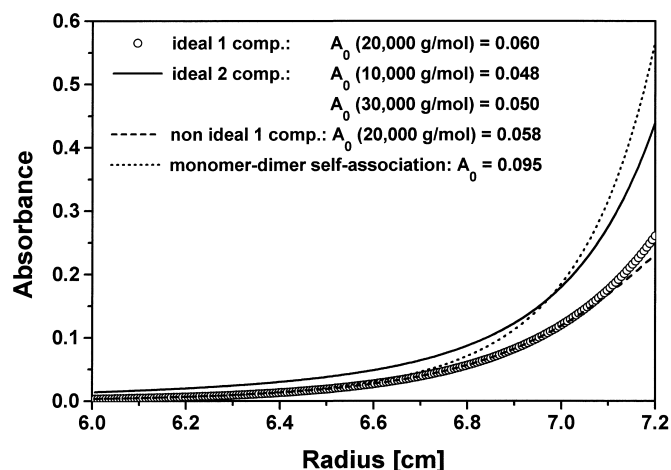
$M_{w,app.}$ determination is when all the data in the cell is taken into account (i.e. the M* data accumulation is taken to completion i.e. to the cell base) and demonstrates the lower precision of the data nearer the meniscus compared to the cell base. In every case, the $M_{w,app.}$ for the system is faithfully reproduced (hardly surprising when the data is of machine accuracy).
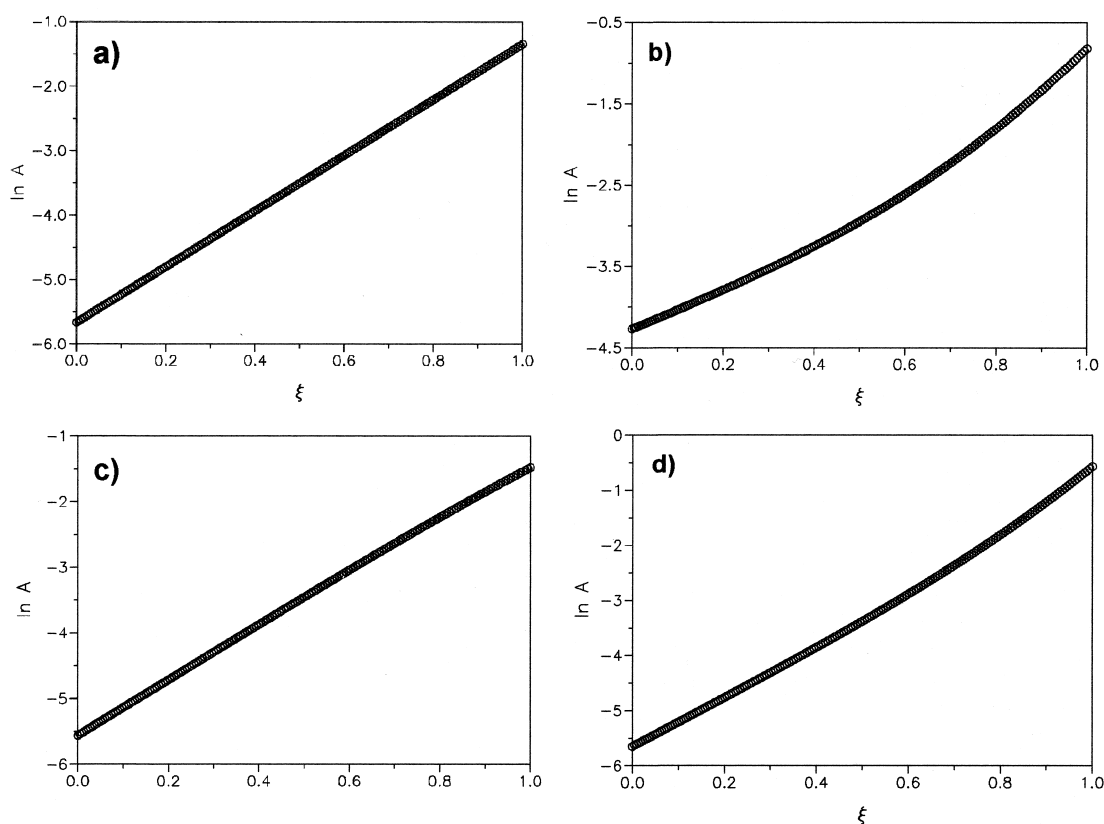

Fig. 6 Concentration gradients for the simulated datafiles (**a**)–(**d**)

Fig. 7 Plots of ln A(r) vs. ξ for simulated datafiles (**a**)–(**d**)

Fig. 8 Plot of M* vs. ξ for simulated datafiles (**a**)–(**d**)
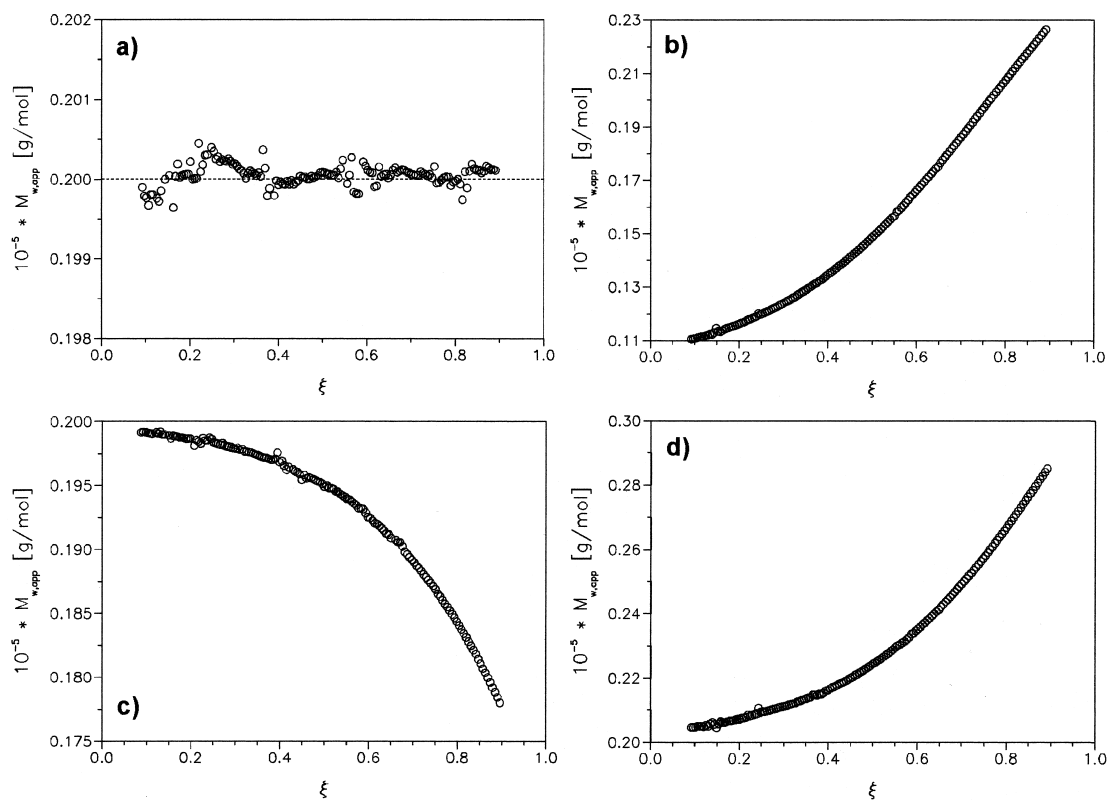

Fig. 9 Plot of $M_{w,app.}(r)$ vs. ξ for simulated datafiles (**a**)–(**d**). The first and last 20 data points are missing due to the sliding strip procedure using a sliding strip of 20
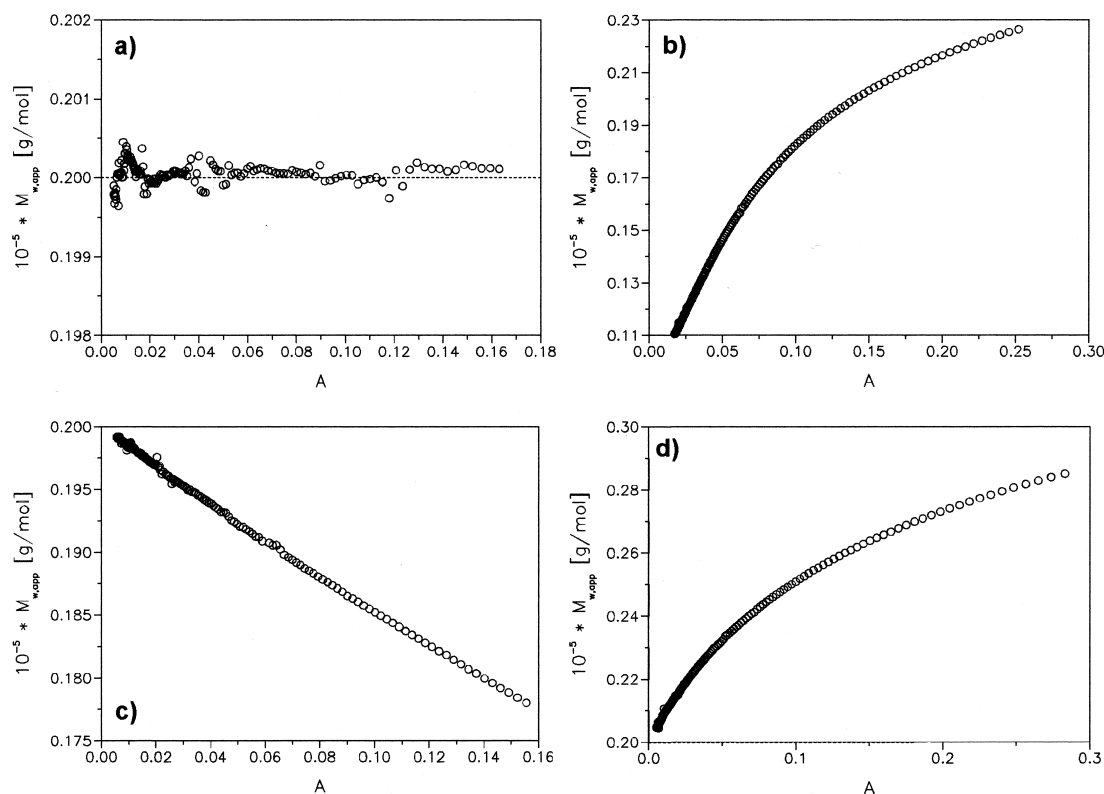
Fig. 10 Plot of $M_{w,app.}(r)$ vs. A(r) for simulated datafiles (**a**)–(**d**)

$M_{w,app.}(r)$ versus ξ and A(r) plots

The next plot in the MSTAR program gives the classical determination of $M_{w,app.}(r)$ by evaluating the local slopes of the plots represented in Fig. 7 using Eq. (8) and applying a sliding strip procedure. For all data files, the sliding strip length was 20 which applies already a moderate data smoothing. It is interesting to observe that for both heterogeneous systems the plots of $M_{w,app}(r)$ are *concave* when plotted against radial displacement squared {or ξ} (Fig. 9b, d) but are *convex* when plotted (Fig. 10b, d) versus A(r): this is evident also in Fig. 4d for the dimerising cytoplasmic domain of the band 3 protein. The reverse is true for the case of non-ideal systems $M_{w,app.}(r)$ versus $\xi(r)$ is convex (Fig. 9c) whereas it is concave when plotted against A(r) (Fig. 10c): again this is illustrated nicely for real data with the chitosan example of Fig. 5d (except that concentration is expressed in Rayleigh fringe rather than uv-absorbance units). Using Eq. (14) to derive $M_{z,app.}$ for the 2 component system (Fig. 10b), an $M_{z,app.}$ of 25,366 g/mol is obtained (error 0.8%) which is close to the real value of 25,124 g/mol. Using this procedure appears to be safer than directly applying a fitting function like a spline function (Lechner and Mächtle 1992) to the concentration gradient, because this has to fit the whole gradient which is not always satisfactorily (see also Fig. 2).

$M_{z,app.}(r)$ versus A(r)

Figures 11a–d show clearly (by comparison with their weight average counterparts of Fig. 10) that $M_{z,app.}$ is a more sensitive function of heterogeneity and also non-ideality (compare e.g. greater concavity of Fig. 11c compared with Fig. 10c).

Further non-ideal plots

The final plots (Fig. 12a, b) show the usefulness of two procedures for manipulating the point average data. Firstly, and providing there is no serious redistribution of solute components at the rotor speed used (an invalid approximation for seriously heterogeneous systems which will tend to accumulate a greater proportion of the higher molecular weight components in a mixture towards the cell base, and vice versa for the meniscus), Fig. 12(a) shows that when non-ideality is of the first order, a linear fit is obtained of $1/M_{w,app.}(r)$ versus concentration. According to Eq. (13) this yields $M_w = 20,002$ g/mol in perfect agreement with the true value and BM = 0.398 (reciprocal absorbance units) corresponding to an error of 0.5%.

The second plot in Fig. 12 shows the use of the Roark-Yphantis (1969) M_{y2} function for first order non-ideal systems. By combining $M_{w,app.}(r)$ and $M_{z,app.}(r)$ together in the way described by Eq. (10), the downward trend of both $M_{w,app.}(r)$ and $M_{z,app.}(r)$ with increase in ξ or c(r) is completely removed.
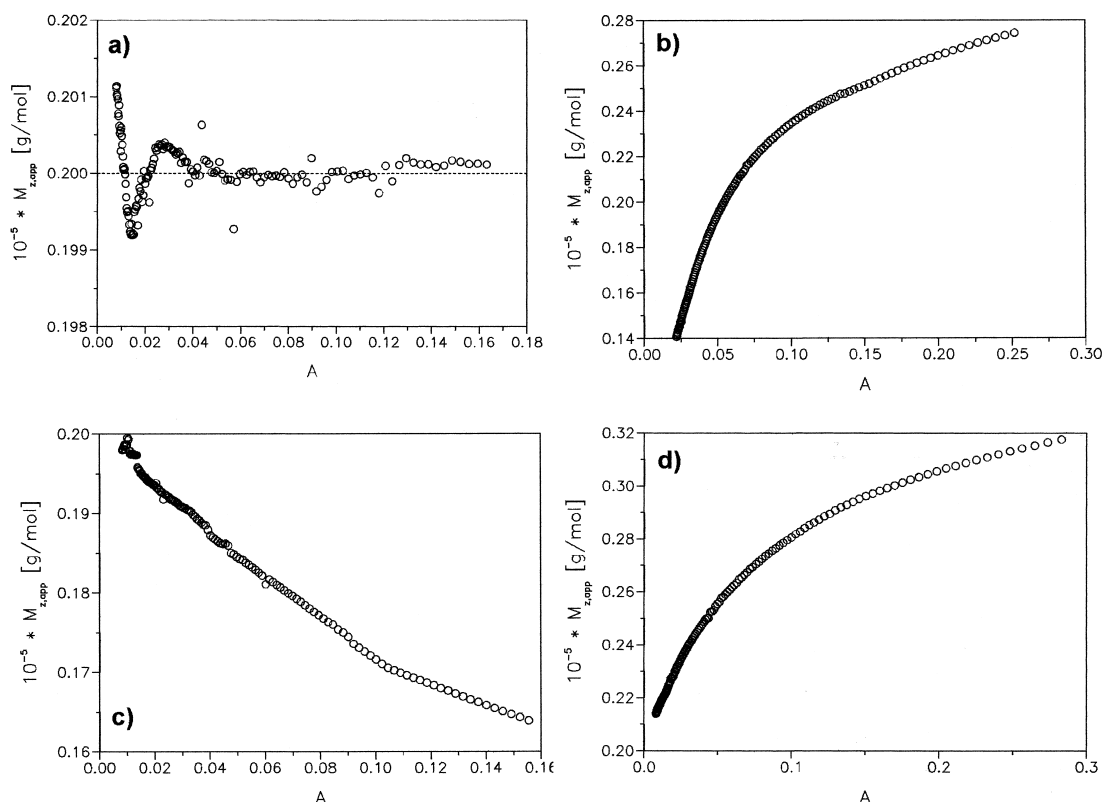
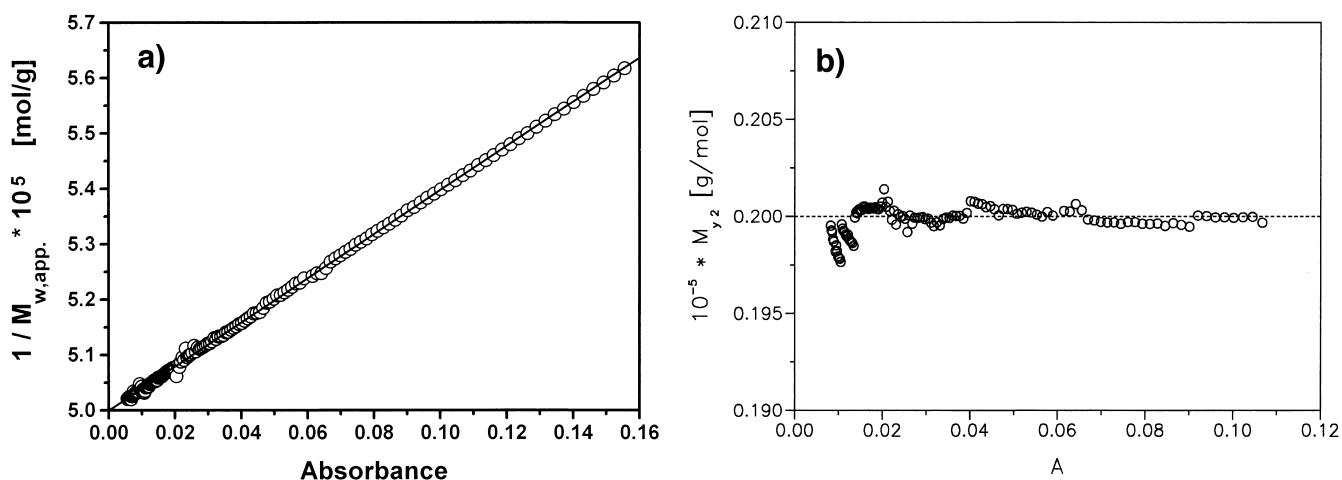**Fig. 11** Plot of $M_{z,app.}(r)$ vs. $A(r)$ for simulated datafiles (**a**)–(**d**)

Fig. 12 Coping with non-ideality (**a**) Plot of $1/M_{w,app.}(r)$ vs. $A(r)$ for simulated datafile (**c**) to yield M_w. (**b**) Plot of $M_{y,2}(r)$ vs. $A(r)$ for the simulated datafile (**c**) to show removal of first order effects of non-ideality

Concluding remarks

The user of the analytical ultracentrifuge, faced with a plethora of possible analysis software, should be able to find these MSTAR routines useful and hopefully simple to

use as a base programme before charging into more advanced packages. MSTAR, although a simple model independent package, does not consider only a selected dataset in the distribution of the macromolecular solute at sedimentation equilibrium, but considers the *complete* solute distribution including those parts of the initial distribution loaded into the centrifuge cell which may be lost from optical registration near the cell base or which cannot be used because of Lambert-Beer law restrictions ($A(r)$ has to be $< \sim 1.4$) in its calculation of the weight average molecular

weight $M_{w,app.}$ (via M*) and other parameters. The form of the other plots {$\ln c(r)$ versus r^2, $M_{w,app.}(r)$ versus r^2 or $A(r)$ etc.} should give the user a fair idea of the nature of any heterogeneity or non-ideality if present, which can then be followed up by easy export of data to specific model dependent programmes like ASSOC4, NONLIN, OMEGA or PSI. The usefulness of the latter, based on the Ω function (Millthorpe et al. 1976) and the PSI function (Wills et al. 1996) as a bolt-on module to MSTAR will be described elsewhere.

Programm availability The MSTAR programs are available free of charge either directly from the authors (email or conventional mail) or from the RASMB data base of ultracentrifuge software. Log in as anonymous ftp on BBRI.HARVARD.EDU and then change to /RASMB/SPIN/MS_DOS/MSTAR-COELFEN.
E-mail: COELFEN@MPIKG-TELTOW.MPG.DE
STEVE.HARDING@NOTTINGHAM.AC.UK

References

Cölfen H, Harding SE, Boulter J, Watts A (1996) Hydrodynamic examination of the dimeric cytoplasmic domain of the human erythrocyte anion transporter, Band-3. Biophys J 71:(in press)

Creeth JM, Harding SE (1982) Some observations on a new type of point average molecular weight. J Biochem Biophys Methods 7:25–34

Creeth JM, Pain RH (1967) The determination of molecular weights of biological macromolecules by ultracentrifuge methods. Progr Biophys Mol Biol 17:217

Fujita H (1975) Foundations of ultracentrifuge analysis. J Wiley, New York, p 423

Furst A (1997) The XL-1 analytical ultracentrifuge with Rayleigh interference optics. Eur Biophys J 25:307–310

Giebeler R (1992) The Optima XL-A: A new analytical ultracentrifuge with a novel precision absorption optical system. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 16–31

Harding SE, Horton JC, Morgan PJ (1992) MSTAR: A FORTRAN program for the model independent molecular weight analysis of macromolecules using low speed or high speed sedimentation equilibrium. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 275–294

Johnson ML, Correia JJ, Yphantis DA, Halvorson HR (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least squares techniques. Biophys J 36:575–588

Laue TM (1992a) Short column sedimentation equilibrium analysis for rapid characterization of macromolecules in solution; Technical information DS-835, Spinco Business Unit, Palo Alto, Cal.

Laue TM (1992b) On-line data acquisition and analysis from the Rayleigh interferometer. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 63–89

Lechner MD, Mächtle W (1992) Sedimentation equilibrium measurements with the new analytical ultracentrifuge Optima XL-A and its digital ultraviolet/visible detector. Macromol Chem Rapid Commun 13:555–563

Lechner MD (1992) Determination of molecular weight averages and molecular weight distribution from sedimentation equilibrium. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 295–310

Milthorpe BK, Jeffrey PD, Nichol LW (1975) The direct analysis of sedimentation equilibrium results obtained with polymerizing systems. Biophys Chem 3:169–176

McRorie DK, Voelker PJ (1993) Self-associating systems in the analytical ultracentrifuge. Beckman Instruments Inc, Fullerton, Cal

Roark DE, Yphantis DA (1969) Studies of self-associating systems by equilibrium ultracentrifugation. Ann NY Acad Sci 164: 245–278

Rowe AJ, Wynne Jones S, Thomas DG, Harding SE (1992) Methods for off-line analysis of sedimentation velocity and sedimentation equilibrium patterns. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 49–62

Schachman HK (1989) The analytical ultracentrifuge reborn. Nature 941:259

Stafford WF (1992) Methods for obtaining sedimentation coefficient distributions. In: Harding SE, Rowe AJ, Horton JC (eds) Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, pp 359–393

Teller DC (1973) Characterization of proteins by sedimentation equilibrium in the analytical ultracentrifuge. Methods Enzymol 27:346–441

Wales M, Adler FT, Van Holde KE (1951) Sedimentationsgleichgewichte polydisperser, nichtidealer gelöster Stoffe. J Phys Coll Chem 55:145–161

Wills PR, Jacobsen MP, Winzor DJ (1996) Direct analysis of solute self-association by sedimentation equilibrium. Biopolymers 38:119–130

Yphantis DA (1964) Equilibrium ultracentrifugation of dilute solutions. Biochemistry 3:297–317