THE REPRESENTATION OF EQUILIBRIUM SOLUTE DISTRIBUTIONS FOR NONIDEAL POLYDISPERSE SYSTEMS IN THE ANALYTICAL ULTRACENTRIFUGE

Application to Mucus Glycoproteins

STEPHEN E. HARDING
Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

ABSTRACT It is relatively easy to represent by computer simulation the observed Rayleigh equilibrium fringe data for systems that are both associative and nonideal in the thermodynamic sense, and to extract the determinant parameters (see, for example, Roark, D., and D. A. Yphantis, 1969, *Ann. NY Acad. Sci.*, 164:245–278; and Johnson M. L., J. J. Correia, D. A. Yphantis, and H. R. Halvorson, 1981, *Biophys. J.* , 36:575–588). It is, however, considerably more difficult to represent systems that are both polydisperse (namely, those that consist of noninteracting species of different molecular weight) and nonideal, although the ideal case has been well described (see, for example, Tindall, S. H., and K. C. Aune, 1982, *Anal. Biochem.* 120:71–84). Here we show that the representation of nonideal polydisperse systems is now possible, after certain assumptions, by using a two-part interdependent minimization routine that uses readily available numerical packages. The method is applied to a well-characterized mucus glycoprotein ($M_r \sim 2 \times 10^6$) from the bronchial secretion of a cystic fibrosis patient. An excellent fit to the observed fringe data is obtained for a polydisperse three-component system, with a value for the second virial coefficient, B, of 0.57 ml mol g^{-2}.

INTRODUCTION

The interpretation of Rayleigh interference fringe patterns from equilibrium analytical ultracentrifugation is often complicated by the presence of any inherent heterogeneity of the sample. By "heterogeneity" we mean the presence of solute particles in chemical equilibrium with each other (i.e., association) or the presence of nonreacting species of different molecular weight or densities (polydispersity). The interpretation may be further complicated by the effects of thermodynamic nonideality, even at low cell-loading concentrations. This problem has been manifested in, for example, mucus glycoproteins, which appear to be highly polydisperse, associative, and have very large excluded volumes in solution.

It is relatively easy to represent the fringe and molecular weight distributions for actual data in terms of, where appropriate, associating systems that are ideal in the thermodynamic sense, and even nonideal single-solute and associative systems (Roark and Yphantis, 1969; Teller, 1973; Kim et al., 1978). Several recent examples of such representations using nonlinear least-squares fitting procedures to simulated data have been given by Johnson et al. (1981). The representation of ideal polydisperse systems is also relatively straightforward (Creeth, 1980). Another recent study has demonstrated the successful application of minimization procedures for the fitting of synthetic data to ideal polydisperse associating systems (Tindall and Aune, 1982). The representation of polydisperse systems that are not ideal in the thermodynamic sense is, however, more difficult but can still be achieved for certain cases using readily available numerical packages. An algorithm using such data-fitting procedures is presented here.

MATHEMATICAL FORMULATION

Self-associating Systems

Before considering polydisperse, nonideal systems, it may be instructive to consider a simple (nonideal) associative system, for example, a nonideal isodesmic association (IDA). The monomer (fringe) concentration $J_1(r)$ is given by (Kim et al., 1977)

$$J_1(r) = J_1(a) \exp\{(r^2 - a^2)A_1 - BM_1[J(r) - J(a)]\}, \quad (1)$$

where r is the radial position, a the corresponding radial position of the meniscus, A_1 the reduced (monomer) molecular weight (Creeth and Harding, 1982), $J(r)$ the total concentration, and B the communal second

Dr. S. Harding's present address is the University of Nottingham, School of Agriculture, Sutton Bonington, LE12 5RD, United Kingdom.

virial coefficient. For an IDA the total concentration, $J(r)$, at a given radial position, r, is given by

$$J(r) = \frac{J_1(r)}{[1 - kJ_1(r)]^2}, \tag{2}$$

where k is the intrinsic association constant (in fringe units). Other simple relations exist for other types of self-association (Kim et al., 1977).

The assumption is made throughout that a second virial coefficient is sufficient to account for all the nonideality (namely, no higher coefficients are needed) and also that the same coefficient can be used to adequately describe the nonideality of all species present in the system. The concentration distribution may therefore be obtained for given values of A_1, B, and k by solving the simple nonlinear equation

$$J_1(r) = J_1(a) \exp \cdot \left(A_1(r^2 - a^2) - BM_1 \left\{ \frac{J_1(r)}{[1 - kJ_1(r)]^2} - \frac{J_1(a)}{[1 - kJ_1(a)]^2} \right\} \right), \tag{3}$$

subject to the conservation-of-mass condition

$$\int_a^b J(r)\, dr = \frac{J^\circ(b^2 + a^2)}{2}, \tag{4}$$

where J° is the initial cell loading concentration and b the radial position at the cell base.

An example of such a fit to a real system is given in Fig. 1, for a well-characterized mucus glycoprotein (Harding and Creeth, 1983) from a cystic fibrosis patient. Fig. 1 also illustrates how the effects of association (and/or polydispersity) and thermodynamic nonideality can often compensate for each other to give a Rayleigh interference pattern that appears to represent a single solute system that is thermodynamically ideal (Teller, 1965). Fitting real data using parameters for a nonideal self-association can be performed relatively quickly using modest computational facilities, and is, therefore, amenable to full nonlinear least-squares types of analyses and investigation of the relationship between data error and the derived parameters (Johnson et al., 1981). For

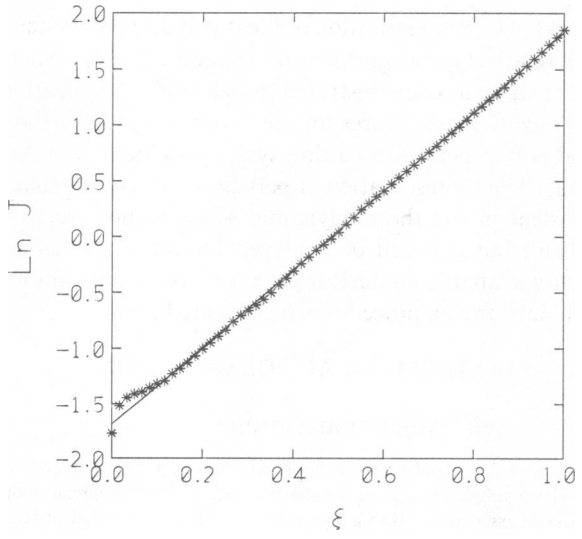

FIGURE 1 Plot of the logarithm of the fringe concentration, J, as a function of the radial displacement parameter $\xi[= (r^2 - a^2)/(b^2 - a^2)]$ for glycoprotein CF PHI: 30-mm cell, 3-mm solution column. Initial loading concentration, $c^\circ \sim 0.2$ mgs/ml. The line fitted is for a nonideal IDA with the following parameters: $M_1 = 2.15 \times 10^6$, $k = 0.26\ dm^3\ g^{-1}$, $B = 1.5 \times 10^{-4}$ ml mol g^{-2}.

thermodynamically nonideal polydisperse representations this is, unfortunately, not the case.

Polydisperse Systems

The concentration distribution for each component, i, of an ideal, noninteracting n-component mixture is given by the Rinde formula (Rinde, 1928; Fujita, 1962)

$$J_i(r) = J_i^\circ \left\{ \frac{A_i(b^2 - a^2) \exp [A_i(r^2 - a^2)]}{\exp [A_i(b^2 - a^2)] - 1} \right\}. \tag{5}$$

The total concentration at each radial position of the cell is given simply by $J(r) = \Sigma_i^n J_i(r)$.

For the generalized thermodynamically nonideal polydisperse case the representation is not so simple. The concentration distribution for each component, i, is given by

$$J_i(r) = J_i(a) \{\exp A_i(r^2 - a^2) - BM_i [\Sigma_j^n J_j(r) - \Sigma_j^n J_j(a)]\} \quad (i, j = 1 \rightarrow n). \tag{6}$$

This is no longer a simple nonlinear equation (unlike Eq. 3) because the nonideal term in Eq. 6 involves the summed concentration over all the components at a particular radial position r (and a). There are now $2n$ unknowns ($J_i[r]$, $J_i[a]$, $i = 1 \rightarrow n$) to solve between the n nonlinear equations for each radial position, subject to the n constraints (over all radial positions)

$$\int_a^b J_i(r)\, dr = \frac{J_i^\circ (b^2 + a^2)}{2}. \tag{7}$$

COMPUTATIONAL METHODS

It is convenient to separate the minimization into two parts. First, to assume a set of the parameters $J_i(a)$, and for this set to minimize the set of nonlinear equations (Eq. 6) that will now be

$$F_i(r) = J_i(r) - J_i(a) \cdot \exp \{A_i(r^2 - a^2) - BM_i (\Sigma_j^n J_j(r) - \Sigma_j^n J_j(a))\} \quad (i, j = 1 \rightarrow n) \tag{8}$$

for each of, say, five more (in addition to the meniscus) radial positions in the cell. The best value of the $J_i(r)$ for a particular choice of the $J_i(a)$ can then be readily found taking advantage of the fact that the differentials $\partial F_i/\partial J_j(r)$ $(i, j = 1 \rightarrow n)$ can be specified. Eq. 8 can then be regarded as a special case of finding a minimum of

$$S(r) = \Sigma_{i=1}^n [F_i(r)]^2. \tag{9}$$

A suitable routine here that makes use of the (user specifiable) first derivatives is the NAG routine E04GEF (1978), based on an algorithm of Gill and Murray (1978). E04GEF is an easy to use modified Gauss-Newton algorithm for finding an unconstrained minimum of a sum of squares of i nonlinear equations in i variables (or less).

From the calculated values of the $J_i(r)$ (for the current estimate for the $J_i[a]$) for each radial position of the cell, the total cell concentration, J_i° of each noninteracting component i can be calculated by numerical integration from Eq. 7 by quadrature, using, for example, NAG (1978) D01GAF. This routine, based on a method of Gill and Miller (1972) uses a four-point finite-difference formula centered on the interval concerned, except in the case of the first and last intervals where four-point forward and backward difference formulae are respectively used.

The values of $G_i = J_{i,\mathrm{CALC}}^\circ - J_i^\circ$ for each component are then fed back into the global minimization route in (e.g., NAG C05NBF), which

initiates a different set of estimates for the $J_i(a)$. C05NBF is a routine for finding a zero of a system of i nonlinear equations in i variables (for which explicit derivatives cannot be supplied by the user) by a modification of the Powell hybrid method (Powell, 1970). It chooses the correction at each step as a combination of the Newton and scaled-gradient directions, normally guaranteeing global convergence for starting points far from the solution and a fast convergence. The whole process (involving E04GEF and D01GAF) is then iterated until $G_i \rightarrow 0$. This procedure has been outlined in the flow chart of Fig. 2.

RESULTS AND DISCUSSION

We have already demonstrated (Fig. 1) that if we neglect polydispersity, a nonideal self-associating system could reasonably represent the observed fringe data for a cystic fibrosis mucus glycoprotein (CF PHI). It is normally impossible, however, to distinguish between the effects of association and polydispersity in a single experiment, even when nonideality is present. However, it was demonstrated earlier (Harding, 1984), using the diagnostic test of nonoverlap of point-weight average molecular weight vs. concentration plots for different initial cell loading concentrations (Roark and Yphantis, 1969), that polydispersity must be significant for this glycoprotein. It was also demonstrated (Harding, 1984) from the lack of effect upon blocking any potential sites for association, that self-association was not significant. These conclusions are consistent with the findings of Creeth and Cooper (1984) for other mucus glycoproteins: the observed heterogeneity in mucins appears to be generally due to polydispersity.

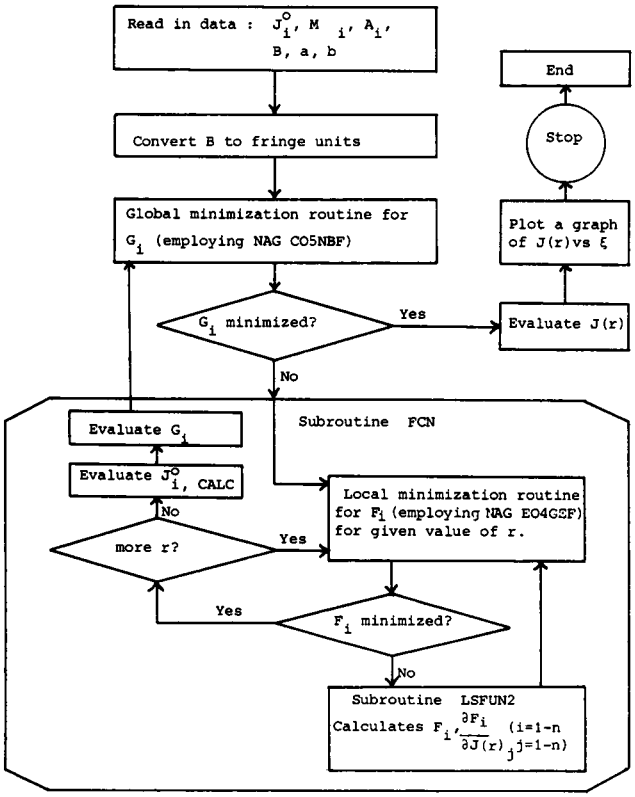

FIGURE 2 Flow chart for the nonideal polydisperse algorithm.

Fig. 3 shows an attempt to represent observed fringe data for the glycoprotein CF PHI (Harding and Creeth, 1983) as a thermodynamically nonideal three-component mixture. It is now widely accepted (Silberberg and Meyer, 1982; Harding et al., 1983*a*) that mucus glycoproteins such as CF PHI are built up from multiples of a 500,000–600,000 molecular weight basic unit. These units are apparently assembled into a linear array (Harding et al., 1983*a*; Carlstedt et al., 1983). For a given mucus glycoprotein sample there is likely to be considerable variability in the numbers of units per molecule, giving rise to a discrete distribution of molecular weights. This is apparently evident for CF PHI, as visualized by electron microscopy (Harding et al., 1983*a*, *b*). To a first approximation, if we take the dominant form (say 67%) as the three-basic-unit form (mol wt = 1,800,000) and assume a lower proportion of two-unit and four-unit forms (16.5% each) it is possible to obtain an excellent fit to the observed fringe data if a value for the second virial coefficient of 0.57×10^{-4} ml mol g^{-2} is chosen (Fig. 3). The value of B estimated earlier (Harding and Creeth, 1982) using the technique of ultrashort column sedimentation equilibrium (1.5×10^{-4}) is found to give a poor fit; an ideal system ($B = 0.0$) also gives a poor fit.

The goodness of fit does not necessarily mean that a three-component nonideal system is the best model for CF PHI. As is always the problem in attempting to represent systems involving terms of an exponential character, there may be other solutions giving equally good fits involving, for example, four, five, or more components. In addition to the discrete form of polydispersity arising from the varia-

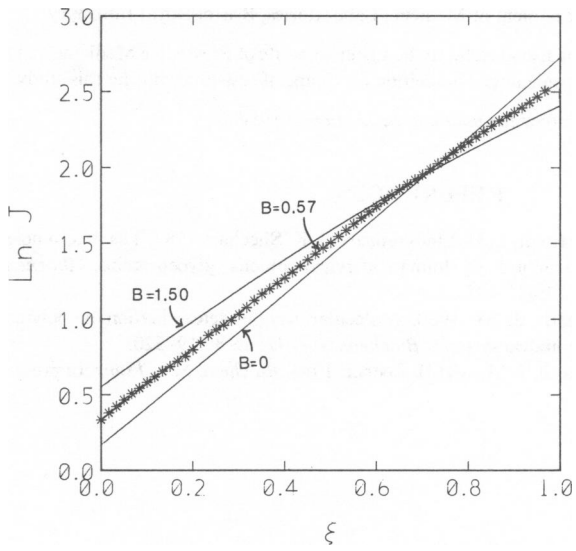

FIGURE 3 Plot of $\ln J$ vs. ξ for glycoprotein CF PHI: 12-mm cell, 3-mm solution column. $J^o = 5.4$ ($\equiv$ ~2.0 mgs/ml). K [$\equiv (1 - \bar{\nu}\rho)/RT$] = 0.87316×10^{-6} mol g^{-1}. The line fitted corresponds to a nonideal three-component mixture with the following component parameters: component 1, $M_1 = 1.2 \times 10^6$, $J_1^o = 0.9$; component 2, $M_2 = 1.8 \times 10^6$, $J_2^o = 3.6$; component 3, $M_3 = 2.4 \times 10^6$, $J_3^o = 0.9$. $A_i = KM_i/2$ ($i = 1 - 3$); units of B: ml mol $g^{-2} \times 10^4$.

bility in the numbers of subunits, there is likely to be superimposed on this a quasi-continuous distribution of molecular weights (and partial specific volumes) arising from variability in the carbohydrate composition. Representation in terms of a nonideal log-normal distribution may prove to be more realistic. Although at present modest computational facilities will allow up to, say, 21 = component fits with, for example, a Gaussian distribution of molecular weights, nonideal log-normal fits provide at the present time insurmountable scaling problems due to the widely varying distribution of molecular weights. Note that the single fit described for the three-component system uses a relatively large amount of central processor unit time, even on a very fast computer (~1 min on an IBM 3081, model B; IBM Instruments, Inc., IBM Corp., Danbury, CT). This would appear to limit at the present time the application of full nonlinear least-squares fits of data to the parameters (B, M_i, c_i^o, etc.), particularly where more than one of these parameters is considered a variable (a single iteration for such a procedure would correspond to this time). This would also therefore appear to limit a full determination of the relationship between data error and the derived parameters.

However the strength of the present technique is when it is supplemented by other information on the polydispersity, such as from electron microscopy as illustrated above, and from photon correlation spectroscopy (Provencher, 1979), and also when we can establish that it is polydispersity and not self-association that is the dominant contribution to observed solute heterogeneity (Harding, 1984).

I am grateful for the invaluable help and advice of Dr. J. M. Creeth, of the Department of Medicine Laboratories, Bristol Royal Infirmary.

I am also grateful to the Lister Institute of Preventive Medicine and to the Oppenheimer Foundation for financial assistance during this study.

Received for publication 23 August 1984.

REFERENCES

Carlstedt, I., H. Lindgren, and J. K. Sheehan. 1983. The macromolecular structure of human cervical mucus glycoproteins. *Biochem. J.* 213:427–435.

Creeth, J. M. 1980. Molecular weight determination in polydisperse nonideal systems. *Biochem. Soc. Trans.* 8:519–520.

Creeth, J. M., and B. Cooper. 1984. *Biochem. Soc. Trans.* In press.

Creeth, J. M., and S. E. Harding. 1982. Some observations on a new type of point average molecular weight. *J. Biochem. Biophys. Methods.* 7:25–34.

Fujita, H. 1962. Mathematical Theory of Sedimentation Analysis. Academic Press, New York. 235–398.

Gill, P. E., and G. F. Miller. 1972. An algorithm for the integration of unequally spaced data. *Comput. J.* 15:80–83.

Gill, P. E., and W. Murray. 1978. Algorithms for the solution of the non-linear least squares problem. *SIAM (Soc. Ind. Appl. Math.) Conf. Ser. Journal on Numerical Analysis.* 15:977–992.

Harding, S. E. 1984. An analysis of the heterogeneity of mucins. No evidence for a self association. *Biochem. J.* 219:1061–1064.

Harding, S. E., and J. M. Creeth. 1982. Self association, polydispersity and thermodynamic non-ideality in a cystic fibrotic glycoprotein. *IRCS (Int. Res. Commun. Syst.) Med. Sci. Libr. Compend.* 10:474–475.

Harding, S. E., and J. M. Creeth. 1983. Polyelectrolyte behavior in mucus glycoproteins. *Biochim. Biophys. Acta.* 746:114–119.

Harding, S. E., J. M. Creeth, and A. J. Rowe. 1983*a*. Modelling the conformation of mucus glycoproteins in solution. *Proc. 7th Int. Conf. Glyconjugates.* Olsson-Reklambyra, Sweden. 558–559.

Harding, S. E., A. J. Rowe, and J. M. Creeth. 1983*b*. Further evidence for a flexible and highly expanded spheroidal model for mucus glycoproteins in solution. *Biochem. J.* 209:893–896.

Johnson, M. L., J. J. Correia, D. A. Yphantis, and H. R. Halvorson. 1981. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* 36:575–588.

Kim, H., R. C. Deonier, and J. W. Williams. 1977. The investigation of self associating reactions by equilibrium ultracentrifugation. *Chem. Rev.* 77:659–690.

NAG Fortran Library manual. 1978. Numerical Algorithms Group, Oxford, United Kingdom.

Powell, M. J. D. 1970. A hybrid method for nonlinear algebraic equations. *In* Numerical Methods for Nonlinear Algebraic Equations. P. Rabinowitz, editor. Gordon and Breach Science Publishers, Inc., New York.

Provencher, S. W. 1979. Inverse problems in polymer characterization: direct analysis of polydispersity with photon correlation spectroscopy. *Makromol. Chem.* 180:201–209.

Rinde, H. 1928. The distribution of the sizes of particles in gold sols prepared according to the nuclear method. Ph.D. thesis. Uppsala, Sweden.

Roark, D., and D. A. Yphantis. 1969. Studies of self associating systems by equilibrium ultracentrifugation. *Ann. NY Acad. Sci.* 164:245–278.

Silberberg, A., and F. A. Meyer. 1982. The structure and physiology of gastrointestinal mucus. *In* Mucus in Health and Disease. Vol. II. E. N. Chantler, J. B. Elder, and M. Elstein, editors. Plenum Publishing Corp., New York. 115–133.

Teller, D. C. 1965. Ph.D. thesis. University of California, Berkeley, CA.

Teller, D. C. 1973. Characterization of proteins by sedimentation equilibrium in the analytical ultracentrifuge. *Methods Enzymol.* 27:346–441.

Tindall, S. H., and K. C. Aune. 1982. A general method for studying homogeneous and/or heterogeneous associations by sedimentation equilibrium. *Anal. Biochem.* 120:71–84.