Static Light Scattering from Concentrated Protein Solutions, I: General Theory for Protein Mixtures and Application to Self-Associating Proteins

Allen P. Minton
Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, Maryland

ABSTRACT Exact expressions for the static light scattering of a solution containing up to three species of point-scattering solutes in highly nonideal solutions at arbitrary concentration are obtained from multicomponent scattering theory. Explicit expressions for thermodynamic interaction between solute molecules, required to evaluate the scattering relations, are obtained using an equivalent hard particle approximation similar to that employed earlier to interpret scattering of a single protein species at high concentration. The dependence of scattering intensity upon total protein concentration is calculated for mixtures of nonassociating proteins and for a single self-associating protein over a range of concentrations up to 200 g/l. An approximate semiempirical analysis of the concentration dependence of scattering intensity is proposed, according to which the contribution of thermodynamic interaction to scattering intensity is modeled as that of a single average hard spherical species. Simulated data containing pseudo-noise comparable in magnitude to actual experimental uncertainty are modeled using relations obtained from the proposed semiempirical analysis. It is shown that by using these relations one can extract from the data reasonably reliable information about underlying weak associations that are manifested only at very high total protein concentration.

INTRODUCTION

The quantitative characterization of proteins in highly concentrated solution is of interest for at least two reasons, one biological and one biotechnological. Some proteins, such as hemoglobin or crystallins, are present at concentrations of hundreds of grams/liter in their normal milieux, and small changes in their self-interaction resulting from changes in environmental variables or covalent modification may result in the formation of pathological aggregates (1,2). When engineered antibodies are formulated as potential biopharmaceutical agents, attention must be paid to the effects of storage and administration in highly concentrated solution upon the association state of the protein and possible consequences for bioactivity and immunogenicity (3).

The characterization of weak associations leading to complex formation at high total protein concentration poses special problems to the experimenter. Weak attractive interactions are likely to be masked by large nonspecific repulsive interactions deriving from excluded volume. The formation of weakly associated complexes at high protein concentration, referred to as hidden associations (4), is revealed only when the contribution of repulsive, or nonideal, interactions to observed solution properties can be assessed explicitly. Techniques for detecting weak protein associations manifested only at high total concentration, based upon novel analysis of sedimentation equilibrium, have been developed by Chatelier and Minton (5,6), Rivas et al. (7), and Zorrilla et al. (8). Other physical-chemical techniques that have been used to characterize concentrated protein solutions include measurement of osmotic pressure (9,10) and static light scattering (11,12). Prior light-scattering investigations revealed that the dependence of scattering intensity upon total concentration in a solution containing a single species of nonassociating protein or a paucidisperse distribution of nonassociating proteins could be well accounted for by a simple model in which the proteins were represented by effective hard spheres, the size of which depended upon the magnitude of repulsive interactions acting between protein molecules in the solution (11–13).

The purpose of the present work is twofold. The first objective is to develop a more complete quantitative description of static light scattering in solutions containing multiple species of proteins at high total concentration, allowing for the possibility of protein equilibrium self-association. The second objective is to develop an approximate analysis of the dependence of light-scattering intensity upon total protein concentration that may be used to detect and qualitatively or semiquantitively characterize weak protein associations in these solutions, without prior knowledge of the composition of the solutions or detailed knowledge regarding the nature of repulsive nonideal interactions acting between individual protein species. The article is organized as follows. In the next section, multicomponent scattering theory is used to develop explicit expressions for the scattering intensity of a solution containing up to three solute species. These expressions are exact, but require specification of the thermodynamic interactions between each pair of scattering species. An effective hard particle model is then introduced that leads to explicit (but approximate) expressions for the thermodynamic interactions between scatterers required by multicomponent scattering theory. Next, a semiempirical analysis is introduced based upon the assumption that the effects of repulsive

Submitted January 4, 2007, and accepted for publication March 26, 2007.

Address reprint requests to Dr. Allen P. Minton, Tel.: 301-496-3604; E-mail: minton@helix.nih.gov.

Editor: Elliot L. Elson.

0006-3495/07/08/1321/08 $2.00

doi: 10.1529/biophysj.107.103895

nonideal interactions upon scattering intensity may be accounted for by a universal correction that is independent (to within an acceptable level of approximation) of the details of the underlying interactions. This assumption is tested by application to analysis of simulated noisy experimental data. The results of the analysis indicate that the approximation is reasonably robust, and permits recovery of information about weak complex formation embedded within the simulated data.

General relations from multicomponent scattering theory

We consider a solution containing multiple species of globular protein, all of which have a maximum dimension that is small relative to the wavelength of incident light (650–700 nm), and therefore behave as point scatterers (14). To simplify computation we shall further assume that all protein species have the same specific refractive increment $d\tilde{n}/dw$, where $\tilde{n}$ denotes the refractive index of the solution and w the w/v concentration of protein. (Extension to mixtures of scattering species with significantly different values of $d\tilde{n}/dw$, such as a solution containing a protein and a nucleic acid, is theoretically straightforward but computationally tedious.) Most proteins that are neither highly glycosylated nor lipidated have the same specific refractive increment (~0.185 cm^3/g) to within reasonable precision (15), and if the different scattering species considered here correspond to different association states of a single protein, then the assumption of equal specific refractive increment is automatically valid. For such a solution, the theory of Rayleigh scattering from multicomponent solutions (14,16) yields

$$\frac{R}{K_o} = \left(\frac{\tilde{n}}{\tilde{n}_o}\right)^2 \sum_{i,j} M_i M_j \langle \Delta c_i \Delta c_j \rangle, \tag{1}$$

where R denotes the (angle-independent) Rayleigh ratio, $\tilde{n}_o$ and $\tilde{n}$ the refractive index of solvent (buffer) and solution, respectively, M_i the molar mass of the i^{th} scattering species, and $\langle \Delta c_i\ \Delta c_j \rangle$ the mean product of the fluctuations of the molar concentrations of the i^{th} and j^{th} scattering species about their respective equilibrium values. R is scaled to an optical constant independent of solution properties, given by

$$K_o = \frac{4\pi^2 \tilde{n}_o^2 (d\tilde{n}/dw)^2}{\lambda_o^4 N_A}, \tag{2}$$

where λ_o denotes the wavelength (in vacuum) of the incident light, and N_A Avogadro's number. Over a broad range of concentration, the refractive index of the solution is well described by a linear function of total w/v protein concentration,

$$\tilde{n} = \tilde{n}_o + \frac{d\tilde{n}}{dw} w_{tot}, \tag{3}$$

where w_i denotes the w/v concentration of the i^{th} species, and $w_{tot} = \sum_i w_i$(17).

For a solution of up to three scattering species, Eq. 1 expands to

$$\frac{R}{K_o} = \left(\frac{\tilde{n}}{\tilde{n}_o}\right)^2 \left[M_1^2\langle \Delta c_1^2 \rangle + M_2^2 \langle \Delta c_2^2 \rangle + M_3^2 \langle \Delta c_3^2 \rangle + 2M_1M_2\langle \Delta c_1 c_2 \rangle + 2M_2M_3\langle \Delta c_2 c_3 \rangle + 2M_1M_3\langle \Delta c_1 c_3 \rangle\right]. \tag{4}$$

Application of concentration fluctuation theory (14) to a mixture of three species, indexed by i, j, and k yields

$$\langle \Delta c_i^2 \rangle = c_i \frac{(1+\zeta_{jj})(1+\zeta_{kk}) - \zeta_{kj}\zeta_{jk}}{D}, \tag{5}$$

$$\langle \Delta c_i \Delta c_j \rangle = \frac{c_k \zeta_{ki}\zeta_{kj} - c_i \zeta_{ij}(1+\zeta_{kk})}{D}, \tag{6}$$

$$D = (1+\zeta_{11})(1+\zeta_{22})(1+\zeta_{33}) + 2\zeta_{12}\zeta_{21}\zeta_{31} - (1+\zeta_{11})\zeta_{32}\zeta_{23} - (1+\zeta_{22})\zeta_{31}\zeta_{13} - (1+\zeta_{33})\zeta_{21}\zeta_{12}, \tag{7}$$

where

$$\zeta_{ij} = c_j \frac{\partial \ln \gamma_i}{\partial c_j} = w_j \frac{\partial \ln \gamma_i}{\partial w_j} = \rho_j \frac{\partial \ln \gamma_i}{\partial \rho_j} \tag{8}$$

and γ_i and ρ_i denote, respectively, the thermodynamic activity coefficient and number density of the i^{th} species. The reader may verify that in the limit of low concentration, i.e., as w_i and ζ_{ij} tend toward zero for all i and j, Eqs. 4–8 reduce to the classical result obtained for ideal solutions (18)

$$\frac{R}{K_o} = \sum_i M_i w_i = M_w w_{tot}, \tag{9}$$

where M_w denotes the weight average molar mass $\sum_i M_i w_i / w_{tot}$, and that for a single species at arbitrary concentration, Eqs. 4–8 reduce to the textbook result (18)

$$\frac{R}{K_o} = \left(\frac{\tilde{n}}{\tilde{n}_o}\right)^2 \frac{M_w}{1 + w\dfrac{\partial \ln \gamma}{\partial w}}. \tag{10}$$

Scattering from a solution of multiple protein species modeled as hard spherical particles

Rationale and justification for hard sphere model

The effective interaction or potential of mean force between two globular protein molecules in solution may be represented as a sum of two contributions: a hard repulsive interaction due to mutual impenetrability (excluded volume), and a soft interaction reflecting forces between the molecules at distances greater than steric contact. While the hard interaction may be assumed to be independent of solution conditions so long as the protein retains its native three-dimensional structure, the soft interaction may be either attractive or repulsive depending upon the relative strengths of electrostatic and hydrophobic contributions under a particular set of experimental conditions (e.g., pH, buffer composition, temperature) (19). Experimental studies of protein solutions at high

concentration carried out under conditions such that soft interactions are largely damped out have revealed that the magnitude of the hard excluded volume interaction between proteins may be estimated semiquantitatively using approximate equations of state for fluids of hard convex particles, in which each protein species is represented by an equivalent hard convex particle of size and shape comparable to that of the actual molecule (7,20). Although such equations of state have been developed for fluids of mixtures of arbitrarily shaped convex particles (21,22), it has been found that so long as a protein is reasonably compact and quasi-spherical (i.e., the largest dimension exceeds the smallest dimension by less than a factor of ~2), it may be represented by a hard spherical particle of similar volume without introduction of major quantitative error (19). We therefore calculate the values of ζ_{ij} appearing in Eqs. 5–7 using expressions given in the Appendix, with geometric constants appropriate for hard sphere fluid mixtures (21–23).

It has been pointed out (24) that even though the intensity of light directly scattered by small cosolutes such as salts is negligible relative to that scattered by macrosolutes, the multicomponent theory of light scattering (14,16) predicts that thermodynamic interactions between small cosolutes and macrosolutes may affect the concentration fluctuations of macrosolutes and thus indirectly influence the dependence of light scattering upon macrosolute concentration. The results of test calculations (available upon request) indicate that the presence of 0.15 M NaCl is likely to produce at most a slight effect on the scattering of typical proteins, and that such an effect may be readily accommodated within the context of the effective hard particle model used here to model the nonideal contribution to the concentration dependence.

Single nonassociating protein species

In Fig. 1, the normalized scattering intensity is plotted as a function of the w/v concentration of a 70,000 MW nonassociating protein, modeled as a hard sphere of specific volume 0.73 cm^3/g. (For brevity, protein molar masses will subsequently be written in units of 1000, e.g., 70 K.) Also plotted for comparison are the ideal scattering, calculated according to

$$\frac{R}{K_o} = \left(\frac{\tilde{n}}{\tilde{n}_o}\right)^2 M_w, \tag{11}$$

and the scattering calculated taking into account only the first-order correction for thermodynamic nonideality (second virial coefficient for hard spheres),

$$\frac{R}{K_o} = \left(\frac{\tilde{n}}{\tilde{n}_o}\right)^2 \frac{M_w}{1 + 8v_{eq}w}, \tag{12}$$

where v_{eq} denotes the specific volume of the equivalent sphere in inverse w/v concentration units. We note that actual scattering deviates significantly from ideal scattering (Eq. 11) at concentrations exceeding ~20 g/l, and deviates significantly from that predicted by the first-order correction to ideality (Eq. 12) at concentrations exceeding ~50 g/l.

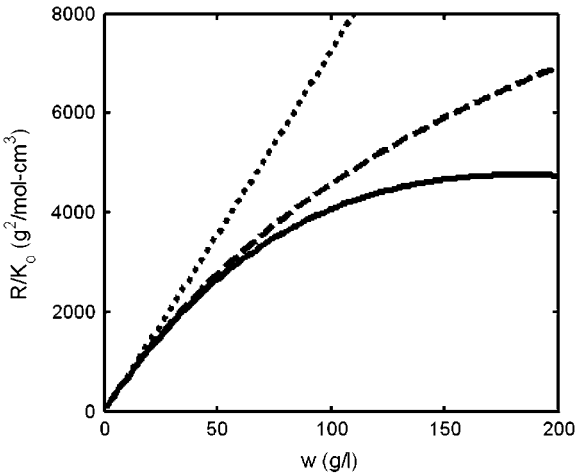

FIGURE 1 Dependence of normalized scattering intensity upon concentration of a monomeric protein (molar mass $M = 70$ K) calculated with full accounting for excluded volume effects as described in text (*solid line*), calculated with a first-order correction for excluded volume (*dashed line*), and calculated with no correction for excluded volume (*dotted line*).

Nonassociating mixture of proteins

In Fig. 2, normalized scattering intensity is plotted as a function of total protein concentration for several solutions of different composition with identical weight-average molar mass. At low concentrations, the concentration dependence of all solutions is identical, but differences in composition result in a 15–20% variation in scattering intensity at total w/v concentrations approaching 200 g/l.

Self-associating protein

We consider the case of a single protein A of molar mass M_1 that may reversibly self-associate to form one or more oligomers A_i with thermodynamic equilibrium association constants given by

$$K_i^o \equiv \frac{a_i}{a_1^i} = \frac{\gamma_i c_i}{\gamma_1^i c_1^i}, \tag{13}$$

where a_i, γ_i, and c_i denote the thermodynamic activity, activity coefficient, and molar concentration of i-mer, respectively. It follows from Eq. 13 that the concentrations of monomer and i-mer are related by an apparent equilibrium constant

$$K_i = \frac{c_i}{c_1^i} = K_1^o \frac{\gamma_1^i}{\gamma_i}. \tag{14}$$

In a concentrated solution, such that a substantial fraction of solution volume is occupied by protein, steric exclusion results in values of γ_1 and γ_i that are significantly greater

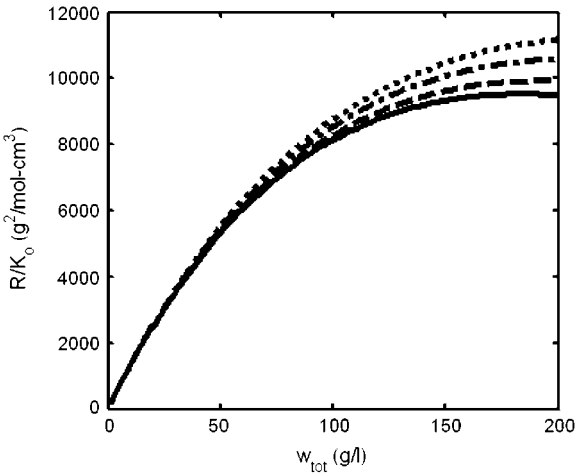

FIGURE 2 Dependence of normalized scattering intensity upon concentration of protein solutions of different composition but identical weight-average molar mass ($M_w = 140$ K). (*Solid line*) A single protein species with $M = 140$ K. (*Dot-dashed line*) A mixture of proteins with $M_1 = 70$ K and $M_2 = 280$ K in the mass ratio 2:1. (*Dotted line*) A mixture containing proteins with $M_1 = 70$ K and $M_2 = 420$ K in the mass ratio 4:1. (*Dashed line*) A mixture containing proteins with $M_1 = 70$ K, $M_2 = 140$ K, and $M_3 = 210$ K in the mass ratio 1:1:1.

than unity, and depending upon the association scheme, the value of K_i may differ significantly from K_i^o (25). Therefore, under highly nonideal conditions, the composition of a solution containing monomer in equilibrium with one or more oligomers should, in principle, be calculated in an iterative fashion, as described previously by Chatelier and Minton (5) and Snoussi and Halle (26). In the present work, the equilibrium solution composition corresponding to a given total w/v protein concentration is calculated as follows:

1. A value of v_{eq} is specified.
2. The value of c_{tot} is set equal to w_{tot}/M_1.
3. Initial values of the K_i are set equal to the K_i^o.
4. The equation of conservation of mass

$$c_1 + i\sum_i K_i c_1^i = c_{tot} \tag{15}$$

is solved numerically for c_1, and the remaining c_i are then calculated via Eq. 14.

5. The values of the $\ln\gamma_i$ are calculated from the c_i and v_{eq} as described in the Appendix, representing each species by a hard sphere with volume proportional to mass.
6. The apparent equilibrium constants are recalculated according to

$$\ln K_i = \ln K_i^o + i\ln\gamma_i - \ln\gamma_i. \tag{16}$$

Steps 4–6 are repeated iteratively until the values of all c_i converge. Our criterion of convergence is that between successive iterations the concentrations of all species remain constant to within one part in 1000. The weight-average molar mass may then be calculated according to

$$M_w = \frac{\sum_i c_i M_i^2}{\sum_i c_i M_i}. \tag{17}$$

In Fig. 3, normalized scattering intensity is plotted as a function of total w/v protein concentration for three self-association schemes and various combinations of equilibrium association constants within each scheme. It is evident that even at very high concentrations, the scattering intensity is sensitive to changes in solution composition resulting from concentration-dependent self-association, and that, at least in principle, information about weak self-associations that are manifested only at high total protein concentration may be obtained from analysis of an experimental measurement of the dependence of scattering intensity on total protein concentration.

A simple semiempirical approximation for inclusion of nonideal effects on scattering

We define an experimentally measurable quantity called the apparent weight-average molar mass,

$$M_{w,app} \equiv \frac{R}{K_o w_{tot}\left(\frac{\tilde{n}}{\tilde{n}_o}\right)^2}, \tag{18}$$

which becomes equal to the true weight-average molar mass in the ideal (low concentration) limit. In Fig. 4, we plot the

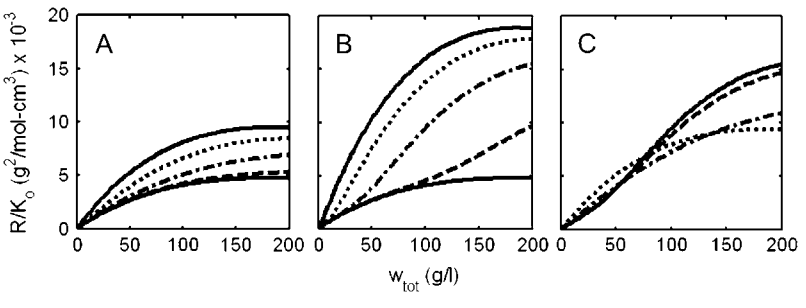

FIGURE 3 Dependence of normalized scattering intensity upon total protein concentration in solutions of a self-associating protein with $M_1 = 70$ K. (*A*) Monomer-dimer. (*Lower solid curve*) Pure monomer; (*upper solid curve*) pure dimer; (*dashed curve*) $K_2 = 10\ M^{-1}$; (*dot-dashed curve*) $K_2 = 100\ M^{-1}$; (*dotted curve*) $K_2 = 1000\ M^{-1}$. (*B*) Monomer-tetramer. (*Lower solid curve*) Pure monomer; (*upper solid curve*) pure tetramer; (*dashed curve*) $K_4 = 10^6\ M^{-3}$; (*dot-dashed curve*) $K_4 = 10^8\ M^{-3}$; (*dotted curve*) $K_4 = 10^{10}\ M^{-3}$. (*C*) Monomer-dimer-tetramer. (*Solid curve*) $K_2 = 0$, $K_4 = 10^8\ M^{-3}$; (*dashed curve*) $K_2 = 10^2\ M^{-1}$, $K_4 = 10^8\ M^{-3}$; (*dot-dashed curve*) $K_2 = 10^3\ M^{-1}$, $K_4 = 10^8\ M^{-3}$; (*dotted curve*), $K_2 = 10^4\ M^{-1}$, $K_4 = 10^8\ M^{-3}$.

value of the ratio $M_w/M_{w,app}$ calculated as a function of w_{tot} for each of the self-association schemes used to create Fig. 3. Although the shapes of these plotted curves depend in detail upon the composition of the scattering species, and can vary by 10–20% at total protein concentrations approaching 200 g/l, the similarity between them suggests that a single universal function, denoted by Z, may be a useful approximation to the actual value of $M_w/M_{w,app}$. It follows that

$$M_{w,app}(w_{tot}) \approx M_w(w_{tot})/Z(w_{tot}). \quad (19)$$

The form of Z is further suggested to be that of the exact ratio for a single species, calculated from (10)

$$Z = 1 + w\frac{\partial \ln\gamma}{\partial w}. \quad (20)$$

The value of this function is known quite accurately for a fluid of identical hard spheres over a broad range of concentration (11), and we propose that it be used as an approximate universal ratio between actual and apparent weight-average molar masses,

$$Z(\varphi) = 1 + 8\varphi + 30\varphi^2 + 73.4\varphi^3 + 141.2\varphi^4 + 238.5\varphi^5 + 395.4\varphi^6, \quad (21)$$

where ϕ denotes an effective fraction of occupied volume, the value of which is calculated according to

$$\varphi = v_{eq}w_{tot}. \quad (22)$$

In this formulation, v_{eq} becomes a semiempirical parameter, the value of which is to be estimated by modeling as described below. Since the volume excluded by one macromolecule to another macromolecule must exceed the volume excluded by that macromolecule to solvent (27), we obtain the lower bound,

$$\varphi = \varphi_{min} \equiv \bar{v}w_{tot}, \quad (23)$$

where $\bar{v}$ denotes the partial specific volume of solute, approximately equal to 0.73 cm^3/g for most polypeptides (Appendix 2 of (28)). Combination of Eqs. 19 and 23 yields the following lower bound to the weight-average molar mass:

$$M_w \geq M_{w,min} \equiv M_{w,app}Z(\varphi_{min}). \quad (24)$$

Testing the semiempirical approximation via analysis of simulated data

On the basis of results of simulations of sedimentation equilibrium in concentrated solutions of multiple macrosolute species, Chatelier and Minton (5) suggested that the effect of repulsive nonideal interactions between multiple solute species upon solution properties may be modeled by an effective single species approximation. Subsequently, Muramatsu and Minton (4) employed this approximation to analyze the results of sedimentation equilibrium experiments carried out on self-associating proteins over a range of concentrations extending to 200 g/l. The validity of the approximate analysis of Muramatsu and Minton was later confirmed by a rigorous analysis of the same data carried out by Zorrilla et al. (8).

Although experimental methods for semiautomated measurement of scattering of protein solutions over a broad concentration range are currently under development in our laboratory, actual scattering data on nonassociating mixtures of proteins and self-associating proteins over a sufficiently broad range of concentration are still lacking as of the date of writing. We therefore test the approximate analysis suggested in the preceding section by generating synthetic data sets of normalized scattering intensity as a function of concentration over a broad concentration range (0.2–200 g/l, in equal logarithmic increments), utilizing the full treatment of nonideality described above and in the Appendix. A value of $v_{eq} = 1.0$ cm^3/g (27) was used for all simulations. Normally distributed pseudo-noise, with a standard deviation equal to 3% of the theoretical value, was added to the calculated scattering to simulate experimental uncertainty of measurement (29).

Nonassociating protein mixtures

Sets of experimental data were generated for four solutions with different composition and equal weight-average molar mass, as described in the caption to Fig. 2. The data set for one mixture is plotted in Fig. 5, and may be compared with the corresponding generating function plotted in Fig. 2. Using Eqs. 3 and 18, each data set was transformed to an equivalent data set of the form $\{\log w_{tot}, M_{w,app}\}$, plotted in Fig. 6, *A–D*. Values of $M_{w,min}$ were calculated according to Eq. 24 and are plotted together with the data in each panel. It is noted that in no case does $M_{w,min}$ exceed the limiting value

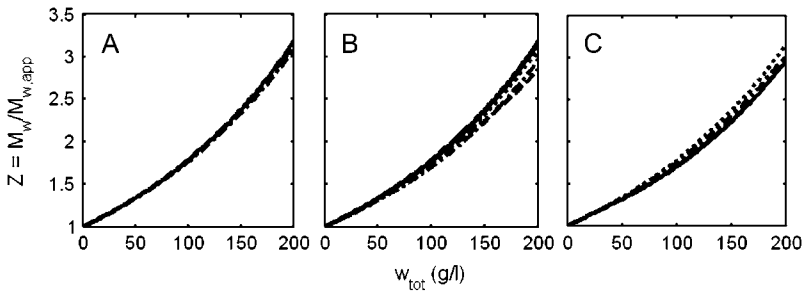

FIGURE 4 Dependence of the ratio $M_w/M_{w,app}$ upon total concentration of protein in solutions of a self-associating protein with $M_1 = 70$ K. Reaction schemes and association constants within a given reaction scheme are the same as given in the caption to Fig. 3.

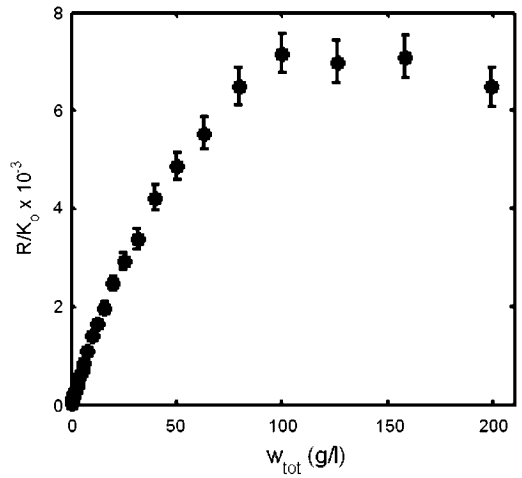

FIGURE 5 Simulated experimental dependence of R/K_o upon w_{tot} for a solution containing an equal amount (by weight) of three proteins with molar masses 70 K, 140 K, and 210 K. Error bars correspond to ± 2 SD.

at low concentration, indicating that it may be possible to account for the data without postulating concentration-dependent association. This was confirmed by fitting Eq. 19 to each of the data sets by nonlinear minimization of χ-square to determine best-fit values of M_w (assumed independent of concentration) and v_{eq}. Curves calculated using best-fit parameter values given in the figure caption are also plotted in each panel.

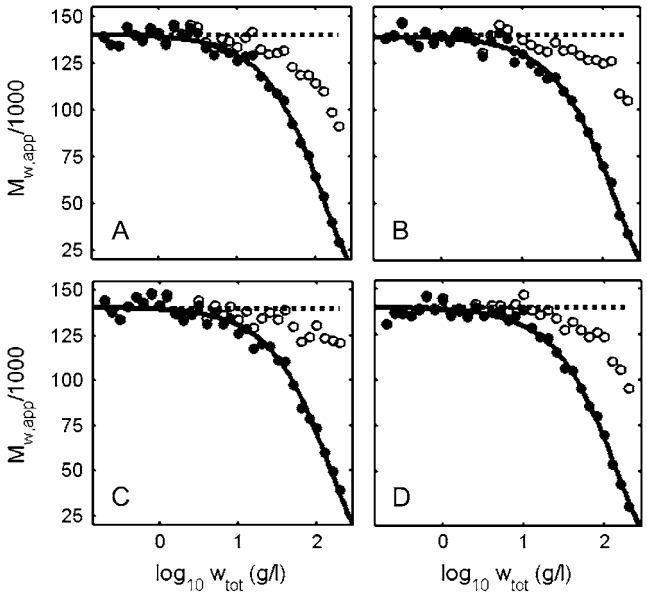

FIGURE 6 Simulated experimental dependence of $M_{w,app}$ (●) and calculated dependence of $M_{w,min}$ (○) upon the logarithm of w_{tot} for the following solutions of nonassociating proteins with $M_w = 140$ K: (*A*) 140 K; (*B*) 70:280 K 2:1; (*C*) 70:420 K 4:1; and (*D*) 70:140:210 K 1:1:1. Solid curves are calculated using Eq. 18 with the following best-fit parameter values: (*A*) $M_w = 140$ K, $v_{eq} = 1.01$. (*B*) $M_w = 139$ K, $v_{eq} = 0.90$. (*C*) $M_w = 140$ K, $v_{eq} = 0.88$. (*D*) 140 K, $v_{eq} = 0.94$.

Self-associating proteins

Association schemes and equilibrium constants used to generate synthetic data sets simulating the scattering behavior of aldolase and ovalbumin are specified in Table 1. These values were taken from Muramatsu and Minton (4), who found that the sedimentation equilibrium of aldolase and ovalbumin up to 200 g/l could be accounted for quantitatively by either of two self-association schemes. To avoid biasing our modeling results in favor of either scheme, we generated synthetic data of the form $\{w_{tot,}\, R/K_o\}$ using each association scheme and then averaged them to obtain the final synthetic data sets plotted in Fig. 7. Using Eq. 18, the simulated data were transformed into equivalent sets of data of the form $\{\log w_{tot}, M_{w,app}\}$. These data are plotted in Fig. 8, along with the corresponding values of $M_{w,min}$ calculated according to Eqs. 23 and 24. The monotonically increasing character of $M_{w,min}$ with concentration clearly indicates that self-association must be invoked to account for the observed concentration dependence of scattering of either protein.

Equation 19, with concentration-dependent M_w calculated via Eq. 17 together with several association schemes, was fitted to each data set via nonlinear minimization of χ^2. Inherent in this approximate analysis is the assumption that the various equilibrium constants are independent of concentration to within an acceptable level of precision (i.e., that $K_i \approx K_i^o$ over the concentration range considered). Then solution composition may be obtained as a function of total protein concentration by simple nonrecursive solution of Eq. 15. Given the concentration-dependent value of M_w, approximate Eq. 19 is used to calculate the concentration-dependent value of $M_{w,app}$. The validity of these assumptions and the utility of the approximate analysis are demonstrated by the excellent agreement, shown in Table 1, between generating schemes and parameter values on the one hand, and the corresponding schemes and parameter values obtained via approximate analysis on the other hand.

SUMMARY AND CONCLUSION

The utility of the approximate analysis introduced here depends upon the validity of the underlying effective hard particle approximation as a description of repulsive nonideal interactions between protein molecules. Although subject to some restrictions, when experiments are carried out under appropriate conditions (19), the effective hard particle model has provided a simple, accurate, and useful description of the behavior of a variety of protein solutions at high total concentration (reviewed in (27) and (19)).

It is emphasized that to extract the maximum amount of information from the concentration dependence of light scattering, the acquisition of an equal density of data along the logarithmic concentration axis is essential, so that data obtained at both very low and very high concentrations receive equal statistical weight. Thus an experiment should be

TABLE 1 Comparison of equilibrium association schemes and parameters used to generate synthetic data sets, and association schemes and parameters obtained via the approximate analysis described in the text

		Generating parameters			Best-fit parameters		
Protein	Self-association scheme	M_1/1000	log K_i^o [$M^{-(i-1)}$]	v_{eq} (cm^3/g)	M_1/1000	log K_i [$M^{-(i-1)}$]	v_{eq} (cm^3/g)
Aldolase	Monomer-dimer	163	2.8	1.0	165 ± 3	2.8 ± 0.2	0.65*
	Monomer-trimer	169	6.1	1.0	170 ± 3	6.1 ± 0.2	0.97
Ovalbumin	Monomer-trimer	46	5.8	1.0	46.5 ± 1	5.9 ± 0.1	0.90
	Monomer-dimer-tetramer	44.5	K_2: 2.7	1.0	46 ± 2	K_2: 2.5 (−0.4,+0.3)	1.09
			K_4: 8.9			K_4: 9.0 ± 0.2	

Generating parameters (excepting v_{eq}) are taken from Muramatsu and Minton (4). Indicated uncertainty in values of best-fit parameters correspond to ±1 standard error of estimate, as determined by the distribution of χ^2.
*This value is within fitting uncertainty of the value of 0.73 cm^3/g used to calculate $M_{w,min}$ (Fig. 8, *left-hand panel*).

designed so that the concentration increases (or decreases) by an equal fraction rather than by an equal amount between samples. The ability to automatically prepare and deliver concentration gradients for light-scattering measurements has recently been developed and utilized to characterize reversible self-associations in dilute protein solution (29). We anticipate that in combination with the analysis presented here, this technology will prove useful in the detection and characterization of weak self-association in concentrated protein solutions as well.

APPENDIX: CALCULATION OF ACTIVITY COEFFICIENTS AND CONCENTRATION DERIVATIVES OF ACTIVITY COEFFICIENTS IN A FLUID MIXTURE OF HARD CONVEX PARTICLES

Equivalent relations have been presented previously by Chatelier and Minton (5). They are reproduced here for convenience, notational consistency, and to correct a typographical error in Eq. A5 of the cited reference.

Consider a fluid containing multiple species of hard convex particles. Let the number density (N/cm^3) of the i^{th} species be denoted by $\rho_i = c_i N_A/1000$. The size and shape of the i^{th} species of convex particles determine the values of a characteristic length r_i expressed in units of centimeters, and three unitless shape descriptors h_i, s_i, and v_i (21). Boublik (22) derived the following relation for the activity coefficient of the ith species in this fluid mixture,

$$\ln\gamma_i = -\ln(1-\langle V\rangle) + \frac{H_i\langle S\rangle + S_i\langle H\rangle + V_i\langle 1\rangle}{1-\langle V\rangle} + \frac{H_i^2\langle S\rangle^2 + 2V_i\langle H\rangle\langle S\rangle}{2(1-\langle V\rangle)^2} + \frac{V_i\langle H^2\rangle\langle S\rangle^2}{3(1-\langle V\rangle)^3}, \quad (25)$$

where $H_i = h_i r_i$, $S_i = s_i r_i^2$, $V_i = v_i r_i^3$, $\langle X\rangle \equiv \sum_i \rho_i X_i$, and $\langle 1\rangle \equiv \sum_i \rho_i$. The values S_i and V_i denote the surface area and volume of the i^{th} species of particle, and H_i denotes an effective radius defined by Kihara (30) to be equal to one-half of the length of the projection of the particle onto a single directional axis, averaged over all particle orientations relative to that axis. This relation may be simplified if all species are assumed to be represented by identically shaped particles, that is, for all i, $h_i = h$, $s_i = s$, and $v_i = v$. Then

$$\langle H\rangle = h\sum_i \rho_i r_i \quad (26a)$$

$$\langle S\rangle = s\sum_i \rho_i r_i^2 \quad (26b)$$

$$\langle V\rangle = v\sum_i \rho_i r_i^3 \quad (26c)$$

and

$$\langle H\rangle^2 = h^2\sum_i \rho_i r_i^2. \quad (26d)$$

Defining $Q \equiv 1 - \langle V\rangle$, we obtain the partial derivatives

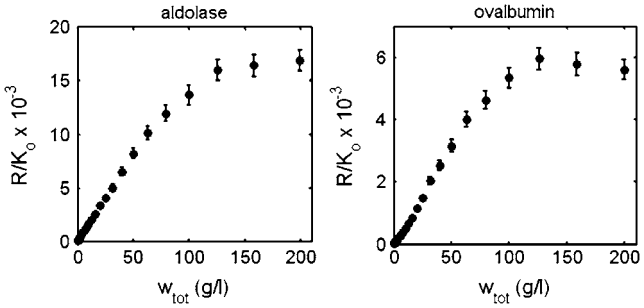

FIGURE 7 Simulated experimental concentration dependence of light scattering of aldolase and ovalbumin solutions, calculated as described in the text. The data set for each protein is the mean of two data sets calculated using each of the alternate reaction schemes and parameter values specified in Table 1. Error bars correspond to ±2 SD.

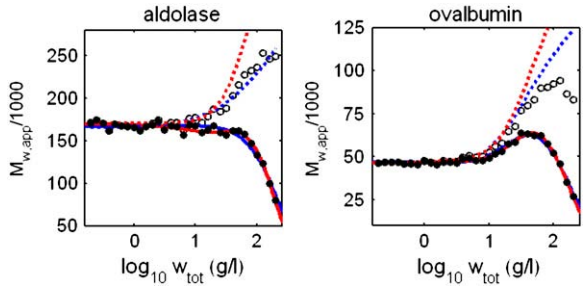

FIGURE 8 Simulated experimental dependence of $M_{w,app}$ (●) and $M_{w,min}$ (○) upon the logarithm of w_{tot} for solutions of aldolase (*left panel*) and ovalbumin (*right panel*), calculated by application of Eqs. 18 and 23, respectively, to the data sets plotted in Fig. 6. Solid curves are calculated using the CM approximation with best-fit equilibrium models and parameter values specified in Table 1. Dashed curves indicate estimates of actual M_w calculated using the corresponding best-fit parameter values. Best fits of different association schemes are distinguished by color as follows. Aldolase: monomer-dimer (*blue*) and monomer-trimer (*red*). Ovalbumin: monomer-trimer (*blue*) and monomer-dimer-tetramer (*red*).

$$\frac{\partial \ln \gamma_i}{\partial \rho_j} = \frac{V_i}{Q} + h\left[\frac{S_i}{Q} + \frac{V_i\langle S\rangle}{Q^2}\right] r_j + \left[s\left(\frac{H_i}{Q} + \frac{H_i^2\langle S\rangle + V_i\langle H\rangle}{Q^2} + \frac{2V_i\langle H^2\rangle\langle S\rangle}{3Q^2}\right) + h^2\frac{V_i\langle S\rangle^2}{3Q^3}\right] r_j^2 + v\left[\frac{1}{Q} + \frac{H_i\langle S\rangle + S_i\langle H\rangle + V_i\langle 1\rangle}{Q^2} + \frac{H_i^2\langle S\rangle^2 + 2V_i\langle H\rangle\langle S\rangle}{Q^3} + \frac{V_i\langle H^2\rangle\langle S\rangle^2}{Q^4}\right] r_j^3. \quad (27)$$

If it is further assumed that all species are spherical, then $h = 1$, $s = 4\pi$, and $v = 4\pi/3$. The value of r_i is estimated from the molar mass M_i and the specific exclusion volume v_{eq}, usually specified in units of cm^3/g:

$$r_i = \left(\frac{v_{eq}M_i}{N_A v}\right)^{1/3}. \quad (28)$$

The author thanks Drs. Peter McPhie (National Institutes of Health) and Daniel Some (Wyatt Technology) for helpful comments on preliminary drafts of this report.

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases.

REFERENCES

1. Ferrone, F. A., and M. A. Rotter. 2004. Crowding and the polymerization of sickle hemoglobin. *J. Mol. Recog.* 17:497–504.
2. Ponce, A., C. Sorensen, and L. Takemoto. 2006. Role of short-range protein interactions in lens opacifications. *Mol. Vis.* 12:879–884.
3. Shire, S. J., Z. Shahrokh, and J. Liu. 2004. Challenges in the development of high protein concentration formulations. *J. Pharm. Sci.* 93: 1390–1402.
4. Muramatsu, N., and A. P. Minton. 1989. Hidden self-association of proteins. *J. Mol. Recog.* 1:166–171.
5. Chatelier, R. C., and A. P. Minton. 1987. Sedimentation equilibrium in macromolecular solutions of arbitrary concentration. I. Self-associating proteins. *Biopolymers.* 26:507–524.
6. Chatelier, R. C., and A. P. Minton. 1987. Sedimentation equilibrium in macromolecular solutions of arbitrary concentration. II. Two protein components. *Biopolymers.* 26:1097–1113.
7. Rivas, G., J. A. Fernandez, and A. P. Minton. 1999. Direct observation of the self-association of dilute proteins in the presence of inert macromolecules at high concentration via tracer sedimentation equilibrium: theory, experiment, and biological significance. *Biochemistry.* 38:9379–9388.
8. Zorrilla, S., M. Jiménez, P. Lillo, G. Rivas, and A. P. Minton. 2004. Sedimentation equilibrium in a solution containing an arbitrary number of solute species at arbitrary concentrations: theory and application to concentrated solutions of ribonuclease. *Biophys. Chem.* 108:89–100.
9. Prouty, M. S., A. N. Schechter, and V. A. Parsegian. 1985. Chemical potential measurements of deoxyhemoglobin S polymerization. Determination of the phase diagram of an assembling protein. *J. Mol. Biol.* 184:517–528.
10. Magid, A. D., A. K. Kenworthy, and T. J. McIntosh. 1992. Colloid osmotic pressure of steer crystallins: implications for the origin of the refractive index gradient and transparency of the lens. *Exp. Eye Res.* 55:615–627.
11. Minton, A. P., and H. Edelhoch. 1982. Light scattering of bovine serum albumin solutions: extension of the hard particle model to allow for electrostatic repulsion. *Biopolymers.* 21:451–458.
12. Xia, J., Q. Wang, S. Tatarkova, T. Aerts, and J. Clauwaert. 1996. Structural basis of eye lens transparency: light scattering by concentrated solutions of bovine α-crystallin proteins. *Biophys. J.* 71:2815–2822.
13. Xia, J., T. Aerts, K. Donceel, and J. Clauwaert. 1994. Light scattering by bovine α-crystallin proteins in solution: hydrodynamic structure and interparticle interaction. *Biophys. J.* 66:861–872.
14. Stacey, K. A. 1956. Light-Scattering in Physical Chemistry. Academic Press, New York.
15. Theisen, A., C. Johann, M. P. Deacon, and S. E. Harding. 2000. Refractive Increment Data-Book. Nottingham University Press, Nottingham, UK.
16. Stockmayer, W. H. 1950. Light scattering in multi-component systems. *J. Chem. Phys.* 18:58–61.
17. Barer, R., and S. Tkaczyk. 1954. Refractive index of concentrated protein solutions. *Nature.* 173:821–822.
18. Tanford, C. 1961. Physical Chemistry of Macromolecules. Wiley & Sons, New York.
19. Hall, D., and A. P. Minton. 2003. Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochem. Biochim. Biophys. Acta.* 1649:127–139.
20. Minton, A. P. 1983. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol. Cell. Biochem.* 55:119–140.
21. Gibbons, R. M. 1969. The scaled particle theory for particles of arbitrary shape. *Mol. Phys.* 17:81–86.
22. Boublík, T. 1974. Statistical thermodynamics of convex molecule fluids. *Mol. Phys.* 27:1415–1427.
23. Lebowitz, J. L., E. Helfand, and E. Praestgaard. 1965. Scaled particle theory of fluid mixtures. *J. Chem. Phys.* 43:774–779.
24. Minton, A. P. 1981. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers.* 20:2093–2120.
25. Asthagiri, D., A. Paliwal, D. Abras, A. M. Lenhoff, and M. E. Paulaitis. 2005. A consistent experimental and modeling approach to light-scattering studies of protein-protein interactions in solution. *Biophys. J.* 88: 3300–3309.
26. Snoussi, K., and B. Halle. 2005. Protein self-association induced by macromolecular crowding: a quantitative analysis by magnetic relaxation dispersion. *Biophys. J.* 88:2855–2866.
27. Zimmerman, S. B., and A. P. Minton. 1993. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* 22:27–65.
28. Attri, A. K., and A. P. Minton. 1983. An automated method for determination of the molecular weight of macromolecules via sedimentation equilibrium in a preparative ultracentrifuge. *Anal. Biochem.* 133: 142–152.
29. Attri, A. K., and A. P. Minton. 2005. New methods for measuring macromolecular interactions in solution via static light scattering: basic methodology and application to nonassociating and self-associating proteins. *Anal. Biochem.* 337:103–110.
30. Kihara, T. 1953. Virial coefficients and models of molecules in gases. *Rev. Mod. Phys.* 25:831–843.