Biophysical Chemistry 24 (1986) 47–52
Elsevier

BPC 01053

AVERAGE QUANTITIES ACCESSIBLE TO THE ANALYSIS OF SEDIMENTATION EQUILIBRIUM DATA FOR SELF-ASSOCIATION SYSTEMS

Enrico BUCCI

Department of Biological Chemistry, University of Maryland, School of Medicine, Baltimore, MD 21201, U.S.A.

Received 1st October 1985
Accepted 1st February 1986

Key words: Self-association; Sedimentation equilibrium; Hemoglobin; Nonideality terms; (Human)

This paper shows that analysis of sedimentation equilibrium data, searching for average molecular weights, gives quantities which are dependent on the total protein concentration of the samples, while knowledge of the molecular weight of the monomeric subunits allows a more meaningful search for the concentrations of the individual polymeric components of the system. From these the various average molecular weights can be construed, and the various dissociation equilibrium constants evaluated. Also, in this paper considerations are proposed on the meaning of nonideality terms in associating systems and possible ways for estimating them. As an example the proposed procedures have been applied to measurements of sedimentation equilibrium in carbonmonoxyhemoglobin.

1. Introduction

The correlation between subunit dissociation properties and ligand-binding affinity is a relevant parameter of allosteric regulation. Recently, it was shown that also the excluded volume properties of hemoglobin, and of sickle cell hemoglobin in particular, play a relevant role in the pathogenesis of certain diseases [1]. Therefore, techniques capable of providing this information are becoming more and more relevant as a tool for biological research. Analytical ultracentrifugation is one of the few techniques which can provide this information.

Sedimentation equilibrium in the ultracentrifuge always appeared to have a great potential for studying dissociation equilibria in proteins; for this reason the theory and practice of its usage have been very well investigated and perfected [2–25]. The theory shows that in the parameters describing the sedimentation equilibrium properties there is a strong correlation between the molecular weight of the monomeric subunits and the protein concentration of the sample. The practical impossibility of separating the two parameters in independent terms limited the analytical power of the technique and forced the investigators to use approaches based on determinations of average molecular weights.

In recent times, techniques have been developed capable of giving the exact amino acid molecular weight of the individual subunits of protein molecules. The techniques are based on a variety of approaches, which, however, denature the samples so that the associating properties of the system are lost. Nevertheless, the availability of the true amino acid molecular weight of the individual subunits provides the investigators with a critical parameter, which simplifies the analysis of sedimentation equilibrium data.

This paper shows that analysis of sedimentation equilibrium data, searching for average molecular weights, gives quantities which are dependent on the total protein concentration of the samples, while knowledge of the molecular weights of the monomeric subunits allows a more meaningful search for the concentrations of the individual polymeric components of the system. From these the various average molecular weights can be construed, and the various dissociation equilibrium constants evaluated.

Also, in this paper considerations are proposed on the meaning of nonideality terms in associating systems and possible ways for estimating them.

As an example the proposed procedures have been applied to measurements of sedimentation equilibrium in carbonmonoxyhemoglobin.

2. Data analysis

2.1. Searching for average molecular weights

The familiar equation, eq. 1, describes the distribution with the radial distance of the concentration of the various polymeric components of an associating system when equilibrium of sedimentation is reached.

$$C_r = \sum_i C_{i,r} = \sum_i C_{i,0} \exp\left[iAM_1\left(r^2 - r_0^2\right)\right] \tag{1}$$

C_r is the total polymer concentration at distance r, $C_{i,r}$ the concentration of component i at distance r, $C_{i,0}$ the concentration of component i at the reference distance r_0, i the degree of polymerization of that component, and M_i the molecular weight of the fundamental monomeric subunit. Also,

$$A = \frac{\omega^2(1 - \bar{v}\rho)}{2RT} \tag{2}$$

in which R and T are the gas constant and absolute temperature, ρ the density of the solvent, which is assumed constant at all radial distances, and $\bar{v}$ the partial specific volume of the polymer, which is assumed to be independent of the degree of polymerization.

A common way of solving eq. 1 is taking its derivative

$$\frac{\mathrm{d}C}{\mathrm{d}r^2} = AM_1 \sum_i iC_{i,r} \tag{3}$$

and dividing both terms by $C_r = \Sigma_i C_{i,r}$ obtaining the working equation

$$\frac{\mathrm{d}C}{\mathrm{d}r^2} \cdot \frac{1}{C_r} = AM_i \frac{\sum_i iC_{i,r}}{\sum_i C_{i,r}} = AM_{w,r} \tag{4}$$

which defines $M_{w,r}$ the weight average molecular weight of the polymer at distance r.

In practice, the numerical value of the gradient $\mathrm{d}C/\mathrm{d}r^2$ is found by averaging increments of C_r with the radial distance, as obtained from determinations at a short distance from one another. Thus, the assumption is made that, in the radial interval used for evaluating the gradient, eq. 1 can be simulated by a single exponential of the kind

$$C_r' = C_0' \exp\left[AM_e\left(r^2 - r_0^2\right)\right] \tag{5}$$

in which the prime indicates simulated quantities.

The quantity M_e would be the weight average molecular weight of the system in the middle of the interval considered, if $M_{e,r} = M_{w,r}$. This is not correct, as follows.

2.2. Average significance of M_e

Any procedure that simulates eq. 1 with a single exponential as in eq. 5 requires that the sum of the square residuals

$$S = \sum_r (C_r' - C_r)^2 \tag{6}$$

is a minimum.

When S is a minimum its derivative with respect to M_e and C_0' is zero. Therefore, we can write

$$\sum_r \left[A\left(r^2 - r_0^2\right) \mathrm{d}M_e + \mathrm{dln}\, C_0'\right] \times \left[(C_r')^2 - C_r'C_r\right] = 0 \tag{7}$$

Also, from eq. 5 we see that

$$A\left(r^2 - r_0^2\right)\mathrm{d}M_e + \mathrm{dln}\, C_0' = \mathrm{dln}\, C'_r \tag{8}$$

Therefore, for equal intervals of r^2 we can write

$$\sum_r (C_r')^2 - \sum_r C_r'C_r = 0 \tag{9}$$

It follows that

$$\frac{\mathrm{d}\sum_r (C_r')^2}{\mathrm{d}r^2} = \frac{\mathrm{d}\sum_r C_r'C_r}{\mathrm{d}r^2} \tag{10}$$

and from eqn. 1, 5 and 10 we obtain

$$M_e\left[2\sum_r (C_r')^2 - \sum_r C_r'C_r\right] = \sum_r\left(C_r'\sum_r iC_iM_1\right). \tag{11}$$

Since from eq. 9 we have

$$2\sum_r (C_r')^2 - \sum_r C_r'C_r = \sum_r C_r'C_r \tag{12}$$

eq. 11 can be rewritten as

$$M_e = M_1\frac{\sum_r\left(C_r'\sum_i iC_i\right)}{\sum_r\left(C_r'\sum_r C_i\right)} = \frac{\sum_r C_r'C_rM_{w,r}}{\sum_r C_r'C_r} \tag{13}$$

which shows that M_e is an average of weight average molecular weights, dependent on the total amount of protein present in the radial interval considered, as implied in the term $\Sigma_r C_r'C_r$.

2.3. Nonideality terms

The concentration dependence of experimentally determined molecular weights can be described by

$$M_{i,\mathrm{app}} = \frac{M_i}{1+Q_iC_i} \tag{14}$$

in which C_i is the concentration of component i and the nonideality term Q_i is

$$Q_i = \frac{\mathrm{d}\ln\gamma_i}{\mathrm{d}\,C_i} \tag{15}$$

where γ_i is the activity coefficient of component i [27]. The term Q has the dimension of $1/C$, in our case 1/g. In self-associating systems, because of the presence of various polymeric species additional terms have to be considered of the kind

$$Q_j = \frac{\mathrm{d}\ln\gamma_i}{\mathrm{d}\,C_j} \tag{16}$$

where $j \neq i$, so that

$$M_{i,\mathrm{app}} = \frac{M_i}{1+Q_iC_i+\sum_j Q_jC_j} \tag{17}$$

When the nonideality of solutions is taken into consideration, eq. 1 can be rewritten as

$$C_r = \sum_i C_{i,0}\exp\left(A\int_{r_o}^{r} M_{i,\mathrm{app}}\mathrm{d}r^2\right) \tag{18}$$

or substituting with eq. 17

$$C_r = \sum_r C_{i,0}\exp\left(iAM_1\int_{r_0}^{r}\frac{\mathrm{d}r^2}{1+Q_iC_i+\sum_j Q_jC_j}\right) \tag{19}$$

Eq. 19 contains too many unknowns and cannot be used as such. A more manageable equation can be obtained by grouping all Q in one term. This entails the definition of an average quantity, the average virial term Q_a as in the equation

$$M_{a,\mathrm{app}} = \frac{M_a}{1+Q_aC_r} \tag{20}$$

Where the subscript a = n or w or z represents any kind of average molecular weight of the protein at distance r. The average term Q_a defined in this way is applicable to all of the components of an associating system when it is multiplied by the total polymer concentration C_r. For a weight average molecular weight we can write

$$M_{w,\mathrm{app}} = \frac{\sum_i C_iM_i}{(1+Q_wC_r)\sum_i C_i} = \frac{\sum_i C_iM^*_{i,\mathrm{app}}}{\sum_i C_i} \tag{21}$$

from which

$$M^*_{i,\mathrm{app}} = \frac{M_i}{1+Q_wC_r} \tag{22}$$

is obtained. Notably $M^*_{i,\mathrm{app}} \neq M_{i,\mathrm{app}}$

The average nonideality term depends on the average procedure used for the molecular weight. For z averages where

$$(M^*_{i,\mathrm{app}})^2 = \frac{M_i^2}{1+Q_zC_r} \tag{23}$$

the ratio between Q_w and Q_z is

$$\frac{Q_z}{Q_w} = \frac{M_i + M^*_{i,\mathrm{app}}}{M^*_{i,\mathrm{app}}} = 2 + Q_wC_r \tag{24}$$

The ratio is 2 when either Q_w or C tend to zero otherwise it is larger.

The average Q_w is dependent on polymer concentrations. However, in proteins it is always much less than 1, therefore, even a substantial change in Q_w does not affect the value of $(1 + Q_w C_r)$ greatly. Also, conditions can be chosen that restrain the variation of the polymer concentration between the meniscus and the bottom of the cell when equilibrium of sedimentation is reached. Therefore, the convenient assumption can be posed that the average Q_w is practically constant at all radial distances.

2.4. *Working equations*

As discussed, M_1 is a known parameter, thus I propose to rewrite eq. 1 as

$$C_r = \sum_i C_{i,0} \exp\left(iAM_1 \int_{r_0}^{r} \frac{\mathrm{d}r^2}{1 + Q_w C_r} \right) \tag{25}$$

or, in its derivative form,

$$\frac{\mathrm{d}C}{\mathrm{d}r^2}\frac{1}{AM_1} = \frac{1}{1 + Q_w C_r} \sum_i iC_{i,0} \times \exp\left(iAM_1 \int_{r_0}^{r} \frac{\mathrm{d}r^2}{1 + Q_w C_r} \right) \tag{26}$$

Both equations can be used to form a set of simultaneous equations so as to overdefine the chosen floating parameters in the iterative nonlinear least-square procedure used to estimate their values.

Either equation can be rewritten in terms of dissociation constants correcting for the concentration of the monomeric component as described by Adams [3,4].

3. Experimental applications

Human hemoglobin was prepared from washed red cells hemolyzed in the presence of chloroform. The protein was recycled for a few hours through a mixed bed resin column.

A Beckman model E centrifuge was used for the experiments. After overspeeding at 2.5-times the equilibrium speed for 2–3 h the sample was left equilibrating for approx. 30 h. Pictures taken at various time intervals showed that the system was already in equilibrium after 20–25 h.

Liquid columns approx. 3 mm long were formed inside a 12 mm centerpiece. Arbitrary fringe numbers of one were assigned to each detectable half fringe. When, near the bottom of the cell, the fringes were too packed to be discernible they were dropped.

A sample of carbonmonoxyhemoglobin in 0.05 M phosphate buffer at pH 7.15 was spun at 9000 rpm for 30 h at 9°C.

The concentration profile of the protein in the cell was measured in terms of arbitrary fringe numbers at equal intervals of r. Alternatively interpolating the data, using a routine based on linear polynomial regression, equidistant square-radial positions were chosen for the subsequent numerical manipulations. Very similar results were obtained with either set of data. Therefore, the interpolation procedures were discontinued.

Both eqs. 25 and 26 were used for the analysis. As the concentration profile was given in arbitrary fringe numbers, eq. 25 was rewritten as

$$y_r = y_r' - D \tag{27}$$

where y' is the true fringe number and D a constant number of fringes at all r. In the equations C_i and Q_w and D were left floating. Also, M_i was left floating when only one molecular species was assumed in the system. The Marquardt algorithm was used for finding the values of the floating parameters [26]. In the algorithm the integral present in eqs. 25 and 27 was approximated, dividing each radial interval in a series of steps as in

$$\int_{r_0}^{r} \frac{\mathrm{d}r^2}{1 + Q_w C_r} = \sum_n \frac{\left(r_n^2 - r_{n-1}^2\right)}{1 + Q_w\left[\left(C_n + C_{n-1}\right)/2\right]} \tag{28}$$

When dimers and tetramers were included in eqs. 25 and 26 the amount of dimers was put very near zero by the algorithm. Therefore, the presence of only tetramers was assumed and the parameter M_1 was either left floating, or fixed at 64500.

Similar data were obtained with either equa-

Table 1

Sedimentation equilibrium of human hemoglobin

Initial protein concentration was 1 mg/ml, 0.05 M phosphate buffer at pH 7.15, 9°C, 12 mm centerpiece. On the basis of a refractive index increment of 190×10^{-6} at 541 nm it was computed that 1 mg/ml = 9.12 half fringes.

Radial distance (cm)	y_r' (experimental) (1/2 fringes)	y_r' (calculated) (1/2 fringes)
6.8308	4.4	4.4
6.8732	5.3	5.2
6.9157	6.3	6.4
6.9582	7.5	7.5
7.0007	8.9	8.9
7.0431	10.7	10.8
7.0856	12.8	12.8
7.1281	15.4	15.5

tion. When M_1 was left floating we obtained $M_w = 64529 \pm 140$ Da and $Q_w = 3.2 \pm 0.06$ ml/g. These data were independent of the initial estimates entered in the algorithm. Table 1 shows a typical fit of calculated to experimental data obtained with eq. 25.

4. Discussion

As reported by Adams and Fujita [2], experiments, conducted on the same protein at different initial concentrations, often failed to give similar M_w values for overlapping regions of protein concentrations. There are many experimental reasons which may explain this phenomenon. The data can be distorted by temperature gradients in the rotors, microheterogeneity of the samples, presence in the system of nonassociating monomers, protein denaturation at the oil interface when false bottoms are used and presence (or absence) or critical metal ions [28–31]. Here we add evidence that mathematical analyses of the concentration profile of sedimentation equilibrium data, conducted searching for average molecular weight values M_w, give instead the quantities M_e, which are dependent on the total protein content of the cells, i.e., on the initial concentrations of the samples.

Although M_e and M_w are probably similar enough in order to justify searches for those parameters, as the parameter M_1 is known, it appears much simpler and more rigorous to search for the concentration of individual molecular species. Actually, both searches can be conducted simultaneously using the same data in programmed algorithms.

From the computational point of view, nonideality terms are parameters which help in adjusting to the experimental data the multiexponential functions applicable to ideal systems. This makes the quantities Q very sensitive to the noise of the data. Also, their large averaging power means that only for low protein concentrations and low values of Q can this term be considered independent of protein concentration. Otherwise, using only one nonideality term will result in poor statistics of the simulations, and additional terms will have to be introduced in eq. 25 or 26 until good statistics are obtained and the terms become concentration-independent. However, in this way the Q terms may assume arbitrary values which are difficult to interpret. A good alternative is to assume a geometrical model of the shape of the molecules of the system and express the nonideality as a series expansion of anticipated virial coefficients, as proposed by Ross and Minton [1].

The experiments reported here were conducted in a relatively simple system. The molecular weight obtained for tetrameric liganded hemoglobin was very near the amino acid molecular weight of the system. The values of $Q_w = 3.2$ ml/g corresponds to the excluded volume of spherical proteins [22], and is consistent with the data obtained by Ross and Minton [1] using high hemoglobin concentrations.

We are now testing these procedures on more complex systems like apohemoglobin and mixtures of hemoglobin and apohemoglobin, which seem to affect the solubility of each other [32].

Acknowledgements

This work was supported in part by NIH grants HL-16134 and HL-33629. The computer facilities were supported by the Computer Science Center of the University of Maryland in College Park,

MD, and its branch at the Medical School in Baltimore.

References

1 P.A. Ross and A.P. Minton , J. Mol. Biol. 112 (1977) 437.
2 E.T. Adams, and H. Fujita in: Ultracentrifugal analysis in theory and experiment, ed. J.W. Williams (Academic Press, New York, 1963).
3 E.T. Adams, Proc. Natl. Acad. Sci. U.S.A. (1964) 509.
4 E.T. Adams, Biochemistry 4 (1965) 1646.
5 E.T. Adams, Biochemistry 4 (1965) 1966.
6 S.M. Creeth and R.N. Pain, Progr. Biophys. Mol. Biol. 17 (1967) 217.
7 P.W. Chun and S.J. Kim, J. Phys. Chem. 74 (1970) 899.
8 P.W. Chun, S.J. Kim, G.D. Williams, W.T. Cope, L.H. Tang and E.T. Adams, Biopolymers 11 (1972) 197.
9 P.W. Chun and S. Kim, Biophys. Chem. 3 (1975) 46.
10 R.J. Goldberg, J. Phys. Chem. 57 (1954) 194.
11 R.H. Hashmeyer and W.F. Bower, Biochemistry 9 (1970) 435.
12 L.A. Holladay and A.J. Sophiannopoulos, J. Biol. Chem. 247 (1972) 477.
13 G.J. Howlett and L.W. Nichol, J. Biol. Chem. 247 (1972) 5681.
14 G.J. Howlett, L.W. Nichol and P.R. Andrews, J. Phys. Chem. 77 (1973) 2907.
15 W.P. Reinhardt and P.G. Squire, Biochim. Biophys. Acta 94 (1965) 566.
16 D.E. Roark and D.A. Yphantis, Biochemistry 10 (1971) 3241.
17 A.J. Sophiannopoulos and K.E. Van Holde, J. Biol. Chem. 239 (1964) 2516.
18 P.G. Squire and C.N. Li, J. Am. Chem. Soc. 83 (1961) 3521.
19 R.F. Steiner, Arch. Biochem. Biophys. 39 (1952) 33.
20 R.F. Steiner, Arch. Biochem. Biophys. 44 (1953) 120.
21 R.F. Steiner, Arch. Biochem. Biophys. 47 (1953) 56.
22 D.C. Teller, T.A. Horbett, E.G. Richard and H.K. Schachman, Ann. N.Y. Acad. Sci. 164 (1969) 66.
23 P.O.P. Ts'o and S.T. Chan, J. Am. Chem. Soc. 86 (1969) 4176.
24 K.E. Van Holde and R.L. Baldwin, J. Am. Chem. Soc. 62 (1958) 734.
25 D.A. Yphantis and D.E. Roark, Biochemistry 11 (1972) 2925.
26 D.W. Marquardt, Soc. Ind. Appl. Met. (Siam) 11 (1963) 431.
27 C. Tanford, Physical chemistry of macromolecules (John Wiley, New York).
28 D.E. Roark and D.A. Yphantis, Ann. N.Y. Acad. Sci. 164 (1969) 245.
29 D.A. Yphantis, J.J. Correia, M.L. Johnson and G.-M. Wu, in: Physical aspects of protein interaction, ed. N. Catsimpoolas (Elsevier/North-Holland, 1978) p. 275.
30 E.T. Adams, Jr. and M.S. Lewis, Biochemistry 7 (1968) 1044.
31 K. Kakiuchi, J. Phys. Chem. 69 (1965) 1829.
32 B. Campbell, C. Fronticelli, A. Zachary and E. Bucci, J. Biol. Chem. 261 (1986) 3931.