COVOL: an answer to your thermodynamic non-ideality problems?

S. E. Harding*, J. C. Horton* and D. J. Winzor†

*National Centre for Macromolecular Hydrodynamics, University of Nottingham, Sutton Bonington LE12 3RD, U.K., and †Centre for Protein Structure, Function and Engineering, Department of Biochemistry, University of Queensland, Brisbane, Queensland 4072, Australia

Introduction

Thermodynamic non-ideality can lead to serious problems in terms of the analysis of macromolecular behaviour in solution, particularly in molar mass and mass action measurements based on the principles of thermodynamic equilibrium. The term 'molar mass' (in g/mol) favoured by the physical chemist is of course equivalent to the molecular mass in daltons, and by 'mass action' we mean interaction phenomena, whether they be self-association or so-called heterologous associations between different molecular species. By the phrase, 'those procedures for examining these phenomena based on thermodynamic equilibrium' we mean osmotic pressure, static light scattering and also sedimentation equilibrium in the ultracentrifuge.

Consequence of non-ideality

The consequence of thermodynamic non-ideality is that if we measure the molecular mass at a finite concentration c (g/ml) the measurement will be an underestimate, i.e. we are measuring only an apparent molecular mass M_{app}. If the solution is dilute enough this underestimate for techniques based on the number average molecular mass, namely osmotic pressure, is represented by a single extra term:

$$1/(M_n)_{app} = 1/M_n + Bc + \ldots \quad (1)$$

and for static light scattering and sedimentation equilibrium in the ultracentrifuge (recorded by using uv absorption or Rayleigh interference) the primary average is the weight average:

$$1/(M_w)_{app} = 1/M_w + 2Bc + \ldots \quad (2)$$

where B is the second thermodynamic virial coefficient (the first is simply $1/M$). By the way, the form of notation that we prefer is to use the traditional B [1] rather than A_2 for the second virial coefficient. Although the latter seems to be preferred by the light scattering community (see [2]), this causes an inconvenience for the study of interacting systems of molecules: the subscript 2 is generally associated with dimer, which is not what is meant here. The classical way of avoiding non-ideality is to measure the apparent molecular mass at a series of concentrations and extrapolate to zero to get the correct value [1], or just to work at very low concentration where the Bc or $2Bc$ term is negligible.

Non-ideality obscures interpretation of molecular interaction phenomena

The difficulty comes, however, when we want to probe phenomena that themselves depend on concentration, namely interactions between macromolecules: to investigate interaction stoichiometries and interaction constants it is necessary to probe the concentration dependence; particularly for weaker interactions, higher concentrations are necessary. To illustrate this consider the simplest situation, a non-ideal dimerization under dilute solution conditions [3]:

$$1/(M_n)_{app} = 1/M_1 + [B'_{11} - (K_2/M_1)^2]c + \ldots \quad (3)$$

$$1/(M_w)_{app} = 1/M_1 + 2[B'_{11} - (K_2/M_1)^2]c + \ldots \quad (4)$$

where the monomer molar mass is M_1, B'_{11} is the monomer–monomer non-ideality (expressed in $ml \cdot mol \cdot g^{-2}$) and K_2 is the association constant in ml/mol.

B'_{11} in $ml \cdot mol \cdot g^{-2}$ is simply related to the Wills and Winzor [3] B_{11} (in ml/mol) by

$$B_{11} = B'_{11} M_1^2. \quad (5)$$

The problem is clear in eqns. (3) and (4): it is impossible to predict association constants properly unless the non-ideality is known or is definitely negligible. Merely allowing B_{11} and K_2 both to float as variables is clearly futile.

The subject of this paper is the prediction of B (or B'_{11}) for quasi-rigid structures (structures that might include globular proteins and some rigid polysaccharides such as many of the polyanionics and polycationics, and the double- and triple-helical ones), but not flexible coil-like molecules: for the latter the reader is referred to the seminal works of, for example, Flory [4] and Tsvetkov et al. [5].

Contributions to the second virial coefficient, *B*

So to be able to predict B for quasi-rigid types of macromolecule we need to look at the contributions to the virial coefficient for a macromole-

cule; there are two: B_{ex}, from excluded volume effects deriving from the large size of the macromolecular solute molecules compared with the surrounding solvent molecules in a solution, and the other, $B_{\not z}$, from polyelectrolyte behaviour if the macromolecule has a net unsuppressed charge. The total virial coefficient is just the sum of these two terms:

$$B = B_{ex} + B_{\not z}. \quad (6)$$

Exclusion volume term

Let us first take the excluded volume term. The excluded volume of a macromolecule is simply the volume blocking or excluding the free movement of another macromolecule into a region of solution because of its presence, averaged (in the case of Brownian solutions) over all possible particle orientations. There will in turn be two contributions to this: a shape term, in which the more asymmetric a structure is the greater the spatially averaged effect, and a hydration term, in which the greater the macromolecule's own volume (swollen by association with solvent), the greater the effect. The simplest shape for which the excluded volume is available is the sphere, where the excluded volume or molecule covolume, u, in ml is just eight times the hydrated volume, V [1].

Because volumes and excluded volumes of macromolecules are very small, we tend to work instead with a quantity known as the molar covolume, U, in ml/mol, which is just u times Avogadro's number, N_A, so for a sphere the molar covolume is just $8N_AV$.

The virial coefficient term is just:

$$B_{ex} = U/(2M^2). \quad (7)$$

For more complicated shapes it is convenient to separate the contributions to U (and B_{ex}) from volume and shape:

$$U = (\bar{v} + \delta/\rho)\, M u_{red} \text{ (ml/mol)} \quad (8)$$

where $(\bar{v} + \delta/\rho)M$ is the volume term, $\bar{v}$ being the partial specific volume, ρ the solvent density and δ the 'hydration' (amount of solvent associated either chemically or physically with the macromolecule); the shape term is the 'reduced excluded volume' defined by:

$$u_{red} = u/V = U/VN_A \quad (9)$$

so, for a sphere, $u_{red} = 8$.

The most general shape for which u_{red} has been worked out exactly is the general triaxial ellipsoid (Figure 1) (of semi-axes $a \geqslant b \geqslant c$ and axial ratios a/b, b/c) according to the Rallison–Harding relation [6], applicable to dilute Brownian solutions averaged over all particle orientations where only two-particle interactions are relevant:

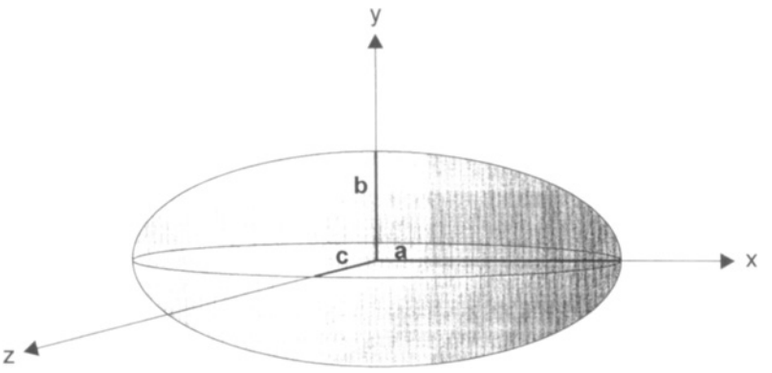
$$u_{red} = 2 + (3/2\pi abc)\, SR \quad (10)$$

where S and R are complicated double integrals [6], which although not solvable analytically can be easily solved using numerical routines. This formula identically reduces to the value of 8 for a sphere and to formulae given for prolate and oblate ellipsoids of revolution (ellipsoids with the restriction of two equal axes) of Isihara [7] and Ogston and Winzor [8], for the case of two equal axes.

Figure 1

Schematic representation of a rigid macromolecule as a triaxial ellipsoid in which all three semi-axes (*a*, *b*, *c*) can differ in length

Its shape is characterized by the two axial ratios (*a*/*b*, *b*/*c*).



Charge term

In studies of charged macromolecules such as proteins and polyelectrolytes in aqueous solution, the effective distance of closest approach is greater than that based on geometrical considerations because of the repulsive force opposing the approach of two particles bearing net charge (valence) Z. This additional contribution to the second virial coefficient, B_Z, has been evaluated explicitly only for impenetrable spheres. For such systems the expression for the second virial coefficient, B, is given by [9,10]:

$$B = B_{ex} + B_Z = uN_A/(2M^2) + [1000Z^2/(4M^2I)] \times [(1+2\kappa r_s)/(1+\kappa r_s)^2 + \ldots \quad (11)$$

where the factor of 1000 is introduced to accommodate the conventional definition of ionic strength I (mol/l); κr_s is the product of the inverse screening [11] and the solvated radius (r_s) of the particle. The Stokes radius provides an acceptable estimate of r_s (cm), irrespective of macromolecular shape; and the magnitude of κ (cm^{-1}) can be evaluated from the expression $\kappa = 3.27 \times 10^7\sqrt{I}$ at 20 °C.

The COVOL algorithm

COVOL is simply a program that builds in the Rallison–Harding excluded volume formula [eqn. (10)] with the Winzor–Wills charge formula [eqn. (11)]; it is written in FORTRAN 77 and runs in PC DOS mode: because it involves some serious number crunching it has to be run in a DOS rather than Windows environment. It evaluates B_{ex} by enumerating S, R and hence u_{red} from user-specified values of the three semi-axes a, b and c (or, alternatively, a/b and b/c because of the sole dependence of u_{red} upon shape), via eqn. (10). The double integrals S and R of eqn. (10) are evaluated by using the NAG [12] numerical integration routine D01DAF.

The next stage is the evaluation of the molar covolume U from u_{red} and user-specified values for the molecular mass (M) and the unsolvated partial specific volume ($\bar{v}$), the solvation (δ) and the solvent density (ρ) [eqn. (8)], or a combined parameter known as the 'swollen specific volume', v_s ($= \bar{v} + \delta/\rho$). The routine prints out the molar excluded volume (U), the molecular excluded volume u ($= U/N_A$), and B_{ex} [eqn. (7)]. At that stage the program asks whether there is an additional contribution to B from charge (polyelectrolyte) behaviour. If so, the user enters the ionic strength (mol/l) and net charge (valence) of the macroion, Z. After evaluation of B according to eqn. (11), the routine concludes by printing out the charge–charge contribution (B_Z) as well as the magnitude of the second virial coefficient, $B = B_{ex} + B_Z$ [eqn. (6)].

Availability

The FORTRAN 77 compiler (Salford [13] FTN77/486 system) and the NAG [12] numerical integration routine D01DAF are built in to the program: no separate FORTRAN or NAG compilers are required. COVOL is available in pre-compiled form from Steve.Harding@nottingham.ac.uk.

Input of shape information (a/b, b/c) onto COVOL

An objective method for evaluating the triaxial shape from an inertial ellipsoid fit to the atomic coordinates of a protein (from X-ray crystallography or high-resolution NMR) has been provided by Taylor et al. [14] in the FORTRAN routine ELLIPSE. The usefulness of this approach for COVOL has been described elsewhere [15].

Application of COVOL

We illustrate the application of COVOL for prediction of the magnitude of second virial coefficients by consideration of ovalbumin, a protein whose high-resolution crystal structure was recently published [16]. Fitting the crystal coordinates from the relevant Protein Data Bank file to the inertial ellipsoid with the FORTRAN routine ELLIPSE [14] yielded axial ratios (a/b, b/c) of (1.87, 1.08) [15]; entry of these respective values for a/b and b/c into COVOL yields a reduced covolume, u_{red}, of 8.996. Conversion of this reduced covolume to B_{ex} depends on the magnitude assigned to the solvation parameter (δ) for this protein with a partial specific volume ($\bar{v}$) of 0.748 ml/g [17] and a molecular mass of 45 kDa [18]. The effect of the extent of solvation on the magnitude of B_{ex} calculated via eqn. (8) is summarized by the solid line in Figure 2, where the intersecting horizontal broken lines denote the estimates of B_{ex} deduced experimentally from sedimentation-equilibrium [18] and size-exclusion chromatography [19] studies of isoelectric ovalbumin (upper and lower lines respectively). It is noted that the consequent estimates of 0.49 ($\pm$0.05) and 0.39 ($\pm$0.18) for the extent of ovalbumin solvation (δ) are at the upper end of, or greater than, the usually accepted range (0.3–0.4) for globular proteins [2,20,21]. Experi-

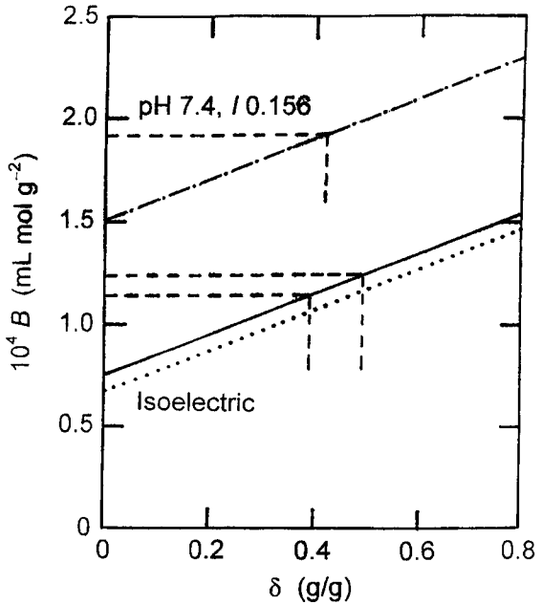
mental support for a higher value is provided by concordance of estimates (2.92 nm) for the Stokes radius and the effective radius deduced from the molar covolume, $U = (32/3)\pi N_A r^3$. A similar conclusion about the extent of solvation stems from size-exclusion chromatography studies [19] in phosphate/chloride buffer, pH 7.4, $I = 0.156$, conditions under which a net charge (Z) of -16 results in a polyelectrolyte contribution to B [eqn. (11)]. The upper dependence (dash–dotted line of Figure 2) summarizes the calculated variation of B with δ, whereas the intersecting horizontal line denotes the experimental value of B obtained by exclusion chromatography. On this basis ovalbumin is also hydrated to the extent of 0.42 (± 0.09).

An obvious difficulty with calculation of the second virial coefficient in this way is the pronounced dependence of B on the magnitude assigned to δ, a parameter for which the value is often very subjective. Indeed, this sensitivity of B to the value of δ relegates to secondary importance the relative magnitudes of the triaxial ellipsoid semi-axes, a factor evident from the dotted dependence in Figure 2, which refers to isoelectric ovalbumin modelled as a sphere ($u_{red} = 8.000$): values of 0.47–0.57 for the extent of ovalbumin hydration are obtained from this model and the exclusion-chromatographic [19] and sedimentation-equilibrium [18] estimates of B.

Figure 2

Effect of the extent of solvation (δ) on the magnitude of the second virial coefficient (B) calculated by COVOL

The basis used for triaxial ellipsoid semi-axes of ratios 1.87 (*a*/*b*) and 1.08 (*b*/*c*) for isoelectric ovalbumin (solid line), and for the same axial ratios under conditions (pH 8.5, $I = 0.156$) where the protein bears a net charge (valence) of -16 (dot–dashed line); the dotted line shows the corresponding dependence for isoelectric ovalbumin modelled as a sphere. Horizontal lines denote experimental estimates of B from sedimentation and exclusion chromatography studies of ovalbumin.



Conclusion

This investigation has demonstrated the use of COVOL to calculate second virial coefficients for macromolecules that can be modelled as impenetrable triaxial ellipsoids, but it has also identified the feature that realization of its full potential must await more definitive means of assessing the magnitude of δ, the extent of macromolecule solvation. In that regard the extent of solvation has usually been considered to be in the range 0.3–0.4 for globular proteins [2,20], whereas experimental measurements of B for ovalbumin signify a higher value (0.4–0.6) for δ. Measurements of B for a range of proteins with known axial dimensions are clearly required to shed further light on the likely magnitude of δ, and hence on its prediction on the geometrical basis of an assigned thickness to the solvation layer extending over the surface of the protein molecules (see, for example, [22]).

It is therefore hoped that this investigation will also stimulate renewed interest in the accurate measurement of osmotic virial coefficients, parameters for which the major use in the past has merely been to guide the elimination of non-ideality by extrapolation of data to infinite dilution.

This investigation has been funded by United Kingdom BBSRC and the EPSRC. Financial support from the Australian Research Council is also gratefully acknowledged. We thank also Dr. Susan Jones and Professor Janet Thornton of University College London for performing the ellipsoid fit to the crystal data of ovalbumin.

1 Tanford, C. (1961) Physical Chemistry of Macromolecules, Wiley, New York
2 Wyatt, P. J. (1992) in Laser Light Scattering in Biochemistry (Harding, S. E., Sattelle, D. B. and Bloomfield, V. A., eds.), chapter 3, Royal Society of Chemistry, Cambridge
3 Wills, P. R. and Winzor, D. J. (1992) in Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S. E., Rowe, A. J. and Horton,

J. C., eds.), chapter 17, Royal Society of Chemistry, Cambridge
4 Flory, P. J. (1953) Principles of Polymer Chemistry, Cornell University Press, New York
5 Tsvetkov, V. N., Eskin, V. E. and Frenkel, S. Ya. (1971) Structure of Macromolecules in Solution, vol. 1, National Lending Library for Science and Technology, Boston, U.K.
6 Rallison, J. M. and Harding, S. E. (1985) J. Colloid Interf. Sci. **103**, 284–289
7 Isihara, A. (1950) J. Chem. Phys. **18**, 1446–1449
8 Ogston, A. G. and Winzor, D. J. (1975) J. Phys. Chem. **79**, 2496–2500
9 Wills, P. R., Nichol, L. W. and Siezen, R. J. (1980) Biophys. Chem. **11**, 71–82
10 Winzor, D. J. and Wills, P. R. (1995) in Protein–Solvent Interactions (Gregory, R. B., ed.), pp. 483–520, Marcel Dekker, New York
11 Debye, P. and Hückel, E. (1923) Phys. Z. **24**, 185–206
12 NAG (1991) Workstation Library Manual, Numerical Algorithms Group, Jordan Hill, Oxford
13 SALFORD (1991) FTN77/486 Manual, Salford Software Ltd., Adelphi Street, Manchester
14 Taylor, W. R., Thornton, J. M. and Turnell, R. J. (1980) J. Mol. Graph. **1**, 30–38
15 Harding, S. E., Horton, J. C., Jones, S., Thornton, J. M. and Winzor, D. J. (1998), submitted
16 Stein, P. E., Leslie, A. G. W., Finch, J. T. and Carrel, R. W. (1991) J. Mol. Biol. **221**, 941–959
17 Dayhoff, M. O., Perlman, G. E. and Macinnes, D. A. (1952) J. Am. Chem. Soc. **74**, 2515–2517
18 Jeffrey, P. D., Nichol, L. W., Turner, D. R. and Winzor, D. J. (1977) J. Phys. Chem. **81**, 776–781
19 Shearwin, K. E. and Winzor, D. J. (1990) Eur. J. Biochem. **190**, 523–529
20 Oncley, J. L. (1941) Ann. N. Y. Acad. Sci. **41**, 121–150
21 Zhou, H. X. (1995) Biophys. J. **69**, 2298–2303
22 Jacobsen, M. P., Wills, P. R. and Winzor, D. J. (1996) Biochemistry **35**, 13175–13179

Received 7 July 1998

Direct analysis of sedimentation equilibrium distributions reflecting complex formation between cytochrome *c* and ovalbumin

Donald J. Winzor*, Michael P. Jacobsen* and Peter R. Wills†

*Department of Biochemistry, University of Queensland, Brisbane, Queensland 4072, Australia and †Department of Physics, University of Auckland, Auckland, New Zealand

Introduction

In recent characterizations of interactions between dissimilar macromolecular reactants by sedimentation equilibrium [1–3], the binding constant has been evaluated by iterative simulation of sedimentation equilibrium patterns to identify the best-fit description of the experimental distributions for the two solute constituents. A problem with such analyses has been their inability to accommodate realistic allowance for the effects of thermodynamic non-ideality in either self-associating or heterogeneously associating systems. Encouraged by the successful development of a direct analysis of solute self-association that incorporates rigorous statistical-mechanical treatment of thermodynamic non-ideality [4], we have looked into the feasibility of employing a similar approach for the characterization of interactions between dissimilar macromolecular solutes.

Theoretical considerations

The use of sedimentation equilibrium to characterize the interaction(s) between an acceptor, A, possessing several binding sites for a macromolecular ligand, S, entails ultracentrifugation of a mixture with defined molar concentrations of the two reactants, $(\bar{C}_A)_0$ and $(\bar{C}_S)_0$, at angular velocity ω and temperature T. Upon attainment of chemical as well as sedimentation equilibrium, the distribution of each individual species (i = A, S, AS, AS_2, etc.) is given by the expression [5,6]:

$$z_i(r) = z_i(r_F)\ \exp\left[M_i\phi_i(r^2 - r_F^2)\right] \qquad (1a)$$

$$\phi_i = (1 - \bar{v}_i\rho)\,\omega^2/(2RT) \qquad (1b)$$

which relates the molar thermodynamic activity, $z_i(r)$, of a species at any radial distance r to its value at a chosen reference radial position r_F: M_i and $\bar{v}_i$ are the respective molecular mass and partial specific volume of species i. Although ρ was considered initially [7] to be the density of the solution, it has now been identified unequivocally as the solvent density [8,9], a conclusion that at last allows the treatment of sedimentation equilibrium data in rigorous thermodynamic terms. Because the distributions of individual species are not recorded separately, the method of analysis depends on the combina-