Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 368 (2007) 168–177
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Characterization of weak protein dimerization by direct analysis of sedimentation equilibrium distributions: The INVEQ approach

Donald J. Winzor [a,*], Peter R. Wills [b]

[a] Department of Biochemistry, School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland 4072, Australia
[b] 37 Fairlands Avenue, Waterview, Auckland 1026, New Zealand

Received 9 March 2007
Available online 27 April 2007

Abstract

Closer scrutiny has been accorded a recently reported procedure for characterizing weak protein dimerization by sedimentation equilibrium (INVEQ) in which the equilibrium distribution is analyzed as a dependence of radial distance on solute concentration rather than of solute concentration on radial distance. By demonstrating theoretically that the fundamental parameter derived from the analysis is simply the difference between the dimerization constant and the osmotic second virial coefficient for monomer-monomer interaction, this investigation refutes the original claim that independent estimates of these two parameters can be obtained by nonlinear curve fitting of the sedimentation equilibrium distribution. This criticism also applies to conventional analyses of sedimentation distributions by the commonly employed Beckman Origin and NONLIN software. Numerically simulated distributions are then analyzed to demonstrate limitations of the procedure and also to indicate a means of improving the reliability of the returned estimate of the dimerization constant. These features are illustrated by applying the original and revised analytical procedures to a sedimentation equilibrium distribution for α-chymotrypsin (pH 4.0, I 0.05 M).

Keywords: Sedimentation equilibrium; Protein dimerization; Thermodynamic nonideality

A recent publication [1] has drawn attention to difficulties encountered in the application of widely used computer software, such as Beckman Origin [2] and NONLIN [3], to characterize reversible protein dimerization by sedimentation equilibrium in situations where thermodynamic nonideality needs to be taken into account. Problems arise in evaluation of the dimerization constant by iterative nonlinear curve fitting of sedimentation equilibrium distributions to expressions in the form $\overline{c} = f(r)$ because of failure to properly effect separation of the dependent variable ($\overline{c}$) from the independent variable (r); terms in $\overline{c}$, the total protein concentration, also appear within $f(r)$, the function of radial distance r. As noted by Rowe [1], this problem can be overcome by inverting the sedimentation equilibrium expression to the form $r = g(\overline{c})$. Although such action identifies radial distance as the dependent variable, it does achieve the separation of variables required to render more robust the nonlinear regression analysis encompassed in the requisite software (INVEQ).

All of the above procedures are open to criticism because of their incorporation of an unrealistic assumption [4] that the activity coefficients of different oligomeric states of a protein are considered to be described by the log–linear relationship

$$\ln \gamma_i = iBM_1\overline{c}, \tag{1}$$

where M_1 is the molecular mass of monomer and γ_i is the activity coefficient of an oligomer consisting of i monomers. The parameter B, regarded as a constant, has twice the magnitude of the second osmotic virial coefficient for a nonassociating species. This description of thermody-

* Corresponding author. Fax: +61 7 3365 4699.
E-mail address: d.winzor@uq.edu.au (D.J. Winzor).

0003-2697/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2007.04.030

namic nonideality has retained popularity because of the simplification it leads to in the form of the identity $K_i(app) = K_i$ between the apparent equilibrium constant, defined as the ratio of species concentrations c_i raised to the appropriate powers, $K_i(app) = c_i/c_1^i$, and the true thermodynamic constant, the corresponding ratio of thermodynamic activities, $K_i = z_i/z_1^i$ [5]. Unfortunately, there is no justification (apart from mathematical expediency) for employing Eq. (1) to describe the thermodynamic nonideality in protein solutions [6].

A more rigorous approach incorporates allowance for effects of thermodynamic nonideality on the statistical–mechanical basis of excluded volume [7,8]. In that regard, a striking feature to emerge from the statistical–mechanical treatment of thermodynamic nonideality in terms of total solute concentration [9,10] is an inability to evaluate independently both the second virial coefficient and the association constant by direct analysis of sedimentation equilibrium distributions for a reversibly dimerizing solute [10,11]. Such an enterprise necessarily fails because the coefficient of the linear term in $\bar{c}$ is simply the difference between these two parameters. It is therefore a matter of concern that the software packages for analyzing sedimentation equilibrium distributions produce as output what are claimed to be optimal estimates of both parameters. The problem clearly warrants further investigation.

Theoretical considerations

Initial applications of the statistical–mechanical approach to sedimentation equilibrium distributions for self-associating systems [5,12] required specification of a species activity coefficient as a virial expansion containing the concentrations of all oligomeric species present and, hence, the adoption of an iterative approach with initial equilibrium concentration estimates based on thermodynamically ideal behavior. However, by regarding solute self-association as a purely thermodynamic phenomenon, in fact another form of thermodynamic nonideality, Hill and Chen [9] developed an analysis whereby the thermodynamic activity of monomer, z_1, defined under conditions of constant temperature and solvent chemical potential, is expressed as a polynomial in total concentration of the single solute component.

Basic expressions for nonideal dimerization

From a theoretical viewpoint, the most logical concentration scale describing thermodynamic nonideality is the molar scale when solute chemical potential is being monitored under the constraints of constant temperature and solvent chemical potential [13], the situation existing in sedimentation equilibrium [14–16]. For nonideal dimerization, the statement of mass conservation is

$$\overline{C} = z_1/\gamma_1 + 2z_2/\gamma_2 = (z_1/\gamma_1)[1 + (2K_2 z_1(\gamma_1/\gamma_2)], \tag{2}$$

where $\overline{C} = \bar{c}/M_1$ is the base molar protein concentration (weight concentration divided by monomer molecular mass), and the concentrations of monomer (species 1) and dimer (species 2) are expressed as ratios of molar thermodynamic activities (z_i) to the corresponding activity coefficients (γ_i). The thermodynamic activity of dimer has then been related to that of monomer via the law of mass action and the true thermodynamic constant ($z_2 = K_2 z_1^2$). On the statistical–mechanical basis of excluded volume [7], the relationships between the monomer and dimer species' activity coefficients and molar concentrations are, correct to the third order in monomer concentration [17],

$$\gamma_1 = \exp[2B_{11}C_1 + B_{12}C_2 + (3/2)B_{111}C_1^2 + B_{112}C_1C_2 + (4/3)B_{1111}C_1^3 + \cdots] \tag{3a}$$

$$\gamma_2 = \exp[2B_{22}C_2 + B_{12}C_1 + (1/2)B_{112}C_1^2 + B_{122}C_1C_2 + \cdots], \tag{3b}$$

where the various osmotic virial coefficients for uncharged spherical molecules are related to species radii (R_1, R_2) by [7,18,19]

$$B_{ii} = 16\pi N_A R_i^3/3 \quad i = 1,2 \tag{4a}$$

$$B_{12} = 4\pi N_A (R_1 + R_2)^3/3 \tag{4b}$$

$$B_{iii} = 160\pi^2 N_A^2 R_i^3/9 \quad i = 1,2 \tag{4c}$$

$$B_{iij} = (16\pi^2 N_A^2/9)[R_i^6 + 2R_i^3 R_j^3 + 3R_i^2 R_j(2R_i + R_j)(R_i^2 + 2R_i R_j)] \quad i,j = 1,2;\ i \neq j \tag{4d}$$

$$B_{iiii} = 1175.35\pi^3 N_A^3 R_i^9/27 \quad i = 1,2, \tag{4e}$$

in which Avogadro's number (N_A) is included to express them on a molar basis rather than on a molecular basis.

From the logarithmic form of Eq. (2), it follows that

$$\ln \overline{C} = \ln z_1 - \ln \gamma_1 + \ln(1 + 2K_2 z_1 \gamma_1/\gamma_2) = \ln z_1 - \ln \gamma_1 + (2K_2 z_1 \gamma_1/\gamma_2) - (2K_2 z_1 \gamma_1/\gamma_2)^2/2 + (2K_2 z_1 \gamma_1/\gamma_2)^3/3 + \cdots \tag{5}$$

As noted previously [9], the incorporation of Eqs. (3a) and (3b), as well as their series-expanded forms, gives

$$\ln z_1 = \ln \overline{C} + 2(B_{11} - K_2)C_1 + [(3/2)B_{111} - 8K_2 B_{11} + 3K_2 B_{12} + 2K_2^2]C_1^2 + \cdots \tag{6}$$

as the expression, correct to the quadratic power of C_1, for monomer activity as a function of its molar concentration. Conversion of this power series in C_1 to the corresponding one in $\overline{C}$ is then effected by making allowance for the extra terms in dimer concentration thereby generated, with the net result being [9]

$$\ln z_1 = \ln \overline{C} + 2[B_{11} - K_2]\overline{C} + (3/2)[B_{111} - 8K_2 B_{11} + 2K_2 B_{12} + 4K_2^2]\overline{C}^2 + \cdots \tag{7}$$

Although the final expression for $\ln z_1$ is a series expansion in $\overline{C}$, the original series expansion [Eq. (6)] is in terms of C_1. Consequently, caution is required in any truncation of Eq. (7), which may converge slowly or even diverge for