Thermodynamics of Protein Aqueous Solutions: From the Structure Factor to the Osmotic Pressure

Luís Fernando Mercier Franco
Chemical Engineering Graduate Program, School of Engineering, Universidade de São Paulo, 05424-970 São Paulo, SP, Brazil

Cristiano Luis Pinto de Oliveira
Dept. of Experimental Physics, Inst. of Physics, Universidade de São Paulo, 05314-970 São Paulo, SP, Brazil

Pedro de Alcântara Pessôa Filho
Dept. of Chemical Engineering, School of Engineering, Universidade de São Paulo, 05424-970 São Paulo, SP, Brazil

DOI 10.1002/aic.14802
Published online April 10, 2015 in Wiley Online Library (wileyonlinelibrary.com)

An analytical expression for the structure factor for globular proteins in aqueous solution is presented. This expression was obtained considering a potential given by the sum of a hard core, a van der Waals attractive, and a screened Coulomb contribution. Experimental data of small angle x-ray scattering for bovine serum albumin (BSA) in aqueous solutions containing sodium salts at different protein concentrations and pH values are also presented. The developed expression for the structure factor describes accurately these experimental data provided a dependence of the attractive parameter on protein concentration is established. An expression for the osmotic pressure was derived from the structure factor. With attractive parameters adjusted from x-ray scattering data, the osmotic pressure of BSA aqueous solutions could be predicted with very good agreement with experimental data. © 2015 American Institute of Chemical Engineers *AIChE J*, 61: 2871–2880, 2015
Keywords: statistical thermodynamics, classical thermodynamics, aqueous solutions, bioprocess engineering

Introduction

Despite the increasing importance of bioprocesses in the chemical industry, the description of proteins in aqueous solutions, aiming at the development of models suitable to correlate and predict phase diagrams, still poses challenges. Phase diagrams of systems constituted by proteins in aqueous solutions may be very different from those of systems that do not contain biomolecules.[1] Moreover, the influence of system conditions (such as temperature, pH, and concentration) on the phase diagram is often counterintuitive.

Systems constituted by solutes and solvent, such as those constituted by a protein in aqueous solutions, are usually studied through a modified Hamiltonian. This modification considers only the solutes as individual particles and treats the solvent as a continuum medium characterized by properties such as permittivity, viscosity, and density. A rigorous formal treatment to this approach was presented by McMillan and Mayer.[2] The McMillan–Mayer approach can be understood as an expanded ensemble in which the number of all component molecules is split in two independent variables: the solvent chemical potential and the number of solute molecules. The McMillan–Mayer framework can be converted in either Gibbs or Helmholtz frameworks.[3,4] The advantage of using the McMillan–Mayer approach is the possibility of describing the behavior of solute molecules in a solvent using the same equations used for describing the behavior of gas molecules in vacuum.[5] This advantage has resulted in its extensive application on the description of colloidal systems.

The basis of the McMillan–Mayer framework is to consider that the interaction between two solute molecules is given by the potential of mean force, that is, the interaction potential averaged over all solvent molecule configurations.[6] Therefore, the key aspect for any calculation within this approach is to determine the potential of mean force between two solute molecules at certain solvent condition. This can be done for protein solutions either through experiments or through molecular simulation. McMillan and Mayer[2] considered the expansion of the grand-partition function in a power series in fugacities in their derivation. Therefore, their framework is usually used through the virial expansion for the osmotic pressure,[2] for which the osmotic second virial coefficient arises as an important parameter. There is great interest in the experimental determination of this coefficient due to its relationship with the outcome of precipitation operations. George and Wilson[7] proposed that protein crystallization would occur if the second virial coefficient lay in a definite range. Since their

Additional Supporting Information may be found in the online version of this article.

Correspondence concerning this article should be addressed to P. de Alcântara Pessôa Filho at pedro.pessoa@poli.usp.br.

work, many investigations have been conducted to get deeper insight on this relation.[8]

Despite recent advances in computational research, much of the experimental research in solution thermodynamics still uses analytical solutions to interpret experimental data. This is due to intrinsic characteristics inherent to analytical solutions. The most important of these characteristics is the fact that a small set of parameters may provide insightful understanding of the raw data. Particularly on the field of light, neutron and x-ray scattering, analytical solutions for the structure factor, which accounts for interparticle interaction, are commonly used to reproduce the experimental scattering intensity.

The development of analytical expressions for the structure factor usually considers the Ornstein and Zernike equation,[9] which defines the direct correlation function. However, this equation can only be solved using approximations known as closure relations. A review on this subject can be found in Caccamo.[10] One of these closure relations, which results in analytical solutions for simple interparticle potentials, was formulated by Percus and Yevick.[11] Such relation was solved independently by Wertheim[12] and Thiele[13] for a hard-sphere fluid. The resulting functional form for the structure factor was presented by Ashcroft and Lekner.[14] Baxter[15] proposed an analytical solution, also through the Percus–Yevick closure relation, for the adhesive hard-sphere fluid. Sharma and Sharma[16] introduced a modification in the Ashcroft-Lekner equation in order to be consistent with the Carnahan and Starling[17] equation of state for hard-sphere fluids.

Lebowitz and Percus[18] proposed the mean spherical approximation, formulated as a generalization of the spherical model for Ising spin systems to classical fluids.[19] Such approximation can be considered a perturbation of the Percus–Yevick closure relation.[20] Expressions for the structure factor for a square well fluid,[20] for charged particles in a neutralizing background[19] and for macroions interacting via a screened Coulomb potential[21,22] were proposed through this closure relation.

For protein solutions, structure factors given by an attractive potential with a hard core have been reported. For γ-crystallin and lysozyme solutions Malfois et al.[23] and Tardieu et al.[24] used a Yukawa potential within the hyper-netted chain approximation, which can only be solved numerically. Barbosa et al.[25] described the structure factor of bovine serum albumin (BSA) through a similar approach with the extension introduced by Narayanan and Liu.[26] However, in Zhang et al.[27] and in Barbosa et al.,[25] the attractive parameter (either the depth from the square well potential or the Yukawa parameter) increases when the protein concentration decreases, even though the salt concentration remains constant.

Therefore, a more realistic attractive potential that allows a better understanding of the behavior of protein solutions is still needed. Here, we present a possible solution to the structure factor considering that the potential comprises a repulsive hard core, an attractive potential proportional to $1/r^6$—a van der Waals potential—and a repulsive screened Coulomb potential. Small-angle x-ray scattering (SAXS) experiments were conducted with BSA in concentrated salt solutions, which is the most important condition for industrial purposes, and these results were used to obtain the parameters of the intermolecular potential. The expression for the structure factor is used to generate an expression for the osmotic equation of state. With this expression and the parameters obtained from regressing SAXS scattering data, the osmotic pressure curve of BSA as a function of protein concentration in aqueous solution with sodium chloride is predicted with excellent agreement with the experimental data.

Materials and Methods

Reagents

The reagents used in the experiments were BSA (Sigma-Aldrich, A3059, $\geq$99.0%), acetid acid (Sigma-Aldrich, 320099, $\geq$99.7%), trihydrated sodium acetate (Sigma-Aldrich, 32318, $\geq$99.5%), sodium chloride (Sigma-Aldrich, S7653, $\geq$99.5%), sodium sulfate (Sigma-Aldrich, 239313, $\geq$99.0%), sodium nitrate (Sigma-Adrich, S5506, $\geq$99.0 %), and deionized water (Milli-Q®). Protein stock-solutions of 100.0 mg mL^{-1} were prepared in buffer of acetic acid and trihydrated sodium acetate. pH values were measured in a pH-meter Digimed®.

SAXS experiments

SAXS experiments were run at room temperature of 23 ± 1°C, using a Bruker's® NANOSTAR® equipment, with wavelength $\lambda = 1.5418$ Å of CuK_α radiation and sample-detector distance of 67.0 mm. Scattering intensity data are presented as a function of the wavevector modulus $q=(4\pi/\lambda)\sin\theta$, in which λ is the radiation wavelength and 2θ is the scattering angle. The momentum transfer range was 0.013–0.33 $Å^{-1}$. Background intensities were obtained from the scattering of buffer solutions measured inside the same capillaries. The scattering data were obtained with 300–1200 s of exposition, depending on the protein concentration, and were analyzed with the software SUPERSAXS.[28] In these experiments, the sample was continuously flown inside the vacuum chamber using an external peristaltic pump. By diluting the original stock solution with buffer solution, this mechanism allowed to obtain curves for several protein concentrations. The initial protein concentration was 100.0 mg mL^{-1}, and the salt concentration of both the stock protein solution and the buffer solution was 1.0 mol L^{-1}. The scattering of BSA in sodium chloride, sodium sulfate, and sodium nitrate solutions was investigated.

The scattering intensity was used to calculate the form and structure factors through the following analysis. Assuming that interactions are independent of the orientation, the scattering intensity $I(q)$ at the wavevector modulus q, for a monodisperse ensemble of anisotropic particles, is given by[29]

$$I(q)=KP(q)[1+\beta(q)(S(q)-1)] \tag{1}$$

where K is a proportionality constant that depends on the number of scattering particles, $S(q)$ is the structure factor, and $P(q)$ is the form factor, defined by

$$P(q)=\langle A^2(q)\rangle \tag{2}$$

where $A(q)$ is the amplitude of scattering at q and $\beta(q)$ is the ratio between the square of the average amplitude and the form factor

$$\beta(q)=\frac{\langle A(q)\rangle^2}{\langle A^2(q)\rangle} \tag{3}$$

For spherical particles, $\beta(q) = 1$.

The BSA form factor was calculated using CRYSOL[30] with the crystallographic structure of monomer A of 4F5S

file from Protein Data Bank.[31] The amplitude was adjusted using the Nelder and Mead algorithm[32] for an oblate ellipsoid of revolution (41.4 Å × 41.4 Å × 18.6 Å), following Guinier's expression[33] with the volume equivalent to a sphere with diameter 63.4 Å.

Monte Carlo simulations

Each Monte Carlo simulation was conducted with 2048 spherical particles of diameter σ. These particles were allocated within a cubic box with edge length of 17.5 σ, such that the packing fraction were 0.2. The initial configuration was that correspondent to a face centered cubic crystal. Each translational trial was attempted following the Metropolis method in canonical ensemble and respecting periodic boundary conditions[34]; the acceptance ratio was set to 0.5. Interaction potentials were truncated at 3.0 σ. For equilibration, 2.048×10^7 steps were performed. For the production stage, the same amount of cycles was used. Space coordinates were stored each 2048 steps. The interaction potential used in the simulations was

$$\frac{U(x)}{\varepsilon}=\begin{cases}\infty, & \text{if } x \le 1\\ -\left(\frac{1}{x}\right)^6, & \text{if } x > 1\end{cases} \tag{4}$$

where ε is the absolute value of the attractive potential at the contact distance, $x=r/\sigma$ and σ is the particle diameter. Lennard-Jones units were used in the simulations.

The calculation of the structure factor was done with a postprocessing code. The structure factor, $S(q)$, is related to the fluctuations of the spatial Fourier transform of the number density,ρ, through[35]

$$S(q)=\frac{1}{N}<\rho(q)\rho(-q)> \tag{5}$$

$$\rho(q)=\sum_{j=1}^{N} e^{i\vec{q}\cdot\vec{r}_j} \tag{6}$$

In a cubic box, $\vec{q}=(2\pi/L)(q_x,q_y,q_z)$ where L is the box edge length and q_x, q_y, and q_z are integers. Thus, following the approach of Frenkel et al.[36] and Cannavacciuolo et al.,[37] the structure factor was calculated at the reciprocal lattice points of the box. A set of 13 directions (h, k, l), generated by 001, 110, and 111 and their equivalents directions, was considered[38]

$$S(q_p)=\frac{1}{N}\left[\sum_{j=1}^{N} e^{-ip2\pi\left(hx_j+ky_j+lz_j\right)/L}\right]^2 \tag{7}$$

$$q_p=\frac{2\pi}{L}p\sqrt{(h^2+k^2+l^2)}, p=1,2,\ldots \tag{8}$$

Structure factors from different directions were averaged. The variance was calculated according to

$$\sigma^2=\frac{1}{M}\sum_{i=1}^{M}\left[S_i(q)-\overline{S(q)}\right] \tag{9}$$

where M is the total number of samples.

Theoretical framework

Using the concept of structure factor is not common in chemical thermodynamics. Therefore, its relationship to other quantities must be established beforehand. The structure factor is simply the Fourier transform of the radial distribution function[10]

$$S(q)=1+\rho\int_0^{+\infty} g(r)e^{-iqr}dr \tag{10}$$

Defining the total correlation function $h(r)$ as

$$h(r)=g(r)-1 \tag{11}$$

The Ornstein–Zernike equation relates the total correlation function, $h(r)$, and the direct correlation function, $c(r)$, through[9]

$$h(r)=c(r)+\rho\int h(|r'-r|)c(r')dr' \tag{12}$$

Applying the Fourier transform on this equation results in

$$H(q)=C(q)+\rho H(q)C(q) \tag{13}$$

where $H(q)$ is the Fourier transform of $h(r)$, and $C(q)$ is the Fourier transform of $c(r)$. Coupling equations 10 and (13)

$$S(q)=\frac{1}{1-\rho C(q)} \tag{14}$$

Therefore, to obtain $S(q)$ we need an expression for $C(q)$. To relate $C(q)$ with the interparticle potential, the mean spherical approximation[18] may be considered

$$c(r)=\begin{cases} g(r)\left[1-e^{U(r)/k_BT}\right], & \text{if } r \le \sigma\\ -\frac{U(r)}{k_BT}, & \text{if } r > \sigma\end{cases} \tag{15}$$

where $U(r)$ is the interparticle interaction potential, k_B is the Boltzmann constant, and T is the absolute temperature. Here, we adopted the procedure developed by Sharma and Sharma.[20] The direct correlation function is equal to a modified Percus–Yevick solution[16] for $r \le \sigma$, and is proportional to the interaction potentical (likewise the random phase approximation) for $r > \sigma$. The interaction potential is considered to be the sum of an attractive contribution and a screened Coulomb contribution

$$U(r)=\begin{cases}\infty, & \text{if } r \le \sigma\\ -\varepsilon\left(\frac{\sigma}{r}\right)^6+\frac{(ze)^2\exp\left[-\kappa(r-\sigma)\right]}{4\pi\epsilon_0\epsilon_r r(1+\kappa\sigma/2)^2}, & \text{if } r > \sigma\end{cases} \tag{16}$$

where ε is the atractive potential, σ is the particle diameter, z is the charge of the particle in units of the elementary charge e, κ is the inverse of Debye's lenght, ϵ_0 is the vacuum permittivity, and ϵ_r is the medium dielectric constant.

Following the approach of Sharma and Sharma,[20] the direct correlation function $c(r)$ can be written as a sum of the direct correlation functions generated by a hard-sphere potential, $c^{hs}(r)$, by the attractive potential, $c^{vdW}(r)$, and by the screened Coulomb potential, $c^{Coul}(r)$

$$c(r)=c^{hs}(r)+c^{vdW}(r)+c^{Coul}(r) \tag{17}$$

The Fourier transform of Eq. 17 is

$$C(q)=C^{hs}(q)+C^{vdW}(q)+C^{Coul}(q) \tag{18}$$

For the hard-sphere contribution, we will consider the solution obtained by Ashcroft and Lekner[14] with the modification introduced by Sharma and Sharma[16]

$$C^{\mathrm{hs}}(k)=-\frac{24\eta}{\rho k^6}\{\alpha k^3[\sin k-k\cos k]+\beta k^2[2k\sin k-(k^2-2)\cos k-2]+\gamma[(4k^3-24k)\sin k-(k^4-12k^2+24)\cos k+24]\} \tag{19}$$

where $\eta=\rho\pi\sigma^3/6$ is the packing fraction, and

$$k=q\sigma \tag{20}$$

$$\alpha=\frac{\left[(1+2\eta)^2+\eta^3(\eta-4)\right]}{(1-\eta)^4} \tag{21}$$

$$\beta=-\frac{\eta}{3}\frac{(18+20\eta-12\eta^2+\eta^4)^2}{(1-\eta)^4} \tag{22}$$

$$\gamma=\frac{\eta\alpha}{2} \tag{23}$$

This expression is consistent with the Carnahan–Starling equation of state.[16] In this case, the Fourier transform depends on the protein concentration, implicit in the packing fraction η; however, we left C^{hs} written simply as function of k, as this parameter is related to q and r (through inverse Fourier transform).

The attractive contribution of the direct correlation function is obtained from the functional form of the interparticle interaction in Eq. 16

$$C^{\mathrm{vdW}}(q)=4\pi\frac{\sigma^6}{T^*}\int_\sigma^\infty\frac{1}{r^6}\frac{\sin(qr)}{qr}r^2dr \tag{24}$$

where T^* is the reduced temperature, defined by $T^*=k_BT/\varepsilon$. This expression results in

$$C^{\mathrm{vdW}}(k)=\frac{\pi\sigma^3}{6T^*k}\left[k^4\left(\frac{\pi}{2}-Si(k)\right)+(6-k^2)\sin(k)+k(2-k^2)\cos(k)\right] \tag{25}$$

where k is defined by Eq. 20 and $Si(x)$ is the integral defined as

$$Si(x)=\int_0^x\frac{\sin(t)}{t}dt \tag{26}$$

From the functional form of the interparticle interation in Eq. 16, the direct correlation function of the electrostatic contribution is

$$C^{\mathrm{Coul}}(q)=-\frac{z^2e^2\exp(\kappa\sigma)}{k_BT\epsilon_0\epsilon_r(1+\kappa\sigma/2)^2}\int_\sigma^\infty\frac{\exp(-\kappa r)}{r}\frac{\sin(qr)}{qr}r^2dr \tag{27}$$

where κ is the inverse of Debye's length, defined as

$$\kappa=\sqrt{\left(\frac{2N_Ae^2I\rho_s}{\epsilon_0\epsilon_rk_BT}\right)} \tag{28}$$

N_A is Avogadro's number, ρ_s is the solvent density, and I is the ionic strength, defined as

$$I=\frac{1}{2}\sum_{i=1}^{N}m_iz_i^2 \tag{29}$$

where m_i is the molality (in mol per kilograms of solvent) of particle i. The influence of protein on the ionic strength is neglected.

From Eq. 27

$$C^{\mathrm{Coul}}(k)=-z^2\frac{\phi}{q\kappa}\left(\frac{\sin k+Q\cos k}{1+Q^2}\right) \tag{30}$$

$$\phi=\frac{e^2}{k_BT\epsilon_0\epsilon_r(1+K)^2} \tag{31}$$

where $k=q\sigma$, $Q=q/\kappa$ and $K=\kappa\sigma/2$.

The Debye–Hückel point ion limit can be recovered from Eq. 30.

Coupling equation 18 with Eq. 19, 25, and 30, the final expression for the direct correlation function $C(k)$ is

$$C(k)=-\frac{24\eta}{\rho k^6}\{\alpha k^3[\sin k-k\cos k]+\beta k^2[2k\sin k-(k^2-2)\cos k-2]+\gamma[(4k^3-24k)\sin k-(k^4-12k^2+24)\cos k+24]-\frac{k^5\varepsilon}{24k_BT}\left[k^4\left(\frac{\pi}{2}-Si(k)\right)+(6-k^2)\sin(k)+k(2-k^2)\cos(k)\right]+\frac{k^6\phi z^2}{4\pi\sigma^3q\kappa}\left(\frac{\sin k+Q\cos k}{1+Q^2}\right)\} \tag{32}$$

Isothermal compressibility and the osmotic equation of state

The isothermal compressibility is related to the limit of $q=0$ in the structure factor so that

$$\lim_{q\to0}S(q)=\rho k_BT\kappa_T=k_BT\left(\frac{\partial\rho}{\partial\Pi}\right)_T \tag{33}$$

where Π is the osmotic pressure and κ_T is the isothermal compressibility. The structure factor presented in Eq. 32 gives the isothermal compressibility

$$\kappa_T=\frac{\pi\sigma^3}{6\eta(\alpha k_BT-8\varepsilon\eta)} \tag{34}$$

where α is given by Eq. 21. The osmotic pressure can be calculated through

$$\Pi=\int_0^\eta\frac{d\eta}{\eta\kappa_T} \tag{35}$$

which results in

$$Z=\frac{\Pi}{\rho k_BT}=\frac{1+\eta+\eta^2-\eta^3}{(1-\eta)^3}-\frac{8}{\eta k_BT}\int_0^\eta\varepsilon(\eta)\eta d\eta \tag{36}$$

We left intentionally ε as a function of η due to the experimental evidence that ε is not constant for proteins in aqueous solutions.[25,27] As Z must approach 1 as η approaches 0, the expression for ε must follow the limit

$$\lim_{\eta\to0}\frac{1}{\eta}\int_0^\eta\varepsilon(\eta)\eta d\eta=0 \tag{37}$$

Equation 36 reduces to the Carnahan–Starling-van der Waals equation of state if ε is constant.

Results

Validation of the structure factor

An example of the structure factor generated by Eq. 32 is presented in Figure 1. This figure was generated using a

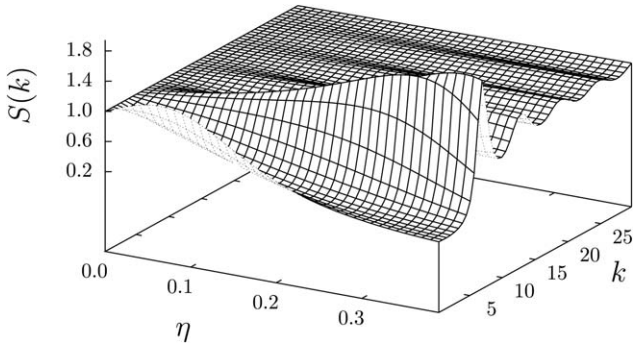

Figure 1. Structure factor (Eq. 32) as a function of η and k (Eq. 20) calculated for a system of uncharged particles and attractive parameter $\varepsilon/k_BT=1.67$.

highly attractive parameter ($\varepsilon/k_BT=1.67$). For high values of k, the structure factor goes to 1 as expected. For small values of η, the structure factor goes to 1 regardless of k, which means that dilute systems approach the ideal behavior. However, the first peak enlarges when η increases, and so does the isothermal compressibility (given by the limit of $k \to 0$) for small values of η. After certain η value, the isothermal compressibility decreases when η increases.

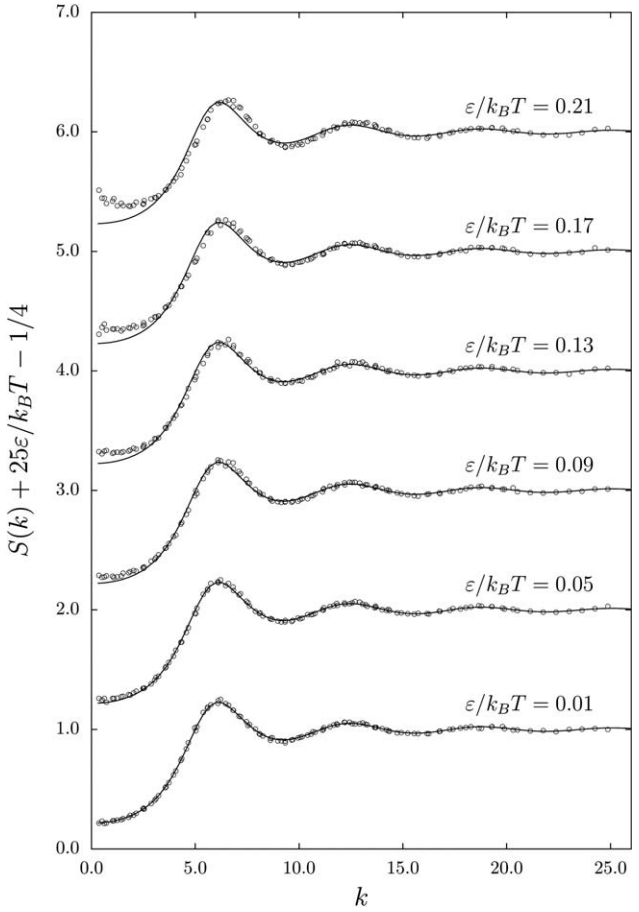

Figure 2. Structure factor as a function of $k=q\sigma$ and ε/k_BT for a system of uncharged particles with $\eta=0.2$.

Open circles, Monte Carlo results (standard deviations are smaller than symbol size). Continuous line, Eqs. 14 and 32.

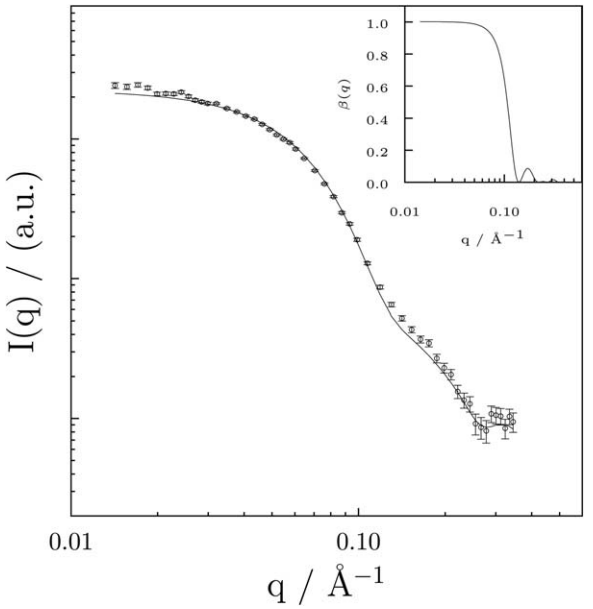

Figure 3. Form factor for BSA in aqueous solution.

Experimental results for the scattering intensity (in a.u., arbitrary units) of BSA in aqueous solution with 6.8 mg mL^{-1}, open circles. Curve calculated using CRYSOL,[30] continuous line. The inlet plot shows $\beta(q)$ (Eq. 3) considering an oblate ellipsoid of revolution.

Figure 1 shows that the expected limits are respected, but it does not allow an assessment of whether the approximations assumed in the derivation of Eq. 32 are adequate. Therefore, comparison with Monte Carlo simulations was carried out. Figure 2 shows the results of the structure factor obtained from Monte Carlo simulations (using the interparticle potential from Eq. 4) and from Eq. 32 for an uncharged particle, calculated for different values of the attractive parameter and $\eta=0.2$.

The agreement between the results from Eq. 32 and from Monte Carlo simulations is remarkable. However, for high values of the attractive parameter a non-negligible difference between the isothermal compressibility calculated via Monte Carlo and Eq. 32 is observed. This means that Eq. 32 fails to predict the large increment on the isothermal compressibility for highly attractive systems. This failure is related to the hypotheses behind the random phase approximation, which considers the attractive potential as a perturbation of the hard-sphere reference system.

Structure factor of BSA in aqueous salt solutions

Figure 3 shows the form factor for BSA in aqueous solution. We considered that the form factor is equal to the scattering intensity in very dilute solutions. In this case, the experimental data were obtained with a protein concentration of 6.8 mg mL^{-1}. The agreement between the calculated curve and the experimental data is high. The value of $\beta(q)$ is close to 1 for small scattering angles, which indicates that at small angles BSA can be modeled as an oblate ellipsoid of revolution.

The experimental data of scattering intensity as a function of protein concentration and wavevector modulus are presented as Supporting Information.

Previous works[25,27] showed that the attractive parameter of proteins in solution may depend on protein concentration.

Table 1. Values of Parameters ε^o and ε', Eq. 38, Values of χ for the Fitting Procedure, Eq. 39, and Values of the Osmotic Second Virial Coefficient B, Eq. 44, Calculated for BSA in Aqueous Solutions with 1.0 mol L^{-1} Sodium Salts

pH	Anion	ε^o/k_BT	ε'/k_BT	χ	$B\times10^4$ (cm^3 mol g^{-2})
4.9	SO_4^{-2}	2.00	8.24	2.52	−2.18
	Cl^-	2.32	11.10	11.35	−2.65
	NO_3^-	2.43	16.95	6.26	−2.81
6.3	SO_4^{-2}	1.44	5.55	1.96	−1.37
	Cl^-	2.39	14.77	8.05	−2.75
	NO_3^-	1.92	11.17	6.18	−2.06

The simplest relation between ε and η that follows the thermodynamic restriction imposed by Eq. 37 is

$$\varepsilon=\varepsilon^o-\varepsilon'\eta \tag{38}$$

where ε^o is the value of the attractive parameter at infinite dilution. The parameters of Eq. 38 were obtained by minimizing the following objective function

$$\chi_r^2=\frac{1}{N_cN-M}\sum_{j=1}^{N_c}\sum_{i=1}^{N}\left(\frac{I_j^{\exp}(q_i)-I_j^{\mathrm{mod}}(q_i)}{c_{p,j}\sigma_{ji}}\right)^2 \tag{39}$$

where N_c is the number of different sets of experimental data, N is the number of experimental intensities for a single-protein concentration, M is the number of parameters used in the fitting procedure, $I_j^{\exp}(q_i)$ is the experimental scattering intensity, $I_j^{\mathrm{mod}}(q_i)$ is the calculated scattering intensity, and σ_{ji} is the standard deviation of the experimental scattering intensity at q_i and protein concentration $c_{p,j}$. Table 1 presents the values of parameters ε^o and ε' adjusted to the experimental data using Eq. 38, as well as the values of χ.

A first question that arises from the analysis of experimental data is whether ε actually depends on η. Figure 4 shows the comparison, for two protein solutions, among the results obtained for the Hard-Sphere Percus–Yevick equation, for Eq. 32 and constant ε, and for Eq. 32 coupled with Eq. 38. For concentrated protein solutions, electrostatic interactions are screened by ions in solution, and even the hard-sphere potential is sufficient to predict the scattering intensity. However, as the protein concentration decreases, the experimental scattering deviates from the Percus–Yevick prediction. The use of a single value of ε does not allow a good correlation for all protein concentrations, and considering that ε depends on the protein concentration is necessary.

Figure 5 shows the scattering intensity of BSA in aqueous solution with $NaNO_3$ (1.0 mol L^{-1}) at the isoelectric point (i.e., the pH value for which the protein net charge is null, pH = 4.9 for BSA) for three protein concentrations, and Figure 6 presents the scattering intensity of BSA in aqueous solution with $NaNO_3$ above the isoeletric point (pH = 6.3). The agreement between the scattering intensities calculated by Eq. 32 and the experimental data is very good. In both figures, the calculated structure factor with the proposed model, Eq. 32, is also presented. Figures comprising all systems studied are presented as Supporting Information.

For the isoelectric point, the net charge is null and the Coulomb contribution to the structure factor vanishes. However, for a pH above the isoelectric point, the protein molecule bears a negative charge that must be considered in Eq. 27. To obtain the net charge profile of BSA as a function of pH, the pK_A values obtained by PROPKA[39] with the crystallographic structure of BSA[31] were considered, and the resulting profile was adjusted so that the isoelectric point was placed at pH = 4.9.

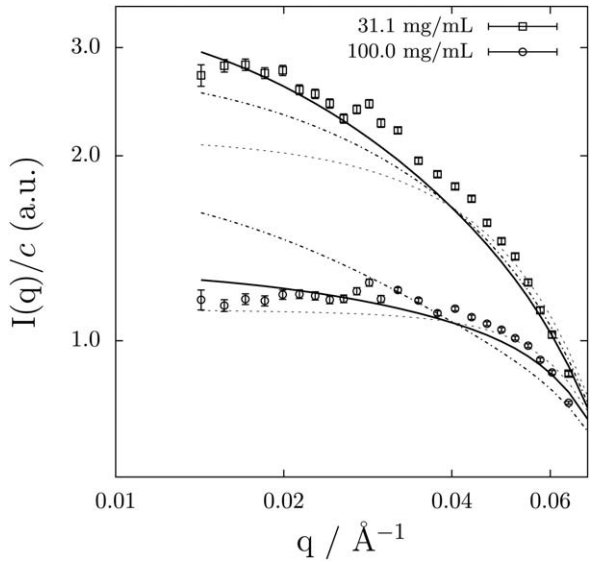

Figure 4. Open symbols, experimental data; continuous line, Eq. 32 with ε calculated using Eq. 38; dotted-dashed line, Eq. 32 with constant ε; dashed line, Hard-Sphere Percus–Yevick solution.

The results for parameters ε^o and ε' shown in Table 1 present two interesting features. The first one is that they depend on the pH. Even considering that this dependence is weak, it shows that the decoupling of the attractive part of $c(r)$ into $c^{\mathrm{vdW}}(r)$ and $c^{\mathrm{Coul}}(r)$, Eq. 17, is imperfect—otherwise, $c^{\mathrm{vdW}}(r)$ should not be influenced by the protein net charge. This may be due either to the expressions considered for $C^{\mathrm{vdW}}(q)$ and $C^{\mathrm{Coul}}(q)$, which involve approximations such as the calculation of the protein net charge, or to the very hypothesis that $c(r)$ can be decoupled in this way. The decoupling of the interaction potential into short-range and long-range terms is certainly an approximation, but is a useful and widely employed one. Conversely, the expressions used for $C^{\mathrm{vdW}}(q)$ and $C^{\mathrm{Coul}}(q)$ also entail approximations. The experimental data are not sufficient to decide which one of these aspects is critical. However, these findings warn against using this model uncritically.

The second feature is that the fitting of Eq. 38 leads to the unexpected conclusion that the more diluted is the protein in solution, the more attractive is its interaction potential. The very fact that the attractive parameter may depend on the solute concentration is not obvious. If the analogy with the nonideal gas behavior, hypothesized by the McMillan and Mayer[2] approach, was complete, then this potential should be independent of the solute concentration. However, as previously shown in Figure 4, considering this dependence is necessary even at the isoelectric point, at which the average charge of protein molecules is null; in this case, the decoupling of the attractive part of $C(q)$ and the expression used for $C^{\mathrm{Coul}}(q)$ play no role in the calculations.

One possible explanation for this contradiction is that the many-body contribution to the interaction potential, which is

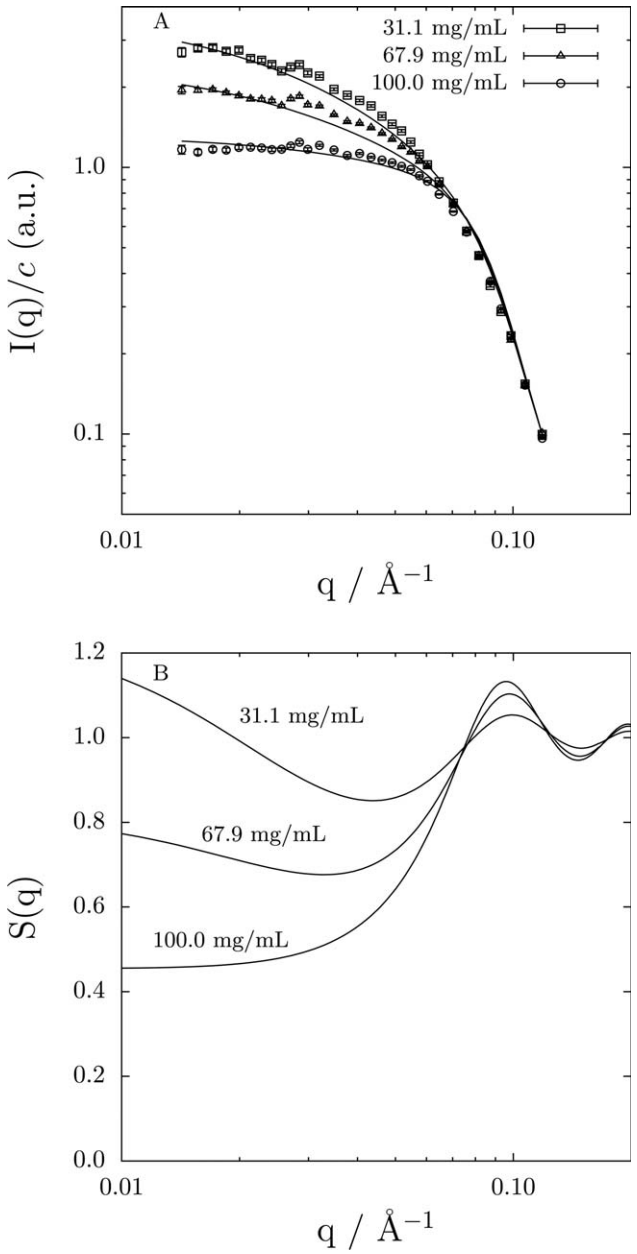

Figure 5. (A) X-ray scattering intensity (in a.u., arbitrary units) of BSA in 1.0 mol L^{-1} $NaNO_3$ aqueous solution at pH = 4.9 and 23°C. Open symbols, experimental data; continuous line, Eq. 32 with ε calculated using Eq. 38. (B) Calculated structure factor. Continuous line, Eqs. 14 and 32 with ε calculated using Eq. 38; dashed line, Hard-Sphere Percus–Yevick solution.

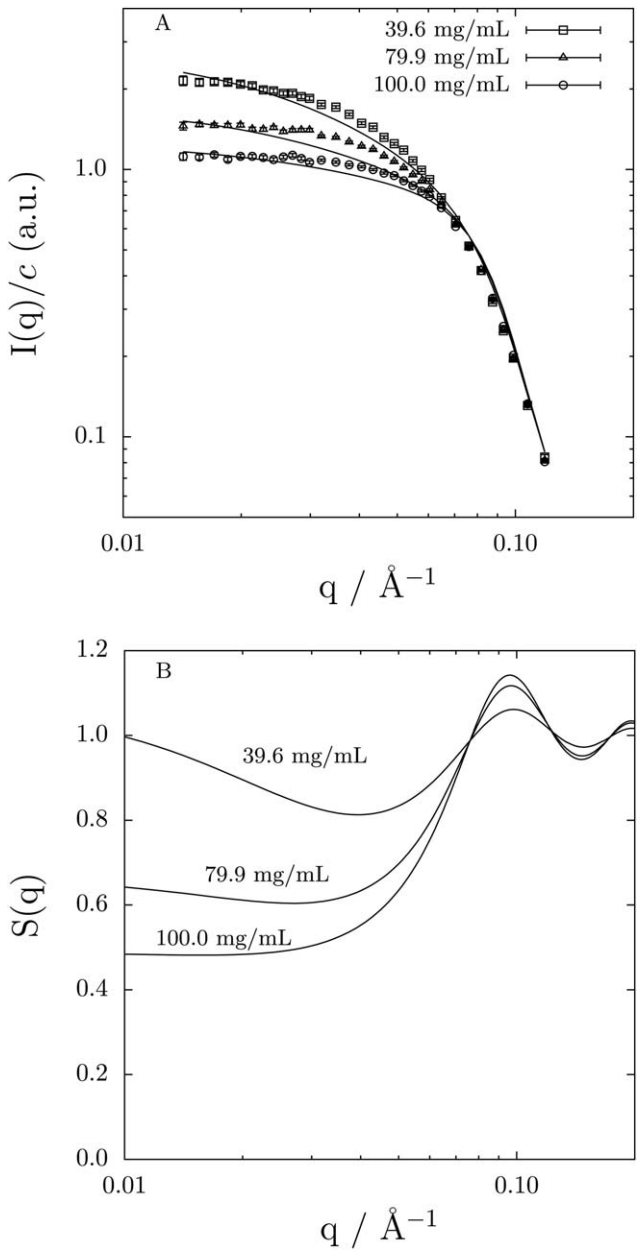

Figure 6. (A) X-ray scattering intensity (in a.u., arbitrary units) of BSA in 1.0 mol L^{-1} $NaNO_3$ aqueous solution at pH = 6.3 and 23°C. Open symbols, experimental data; continuous line, Eq. 32 with ε calculated using Eq. 38. (B) Calculated structure factor. Continuous line, Eqs. 14 and 32 with ε calculated using Eq. 38; dashed line, Hard-Sphere Percus–Yevick solution.

more relevant for more concentrated protein solutions, would lead to an apparent decrease (in absolute value) of the value of the interaction parameter. However, this contribution always leads to an increase in the attraction between molecules, which would result in an increase in the absolute value of the interaction parameter. A brief general demonstration of this fact is presented in Appendix.

Another possible explanation for this fact lies in the charge-dipole interactions. According to Striolo et al.[40,41] and Bratko et al.,[42] accounting for charge-dipole interactions improves the description of protein–protein interactions. In this case, the following term should be added to the interaction potential[43]

$$U^{c-d}(r) = -\zeta\left(\frac{\sigma}{r}\right)^4, \text{ if } r > \sigma \tag{40}$$

where ζ is a parameter proportional to the square of the product between the protein charge, z, and the protein dipole moment, μ.

This would mean that the following term should be added to the direct correlation function, Eq. 32

$$C^{c-d}(k)=\frac{2\pi\sigma^3\zeta}{k_BT}\left[\frac{\sin k}{k}-\cos k+k^2\left(2Si(k)-\frac{\pi}{2}\right)\right] \quad (41)$$

However, this term does not provide any insight in why ε changes with η, as it does not depend on protein concentration and would have the same effect on $S(q)$ for all protein concentrations. Particularly at the isoelectric point, this contribution is null, as the protein has no net charge and ζ is zero.

Finally, the other possible explanation for this finding is that the implicit solvent hypothesis behind the McMillan–Mayer framework is not adequate to describe the behavior of protein solutions. The fact that different salts have different effects on the protein molecules has long been known.[44] The conclusion of this work is subtler: the effect of the same salt at the same concentration may not be the same, and may depend on the concentration of the protein, or likewise on the proportion between salt ions to protein molecules. In this sense, the inclusion of the salt ions as solutes in the study of protein solutions, instead of a cosolvent, seems to be unavoidable even within the McMillan and Mayer framework.

Osmotic equation of state

By coupling Eqs. 38 and 36, the following expression for Z is obtained

$$Z=\frac{1+\eta+\eta^2-\eta^3}{(1-\eta)^3}-\frac{4\varepsilon^o\eta}{k_BT}+\frac{8\varepsilon'\eta^2}{3k_BT} \quad (42)$$

The osmotic second virial coefficient, B, can be calculated through

$$B=\lim_{\rho\to 0}\frac{Z-1}{\rho} \quad (43)$$

From Eq. 42

$$B=\frac{2\pi\sigma^3}{3}\left(1-\frac{2\varepsilon^o}{k_BT}\right) \quad (44)$$

Table 1 presents the value of the osmotic second virial coefficient calculated for BSA in these salt solutions. All values are negative, which indicates the predominance of attractive forces over repulsive ones. The osmotic second virial coefficient is known to depend on the pH, due to the dependence of the average net charge on the pH, which shifts the electrostatic repulsion.[45] For BSA, the calculated value of B at pH 6.3 is lower (in absolute value) than the calculated value of B at the isoelectric point, except for sodium chloride. The calculated values of B are similar to those previously reported for BSA in aqueous salt solutions. For instance, the experimental data by Wu and Prausnitz[46] suggests a value of $B=-1.7\times10^{-4}$ cm^3 mol g^{-2} for BSA in aqueous solution with 1.0 mol L^{-1} sodium chloride at pH = 4.5. For BSA in aqueous solution with 1.0 mol L^{-1} ammonium sulfate at pH = 4.8, a value of $B=-0.77\times10^{-4}$ cm^3 mol g^{-2} was reported.[47]

Further analysis can be done by considering experimental data on osmotic pressures. Figure 7 shows the results for BSA osmotic pressure as a function of protein concentration in 1.0 mol L^{-1} NaCl aqueous solution. The continuous line was drawn using Eq. 42 with parameters presented in Table 1. The developed equation can reproduce with good agreement the experimental data, despite the slight difference in pH values. The comparison with the Hard-Sphere Percus–Yevick equation shows that accounting for the attractive intermolecular interactions is important when calculating the osmotic pressure even at low protein concentrations. The agreement between the proposed equation for Π and the experimental results shows that the simplified hypothesis adopted in the development do not result in a loss of accuracy for the description of macroscopic quantities.

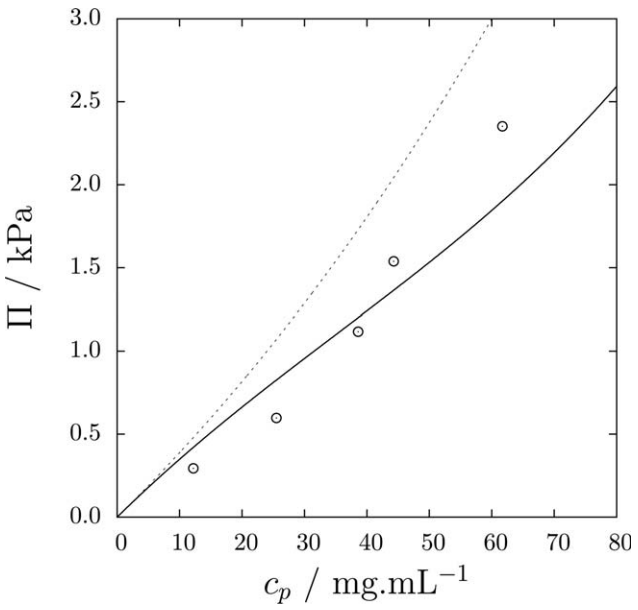

Figure 7. Osmotic pressure of BSA in 1.0 mol L^{-1} NaCl aqueous solution.

Open circles, experimental data from Wu and Prausnitz[46] (pH = 4.5); continuous line, Eq. 42 with parameters calculated at pH = 4.9; dashed line, Hard-Sphere Percus–Yevick equation of state.

To conclude, we would like to recall a famous statement by John Prausnitz, to whom this Tribute Issue is dedicated[1]:

> *"(.) Don't lose sight of the goal. Thermodynamics comes second. First comes chemical engineering."*

In this article, we have shown how to relate a fundamental physical measure (the scattering of x-ray) into useful information for Chemical Engineers—an osmotic equation of state, which entails the necessary information to construct phase diagrams. By pursuing the spirit of his teaching, we hope to have paid suitable homage to one of the greatest minds of Chemical Engineering.

Conclusions

An analytical expression for the structure factor of a fluid described by a van der Waals-like attractive potential was developed. The expression predicts the structure of a van der Waals generated through Monte Carlo simulations. Introducing an empirical relation to account for the dependence of the attractive parameter on the protein concentration allowed to correlate accurately experimental x-ray scattering intensities of BSA aqueous solution ate relatively high salt concentration. The results posed the question on how far the

implicit solvent hypothesis can be taken for granted. The expression for the structure factor was also used to develop an osmotic equation of state, which was used to predict experimental data of osmotic pressure of BSA solutions with good agreement.

Acknowledgment

Financial support of the Brazilian agencies São Paulo Research Foundation (FAPESP, grants 2011/22070-5 and 2013/18057-2), CAPES, and CNPq is gratefully acknowledged.

Literature Cited

1. Prausnitz JM. Thermodynamics and the other chemical engineering sciences: old models for new chemical product and processes. *J Phys: Condens Matter*. 2008;20:1–4.
2. McMillan WG, Mayer JE. The statistical thermodynamics of multicomponent systems. *J Chem Phys*. 1945;13:276–305.
3. Mollerup JM, Breil MP. On the thermodynamics of the McMillan-Mayer state function. *Fluid Phase Equilib*. 2009;276:18–23.
4. Mollerup JM, Breil MP. The osmotic second virial coefficient and Gibbs-McMillan-Mayer framework. *Fluid Phase Equilib*. 2009;286: 88–94.
5. Hill TL. *An Introduction to Statistical Thermodynamics*. New York: Dover, 1986.
6. Prausnitz JM. Molecular thermodynamics for some applications in biotechnology. *Pure Appl Chem*. 2003;75:859–873.
7. George A, Wilson WW. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr D*. 1994;50:361–365.
8. Tessier PM, Johnson HR, Pazhianur R, Berger BW, Prentice JL, Bahnson BJ, Sandler SI, Lenhoff AM. Predictive crystallization of ribonuclease A via rapid screening of osmotic second virial coefficients. *Proteins*. 2003;50:303–311.
9. Ornstein LS, Zernike F. Accidental deviations of density and opalescence at the critical point of a single substance. *Proc Sect Sci K ned Acad Wet*. 1914;17:793–806.
10. Caccamo C. Integral equation theory description of phase equilibria in classical fluids. *Phys Rep*. 1996;274:1–105.
11. Percus JK, Yevick GJ. Analysis of classical statistical mechanics by means of collective coordinates. *Phys Rev*. 1958;110:1–13.
12. Wertheim E. Exact solution of the Percus-Yevick integral equation for hard spheres. *Phys Rev Lett*. 1963;10:321–323.
13. Thiele E. Equation of state for hard spheres. *J Chem Phys*. 1963;39: 474–479.
14. Ashcroft NW, Lekner J. Structure and resistivity of liquid metals. *Phys Rev*. 1966;145:84–90.
15. Baxter RJ. Percus-Yevick equation for hard spheres with adhesion. *J Chem Phys*. 1968;49:2770–2774.
16. Sharma RV, Sharma KC. On generalization of the hard-sphere model for the the structure factor. *Phys Lett*. 1976;56:107–108.
17. Carnahan NF, Starling KE. Equation of state for nonattracting rigid spheres. *J Chem Phys*. 1969;51:635–636.
18. Lebowitz JL, Percus JK. Mean spherical model for lattice gases with extended hard cores and continuum fluids. *Phys Rev*. 1966;144:215–258.
19. Palmer RG, Weeks JD. Exact solution of the mean spherical model for charged hard spheres in uniform neutralizing background. *J Chem Phys*. 1973;58:4171–4174.
20. Sharma RV, Sharma KC. The structure factor and the transport properties of dense fluids having molecules with square well potential, possible generalization. *Physica*. 1977;89:213–218.
21. Hayter JB, Penfold J. An analytic structure factor for macroion solutions. *Mol Phys*. 1981;42:109–118.
22. Hansen JP, Hayter JB. A rescaled MSA structure factor for dilute charged colloidal dispersions. *Mol Phys*. 1982;46:651–656.
23. Malfois M, Bonneté F, Belloni L, Tardieu A. A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. *J Chem Phys*. 1996;105:3290–3300.
24. Tardieu A, LeVerge A, Malfois M, Bonneté F, Finet S, Riès-Kautt M, Bellonic L. Proteins in solution: from X-ray scattering intensities to interaction potentials. *J Cryst Growth*. 1999;196:193–203.
25. Barbosa LRS, Ortore MG, Spinozzi F, Mariani P, Bernstorff S, Itri R. The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: a small-angle X-ray scattering study. *Biophys J*. 2010;98:147–157.
26. Narayanan J, Liu XY. Protein interactions in undersaturated and supersaturated solutions: a study using light and X-ray scattering. *Biophys J*. 2003;84:523–532.
27. Zhang F, Skoda MWA, Jacobs RMJ, Martin RA, Martin CM, Schreiber F. Protein interactions studied by SAXS: effect of ionic strength and protein concentration for BSA in aqueous solutions. *J Phys Chem B*. 2007;111:251–259.
28. Oliveira CLP. SUPERSAXS – program package for data treatment, analysis and modeling. Aarhus: University of Aarhus, 2009.
29. Kotlarchyk M, Chen SH. Analysis of small angle neutron scattaring spectra from polydisperse interacting colloids. *J Chem Phys*. 1983; 79:2461–2469.
30. Svergun DI, Barberato C, Koch MHJ. *CRYSOL* – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst*. 1995;28:768–773.
31. Bujacz A. Structures of bovine, equine and leporine serum albumin. *Acta Crystallogr D*. 2012;68:1278–1289.
32. Nelder JA, Mead RA. A simplex method for function minimization. *Comput J*. 1965;7:308–313.
33. Pedersen JS. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv Colloid Interface Sci*. 1997;70:171–210.
34. Frenkel D, Smit B. *Understanding Molecular Simulation: From Algorithms to Applications*. San Diego: Academic Press, 2002.
35. Allen MP, Tildesley DJ. *Computer Simulation of Liquids*. New York: Oxford Science Publications, 1987.
36. Frenkel D, Vos RJ, Kruif CG, Vrij A. Structure factors of polydisperse systems of hard spheres. A comparison of Monte Carlo simulations and Percus-Yevick theory. *J Chem Phys*. 1986;84:4625–4630.
37. Cannavacciuolo L, Sommer C, Pedersen JS, Schurtenberger P. Size, flexibility, and scattering functions of semiflexible polyelectrolytes with excluded volume effects: Monte Carlo simulations and neutron scattering experiments. *Phys Rev E*. 2000;62:5409–5419.
38. Oliveira CLP. Small angle X-ray scattering from biological systems: theory and applications. PhD Thesis. Campinas: University of Campinas, 2005.
39. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins*. 2005;61:704–721.
40. Striolo A, Bratko D, Wu JZ, Elvassore N, Blanch HW, Prausnitz JM. Forces between aqueous nonuniformly charged colloids from molecular simulation. *J Chem Phys*. 2002;116:7733–7743.
41. Striolo A, Tavares FW, Bratko D, Blanch HW, Prausnitz JM. Analytic calculation of phase diagrams for charged dipolar colloids with orientation-averaged pair potentials. *Phys Chem Chem Phys*. 2003;5: 4851–4857.
42. Bratko D, Striolo A, Wu JZ, Blanch HW, Prausnitz JM. Orientation-averaged pair potentials between dipolar proteins or colloids. *J Phys Chem B*. 2002;106:2714–2720.
43. Israelachvili JN. *Intermolecular and Surface Forces*, 3rd ed. San Diego, CA: Academic Press, 2011.
44. Zhang Y, Cremer PS. Interactions between macromolecules and ions: the Hofmeister series. *Curr Opin Chem Biol*. 2006;10:658–663.
45. Neal BL, Asthagiri D, Lenhoff AM. Molecular origins of osmotic second virial coefficients of proteins. *Biophys J*. 1998;75:2469–2477.
46. Wu J, Prausnitz JM. Osmotic pressures of aqueous bovine serum albumin solutions at high ionic strength. *Fluid Phase Equilib*. 1999; 155:139–154.
47. Lu Y, Chen DJ, Wang GK, Yan CL. Study of interactions of bovine serum albumin in aqueous $(NH_4)_2SO_4$ solution at 25 °C by osmotic pressure measurements. *J Chem Eng Data*. 2009;54:1975–1980.
48. McQuarrie DA. *Statistical Mechanics*. California: University Science Books, 2000.

Appendix : The Many Body Effect on the Interaction Potential

Considering that the radial distribution function may be written as an expansion on densities, the same approach used to derive pressure as a virial expansion,[48] the following equation is obtained

$$g(r,T,\rho)=\sum_{j=0}^{N-2}\rho^j g_j(r,T) \tag{A1}$$

By the definition of the potential of mean force,[48] W, one may rewrite Eq. A1 as

$$e^{-W(r,\rho)/k_BT}=\sum_{j=0}^{N-2}\rho^j e^{-W_j(r)/k_BT} \tag{A2}$$

or

$$e^{-[W(r,\rho)-W_0(r)]/k_BT}=1+\sum_{j=1}^{N-2}\rho^j e^{-[W_j(r)-W_0(r)]/k_BT} \tag{A3}$$

where W_0 is the two-body contribution. Taking the natural logarithm leads to

$$W(r,\rho)=W_0(r)-k_BT\ln\left[1+\sum_{j=1}^{N-2}\rho^j e^{-[W_j(r)-W_0(r)]/k_BT}\right] \tag{A4}$$

The logarithm value on the right side of Eq. A4 is always positive, thus W is always lower than W_0; they will be equal at infinite dilution. Thus, all many-body contributions are somehow attractive in relation to the two-body interaction potential, provided the radial distribution function can be written as an expansion on densities.

Manuscript received Dec. 5, 2014, and revision received Feb. 20, 2015.