Effective interactions in lysozyme aqueous solutions: A small-angle neutron scattering and computer simulation study

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · JANUARY 2012

Impact Factor: 2.95 · DOI: 10.1063/1.3677186 · Source: PubMed

Effective interactions in lysozyme aqueous solutions: A small-angle neutron scattering and computer simulation study

M. C. Abramo, C. Caccamo, D. Costa, G. Pellicane, R. Ruberto et al.

Citation: J. Chem. Phys. **136**, 035103 (2012); doi: 10.1063/1.3677186
View online: http://dx.doi.org/10.1063/1.3677186
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v136/i3
Published by the American Institute of Physics.

Effective interactions in lysozyme aqueous solutions: A small-angle neutron scattering and computer simulation study

M. C. Abramo,[1,a)] C. Caccamo,[1] D. Costa,[1] G. Pellicane,[2] R. Ruberto,[1] and U. Wanderlingh[1]

[1]*Dipartimento di Fisica, Università degli Studi di Messina and CNISM (Consorzio Nazionale Interuniversitario di Struttura della Materia) Viale F. Stagno d'Alcontres 31, 98166 Messina, Italy*
[2]*School of Chemistry and Physics, University of Kwazulu-Natal Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa*

(Received 12 October 2011; accepted 24 December 2011; published online 20 January 2012)

We report protein-protein structure factors of aqueous lysozyme solutions at different *p*H and ionic strengths, as determined by small-angle neutron scattering experiments. The observed upturn of the structure factor at small wavevectors, as the *p*H increases, marks a crossover between two different regimes, one dominated by repulsive forces, and another one where attractive interactions become prominent, with the ensuing development of enhanced density fluctuations. In order to rationalize such experimental outcome from a microscopic viewpoint, we have carried out extensive simulations of different coarse-grained models. We have first studied a model in which macromolecules are described as soft spheres interacting through an attractive r^{-6} potential, plus embedded *p*H-dependent discrete charges; we show that the uprise undergone by the structure factor is qualitatively predicted. We have then studied a Derjaguin-Landau-Verwey-Overbeek (DLVO) model, in which only central interactions are advocated; we demonstrate that this model leads to a protein-rich/protein-poor coexistence curve that agrees quite well with the experimental counterpart; experimental correlations are instead reproduced only at low *p*H and ionic strengths. We have finally investigated a third, "mixed" model in which the central attractive term of the DLVO potential is imported within the distributed-charge approach; it turns out that the different balance of interactions, with a much shorter-range attractive contribution, leads in this latter case to an improved agreement with the experimental crossover. We discuss the relationship between experimental correlations, phase coexistence, and features of effective interactions, as well as possible paths toward a quantitative prediction of structural properties of real lysozyme solutions. © *2012 American Institute of Physics*. [doi:10.1063/1.3677186]

I. INTRODUCTION

Effective interactions in aqueous protein solutions depend on a complex balance between competing mechanisms and effects, finely influenced by the physicochemical preparation of the solutions (see, e.g., Refs.1–3). Forces between the proteins of electrostatic nature, arising from the total charge of the macromolecules, give rise to intermediate and long-range repulsions that can be substantially screened out through the addition of salts of various nature. Non-Coulombic interactions also play a prominent role, especially at low protein charge or high ionic strength: beside attractive van der Waals (vdW) forces, water-mediated forces are in fact present, including both hydration (repulsive) and hydrophobic (attractive) contributions.[4]

Understanding the effective protein-protein interactions is essential to explain the properties of such macromolecular solutions, in particular to predict their stability and phase behavior,[5,6] as well as to develop systematic protocols to grow protein crystals.[2,7] An insight into this issue is provided by small-angle neutron scattering (SANS) and x-ray investigations. Such experiments have allowed to establish that short-range attractions, and electrostatic repulsions induced by residual charges, are active between the proteins.[1,3,8] Recently, we have reported SANS results for lysozyme solutions[9] showing that, upon suitable physicochemical conditions, a marked rise of the scattered intensity $I(q)$ occurs at small wavevectors. Such an increase has been detected also for different protein solution conditions by other authors,[8,10–14] and its occurrence debated in relation to different interaction mechanisms and preparation protocols (see Ref. 15 for a review). Our SANS experiments show that the upturn of $I(q \to 0)$ marks a crossover between two different regimes, occurring around *p*H = 4–5, whereupon the solution passes from a behavior mostly dominated by repulsive forces toward a regime where attractive interactions become progressively relevant, with the ensuing development of enhanced density fluctuations, that might be associated with the presence of aggregates in the solution.[9] The interplay between cluster formation and phase separation, in relation with the long-range attractive/repulsive features of microscopic interaction has also been discussed.[11,14,16,17]

A support to the interpretation of experimental results is offered by simulations of microscopic models;

a)Author to whom correspondence should be addressed. Electronic mail: mcabramo@unime.it

because of the complexity of the protein solution environment, coarse-grained and effective descriptions of the intermolecular forces are frequently adopted (see our recent work[18] for an up to date bibliography). At the simplest level, one can assume that the macromolecules interact via an effective central pair potential consisting of a repulsive electrostatic term combined with a van der Waals attractive contribution, as envisaged for instance in the Derjaguin-Landau-Verwey-Overbeek (DLVO) theory of charge-stabilized colloids.[19,20] The pros and cons of such "colloidal representation" of globular protein solutions have been discussed in several studies;[1,6,21–26] in particular, we have shown that a DLVO description of microscopic interactions predicts with overall satisfactory results, the metastable protein-rich/protein-poor phase separation and solubility line of aqueous lysozyme and γ-crystallin solutions.[24–26] Other simple models do not seem equally able to reproduce the liquid phase coexistence lines of real solutions, and in particular their observed flat shape.[27–29]

In common with other simple, spherically symmetric potentials, the DLVO model cannot cope with many, highly specific aspects of protein interaction,[3,28–33] and fully atomistic simulations would be highly desirable to such a purpose; unfortunately, the computational costs pose severe limitations to a systematic use of this latter approach. Intermediate coarse-grained models seem then to constitute a suitable and convenient tool to investigate protein solutions;[18] in particular, such models take into account, with various degrees of approximation, directionality effects, which seem to play a fundamental role in the overall behavior of real systems.[34–36] We make specific reference in this work to the multi-site model early introduced by Carlsson, Malmsten, and Linse (CML hereafter),[37,38] in which the solution is represented as an effective one-component fluid, and lysozymes are modeled as hard spheres interacting by means of a central attractive short-range potential plus a directional force term, arising from a collection of embedded screened charged sites roughly reproducing the positions and charges of amino acid residues. The dependence of the structure factor of this model on the solution conditions has been early examined through Monte Carlo (MC) simulations[37] and considered qualitatively good; without, however, a detailed comparison with experimental data. The phase coexistence of the CML model (with a choice of parameters different from the original prescription) has been later investigated by Rosch and Errington,[39] with a resulting liquid-vapor coexistence curve essentially parabolic in shape, at variance with the rather flat experimental shape. Recently, we have carried out a molecular dynamics (MD) simulation study of a model similar to the CML one,[9,18] except for the fact that the hard-core excluded volume term has been replaced by a continuous soft-core term, as more suitable for the MD algorithm; we have obtained in Ref. 9 a qualitative reproduction of SANS intensities for lysozyme solutions, though our results turn out less satisfactory as for the evolution of the total diffracted intensity patterns as a function of the *p*H of the solution.

This work is motivated by the previous overview. On the one side, the CML approach is susceptible of a systematic application to different proteins, and can be refined to better describe the molecular shape, the aqueous solvent, and the electrostatic forces; then, it seems worth trying to improve its performances by starting from its structural predictions.[9,38] On the other side, we investigate to what extent the DLVO representation can be employed to predict the SANS structure factor, and whether any improvement of the distributed charge model can be suggested by such a simplified approach.

Our study is carried out according to the following plan:

1. We describe in detail the procedure we follow in order to obtain structure factors $S(q)$ from the SANS data partially reported in Ref. 9. We discuss on this basis the emerging structural features of the solutions at different *p*H and ionic strengths.
2. We report MD simulations for the CML-like model envisaged in Refs. 9 and 18 in order to show whether and to what extent the MD predictions improve if one chooses a hydrophobic interaction more attractive than the one previously adopted in Ref. 9.
3. We then investigate, by Monte Carlo simulations, the structure factors of a DLVO model in the physicochemical conditions of the SANS experiments. As in our previous works,[24–26] we fix the model parameters through the fit of the second virial coefficient, by showing that the liquid-vapor coexistence curve, determined by means of Gibbs ensemble Monte Carlo simulations (GEMC), satisfactorily reproduces the experimental protein-rich/protein-poor coexistence line reported in Ref. 40.
4. Finally, we investigate by MD a third "mixed" model, in which we keep the distributed-charge representation of the CML approach, but adopt a different central interaction inspired by the DLVO approach.

On the basis of the above analysis, we discuss the relationship between experimental results and features of the effective interactions, as well as possible paths toward a quantitative prediction of structural properties of real lysozyme solutions.

The paper is organized as follows: in Sec. II we introduce the coarse-grained models. Details of simulations are provided in Sec. III. The SANS structure factors are calculated and discussed in Sec. IV. The comparison between experimental and simulations data is reported and commented in Sec. IV. Conclusions follow in Sec. V.

II. DEFINITION OF THEORETICAL MODELS

As for the CML-like model mentioned in point 2 of Introduction, we recall that in our previous papers[9,18] we have considered an interaction potential written as the sum of two different contributions.

A central soft-core interaction in the form

$$V_{\mathrm{CML}}(r) = \varepsilon_{\mathrm{cml}} \left[\left(\frac{\sigma_{\mathrm{cml}}}{r} \right)^{48} - \left(\frac{\sigma_{\mathrm{cml}}}{r} \right)^{6} \right]. \qquad (1)$$

In Refs. 9 and 18, we have fixed the value $\sigma_{\mathrm{cml}} = 36.72$ Å so to represent an effective diameter of the lysozyme molecule and $\varepsilon_{\mathrm{cml}} = 4.84$ kJ/mol so to reproduce the experimental critical temperature of the protein-rich/protein-poor phase

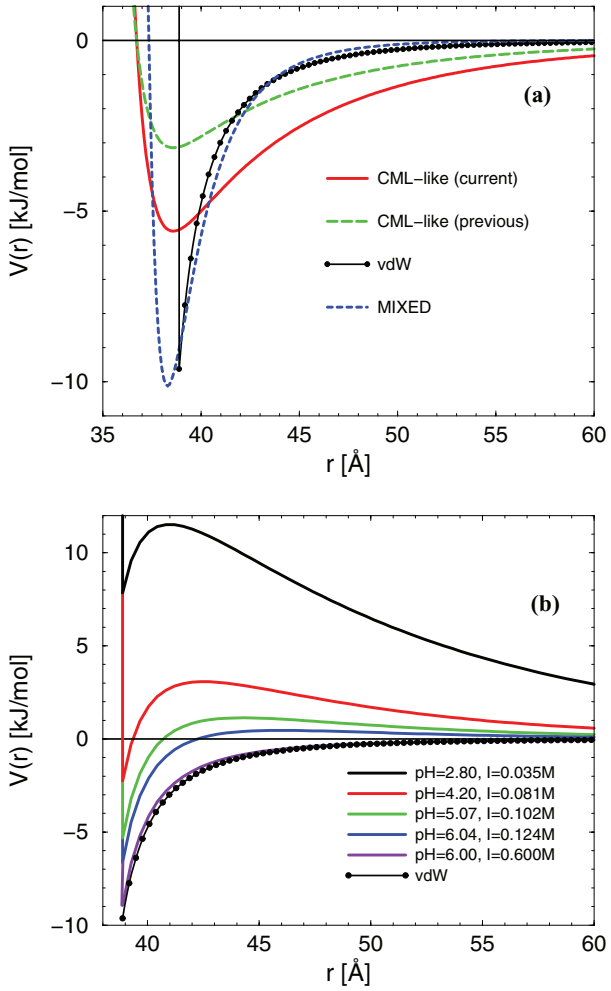

FIG. 1. Panel (a): Full line and dashed line: central parts of the CML-like model [see Eq. (1)] adopted, respectively, in this work ($\sigma_{\rm cml} = 36.72$ Å, $\varepsilon_{\rm cml} = 8.6$ kJ/mol) and in Refs. 9 and 18 ($\sigma_{\rm cml} = 36.72$ Å, $\varepsilon_{\rm cml} = 4.84$ kJ/mol). Line with circles: van der Waals contribution to the DLVO potential [see Eq. (4)] ($\sigma_{\rm dlvo} + \delta = 38.88$ Å, $A_{\rm H} = 18.8$ kJ/mol). Dotted line: central part of the mixed model [see Eq. (7)] ($\sigma_{\rm mix} = 38.0$ Å, $\varepsilon_{\rm mix} = 9.6$ kJ/mol). Panel (b): DLVO models ($\sigma_{\rm dlvo} + \delta = 38.88$ Å, $A_{\rm H} = 18.8$ kJ/mol) at pH and salt molarities reported in Table I; the case pH = 6 and $I = 0.6$ M corresponds to the experimental conditions of Taratuta *et al.*[40] The line with symbols is the van der Waals contribution also reported in panel (a).

separation in typical lysozyme solutions.[39] This central term is shown in Fig. 1(a) (dashed line).

A second contribution in the form of a distribution of charged sites, fixed at a distance of 16.54 Å from the center of the macromolecule.[37–39] The charges are determined according to the experimental titration curve[41] and depend on the solution pH (see Table I); they are treated in a linearized Poisson-Boltzmann framework giving rise to the Debye-Hückel (DH) screened interaction,

$$V_{\rm DH}(r) = \frac{Z_i Z_j e^2}{4\pi\epsilon_0\epsilon_{\rm r}} \frac{\exp(-\kappa r)}{r}, \tag{2}$$

where r is the distance between two charged sites on different macromolecules, Z_i is the charge of site i, e is the electronic charge, ϵ_0 is the permittivity of vacuum, and $\epsilon_{\rm r} = 86.765 - 0.3232 \times T$ (°C) (Ref. 39) is the relative permittivity of the solvent considered as a continuous medium. The Debye screening length κ is given by

$$\kappa = \sqrt{\frac{2Ie^2}{\epsilon_0\epsilon_{\rm r}k_{\rm B}T}}, \tag{3}$$

where I is the ionic strength of the solution (reported in Table I), taking into account the presence of ions due to the buffer and to the added salts.

As can be noted from Eqs. (2)–(3), once the solution conditions (protein charge, salt concentration, and temperature) are fixed, the electrostatic part of the full CML-like interaction is unequivocally determined. On the other hand, the parameters $\sigma_{\rm cml}$ and $\varepsilon_{\rm cml}$, entering the central interaction of Eq. (1), can be used to fit some appropriate property of the solution. In this work we leave $\sigma_{\rm cml} = 36.72$ Å unchanged, and we fix $\varepsilon_{\rm cml}$ so that our MD calculated scattered intensity reasonably agrees with the SANS data. We defer providing the details of such fitting procedure until Sec. IV; for comparison with the previous choice, we simply anticipate that we obtain the value $\varepsilon_{\rm cml} = 8.6$ kJ/mol and plot the resulting central potential in Fig. 1(a) (full line).

As for the DLVO description of the lysozyme solution (point 3 of Introduction), we recall that this model consists of (i) a short-range attractive van der Waals term,

$$V_{\rm vdW}(r) = -\frac{A_{\rm H}}{12}\left[\frac{\sigma_{\rm dlvo}^2}{r^2} + \frac{\sigma_{\rm dlvo}^2}{r^2-\sigma_{\rm dlvo}^2} + 2\ln\frac{r^2-\sigma_{\rm dlvo}^2}{r^2}\right], \tag{4}$$

where $\sigma_{\rm dlvo}$ is an effective hard-core diameter and $A_{\rm H}$ is the Hamaker constant, setting the depth of the potential well, and (ii) a Debye-Hückel term

$$V_{\rm DH}(r) = \frac{1}{4\pi\epsilon_0\epsilon_{\rm r}}\left[\frac{Ze}{1+\kappa\sigma_{\rm dlvo}/2}\right]^2 \frac{\exp[-\kappa(r-\sigma_{\rm dlvo})]}{r}, \tag{5}$$

where Ze is the net charge of the lysozyme macromolecule in electronic units, and with ϵ_0, $\epsilon_{\rm r}$, and κ having the same meaning previously introduced. The total DLVO potential is, therefore, written as

$$V_{\rm DLVO}(r) = \begin{cases} \infty & r < \sigma_{\rm dlvo} + \delta \\ V_{\rm vdW}(r) + V_{\rm DH}(r) & r \geq \sigma_{\rm dlvo} + \delta, \end{cases} \tag{6}$$

where the cut-off value δ, corresponding to the thickness of the Stern layer, is introduced to avoid the singularity of the van der Waals contribution at $r = \sigma_{\rm dlvo}$.

There are similarities between the DLVO and CML-like models. Both approaches treat the full solution as an effective one-component fluid. Besides the different form used to describe the cumulative effects of short-range forces, in both models the electrostatic interactions are represented in the same effective, salt-mediated form: in the DLVO approach only the net charge on the protein is taken into account, and considered as uniformly distributed over the macromolecular surface; in the CML multi-site approach all distinct (positive and negative) residual charges are explicitly considered. As in the CML case, the DLVO electrostatic contribution is completely determined by the properties of the solution, and one is left free only with the parameters describing the short-range attraction, namely $\sigma_{\rm dlvo}$ and $A_{\rm H}$.

We will assume in what follows $\sigma_{\rm dlvo} = 37.08$ Å, a value slightly larger than $\sigma_{\rm cml}$, and a Stern-layer thickness $\delta = 1.8$ Å, a value borrowed from previous DLVO calculations.[24–26] $A_{\rm H}$ is deduced in such a way that B_2, the second virial coefficient of the system (representing the overall strength of interaction), has a well-defined value at the liquid-vapor critical temperature of the model. Specifically, we recall that for a wide class of simple potentials $B_2^*(T_{\rm cr}) = B_2(T_{\rm cr})/B_2^{\rm HS} \approx -6.0$, where $B_2(T_{\rm cr})$ is the value of B_2 at the critical temperature $T_{\rm cr}$ and $B_2^{\rm HS}$ is the second virial coefficient of a fluid of hard spheres of diameter $\sigma_{\rm dlvo} + \delta$.[42] A similar analysis of the critical temperatures of different DLVO models yields:[24] $B_2^*(T_{\rm cr}) \approx -5.1$, a value close to the experimental outcome $B_2^*(T_{\rm cr}) = -5.3$ relative to lysozyme in water/NaCl solutions.[43,44] Therefore, once the critical temperature is known, such a condition on B_2 can be used to fix $A_{\rm H}$. We make reference to the phase diagram determinations by Taratuta *et al.*;[40] these authors estimated in particular $T_{\rm cr} \approx 273$ K for lysozyme solutions in a 0.6 M sodium phosphate buffer at pH = 6.0. Such pH is close to one of those investigated by SANS (see Table I); at the same time, the analysis of Taratuta *et al.*[40] demonstrates that the overall coexistence line is almost insensitive to the ionic strength of the solution. If we then assume pH = 6 [and hence $Z = 9$ in Eq. (5), see Table I], $T_{\rm cr} \approx 273$ K and $B_2^*(T_{\rm cr}) = -5.3$, we finally get $A_{\rm H} = 18.8$ kJ/mol. The Hamaker constant determined in this manner is then used for all cases considered in this work. Several considerations justify this choice: in the DLVO theory, electrostatic effects (Eq. (5)) and cumulative residual attractions (Eq. (4)) are treated at a separate level; moreover, the theory is insensitive to the nature of different salts used (sodium phosphate in Ref. 40 and citrate phosphate in our SANS experiments), which only determine the total ionic strength of the solution. Hence, it is reasonable to assume as a first approximation that changing salt and pH of the solution only affects the electrostatic contribution to the DLVO model, leaving unaltered the properties of the short-range attractions. As a proof of the validity of our reasoning, we have demonstrated in Refs. 24–26 that the experimental B_2 vs. I behavior can be accurately reproduced with a unique choice of the Hamaker constant. The fixed van der Waals contribution to the DLVO potential (Eq. (4)) is shown in both panels of Fig. 1 (line with circles); the different DLVO potentials for all cases listed in Table I, and for the cases envisaged by Taratuta *et al.*[40] are reported in Fig. 1(b). As visible, the conditions investigated in Ref. 40 give rise to an interaction potential substantially dominated by the van der Waals contribution. As the pH decreases, the net repulsive charge increases while, as visible in Table I, the ionic strength of the buffer decreases (resulting in a reduced screening length), so that the electrostatic repulsive contribution progressively dominates the overall microscopic interactions.

As for point 4 of the Introduction, we study a mixed model, taking into account the attractive part of the DLVO model determined above, and the advantages offered by the charge distribution of the CML-like model. As in Refs. 9 and 18 we replace the hard-core plus the attractive van der Waals contribution to the DLVO potential (see Eq. (4)) with a continuous analytic expression, suitable for MD simulations.

TABLE I. The different pH investigated in this work, with the corresponding lysozyme positive and negative charges (in units of e). I is the ionic strength of the solution (in mM).

pH	Z^+	Z^-	I
2.0	19	2	6
2.8	19	5	35
4.2	19	7	81
5.07	19	9	102
6.04	18	9	124

Specifically, we employ a soft-sphere potential defined as

$$V_{\rm MIX}(r) = \varepsilon_{\rm mix}\left[\left(\frac{\sigma_{\rm mix}}{r}\right)^{64} - \left(\frac{\sigma_{\rm mix}}{r}\right)^{32} - \left(\frac{\sigma_{\rm mix}}{r}\right)^{16}\right], \tag{7}$$

and calculate $\sigma_{\rm mix}$ and $\varepsilon_{\rm mix}$ in such a way that Eq. (7) yields a best overall fit both of the depth and of the large-distance behavior of the van der Waals contribution to the DLVO. This is obtained for $\sigma_{\rm mix} = 38$ Å and $\varepsilon_{\rm mix} = 9.6$ kJ/mol. As visible from Fig. 1(a) (dotted line), the minimum of potential (7) falls in the same position of CML-like models due to the chosen values of $\sigma_{\rm dvlo}$ and δ for the DLVO; its long-range decay appears considerably faster and its minimum considerably deeper than that of the previous central potential. To summarize, the total mixed model is constituted by the central soft-core interaction in Eq. (7), supplemented by the distribution of residual charges described by Eqs. (2) and (3).

III. SIMULATION TECHNIQUES

We have carried out molecular dynamics simulations for the CML-like and mixed multi-site models, by employing the program `moldy` (Ref. 45) with samples composed of $N = 500$ particles (with a total number of ≈15 000 interaction sites), enclosed in a cubic box with standard periodic boundary conditions. The interaction cutoff has been fixed to 130 Å, roughly corresponding to 3.5 times the protein diameter, with standard long-range corrections. We have adopted a time step $\Delta t = 20$ fs with cumulation times of about 10 ns, divided into blocks of 2 ns, in order to calculate the variance over the calculated average values. We have calculated the structure factors both as Fourier Transform of the center-center pair distribution functions, and by direct estimate of average correlations of density fluctuations. Tests of the dependence of results on the system size have been performed on samples of $N = 1000$ lysozyme molecules.

As for the DLVO model, we have carried out Monte Carlo simulations with samples of 512 particles and, occasionally, with 2048 particles, in order to better characterize the structure factor at small wavevectors. After equilibration periods of generally 10^5 MC steps, we have followed the samples for further 10^4–10^5 cumulation steps in order to calculate the structural properties of the fluid. GEMC simulations[46,47] have been carried out on samples composed by $N = 1024$ particles. The trial number of sweeps of particles between the simulation boxes is sufficient to ensure that at least 2%–3% of the total number of particles are exchanged during the transfer routine. We have calculated averages over 10^5 GEMC cycles, generally divided into three blocks to calculate the statistical

uncertainties. Starting from the GEMC coexisting densities along several isotherms, we have estimated the critical point by means of the law of rectilinear diameter and the scaling law for the densities with a non-classical critical exponent $\beta = 0.32$.[47]

IV. RESULTS AND DISCUSSION

A. SANS structure factors

A presentation of SANS experiments concerning this paper is given in Ref. 9 which we refer to for full details. We here shortly recall that we have analyzed lysozyme samples prepared at ambient conditions and protein concentration of 10% by weight (100 mg/ml, i.e., 6.983 mM, corresponding to a number density of 4.2×10^{-6} Å^{-3}) in D_2O citrate-phosphate buffer at pH = 2.8, 4.2, 5.07, and 6.04; the latter are reported in Table I, together with new data at pH = 2.0, and with the ionic strengths of the solutions and the number of residual positive and negative charges on lysozyme. In a SANS experiment, the scattered intensity is expressed as

$$I(q) = n\Delta\rho^2 V^2 F(q) S(q) + B(q), \tag{8}$$

where n is the number density of the scatterers; V is the volume of a single scatterer; $\Delta\rho$ is the contrast factor accounting for the difference between the average (over the volume V) scattering cross section of the protein and the solvent; $F(q)$ and $S(q)$ are, respectively, the form factor and the structure factor; $B(q)$ cumulatively represents the contribution to the scattered intensity due to incoherent and instrumental background. Equation (8) strictly holds for monodisperse spherical particles,[48] but corrections turn out negligibly small in our case.[10] Experimental $I(q)$ at various pH and ionic strengths[9] are shown in Fig. 2(a), along with the form factor $F(q)$ calculated by using a diluted sample at high ionic strength (2% by weight concentration and 0.07 M NaCl of added salt) so to make protein-protein correlations practically negligible. Taking into account that lysozyme has approximately the shape of a prolate ellipsoid,[49,50] we fit the experimental form factor against the form factor of a biaxial ellipsoidal particle,[9] with axes a, b, $c \equiv b$ and volume $V = 4\pi(abc)/3$, averaged over all possible orientation, given by[51]

$$F(q) = \int_0^1 \left(\frac{3J_1(u)}{u}\right)^2 \mathrm{d}u, \tag{9}$$

where J_1 denotes the first-order spherical Bessel function and u is a function of the angle α between q and a, defined as $u = q\sqrt{a^2\alpha^2 + b^2(1-\alpha^2)}$. The best-fit is also reported in Fig. 2(a); values for a and b are, respectively, 13.5 Å and 21.9 Å,[9] in good agreement with previous findings.[52]

Once $I(q)$ and $F(q)$ are known, Eq. (8) can be used to calculate the experimental structure factor $S(q)$. This task poses some difficulties related to the exact determination of the prefactor $n\Delta\rho^2V^2$ and the background function $B(q)$, as well as to the limited range of explored wavevectors. Moreover, in the case of lysozyme, the form factor attains a value of ≈0.4% of its initial values at $q \approx 0.3$ Å^{-1} making the signal noisy in the high-q portion of the spectrum. Here, in order to determine the best choice for the prefactor and background, we have enforced the following physical constraints in a least-square

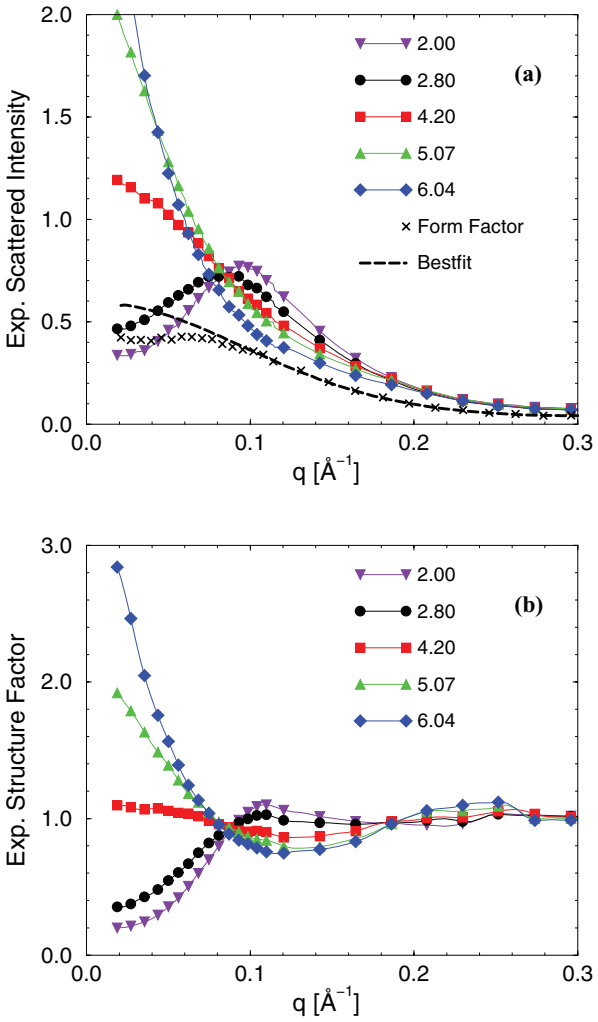

FIG. 2. SANS scattered intensity (a) and structure factor (b) at different pH. In (a) the form factor $F(q)$ and the corresponding best-fit are also shown.

minimization routine:[53,54]

$$S(q) = 1 \text{ for } q > q_{\max}, \tag{10}$$

$$\int_0^\infty (S(q) - 1)\, q^2\, \mathrm{d}q = -2\pi^2 n, \tag{11}$$

$$g(r) = 0 \text{ for } r \le 0.8 D_{\min}, \tag{12}$$

where $q_{\max} = 0.28$ Å^{-1} is the value beyond which we assume that the structure factor attains its high-q behavior and $D_{\min} = 27$ Å is a minimum approach distance, set equal to the smaller diameter of the ellipsoid obtained from the fit of the form factor. The center-center radial distribution function $g(r)$ is related to the Fourier transform of $S(q)$ according to

$$g(r) - 1 = (2\pi^2 nr)^{-1} \int_0^\infty [S(q) - 1]\, q \sin(qr)\, \mathrm{d}q\,. \tag{13}$$

In the fitting procedure, the total chi-square has been calculated as the weighted sum of the partial chi-squares obtained by each constraint. For the first few iterations, higher weights have been given to the first and second conditions, Eqs. (10) and (11), in order to obtain an approximate $S(q)$. Then the third condition, Eq. (12) has been enforced to the

derived $g(r)$ (then back-transformed into $S(q)$), with an increasing weight of the relative chi-square toward the end of the minimization procedure. Eventually, one gets a structure factor that consistently derives from the experimental intensity. As for the prefactor in Eq. (8), we have obtained values ranging from 1.11 cm^{-1} to 0.84 cm^{-1} as the pH increases, closely matching the theoretical estimate given in Ref. 9, $n\Delta\rho^2 V^2 \approx 1.02\ cm^{-1}$, based on the assumption that the scattering cross section of the solvent can be approximated with that of pure water. As for the background contribution $B(q)$ in Eq. (8), we have obtained a weak linear dependence on q.

All structure factors at different pH are reported in Fig. 2(b). In our experimental setup, the increase of pH, giving rise to a progressive decrease of the net protein charge, is associated with an increase of the ionic strength of the buffer (see Table I), resulting in a further reduction of the electrostatic repulsions. Upon such concurrent effects, the protein solution undergoes a change of its structural arrangement due to the progressive prevalence of attractive over repulsive forces, reflected in the pronounced uprise of $I(q)$ and $S(q)$ as q goes to zero, visible in Fig. 2. At small pH, $S(q)$ develops a maximum around 0.1 $Å^{-1}$, testifying the presence of a correlation distance of about 60 Å in direct space; this distance corresponds to the average interparticle distance at the assumed protein density, and is roughly equal to two protein (average) diameters; this outcome suggests the dominance, in this pH regime, of repulsive effects that tend to correlate the macroparticles on large distances. At the intermediate value $p\mathrm{H} = 4.2$, $S(q)$ remains almost horizontal and practically featureless over the whole wavevector range; this seems to signal that repulsive effects at such pH are more effective than it could be concluded on the basis of the corresponding $I(q)$ alone, showing instead a well-defined uprise as $q \to 0$. When the pH further increases, $S(q \to 0)$ increases, thus indicating that density fluctuation correlations are established on scale considerably greater than a single macroparticle diameter; at the same time the maximum in $S(q)$ at $q = 0.1\ Å^{-1}$ shifts to $q = 0.2\ Å^{-1}$, corresponding to a single protein diameter in direct space. Fluctuations possibly indicate the formation of protein aggregates, while the intermediate-q peak can be associated with the correlation distances between two macromolecules inside a given aggregate.[11] It is not obvious to establish whether the density fluctuations prelude to a protein-rich/protein-poor phase separation; this might sensitively depends on how far the thermodynamic point is from the actual phase coexistence line: if we assume for reference a phase behavior similar to that investigated in Ref. 40 (see also Fig. 4), our conditions should fall $\sim 30^\circ$ above the cloud point temperature corresponding to the actual protein concentration (100 mg/ml). On the other hand, the decrease of the net repulsive charge and the increase of the ionic screening may herald the flocculation occurring at the isoelectric point.[10]

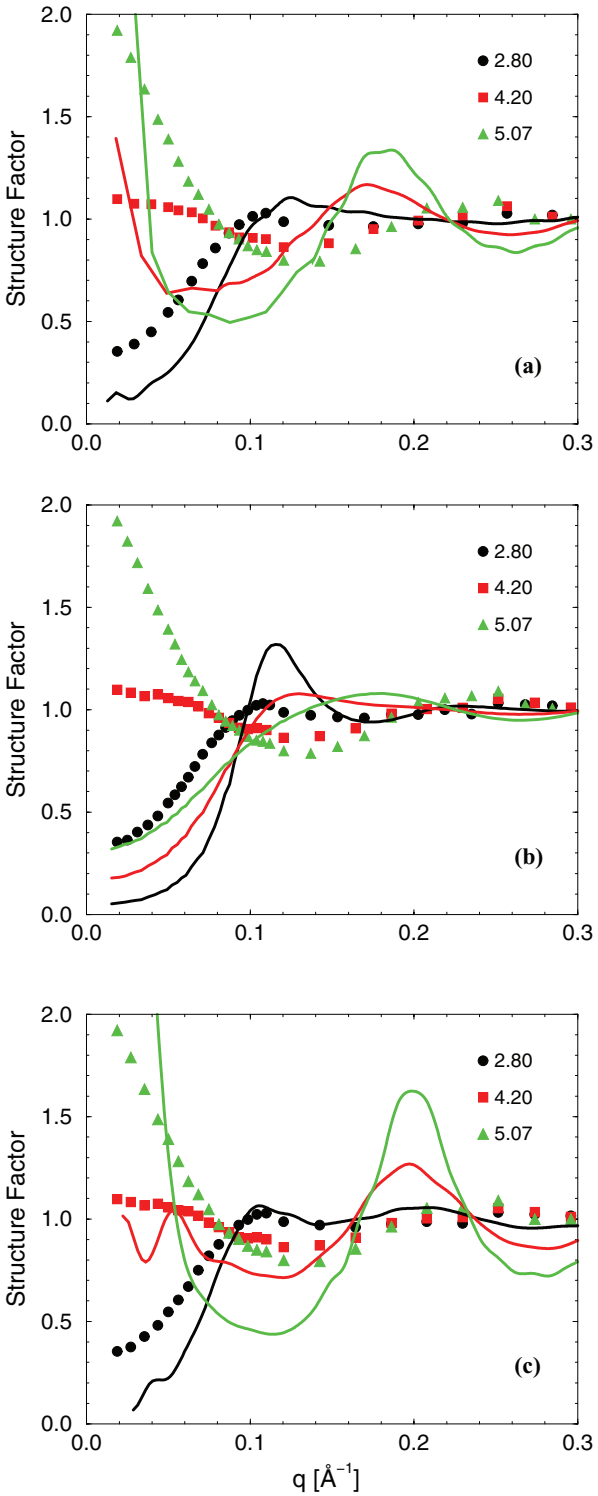

FIG. 3. Structure factor of the CML-like (a), DLVO (b), and mixed (c) models (lines), with corresponding SANS data (symbols), at $p\mathrm{H} = 2.8$, 4.2, and 5.07.

B. Comparison between simulation and experiments

Structural results for all models envisaged in this work are collectively displayed in Fig. 3, along with their experimental counterpart; specifically, we show the $S(q)$ for the CML-like, the DLVO, and the mixed models only at the three intermediate pH values 2.8, 4.2, and 5.07, though simulations have been systematically carried for all entries reported in Table I; such a choice, while avoiding to overcrowd the figure, turns out completely sufficient to present and assess the main trends emerging from our study.

As for the MD simulations of the CML-like model, we take in Eq. (1) $\sigma_{\mathrm{cml}} = 36.72$ Å,[9,18] while we progressively

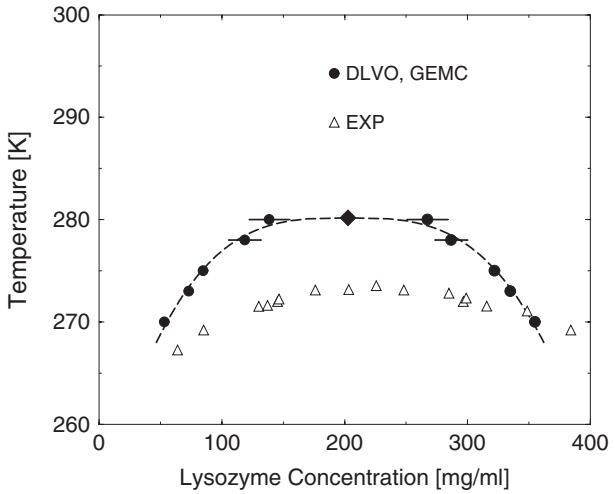

FIG. 4. GEMC liquid-vapor coexistence of the DLVO model at pH = 6.0 and $I = 0.6$ M (circles with error bars) and corresponding experimental protein-poor/protein-rich phase coexistence (triangles).[40] The dashed line is the interpolation of GEMC points based on the scaling law for the densities and the law of rectilinear diameters,[47] with the resulting critical point (diamond).

increase $\varepsilon_{\rm cml}$ from the previously used 4.84 kJ/mol value until we obtain—for $\varepsilon_{\rm cml} = 8.6$ kJ/mol —the crossover between the different behaviors observed in the SANS data at pH = 4.2 (see Fig. 2). As can be seen from Fig. 3(a), the MD correlations reproduce the experimental crossover at pH = 4.2, even if the general agreement between theoretical patterns and SANS data is only qualitative. At the lowest pH in Fig. 3(a) (i.e., pH = 2.8), the CML-like model reproduces fairly well the overall shape of the experimental pattern, with the presence of a correlation peak at $q \approx 0.12$ Å^{-1}. The accuracy of the MD correlations appear less satisfactory at higher pH; in particular, the MD patterns exhibit too marked an uprise at small wavevectors, and a peak at $q \approx 0.18$ Å^{-1}, whereas the experimental curves show less defined features. To summarize, the overall reproduction of experimental trends obtained through the current CML-like approach appears only qualitatively adequate. We recall in this instance that a similar behavior was early observed in the original CML model investigations[37] and referred to as in qualitative agreement with the experimental data of Refs. 10 and 55. Our results illustrate that such agreement can be improved if a relatively deeper minimum of the central soft-core interaction is adopted, in comparison with previous prescriptions[9,37,39] (see Fig. 1(a)).

The DLVO structure factors are shown in Fig. 3(b), while the liquid-vapor coexistence (deduced at pH = 6 and $I = 0.6$ M) is reported in Fig. 4, along with the experimental protein-poor/protein-rich coexistence line calculated in the same conditions.[40] It appears that the DLVO $S(q)$ for small wavevectors shifts upward with increasing pH but fails to reproduce the pronounced uprise observed in the experiments. Moreover, the maxima visible at higher wavevectors only roughly correspond to the features of experimental patterns, by also showing a sensible dephasing at high pH. On the other hand, the qualitative aspects of the experimental phase separation are well reproduced; in particular, the relatively flat shape of the coexistence curve. This evidence further corroborates the conclusions drawn in our previous studies,[24–26] about the good quality of DLVO predictions for phase coexistences in real protein solutions. According to our estimate of critical temperature and lysozyme concentration, $T_{\rm cr} \approx 280$ K and $c_{\rm cr} \approx 200$ mg/ml, theoretical predictions bring a 6°–7° overestimate of the critical temperature, a result we consider as a relatively successful outcome, if one takes into account that we have determined the Hamaker constant through the empirical choice $B_2^* = -5.3$ at $T = 273$ K. Test simulations indicate that the liquid-vapor coexistence curve weakly depends on the Hamaker constant adopted, whereas the $S(q)$ turns out to sensitively depend on such a choice. In this respect, a more refined tuning of the Hamaker constant might lead to improved predictions for both the thermodynamic and the structural quantities.

The structure factors obtained by MD simulation of the mixed model are reported in Fig. 3(c). As visible, the correlation pattern at pH = 4.2 is better approximated than by other models. The overall monotonic increase of $S(q)$ as q goes to zero is also correctly displayed by all MD curves, although a sharp correlation peak at $q \approx 0.2$ Å^{-1} is predicted (similarly to the previous CML-like model), at variance with the experimental patterns.

The body of our results indicates the emergence of specific relationships between microscopic interactions, structure factors, and phase coexistence lines. In particular, the different $S(q)$ behavior between the CML-like and DLVO models appears due to an overestimate in the latter approach of electrostatic repulsive effects with respect to attractive contributions. In fact, the attractive (van der Waals) contribution to the DLVO model is deeper than the one entering the CML-like model, but it decays much faster; at the same time we have in the DLVO a net positive charge, rather than a set of charges of opposite sign, and this results in an increased repulsion. Then, a remarkable feature emerges from the comparison of the $S(q \to 0)$ behavior of the CML-like model and the mixed one: as visible in Fig. 1(a) the latter is characterized by a central potential-well considerably deeper than the former. Now, the deepening of the potential in the CML-like model results in a considerable enhancement of the $S(q)$ at small wavevectors (see for comparison Ref. 9); one could then expect that in the mixed model the upward bending of $S(q \to 0)$ should be even greater. However, this is not what happens: in the mixed model a different balance of interactions—more repulsive because of the shorter range of the attractive potential, see Fig. 1(a)—leads to a better prediction of small-wavevector correlations, i.e., to a better reproduction of the experimental crossover between different regimes as a function of pH, though correlations at larger wavevectors appear generally overemphasized.

It is worth noting that our approach relies in all cases on a spherical description of lysozymes, against the actual ellipsoidal shape of the real protein. We then expect that the theoretical $S(q)$ may depend upon such a choice, and possibly our comparison with experimental data becomes only indicative beyond $q \approx 0.2$ Å^{-1}, corresponding to interparticle distances $\simeq 31$Å, roughly comparable to the effective core diameter of our models. We also expect that the same spherical description sensibly affects the structural results as the pH

increases, since the corresponding decrease of electrostatic repulsive effects allows particles to get close enough so to majorly test the assumed approximate shape and size.

As for the phase coexistence, we recall that CML-like models give rise to narrow, parabolic-like liquid-vapor coexistence curves, at variance with the experimental flat protein-rich/protein-poor phase separation;[39] this evidence has been attributed to the prevalence of attractive over electrostatic repulsive interactions.[39] In our case, the deeper potential-well employed to reproduce the crossover at $p\mathrm{H} = 4.2$ plausibly leads to an enhancement of the parabolic shape of the phase diagram. On the other hand, the balance between attractive and repulsive contributions associated with the DLVO interaction discussed before is able to produce a flat coexistence curve in the corresponding phase diagram.

It appears that the possibility to reproduce both the phase portrait and the structural correlations of real solutions by means of a unique model still remains an elusive task. In this context, the question naturally arises whether bringing together the advantage of the charge distribution and the DLVO short-range features, as we have done for the mixed model, is enough to represent accurately both such aspects. The answer requires an extensive test of distributed-charge models with different choices of their central potential interaction, with the contextual determination of the related phase diagram, a computationally demanding task to deal with in future work.

V. CONCLUSIONS

We have reported SANS data for the structure factors $S(q)$ of aqueous lysozyme solutions at ambient temperature conditions, for a fixed 10% by weight protein concentration, at pH in the range of 2.0–6.04 and ionic strength of the buffer in the range of 6–124 mM. We have described in detail the best fit procedure used to reconstruct the structure factors from the SANS intensities. The $S(q)$ shows a marked increase at small wavevectors as the pH and ionic strength of the solution increase, signaling a crossover from a physical regime mostly dominated by repulsive forces to one in which attractive interactions become prominent.

The experimental results have allowed us to assess the predictions of various effective microscopic models with different levels of coarse-graining. In the first model, proteins have been represented as soft spheres characterized by an attractive central r^{-6} potential plus a distribution of interacting sites reproducing the positions and pH-dependent charges of amino-acid residues. Molecular dynamics simulations qualitatively reproduce the uprise of the $S(q)$ as a function of the pH. We have also assessed the predictions of a DLVO model of the solution. We have preliminary shown through Gibbs ensemble Monte Carlo simulations that the protein-rich/protein-poor phase coexistence line of a lysozyme solution with $p\mathrm{H} = 6.0$ and 0.6 M ionic strength is satisfactorily reproduced. The structure factors, determined through Monte Carlo simulations, miss the uprise of the scattering patterns at higher pH, although the structural results turn out to sensitively depend on the specific values of DLVO parameters. Hinging on the previous analysis, we have examined by molecular dynamics a third model in which the distributed charge framework is taken together with the van der Waals interaction borrowed from the DLVO model. It turns out that the much shorter range of this potential leads to a better reproduction of the experimental crossover than previously obtained.

Overall, the comparison between experiment and simulation, though not being satisfactory, allows us to identify trends and to conjecture on the roles of the shape of the potential as well as of the charge distribution, in determining the accuracy of the envisaged coarse-grained models; it also suggests possible strategies to improve such a description. In particular, as we have alluded in Sec. IV, an extensive analysis concerning both the liquid phase boundaries and the structural correlations, is currently underway. Moreover, by hinging on the manner in which the envisaged models are built, we plan to improve the description of electrostatic interactions and of the molecular shape. A promising path for such a purpose is the calculation of effective interactions between "patches" on the surface of real proteins,[56] based on fully atomistic simulations of proteins in solution. Also, we are investigating the possibility to take into account the ellipsoidal form factor of lysozyme, whose importance in interpreting our simulation results in comparison with experimental data has been illustrated in the text.

ACKNOWLEDGMENTS

We gratefully acknowledge the computer resources (AMD Opteron Cluster of the former TriGrid project) and the technical support provided by the Centro di Calcolo Elettronico dell'Università di Messina "A. Villari". G.P. acknowledges NRF support under the KIC Grant No. 76513 and the kind support by the College of AES, Dean of Research, Professor D. Jaganyi.

[1] A. Tardieu, A. Le Verge, M. Malfois, F. Bonneté, S. Finet, M. Ries-Kautt, and L. Belloni, J. Cryst. Growth **196**, 193 (1999).
[2] R. A. Curtis, C. Steinbrecher, M. Heinemann, H. W. Blanch, and J. M. Prausnitz, Biophys. Chem. **98**, 249 (2002).
[3] A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf, and P. Schurtenberger, Nature (London) **432**, 492 (2004).
[4] J. N. Israelachvili, *Intermolecular and Surface Forces, 2nd ed.* (Academic, London, 1992).
[5] R. Piazza, Curr. Opin. Colloid Interface Sci. **5**, 38 (2000).
[6] A. J. Rowe, Biophys. Chem. **93**, 93 (2001).
[7] A. McPherson, *Crystallization of Biological Macromolecules* (Cold Spring Harbor, New York, 1999).
[8] B. Lonetti, E. Fratini, S.-H. Chen, and P. Baglioni, Phys. Chem. Chem. Phys. **6**, 1388 (2004).
[9] M. C. Abramo, C. Caccamo, M. Calvo, V. C. Nibali, D. Costa, R. Giordano, G. Pellicane, R. Ruberto, and U. Wanderlingh, Philos. Mag. **91**, 2066 (2011).
[10] R. Giordano, A. Grasso, J. Teixeira, F. Wanderlingh, and U. Wanderlingh, Phys. Rev. A **43**, 6894 (1991).
[11] Y. Liu, E. Fratini, P. Baglioni, W.-R. Chen, and S.-H. Chen, Phys. Rev. Lett. **95**, 118102 (2005).
[12] A. Stradner, F. Cardinaux, and P. Schurtenberger, Phys. Rev. Lett. **96**, 219801 (2006).
[13] Y. Liu, E. Fratini, P. Baglioni, W.-R. Chen, L. Porcar, and S.-H. Chen, Phys. Rev. Lett. **96**, 219802 (2006).
[14] Y. Liu, L. Porcar, J. Chen, W.-R. Chen, P. Falus, A. Faraone, E. Fratini, K. Hong, and P. Baglioni, J. Phys. Chem. B **115**, 7238 (2011).
[15] L. Ianeselli, F. Zangh, M. W.A. Skoda, R. M. J. Jacobs, R. A. Martin, S. Callow, S. Prevost, and F. Schreiber, J. Phys. Chem. B **114**, 3776 (2010).
[16] F. Sciortino, S. Mossa, E. Zaccarelli, and P. Tartaglia, Phys. Rev. Lett **93**, 055701 (2004).

[17]F. Cardinaux, E. Zaccarelli, A. Stradner, S. Bucciarelli, B. Farago, S. U. Egelhaaf, F. Sciortino, and P. Schurtenberger, J. Phys. Chem. B **115**, 7227 (2011).
[18]M. C. Abramo, C. Caccamo, D. Costa, G. Pellicane, and R. Ruberto, J. Phys. Chem. B **114**, 9109 (2010).
[19]E. J. W. Verwey and J. T. G. Overbeek, *Theory of the Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
[20]B. V. Derjaguin and L. Landau, Acta Physicochim. USSR **14**, 633 (1941).
[21]R. Piazza, J. Cryst. Growth **196**, 415 (1999).
[22]D. N. Petsev and P. G. Vekilov, Phys. Rev. Lett. **84**, 1339 (2000).
[23]M. Böstrom, D. R. M. Williams, and B. W. Ninham, Phys. Rev. Lett. **87**, 168103 (2001).
[24]G. Pellicane, D. Costa, and C. Caccamo, J. Phys.: Condens. Matter **15**, 375 (2003).
[25]G. Pellicane, D. Costa, and C. Caccamo, J. Phys.: Condens. Matter **15**, S3485 (2003).
[26]G. Pellicane, D. Costa, and C. Caccamo, J. Phys. Chem. B **108**, 7538 (2004).
[27]M. Malfois, F. Bonneté, L. Belloni, and A. Tardieu, J. Chem. Phys. **105**, 3290 (1996).
[28]A. Lomakin, N. Asherie, and G. B. Benedek, J. Chem. Phys. **104**, 1646 (1996).
[29]D. F. Rosenbaum, A. Kulkarni, S. Ramakrishnan, and C. F. Zukoski, J. Chem. Phys. **111**, 9882 (1999).
[30]D. Rosenbaum, P. C. Zamora, and C. F. Zukoski, Phys. Rev. Lett. **76**, 150 (1996).
[31]D. Costa, P. Ballone, and C. Caccamo, J. Chem. Phys. **116**, 3327 (2002).
[32]M. Broccio, D. Costa, Y. Liu, and S.-H. Chen, J. Chem. Phys. **124**, 084501 (2006).
[33]F. Cardinaux, A. Stradner, P. Schurtenberger, F. Sciortino, and E. Zaccarelli, Europhys. Lett. **77**, 48004 (2007).
[34]A. Lomakin, N. Asherie, and G. B. Benedek, Proc. Natl. Acad. Sci U.S.A. **96**, 9465 (1999).
[35]H. Liu, S. K. Kumar, and F. Sciortino, J. Chem. Phys. **127**, 084902 (2007).
[36]A. Shiryayev, X. Li, and J. D. Gunton, J. Chem. Phys. **125**, 024902 (2006).
[37]F. Carlsson, M. Malmsten, and P. Linse, J. Phys. Chem. B **105**, 12189 (2001).
[38]F. Carlsson, P. Linse, and M. Malmsten, J. Phys. Chem. B **105**, 9040 (2001).
[39]T. W. Rosch and J. R. Errington, J. Phys. Chem. B **111**, 12591 (2007).
[40]V. G. Taratuta, A. Holschbach, G. M. Thurston, D. Blankschtein, and G. B. Benedek, J. Phys. Chem. **94**, 2140 (1990).
[41]M. Ramanadham, L. C. Sieker, and L. H. Jensen, Acta Crystallogr. B **46**, 63 (1990).
[42]G. A. Vligenthart and H. N. W. Lekkerkerker, J. Chem. Phys. **112**, 5364 (2000).
[43]M. Muschol and F. Rosenberger, J. Chem. Phys. **107**, 1953 (1997).
[44]A. George and W. Wilson, Acta Crystallogr. D **50**, 361 (1994).
[45]K. Refson, Comput. Phys. Commun. **126**, 310 (2000).
[46]A. Z. Panagiotopoulos, Mol. Phys. **61**, 813 (1987).
[47]D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic, New York, 1996).
[48]M. Kotlarchyk and S.-H. Chen, J. Chem. Phys. **79**, 2461 (1983).
[49]N. C. J. Strynadka and M. N. G. James, J. Mol. Biol. **220**, 401 (1991).
[50]H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, and L. J. Smith, Protein Sci. **10**, 677 (2001).
[51]A. Guinier and G. Fournet, *Small Angle Scattering of X-rays* (Wiley, New York, 1955).
[52]B. Jacrot, Rep. Prog. Phys. **39**, 911 (1976).
[53]J. L. Yarnell, M. J. Katz, R. G. Wenzel, and S. H. Koenig, Phys. Rev. A **7**, 2130 (1973).
[54]C. D. Andriesse and E. Legrand, Physica **57**, 191 (1972).
[55]Y. Georgalis, P. Umbach, W. Saenger, B. Ihmels, and D. M. Soumpasis, J. Am. Chem. Soc. **121**, 1627 (1999).
[56]G. Pellicane, G. Smith, and L. Sarkisov, Phys. Rev. Lett. **101**, 248102 (2008).