Theory and simulation of short-range models of globular protein solutions

2004 J. Phys.: Condens. Matter 16 S4923

(http://iopscience.iop.org/0953-8984/16/42/010)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **16** (2004) S4923–S4936 PII: S0953-8984(04)78938-9

Theory and simulation of short-range models of globular protein solutions

G Pellicane, D Costa and C Caccamo

Istituto Nazionale per la Fisica della Materia (INFM) and Dipartimento di Fisica, Università di Messina, Contrada Papardo, CP 50, 98166 Messina, Italy

E-mail: Carlo.Caccamo@unime.it

Received 1 April 2004
Published 8 October 2004
Online at stacks.iop.org/JPhysCM/16/S4923
doi:10.1088/0953-8984/16/42/010

Abstract

We report theoretical and simulation studies of phase coexistence in model globular protein solutions, based on short-range, central, pair potential representations of the interaction among macro-particles. After reviewing our previous investigations of hard-core Yukawa and generalized Lennard-Jones potentials, we report more recent results obtained within a DLVO-like description of lysozyme solutions in water and added salt. We show that a one-parameter fit of this model, based on static light scattering and self-interaction chromatography data in the dilute protein regime, yields demixing and crystallization curves in good agreement with experimental protein-rich–protein-poor and solubility envelopes. The dependence of cloud and solubility point temperatures of the model on the ionic strength is also investigated. Our findings highlight the minimal assumptions on the properties of the microscopic interaction sufficient for a satisfactory reproduction of the phase diagram topology of globular protein solutions.

1. Introduction

Phase coexistence in aqueous solutions of globular proteins, like for instance lysozyme in water and added salt, has been intensively investigated in most recent years from experimental and theoretical points of view (see e.g. [1–7] and references therein). Such an interest is primarily intended to clarify the mechanism of protein crystallization, a process whose control, though of crucial importance for the study of protein structure, has not yet been characterized in terms of rigorous and predictive experimental protocols [8, 9].

Indeed, the literature documents a diffuse resort to trial-and-error procedures in the attempt to grow 'good' crystals from the mother solutions [9]; this is also due to a lack of an accurate microscopic description of the overall phase behaviour, which depends on many factors: for instance, even assuming that a satisfactory description of the interactions among 'bare' proteins

were available, understanding how these interactions are altered by the solution variables is a difficult task, also considering that specific salt effects determine a ranking of various ions, which is well-known as the Hofmeister series [10]. Despite the strong dependence on solution variables, several globular protein solutions share important features, when crystallization conditions are approached. In particular, the second virial coefficient B_2 of many protein solutions falls in a narrow window of small negative values, whenever the transition takes place [11]. In close relationship with such an occurrence, the phase diagram is characterized by a metastable protein–protein demixing region located just below the solubility line [4], a peculiar property of crucial interest, since experimental and numerical studies have shown that the proximity of the critical point to the solubility line should favour the formation of good crystals [4, 12, 13]. Besides the crystallization properties, the protein demixing binodal also plays an important role in several human diseases [14].

These experimental evidences have been rationalized in terms of effective protein–protein interactions described through short-range, centrosymmetric potentials [2, 3, 6, 15–17], and such models have been intensively investigated by means of simulation and liquid state theories. On the other hand, the rich phase behaviour of short-range attractive models has challenged a straightforward generalization to such systems of concepts and methods derived from the theory of simple fluids. In fact, it has been recognized (see [18] and references therein) that the onset of freezing, and generally the overall phase behaviour, is substantially affected by the 'perturbative' part (with respect to the repulsive core) of the potential, rather than being dominated by excluded-volume and packing (i.e. by entropic) effects, as is the case in the *van der Waals picture* of simple liquids [19]. The observation of two glassy states, originating from different cage mechanisms which drive the structural arrest of the system (see [20, 21] and references therein), recently confirmed by experiments [22], is another aspect of the unusual physical properties of systems interacting through short-range forces.

In this context, one of the most extensively studied model is represented by the hard-core Yukawa fluid (HCYF). In fact, early simulations [23] for this system showed that, when the decay of the attractive tail is sufficiently fast, the phase diagram is topologically similar to that of most crystallizing protein solutions, i.e. with a metastable liquid–vapour binodal lying beneath the sublimation line. Yukawa fluids have been studied by other authors (see [24] and references therein) and by us [25–27], in terms of integral equation theories (IET) of liquids [28]. The utility of theoretical approaches, along with computer simulation investigations, arises from several circumstances. In fact, simulations of systems characterized by short-range attractive forces, as in the case of charge-stabilized protein solutions, might be affected by ergodicity problems; similarly, the extension to more realistic multi-component cases (where proteins, solute ions and water molecules are explicitly considered) faces severe difficulties, related to the strong size asymmetry of the particle species, and to the usually high dilution of the macromolecules. Such studies would greatly benefit from the availability of reliable theoretical approaches to deal, for instance, with the multicomponent configuration of the system. We have devoted a considerable effort [25] to test refined IETs against rigorous simulation data, starting from the simplest one-component HCYF case. We succeeded in particular in the optimization of the thermodynamic consistency, a constraint which is often imposed in the solution of IETs in order to improve their accuracy [29]. In this case, we have investigated both the HCYF and a simple DLVO model [30] of proteins, widely used in colloid physics.

As for the determination of the solid–fluid coexistence, we have calculated the freezing line of envisaged models [25] in terms of one-phase freezing criteria proposed by various authors [31, 32]. We have assessed the accuracy of such criteria in [33], where theoretical predictions for the Girifalco interaction potential of C_{60} fullerene [34], characterized by a short-range decay, have been compared with the solid–fluid equilibrium determined from free

energy Monte Carlo simulations. Interest in such calculations stems from the observation [23] that the Boyle temperature of the C_{60} model [34] is reproducible with a somewhat short-range HCYF potential. Moreover, the Girifalco potential is currently employed as a robust benchmark to test the performances of new simulation strategies and theoretical approaches, when applied to fluids with short-range interactions (see [35–37]). We summarize in section 2 our most recent investigations of the HCYF and the Girifalco potentials.

We also review in section 2 our results [38] on the crystallization processes occurring in another model for protein solutions, namely the generalized Lennard-Jones potential proposed by ten Wolde and Frenkel [12]. The phase diagram of the model was calculated in [12] through computer simulation estimates of the free energy, and the most favourable conditions for crystal growth from the solution were identified. We have carried out extended molecular dynamics calculations, in order to compare our results for the crystallization kinetics with the free energy calculations reported in [12], and with experimental evidences in real protein solutions.

In section 3 we report on recent investigations [39–41] of the phase diagram of lysozyme solutions in water and sodium chloride added salt, modelled in terms of a flexible representation of protein–protein interactions, where the forces acting among macromolecules are described by means of a DLVO-like potential [30]. We have been prompted to such an investigation since other models—including the HCYF and the generalized Lennard-Jones potentials—have hitherto yielded only a qualitative reproduction of the experimental phase diagram, and allowed at best to fit the demixing line [3, 15]; these models also fail to capture the sensitivity of the metastable demixing line to solution conditions [6]. We have performed both IET and simulation calculations for the protein-rich–protein-poor coexistence curve (the metastable 'liquid–vapour' binodal), and determined the solubility line through a simplified version [42–45] of the cell theory [46] to estimate the free energy of the solid phase. We also show theoretical results for the dependence of the cloud and solubility temperatures on the ionic strength [40], in comparison with experimental data for lysozyme solutions. Our conclusions follow in section 4.

2. Theoretical and simulation studies of model fluids interacting through short-range forces

In the last few years we have undertaken an extensive investigation of the phase behaviour of the HCYF [25–27], and assessed at the same time the performances of most advanced liquid state theories presently available (both semi-analytic and numerically solvable). We recall that the hard-core Yukawa potential is written as

$$v(r) = \begin{cases} \infty & r < \sigma \\ -\sigma\epsilon \exp[-z(r-\sigma)]/r & r \geqslant \sigma, \end{cases} \tag{1}$$

where σ is the hard-core diameter, ϵ is the potential depth at closest contact and z is the inverse screening length (see figure 1). The properties of the HCYF are calculated in the context of several liquid state theories, like for instance the modified hypernetted chain approximation (MHNC, [47]), the generalized mean spherical approximation (GMSA, [49]) and the self-consistent Ornstein–Zernike approximation (SCOZA, [48]). We briefly review the basic features of such theories in the appendix, and refer to other papers (see [24, 47–49] and references therein) for a detailed presentation of their solution schemes.

The theoretical results reviewed here concern an interaction range of the HCYF corresponding to a realistic model of colloidal suspensions and protein solutions [2, 3, 23]. We compare our predictions with computer simulations carried out by other authors [23, 48, 50] or by ourselves. Our investigations [25] show that the MHNC energies are quite accurate for z

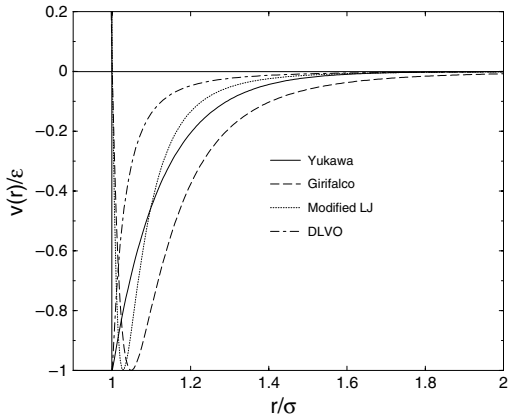

Figure 1. Model potentials presented in this work. Full curve: Yukawa potential with $z = 7$, see equation (1); dashed curve: Girifalco potential for C_{60}, equation (2); dotted curve: modified Lennard-Jones potential with $\alpha = 50$, equation (3); dot–dashed curve: DLVO model with $\sigma = 3.6$ nm, $Q = 10$ e, $A_H = 8.61 k_B T_{20}$ and $\delta = 0.18$ nm, see equations (4)–(6). In order to compare among different models, all potentials are drawn in their reduced units of length and energy, so that $v(r/\sigma = 1) = 0$ for the Girifalco and the modified Lennard-Jones potentials, respectively, and the minimum of the well depth is -1.

ranging from four to nine; the GMSA energies are reasonably good at $z = 4$ and show a $\sim$10% maximum discrepancy from Monte Carlo (MC) data at the highest z. The MHNC and GMSA equations of state and compressibility are also within a few per cent of the simulation results, at not too low temperatures. The SCOZA results for the equation of state are quantitatively accurate in any case. At lower z all theories become almost quantitative.

Results for the phase diagrams are reproduced in figure 2, in the temperature–density (T–ρ) representation, with the freezing lines estimated through the one-phase criterion of [32], amounting to the fulfilment of the condition that the so-called residual multi-particle entropy Δs vanishes at the phase transition. The limited applicability of such structural indicators in the context of 'energetic' fluids [18] has been carefully discussed in [33]. As is visible in figure 2, at $z = 7$ the GMSA qualitatively reproduces the relative location of the binodal and sublimation lines estimated in simulations of [23]. SCOZA results look fairly good; such an accuracy seems useful for the investigation of the $z = 9$ binodal, where no simulation results are available. Also in this case the SCOZA binodal [51] lies beneath the GMSA sublimation line; the latter one in turn reproduces the high-density portion of the sublimation line fairly well. It appears that the combination of GMSA and SCOZA may provide a description of the relative position of phase coexistence lines of the HCYF in the very short-range interaction regime.

In a subsequent study [29], we have exploited the sensitivity of integral equation approaches like the MHNC and the hybrid mean spherical approximation (HMSA, [52], see the appendix), to the procedure adopted in order to achieve the (partial) thermodynamic consistency of the theory. We have analysed in particular the application of MHNC and HMSA to the one-component HCYF and to a system interacting through a short-range DLVO potential [30] (see figure 1 and the next section for details). We have found that the enforcement of a global consistency constraint, i.e. a procedure in which one imposes the equality of the equation of state, as calculated along different routes from structure to thermodynamics, is in general more accurate than the requirement of a local consistency, based on the equality

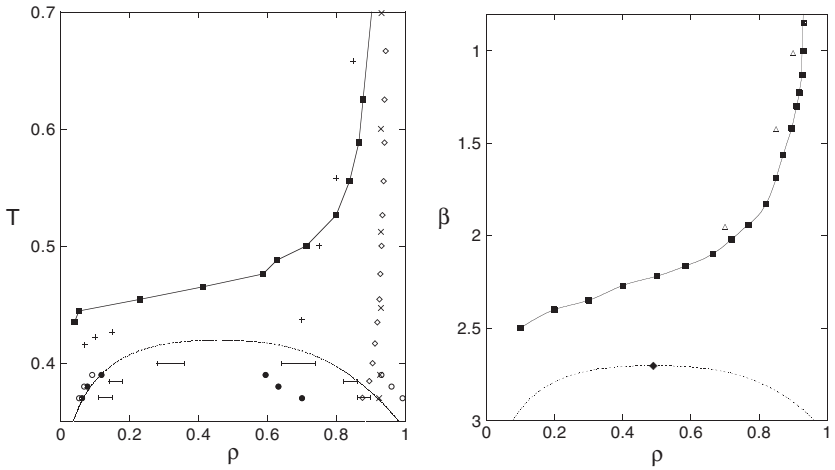

Figure 2. Phase diagram of the HCYF; temperatures are reported in units of ϵ/k_B, densities in units of σ^{-3}, $\beta = \epsilon/k_B T$. Left: $z = 7$. Liquid–gas coexistence: horizontal bars, simulation results [23]; dotted curve, SCOZA; full circles, GMSA; open circles, MHNC. $\Delta s = 0$ locus: diamonds, SCOZA; pluses, GMSA; crosses MHNC; line with squares, sublimation line of [23]. Right: $z = 9$. Triangles: GMSA predictions; dotted curve with diamond, SCOZA binodal and corresponding critical point [51]; curve with squares: simulation results [23]. See [25] for details.

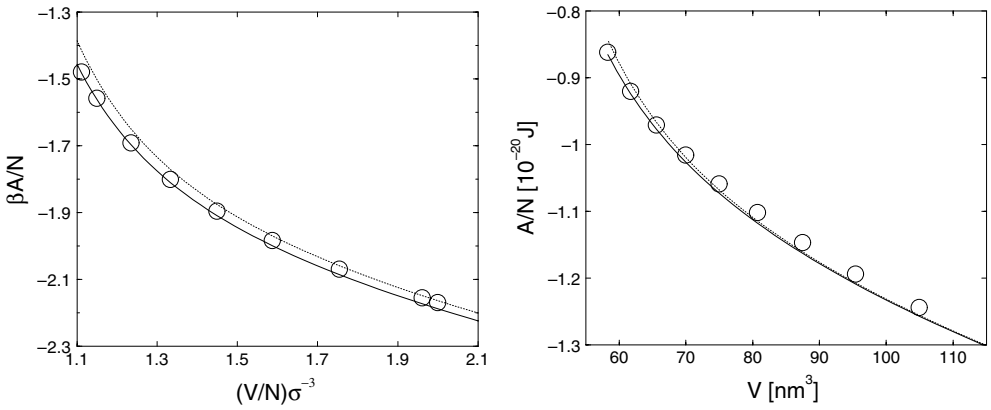

Figure 3. Helmholtz free energy as a function of the system volume along supercritical isotherms for the HCYF with $z = 7$ at $k_B T/\epsilon = 0.5$ (left), and a DLVO model with $\sigma = 3.6$ nm, $Q = 10$ e, $A_H = 8k_B T_{20}$ and $\delta = 0.144$ nm (see equations (4)–(6) for symbols) at $T = 442.5$ K (right). Open circles: MC simulations; dotted curve: MHNC; full curve: HMSA. See [29] for details.

of the isothermal compressibilities. The improvement associated to the global consistency becomes crucial in the short-range interaction regime; in particular, as is visible in figure 3, the Helmholtz free energy and the chemical potential are almost quantitatively predicted in the global consistency approach [29]. These results may help in the selection of integral equation approaches appropriate to determine in a confident manner the phase diagram of both pure fluids and binary mixtures of model short-range systems.

As far as a theoretical characterization of solid–fluid equilibrium is concerned, we have recently investigated [53] the solid phase of a pair, short-range model introduced by

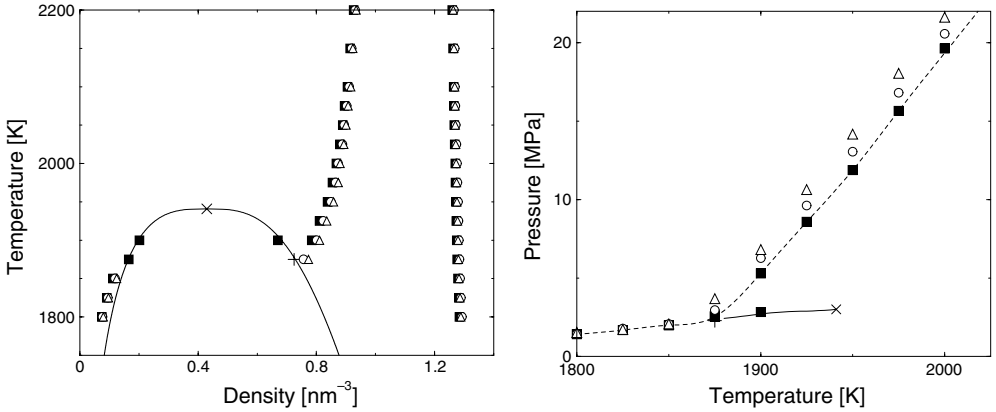

Figure 4. Fluid–solid coexistence densities (left) and pressures (right) of the Girifalco C_{60} model. PT predictions are used for the solid phase and MC data for the fluid phase [33]. Open symbols refer to different prescriptions for the effective hard-core diameter entering the theory—see equation (A.8) in the appendix. Squares are full Monte Carlo results [33]. The Gibbs ensemble Monte Carlo liquid–vapour coexistence line [57] is also shown. Dashed lines are guides to the eye. See [53] for details.

Girifalco [34] to describe particle interactions in C_{60}:

$$v(r) = -\alpha_1 \left[\frac{1}{s(s-1)^3} + \frac{1}{s(s+1)^3} - \frac{2}{s^4} \right] + \alpha_2 \left[\frac{1}{s(s-1)^9} + \frac{1}{s(s+1)^9} - \frac{2}{s^{10}} \right], \tag{2}$$

where $s = r/d$, $\alpha_1 = N^2 A/12d^6$, and $\alpha_2 = N^2 B/90d^{12}$; N and d are the number of carbon atoms and the diameter, respectively, of the fullerene particles, and $A = 32 \times 10^{-60}$ erg cm^6 and $B = 55.77 \times 10^{-105}$ erg cm^{12} are constants entering the Lennard-Jones 12-6 potential through which two carbon sites on different spherical molecules are assumed to interact. For C_{60}, $d = 0.71$ nm, while the node of the potential (2), the minimum, and its position, are $r_0 \simeq 0.959$ nm, $\varepsilon \simeq 0.444 \times 10^{-12}$ erg and $r_{\min} = 1.005$ nm, respectively (see figure 1). The phase diagram of the model has been recently characterized in terms of MC calculations of the free energies of the fluid and solid phases [33, 54]. We have also determined theoretically the free energy of the solid phase in terms of a standard perturbation theory (PT, [28], see the appendix). As is visible in figure 4, such an approach yields reliable predictions for the coexistence properties and, more generally, for the whole solid phase behaviour. Since the PT embodies a first-order expansion of the model's free energy around the free energy of a reference hard-sphere crystal, our results noticeably imply that the hard-sphere properties dominate, to a large extent, the structure of the solid phase; conversely, as discussed in [33], the structure of the fluid phase is markedly affected by energetic aspects, related to the attractive part of the potential. Our results further support the use of PT to characterize the solid phase of systems interacting through short-range forces, already documented in [21, 55, 56].

We have also investigated in a recent study [38] the phase transformations occurring in a system of spherical particles interacting through a short-range potential, obtained as a generalization of the Lennard-Jones interaction law [12]:

$$v(r) = \frac{4\epsilon}{\alpha^2} \left\{ \left[\left(\frac{r}{\sigma} \right)^2 - 1 \right]^{-6} - \alpha \left[\left(\frac{r}{\sigma} \right)^2 - 1 \right]^{-3} \right\}. \tag{3}$$

By changing the parameter α, this expression gives rise to a family of potentials whose shape ranges from a Lennard-Jones-like potential (for $\alpha \sim 1$) to potentials having a steep repulsive

part, followed by a deep and narrow attractive well, approaching the sticky sphere limit with increasing values of α. In our computations, we set $\alpha = 50$, thus recovering the model previously proposed in [12] to study the phase behaviour of globular protein solutions (see figure 1). We have focused on the crystallization processes of this system [38]. In particular, we have performed extensive molecular dynamics (MD) simulations of the metastability induced by progressively cooling the fluid below the fluid–solid and the (metastable) liquid–vapour coexistence at fixed density. We have analysed in detail how the crystallization process evolves in the simulated system, and the interplay of crystallization with the metastable liquid–vapour separation. In figure 5 we show a sequence of snapshots of MD simulations at $\rho\sigma^3 = 0.5$ (i.e. a density moderately higher than the critical density of the liquid–vapour binodal), for various decreasing temperatures in a path which crosses the binodal line at a reduced temperature $T \simeq 0.41$. Model (3) for globular proteins is admittedly a crude one. Nonetheless, qualitative similarities with the experimental behaviour still survive, as for instance the enhancement of the crystallization kinetics due to the presence of the metastable liquid–vapour separation; this phenomenon is very likely determined by the presence of the high-density fluid, which decreases both the interfacial free energy required to nucleate the crystal, and the amplitude of the density fluctuations, necessary to reach the solid density. Another interesting outcome is that the simulated crystallization process does not appear to be accompanied by any precursor effect: in fact, when any of the criteria we devised to this aim was met, the nucleation process was in any case already well underway. Our observations do not strictly prove, however, that precursors do not exist, but only that the chain of events leading to the transition involves the non-trivial coupling of several variables, as shown experimentally in [58].

3. Colloidal models for lysozyme solutions

In a recent series of studies [39–41] we have considered a prototype globular protein solution, namely lysozyme in water and sodium chloride added salt, extensively characterized from the experimental point of view (see e.g. [1–7] and references therein). In particular, we have parameterized the physical properties of the protein solution in the framework provided by the DLVO theory of charged colloidal suspensions [30]. We recall that the DLVO potential of mean force is given by the sum of a repulsive, Debye–Hückel contribution,

$$v_{\mathrm{DH}}(r) = \frac{1}{4\pi\epsilon_0\epsilon_{\mathrm{r}}}\left[\frac{z_{\mathrm{p}}e}{1+\chi_{\mathrm{DH}}\sigma/2}\right]^2 \frac{\exp[-\chi_{\mathrm{DH}}(r-\sigma)]}{r}, \tag{4}$$

and an attractive part, represented by a short-range van der Waals term:

$$v_{\mathrm{HA}}(r) = -\frac{A_{\mathrm{H}}}{12}\left[\frac{\sigma^2}{r^2} + \frac{\sigma^2}{r^2-\sigma^2} + 2\ln\frac{r^2-\sigma^2}{r^2}\right]. \tag{5}$$

The resulting total interaction is

$$v_{\mathrm{DLVO}}(r) = \begin{cases} \infty & r < \sigma + \delta \\ v_{\mathrm{HA}}(r) + v_{\mathrm{DH}}(r) & r \geqslant \sigma + \delta. \end{cases} \tag{6}$$

Here σ represents the effective diameter, $Q = z_{\mathrm{p}}e$ is the net charge (in electron units) of the macroparticle and A_{H} is the Hamaker constant; ϵ_{r} and ϵ_0 are, respectively, the (solution) relative and the vacuum dielectric constants; χ_{DH} is the inverse Debye screening length. The Stern layer thickness, δ, related to the intrinsic size of counterions which condense on the protein surface, is introduced in equation (6) to circumvent the singularity of the attractive term at $r = \sigma$ [1].

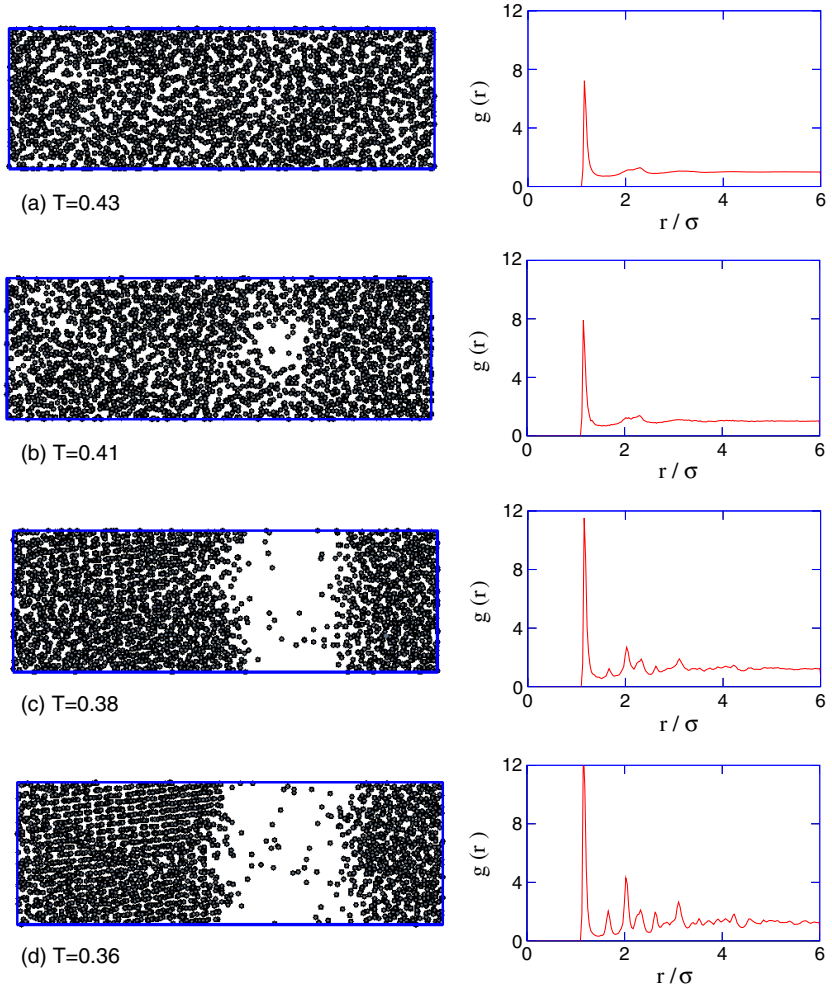

Figure 5. Generalized Lennard-Jones potential—see equation (3). Snapshots of MD configurations at progressively decreasing temperatures (in units of ϵ/k_B), together with the corresponding radial distribution functions; (a) metastable homogeneous fluid; (b) low-density bubble in a high-density fluid, just below the metastable liquid–vapour separation; (c) crystal nucleus in the high-density fluid; (d) extended defective crystal phase. See [38] for details.

(This figure is in colour only in the electronic version)

In our analysis, we have fixed all parameters entering equation (6), except the Hamaker constant, to values commonly accepted on the basis of several experimental evidences for lysozyme, i.e. $\sigma = 3.6$ nm, $Q = 10$ e, and $\delta = 0.18$ nm [1, 59, 60]. The Hamaker constant is used as a free parameter to fit the experimental interaction factor defined as

$$K_S = K_S^{HS} + 24 \int_{\delta/\sigma}^{\infty} x^2 \{1 - \exp[-\beta v_{DLVO}(x)]\}\, dx, \tag{7}$$

where $x = (r/\sigma - 1)$ and $K_S^{HS} = 8$ is the interaction factor corresponding to a fluid of hard spheres. The interaction factor is proportional to the second virial coefficient, $B_2 = K_S \nu/2M$, where $\nu = 0.703$ ml g^{-1} and $M = 14\,300$ Da are respectively the specific volume and the

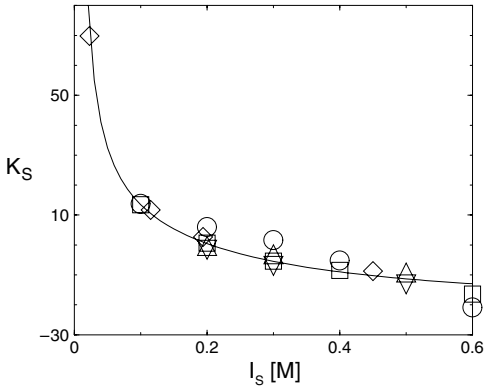

Figure 6. Interaction factor as a function of the solution ionic strength for lysozyme in water and NaCl added salt. Circles: SIC [7]; squares: SLS [2, 6]; diamonds: SLS [1]; triangles up: SLS [5] at $T = 303.15$ K; triangles down: SLS [5] at $T = 293.15$ K. Full curve: DLVO potential with $A_H = 8.61 k_B T_{20}$. See [41] for details.

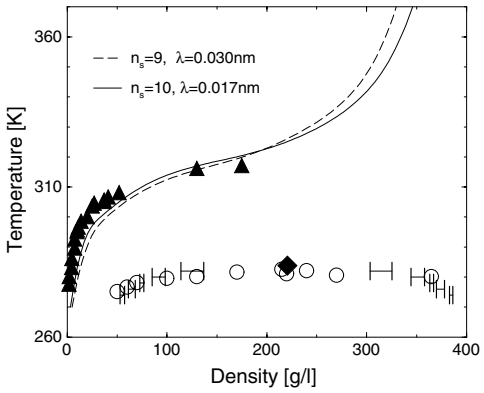

Figure 7. Phase diagram of a lysozyme solution in water and NaCl at $I_s = 0.51$ M. Open circles and full triangles: cloud and solubility points, respectively [4]. Error bars with diamond: DLVO demixing envelope and corresponding critical point. Full and dashed curves: DLVO crystallization curves with parameters for equation (8) in the legend.

atomic mass of lysozyme. As shown in figure 6, we best-fitted self-interaction chromatography (SIC) results, reported in [7], as well as static light scattering data from various sources (SLS [1, 2, 5, 6]), for a lysozyme solution in the range [0.1–0.6] M NaCl at a pH close to that of the experimental phase diagram [4]. Our fits are carried out in a temperature range relatively far from the demixing region; under these conditions the second virial coefficient of the lysozyme solution is expected to vary moderately, and almost linearly, with the increase of the temperature. An average Hamaker constant is then obtained, $A_H = 8.61$ (in $k_B T_{20}$ units) with a maximum half-dispersion equal to 0.26.

The extension of a 'colloid physics approach', as the DLVO theory, to globular protein solutions has long been debated. In fact, the literature reports either successful applications (see e.g. [17, 60–62]), or evidences of opposite sign, concerning the accuracy and the theoretical foundation of this picture [63–65]. Our observation is that, in agreement with previous studies, the validation of the present approach mainly derives from a positive, non-trivial, parameterization of phenomenological observations. In particular, we envisage the functional form of the attractive potential in equation (5) as being flexible enough to accommodate not only the van der Waals forces, but also other non-DLVO interactions, such as hydration effects [60, 61, 66]. We assume in fact that these forces act to artificially inflate the value of the Hamaker constant, which plays a different role from the original one, only related to dipolar attractions. Within this framework, specific effects, related for instance to the salt identity, are taken into account only indirectly, through the fit of experimental data. Although one could introduce additional interaction terms (which can be salt-specific), in order to increase the flexibility of the phenomenological potential, we shall show that a fine tuning of the van der Waals amplitude is sufficient to set the proper order of magnitude of the effective protein–protein interaction. In particular, this attractive term is the leading contribution to the effective potential because repulsive electrostatic interactions are almost screened out at biological salt molarities, thus constituting a minor correction to the potential of mean force.

On the basis of the above model, we have calculated the phase diagrams of the lysozyme solution at $I_s = 0.51$ M reported in figure 7. The protein-rich–protein-poor phase coexistence

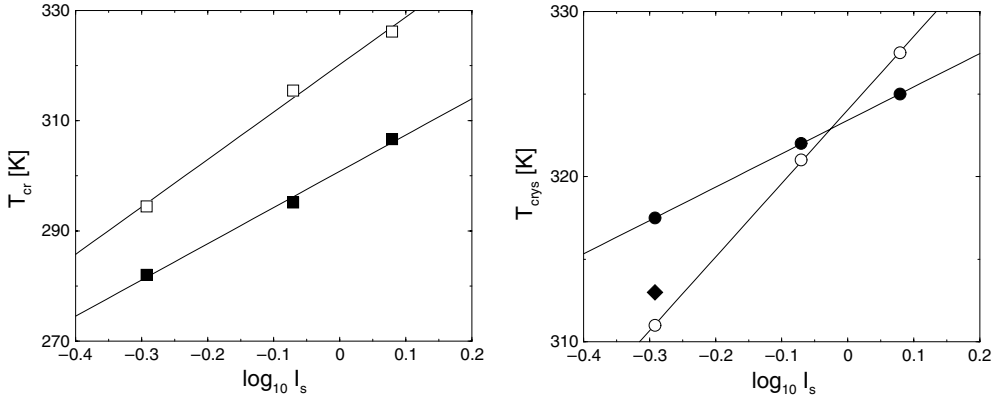

Figure 8. Critical (left) and solubility (right) points as a function of the ionic strength for lysozyme solutions at fixed protein concentration $\rho = 90$ g l^{-1}. Full symbols: experimental data from [4] (squares), [10] (circles, pH = 7.8), and [4] (diamond, pH = 4.5). Open symbols: DLVO results. Curves are linear fits of theoretical and experimental data.

line is determined by means of Gibbs ensemble Monte Carlo [67] simulations, carried out on a system composed of 1024 particles (with a 3σ interaction cutoff). As for the solubility line, the free energy of the fluid phase is calculated in the framework provided by the refined HMSA theory [52] whereas, for the solid phase, we have used for the chemical potential a simplified expression [42–45] derived in the framework of the cell theory [46],

$$\mu_{\mathrm{CRY}} = \mu_0 - n_{\mathrm{s}}\epsilon_{\mathrm{EFF}}/2 - 3k_{\mathrm{B}}T\ln(\lambda^*). \tag{8}$$

This equation provides a direct link with the essential properties of the protein crystals, namely the average number of contacts, n_{s}, and the translational freedom along one axis, λ [with $\lambda^* = \lambda/(\sigma+\delta)$], of the protein inside the unit cell, so that λ^3 constitutes a rough estimate of the volume accessible to the protein inside the cage formed by its first neighbours; μ_0 is the standard part of the chemical potential, and ϵ_{EFF} is the minimum of the effective potential (6). A constant value $\log[V_{\mathrm{p}}/(\sigma+\delta)^3]$ [45] has been included in μ_0, where V_{p} is the protein volume. The coexisting fluid branch is then found by equating the chemical potentials of both phases [43]. We fix the number of contacts between nine and ten, in agreement with the estimates of [68]; a comparable number of contacts has been specifically reported for lysozyme crystals in [69]. Our results for the phase diagram are compared in figure 7 with the experimental demixing and solubility lines reported in [4]. As is visible, the theoretical demixing envelope coincides with the experimental *locus* of cloud points with a critical temperature $T_{\mathrm{cr}} \simeq 284$ K, very close to the experimental outcome $T_{\mathrm{cr}} = 282.5$ K. The theoretical crystallization boundary, calculated with $n_{\mathrm{s}} = 9$ and $\lambda = 0.03$ nm in equation (8), is also in good agreement with the experimental curve. We obtain similar results if another plausible number of contacts, namely $n_{\mathrm{s}} = 10$ (with $\lambda = 0.017$ nm), is assumed (see figure 7). All values of λ are compatible with the temperature factors of the Bragg intensities in lysozyme crystals, where the typical range for λ is [0.014–0.033] nm [70].

As is visible from figure 8, satisfactory results are also obtained for the variation of cloud and solubility temperatures as a function of the ionic strength. These results are reported in a very recent work [40], where we have fitted the Hamaker constant against the collective diffusion coefficient measurements carried out by Beretta and co-workers [60]. It appears that the theory substantially captures the experimental dependence [4, 10] of the critical and solubility temperatures on the logarithm of the ionic strength.

In summary, although a protein model where we only consider central forces might appear oversimplified, a reasonable prediction of the phase diagram of the lysozyme solution may be obtained through a 'colloidal' model, provided a judicious fit of the experimental second virial coefficient of the solution [1, 2, 7] is performed.

On the other hand, it is evident that a detailed analysis of the complex behaviour of protein solutions would require a refined, highly specific treatment of the microscopic interactions; several improvements to the DLVO-like model have been recently proposed such as, for instance, different shape [71] or multi-site representations of protein interactions (see e.g. [43, 72–74]). In comparison with such studies, our results support the observation that in the fluid phase, close to the critical point, a distributed approach, as envisaged for instance in [72], may be well approximated by its spherically averaged counterpart.

4. Conclusions

We have reported results of extensive investigations of short-range potential models suited to describe, in a realistic albeit qualitative manner, the physical properties of prototype globular protein solutions, like for instance lysozyme in water and added salt. We have focused on the phase diagram of these systems, and first reviewed predictions from simulations and integral equation theories of the fluid state for the hard-core Yukawa fluid. We have carried out a systematic assessment of various refined theories—like the MHNC, the SCOZA and the GMSA—against simulation data, and investigated the improvement of the strategy adopted to enforce their thermodynamic consistency. It emerges that integral equations can reach a good degree of accuracy in the prediction of phase coexistence properties. In particular, the basic property of the phase diagram in many protein solutions, namely a liquid–vapour coexistence line metastable with respect to the fluid–solid equilibrium, is correctly reproduced. Hopefully, these theoretical approaches may be used as a complementary tool, especially when simulations are applied with difficulty, as in the treatment of multicomponent systems characterized by the high dilution of one of the components, or by strong charge and/or size asymmetries.

We have also reviewed a recent molecular dynamics investigation of crystallization processes in a generalized Lennard-Jones model of protein solutions. This study shows that some features of the true crystal formation kinetics are qualitatively represented, even if a realistic description of the natural process (especially as far as the crystallization timescale is concerned) is still outside the possibilities offered by such an approach.

We have then reported on an investigation of protein solutions based on the DLVO potential, widely adopted in the representation of colloidal systems. We have shown that the phase coexistence lines of lysozyme solutions in water and sodium chloride added salt can be reproduced through a combined use of theories and simulations, when the DLVO potential is parameterized on the experimental second virial coefficient of the protein solution at moderate ionic strengths. Preliminary results for γ-crystallin in water and added sodium phosphate also turn out to be quantitatively accurate. Our procedure is therefore potentially suited for a variety of protein systems, and presently its extensive test appears to us as essentially hampered by the scarcity of phase diagram and B_2 measurements in the same solution conditions.

In conclusion, basic assumptions on the effective interaction among globular proteins in solution are sufficient to accurately reproduce the phase boundaries of such systems. In particular, a parameterization of a central, pair potential, based on the experimental second virial coefficient allows for a confident prediction of the true cloud-points and solubility lines. Since much less effort is required for the experimental determination of B_2, the advantage of the procedure here exploited becomes evident. Our predictions are accurate also for high protein concentrations, although experimental measurements of B_2 are carried out in the dilute

regime. These findings agree with the linearity observed in several Debye plots of scattering data [1], indicating that B_2 values measured in undersaturated solutions slightly change in the supersaturated regime.

Appendix. Theoretical methods

The properties of model fluids discussed in this work have been determined theoretically in the context of integral equation theories of liquids [28] like the modified hypernetted chain approximation (MHNC, [47]), the generalized mean spherical approximation (GMSA, [49]), the self-consistent Ornstein–Zernike approximation (SCOZA, [48]), and the hybrid mean spherical approximation (HMSA, [52]). In the MHNC scheme one considers the Ornstein–Zernike equation

$$h(r) = c(r) + \rho \int c(|\mathbf{r} - \mathbf{r}'|)h(r')\,\mathrm{d}\mathbf{r}', \tag{A.1}$$

(where ρ is the density of the fluid, $g(r)$ is the radial distribution function, rdf, and $h(r) = g(r) - 1$ and $c(r)$ are the pair and direct correlation function, respectively) coupled to the formally exact expansion result for the rdf

$$g(r) = \exp[-\beta v(r) + h(r) - c(r) + B(r)] \tag{A.2}$$

where $v(r)$ and $B(r)$ are the pair potential and the bridge function, respectively, of the system under study, while $\beta = 1/k_\mathrm{B}T$ with k_B the Boltzmann constant and T the temperature. A closure to equation (A.1) is then obtained through equation (A.2) by approximating the bridge function for the potential $v(r)$ in terms of that of a hard sphere fluid characterized by a hard core diameter σ^* chosen so as to enforce the (partial) thermodynamic consistency of the theory.

In the GMSA, a solution to the Ornstein–Zernike equation (2) is obtained by hinging on the exact core condition $g(r) = 0$ for $r < \sigma$, descending from the fact that $v(r) = \infty$ on such a distance range, coupled to the approximation for the direct correlation function outside the hard-core diameter:

$$c(r) = -\beta v(r) + K \exp[-z(r - \sigma)]/r \qquad r \geqslant \sigma \tag{A.3}$$

where K and z are used as adjustable parameters in order to impose the internal thermodynamic consistency of the theory.

In the SCOZA scheme $c(r)$ is written as

$$c(r) = -A\beta v(r) + K_\mathrm{HS} \exp[-z_\mathrm{HS}(r - \sigma)]/r \qquad r \geqslant \sigma; \tag{A.4}$$

K_HS and z_HS are predetermined by setting $v(r) = 0$ in equation (A.4) and requiring that both the compressibility and the virial routes to thermodynamics yield the Carnahan–Starling equation of state for a hard-sphere fluid. The factor A is then obtained by imposing that the compressibility and energy routes yield the same thermodynamics.

The HMSA is so called because it interpolates between the hypernetted chain (HNC) [24] approximation and the soft mean spherical approximation (SMSA) [24]: it is written as

$$g(r) = \exp[-\beta v_1(r)] \left\{ 1 + \frac{\exp\{f(r)[h(r) - c(r) - \beta v_2(r)]\} - 1}{f(r)} \right\}. \tag{A.5}$$

In equation (A.5) the potential $v(r)$ is split into a soft core repulsive part, $[v_1(r)]$, and an attractive part, $[v_2(r)]$, according to

$$v_1(r) = \begin{cases} v(r) - v(r_\mathrm{min}) & r \leqslant r_\mathrm{min} \\ 0 & r > r_\mathrm{min} \end{cases} \qquad v_2(r) = \begin{cases} v(r_\mathrm{min}) & r \leqslant r_\mathrm{min} \\ v(r) & r > r_\mathrm{min} \end{cases} \tag{A.6}$$

where $r_{\min}$ denotes the position of the minimum of potential $v(r)$ and $f(r) = 1 - \exp[\xi r]$. The parameter ξ is used to enforce thermodynamic consistency.

In order to determine theoretically the properties of the solid phase of the Girifalco C_{60} model we have employed a standard perturbation theory (PT, [28]). In this scheme, the model's free energy per particle, f, is calculated as a first-order expansion around the free energy of a reference hard-sphere solid, $f_{\rm HS}$, according to

$$\beta f = \beta f_{\rm HS} + \frac{\beta\rho}{2} \int v_{\rm pert}(r) \overline{g}_{\rm HS}(r)\, {\rm d}\mathbf{r}, \tag{A.7}$$

where $\overline{g}_{\rm HS}(r)$ is the radial distribution function of the reference hard-sphere solid. In order to determine the perturbative part of the potential, $v_{\rm pert}(r)$, and the diameter σ^* of the reference hard-sphere system entering equation (A.7), we have followed the WCA prescription [19], that amounts to splitting the original potential according to equation (A.6) with $v_{\rm pert}(r) \equiv v_2(r)$ and $v_{\rm ref} \equiv v_1(r)$; then [75]

$$\sigma^* = \int_0^\infty \{1 - \exp[-\beta v_{\rm ref}(r)]\}\, {\rm d}r. \tag{A.8}$$

References

[1] Muschol M and Rosenberger F 1995 *J. Chem. Phys.* **103** 10424
[2] Rosenbaum D F, Zamora P C and Zukoski C F 1996 *Phys. Rev. Lett.* **76** 150
[3] Malfois M, Bonneté F, Belloni L and Tardieu A 1996 *J. Chem. Phys.* **105** 3290
[4] Muschol M and Rosenberger F 1997 *J. Chem. Phys.* **107** 1953
[5] Bonneté F, Finet S and Tardieu A 1999 *J. Cryst. Growth* **196** 403
[6] Rosenbaum D F, Kulkarni A, Ramakrishnan S and Zukoski C F 1999 *J. Chem. Phys.* **111** 9882
[7] Tessier P M, Lenhoff A M and Sandler S I 2002 *Biophys. J.* **82** 1620
Tessier P M *et al* 2002 *Acta Crystallogr.* D **82** 1531
[8] Chayen N E 2002 *Trends Biotechnol.* **20** 98
[9] McPherson A 1982 *Preparation and Analysis of Protein Crystals* (Malabar, FL: Krieger)
[10] Broide M, Tominc T M and Saxowsky M D 1996 *Phys. Rev.* E **53** 6325
Grigsby J J, Blanch H W and Prausnitz J M 2001 *Biophys. Chem.* **91** 231
[11] George A and Wilson W 1994 *Acta Crystallogr.* D **50** 361
[12] ten Wolde P R and Frenkel D 1997 *Science* **277** 1975
[13] Haas C and Drenth J 2000 *J. Phys. Chem.* **104** 368
[14] Eaton W A and Hofrichter J 1990 *Advances in Protein Chemistry* vol 40 (San Diego, CA: Academic) p 63
[15] Lomakin A, Asherie N and Benedek G B 1996 *J. Chem. Phys.* **104** 1646
[16] Piazza R, Peyre V and Degiorgio V 1998 *Phys. Rev.* E **58** R2733
[17] Poon W C K 1997 *Phys. Rev.* E **55** 3762
Poon W C K *et al* 2000 *J. Phys.: Condens. Matter* **12** L569
[18] Louis A A 2001 *Phil. Trans. R. Soc.* A **359** 939
[19] Chandler D, Weeks J D and Andersen H C 1983 *Science* **220** 787
[20] Dawson K 2002 *Curr. Opin. Colloid Interface* **7** 218
[21] Foffi G *et al* 2002 *Phys. Rev.* E **65** 031407
[22] Chen S H, Chen W R and Mallamace F 2003 *Science* **300** 619
[23] Hagen M H J and Frenkel D 1994 *J. Chem. Phys.* **101** 4093
[24] Caccamo C 1996 *Phys. Rep.* **274** 1
[25] Caccamo C, Pellicane G, Costa D, Pini D and Stell G 1999 *Phys. Rev.* E **60** 5533
[26] Caccamo C, Costa D and Pellicane G 1999 *J. Chem. Phys.* **106** 4498
Caccamo C, Costa D and Pellicane G 1999 *Proc. NATO-ASI New Approaches to Problems in Liquid State Theories* ed C Caccamo, J P Hansen and G Stell (Dordrecht: Kluwer)
[27] Caccamo C, Pellicane G and Costa D 2000 *J. Phys.: Condens. Matter* **12** A437
[28] Hansen J-P and McDonald I R 1986 *Theory of Simple Liquids* 2nd edn (London: Academic)
[29] Caccamo C and Pellicane G 2002 *J. Chem. Phys.* **117** 5072
[30] Derjaguin B V and Landau L V 1941 *Acta Phys. Chim. USSR* **14** 633
Verwey E J W and Overbeek J T G 1948 *Theory of Stability of Lyophobic Colloids* (Amsterdam: Elsevier)

[31] Hansen J P and Verlet L 1969 *Phys. Rev.* **184** 151
[32] Giaquinta P V and Giunta G 1992 *Physica* A **187** 145
Giaquinta P V, Giunta G and Prestipino Giarritta S 1992 *Phys. Rev.* A **45** R6966
[33] Costa D, Pellicane G, Caccamo C and Abramo M C 2003 *J. Chem. Phys.* **118** 304
[34] Girifalco L F 1992 *J. Phys. Chem.* **96** 858
[35] Chen B, Siepmann I, Karaborni S and Klein M 2003 *J. Phys. Chem.* B **107** 12320
[36] Fartaria R P S, Silva Fernandes F M S and Freitas F F M 2002 *J. Phys. Chem.* B **106** 10227
[37] Ben-Amotz D and Stell G 2003 *J. Chem. Phys.* **119** 10777
[38] Costa D, Ballone P and Caccamo C 2002 *J. Chem. Phys.* **116** 3327
[39] Pellicane G, Costa D and Caccamo C 2003 *J. Phys.: Condens. Matter* **15** 375
[40] Pellicane G, Costa D and Caccamo C 2003 *J. Phys.: Condens. Matter* **15** S3485
[41] Pellicane G, Costa D and Caccamo C 2004 *J. Phys. Chem.* B **108** 7538 (Letter)
[42] Asherie N, Lomakin A and Benedek G B 1996 *Phys. Rev. Lett.* **77** 4832
[43] Sear R P 1999 *J. Chem. Phys.* **111** 4800
[44] Curtis R A, Blanch H W and Prausnitz J M 2001 *J. Phys. Chem.* B **105** 2445
[45] Warren P B 2002 *J. Phys.: Condens. Matter* **14** 7617
[46] Lennard-Jones J E and Devonshire A F 1937 *Proc. R. Soc.* A **163** 53
[47] Rosenfeld Y and Ashcroft N W 1979 *Phys. Rev.* A **20** 1208
[48] Pini D, Stell G and Wilding N B 1998 *Mol. Phys.* **95** 483
[49] Waisman E 1973 *Mol. Phys.* **25** 45
Høye J S, Lebowitz J L and Stell G 1974 *J. Chem. Phys.* **61** 3253
Høye J S and Blum L 1978 *J. Stat. Phys.* **19** 317
Blum L 1980 *J. Stat. Phys.* **22** 661
[50] Lomba E and Almarza N E 1994 *J. Chem. Phys.* **100** 8367
Meijer E J and Elazhar F 1997 *J. Chem. Phys.* **106** 4678
[51] Pini D 2000 private communication
[52] Zerah G and Hansen J-P 1986 *J. Chem. Phys.* **84** 2336
[53] Costa D, Pellicane G, Caccamo C, Schöll-Paschinger E and Kahl G 2003 *Phys. Rev.* E **68** 021104
[54] Hasegawa M and Ohno K 1999 *J. Chem. Phys.* **111** 5955
Hasegawa M and Ohno K 2000 *J. Chem. Phys.* **113** 4313
[55] Germain P and Amokrane S 2002 *Phys. Rev.* E **65** 031109
[56] Velasco E, Navascués G and Mederos L 1999 *Phys. Rev.* E **60** 3158
[57] Caccamo C, Costa D and Fucile A 1997 *J. Chem. Phys.* **106** 255
[58] Schätzel K and Ackerson B J 1992 *Phys. Rev. Lett.* **68** 337
[59] Kuehner D E *et al* 1997 *Biophys. J.* **73** 3211
Kulkarni A M, Chatterjee A P, Schweizer K S and Zukoski C F 1999 *Phys. Rev. Lett.* **83** 4554
[60] Beretta S, Chirico G and Baldini G 2000 *Macromolecules* **33** 8663
[61] Farnum M and Zukoski C 1999 *Biophys. J.* **76** 2716
[62] Rowe A J 2001 *Biophys. Chem.* **93** 93
[63] Piazza R 1999 *J. Cryst. Growth* **196** 415
[64] Petsev D N and Vekilov P G 2000 *Phys. Rev. Lett.* **84** 1339
[65] Böstrom M, Williams D R M and Ninham B W 2001 *Phys. Rev. Lett.* **87** 168103
[66] Chernov A A 1997 *Phys. Rep.* **288** 61
[67] Panagiotopoulos A Z 1987 *Mol. Phys.* **61** 813
[68] Janin J and Rodier F 1995 *Proteins: Struct. Funct. Genet.* **23** 580
[69] Pedersen T G *et al* 1991 *J. Mol. Biol.* **218** 413
[70] Doucet J and Benoit J 1987 *Nature* **325** 643
[71] Neal B L *et al* 2001 *J. Cryst. Growth* **196** 377
Sun N and Waltz J Y 2001 *J. Colloid Interface Sci.* **234** 90
[72] Lomakin A, Asherie N and Benedek G B 1999 *Proc. Natl Acad. Sci.* **96** 9465
[73] Allahyarov E, Löwen H, Louis A A and Hansen J P 2002 *Europhys. Lett.* **57** 731
Allahyarov E, Löwen H, Louis A A and Hansen J P 2003 *Phys. Rev.* E **67** 041801
[74] Chang J, Lenhoff A M and Sandler S I 2004 *J. Chem. Phys.* **120** 3003
[75] Barker J A and Henderson D 1967 *J. Chem. Phys.* **47** 4714