Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2014, **10**, 894

Protein cluster formation in aqueous solution in the presence of multivalent metal ions – a light scattering study

Daniel Soraruf,[a] Felix Roosen-Runge,[a] Marco Grimaldo,[ab] Fabio Zanini,[c] Ralf Schweins,[b] Tilo Seydel,*[b] Fajun Zhang,*[a] Roland Roth,[d] Martin Oettel[a] and Frank Schreiber[a]

Received 17th September 2013
Accepted 20th November 2013

DOI: 10.1039/c3sm52447g

www.rsc.org/softmatter

The formation of protein clusters as precursors for crystallization and phase separation is of fundamental and practical interest in protein science. Using multivalent ions, the strengths of both long-range Coulomb repulsion and short-range attraction can be tuned in protein solutions, representing a well-controlled model system to study static and dynamic properties of clustering during the transition from a charge-stabilized to an aggregate regime. Here, we study compressibility, diffusion, and size of solutes by means of static (SLS) and dynamic light scattering (DLS) in solutions of bovine serum albumin (BSA) and YCl_3. For this and comparable systems, an increasing screening and ultimately inversion of the protein surface charge induce a rich phase behavior including reentrant condensation, liquid–liquid phase separation and crystallization, which puts the cluster formation in the context of precursor formation and nucleation of liquid and crystalline phases. We find that, approaching the turbid aggregate regime with increasing salt concentration c_s, the diffusion coefficients decrease and the scattered intensity increases by orders of magnitude, evidencing increasing correlation lengths likely associated with clustering. The combination of static and dynamic observations suggests the formation of BSA clusters with a size on the order of 100 nm. The global thermodynamic state seems to be stable over at least several hours. Surprisingly, results on collective diffusion and inverse compressibility from different protein concentrations can be rescaled into master curves as a function of c_s/c^*, where c^* is the critical salt concentration of the transition to the turbid aggregate regime.

1 Introduction

The formation of protein clusters is of fundamental and practical interest in protein science. Depending on the environmental control parameters, protein solutions exhibit a rich phase diagram, leading to the formation of gels, liquid–liquid phase separation (LLPS), and amorphous aggregation.[1,2] Clustering proves to be promising for drug delivery *via* concentrated solutions of antibodies.[3] For a better understanding and an improved efficiency of protein crystallization, protein clusters might play an essential role: crystallization can occur *via* a two-step nucleation, a process in which dense liquid droplets or small clusters serve as nucleation precursors.[4–6] The study of LLPS and clusters in solution provides therefore insight into the physical processes underlying protein crystallization,[7] which is known to be driven by a subtle balance between repulsive and attractive interactions[8] and to be accompanied by a rise in density and a steep reduction in free energy.[9] The same signatures are expected for cluster formation, but no thorough and complete theoretical understanding has been achieved so far. Theoretical approaches focus on the formation and stabilization mechanism of clusters in solution, generally as a balance between a short-range attractive aggregation potential and the long-range Coulomb repulsion.[10–13] Computer simulations support this notion of cluster formation due to competing interactions.[14–19] The importance of the range of the attractive interaction for clustering has been demonstrated in uncharged colloid-polymer systems.[20] Further studies have addressed clustering in charged colloid-polymer systems with both adsorbing and non-adsorbing polymers.[21,22] Also, a recent study comparing clustering in β-lactoglobulin and BSA has shown the significant role of charge asymmetry in these phenomena.[23]

Experimental studies on several protein systems have revealed different features of protein clustering in solution. For example, experiments on lysozyme indicate the formation of equilibrium clusters in concentrated protein solutions,

[a]*Institut für Angewandte Physik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany. E-mail: fajun.zhang@uni-tuebingen.de*

[b]*Institut Laue-Langevin, B.P.156, 38042 Grenoble, France. E-mail: seydel@ill.eu*

[c]*Max-Planck Institute for Developmental Biology, Spemannstraße 20, 72076 Tübingen, Germany*

[d]*Institut für Theoretische Physik, Universität Tübingen, Auf der Morgenstelle 14, 72076 Tübingen, Germany*

including notably the results obtained by Stradner *et al.*[24] Further studies by combining small-angle scattering techniques (SAXS and SANS), nuclear magnetic resonance (NMR), neutron spin echo (NSE), and dynamic light scattering (DLS) suggest that the protein clusters are rather dynamic instead of static.[25–31] In β-lactoglobulin solutions, Piazza and Iacopini observed the spontaneous formation of transient clusters. This transient state has a very short lifetime and cannot be regarded as chemically bound.[32] Large metastable clusters of proteins have been studied experimentally by SLS/DLS and Brownian microscopy for proteins such as hemoglobin, lumazine synthase, and lysozyme.[31,33–35]

Our group has studied the phase behavior of model globular proteins in solution in the presence of multivalent counterions. For negatively charged globular proteins and with increasing salt concentration c_s, binding of multivalent counterions inverts the protein surface charge as confirmed by zeta potential measurements and supported by Monte Carlo simulations and analytical calculations.[36–38] Macroscopically, the charge inversion is reflected in a reentrant condensation (RC) phase behavior[36,37] as depicted in Fig. 1, *i.e.* a turbid region is observed in between two critical salt concentrations, $c^* < c_s < c^{**}$, also including a metastable liquid–liquid phase separation (LLPS) at high protein concentrations.[39]

Crystallization from the condensed regime follows different mechanisms. Near c^*, crystals grow following a classic nucleation and growth mechanism; near c^{**}, the crystallization follows a two-step mechanism, *i.e.*, crystal growth follows a metastable LLPS.[7,39,40] SAXS measurements demonstrate that protein clusters act as precursors for crystal growth, reducing the energy barrier for nucleation.[7] X-ray diffraction analyses on the high quality single crystals provide direct evidence of the crystal structure and cation binding sites.[40] In comparison with our studies on both static[41,42] and dynamic[43–45] behavior of protein solutions with mono- and divalent ions, the effects from multivalent ions cannot be explained only by charge screening, salting-in or salting-out, but rather by a bridging effect which may lead to the formation of protein clusters.[46]

In this paper, we report on a systematic experimental study of the protein cluster formation approaching the boundary c^* towards the turbid regime as a function of protein and salt concentrations. The goal here is to explore the properties of protein clusters, which are presumably induced by ion-bridging in solution, such as cluster size and stability. We therefore systematically characterize both static and dynamic properties of protein solutions using static and dynamic light scattering. In particular, we focus on the forward scattering intensity, radius of gyration, and collective diffusion coefficient. We discuss the results in terms of mechanisms of cluster formation.

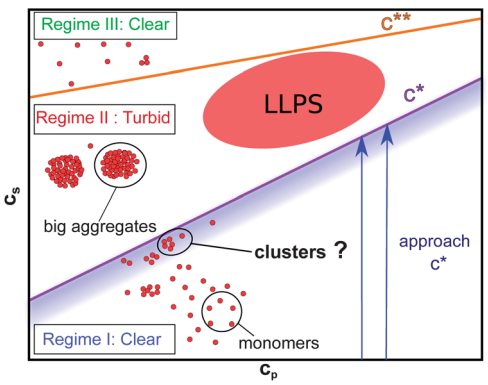

Fig. 1 Sketch of the phase diagram of aqueous solutions of BSA with YCl_3.[36,37] c_p and c_s denote logarithmic scales for the concentrations of the solutes BSA and YCl_3, respectively, and c^* and c^{**} the boundaries of the turbid region. In part of the region $c^* < c_s < c^{**}$ a liquid–liquid phase separation (LLPS), and for $c > c^{**}$ a reentrant dissolution are observed. The vertical arrows indicate schematically the type of approach to c^* adopted in this work, *i.e.* the variation of $c_s < c^*$ at different c_p approaching c^*.

2 Materials and methods

2.1 Sample preparation

BSA and YCl_3 were purchased from Sigma Aldrich with a purity of 99% (A3059) and 99.99% (451363), respectively. The samples were prepared under a flow box filtering the stock solutions of BSA, YCl_3 and water using 100 nm Wattman filters before mixing them in the borosilicate cuvette for the light scattering investigation. The concentration of the BSA stock solution was determined by UV-Vis with an extinction coefficient of 0.667 ml mg^{-1} (ref. 47).

2.2 Static and dynamic light scattering

Light scattering using coherent visible light is a well established technique to gain information about molecular weight, interaction strength and diffusion behavior of particles in solution.[48] The magnitude of the scattering vector is

$$Q = \frac{4\pi n}{\lambda}\sin\left(\frac{\theta}{2}\right) \quad (1)$$

where θ is the scattering angle, λ the wavelength of the light and n the refractive index of the solvent.

Static light scattering (SLS) analyzes the absolute scattering intensity, and the Rayleigh ratio R is introduced as a renormalized intensity independent of the instrumental setup. In general, the behavior at small Q^2 allows the correlation length ξ in the system to be extracted *via* the Zimm equation:

$$\frac{Kc}{R(Q)} = \frac{1}{M}(1 + 2A_2c)(1 + Q^2\xi^2) \quad (2)$$

where $K = 4\pi^2n^2(\partial n/\partial c)^2/\lambda^4N_A$ is the optical constant, c the concentration of the solute, M its molecular weight and A_2 the second virial coefficient. For a dilute solution of monodisperse particles (monomers or clusters) with R_g being their radius of gyration, $\xi^2 = {R_g}^2/3$ holds for small Q^2. The assumption that ξ gives roughly the size of the clusters holds only for a statistical ensemble average. From the SLS data, no information about a possible lifetime of the clusters can be extracted. The left-hand side of eqn (2) is related to the isothermal compressibility χ_T *via*[48]

$$\frac{Kc}{R(Q\to 0)} = \frac{\chi_{ideal}}{M\chi_T}, \quad (3)$$

where χ_{ideal} is the compressibility of an ideal gas.

Dynamic light scattering (DLS) investigates the motion of the suspended particles in solution. The time-dependent intensity auto-correlation function $g^{(2)}(\tau)$ related to the speckle fluctuations induced by the coherently illuminated moving particles is recorded in DLS measurements *via*

$$g^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle/\langle I(t)\rangle^2, \quad (4)$$

where $I(t)$ is the time-dependent scattering intensity. The particle diffusion coefficient D can be related to $g^{(2)}(\tau)$ *via* the Siegert relation

$$g^{(2)}(\tau) = 1 + \beta|g^{(1)}(\tau)|^2 = 1 + \beta\exp(-DQ^2\tau)^2 \quad (5)$$

where β is an instrumental factor, $g^{(1)}(\tau)$ the first-order correlation function, and D the diffusion coefficient of the solutes. Due to the small accessible scattering vectors and the small size of protein molecules, DLS generally probes relaxation of concentration gradients on the length scale $2\pi/Q$, which are highly affected by interactions between the proteins. Due to the long observation times, the apparent collective diffusion of protein particles is in general observed in the long-time limit. The investigation of multicomponent solutions results in multiple decay times. For the analyzed samples we observe two well-separated decay times, and use a double exponential fit with two distinct relaxation rates Γ_1 and Γ_2 to model the correlation decay:

$$g^{(2)}(\tau) - 1 = \beta[A_1\exp(-\Gamma_1\tau) + A_2\exp(-\Gamma_2\tau)]^2. \quad (6)$$

We denote the fast and slow contributions to $g^{(2)}(\tau)$ by the indices 1 and 2, respectively, tentatively attributing them to monomers and aggregates, respectively. We carried out all fits using MATLAB. In addition, we compared several MATLAB fit results with those obtained by CONTIN[49] performing a numerical regularized inverse Laplace transform, obtaining comparable results for the decay rates.

The SLS/DLS experiments were performed using an ALV-CGS3 setup with a wavelength $\lambda = 632.8$ nm. The CONTIN analysis was performed using the light scattering software provided by ALV.

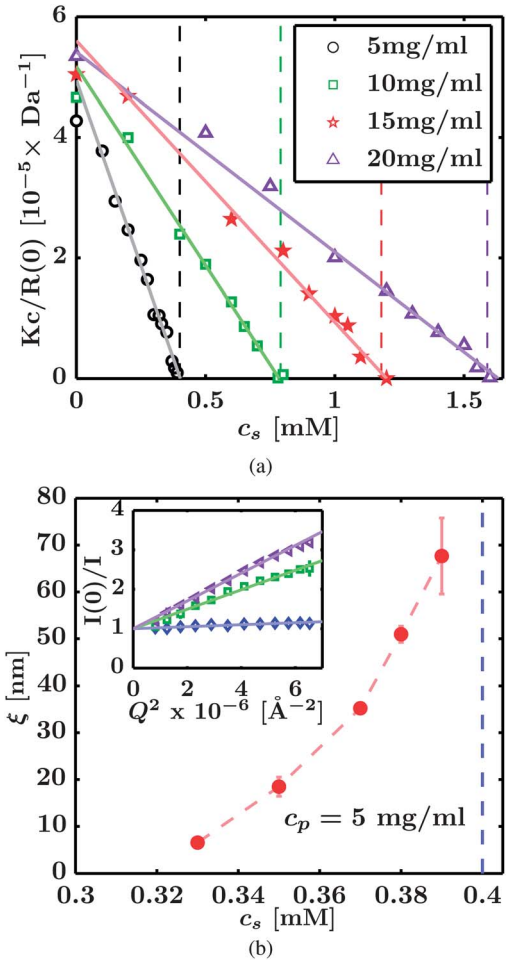

Fig. 2 (a) Forward inverse scattering intensity $1/I(0) = Kc/R(Q = 0)$ from SLS (symbols) and (b) correlation length ξ for a BSA concentration of 5 mg ml^{-1} (circle symbols) over c_s. The inset of (b) depicts the dependence of $I(0)/I(Q)$ on Q^2 for BSA at 5 mg ml^{-1} for salt concentrations of 0.33 (diamond symbols), 0.38 (square symbols), and 0.39 mM (triangle symbols), respectively. Solid lines indicate linear fits in all parts of the figures. The vertical dashed lines indicate c^* for the different c_p. When not visible, the error bars are smaller than the symbols.

3 Results

3.1 Static behavior probed by SLS

We first present the static properties of the solutions, *i.e.* the inverse forward intensity $Kc/R(0)$ and correlation length ξ as functions of protein and salt concentrations. Exemplary results for $Kc/R(0)$ as a function of c_s for different protein concentrations are depicted in Fig. 2(a). The observation is clearly consistent with a linear decrease in c_s. From eqn (3) it follows that the inverse compressibility decreases linearly with increasing salt concentration. The intersection point $Kc/R(0) = 0$ can be used to define the critical salt concentration c^*. The results are in good agreement within the errors with other methods reported before,[36] such as visual inspection and laser transmission. To compare the determined values of c^* from visual inspection of the sample turbidity and from the fits in Fig. 2(a), respectively, the results for c^* are summarized in Table 1.

We observe that a notable Q-dependence of $Kc/R = 1/I$ is visible close to c^*. In good approximation, $1/I$ is proportional to Q^2 (see the inset of Fig. 2(b)). From this dependence we can estimate the correlation length ξ, which is given by eqn (2). In Fig. 2(b) a clear increase of ξ with the salt concentration approaching c^* is observed. Such an increase is typical for an approach to a critical point or a spinodal line in a solution. In Section 4 we discuss that indeed the static and dynamic scattering data point towards a spinodal near c^*. Additionally, we attribute the strong increase of ξ to the emergence of protein clusters for which the dynamic data give further evidence through the appearance of a second component (see below). From the values of ξ in Fig. 2(b), an aggregate radius of about 100 nm close to c^* can be estimated. For the interpretation of ξ (eqn (2)) in terms of a correlation length extending over several

Table 1 Values of c^* for different BSA concentrations determined from visual inspection (center column) and by a linear fit to the SLS inverse scattering intensity (right column), respectively. Results from visual inspection are obtained by preparing a series of samples and judging their turbidity by eye. Errors (a) are given by the increment in c_s within a specific series. However, a systematic error (b) is present which is mainly due to the variation of turbidity onset across different batches[36,37]

BSA [mg ml^{-1}]	c^* [mM]	c^*_{SLS} [mM]
5	$0.40 \pm 0.02^a \pm 0.1^b$	0.42 ± 0.03
10	$0.78 \pm 0.03^a \pm 0.1^b$	0.79 ± 0.08
15	$0.17 \pm 0.04^a \pm 0.1^b$	1.20 ± 0.11
20	$1.58 \pm 0.03^a \pm 0.1^b$	1.63 ± 0.29

molecules we plot the results for $c_s > 0.32$ mM. We have observed a similar behavior for other protein concentrations close to c^* (data not shown), also resulting in $\xi \approx 150$ nm, corroborating the picture of formation of aggregates of considerable size ~100 nm close to c^*.

We note that measurements after several days on the same samples confirm results of the same order of magnitude as reported above as long as $c_s < c^*$ holds, confirming the stability (in a colloidal sense) of the aggregates in this regime. For $c_s > c^*$, however, the samples show precipitated aggregates after some days, indicating continuously growing aggregates.

3.2 Dynamic behavior probed by DLS

In this section we present the results obtained by dynamic light scattering for the previously discussed samples. We show the application of a two-exponential model (eqn (6)) to a series of BSA solutions with YCl_3. In addition, we will discuss the inverse scattering ratio resulting from SLS in this section, because this ratio equally follows a master curve.

Fig. 3(a) depicts examples of a selected set of raw data with changing salt concentrations c_s at fixed BSA protein concentration $c_p = 15$ mg ml^{-1} and scattering angle $\theta = 60°$. A general trend of an increasing decay time with increasing c_s can be observed. In the depicted example dataset, two exponentials are required for all conditions for reasonable fits. The two-exponential character of the decay is most pronounced for $c_s = 1.1$ mM. The good agreement of the two-exponential fits with the measured autocorrelation functions confirms the validity of the model function (eqn (6)). We note that the sample with $c_s = 1.2$ mM depicted in Fig. 3(a) was slightly turbid, and hence has to be considered to be above c^*. However, very big aggregates were not seen, which normally result in an apparent size determined by DLS around 1 μm for the studied system.

We briefly discuss the initial qualitative observation of the general trend towards a decrease of the overall dynamics upon addition of salt. We attribute this trend to charge-screening at least for low salt concentrations. On the other hand, the existence of slowly diffusing particles at least at higher salt concentrations can also be deduced from the pronounced second decay in Fig. 3(a) (modeled by the second exponential with relaxation rate Γ_2). In combination with the increase in scattering intensity this indicates the presence of a transition,

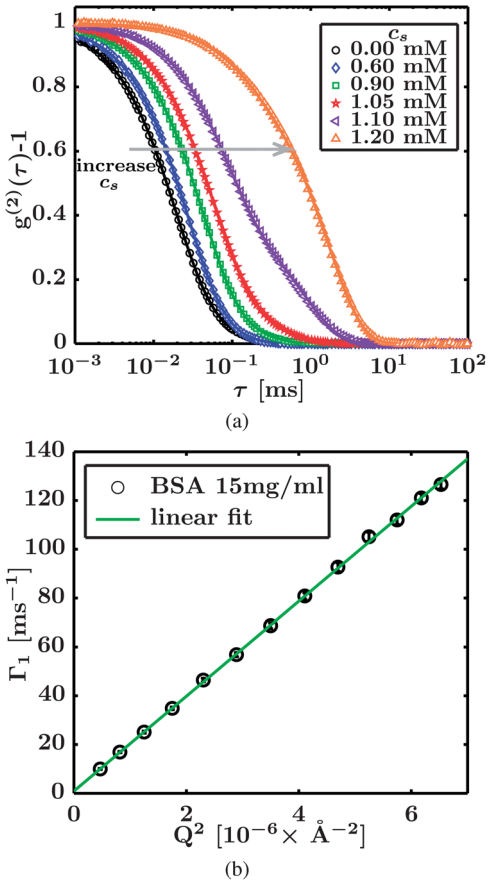

Fig. 3 (a) Autocorrelation functions $g^{(2)}(\tau) - 1$ (*cf.* eqn (5)) *versus* lag time τ recorded on samples at $c_p = 15$ mg ml^{-1} at the scattering angle $\theta = 60°$ ($Q^2 = 1.75 \times 10^{-6}$ Å^{-2}) for different c_s as specified in the legend (symbols) and fits of eqn (6) (lines superimposed on the symbols), and (b) relaxation rate Γ_1 (symbols, *cf.* eqn (6)) *versus* Q^2 for a BSA solution at $c_p = 15$ mg ml^{-1} without salt and fit of the function $\Gamma_1 = D_1Q^2$ (solid line).

which we tentatively attribute to the formation of big aggregates close to c^*.

In Fig. 3(b) the fit results for the relaxation rates Γ_1 corresponding to the first exponential (eqn (6)) are displayed for one sample. The fit to these results and the intersection at $Q = 0$ confirm the relation $\Gamma_1 = D_1Q^2$. This relation which is indicative of simple Brownian diffusion holds within experimental accuracy for most of our samples.

3.3 Master curves of the diffusion coefficients as a function of c_s/c^*

In parts (a), (b), and (d) of Fig. 4 we display the fit results of our DLS data together with SLS results [part (c) of the figure, *cf.* subsection 3.4] against the renormalized axis c_s/c^*: In part (a), the diffusion coefficient $D_1 = \Gamma_1/Q^2$ associated with the fast relaxation contribution of the two-exponential model (eqn (6)) is given.

From a linear fit in Fig. 4(a),

$$D_1(\tilde{c}) = a \cdot \tilde{c} + b \quad (7)$$

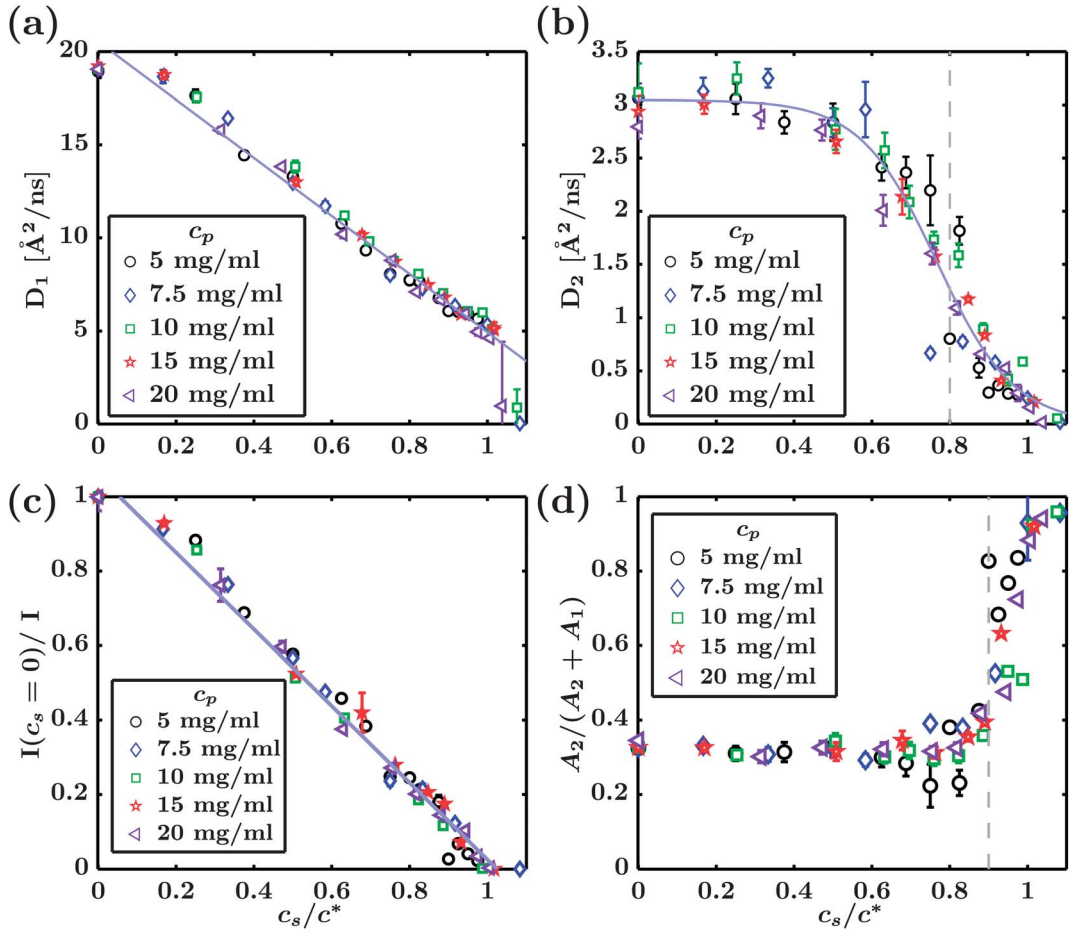

Fig. 4 Overview of results from both DLS (parts (a), (b), and (d)), and SLS (part (c)) plotted *versus* the renormalized YCl_3 salt concentration c_s/c^* for different protein concentrations c_p (note that $c_s = c^*(c_p)$): (a) diffusion coefficient D_1 corresponding to the faster component resulting from the two-exponential fit (eqn (6)) (symbols). The solid line is a linear fit to the entire dataset. (b) Diffusion coefficient D_2 corresponding to the slower component resulting from the two-exponential fit (eqn (6)) (symbols). The solid line is a fit of the heuristic two-state model (eqn (8)) to the entire dataset. (c) Normalized inverse scattering intensity (symbols) and linear fit (solid line). (d) Weight ratio of the fast and the sum of both fast and slow components in the fit (eqn (6), measured at $\theta = 60°$). Vertical dashed lines are guides to the eye. The error bars are the 95% confidence limits from the fits. When not visible, the error bars are smaller than the symbols.

with $\tilde{c} = c_s/c^*$ we obtain $a = (-15.6 \pm 1)$Å^2 ns^{-1} and $b = (20.5 \pm 0.9)$Å^2 ns^{-1}. The results in Fig. 4(a) indicate that the decrease of the fast component D_1 is universal, *i.e.* the protein concentration only enters *via* $c^* = c^*(c_p)$, and the linear functional form $D_1(c_s/c^*)$ is general within the investigated concentration range and accuracy of the experiment. We tentatively attribute the fast diffusion component D_1 to protein monomers. The diffusion coefficients D_1 near c^* are of the order of 5 Å^2 ns^{-1}, *i.e.* slightly smaller than the dilute-limit diffusion coefficient D_0 of BSA, $D_0 = (6.32 \pm 0.07)$Å^2 ns^{-1} at $T = 296$ K.[50]

The fact that $D_1(c_s/c^* \simeq 0) > D_0$ and decreases with increasing salt concentration indicates that the repulsive potential between the monomers is reduced, although the results very close to c^* are less reliable due to low exponent weights A_1 in the fits (*cf.* Fig. 4(d), discussed further below).

We apply the same scaling of the salt axis c_s/c^* to the slower component in the two-exponential decay. The resulting diffusion coefficients $D_2 = \Gamma_2/Q^2$ are depicted in part (b) of Fig. 4. The absolute values of D_2 are systematically lower than those of D_1, corroborating the assumption that D_2 is associated with higher n-mers or aggregates of proteins. The results for D_2 again indicate a universal functional form, *i.e.* $D_2 = D_2(c_s/c^*)$. However, in the case of D_2, this function is far from linear. Instead, following an initially weak salt-dependence, a strong decrease is observed above $c_s/c^* \approx 0.5$, with a change of curvature at $c_s/c^* \approx 0.8$ and a subsequent flattening.

Observing that $D_2 = D_2(c_s/c^*)$ drops from an apparent plateau for $c_s/c^* < 0.4$ to 0 for $c_s/c^* \geq 0.4$, in the heuristically simplest approach we fit the data with a hypothetical two state transition model based on a smeared-out step function,

$$D_2(\tilde{c}) = D_{2,0}(1 - \Theta[\tilde{c} - c_0;\ \Delta c])$$
$$\Theta[\tilde{c};\ \Delta c] = \frac{1}{\left(1 + \exp\left(-2\frac{\tilde{c}}{\Delta c}\right)\right)}, \quad (8)$$

where $\tilde{c} = c_s/c^*$. The fit with this function is shown as a solid line in Fig. 4(b), where $D_{2,0}$, c_0, and Δc are free parameters. Δc defines the 'smoothness', *i.e.* decay interval of the curve, and

$\Theta(\tilde{c};\ \Delta c \rightarrow 0)$ is the Heaviside step function. The fit yields $D_2(c_s/c^* = 0) = D_{2,0} = (3.03 \pm 0.14)$ mM, $\Delta c = 0.20 \pm 0.30$, and $c_0 = 0.77 \pm 0.02$. The change of sign in the curvature of D_2 is at $c_0 \approx 0.8$.

3.4 Master curve of the normalized intensity as a function of c_s/c^*

In Fig. 4(c) the ratio of $1/I(0) = Kc/R(c_s = 0)$ and $1/I(c_s) = K_c/R(c_s)$ as obtained from SLS is plotted against the renormalized salt concentration c_s/c^*. We emphasize that c^* depends on c_p, *i.e.* $c^* = c^*(c_p)$ (*cf.* Fig. 2). We observe that following this renormalization, the results for different protein concentrations coincide almost perfectly. In view of eqn (3) we interpret this observation as a universal scaling behavior of the renormalized compressibility as a function of the renormalized salt concentration. A linear fit of the entire dataset using eqn (7) yields: $a = -1.10 \pm 0.04$, $b = 1.11 \pm 0.03$, resulting in a value of $-a/b = 1.01 \pm 0.03$ as intersection with the abscissa, which corresponds within the errors to the value of $c_s/c^* = 1$.

Finally, we also present and discuss the ratio of the amplitude coefficients A_1 and A_2 in eqn (6) (Fig. 4(d)). We observe that $A_2/(A_2 + A_1)$ is nearly constant up to $c_s/c^* \approx 0.8$. For $c_s/c^* > 0.9$, $A_2/(A_2 + A_1)$ increases strongly. A difficulty in the interpretation of this observation arises from the fact that both an increasing particle size at a constant number density as well as an increasing number density at a constant particle size of a large-size species in suspension can cause an increasing contribution to the scattering signal. Since A_2 is the amplitude coefficient associated with the slow diffusion component D_2 which we tentatively assign to protein n-mers or clusters, the result for $A_2/(A_2 + A_1)$ agrees with the assumption that close to c^* aggregates form. In this picture, these aggregates have the same scattering weight in DLS as the monomers around $c_s/c^* \approx 0.9$ and cannot be attributed to dust or other impurities related to sample preparation. Therefore, the weight ratio provides further evidence that aggregates grow close to c^* and contribute increasingly to the scattered intensity when c_s approaches c^*, as is expected for particles with increasing size.

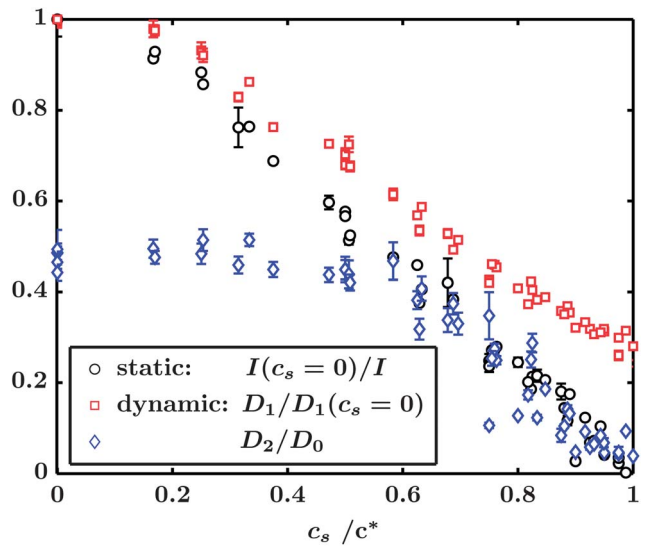

Fig. 5 Normalized SLS intensity data together with normalized DLS data for D_1 and D_2. The D_1 data have been normalized by an average value of D_1 at $c_s = 0$ whereas the D_2 data have been normalized by D_0, the diffusion coefficient of BSA in the dilute limit. See text for a discussion. When not visible, the error bars are smaller than the symbols.

4 Discussion

The collapsed data for $Q = 0$ from static light scattering (Fig. 4(c)) show a linear decrease of the pressure derivative $\partial p/\partial c$ with increasing salt concentration c_s towards zero near $c_s = c^*$ (since $1/I \propto \chi_T^{-1} = c(\partial p/\partial c)$). At the same time, the correlation length in the system increases considerably when $c_s \rightarrow c^*$ (Fig. 2(b)). This is indicative of approaching a spinodal line (where c_s takes the role of the inverse temperature in the picture of a simple liquid). However, it is not certain if c^* exactly coincides with a spinodal since we have identified c^* by the onset of a light absorption in a transmission experiment.[36] Interestingly, the results for *both* collective diffusion coefficients D_1 and D_2 are consistent with this thermodynamic behavior upon an appropriate rescaling, see Fig. 5. In order to understand this rescaling, we note that the collective diffusion coefficient $D_c = D(Q^2 \rightarrow 0)$ describes the diffusion of the perturbations δc of the protein concentration field c *via* $\partial\delta c/\partial t = D_c\Delta\delta c$. From linear response we find†

$$\frac{\partial\delta c}{\partial t} = \frac{\Gamma(c, c_s)}{kT}\left(\frac{\partial p}{\partial c}\right)\Delta\delta c, \tag{9}$$

with the salt and protein concentration dependent mobility $\Gamma(c, c_s)$. Thus, the collective diffusion coefficient is identified as

$$D_c = (\Gamma(c,\ c_s)/kT)(\partial p/\partial c). \tag{10}$$

Therefore, it is clear that the behavior of D_c with increasing salt concentration is proportional to the SLS quantity $1/I(c_s)$. The particular feature of the present system is the appearance of the "slow" diffusion component (which is not present in a simple liquid and which, in our opinion, signals protein clustering in some form). However, for $c_s < c^*$ the system is a single-component system in equilibrium and thus (in the limit $Q^2 \rightarrow 0$) only one collective diffusion process (as described by eqn (9)) with associated diffusion coefficient should be present. We take a pragmatic point of view and recall that for $c_s \rightarrow 0$ the "slow" component is present in the scattering signal but it is weak compared with the "fast" (monomer) component, see Fig. 4(d). The collective diffusion process (eqn (9)) should be related to the "fast" component and the associated diffusion coefficient is D_1. On the other hand, for $c_s \rightarrow c^*$ the "slow" component dominates and thus the collective diffusion process (eqn (9)) should be related to this component (with coefficient D_2). To compare D_1 and D_2 quantitatively to the normalized quantity $I(c_s = 0)/I(c_s)$ one must divide by the diffusion coefficients at zero salt concentration, $D(c_s = 0)$. We recall that the monomer

† Alternatively, one can derive this diffusion equation from dynamic density functional theory:[51,52] $(\partial c/\partial t) = \Gamma\nabla\cdot(c\nabla[\delta\mathcal{F}[c]/\delta c])$, where $\mathcal{F}[c]$ is the equilibrium free energy functional. For small deviations δc from equilibrium this is rigorously valid and on large time and length scales one can use a local gradient expansion and eqn (9) follows.

diffusion coefficient in the dilute ("ideal gas") limit at $c_s = 0$ has been determined previously[50] as $D_0 = (6.32 \pm 0.07)$Å^2 ns^{-1} at $T = 296$ K. Our results for $D_1(c_s = 0) \approx 20$ Å^2 ns^{-1} indicate that the measured c_p values are above the dilute limit, in a range where $D_1(c_s = 0)$ is roughly constant, on a plateau, before it decreases with further increase of c_p.[43] For the "slow" component, the associated protein density at $c_s = 0$ is small and close to the dilute ("ideal gas") limit. Therefore we normalize D_2 by the dilute limit diffusion coefficient D_0 since this is consistent with eqn (9) where we consider the diffusion of a single-component protein concentration field (and not of a multi-component field associated with clusters). This normalization has been done in Fig. 5, and we find a nice agreement of the static data $I(c_s = 0)/I$ with the normalized D_1 and D_2, respectively, in the regions where the corresponding components are dominant, that is, the low-c_s region for D_1 and the region $c_s \to c^*$ for D_2. Thus the "slow" (cluster) component also signals the approach to the spinodal as observed in the static data and becomes the dominant component there.

Considering the fact that $D_2(c_s = 0) \approx 3$ Å^2 ns$^{-1} < D_0$, one might be tempted to associate this with the diffusion of particles larger than a single protein, *i.e.* clusters. However, we do not have solid information on the size distribution of clusters and their temporal stability. Further understanding will require future enhanced theoretical models of clustering fluids. The analysis of time-dependent correlation functions in such models would hint at whether a simple picture of additional cluster components with associated, effective diffusion coefficients is viable.

For a better understanding of possible cluster formation mechanisms in BSA solutions with yttrium, we recall our recent crystallographic studies of the globular protein β-lactoglobulin (BLG) showing that Y^{3+} binds to specific patches on the protein surface.[40] For BLG, 4 binding sites are attributed to each protein in the crystal structure, and they function as bridges between the proteins.[40] When bound to the surface, trivalent counterions reduce the global charge of the protein and can locally invert the surface charge. Bound counterions thus represent highly orientation-dependent attractive interactions due to enthalpic binding energy and electrostatic interaction. Furthermore, binding of counterions also varies the hydrophobicity patterns on the protein surface, which might be another source of anisotropic interaction.[46]

A promising interpretation for the protein–protein interaction induced by counterions thus involves the concept of patchy particles.[53] At least qualitatively, the observed behavior in BSA solutions is in accordance with the work by F. Sciortino, E. Bianchi and collaborators, who discuss the formation of clusters[54–56] from patchy particles. In this theoretical limit, electrostatic and isotropic interactions such as van der Waals attraction are neglected, and clusters are mainly stabilized against macroscopic aggregation by entropic effects.

However, the balance of long-range Coulomb repulsion and short-range attraction of any kind represents a mechanism of charge stabilization for the cluster sizes.[10] This theory has been applied to lysozyme clusters.[12,57] For our system, we find very large correlation lengths and thus mean cluster sizes of ~100 nm close to the point of zero charge of the global protein. Thus, the explanation for cluster formation in our system is a complicated topic, combining presumably two mechanisms for stabilization (charge and entropy) as well as different kinds of attraction (electrostatic, binding, hydrophobic, and van der Waals).

Following the discussion on the mechanisms of cluster formation, the lifetime of clusters is another inspiring topic. We observe in this work that, approaching the spinodal line by increasing the salt concentration c_s, the apparent diffusion coefficients decrease and the scattered intensity increases by orders of magnitude. We thus propose that protein clusters can form in solution *via* metal ion bridging. In a simple approximation, the binding energy, E, of proteins in clusters is proportional to the number of bridging ions and further affected by the global charge of the proteins. From this energy, an Arrhenius-like escape rate for a single protein at the cluster surface is proportional to $\exp(E/k_BT)$,[58] suggesting that the lifetime of clusters increases with salt concentration.

Approximately 3 to 4 Y^{3+}-ions per protein are required for neutralization of the BSA protein surface charge.[40] At low salt concentrations or small c_s, the binding energy is very weak, since the proteins are still overall negatively charged. The initially limited number of bridging bonds can only hold small clusters for a short time. With an increasing number of ions per protein, the overall net charge is reduced and more bridging bonds become available. The clusters may then exist for longer times, and eventually the cluster lifetime can exceed the time-scale of the DLS experiment.

Reconsidering the two distinct diffusion coefficients D_1 and D_2 (Fig. 4) identified in our DLS data, we assume that protein monomers and clusters coexist over a wide range of protein and salt concentrations. This result is in agreement with similar observations and theory by Pan *et al.*[59] on lysozyme solutions. Looking at Fig. 4, we may further hypothesize that due to the smeared-out two-step transition character (eqn (8)) of the slower diffusion coefficient D_2 attributed to clusters, few or no clusters exist for $c_s/c^* \ll 0.5$. Within the range centered around $c_s/c^* \approx 0.8$, clusters occur and become increasingly stable on the time-scale of the DLS experiment for $c_s/c^* > 0.9$. Pan *et al.*[59] provide further theoretical arguments that the coexistence of protein monomers and transient clusters may be expected in relatively dilute protein solutions in agreement with our observations. By contrast, for highly concentrated protein solutions beyond the range investigated in our study, Pan *et al.*[59] predict the emerging of dynamic density fluctuations.

5 Summary and conclusion

Our study of BSA protein solutions containing the trivalent salt YCl_3 approaching the precipitate boundary c^* provides new and systematic data concerning the salt-induced protein cluster formation and aggregation behavior. Our experimental study using multi-angle light scattering gives quantitative information on several parameters: the static scattering clearly shows a substantial increase of the correlation length with increasing salt concentration. Further, the inverse compressibility decreases linearly over the entire salt concentration range c_s up to c^*, indicative of an approach to a spinodal. The dynamic scattering

data show the presence of two components, "fast" and "slow". At small c_s, the "fast" component is dominant and corresponds to monomer diffusion, whereas the "slow" component diffuses considerably slower than the monomers, consistent with being composed of small clusters. The "slow" component becomes stronger with increasing c_s and dominates near $c_s = c^*$. The collective diffusion coefficients extracted from both components confirm the approach to the spinodal line as seen in the static data. The results for the two diffusion coefficients as well as the inverse compressibility ratio $\chi(c_s/c^*)/\chi(c_s = 0)$ can be described by universal functional forms $f = f(c_s/c^*)$ denoted by master curves. We note that our observations regarding the master curves hold for protein concentrations below 100 mg ml^{-1}. At higher protein concentrations, caging or crowding effects appear to cause deviations of the observable apparent long-time collective diffusion from the master curves.[45]

The faster diffusion component attributed to monomers decays linearly with c_s, while the slower diffusion component attributed to aggregates agrees with a phenomenological two-step transition model, supporting an onset of aggregate formation for $c_s/c^* > 0.5$ which remains presumably transient up to $c_s/c^* \approx 0.8$. The scattered light of the clusters with respect to that of the monomers increases significantly above $c_s/c^* > 0.9$ up to values which indicate a solution dominated by increasingly stable clusters close to c^*. We hope that our experimental study will inspire efforts for a full theoretical understanding of our observations, which is at present not available. Both from an experimental and a theoretical point of view, it is clearly desirable to better characterize cluster properties, such as *e.g.* cluster size and lifetime distributions.

Acknowledgements

M. Grimaldo, D. Soraruf, and F. Zanini gratefully acknowledge ILL studentships. F.R.-R. acknowledges a fellowship from the Studienstiftung des Deutschen Volkes. This work was supported by the DFG.

References

1 P. G. Vekilov, *Soft Matter*, 2010, **6**, 5254.
2 A. M. Kulkarni, N. M. Dixit and C. F. Zukoski, *Faraday Discuss.*, 2003, **123**, 37–50.
3 K. P. Johnston, J. A. Maynard, T. M. Truskett, A. U. Borwankar, M. A. Miller, B. K. Wilson, A. K. Dinin, T. A. Khan and K. J. Kaczorowski, *ACS Nano*, 2012, **6**, 1357–1369.
4 P. G. Vekilov, *Nanoscale*, 2010, **2**, 2346–2357.
5 J. M. Garcia-Ruiz, *J. Struct. Biol.*, 2003, **142**, 22–31.
6 P. N. Segrè, V. Prasad, A. B. Schofield and D. A. Weitz, *Phys. Rev. Lett.*, 2001, **86**, 6042–6045.
7 F. Zhang, F. Roosen-Runge, A. Sauter, R. Roth, M. W. A. Skoda, R. Jacobs, M. Sztucki and F. Schreiber, *Faraday Discuss.*, 2012, **159**, 313–325.
8 A. George and W. W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1994, **50**, 361–365.
9 P. G. Vekilov, *Cryst. Growth Des.*, 2010, **10**, 5007–5019.
10 J. Groenewold and W. K. Kegel, *J. Phys. Chem. B*, 2001, **105**, 11702–11709.
11 J. Groenewold and W. K. Kegel, *J. Phys.: Condens. Matter*, 2004, **16**, S4877.
12 S. B. Hutchens and Z.-G. Wang, *J. Chem. Phys.*, 2007, **127**, 084912.
13 Y. Liu, W. R. Chen and S. H. Chen, *J. Chem. Phys.*, 2005, **122**, 044507.
14 F. Sciortino, S. Mossa, E. Zaccarelli and P. Tartaglia, *Phys. Rev. Lett.*, 2004, **93**, 055701.
15 S. Mossa, F. Sciortino, P. Tartaglia and E. Zaccarelli, *Langmuir*, 2004, **20**, 10756–10763.
16 P. Kowalczyk, A. Ciach, P. Gauden and A. Terzyk, *J. Colloid Interface Sci.*, 2011, **363**, 579–584.
17 E. Spohr, B. Hribar and V. Vlachy, *J. Phys. Chem. B*, 2002, **106**, 2343–2348.
18 T. Jiang and J. Wu, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2009, **80**, 021401.
19 A. Fierro, T. Abete, A. Coniglio and A. de Candia, *J. Phys. Chem. B*, 2011, **115**, 7281–7287.
20 P. J. Lu, J. C. Conrad, H. M. Wyss, A. B. Schofield and D. A. Weitz, *Phys. Rev. Lett.*, 2006, **96**, 028306.
21 J. M. Tam, A. K. Murthy, D. R. Ingram, R. Nguyen, K. V. Sokolov and K. P. Johnston, *Langmuir*, 2010, **26**, 8988–8999.
22 A. K. Atmuri and S. R. Bhatia, *Langmuir*, 2013, **29**, 3179–3187.
23 Y. Yan, D. Seeman, B. Zheng, E. Kizilay, Y. Xu and P. L. Dubin, *Langmuir*, 2013, **29**, 4584–4593.
24 A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf and P. Schurtenberger, *Nature*, 2004, **432**, 492–495.
25 A. Stradner, F. Cardinaux and P. Schurtenberger, *J. Phys. Chem. B*, 2006, **110**, 21222–21231.
26 Y. Liu, L. Porcar, J. Chen, W.-R. Chen, P. Falus, A. Faraone, E. Fratini, K. Hong and P. Baglioni, *J. Phys. Chem. B*, 2011, **115**, 7238–7247.
27 L. Porcar, P. Falus, W.-R. Chen, A. Faraone, E. Fratini, K. Hong, P. Baglioni and Y. Liu, *J. Phys. Chem. Lett.*, 2010, **1**, 126–129.
28 S. Barhoum and A. Yethiraj, *J. Phys. Chem. B*, 2010, **114**, 17062–17067.
29 A. Shukla, E. Mylonas, E. Di Cola, S. Finet, P. Timmins, T. Narayanan and D. I. Svergun, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5075–5080.
30 F. Cardinaux, E. Zaccarelli, A. Stradner, S. Bucciarelli, B. Farago, S. U. Egelhaaf, F. Sciortino and P. Schurtenberger, *J. Phys. Chem. B*, 2011, **115**, 7227–7237.
31 Y. Li, V. Lubchenko and P. G. Vekilov, *Rev. Sci. Instrum.*, 2011, **82**, 053106.
32 R. Piazza and S. Iacopini, *Eur. Phys. J. E*, 2002, **7**, 45–48.
33 W. Pan, O. Galkin, L. Filobelo, R. L. Nagel and P. G. Vekilov, *Biophys. J.*, 2007, **92**, 267–277.
34 O. Gliko, N. Neumaier, W. Pan, I. Haase, M. Fischer, A. Bacher, S. Weinkauf and P. G. Vekilov, *J. Am. Chem. Soc.*, 2005, **127**, 3433–3438.

35 O. Gliko, W. Pan, P. Katsonis, N. Neumaier, O. Galkin, S. Weinkauf and P. G. Vekilov, *J. Phys. Chem. B*, 2007, **111**, 3106–3114.
36 F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, S. Zorn, R. A. Martin, C. M. Martin, G. F. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher and F. Schreiber, *Phys. Rev. Lett.*, 2008, **101**, 148101.
37 F. Zhang, S. Weggler, M. J. Ziller, L. Ianeselli, B. S. Heck, A. Hildebrandt, O. Kohlbacher, M. W. A. Skoda, R. M. J. Jacobs and F. Schreiber, *Proteins*, 2010, **78**, 3450–3457.
38 F. Roosen-Runge, B. S. Heck, F. Zhang, O. Kohlbacher and F. Schreiber, *J. Phys. Chem. B*, 2013, **117**, 5777–5787.
39 F. Zhang, R. Roth, M. Wolf, F. Roosen-Runge, M. W. A. Skoda, R. M. J. Jacobs, M. Sztucki and F. Schreiber, *Soft Matter*, 2012, **8**, 1313–1316.
40 F. Zhang, G. Zocher, A. Sauter, T. Stehle and F. Schreiber, *J. Appl. Crystallogr.*, 2011, **44**, 755–762.
41 F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, R. A. Martin, C. M. Martin and F. Schreiber, *J. Phys. Chem. B*, 2007, **111**, 251–259.
42 F. Zhang, F. Roosen-Runge, M. W. A. Skoda, R. M. J. Jacobs, M. Wolf, P. Callow, H. Frielinghaus, V. Pipich, S. Prevost and F. Schreiber, *Phys. Chem. Chem. Phys.*, 2012, **14**, 2483–2493.
43 M. Heinen, F. Zanini, F. Roosen-Runge, D. Fedunova, F. Zhang, M. Hennig, T. Seydel, R. Schweins, M. Sztucki, M. Antalik, F. Schreiber and G. Nägele, *Soft Matter*, 2012, **8**, 1404–1419.
44 F. Roosen-Runge, M. Hennig, F. Zhang, R. M. J. Jacobs, M. Sztucki, H. Schober, T. Seydel and F. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11815–11820.
45 F. Roosen-Runge, M. Hennig, T. Seydel, F. Zhang, M. W. Skoda, S. Zorn, R. M. Jacobs, M. Maccarini, P. Fouquet and F. Schreiber, *Biochim. Biophys. Acta, Proteins Proteomics*, 2010, **1804**, 68–75.
46 F. Roosen-Runge, F. Zhang, F. Schreiber and R. Roth, submitted.
47 H. A. Sober, *CRC Handbook of Biochemistry: selected data for molecular biology*, CRC, 1970.
48 W. Brown, *Light Scattering: Principles and Development*, Clarendon Press, 1996.
49 S. Provencher, *Comput. Phys. Commun.*, 1982, **27**, 213–227.
50 A. K. Gaigalas, J. B. Hubbard, M. McCurley and S. Woo, *J. Phys. Chem.*, 1992, **96**, 2355–2359.
51 U. Marconi and P. Tarazona, *J. Chem. Phys.*, 1999, **110**, 8032–8044.
52 A. Archer and M. Rauscher, *J. Phys. A: Math. Gen.*, 2004, **37**, 9325–9333.
53 G. Jackson, W. G. Chapman and K. E. Gubbins, *Mol. Phys.*, 1988, **65**, 1–31.
54 F. Sciortino, C. De Michele, S. Corezzi, J. Russo, E. Zaccarelli and P. Tartaglia, *Soft Matter*, 2009, **5**, 2571–2575.
55 E. Bianchi, P. Tartaglia, E. La Nave and F. Sciortino, *J. Phys. Chem. B*, 2007, **111**, 11765–11769.
56 E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli and F. Sciortino, *Phys. Rev. Lett.*, 2006, **97**, 168301.
57 J. Groenewold, T. Zhang and W. K. Kegel, *J. Phys. Chem. B*, 2011, **115**, 7264–7267.
58 J. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, London, 1991.
59 W. Pan, P. G. Vekilov and V. Lubchenko, *J. Phys. Chem. B*, 2010, **114**, 7620–7630.

