The role of cluster formation and metastable liquid—liquid phase separation in protein crystallization

Fajun Zhang,*[a] Felix Roosen-Runge,[a] Andrea Sauter,[a] Roland Roth,[b] Maximilian W. A. Skoda,[c] Robert M. J. Jacobs,[d] Michael Sztucki[e] and Frank Schreiber[a]

Received 15th February 2012, Accepted 19th March 2012
DOI: 10.1039/c2fd20021j

We discuss the phase behavior and in particular crystallization of a model globular protein (beta-lactoglobulin) in solution in the presence of multivalent electrolytes. It has been shown previously that negatively charged globular proteins at neutral pH in the presence of multivalent counterions undergo a "re-entrant condensation (RC)" phase behavior (Zhang *et al.*, *Phys. Rev. Lett.*, 2008, **101**, 148101), *i.e.* a phase-separated regime occurs in between two critical salt concentrations, $c^* < c^{**}$, giving a metastable liquid–liquid phase separation (LLPS). Crystallization from the condensed regime has been observed to follow different mechanisms. Near c^*, crystals grow following a classic nucleation and growth mechanism; near c^{**}, the crystallization follows a two-step crystallization mechanism, *i.e*, crystal growth follows a metastable LLPS. In this paper, we focus on the two-step crystal growth near c^{**}. SAXS measurements indicate that proteins form clusters in this regime and the cluster size increases approaching c^{**}. Upon lowering the temperature, *in situ* SAXS studies indicate that the clusters can directly form both a dense liquid phase and protein crystals. During the crystal growth, the metastable dense liquid phase is dissolved. Based on our observations, we discuss a nucleation mechanism starting from clusters in the dilute phase from a metastable LLPS. These protein clusters behave as the building blocks for nucleation, while the dense phase acts as a reservoir ensuring constant protein concentration in the dilute phase during crystal growth.

Introduction

Crystallization plays a decisive role in many processes in nature and industry. For example, the pharmaceutical industry requires production of the desired crystal form of the drug molecules which is important for their biofunction and stability.[1] However, it has long been known that the unique features of crystallization, including crystal lattice and polymorphism, particle size, and its distribution, are defined in the nucleation stage. In spite of the huge efforts over the last decades, our understanding of the early stage of crystallization is still limited. A breakthrough over the last decade has revealed new insight into this step: studies have shown that

[a]*Institut für Angewandte Physik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany. E-mail: fajun.zhang@uni-tuebingen.de*
[b]*Institut für Theoretische Physik, Universität Erlangen-Nürnberg, Germany*
[c]*STFC, ISIS, Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0OX, UK*
[d]*Department of Chemistry, Chemistry Research Laboratory, University of Oxford, UK*
[e]*European Synchrotron Radiation Facility, 6 rue Jule Horowitz, F-38043 Grenoble, France*

in colloid and protein solutions, the attractive potential is short-ranged as compared to their size, which is crucial for their phase behavior, and a "two-step" mechanism has been proposed to explain the crystallization behavior in these systems under suitable conditions, *i.e.* nucleation events follow a metastable liquid–liquid phase separation (LLPS).[1–3]

In the region of LLPS, however, colloidal systems with spherical isotropic short-range potential will become a gel or arrested phase instead of a dense liquid phase. In other words, the gelation line coincides with the phase separation boundary.[4] Proteins, which differ from conventional colloids by having a non-spherical shape and inhomogeneous charge pattern, have shown liquid–liquid coexistence under suitable conditions[5–17] and gelation at higher volume fraction or lower temperature.[18–21] Near LLPS, recent studies reveal the formation of dense protein clusters. The formation and physical origin of clusters in colloidal and protein solutions have attracted significant attention in soft matter studies during the last few years. For instance, it is a crucial step for understanding the mineralization process.[22] The cluster phenomenon is closely related to the interactions and phase behavior of these systems. For example, counter-balanced interactions have been reported to lead to equilibrium as well as transient clusters in concentrated protein solution.[23–25] Recent studies have revealed another type of cluster phase existing in concentrated protein solutions, such as lysozyme, hemoglobin, *etc.*[26–28] The clusters can be very big with 10^5 to 10^6 molecules. A theoretical model has been proposed to explain the origin of such a long-living cluster phase.[28]

The consequence of the formation of protein clusters (transient or equilibrium) and metastable LLPS is that it changes the kinetic pathway of crystal nucleation significantly. There is increasing evidence that clusters, nanoscale amorphous precipitates, and other more complex precursors in the aqueous phase play an important role in crystallization.[22,29–31] Computer simulations by ten Wolde and Frenkel show that far from the metastable LLPS critical point, the nucleation follows the classical nucleation mechanism. However, when approaching the critical point, the critical nucleus becomes highly disordered, liquid-like droplets which further follow a structural change to eventually become crystalline.[32] This finding inspires a two-step model, *i.e.* nucleation occurs within the dense liquid phase, which corresponds to the separation of order parameters (density and structure) during crystallization. It applies not only to proteins and colloidal systems, but also to small molecular systems.[33,34] However, the role of the protein cluster as well as the dense liquid phase during nucleation and protein crystallization is still not entirely clear.[3]

We have recently studied the phase behavior of globular proteins in solution in the presence of multivalent metal ions. It has been shown that solutions of negatively charged globular proteins at neutral pH in the presence of multivalent counterions undergo a "re-entrant condensation (RC)" phase behavior,[16,17,35–37] *i.e.* a phase-separated regime occurs in between two critical salt concentrations, $c^* < c^{**}$, including a metastable liquid–liquid phase separation (LLPS).[17] Crystallization from the condensed regime follows different mechanisms. Near c^*, crystals grow following a classic nucleation and growth mechanism; near c^{**}, the crystallization occurs from phase-separated protein solutions, indicating a two-step mechanism, *i.e.* crystal growth follows a metastable LLPS.[16]

In this work, we aim to achieve new insights into the role of LLPS and clustering in protein crystallization by presenting a study of the structural evolution during a two-step process. The questions we are interested in are the following: (1) what is the structure of proteins in solution near c^{**}? Are they still in their dimeric state or forming clusters? (2) If clusters are formed, how do the size and structure of the protein clusters depend on the location in the phase diagram? (3) Can the nuclei of crystals be formed *via* cluster–cluster aggregation? (4) What is the relationship between clusters and the LLPS? With these questions in mind, we have performed systematic SAXS measurements on a series of solutions, and the crystal growth has been followed using *in situ* measurements as a function of temperature.

Experimental

Materials

Globular β-lactoglobulin (BLG) from bovine milk (L3908) and yttrium chloride (YCl_3) were purchased from Sigma-Aldrich. Solutions were prepared by mixing stock solutions of BLG (67 mg ml^{-1}) and YCl_3 (100 mM). The phase diagram (protein concentration c_p *vs.* salt concentration c_s) was determined at room temperature (~22 °C) by monitoring the optical transmission of a series of protein solutions containing different salt concentrations.[35] The c_p values were determined by UV absorption using an extinction coefficient of 0.96 ml mg^{-1} at a wavelength of 278 nm.[38] Note that the presence of high concentration of buffer (such as HEPES and Tris buffer) can affect the phase behavior and the solubility of yttrium salts. However, with lower buffer concentration (about 5 mM), the effect on the solubility of yttrium salts is negligible. To avoid the effect of other ions, no buffer was used in this work for sample preparation except for specific reference data set in Fig. 2.[16]

Small-angle X-ray scattering (SAXS) measurements

The SAXS measurements were performed at the ESRF (Grenoble, France) on the beamline ID02 with a sample-to-detector distance of 2 m or a combination of 0.85 and 5 m in order to cover a larger q range of 0.04 to 7.8 nm^{-1}.[39] The data were collected by a high sensitivity fiber-optic coupled CCD (FReLoN) detector placed in an evacuated flight tube. The protein solutions were loaded using a flow-through capillary cell (diameter ~2 mm; wall thickness ~10 μm). No variation that would indicate radiation damage of SAXS profiles was observed comparing successive short exposures of cumulated time up to 3 s. For the temperature scan, each temperature was held for 15 min to allow equilibration. The incident and transmitted beam intensities were simultaneously recorded for each SAXS pattern with an exposure of 0.3 s. The 2D data were normalized to an absolute scale and azimuthally averaged to obtain the intensity profiles, and the solvent background was subtracted. For more detailed information on data reduction and q-resolution calibration, see ref. 40.

Results and discussions

Phase behavior of protein solutions in the presence of multivalent counterions

We first briefly summarize the phase diagram of BLG (c_p) as a function of the YCl_3 concentration (c_s) at room temperature (Fig. 1) with an extended range of c_p. This diagram provides a guide for optimizing the conditions for protein crystallization. For a given protein concentration c_p, an increase of the salt concentration c_s above a certain threshold (c^*) results in the protein solution becoming turbid and entering a two-phase state. When c_s is increased further (above c^{**}), the protein solution turns clear again. Thus, the two salt concentrations c^* and c^{**} divide the phase diagram into three regimes (Fig. 1). Regimes I and III contain clear protein solution, whereas the protein condenses (or aggregates) in Regime II. The addition of YCl_3 leads to a charge inversion on the protein surface, which has been proved using zeta potential measurements.[16,37] The experimental observations on the crystal growth from different regions indicate different growth mechanisms:[16] near c^*, crystal growth follows the classical nucleation theory, *i.e.* nucleation occurs from homogeneous supersaturated solutions. Near c^{**}, the solution has a transition temperature, T_h. Below T_h, a metastable LLPS occurs before crystallization, apparently resembling the so-called "two-step" nucleation mechanism. In the remaining part of paper, we will focus on the structure of protein clusters and crystallization in Regime III.

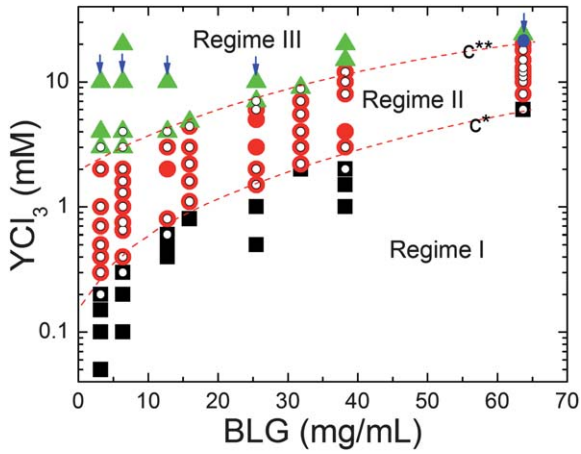

Fig. 1 Extended phase diagram of protein BLG as a function of protein (c_p) and salt, YCl_3 (c_s).[16] (a) Phase diagram at room temperature (22 °C). Solid symbols present the sample solutions at different regimes (see text). Small open symbols present the samples solutions where crystallization was observed at 4 °C. The small arrows in Regime III indicate the samples measured by SAXS.

Protein clusters and their structure in Regime III studied by SAXS

We now examine the solution structure of proteins in Regime III using SAXS. By fixing the salt concentration and varying the protein concentration, we approach the phase transition boundary, c^{**}. The selected solutions are shown in Fig. 1 as indicated by vertical arrows. The additional sample with higher protein concentration (67.0 mg mL^{-1}) with 15 mM YCl_3 near c^{**} is chosen for a temperature dependent measurement. Fig. 2 shows a series of SAXS profiles of BLG in solution with 10 mM YCl_3 and different protein concentrations. The scattering curves have been scaled by their protein concentrations in terms of the absolute intensity.

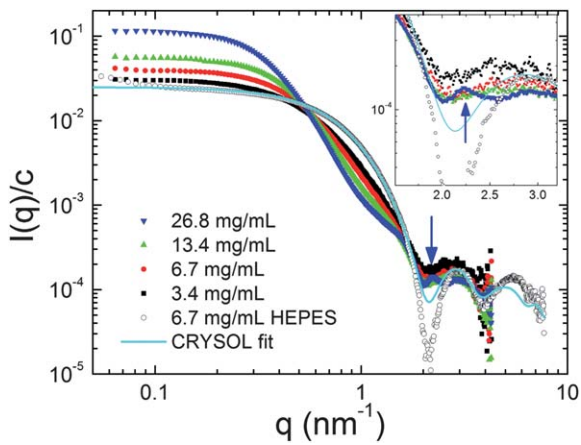

Fig. 2 SAXS profiles of BLG solutions with YCl_3 in the re-entrant regime. For a given value of c_s, a clear transition from dimer to cluster was observed with increasing protein concentration. For comparison, the SAXS curve of BLG in HEPES buffer is also shown, which can be well described by a dimer using the crystal structure (PDB code of 1BEB). Inset shows the expanding of the region with the arrow.

For comparison, a SAXS profile for BLG in 20 mM HEPES buffer is also plotted. These data were collected at two sample-to-detector distances (0.85 and 5 m) covering a large q range from 0.03 to 7.8 nm^{-1}. Two scattering minima are clearly visible at $q = 2.2$ and 4.2 nm^{-1}, respectively. In HEPES buffer (pH 7.0), BLG is dissolved as dimers,[41–43] which is further confirmed by fitting the SAXS data using CRYSOL[44] with its crystal structure (PDB code 1BEB). The radius of gyration, R_g, of the BLG dimer obtained by the fitting procedure is 2.33 nm, which is in good agreement with those reported in the literature.[45] For solutions with 10 mM YCl_3 and increasing BLG concentration, the scattering curves change smoothly. At 3.4 mg mL^{-1} protein, well below the phase boundary, the scattering intensity in the low q region ($q < 0.5$ nm^{-1}) is higher as compared to the dimeric state, whereas in the intermediate q region (0.5 $nm^{-1} < q < 1.4$ nm^{-1}), the scattering intensity is lower. With increasing the protein concentration, this trend becomes more significant and two crossing points are observed at $q = 0.5$ and 1.4 nm^{-1}, respectively. With 26.8 mg mL^{-1} protein, a kink is well-developed in the q region between the crossing points, indicating the formation of clusters. Basic structural parameters of these clusters, such as R_g, and the normalized forward intensity, $I(0)$, are obtained from a Guinier analysis.[46] As shown in Table 1, R_g increases when approaching the phase boundary. $I(0)$ is related to the molecular weight of the cluster, which can be used to estimate the number of dimers within the clusters. Simply dividing the normalized $I(0)$ by 0.0248 (the value obtained from the form factor), the numbers are estimated as 1.3, 1.7, 2.7, 4.5, respectively. These results are listed in Table 1. A new maximum appears at $q = 2.2$ nm^{-1} for $c_p > 20$ mg mL^{-1} as indicated by the arrow in Fig. 2 (expanded in the figure). This maximum, which corresponds to a center-to-center distance of $d = 2\pi/q = 2.82$ nm, is very close to the monomer–monomer distance within a dimer (3.08 nm) (calculated from the crystal structure: 1BEB). Thus, the appearance of this peak is attributed to the nearest neighbour correlation within clusters, indicating the liquid-like structure within the clusters.

The cluster phase observed in our system is due to the balance of electrostatic repulsion and the attraction.[47] We have demonstrated that in the re-entrant regime, the effective surface charge of proteins is inverted from negative to positive, which provides the long range repulsion.[35,37] On the other hand, the bridging effect of the yttrium cations provides the short range attraction. Although other interactions, such as van der Waals attraction may also contribute to the overall attractive potential, we assume that the bridging effect dominates, which is consistent with the crystallography study.[16] Clustering for qualitatively comparable interactions has been observed also for lysozyme and colloids in solution.[25]

Since the bridging effect seems to contribute anisotropic attractions, not only the crystal structure, but also the structure of these clusters should reflect the anisotropic interaction between the proteins being bridged by counterions. Indeed, the SAXS profiles of the cluster phases can be reasonably fitted by typical cluster structures created based on the crystal structure determined in our previous work

Table 1 Structural parameters obtained by Guinier analysis on the SAXS data

Sample	$I(0)/c/cm^{-1}$	R_g/nm	Number of dimer in cluster[a]
BLG 6.7 HEPES	0.0248	2.33	Monodisperse dimer
BLG 3.4 YCl_3 10 mM	0.0321	3.08	1.3
BLG 6.7 YCl_3 10 mM	0.0432	3.36	1.7
BLG 13.4 YCl_3 10 mM	0.0663	3.67	2.7
BLG 27.0 YCl_3 10 mM	0.1126	4.90	4.5

[a] These values are calculated by dividing $I(0)$ by 0.0248, $I(0)$ of the pure dimer solution.

(PDB 3PH5)[16] with 2, 3 and 4 dimers, as shown in Fig. 3. The clusters are created as follows: we first choose one dimer from the lattice and set its center of mass to be the origin. Then the positions of the other dimers within the clusters are chosen based on the shortest center-to-center distance to the origin. The experimental SAXS profile of BLG 13.4 mg mL^{-1} with 10 mM YCl_3 is used to compare with these cluster structures using CRYSOL.[44] The best fitting comes from the cluster with 3 dimers (consistent with Table 1) which reproduces the R_g and $I(0)$ quantitatively, and other features of the SAXS profile qualitatively, such as the position of the kink, the shoulder for dimer, the minimum and maximum from the form factor as indicated by arrows in Fig. 3. Of course, this simple structural model of the cluster cannot reproduce all the details since the size and structure of the clusters will have a distribution. In spite of these difficulties, it is plausible to assume that the protein clusters have a "pre-crystalline" structure (Fig. 3) similar to those created from the crystal structure.

Having elaborated on the formation of small clusters with dimers as the principle building blocks from Fig. 2 and 3, we now review the full hierarchic structure of protein assembly in solution. Fig. 4 shows the SAXS profile of a sample close to the phase boundary with BLG 67 mg mL^{-1} with 15 mM YCl_3. A pronounced maximum at $q = 2.2$ nm^{-1} is clearly visible. As mentioned above, the scattering peak at $q = 2.2$ nm^{-1} is attributed to the monomer–monomer (M–M) correlation within clusters, which cannot be reproduced from the crystal structures in Fig. 3. The shoulder at $q \sim 1.8$ nm^{-1} corresponds to the dimer scattering, and the shoulder at $q \sim 0.3$ nm^{-1} corresponds to the form factor of the protein cluster. The number of dimers within the clusters is estimated to be around 3–4 for this sample since the intensities continuously increase in the low q region and both the forward intensity and the radius of gyration cannot be precisely determined by the Guinier analysis. The continuous increase indicates that the clusters may further build up higher level structures with a larger length scale, which makes the M–M correlation more significant.

Crystal growth followed by *in situ* SAXS

We now present the results of crystal growth directly from the cluster solution by an *in situ* SAXS observation. Fig. 5 shows a series of SAXS profiles for BLG

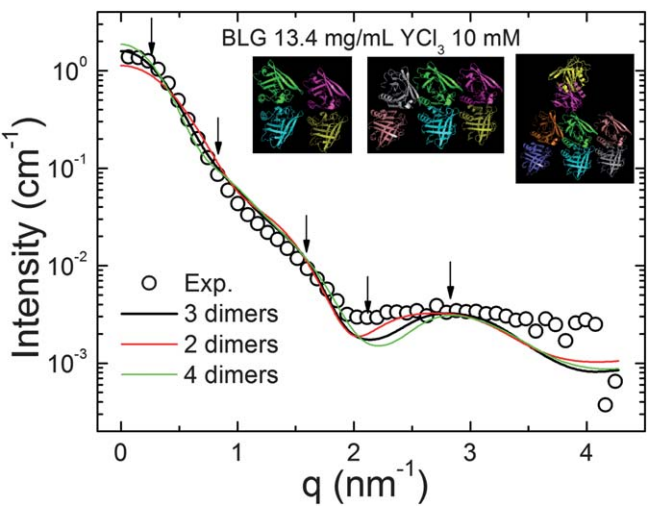

Fig. 3 Experimental SAXS data of protein clusters compared to those created using the crystal structure (PDB code of 3PH5). Only 10% of experimental data are plotted for clarity. The inset shows the structure of protein clusters created by the crystal structure with 2, 3 and 4 dimers.

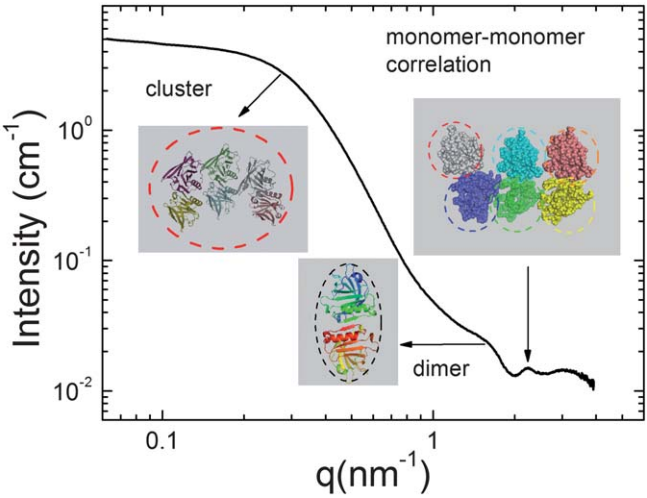

Fig. 4 Structural hierarchy of BLG solutions in re-entrant regime revealed by SAXS. The peak and shoulders in the SAXS curve at q equals $q = 2.2$, 1.8 and 0.3 nm^{-1} correspond to the monomer–monomer correlation, the form factor of a dimer, and the cluster, respectively.

67 mg mL^{-1} with YCl_3 15 mM. The transition temperature, T_h, is about 22 °C. At 25 °C (after preparation), the SAXS curve is the same as shown in Fig. 4. The sample is firstly quenched to 10 °C, then heated up stepwise to 22 °C, and then cooled down stepwise to 10 °C. The increment of heating and cooling is 1 °C and the time for equilibration is 15 min. After quenching to 10 °C, the scattering intensity in the low q region (region (1) in Fig. 5) increases as compared to that at 25 °C, indicating the formation of larger clusters. The solution becomes turbid. In the meantime, the M–M correlation peak (region (3) in Fig. 5) becomes weaker. In the intermediate q region (region (2) in Fig. 5), where the scattering is mainly contributed by the form factor of a dimer, no changes could be observed. The subsequent stepwise heating

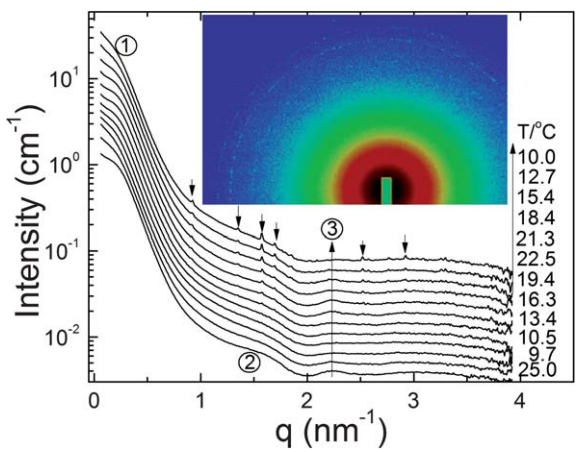

Fig. 5 SAXS curves at different temperatures during cooling. Crystallization occurs below 25 °C. The intensity of the maximum at $q = 2.2$ nm^{-1} decreases with lowering temperature, and the low q intensity increases steadily. Bragg peaks appearing in the intermediate q range have been indexed using the crystal structure (see text). The curves are shifted upward for clarify. The inset shows the 2D scattering pattern at 10 °C.

shows the inverse trend, *i.e.* with increasing temperature, the intensity in the low q region decreases, and the M–M correlation peak becomes stronger. The solution becomes clear again. This observation suggests that the formation of larger clusters is a reversible process, which is consistent with our previous measurement on T_h.[16] So far, no Bragg peaks could be observed. Upon further lowering the temperature, one observes the following significant changes: (i) Bragg peaks appear, indicating the crystallization of proteins, indicating that protein crystals can grow directly from the clusters. The inset 2D pattern of the final state shows sharp circles indicating the isotropic distribution of the orientation in the solution. We have recently determined the structure of crystals grown under similar conditions in the phase diagram. All crystals have an orthorhombic structure with space group $P2_12_12_1$ and contain one dimer in their asymmetric unit (PDB code 3PH5 and 3PH6), which can be used to index the Bragg peaks in the SAXS profiles.[16] For example, the first three Bragg peaks at $q = 0.924\ \mathrm{nm}^{-1}$, $q = 1.348\ \mathrm{nm}^{-1}$ and $q = 1.571\ \mathrm{nm}^{-1}$ can be assigned as (002), (012) and (100) of the crystal structure, respectively. This means that although M–M correlation is visible in the cluster phase, the dimer represents the building block of the crystals. (ii) In the low q region, the scattering intensity increases with decreasing temperature, indicating the formation of larger objects, *i.e.* clusters and crystals. (iii) No change is observed in region (2) except the appearance of Bragg peaks. (iv) the M–M correlation peak at $q = 2.2\ \mathrm{nm}^{-1}$ continuously decreases in intensity as the temperature decreases. The temperature dependence of the M–M correlation peak may be related to the flexibility of the monomers within the cluster.

The crystal growth of similar samples has been followed using an optical microscope.[16] It shows that after quenching the solution below T_h, the solution becomes turbid instantly. After some time (hours) the dense liquid phase appears, then crystal growth appears to start from the dilute phase. Over time, the crystal growth consumes the material and leads to the dissolution of the droplets. Our observations of crystal growth in the presence of YCl_3 suggest that the most common way is the growth from the dilute phase directly leading to the coexistence of a dense liquid phase and crystals. From the results of the SAXS measurements, we now know that even in the dilute phase, proteins form clusters (Fig. 2, 3) and crystals grow directly from such protein clusters (Fig. 5). These observations suggest that the protein clusters undergo two competing pathways during cooling, *i.e.* either forming a dense liquid phase or reordering into a crystal, which can modify the pathways of crystal growth as to be discussed below.

Discussion

Pathway of the two-step crystallization

We first discuss the possible pathway of protein crystallization during a two-step growth procedure, which leads to the discussion of the role of protein clusters and a metastable LLPS. A two-step mechanism following a metastable LLPS suggests that the nucleation occurs within the dense liquid phase instead of the dilute phase.[3,32,48] Because the surface free energy at the interface between the crystal and the solution is significantly higher than at the interface between the crystal and the dense liquid, the barrier for nucleation of crystals from the solution would be much higher. This would lead to much slower nucleation of crystals directly from the solution than inside the clusters. While indeed some experimental observations are consistent with this prediction, *e.g.* for glucose isomerase and hemoglobin,[10,49] other observations however, indicate that crystallization prefers to start in the dilute phase.[50] While this has often been attributed to the high viscosity or gelation of the dense liquid phase, a clear understanding of their exact role is still missing.

In our system, we observe in addition to the metastable LLPS a non-negligible effect of clustering. Although the mechanism of clustering observed here may not

be the same as in the concentrated protein solutions,[26–28] it can be assumed to play a similar role on protein crystallization. The protein clusters observed show interesting features with respect to crystallization, which make them suitable as building blocks for crystal formation. Firstly, the clusters have a pre-crystalline structure, since the cation binding sites represent specific interaction patches. These imprint a favorable structure, as implied by the SAXS scattering curves of the clusters being simulated well using the clusters created from the crystal structure. Secondly, clusters seem to have an internal flexibility, as evidenced by the pronounced M–M correlation peak in larger clusters. This flexibility enables local reorientation within the clusters, rendering the pre-crystalline structure of clusters to be an ideal building block for crystallization, since the enthalpy cost of nucleation *via* local reorientation is much lower than for a hypothetical nucleation directly from a dimer in solution. Thirdly, the cluster size and its distribution vary throughout the phase diagram; in particular, the clusters grow when the solution conditions approach the coexistence region. Large pre-crystalline clusters can serve as the precursor (stabilized nuclei) for further crystal growth, as recently observed in various systems.[22,29–31]

Inspired by our observations, we propose an alternative pathway of protein crystallization in Fig. 6. The monodisperse dimeric proteins in solution form clusters upon adding YCl_3 due to the bridging effect of cations (Fig. 6a), as indicated by SAXS measurements (Fig. 2). Upon lowering the temperature across the phase boundary (below T_h), LLPS takes place, leaving a dense phase and a dilute phase with clusters (Fig. 6b). In solution, clusters can move freely and restructure themselves internally. Furthermore, clusters can assemble to larger clusters with a pre-crystalline structure. As a next step, internal reorientation turns the clusters to stable precursor nuclei (Fig. 6c). Because the dense liquid phase is metastable with respect to the crystalline phase, the crystals grow and consume the protein molecules from the dense liquid phase which dissolves and disappears (Fig. 6d). This pathway differs from the previous view of the two-step mechanism involving a metastable LLPS by the phases where nucleation starts. While the previous view predicts the crystals to be nucleated within the dense liquid phase, our observations suggest that both condensed phases can arise from the protein clusters in the protein-poor phase.

The role of clusters and dense liquid phase in protein crystallization

The possible pathway of protein crystallization discussed above (Fig. 6) leads to a further discussion of the role of the protein clusters and the metastable dense liquid phase during the two-step crystallization procedure. In our system, both the dense liquid phase and crystals grow directly from the protein clusters (Fig. 6): we thus speculate that while the protein clusters enhance the nucleation, the dense liquid phase may play a role in optimizing the conditions for crystallization. Some clues

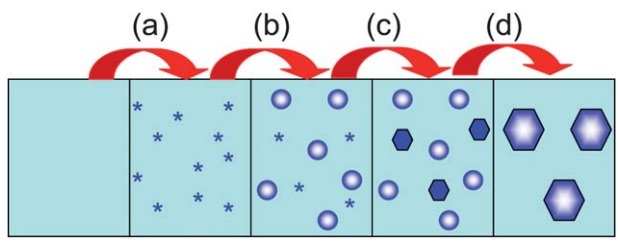

Fig. 6 Schematic illustration of the phase transitions (LLPS and crystallization) from protein clusters (see text for details). (a) Upon addition of YCl_3, protein clusters formed *via* ion bridging. (b) Lowering temperature to $T < T_h$ and crossing the phase boundary, the clusters first aggregate with a liquid-like structure. (c) After the induction time protein crystals grow directly from the clusters. (d) With time increasing, while the crystals grow steadily, the dense liquid phase dissolves and disappears.

arising from the detailed studies on lysozyme in solution support this explanation: Pan *et al.* have shown that the protein clusters exist between the solubility line and the LLPS in the phase diagram,[23] which coincides with the optimum conditions for protein crystallization predicted using the second virial coefficient as predictor by Vliegenthart and Lekkerkerker.[51] Our observations from *in situ* SAXS (Fig. 5) indicate that nucleation is possible from protein clusters, and microscope observations indicate that crystals grow from the dilute solution instead of the dense liquid phase. Employing this picture, the metastable LLPS occurs as a helpful side effect at these thermodynamic conditions, since the LLPS indicates suitable interaction conditions and ensures that the dilute protein solutions stay at a defined concentration. Furthermore, the dense liquid phase potentially acts as a reservoir feeding the dilute phase with protein molecules and thus stabilizing the thermodynamic conditions during the ongoing crystal growth.

Connection of two-step crystallization to the phase diagram

In addition to the two-step nucleation mechanism,[3,32,48] we have discussed an alternative pathway of two-step crystallization *via* cluster precursors. Considering the complexity of protein phase behavior, we expect both to be relevant under suitable conditions, which we discuss with the schematic phase diagram in Fig. 7. Vliegenthart and Lekkerkerker[51] elaborated on earlier findings of George and Wilson[52] to provide a criterion for optimum crystallization conditions based on the reduced second virial coefficient. Besides the metastable critical point enhancing nucleation *via* critical density fluctuations,[32] the identified optimum conditions in terms of the reduced second virial coefficient correspond to a temperature window, *i.e.* both the dilute and the dense phase (indicated with grey shading). Considering the lower nucleation barrier in the dense phase, the optimum conditions for protein crystallization should be represented by the dense solution as predicted in the two-step nucleation mechanism,[3,32,48] although in practice, such condition often leads to polycrystallites instead of single crystals, which are needed for the crystallographic study.

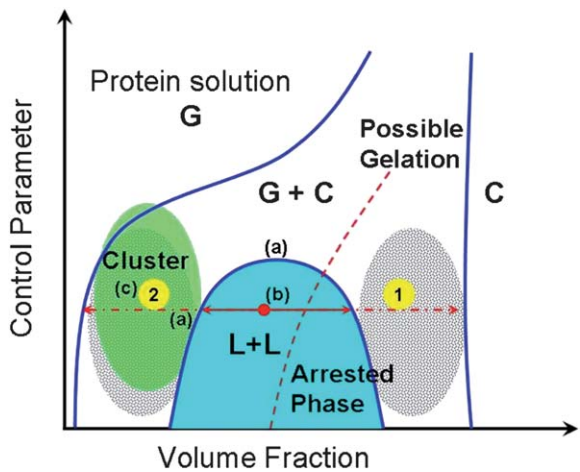

Fig. 7 Schematic phase diagram of protein solutions. The region of clustering[35] is marked in green and the grey shaded area corresponds to the optimum condition for crystallization.[46] Two-step nucleation mechanism predicts that protein solutions located within LLPS region (red point) undergo a LLPS in the first step and then nucleation occurs within the dense liquid phase (along pathway (1)). An alternative pathway proposed in this work follows pathway (2): after LLPS, protein clusters in the dilute phase with the optimized thermodynamic conditions for crystallization can initialize crystal growth with a reduced energy barrier. Note that the exact conditions or the location of the gelation line will vary with the specific system.

These conditions, however, are not necessarily accessible for all systems, since arrested phase behaviour may render the high-concentration coexistent phase a gel instead of an equilibrium solution.[4,19–21,53] While for isotropic interaction potentials, a metastable LLPS is generally arrested,[4,20] anisotropic and patchy spheres have been shown to phase separate completely into a metastable "gas" and "empty liquid" phase.[21,53] At least for the arrested systems, however, protein clusters in the dilute phase could step in as precursors, providing an alternative (even dominate) way of crystallization. Interestingly, the cluster phase is expected to overlap with the optimum crystallization conditions based on the second virial coefficient. The proposed mechanism *via* cluster precursors could thus explain the frequent observation of crystal growing in the dilute phase.

Finally, since there appear to be different views in the literature regarding the terminology of the "two-step" crystallization process, we would like to comment on the different scenarios. In protein solutions, depending on the position in the phase diagram (labeled in Fig. 7 as a, b and c), at least three scenarios for the intermediate state could be realized: (a) a transient density fluctuation or cluster near the critical point or near the G–L binodal, this scenario is supported by simulations and theoretical studies;[32,54,55] (b) a metastable co-existence of liquid phases, which has been observed in several protein systems;[10,49] and (c) protein clusters as shown in our system and other proteins in solution.[27,28] In real protein solutions, these mechanisms might also occur at the same time, rendering a deeper understanding of protein crystallization an interesting challenge for future research.

Conclusions and outlook

In summary, our SAXS study of the structure and crystallization from bovine β-lactoglobulin (BLG) solutions in the presence of YCl_3 leads to the following conclusions. First, near the phase boundary, the balance between the bridging effect of yttrium cations and effective charge inversion of proteins leads to the formation of small protein clusters which have a pre-crystalline structure. Second, *in situ* SAXS measurements on the crystallization together with previous observation by optical microscopy indicate that these protein clusters can form both a dense liquid phase and crystals. The crystals occur later compared to the dense liquid phase due to the higher energy barrier of nucleation. This procedure leads to the competing coexistence of these two structures. Third, due to the metastable character of the dense liquid phase, it re-dissolves during the ongoing crystal growth and disappears in the post-crystal growth stage. The observed two-step crystal growth procedure suggests that clusters play the main mechanistic role for nucleation, while LLPS allows the system to access and stabilize suitable conditions.

Acknowledgements

We acknowledge the useful discussion with Dr Zocher (IFIB, Universität Tübingen) and Dr T. Narayanan (ESRF, Grenoble, France). We acknowledge financial support from Deutsche Forschungsgemeinschaft (DFG) and beamtime allocation at ESRF, Grenoble, France.

References

1 J. D. Gunton, A. Shiryayev and D. L. Pagan, *Protein Condensation—kinetic pathways to crystallization and disease*, Cambridge University Press, New York, 2007.
2 P. G. Vekilov, *Cryst. Growth Des.*, 2004, **4**, 671–685.
3 P. G. Vekilov, *Nanoscale*, 2010, **2**, 2346–2357.
4 P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino and D. A. Weitz, *Nature*, 2008, **453**, 499–503.
5 M. Delaye, J. I. Clark and G. B. Benedek, *Biochem. Biophys. Res. Commun.*, 1981, **100**, 908–914.

6 M. L. Broide, C. R. Berland, J. Pande, O. Ogun and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 5660–5664.
7 M. Muschol and F. Rosenberger, *J. Chem. Phys.*, 1997, **107**, 1953–1962.
8 O. Galkin and P. G. Vekilov, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6277–6281.
9 O. Annunziata, N. Asherie, A. Lomakin, J. Pande, O. Ogun and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14165–14170.
10 O. Galkin, K. Chen, R. L. Nagel, R. E. Hirsch and P. G. Vekilov, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8479–8483.
11 S. Grouazel, J. Perez, J. P. Astier, F. Bonneté and S. Veesler, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 1560–1563.
12 O. Annunziata, O. Ogun and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 970–974.
13 Q. Chen, P. G. Vekilov, R. L. Nagel and R. E. Hirsch, *Biophys. J.*, 2004, **86**, 1702–1712.
14 N. Dorsaz, G. M. Thurston, A. Stradner, P. Schurtenberger and G. Foffi, *Soft Matter*, 2011, **7**, 1763–1776.
15 Y. Wang, A. Lomakin, R. F. Latypov and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 16606–16611.
16 F. Zhang, G. Zocher, A. Sauter, T. Stehle and F. Schreiber, *J. Appl. Crystallogr.*, 2011, **44**, 755–762.
17 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.
18 F. Cardinaux, T. Gibaud, A. Stradner and P. Schurtenberger, *Phys. Rev. Lett.*, 2007, **99**, 118301.
19 T. Gibaud, F. Cardinaux, J. Bergenholtz, A. Stradner and P. Schurtenberger, *Soft Matter*, 2011, **7**, 857–860.
20 T. Gibaud and P. Schurtenberger, *J. Phys.: Condens. Matter*, 2009, **21**, 322201.
21 F. Sciortino and E. Zaccarelli, *Curr. Opin. Solid State Mater. Sci.*, 2011, **15**, 246–253.
22 A. Navrotsky, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12096–12101.
23 Y. Liu, E. Fratini, P. Baglioni, W. R. Chen and S. H. Chen, *Phys. Rev. Lett.*, 2005, **95**, 118102.
24 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.
25 A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf and P. Schurtenberger, *Nature*, 2004, **432**, 492–495.
26 O. Gliko, W. Pan, P. Katsonis, N. Neumaier, O. Galkin, S. Weinkauf and P. G. Vekilov, *J. Phys. Chem. B*, 2007, **111**, 3106–3114.
27 W. Pan, O. Galkin, L. Filobelo, R. L. Nagel and P. G. Vekilov, *Biophys. J.*, 2007, **92**, 267–277.
28 W. Pan, P. G. Vekilov and V. Lubchenko, *J. Phys. Chem. B*, 2010, **114**, 7620–7630.
29 J. F. Banfield, S. A. Welch, H. Zhang, T. T. Ebert and R. Lee Penn, *Science*, 2000, **289**, 751–754.
30 G. Furrer, B. L. Phillips, K.-U. Ulrich, R. Pöthig and W. H. Casey, *Science*, 2002, **297**, 2245–2247.
31 S. Mintova, N. H. Olson, V. Valtchev and T. Bein, *Science*, 1999, **283**, 958–960.
32 P. R. ten Wolde and D. Frenkel, *Science*, 1997, **277**, 1975–1978.
33 D. Erdemir, A. Y. Lee and A. S. Myerson, *Acc. Chem. Res.*, 2009, **42**, 621–629.
34 P. E. Bonnett, K. J. Carpenter, S. Dawson and R. J. Davey, *Chem. Commun.*, 2003, 698–699.
35 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.
36 L. Ianeselli, F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, R. A. Martin, S. Callow, S. Prévost and F. Schreiber, *J. Phys. Chem. B*, 2010, **114**, 3776–3783.
37 F. Zhang, S. Weggler, M. Ziller, L. Ianeselli, B. S. Heck, A. Hildebrandt, O. Kohlbacher, M. W. A. Skoda, R. M. J. Jacobs and F. Schreiber, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 3450–3457.
38 *Handbook of Biochemistry: Selected Data for Molecular Biology*, ed. H. A. Sober, The Chemical Rubber Co., Cleveland, OH, 1970.
39 M. Sztucki, T. Narayanan, G. Belina, A. Moussaïd, F. Pignon and H. Hoekstra, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 051504.
40 T. Narayanan, in *Soft Matter: Characterization*, ed. R. Borsali and R. Pecora, Springer, Berlin-Heidelberg, 2008, vol. II, ch. 17, pp. 899–952.
41 U. M. Elofsson, M. A. Paulsson and T. Arnebrant, *Langmuir*, 1997, **13**, 1695–1700.
42 M. Gottschalk, H. Nilsson, H. Roos and B. Halle, *Protein Sci.*, 2003, **12**, 2404–2411.

43 K. Vogtt, N. Javid, E. Alvarez, J. Sefcik and M.-C. Bellissent-Funel, *Soft Matter*, 2011, **7**, 3906–3914.
44 D. I. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.
45 C. Moitzi, L. Donato, C. Schmitt, L. Bovetto, G. Gillies and A. Stradner, *Food Hydrocolloids*, 2011, **25**, 1766–1774.
46 O. Glatter and O. Kratky, *Small angle X-ray scattering*, Academic Press London, 1982.
47 J. Groenewold and W. K. Kegel, *J. Phys. Chem. B*, 2001, **105**, 11702–11709.
48 V. Talanquer and D. W. Oxtoby, *J. Chem. Phys.*, 1998, **109**, 223–227.
49 D. Vivares, E. W. Kaler and A. M. Lenhoff, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **61**, 819–825.
50 Y. Liu, X. Wang and C. B. Ching, *Cryst. Growth Des.*, 2010, **10**, 548–558.
51 G. A. Vliegenthart and H. N. W. Lekkerkerker, *J. Chem. Phys.*, 2000, **112**, 5364–5469.
52 A. George and W. W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1994, **50**, 361–365.
53 E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli and F. Sciortino, *Phys. Rev. Lett.*, 2006, **97**, 168301.
54 D. W. Oxtoby, *Acc. Chem. Res.*, 1998, **31**, 91–97.
55 J. F. Lutsko, *Adv. Chem. Phys.*, 2012, **151**, arXiv:1104.3413v1101.