PAPER

On the question of two-step nucleation in protein crystallization†

Andrea Sauter,[a] Felix Roosen-Runge,[b] Fajun Zhang,*[a] Gudrun Lotze,[c] Artem Feoktystov,[d] Robert M. J. Jacobs[e] and Frank Schreiber[a]

We report a real-time study on protein crystallization in the presence of multivalent salts using small angle X-ray scattering (SAXS) and optical microscopy, focusing particularly on the nucleation mechanism as well as on the role of the metastable intermediate phase (MIP). Using bovine beta-lactoglobulin as a model system in the presence of the divalent salt $CdCl_2$, we have monitored the early stage of crystallization kinetics which demonstrates a two-step nucleation mechanism: protein aggregates form a MIP, which is followed by the nucleation of crystals within the MIP. Here we focus on characterizing and tuning the structure of the MIP using salt and the related effects on the two-step nucleation kinetics. The results suggest that increasing the salt concentration near the transition zone *pseudo-c** enhances the energy barrier for both MIPs and crystal nucleation, leading to slow growth. The structural evolution of the MIP and its effect on subsequent nucleation is discussed based on the growth kinetics. The observed kinetics can be well described, using a rate-equation model based on a clear physical two-step picture. This real-time study not only provides evidence for a two-step nucleation process for protein crystallization, but also elucidates the role and the structural signature of the MIPs in the nonclassical process of protein crystallization.

1.   Introduction

Studies of the early stage of nucleation in various systems have revealed new features which cannot be explained using the classical nucleation theory.[1–6] A large body of experimental results supported by theory and simulations suggest

[a]Institut für Angewandte Physik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany. E-mail: fajun.zhang@uni-tuebingen.de; Fax: +49 7071 29 5110; Tel: +49 7071 29 78670

[b]Institut Laue-Langevin, 6, rue Jules Horowitz, 38042 Grenoble Cedex 9, France

[c]European Synchrotron Radiation Facility, 71 avenue des Martyrs, 38043 Grenoble Cedex 9, France

[d]Jülich Centre for Neutron Science JCNS, Forschungszentrum Jülich GmbH, Outstation at MLZ, Lichtenbergstrasse 1, 85747 Garching, Germany

[e]Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford, OX1 3TA, UK

that a metastable intermediate phase (MIP) exists before the final crystal structure is formed,[7–21] *i.e.* the solutes in the supersaturated solution form, in a first step, either small clusters or a macroscopic dense liquid phase. In the second step, the nucleation occurs within the MIPs. This two-step nucleation mechanism was originally proposed by ten Wolde and Frenkel for crystallization of a colloidal system with short-range attraction and near the critical point of the metastable liquid–liquid coexistence line.[14] The two-step nucleation mechanism can be considered as an example of Ostwald's step rule in the microscopic world.[22] Since then, this concept has been widely used under various conditions.[1–6,17–20,23]

While the two-step mechanism seems plausible for certain experiments, direct observation of such a process is not easy. Recently, direct visualization of the crystallization kinetics and pathways of nucleation in colloidal crystallization became possible. This provided detailed information on the MIP and the transition in colloidal suspensions. Colloidal systems exhibit similar phase behavior to atomic and molecular systems, and their large particle sizes enable visualization on a single-particle level. Using this technique, Tan *et al.* studied the liquid–solid phase transition and observed the formation of a metastable precursor under their experimental conditions, regardless of the final state and the interaction potential.[24] Peng *et al.* studied the kinetics of a solid–solid phase transition using single-particle resolution video microscopy. They observed that the transition between two different solid states occurs *via* a two-step diffusive nucleation pathway involving liquid nuclei.[25] This pathway is favored in comparison to one-step nucleation, because the energy of the solid–liquid interface is lower than that between the solid phases.

While these experimental observations on colloidal systems demonstrate that two-step nucleation follows Ostwald's step rule for simple liquids,[16,19,26,27] the application of this concept to other systems, in particular protein crystallization, is still challenging. The dimensions of proteins on the nanometer scale limit the applicability of optical methods for the study of MIP formation. Because of the larger size and slow dynamics of colloids, the microstructural arrangement of colloidal particles typically relaxes in a time scale of seconds, which leads to various non-equilibrium phenomena in these systems. Moreover, the interaction potentials in these colloidal systems are isotropic, whereas the effective protein–protein interactions are often non-isotropic, involving hydrophobic and electrostatic patches as well as ion-bridges. The quantitative description remains poorly understood.[28–30]

For these systems, in which direct visualization is not possible, another method has to be developed to characterize the MIP and the nucleation and growth kinetics. Here, we argue that two-step nucleation can be distinguished from the classical one-step process by following the overall crystallization kinetics. When a MIP exists, particular care should be paid to distinguishing the sequential and parallel pathways. The sequential pathway corresponds to real two-step nucleation, in which the crystals nucleate from the MIP. The parallel pathway consists of the parallel formation of the MIP and crystals, in two one-step nucleation events from the liquid. The different pathways are illustrated in Fig. 1.

For a one-step nucleation and growth mechanism with or without MIPs, nucleation and growth are mainly determined by the supersaturated initial solution. The overall growth kinetics is expected to show an increase first and then to saturate later. In contrast, for a two-step mechanism, nucleation may
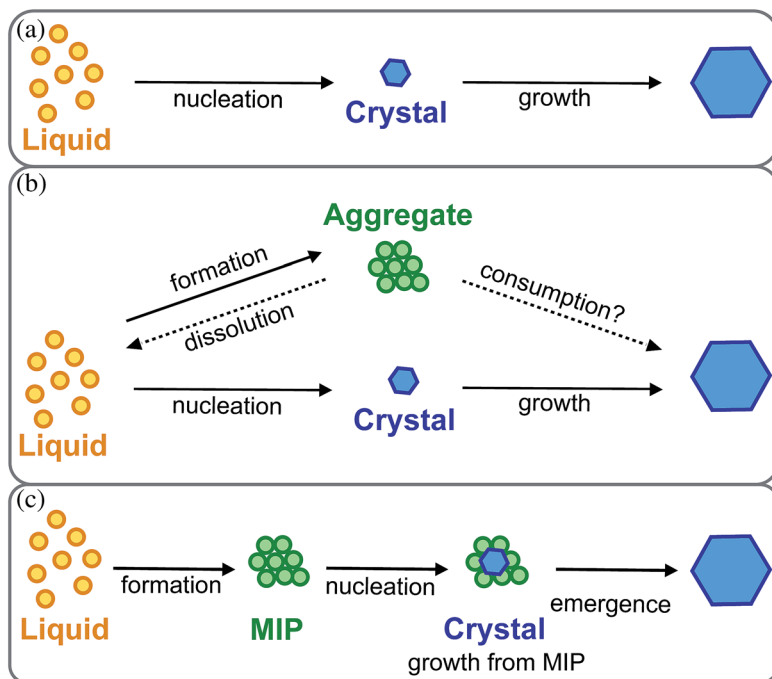
Fig. 1 Illustrations of different pathways for crystal nucleation and growth. (a) Classical one-step pathway. (b) Parallel process of aggregate formation and crystal nucleation. (c) Multi-step process with a precursor (MIP) for nucleation.

occur in both the MIPs and the initial supersaturated solution, and the nucleation rates are different due to the different interfacial energies. Furthermore, the subsequent crystal growth rates are different too. Therefore, a characteristic two-step growth kinetics is expected, dominated by the MIP and the initial solution, respectively. More details will be discussed based on the rate equation model in this paper. Indeed, in our recent study using time-resolved SAXS and optical microscopy, we identified the kinetic signature of two-step nucleation in the crystallization of the protein β-lactoglobulin (BLG) in the presence of $CdCl_2$.[31] The first step of kinetics is dominated by the two-step nucleation and growth within the MIP, *i.e.* protein aggregates form as a MIP, followed by the nucleation of crystals within the MIP. During this period, the number of crystals increases with time, but the growth rate is slow due to the low mobility of surrounding aggregates. The second step of kinetics is dominated by the growth in the dilute phase: the consumption of the MIP leads to the exposure of crystals to free molecules in the dilute phase, which accelerates the crystal growth. In this work, we focus on the effect of the structural property of the MIP on the nucleation and growth of the two-step nucleation mechanism.

2. Experimental

2.1 Materials and sample preparation

The globular protein β-lactoglobulin (BLG) from bovine milk (product no. L3908), $CdCl_2$ (202908) was purchased from SIGMA-ALDRICH. For sample preparation,

appropriate amounts of salt stock solution, millipore water and protein stock solution were mixed. Stock solutions were prepared by dissolving the salt or protein powder in deionized (18.2 MΩ) and degassed millipore water. The protein concentration of stock solutions was determined by UV absorption measurements using an extinction coefficient of 0.96 l g^{-1} cm^{-1}, at a wavelength of 278 nm.[32] All samples in this work were prepared without additional buffer, since buffers can affect the phase behavior of proteins and the solubility of salts. The pH of the solutions was monitored using a Seven Easy pH instrument from Mettler Toledo. The pH values for all experimental conditions were above the pI = 5.2 of BLG.[33] Therefore, cation binding was the main driving force of charge inversion instead of pH.[33,34,41] All experiments were performed at a room temperature of 293 ± 1 K.

2.2 Optical microscopy

Time-dependent protein crystallization was followed by transmission optical microscopy using an AXIOSCOPE.A1 from ZEISS. The protein stock solution was filtered (pore size 100 μm) in advance. Samples were prepared using a micro-batch setup with two hydrophobically coated glass slides sealed by silicone (sample thickness ≈ 250–300 μm). Images were taken by an integrated AXIOCAM ICC5 camera.

2.3 Small-angle X-ray and neutron scattering (SAXS and SANS)

Small angle X-ray scattering (SAXS) measurements were performed at the ESRF, Grenoble, France at the beamline ID02. Different energies of 16 047 eV and 12 460 eV were used for two different beamtimes, and the sample-to-detector distance was 2 m with an accessed q-range of 0.06 to 4.3 nm^{-1} or 0.04 to 3.9 nm^{-1}, respectively. *Ex situ* measurements were performed using a flow capillary cell. For real-time measurements, samples were measured using quartz capillaries in a vertical capillary holder; these were quickly loaded and transferred to the sample station after sample preparation. Measurements started about 2–3 min after mixing and took place every couple of minutes during the entire crystallization process. The beam position in the sample was shifted after each exposure (duration 0.05 s) to avoid radiation damage. For further details on the beamline, calibration and data collection, see ref. 35.

SANS measurements were carried out at KWS1 at FRMII, Munich, Germany. The applied sample-to-detector distances were 1.5 and 8 m, which covers a q-range from 0.04 nm^{-1} to 3.1 nm^{-1} at a wavelength of 7 Å (Δλ/λ = 10%). Protein-salt solutions in D$_2$O were filled in rectangular quartz cells with a path-length of 2 mm. The beam size on the sample was 6 mm × 12 mm. Plexiglas was used as a secondary standard to calibrate the absolute scattering intensity. The data correction and absolute intensity calibration were obtained using QtiKWS software.[36]

3. Results

3.1 Experimental phase diagram of BLG with CdCl$_2$ or ZnCl$_2$

We first describe the experimental phase diagram of our system which provides the basis for the following kinetic studies on protein crystallization. Our studies of globular proteins in solutions containing multivalent metal ions have revealed

complex phase behavior including reentrant condensation (RC), metastable liquid–liquid phase separation (LLPS) and crystallization.[34,37–43] A similar experimental phase diagram, like the one shown in Fig. 2, has been observed for several proteins in solution in the presence of trivalent metal ions. The physical mechanism of this RC behavior is due to the effective charge inversion of proteins and a cation-mediated attraction, presumably by intermolecular bridges of multivalent cations.[38,44] With an isoelectric point below 7, the proteins used in our work are acidic at a neutral pH. At low salt concentrations, proteins carry negative net charges, and the electrostatic repulsion stabilizes the solution. When trivalent metal ions are added into the solution, the binding of metal ions to the carboxyl groups on the protein surface reduces the effective net charge. Above a certain salt concentration c^*, electrostatic repulsion is not strong enough to balance the attractive potential, and samples phase separate and become turbid ("regime II"). The interesting observation for this system is that with a further increase in salt concentration, the continuous binding of metal ions to the protein surface leads to a charge inversion, which again establishes the long-range electrostatic repulsion. Therefore above a second boundary (experimentally rather broad), c^{**}, the solutions become completely clear again. The charge inversion and the effective attraction mediated *via* multivalent metal ions have been further investigated by experiments, simulations and theoretical studies.[41,42,44–46]

Previous studies on β-lactoglobulin (BLG) systems with divalent salts ZnCl$_2$ and CdCl$_2$ showed a similar experimental phase behavior (Fig. 2).[31,47] For these divalent salts, the samples above a certain salt concentration become gradually less turbid, but not completely clear again, and this transition zone is denoted as *pseudo-c^{**}*. Both boundaries induced by CdCl$_2$ and ZnCl$_2$ are remarkably similar as shown in Fig. 2. In comparison to the trivalent salt YCl$_3$, both transitions are
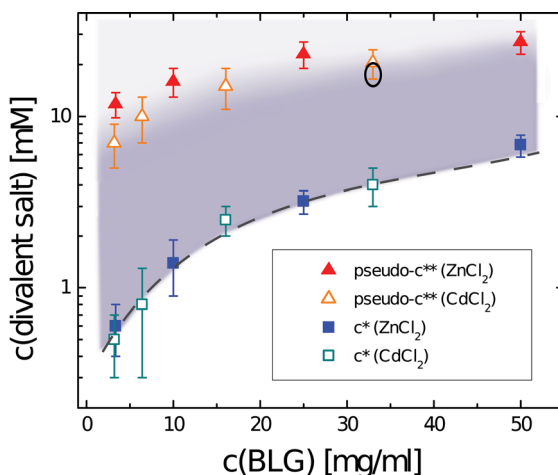

Fig. 2 Experimental phase diagram of BLG with ZnCl$_2$ and CdCl$_2$.[31,47] Both salts share a similar c^*, below and above which the protein solutions are clear and turbid, respectively. With a further increase in salt concentration, samples become gradually less turbid, but not completely clear again, and this transition zone is denoted as *pseudo-c^{**}*. The black ellipse marks the conditions mainly used in this work (33 mg ml^{-1} BLG with 15–20 mM CdCl$_2$).

shifted towards higher salt concentrations.[47] Although the reentrant effect is not complete, a charge inversion with increasing divalent salt concentration has been observed in both cases (ESI, Fig. S1†). Note that the phase behavior shown in Fig. 2 was only observed for BLG, not for bovine or human serum albumin (data not shown), which suggests a more specific interaction between these divalent ions and BLG.[47]

We emphasize that the observed protein condensation is not caused by a change in the protein structure induced by $CdCl_2$, $ZnCl_2$ or other multivalent salts; this was demonstrated in previous work using Fourier transform infrared (FTIR) and circular dichroism spectroscopy for a broad protein and salt concentration range.[31,48] Both techniques indicate no significant change in the secondary structure of the protein. Moreover, the successful growth of high-quality crystals and fine structural analysis confirm that the proteins are still in their native state.[38,47]

3.2 Crystallization followed by optical microscopy

Crystallization of BLG in the presence of $CdCl_2$ and $ZnCl_2$ was observed within a wide range of the above shown phase diagram (Fig. 2).[31,47] Slightly below c^*, no indications of a multi-step process were observed: crystals nucleate directly from a solution and grow larger without any visible aggregation or LLPS. In contrast, in the transition zone of *pseudo-c^{**}*, protein aggregates quickly form after sample preparation and the solution observed by optical microscopy becomes more turbid. In the case of $ZnCl_2$, the aggregates further relax into a liquid-like network before crystallization starts.[47] During crystallization, the turbid solution becomes gradually clearer, indicating that the aggregates are consumed. If the aggregates are removed by centrifugation, the number of crystals or nucleation rate can be significantly reduced, as has also been shown in other protein systems.[20] The two different growth pathways have been observed for the same protein, BLG, with a trivalent salt, YCl_3.[38,47]

Within the *pseudo-c^{**}* region, the crystallization of BLG in the presence of $CdCl_2$ follows a two-step pathway *via* aggregates.[31] Here, we study a series of samples with a fixed protein concentration but varying salt concentrations. Variation of the salt concentration changes the strength of the attractive potential, and therefore the driving force of formation, of both MIP and nucleation. Crossing *pseudo-c^{**}* from low to high salt concentrations, the driving force is reduced. Fig. 3 shows a time series of images of crystallization, for samples with 33 mg ml^{-1} BLG and $CdCl_2$ concentrations of 17, 18, 19 and 20 mM. Below 17 mM, the sample solution is deep in regime II, where massive amorphous aggregates form and precipitate quickly from the solution due to the strong attraction. Although crystallization can still be observed, only a small fraction of the amorphous aggregates can turn into the crystalline phase. Above 20 mM, solutions become nearly clear, and only small aggregates are visible under a microscope. Crystallization in these solutions becomes extremely slow, or they do not crystallize at all.

From the images shown in Fig. 3, one can see that with 17 mM salt, large aggregates are still formed, but the nucleation rates are also high. We emphasize that nearly all crystals are associated with the network. Hardly any crystals form in the dilute phase. In contrast to the high nucleation rate, the crystal growth period is short. After about 2 h, visible change can no longer be observed, and the
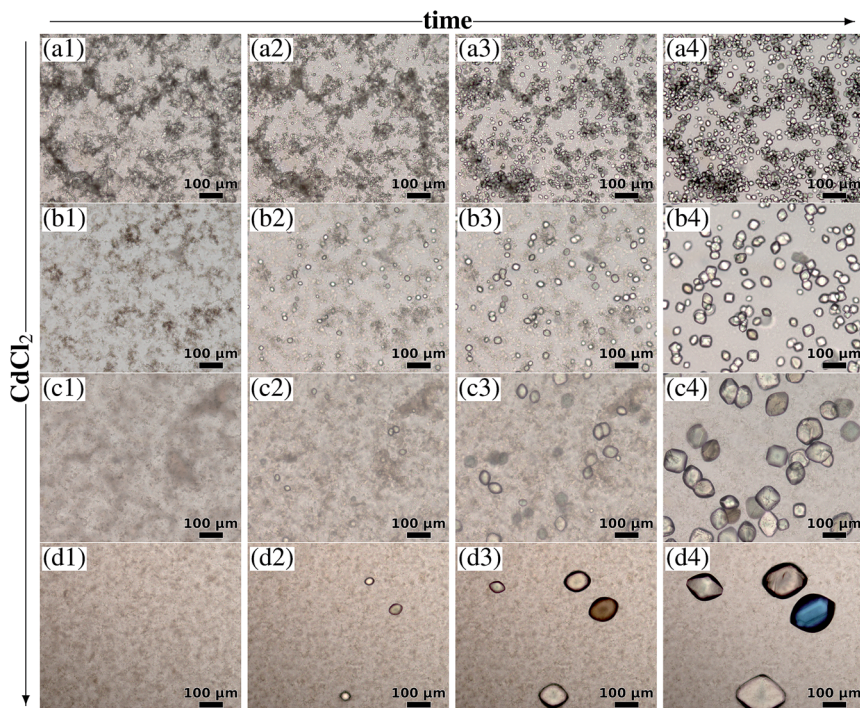
Fig. 3 Image series of protein crystallization for samples with a BLG concentration of 33 mg ml^{-1} and CdCl$_2$ at different times: (a) 17 mM, (b) 18 mM, (c) 19 mM and (d) 20 mM (increasingly less turbid). (a1) 0, (a2) 50 min (a3) 75 min (a4) 100 min after preparation. (b1) 0 (b2) 75 min (b3) 100 min (b4) 3 h. (c1) 0, (c2) 100 min, (c3) 3 h, (c4) 6 h. (d1) 0, (d2) 3 h, (d3) 6 h and (d4) 10 h. Videos of the crystallization process in (a) and (d) can be found in the ESI.†

resulting system contains a large number of small crystals, while most of the network of amorphous aggregates has turned into the crystalline phase. When the salt concentration is increased by just 1 mM, to 18 mM, the overall phenomenological picture changes dramatically. Large aggregates are still visible, but are not well connected to each other. The number of crystals is reduced, but the final crystals are bigger. When the salt concentration is further increased, protein aggregates become smaller and the number of crystals is further reduced, but the crystal size is larger. In the end, the MIP is consumed by crystal growth and the solutions become clear. The average number of crystals, normalized for an area of 1 mm^2, as a function of time is plotted in Fig. 4 for three conditions. The number of crystals initially increases with time, then saturates. The nucleation rates in the early stage are obtained from the slope of a linear fit; they are 1.44 ± 0.08, 0.32 ± 0.03 and 0.08 ± 0.01 min^{-1} for 18 mM, 19 mM and 20 mM CdCl$_2$, respectively. This decrease in the nucleation rate is expected, as the driving force is reduced with increasing salt concentration.

3.3 Structure of MIP revealed by SAXS and SANS

Due to the limited resolution of optical microscopy, SAXS and SANS were used to study the structure and the role of the MIP on the crystallization process. Firstly,
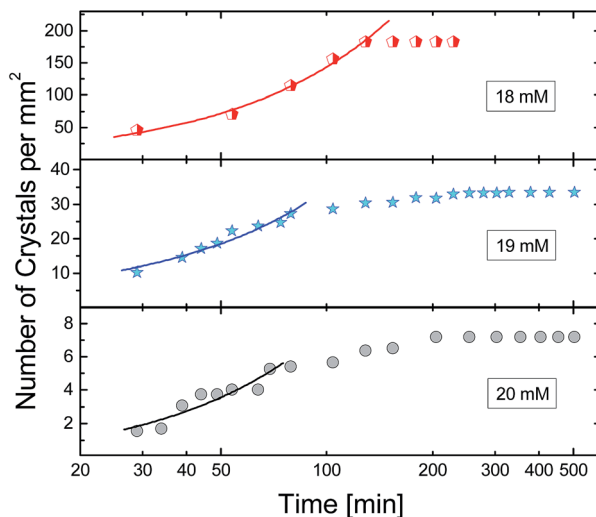
Fig. 4 Number of crystals as a function of time in a normalized area from Fig. 3 for samples with 18 mM, 19 mM and 20 mM CdCl₂. For the 19 and 20 mM plots, three series of picture sets were analyzed and averaged for better statistics.

typical SAXS results are shown in Fig. 5a for samples with a low protein concentration of 6.5 mg ml^{-1} and CdCl₂ concentrations covering all three regimes. In regime I, the scattering curve (with 0.5 mM salt) is dominated by the form factor of the BLG dimer, which is in agreement with the literature that BLG occurs predominantly as a dimer at room temperature and a pH between 3.5 and 7.5.[49] In regime II, with 1 mM salt, the high q part is still dominated by the form factor, but in the low q region, the increase in intensity indicates the formation of protein aggregates. In the third regime (12–90 mM, not all data are shown), a new feature forms shortly after preparation at $q \approx 0.7$ nm^{-1} and a sharp peak occurs at 2.2 nm^{-1}, as indicated by the arrows in this figure. A previous study of BLG in solution in the presence of YCl₃ has shown that this maximum corresponds to the monomer–monomer (M–M) correlation, due to the bridging effect.[50] Here, in the presence of CdCl₂, the peak is sharper. A possible explanation is the formation of a highly ordered fiber-like structure, which gives such a diffraction peak corresponding to the axial translation of the subunit (BLG monomer). Similar diffraction peaks have been observed in the solution scattering of F-actin.[51]

Fig. 5b presents SAXS data for samples under similar conditions as for the real-time optical microscopy study in Fig. 3. The SAXS profiles were collected during the crystallization process at a time at which the broad peak at $q \approx 0.7$ nm^{-1} has already developed, but Bragg peaks have not yet appeared. With 15 mM salt, the sample is located deep in regime II. While large aggregates are visible using optical microscopy (Fig. 3a), the SAXS profile shows a rather weak shoulder at $q \approx 0.7$ nm^{-1}, indicating less order within the large aggregates. Increasing the salt concentration increases the ordering inside of the aggregates, as the broad peak becomes more pronounced. The data set with 17 mM salt was collected at a different beamtime with higher resolution, and also shows the monomer–monomer correlation peak at q around 2.2 nm^{-1}.

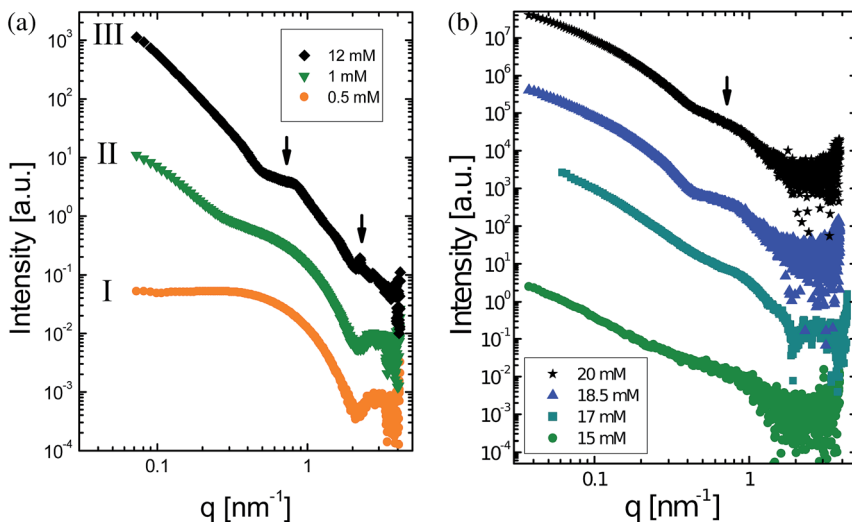Fig. 5 (a) SAXS measurements at a protein concentration of 6.5 mg ml^{-1}, with CdCl$_2$ concentrations covering all three regimes of RC (marked by I, II, III). (b) Selected SAXS profiles at the time when the broad peak is well developed for samples in regime III with 33 mg ml^{-1} BLG and 15, 17, 18.5 and 20 mM CdCl$_2$, respectively. Data are shifted upward for clarity.

We further measured a similar set of samples using SANS. In this case, the samples were prepared in D$_2$O instead of H$_2$O for better contrast. Neutron scattering causes no radiation damage to the proteins in solution, and most importantly, SANS measurements apply a large beam size (6 mm × 12 mm), which covers a much larger volume containing a reasonable number of crystals. This is in contrast to SAXS measurements with a beam size of 0.3 × 0.15 mm, which contains a limited number of crystals in the case of the crystallization in Fig. 3c and d. The measurements were carried out for freshly prepared samples, 2 h and 6 h after preparation, respectively. The results shown in Fig. 6 suggest that the samples behave very similarly as to how they do in H$_2$O. First of all, in all cases, the fresh samples show the broad peak at $q \approx 0.7$ nm^{-1} which has been identified as the characteristic structural feature of the MIP. However, due to the low time resolution, the details of the development of this peak cannot be obtained. Nevertheless, for the sample with 20 mM salt, it is clear that this broad peak develops and becomes more pronounced after 2 h. In addition to this broad peak, the monomer–monomer correlation peak at q around 2 nm^{-1} is pronounced in all cases, which is in good agreement with the SAXS measurements. The slight shift of the monomer–monomer peak to a low q value in SANS is due to the hydration effect.[52] Secondly, after a certain time, smeared Bragg peaks appear for all samples. At the same time, the broad peak (MIP) reduces in its intensity or completely disappears. Although the low resolution of SANS at the high-q region smears the Bragg peaks, their positions are consistent with the SAXS measurements.

Both SAXS and SANS measurements reveal a similar structural feature of the MIP, *i.e.* the local ordering within the large protein aggregates, characterized by a
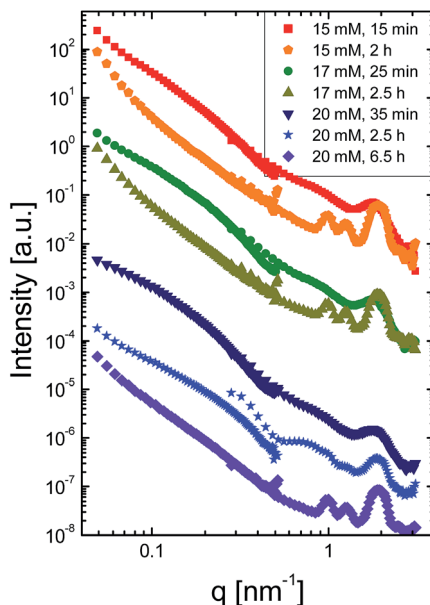
Faraday Discussions

Fig. 6 BLG 33 mg ml^{-1} with 15, 17 and 20 mM CdCl$_2$ in D$_2$O measured by SANS. Data are shifted upward for clarity. The mismatch of the SANS curves (especially visible for 20 mM after 2.5 h) was caused by measurements at different detector positions, together with the fast kinetics occurring in the sample.

broad peak at $q \approx 0.7$ nm^{-1} and the monomer–monomer correlation peak at q around 2 nm^{-1}. As discussed in the following section, the broad peak is closely related to nucleation and crystal growth. It thus becomes the structural signature of the MIP.

3.4 Crystallization kinetics followed by real-time SAXS

To extract information on the underlying crystallization process, we employed real-time SAXS measurements of the crystallization kinetics with high time and structural resolution. Fig. 7 shows examples of time-resolved SAXS measurements for 33 mg ml^{-1} BLG with 17 (a) and 20 mM CdCl$_2$ (b) in 3D surface illustration and 2D projection; additional data for samples with 15 and 18.5 mM CdCl$_2$ are shown in the ESI, Fig. S2.† The bottom 2D projections are created by dividing all curves by the first one, and therefore visualize the ongoing changes in the system with time. Selected $I(q,t)/I(q,t = 0)$ curves are further presented in Fig. 7c and d. The SAXS curves of both samples feature a strong increase at a low q which hardly changes with time; these indicate the presence of large aggregates, consistent with the observation by optical microscopy. With increasing time, a broad peak located at $q \approx 0.7$ nm^{-1} develops, which has been assigned to the nucleation precursors (MIP).[31] Once Bragg peaks appear, the most pronounced ones at 1.01 and 1.27 nm^{-1} overlap with the broad peak. The intensity of the Bragg peaks increases with time, while the broad peak shrinks. In the $I(q,t)/I(q,t = 0)$ plots (Fig. 7c and d), it is clear that the broad peak appears before the Bragg peaks and becomes stronger with time. At the end of crystallization, the broad peak shrinks and eventually disappears (Fig. 7a and c).
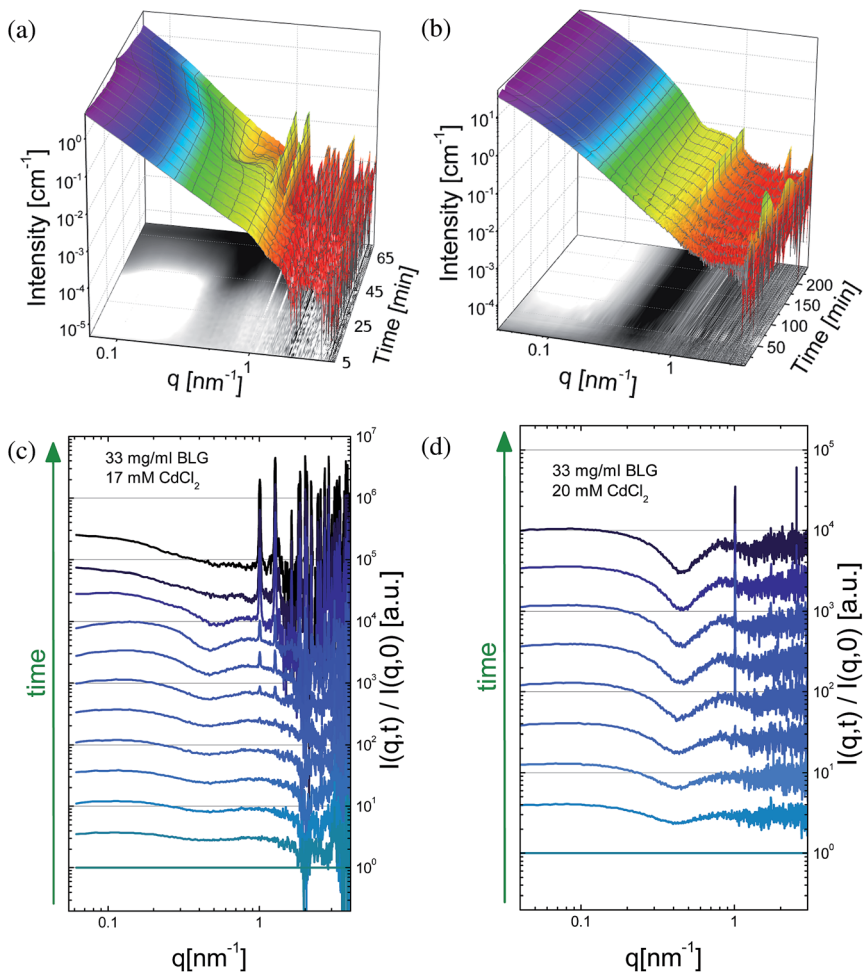
Fig. 7 Protein crystallization followed by real-time SAXS measurements. Samples contain 33 mg ml^{-1} BLG with 17 mM CdCl$_2$ (a & c) and 20 mM CdCl$_2$ (b & d), located in the transition zone of *pseudo-c*** (*i.e.* at the upper end of regime II). (a & b) Three-dimensional plots of $I(q,t)$. The bottom projections are created from $I(q,t)/I(q,t=0)$ and thus monitor the evolution of the curves with time. (c & d) Selected $I(q,t)/I(q,t=0)$ curves for the evolution of Bragg peaks and the amorphous peak. Data are shifted upward by a constant factor for better visibility.

From the optical microscopy experiments (Fig. 3) one observes that the MIP forms before crystallization starts, and is consumed during crystal growth. From real-time SAXS measurements, one can see that the typical broad peak for the MIP follows the same development: it appears first and develops, and once crystallization starts it reduces in intensity and eventually disappears. Based on these observations, we propose using the relative change of the area of this peak (representative of the MIP) and the two Bragg peaks which overlap with it, to quantify the relationship between the MIP and the crystalline phase as a function of time. At this point, we use the concept of crystallinity from

semi-crystalline polymer systems for further data analysis.[53] After subtraction of the intensity at the minimum, the broad peak in the $I(q,t)/I(q,t=0)$ curves was fitted by a scaled Gaussian function and the Bragg peaks by two further (sharp) Gaussians.[31] The crystallization kinetics can be followed by the enveloped area of the broad region, A_{interm}, and the area of the Bragg peaks, A_{Bragg}, as a function of time. This method is further illustrated in an animation that can be found in the ESI†. Fig. 8a displays an example of such an analysis, for a sample with 33 mg ml^{-1} BLG and 17 mM CdCl$_2$. The development of the MIP (A_{interm}) shows a maximum at around 40 min, and the overall crystallinity (A_{Bragg}) has a plateau between 40 and 60 min, and then grows faster. Interestingly, the overall growth rate, $i.e.$ the first derivative of A_{Bragg} on time, gives a maximum that is also located around 40 min, indicating that the overall crystal growth rate in the early stage strongly depends on the development of the MIP.

We have performed real-time SAXS measurements on all four salt conditions followed by optical microscopy. However, as seen from Fig. 3, the number of crystals decreases and the size of crystals increases with increasing salt concentration. This makes the real-time SAXS measurements challenging, as the number of crystals within the illuminated volume drops significantly. We have tried to compensate for this by measuring more positions from the sample. This is partially successful. However, the time resolution is reduced, as only one or two out of ten spots show the development of the Bragg peaks. As shown in Fig. 8b, less data for Bragg peaks than the MIP are shown. Nevertheless, one can still recognize the interesting kinetics: first the MIP develops relatively fast, and it then becomes saturated after 120 min. Within the current experimental time scale, only a minor fraction of crystalline phase was detected. The experimental observations on the kinetics, particularly including the non-monotonous crystallization rate (red dashed line in Fig. 8a), agree well with a simple model, which will be discussed in the following section.
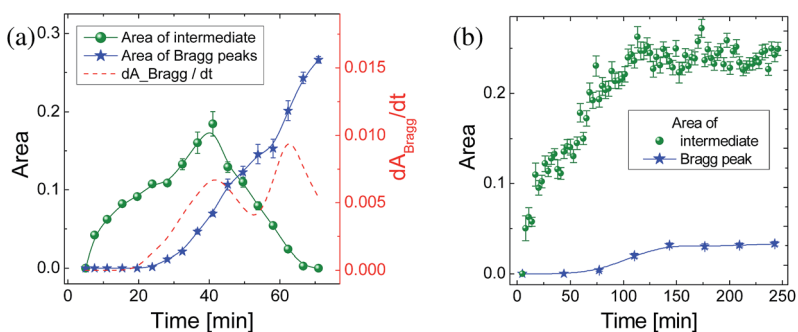

Fig. 8 Time-dependence of the area of the broad peak (A_{interm}) and the Bragg peaks (A_{Bragg}). Green spheres (left axis) show $A_{interm}(t)$, the integral of the broad Gaussian function connected to the intermediate. Blue stars (left axis) show $A_{Bragg}(t)$, the integral of the two Bragg peaks in this area. (a) BLG 33 mg ml^{-1} with 17 mM CdCl$_2$. For clarity and better statistics, at each time point three data points of $A_{interm}(t)$ and $A_{Bragg}(t)$ were merged into one. The dashed red line corresponds to the first derivative of the area of Bragg peaks as a function of time (right axis). (b) BLG 33 mg ml^{-1} with 20 mM CdCl$_2$.

Paper

4. Modeling with rate equations

In order to also compare the observed kinetic features more quantitatively to possible crystallization scenarios, we employed rate equation models. As essential variables of the modeling, L, I, C_I and C_L denote the mass fractions of free monomers in the liquid, in the intermediate, crystals in the intermediate and crystals in the liquid, respectively. The three paradigmatic cases for the crystallization process shown in Fig. 1 are presented as follows:

Classical nucleation $L \rightarrow C_L$

The classical nucleation theory contains a one-step nucleation process of the critical nucleus from the homogeneous solution. After nucleation, crystallites grow larger from the solution. In terms of modeling, $\Delta_{nL} = k_n L$ and $\Delta_{gL} = k_{gL} L C_L$ represent the nucleation and the growth term with rates k_n and k_{gL}. Using these, the process can be easily modeled by the following set of rate equations:

$$\partial_t L = -\Delta_{nL} - \Delta_{gL} \tag{1}$$

$$\partial_t C_L = \Delta_{nL} + \Delta_{gL}$$

Parallel process $I \leftrightarrow L \rightarrow C_L$

In this nonclassical process, we assume that in addition to the one-step crystallization process, a reversible intermediate is formed in the solution, competing with crystallization for the free monomers. In addition to the nucleation and growth term from the classical one-step nucleation, we include the formation of the intermediate with the term $\Delta_I = k_I(L - I)$. The corresponding set of rate equations reads

$$\partial_t L = -\Delta_I - \Delta_{nL} - \Delta_{gL} \tag{2}$$

$$\partial_t I = \Delta_I$$

$$\partial_t C_L = \Delta_{nL} + \Delta_{gL}$$

Two-step nucleation process $L \rightarrow I \rightarrow C_I \rightarrow C_L$

For true two-step nucleation, the formation of an intermediate is followed by the nucleation of crystallites within the intermediate. After the subsequent growth of the crystallites has consumed the intermediate, crystals emerge into the remaining liquid and grow further. Table 1 describes the steps in more detail and lists the rate terms. Using these, the rate equations read

$$\partial_t L = -\Delta_I - \Delta_{gL} \tag{3}$$

$$\partial_t I = \Delta_I - \Delta_{nI} - \Delta_{gI}$$

$$\partial_t C_I = \Delta_{nI} - \Delta_e + \Delta_{gI}$$

$$\partial_t C_L = \Delta_e + \Delta_{gL}$$

The results of the three models are shown in Fig. 9 with the model parameters from another study[31] (see also figure caption). For the classical nucleation (Fig. 9a), nucleation and growth of crystals speed up, until the liquid is halfway consumed, and saturate until complete consumption of the liquid. In the parallel process (Fig. 9b), the intermediate competes with crystallization and slows down the crystal nucleation and growth considerably. It is important to notice, however, that the crystallization rate is homogeneously increasing until saturation. By contrast, the two-step process (Fig. 9c) shows the formation of a plateau in the crystal mass fraction. After increasing at small times, the crystallization rate drops again considerably, whereas it increases monotonically for both one-step cases until saturation (Fig. 9d). This particular feature of the two-step process is caused by the nucleation and slow growth in the intermediate, while the crystals grow faster once emerged into solution. Thus, the occurrence of the plateau indicates the presence of a multi-step nucleation process.

The large number of model parameters does not allow for a reliable extraction of nucleation rates *via* model fits to the kinetic analysis of the SAXS data. We emphasize that this is not a problem of the model or the data quality, but the complex pathway, which involves many coupled processes. As an example, the amount of crystal nucleation within the MIP depends not only on the rate k_n, but also on the amount of MIP. The latter is mainly determined by the supersaturation represented by L_0, at least for fast formation of the MIP compared to crystal

Table 1 Summary of the proposed two-step process. From the intermediate formed firstly from the liquid, crystals nucleate and grow until they consume the intermediate. Subsequently, the crystals emerge into the liquid and grow from the free monomers. $(x)_+$ stands for x for positive x and zero otherwise. L, I, C_I and C_L denote the mass fractions of free monomers in the liquid, in the intermediate, crystals in the intermediate and crystals in the liquid, respectively

	Processes	Rates	Parameters
	1. Formation of intermediate until the remaining solution is stable	$\Delta_I = k_I(L - L_0)_+$	L_0 mass fraction of stable solution k_I formation rate
	2. Crystal nucleation within the intermediate	$\Delta_{nI} = k_n I$	k_n nucleation rate
	3. Crystal growth within the intermediate	$\Delta_{gI} = k_{gI} I C_I$	k_{gI} growth rate in intermediate
	4. Crystals emerge into the liquid once intermediate is consumed	$\Delta_e = k_e C_I(\alpha_I C_I - I)_+$	α_I critical ratio of intermediate and crystals k_e emergence rate
	5. Crystal growth in the liquid	$\Delta_{gL} = k_{gL} L C_L$	k_{gL} growth rate in liquid

nucleation. Based only on fitting the kinetic data from SAXS, an decrease in L_0 will consequentially cause a decrease in k_n. While additional information *e.g.* on the supersaturation are therefore required for a real quantitative fit, the qualitative signatures of the two-step process, such as the plateau in the crystal fraction, can still be used to provide evidence for the scenario of a two-step nucleation process.

Nevertheless, this simple model can reproduce the experimental crystallization kinetics at different conditions. This can be achieved either by varying the rate parameters or by choosing the amount of MIP which approximately approaches $L - L_0 = 0.2$. This value can be determined experimentally by following the protein concentration in the supernatant over time using UV-visible
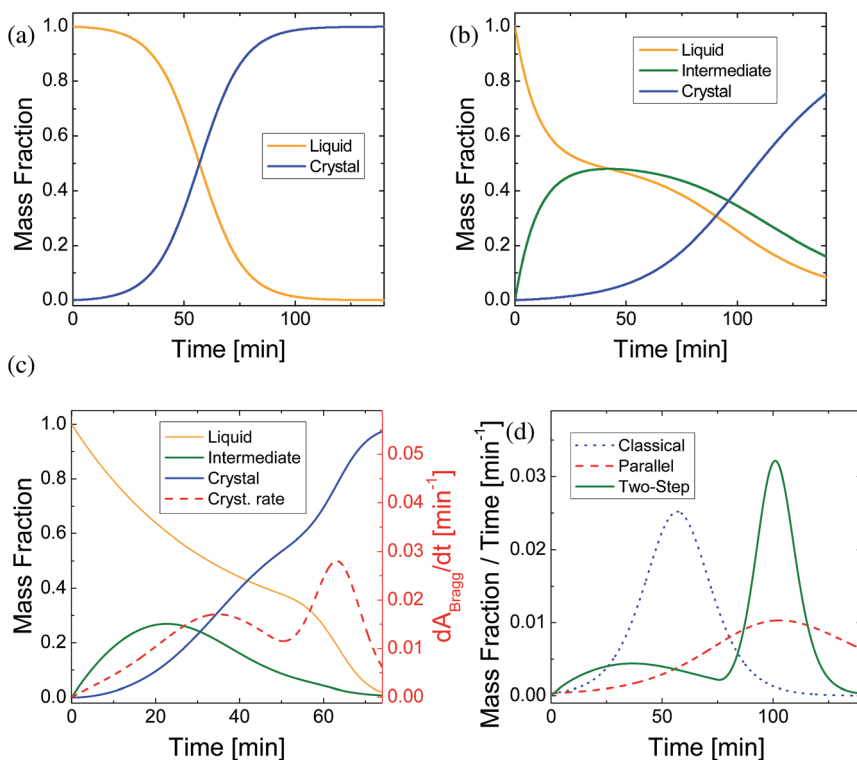

Fig. 9 Comparison between different crystallization scenarios: (a) classical nucleation from a homogeneous liquid and subsequent growth of the crystallites. (b) Parallel processes, representing the formation of an intermediate and independent nucleation from the liquid with subsequent growth of the crystallites. (c) Two-step nucleation process involving the formation of an intermediate from which the crystals nucleate (see Table 1 for further details). The following parameter values were used to show the good qualitative agreement of the model with the data set in Fig. 8a: $k_e = 0.15$ min^{-1}, $L_0 = 0.2$, $\alpha_l = 0.2$, $k_{gL} = 0.6$ min^{-1}, $k_l = 0.03$ min^{-1}, $k_n = 0.02$ min^{-1}, $k_{gl} = 0.2$ min^{-1}. An additional model plot reproducing the data set in Fig. 8b can be found in the ESI, Fig. S3.† (d) Comparison of the crystallization rates dC/dt. While one-step and parallel nucleation processes show a monotonous speed-up until saturation, the two-step process can have a non-monotonous signature with two maxima. The model parameters were chosen for all three models to be the same: $k_l = 0.05$ min^{-1}, $k_n = 0.02$ min^{-1}, $k_{gL} = 1.0$ min^{-1}, $k_{gL} = 0.2$ min^{-1}, $k_e = 1.0$ min^{-1}, $L_0 = 0.7$, $\alpha_l = 0.2$.

spectroscopy. A tentative experiment for a sample with 20 mM $CdCl_2$ leads to a value of 0.2. Further experiments are needed to refine this parameter.

5. Conclusions and outlook

We have investigated the two-step nucleation process of protein crystallization in solutions by following the overall crystallization kinetics, using real-time optical microscopy and SAXS. The experimental results, together with a rate equation model, provide evidence of two-step nucleation in the early stage of crystallization. The BLG-salt ($CdCl_2$) solutions were chosen at the transition zone of *pseudo-c***, where small aggregates form after sample preparation. These protein aggregates serve as the metastable intermediate phase (MIP) during crystallization.

SAXS and SANS reveal that the MIP shows a certain local ordering instead of random aggregates, as monitored by a broad shoulder at intermediate $q \approx 0.7$ nm^{-1}, and a monomer–monomer correlation peak at q around 2 nm^{-1}. Real-time SAXS results show that the crystallization kinetics is proportional to the development of the MIP in the early stage of crystallization, *i.e.* there appears a local maximum in the crystallization rate at the maximum quantity of the intermediate. In the late stage of crystallization, a plateau develops due to the transition from nucleation controlled in the early stage, to growth controlled after the consumption of MIP. This transition in the overall crystallization kinetics is a typical feature for two-step nucleation in the early stage. These experimentally observed kinetics can be reproduced using a rate equation model.

For further real-time measurements, we note that the smaller beam size and scattering volume of SAXS can be compensated by using SANS. The combination of real-time SAXS and SANS could provide more systematic information of the crystallization kinetics.

Acknowledgements

References

1 L. C. Jacobson, W. Hujo and V. Molinero, *J. Am. Chem. Soc.*, 2010, **132**, 11806–11811.

2 H. Cölfen and M. Antonietti, *Mesocrystals and nonclassical crystallization*, John Wiley & Sons, West Sussex, 2008.

3 D. Gebauer, A. Völkel and H. Cölfen, *Science*, 2008, **322**, 1819–1822.

4 D. Gebauer, M. Kellermeier, J. D. Gale, L. Bergström and H. Cölfen, *Chem. Soc. Rev.*, 2014, **43**, 2348–2371.

5 A. F. Wallace, L. O. Hedges, A. Fernandez-Martinez, P. Raiteri, J. D. Gale, G. A. Waychunas, S. Whitelam, J. F. Banfield and J. J. De Yoreo, *Science*, 2013, **341**, 885–889.

6 R. P. Sear, *Int. Mater. Rev.*, 2012, **57**, 328–356.

7 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.

8 O. Gliko, W. Pan, P. Katsonis, N. Neumaier, O. Galkin, S. Weinkauf and P. G. Vekilov, *J. Phys. Chem. B*, 2007, **111**, 3106–3114.

9 W. Pan, O. Galkin, L. Filobelo, R. L. Nagel and P. G. Vekilov, *Biophys. J.*, 2007, **92**, 267–277.

10 W. Pan, P. G. Vekilov and V. Lubchenko, *J. Phys. Chem. B*, 2010, **114**, 7620–7630.

11 Q. Chen, P. G. Vekilov, R. L. Nagel and R. E. Hirsch, *Biophys. J.*, 2004, **86**, 1702–1712.

12 Y. Liu, X. Wang and C. B. Ching, *Cryst. Growth Des.*, 2010, **10**, 548–558.

13 T. Schilling, H. J. Schöpe, M. Oettel, G. Opletal and I. Snook, *Phys. Rev. Lett.*, 2010, **105**, 025701.

14 P. R. T. Wolde and D. Frenkel, *Science*, 1997, **277**, 1975–1978.

15 S. B. Hutchens and Z.-G. Wang, *J. Chem. Phys.*, 2007, **127**, 084912.

16 G. I. Tóth, T. Pusztai, G. Tegze, G. Tóth and L. Gránásy, *Phys. Rev. Lett.*, 2011, **107**, 175702.

17 H. J. Schöpe, G. Bryant and W. van Megen, *Phys. Rev. Lett.*, 2006, **96**, 175701.

18 T. H. Zhang and X. Y. Liu, *J. Am. Chem. Soc.*, 2007, **129**, 13520–13526.

19 J. R. Savage and A. D. Dinsmore, *Phys. Rev. Lett.*, 2009, **102**, 198302.

20 M. Sleutel and A. E. S. Van Driessche, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E546–E553.

21 A. Bijelic, C. Molitor, S. G. Mauracher, R. Al-Oweini, U. Kortz and A. Rompel, *ChemBioChem*, 2015, **16**, 233–241.

22 P. Rein ten Wolde and D. Frenkel, *Phys. Chem. Chem. Phys.*, 1999, **1**, 2191–2196.

23 P. G. Vekilov, *Cryst. Growth Des.*, 2004, **4**, 671–685.

24 P. Tan, N. Xu and L. Xu, *Nat. Phys.*, 2014, **10**, 73–79.

25 Y. Peng, F. Wang, Z. Wang, A. M. Alsayed, Z. Zhang, A. G. Yodh and Y. Han, *Nat. Mater.*, 2015, **14**, 101–108.

26 T. Kawasaki and H. Tanaka, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14036–14041.

27 J. F. Lutsko and G. Nicolis, *Phys. Rev. Lett.*, 2006, **96**, 046102.

28 R. Piazza, *Curr. Opin. Colloid Interface Sci.*, 2004, **8**, 515–522.

29 R. P. Sear, *Curr. Opin. Colloid Interface Sci.*, 2006, **11**, 35–39.

30 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.

31 A. Sauter, F. Roosen-Runge, F. Zhang, G. Lotze, R. M. J. Jacobs and F. Schreiber, *J. Am. Chem. Soc.*, 2015, **137**, 1485–1491.

32 H. A. Sober, *CRC Handbook of Biochemistry: Selected data for molecular biology*, CRC, 1970.

33 U. M. Elofsson, M. A. Paulsson and T. Arnebrant, *Langmuir*, 1997, **13**, 1695–1700.

34 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: Struct., Funct., Bioinf.*, 2010, **78**, 3450–3457.

35 T. Narayanan, Synchrotron small-angle X-ray scattering, in *Soft Matter Characterization*, ed. R. Borsali and R. Pecora, Springer, 2008, p. 899ff.

36 V. Pipich, QtiKWS: *user-friendly program for reduction, visualization, analysis and fit of SA(N)S data*, http://www.qtikws.de, 2012.

37 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.

38 F. Zhang, G. Zocher, A. Sauter, T. Stehle and F. Schreiber, *J. Appl. Crystallogr.*, 2011, **44**, 755–762.

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, F. Roosen-Runge, A. Sauter, M. Wolf, R. M. J. Jacobs and F. Schreiber, *Pure Appl. Chem.*, 2014, **86**, 191–202.

41 F. Roosen-Runge, B. S. Heck, F. Zhang, O. Kohlbacher and F. Schreiber, *J. Phys. Chem. B*, 2013, **117**, 5777–5787.

42 E. Jordan, F. Roosen-Runge, S. Leibfarth, F. Zhang, M. Sztucki, A. Hildebrandt, O. Kohlbacher and F. Schreiber, *J. Phys. Chem. B*, 2014, **118**, 11365–11374.

43 D. Soraruf, F. Roosen-Runge, M. Grimaldo, F. Zanini, R. Schweins, T. Seydel, F. Zhang, R. Roth, M. Oettel and F. Schreiber, *Soft Matter*, 2014, **10**, 894–902.

44 F. Roosen-Runge, F. Zhang, F. Schreiber and R. Roth, *Sci. Rep.*, 2014, **4**, 7016.

45 A. Kubíčková, T. Křížek, P. Coufal, M. Vazdar, E. Wernersson, J. Heyda and P. Jungwirth, *Phys. Rev. Lett.*, 2012, **108**, 186101.

46 R. Akiyama and R. Sakata, *J. Phys. Soc. Jpn.*, 2011, **80**, 123602.

47 A. Sauter, M. Oelker, G. Zocher, F. Zhang, T. Stehle and F. Schreiber, *Cryst. Growth Des.*, 2014, **14**, 6357–6366.

48 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.

49 M. Verheul, J. S. Pedersen, S. P. F. M. Roefs and K. G. de Kruif, *Biopolymers*, 1999, **49**, 11–20.

50 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.

51 T. Sato, T. Shimozawa, T. Fukasawa, M. Ohtaki, K. Aramaki, K. Wakabayashi and S. Ishiwata, *Biophysics*, 2010, **6**, 1–11.

52 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.

53 G. Strobl, in *The physics of polymers*, Springer, Berlin, Heidelberg, New York, 3rd edn, 2007, ch. 5.