The nucleation of protein crystals as a race against time with on- and off-pathways

Cecilia Ferreira,[a] Silvia Barbosa,[b,c] Pablo Taboada,[b,c] Fernando A. Rocha,[a] Ana M. Damas[d] and Pedro M. Martins[a,d]*

[a]LEPABE, Laboratório de Engenharia de Processos, Ambiente, Biotecnologia e Energia, Departamento de Engenharia Química, Faculdade de Engenharia da Universidade do Porto, Rua Dr Roberto Frias, 4200-465 Porto, Portugal, [b]Grupo de Física de Coloides y Polímeros, Facultad de Física, Universidad de Santiago de Compostela, Spain, [c]Instituto de Investigación Sanitaria (IDIS) de Santiago de Compostela, Spain, and [d]ICBAS – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal. *Correspondence e-mail: pmartins2106@gmail.com

High supersaturation levels are a necessary but insufficient condition for the crystallization of purified proteins. Unlike most small molecules, proteins can take diverse aggregation pathways that make the outcome of crystallization assays quite unpredictable. Here, dynamic light scattering and optical microscopy were used to show that the nucleation of lysozyme crystals is preceded by an initial step of protein oligomerization and by the progressive formation of metastable clusters. Because these steps deplete the concentration of soluble monomers, the probability of obtaining protein crystals decreases as time progresses. Stochastic variations of the induction time are thus amplified to a point where fast crystallization can coexist with unyielding regimes in the same conditions. With an initial hydrodynamic radius of ~100 nm, the metastable clusters also promote the formation of protein crystals through a mechanism of heterogeneous nucleation. Crystal growth (on-pathway) takes place in parallel with cluster growth (off-pathway). The Janus-faced influence of the mesoscopic clusters is beneficial when it accelerates the formation of the first precrystalline nuclei and is detrimental as it depletes the solution of protein ready to crystallize. Choosing the right balance between the two effects is critical for determining the success of protein crystallization trials. The results presented here suggest that a mild oligomerization degree promotes the formation of a small number of metastable clusters which then catalyze the nucleation of well differentiated crystals.

1. Introduction

The progress of macromolecular crystallography is often dependent on obtaining well diffracting protein crystals for X-ray crystallographic analysis. Protein molecules are characterized by a complex energy landscape involving different folding states besides the native conformation of lowest entropy (Tyka *et al.*, 2011). In addition, noncovalent bonding of macromolecules gives rise to different sized oligomers, with molar concentrations highly influenced by temperature, ionic strength, pH and total concentration of the protein sample (Wilson *et al.*, 1996). Higher-order aggregates such as metastable clusters (Tanaka *et al.*, 1999) and amyloid fibrils (Crespo *et al.*, 2016) are also produced depending on the solution conditions. Absent during the crystallization of salts, minerals and other model systems, these pathways help to explain why the crystallization of macromolecules remains unsatisfactorily described by classical theories (García-Ruiz, 2003). Finally, the nucleation step by which any crystal is formed is itself one of the most poorly understood processes in nature (Vekilov, 2016), whereas alternative outcomes such as liquid–liquid

① Oligomerization
② Cluster nucleation
③ Crystal nucleation
④ Cluster growth
⑤ Crystal growth

MONOMERS
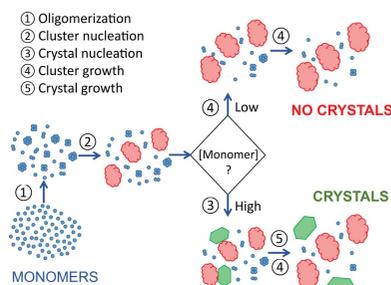
separation (Wallace *et al.*, 2013) and amorphous precipitation (Veesler *et al.*, 2003) are also expected for other smaller molecules.

Much of the knowledge about the critical prenucleation stages has been obtained using dynamic light scattering (DLS), as this technique can be directly applied to supersaturated protein solutions without causing denaturation or major interference (Wilson, 2003). Hen egg white lysozyme (HEWL) has been particularly analyzed in these studies because it is an available protein well characterized in terms of thermodynamic phase diagrams (Muschol & Rosenberger, 1997), crystal nucleation kinetics (Akella *et al.*, 2014) and crystal growth kinetics (Liu *et al.*, 2010). The DLS signature of prenucleation HEWL solutions is common to several other proteins (Schubert *et al.*, 2015; Vekilov, 2016) and reveals the presence of two groups of scatterers, one consisting of protein monomers and low-order oligomers, and the other, commonly referred to as 'protein clusters', containing over 10^5 aggregated molecules (Pan *et al.*, 2010; Tanaka *et al.*, 1999; Georgalis *et al.*, 1995; Onuma & Kanzaki, 2007; Saridakis *et al.*, 2002). The average hydrodynamic radius of the latter group is more than $50\times$ larger than that of the first one ($R_2 > 50R_1$), where R_1 is of the same order of magnitude as the molecular size (Pan *et al.*, 2007). There are, however, several questions about the composition and nature of the two populations that need further clarification. Small angle X-ray scattering (SAXS) was recently used to determine the oligomeric composition of HEWL and test a structure-based model dominated by monomers, dimers and octamers under crystallization conditions, and only monomers when crystal growth was not possible (Kovalchuk *et al.*, 2016). These measurements indicated that the likely crystal growth unit of HEWL is an octamer (Kovalchuk *et al.*, 2016), in agreement with previous estimations derived from growth-rate modeling of tetragonal lysozyme crystals by Li *et al.* (1995). The so-called Li *et al.* model of crystal growth includes an oligomerization pathway of monomer ↔ dimer ↔ tetramer ↔ octamer ↔ 16-mer that was subsequently confirmed by measurements of dialysis flux data (Wilson *et al.*, 1996).

Concerning the higher-order aggregates, a distinction should be made between metastable clusters formed at moderate supersaturation levels (Tanaka *et al.*, 1999; Georgalis *et al.*, 1995) and the dense phase droplets formed after liquid–liquid separation at high protein concentrations (often >100 mg ml^{-1}) (Muschol & Rosenberger, 1997; Vekilov & Vorontsova, 2014; Pan *et al.*, 2010). Both have similar values of initial hydrodynamic radius but, while metastable clusters are formed after a lag phase of variable duration (Tanaka *et al.*, 1999) and then grow steadily with time (Georgalis *et al.*, 1995), the dense phase droplets appear immediately after solution preparation and their mean radius remains unchanged after several hours of monitoring (Vekilov & Vorontsova, 2014; Pan *et al.*, 2010). From the power-law variation of the hydrodynamic radius R_2 with time t, attempts have been made to classify the mechanism of cluster growth as diffusion or reaction limited according to the value of the power exponent of t (Tanaka *et al.*, 1999; Georgalis *et al.*, 1995). The reciprocal

of the power exponent corresponds to the fractal dimension d_f, whose values for crystallization solutions of HEWL have been determined to be in the range of 1.7–2.0 (Tanaka *et al.*, 1999; Georgalis *et al.*, 1995; Parmar *et al.*, 2007). Although this may suggest the aggregation of nucleated clusters into fractal structures, similar values of d_f were obtained from the size evolution of pre-assembled clusters present in the lyophilized HEWL used in nucleation studies (Parmar *et al.*, 2007). If the contaminant aggregates are removed before nucleation, longer periods of time are required to produce fewer (and larger) crystals, indicating that the pre-assembled clusters act as heterogeneous nucleation centers (Parmar *et al.*, 2007). In an analogous way, larger and morphologically better crystals could be produced by diluting crystallization drops from nucleation to metastable conditions after the formation of mesoscopic clusters (Saridakis *et al.*, 2002). This procedure caused the partial reversion of the high-order aggregates into soluble proteins (Saridakis *et al.*, 2002) which, in turn, decreases the number of heterogeneous nucleation centers and increases the protein pool available for the production of larger crystals.

Differently from the metastable clusters, dense liquid droplets are an intermediate state during the crystallization process in the sense that crystal nucleation occurs within the protein-rich phase (Vekilov & Vorontsova, 2014; Vorontsova *et al.*, 2015). In comparison with the conditions commonly adopted by structural biologists, liquid–liquid separation requires atypically high supersaturation levels corresponding to high concentrations of protein and precipitant and low solution temperatures (Pan *et al.*, 2010). The nucleation kinetics of HEWL in the vicinity of the liquid–liquid phase boundary deviate markedly from the predictions of classical nucleation theory (Galkin & Vekilov, 1999), whereas these differences become less evident at slightly lower supersaturation levels (Akella *et al.*, 2014; Liu *et al.*, 2010). In this study we went further down in the HEWL phase diagram to study regions of indeterminate crystallization outcome regarding, for example, the presence or absence of protein crystals. It is commonly observed in macromolecular crystallography that experiments set up under seemingly identical conditions produce very different results, regardless of the protein under investigation (Luft *et al.*, 2011, 2014; Gorrec, 2013; Newman *et al.*, 2007). Alone, the stochastic nature of nucleation seems insufficient to explain why some crystallization replicates produce visible crystals at the end of a few hours while others produce no crystals at all after several days of observation (Newman *et al.*, 2007; Martins *et al.*, 2008). Our measurements of HEWL nucleation rates using optical microscopy indicate the existence of concurrent aggregation pathways competing with crystal nucleation. We confirmed by DLS analysis that the monomeric protein is depleted by the sequential formation of low-order oligomers and metastable clusters, with the latter growing above the submicrometre size. Off-pathway aggregates also have positive effects on crystallization since protein oligomers seem to favor crystal growth and the mesoscopic clusters are confirmed as heterogeneous nucleation centers.

2. Materials and methods

2.1. Protein Preparation

HEWL powder obtained from Merck KGaA (Darmstadt, Germany) was dissolved in 0.2 M sodium acetate buffer pH 4.7, containing 3%(w/v) sodium chloride. The solutions were prepared to final concentrations of 25, 30, 35, 40, 45 and 50 mg ml^{-1} HEWL and filtered through sterile 0.22 μm pore size filters (Millipore Millex-GV). The steps of sample preparation were carried out at 301 K to guarantee total protein dissolution and prevent the formation of any precursors of crystalline material.

2.2. Observation of crystal formation

A calibrated Nikon SMZ1500 stereomicroscope equipped with a Nikon DS-Ri1 digital camera and Nis-Elements BR 3.1 software (Ferreira *et al.*, 2013, 2012) was used to determine the apparent induction time for the occurrence of the first crystal (t_1) from the analysis of time-lapse image acquisition of different crystallization drops. 0.7 μl drops of freshly prepared HEWL solutions were placed under 1.0 ml of thermostatized paraffin oil onto glass sitting-drop rods (HR3-146, Hampton Research), in a jacketed glass growth cell at constant temperatures of 285.8, 287.2, 288.2, 289.2, 290.2, 291.2 and 293.2 K. If no crystals were produced at the end of the first 24 h of continuous observation, the experiment was interrupted and assigned to an indeterminately high value of induction time. Crystallization success rates were calculated as the fraction of drops producing crystals during the period of observation. Crystal nucleation rates were calculated as the number of crystals formed per unit volume of drop and per unit of apparent induction time. A variable number of replicate conditions were studied to evaluate the effect of temperature and HEWL concentration on the induction times and nucleation rates. Only conditions of uncertain crystallization outcome (for which the success rates were different from 0 or 100%) were chosen. As in a previously reported control analysis, 0.7 μl crystallization drops under comparable experimental conditions showed no major variation of protein concentration and enzymatic activity during the first 24 h of incubation (Ferreira *et al.*, 2016).

2.3. DLS

DLS measurements were performed using an ALV/DLS/SLS-5000F, SP-86 goniometer system (ALV-GmbH, Langen, Germany) equipped with a continuous wave diode-pumped Nd:YAG solid-state Compass-DPSS laser with a symmetrizer (Coherent Inc., Santa Clara, CA). The laser operates at 532 nm with an output power of 400 mW. The intensity scale was calibrated against scattering from toluene. Samples of 1.0 ml of freshly prepared HEWL solutions, incubated in DLS glass cuvettes at constant temperatures of 289.2 and 298.2 K, were periodically analyzed at a scattering angle of 90° with respect to the incident beam. Hydrodynamic radii of the particles in solution were estimated from the diffusion coefficient(s) obtained from a *CONTIN* analysis (Provencher, 1982) of the auto-correlation function of the scattered light intensity, and taking into account the effects of temperature and concentration of sodium chloride on the solution viscosity.
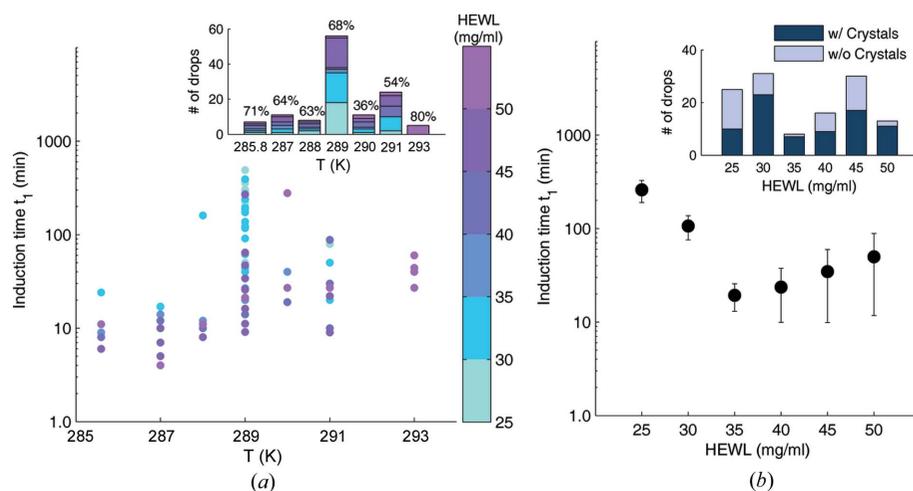
Figure 1
Influence of temperature (T) and HEWL concentration (C_{HEWL}) on the apparent induction time for the formation of the first crystal (t_1). (*a*) Symbols: influence of T on the measured values of t_1 for solutions containing different HEWL concentrations (indicated in the color bar) in the presence of 3%(w/v) NaCl. Inset: number of crystallization drops observed at each temperature for the HEWL concentrations indicated in the color bar. The percentages of drops producing crystals at the end of periods of observation of at least 24 h are given on the top of the bars. (*b*) Symbols and error bars: mean and 90% confidence intervals of t_1 values obtained for each C_{HEWL} excluding cases where no crystals were produced. Inset: number of crystallization drops observed for each C_{HEWL}. Legend: cases with (w/) or without (w/o) the presence of crystals at the end of the period of observation. (*a*) and (*b*) are different log–linear plots of the same experimental data.

3. Results and discussion

3.1. Nucleation of HEWL crystals followed by optical microscopy

To investigate the alternative aggregation pathways coexisting with the nucleation of protein crystals, we selected HEWL crystallization conditions for which the nucleation probability is less than 1. Specifically, we used optical microscopy to follow the evolution of 0.7 μl drops of protein solution containing 25–50 mg ml^{-1} HEWL, 3%(w/v) NaCl in 0.2 M sodium acetate buffer pH 4.7, at constant temperatures in the range between 285.8 and 293.2 K. As expected, the time required for the observation of the first crystal (t_1) greatly varied, from a few minutes to indeterminately high values (>24 h), for the same replicate experiments (Fig. 1*a*).

The condition of steady-state nucleation rates that would allow

calculation of a finite average value of the nucleation rate valid for all replicates (Galkin & Vekilov, 1999, 2001; Ildefonso *et al.*, 2013) is not applicable here, since only a fraction of the drops were able to produce crystals in a finite period of time. Instead, the data in Fig. 1(*a*) suggest the existence of at least two nucleation regimes coinciding with finite and indeterminate values of t_1. Moreover, conditions with high success rates (Fig. 1*a* inset) do not necessarily correspond to faster nucleation rates (and lower values of t_1). In this section, we will continue to focus on kinetic aspects of the productive regime; we will return to the discussion of unyielding crystallization drops in the following sections.

The effect of temperature T on the time required for nucleation is too slight to be discernible from the scattered results in Fig. 1(*a*). In fact, according to systematic measurements of the nucleation kinetics under comparable experimental conditions, the range of temperatures 285.8−293.2 K coincides with the transition from positive to negative effects of T on the nucleation rate J (Galkin & Vekilov, 2001). In Fig. 1(*b*), mean values of t_1 are represented as a function of HEWL concentration, taking into consideration the negligible T effect and excluding the cases in which crystals were not produced. As the protein concentration increases, the t_1 values initially decrease from >4 h to a minimum of ∼20 min at $C_{HEWL} = 35$ mg ml^{-1}. Above this limit concentration, the induction times progressively increase to reach t_1 values close to 1 h for $C_{HEWL} = 50$ mg ml^{-1}. This behavior is unexpected since, in principle, the higher supersaturation levels for $C_{HEWL} > 35$ mg ml^{-1} should continue to shorten the duration of the lag phase to values lower than a few minutes. Although t_1 is the apparent induction time for nucleation, a period of crystal growth t_g is required until the first nuclei reach tech-

nically observable dimensions (García-Ruiz, 2003). The duration of t_g is also expected to decrease with C_{HEWL}. Therefore, the trend observed in Fig. 1(*b*) suggests a nucleation mechanism composed of a series of events, some of which are unfavorably affected by higher protein concentrations.

Secondary nucleation, *i.e.* the formation of new nuclei because of the presence of other HEWL crystals in solution, is not likely to have affected the induction time results since the number of crystals per drop was small (most of the time <5 and always <16), and the continuous observation of the drops showed simultaneous formation of the crystals in each experiment. The paraffin oil used to prevent drop evaporation noticeably reduced the occurrence of heterogeneous nucleation as compared with the extensive formation of HEWL crystals at the solution–cuvette interface in DLS experiments (see below). Only drops that produced crystals within 24 h were taken into account when estimating the mean nucleation rates (J) represented by the black symbols in Fig. 2. The inset graph illustrates an alternative method to calculate J, based on the time evolution of the average number of crystals per drop considering both the productive and clear crystallization drops. Since the two nucleation regimes are not decoupled by the latter method, the limit nucleation rate given by the initial slope of the curve is significantly lower than the result obtained by the first method (∼1 order of magnitude in the example considered in Fig. 2). The hyperbolic type progress curve shown in the inset plot of Fig. 2 denotes a gradual decrease of the homogeneous nucleation rate towards complete cessation as the incubation time exceeds a limit value t^*. In the experiments carried out at 289.2 K the values of t^* decreased from ∼8 to ∼4 h as C_{HEWL} increased from 25 to 40 mg ml^{-1}.

In contrast with the induction time results shown in Fig. 1(*b*), the values of J corresponding to the fastest nucleation are observed at higher HEWL concentrations than 35 mg ml^{-1}. Thus at $C_{HEWL} > 35$ mg ml^{-1} an increasing number of nuclei are formed, but the lag time before the appearance of the first nuclei also increases. This influence of C_{HEWL} on the nucleation rates is very far from the exponential-type relationship traditionally associated with homogeneous nucleation kinetics (Akella *et al.*, 2014; Galkin & Vekilov, 1999, 2001). On the whole, the non-monotonic trends in Figs. 1(*b*) and 2 converge to support a multistep mechanism of crystal nucleation that can be positively or negatively affected by protein concentration, depending on which step is rate limiting.
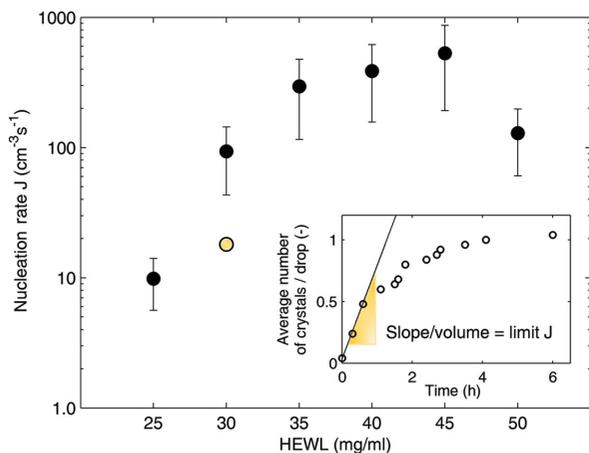
Figure 2
Influence of protein concentration on the crystal nucleation rates (J) of HEWL represented in a log–linear scale. Black circles and error bars: mean and 90% confidence intervals of J values calculated as the number of crystals formed per unit volume of drop and per unit of apparent induction time. Yellow circle: nucleation rate calculated by the method illustrated in the inset graph. Inset: average number of crystals per drop as a function of the time elapsed since the first observation of crystals for 30 mg ml^{-1} HEWL. In this example, 24 different drops of 30 mg ml^{-1} HEWL solution were followed at 289.2 K; 12.5% of the drops produced no crystals at the end of 24 h. The initial slope corresponds to the limit nucleation rate.

3.2. DLS results indicate different aggregation pathways

The hypothesis that higher C_{HEWL} values promote aggregation pathways other than the formation of precrystalline nuclei was tested using DLS data. Our DLS experiments do not fully reproduce the crystallization conditions in the submicrolitre HEWL drops, given the distinct sample volumes (1.0 ml *versus* 0.7 μl) and dominant interfaces (glass cuvette *versus* paraffin oil) in each technique. Nevertheless, DLS was used to investigate whether the results shown in Figs. 1 and 2
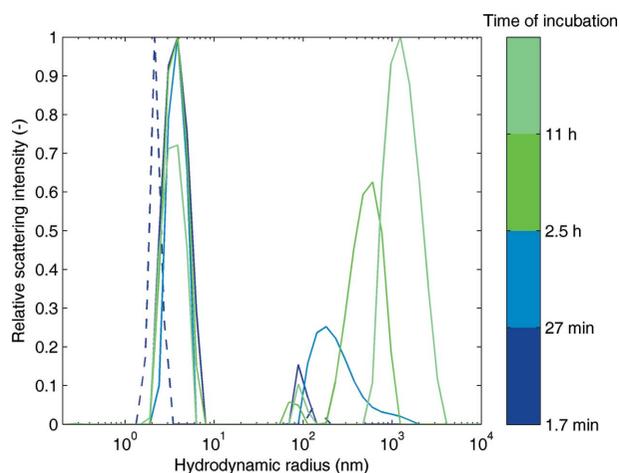
Figure 3
DLS data obtained for 40 mg ml^{-1} HEWL solutions containing 3%(w/v) NaCl at pH 4.7 and 289.2 K (solid lines) and for 50 mg ml^{-1} HEWL with no added NaCl (dashed line). Size distributions obtained from the analysis of the decay-time correlation functions using the *CONTIN* algorithm. Different colors represent different incubation times, as indicated in the color bar.

are determined by alternative aggregation pathways taking place in the solution bulk. In accordance with previous reports (Parmar *et al.*, 2007), most of the HEWL crystals were formed along the walls of the DLS cuvette and some of them were deposited at the bottom of the container. Since the population of immobilized crystals produces minor interference in the scattered light, we were able to measure the decay times of the main scattering species in solution during the pre- and post-nucleation phases of HEWL crystals. Fig. 3 confirms the presence of the expected two groups of prenucleation scatterers under moderately supersaturated conditions: soluble protein with an average hydrodynamic radius of $R_1 \simeq 3$ nm and metastable clusters with an average hydrodynamic radius of $R_2 > 100$ nm. DLS analysis of the protein solutions containing no added NaCl (dashed line in Fig. 3) shows a narrow distribution of soluble protein centered at $R_1 = 2.15$ nm, a value that is in good agreement with the estimated hydrodynamic radius of lysozyme monomers (Georgalis *et al.*, 1995; Tanaka *et al.*, 1999). These samples revealed no signs of protein aggregation upon incubation at 289.2 K. Therefore, the presence of precipitant not only decreases the protein solubility but promotes the rapid formation of small-order oligomers responsible for the increased value of R_1 by the time the first dataset is collected. In turn, the oligomerization step is followed by the formation of metastable clusters less than 6 min after mixing the protein

solutions and precipitant at C_{HEWL} between 25 and 50 mg ml^{-1}, 3%(w/v) NaCl at pH 4.7 and 289.2 K. Much longer lag phases preceding the formation of metastable clusters were observed for experiments carried out at higher temperatures (see below an example for $T = 298$ K in Fig. 6). Fig. 3 also shows a progressive decline of the relative scattering intensity of the R_1 peak as a natural result of the formation and growth of metastable clusters at the expense of protein monomers, but also as a consequence of the formation of HEWL crystals that became visible on the container walls less than 1 h after the beginning of the experiment.

3.2.1. The oligomerization pathway. Whether soluble oligomers and metastable clusters favor or oppose crystal nucleation is discussed in the following two sections on the basis of a separate analysis of the R_1 and R_2 peaks. Fig. 4 specifically addresses the influence of C_{HEWL} on the size and composition of small scattering species consisting of soluble protein in different oligomerization states. In all the cases presented in Fig. 4(*a*), stationary size distributions were attained very rapidly in spite of some initial fluctuations that had ceased within 30 min. By increasing the total protein concentration, a noticeable increase in the presence of soluble oligomers larger than 3 nm is observed. Although the DLS technique alone cannot resolve the population of the different n-mers present in solution, we combined the results in Fig. 4(*a*) with established equilibrium models comprising HEWL monomers $\leftrightarrow$ dimers $\leftrightarrow$ tetramers $\leftrightarrow$ octamers $\leftrightarrow$ 16-mers (Li *et al.*, 1995) to estimate the different oligomeric compositions for each value of C_{HEWL}. On the basis of the monomeric dimension obtained in Fig. 3 (2.15 nm), the Rayleigh law of light scattering was used to estimate the hydrodynamic sizes of dimers (2.71 nm), tetramers (3.41 nm), octamers (4.30 nm)
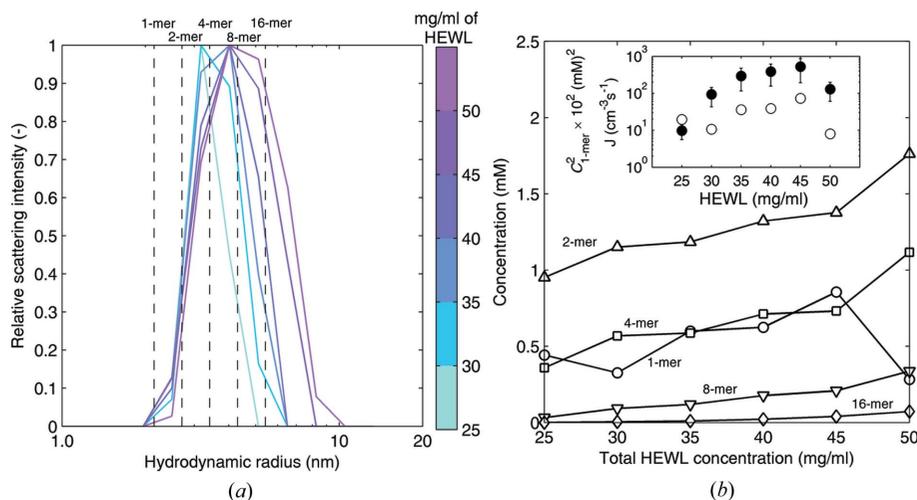
Figure 4
(*a*) Influence of the initial HEWL concentration on the size distribution of low-order protein oligomers. DLS datasets were collected after initial fluctuations had ceased during incubation of HEWL solutions containing 3%(w/v) NaCl at 289.2 K. In all cases the fluctuations ceased within 30 min of incubation. Different colors represent the different HEWL concentrations indicated in the color bar. Dashed lines: estimated hydrodynamic sizes of monomers, dimers, tetramers, octamers and 16-mers. (*b*) Oligomeric compositions calculated from the DLS data represented in (*a*) (see text for details). Inset: log–linear representation of the effect of C_{HEWL} on (solid circles) the crystal nucleation rates (J) and (open circles) the equilibrium concentrations of monomer raised to the power of 2 ($C_{1\text{-mer}}^2$); error bars indicate 90% confidence intervals of J values.

research papers

Table 1
Equilibrium constants calculated from equation (1) for different HEWL concentrations.

The mean values of $K_{n \to 2n}$ are compared with the data obtained by Li *et al.* (1995) for similar lysozyme concentrations in the presence of 3%(*w*/*v*) NaCl pH 4.6. The literature values were calculated for 289.2 K using the pre-exponential constants and reaction enthalpies obtained by Li *et al.* (1995) and assuming the crystal growth unit to be an octamer.

	C_{HEWL} (mg ml^{-1})							Li *et al.*
	25	30	35	40	45	50	Mean ± SD	(1995)
$K_{1 \to 2}$ ($M^{-1/2}$)	69.6	104.2	57.4	58.2	43.3	148.4	66.5 ± 39.3	30.3
$K_{2 \to 4}$ ($M^{-1/4}$)	4.46	4.55	4.52	4.49	4.43	4.35	4.47 ± 0.07	5.20
$K_{4 \to 8}$ ($M^{-1/8}$)	2.00	2.03	2.08	2.08	2.11	2.01	2.05 ± 0.04	2.50
$K_{8 \to 16}$ ($M^{-1/16}$)	0.00†	1.47	1.51	1.51	1.53	1.50	1.50 ± 0.02	1.83

† Value not considered in the mean calculation.

and 16-mers (5.42 nm) assuming spherical geometry. The scattering intensity corresponding to each size was directly interpolated from the distributions in Fig. 4(*a*) and normalized by the sixth power of the *n*-mer size. The intensity values thus obtained were then used to estimate the number fraction of each oligomer, which in turn was multiplied by the total protein concentration to obtain the absolute compositions shown in Fig. 4(*b*).

To compare our predictions with those resulting from previously validated models (Li *et al.*, 1995; Wilson *et al.*, 1996), the equilibrium constants $K_{n \to 2n}$ of each $n \leftrightarrow 2n$ oligomerization reaction were calculated as described by Li *et al.* (1995) using the values of the oligomeric concentration in molar (*M*) units ($C_{n\text{-mer}}$):

$$K_{n \to 2n} = \left(C_{2n\text{-mer}}^{1/2} / C_{n\text{-mer}} \right)^{1/n}, \qquad (1)$$

where *n* is an integer value between 1 and 8. Table 1 summarizes the results obtained in this work and by Li *et al.* (1995) under similar crystallization conditions. The agreement between the two estimations is worthy of note, especially taking into account the totally different methods compared and that the Li *et al.* model is supported by measurements of dialysis flux data (Wilson *et al.*, 1996). As the values of the standard deviations illustrate, different protein concentrations resulted in only minor variations in the estimated equilibrium constants. The exception to this is the dimerization constant; however, its scattered values are not affected by C_{HEWL} in any obvious way. The value of $K_{8 \to 16} = 0$ obtained for 25 mg ml^{-1} HEWL is a reasonable consequence of an insufficient concentration of 16-mers for a detectable scattering signal to be produced.

The influence of C_{HEWL} on the measured induction times and crystal nucleation rates (Figs. 1*b* and 2) can now be interpreted in the light of the oligomeric compositions presented in Fig. 4(*b*). In all the cases considered in Fig. 4(*b*), the concentration of HEWL monomers is the only value changing non-monotonically with the total protein concentration in a way that resembles the tendency followed by the crystal nucleation rates shown in Fig. 2. This is illustrated in

the inset graph of Fig. 4(*b*), where the monomer concentration is raised to the power of 2 under the assumption of a bimolecular nucleation mechanism. The similar variations of *J* and $C_{1\text{-mer}}^2$ with C_{HEWL} seem to suggest a preferential crystallization pathway departing from HEWL monomers, even though the formation of precrystalline nuclei at the expense of other oligomeric species is a possibility that cannot be excluded. Recent SAXS measurements on the oligomeric composition of HEWL under crystallization conditions identified no significant quantity of tetramers in solution (Kovalchuk *et al.*, 2016). The reexamination of the DLS data in Fig. 4(*a*) using an equilibrium model with no tetramers included did not change the main conclusions drawn thus far.

3.2.2. The cluster formation pathway. The apparent nonlinear influence of protein concentration on the induction times and nucleation rates shown in Figs. 1(*b*) and 2 is well described by the variation of monomer concentration obtained in Fig. 4(*b*). However, this does not yet determine whether protein aggregation into metastable clusters also affects the measured nucleation kinetics and, above all, why there are crystal-productive and nonproductive outcomes for experiments carried out under replicate crystallization conditions. To answer the first question, HEWL concentrations between 25 and 50 mg ml^{-1} did not produce major differences in the initial dimensions of the observed mesoscopic clusters [$R_2(0) \simeq 100$ nm] or in the cluster growth rates (fractal dimensions $d_f \simeq 2.0$) during the precrystallization phase (initial 30 min). Consequently, cluster formation does not seem to account for the atypical nucleation-rate curves, which require uneven kinetic effects in order to be explained. The cluster pathway might, however, explain the unsteady-state nucleation rates characterized by decreasing values of *J* as time elapses (inset graph in Fig. 2). As illustrated in Fig. 5(*a*), cluster growth precedes the formation of HEWL crystals and continues to occur in parallel with crystal growth. Thus, the progressive depletion of soluble protein is the result of two competing pathways producing, in one case, metastable clusters and, in the other, protein crystals. We hypothesize that the unsteady-state nucleation of HEWL crystals follows from a gradual decrease in the concentration of soluble protein as a result of the noncrystallizing pathway. In this way, the natural variability of stochastic nucleation is amplified so that long induction times become even longer owing to the continuous loss of free protein molecules in solution. After a certain period of time devoid of crystalline nuclei, the concentration of soluble protein reaches a limit below which crystal nucleation is not possible anymore. This model suggests that productive and unyielding nucleation regimes result from a common molecular mechanism that involves both on- and off-pathway protein aggregation, and that the different experimental outcomes represent the effects of stochastic nucleation further amplified by the continuous depletion of soluble protein during cluster growth.

The variations of the relative weight of the R_1 peak (Fig. 5*b*) and of the total scattering intensity (Fig. 5*c*) primarily reflect the growth of metastable clusters. While the growth of crystals spontaneously attached to the container walls is not expected

to directly interfere with the scattered light, the subsequent decrease in the concentration of soluble protein will contribute to a reduction in the total scattering intensity. This effect is, mostly, hidden in Fig. 5(c) by the prevailing influence of cluster growth. At the end of ~5 h, the metastable clusters exceed the submicrometre size (Fig. 5a) and start sedimenting, as indicated by the abrupt decrease in the total scattering intensity observed in Fig. 5(c). Since the weight % of the R_1 peak and the total scattering intensity decrease simultaneously during this period, we conclude that the nucleated crystals continue to grow after 5 h of incubation (also confirmed by time-lapse imaging).
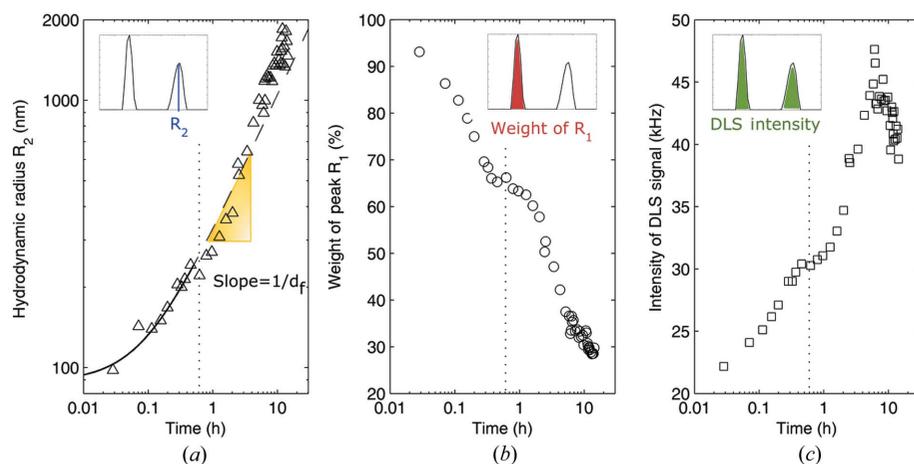
Figure 5
Growth of metastable clusters in solutions of 40 mg ml^{-1} HEWL, 3%(w/v) NaCl at pH 4.7 and 289.2 K (supersaturation ratio ~8.9). (a) Average hydrodynamic radius of metastable clusters, R_2, represented as a function of time in a log–log scale. Solid line: nonlinear least-squares fit of the power law $R_2 = R_2(0)(1 + \text{constant} \times \text{time})^{1/d_f}$ to the cluster growth data measured during the precrystallization phase (Georgalis et al., 1995). Dashed line: extrapolation of the fitted curve to later aggregation phases; the fractal dimension, d_f, corresponds to the reciprocal of the slope ($d_f = 1.93$). (b), (c) Linear–log plots of (b) the % weight of the low-order oligomer protein peak, R_1, and (c) the total light scattered, as a function of time. Vertical dotted lines indicate the time of formation of the first crystalline material as estimated by visual inspection of the DLS cuvette.
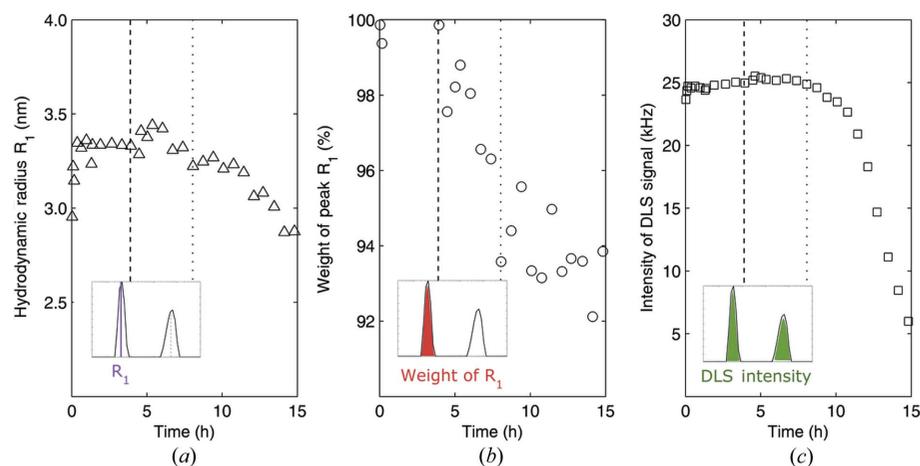
Figure 6
Protein aggregation in solutions of 50 mg ml^{-1} HEWL, 3%(w/v) NaCl at pH 4.7 and 298.2 K (supersaturation ratio ~4.5). Time evolution of the (a) average hydrodynamic radius of the soluble protein R_1, (b) % weight of the soluble protein peak and (c) total light scattered. Vertical lines indicate the estimated time of formation of the first (dashed lines) metastable clusters and (dotted lines) crystalline material.

Despite the widespread formation of 'off-pathway' clusters in the DLS experiments, numerous crystals were always observed in the glass cuvette walls a few hours after the solutions were mixed. This emphasizes the role of interfaces during the crystallization of proteins (Ildefonso et al., 2013; Silver et al., 2011) given that 0.7 µl droplets of equivalent HEWL solutions showed much lower success rates (Fig. 1a). On the other hand, metastable clusters seem to be more than just competing aggregates as their presence always precedes the formation of HEWL crystals. This aspect is further elucidated by the DLS analysis of crystallization solutions associated with a lower nucleation probability. To obtain the results shown in Fig. 6, the supersaturation level was lowered by increasing the incubation temperature to 298.2 K. As a first consequence, the initial average size of the low-order protein oligomers decreased from $R_1 \simeq 4.4$ nm at 289 K to $R_1 \simeq 2.9$ nm at 298 K. Also, in contrast to the behavior observed at 289.2 K, Fig. 6(a) shows a progressive increase of R_1 with time, reaching a plateau value of ~3.3 nm at the end of the first hour of incubation, followed by a slight increase to ~3.4 nm simultaneously with the apparent formation of the first metastable clusters ~4.5 h after the beginning of the experiment. The increasing values of R_1 suggest the gradual formation of soluble oligomers such as tetramers, octamers or 16-mers, which are apparently required for the production of metastable clusters, as demonstrated by the subsequent return of R_1 to values below 3.4 nm. The ~4 h period that succeeded the appearance of cluster aggregates is characterized by a gradual decrease of ~5% in the relative weight of the first peak (Fig. 6b) and by approximately constant values of the total scattering intensity (Fig. 6c). This is in contrast with the mirrored trends previously observed in Figs. 5(b) and 5(c), and is indicative of a less extensive cluster formation at 298.2 K. Because the weight % of the R_1 peak steadily decreases with no significant variation in the total light scattered, we can conclude that metastable clusters grow at the expense of soluble protein and not by the self-aggregation of smaller clusters. This type of mechanism was difficult to discern from the data in Figs. 5(b) and 5(c) since the predominant influence of cluster growth on the total scattering intensity rendered the

experiment less sensitive to variations in the number of monomers and small oligomers.

The decreasing values of the scattering signal in Fig. 6(c) reflect the formation of HEWL crystals, which were observed to grow in lower number but with larger average sizes than in any of the experiments carried out at 289.2 K. At 298.2 K, the soluble protein is primarily consumed by the crystal growth process since the total scattering intensity hardly increased during the preceding period of cluster formation and markedly decreased after crystal nucleation. The latter phase is accompanied by a decrease of the average size of soluble protein to values of R_1 below 3 nm (Fig. 6a), thus confirming the preferential incorporation of oligomers, most probably octamers (Kovalchuk et al., 2016; Li et al., 1995), during crystal growth. Given the reduced number of mesoscopic clusters in solution, accurate R_2 dimensions and rates of cluster growth could not be determined at 298.2 K. The correspondence between the number of clusters in solution and the number of crystals produced at 289.2 and 298.2 K, and the fact that crystallization is always preceded by the formation of metastable clusters in solution, are both in line with previous reports of heterogeneous nucleation of HEWL crystals induced by the presence of the precursor aggregates (Parmar et al., 2007; Saridakis et al., 2002).

3.3. Possible mechanism of intermittent crystallization

Rationalizing the phenomenon of intermittent protein crystallization is now possible on the basis of the measured crystal nucleation rates and protein aggregation dynamics. Fig. 7 summarizes the on- and off-pathways, illustrating the coexistence of productive and unproductive crystallization regimes. Some of the oligomers formed during the first step will later incorporate into the crystal lattice, while others will participate in the nucleation and growth of metastable clusters. As such, the oligomerization step is simultaneously on- and off-pathway for crystal formation. Crucial for the final outcome of the crystallization trial, the concentration of monomers in the moments that follow cluster formation determines whether the heterogeneous nucleation of crystals is possible or not. The threshold concentration value is difficult to define accurately since the stochastic variability in the time required for monomer assembly is continuously amplified by the parallel depletion of monomers.

The formation of mesoscopic clusters is a nucleation process whose induction time – a few minutes in Fig. 5, and >4 h in Fig. 6 – seems to be supersaturation dependent. The succeeding cluster growth phase occurs by the incorporation of soluble protein monomers and low-order oligomers and not by the coalescence of smaller aggregates. With hydrodynamic sizes rapidly exceeding the submicrometre range (Fig. 5a), the metastable clusters are in permanent competition for soluble protein, but they also act as heterogeneous nucleation centers for the formation of the precrystalline nuclei. While the role of interfaces (air/solution, oil/solution, glass/solution etc.) has not been considered in the schematic mechanism in Fig. 7, concentration gradients created close to the boundary regions

can promote heterogeneous nucleation (Martins et al., 2008), alter the local effective concentrations of protein and precipitant (García-Ruiz et al., 2016), and/or accelerate protein loss by denaturation (Ferreira et al., 2016). Liquid–liquid separation occurring at higher supersaturation levels (Muschol & Rosenberger, 1997; Vekilov & Vorontsova, 2014; Pan et al., 2010) is also absent from the simplified crystallization model represented in Fig. 7. Care must be taken in expanding the mechanism of intermittent crystallization to other proteins. For example, it may not always be the case that crystal nucleation only occurs through the initial assembly of monomeric molecules as suggested by the inset graph in Fig. 4(b). We speculate, however, that heterodisperse populations of soluble protein are more prone to self-aggregate into anisotropic clusters than to form highly organized precrystalline structures.

Our findings raise fundamental questions about the way protein crystallization is understood. Are the energetic barriers for nucleation and growth significantly different for monomers and oligomers? Should the definition of supersaturation change according to the oligomeric state of the macromolecules? To what extent is the nucleation of protein

1. Oligomerization
2. Cluster nucleation
3. Crystal nucleation
4. Cluster growth
5. Crystal growth

CRYSTALS   **NO CRYSTALS**
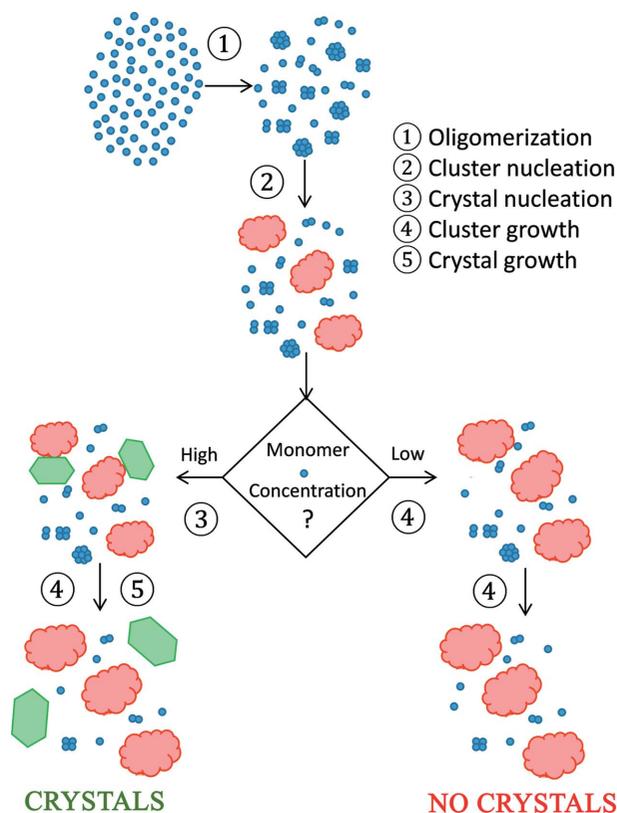
Figure 7
Unsteady-state nucleation of HEWL crystals (step 3) is contingent on the earlier steps of protein oligomerization (1) and cluster nucleation (2). Metastable clusters (red clouds) act as nucleants of HEWL crystals (green polygons) as long as the monomer concentration is above the critical limit for heterogeneous nucleation. In these cases, cluster growth (step 4) will compete with crystal growth (step 5) for the population of monomers (single dots) and oligomers (grouped dots). However, if the concentration of monomers is too greatly depleted by steps 1 and 2, no HEWL crystals will be produced.

crystals heterogeneous? The answers to these questions require more work to be done on the characterization of the kinetic signatures associated with different compositions of low- and high-order aggregates (Crespo *et al.*, 2017). Some aspects of nonclassical nucleation mechanisms can be unveiled with the help of advanced microscopic techniques (Fermani *et al.*, 2013). For now, it is important to note the critical role of protein oligomers and metastable clusters in the success of the crystallization step in structural biology pipelines. Since the concentration of intermediate and off-pathway species is markedly influenced by elemental parameters such as protein concentration, ionic strength and temperature, a change of paradigm is conceivable where the search for crystallization conditions becomes more centered in the biophysical screening space and less dispersed over an infinite number of chemical possibilities. The miniaturization of fluorescence (Ericsson *et al.*, 2006; Reinhard *et al.*, 2013; Ristic *et al.*, 2015) and DLS techniques (Bolanos-Garcia & Chayen, 2009; Dierks *et al.*, 2008) should facilitate a rational high-throughput strategy comprising a first round of selection of the better chemical environment favoring protein stability (using, for example, fluorescent dyes to measure the denaturation midpoint) and, then, a second round of screening searching for the protein concentration, precipitant/additive concentration and temperature that lead to the formation of the right number of precursor protein aggregates. In this respect, the present results highlight the negative correlation between crystallizability and sample polydispersity (Zulauf & D'Arcy, 1992): pre-existing off-pathway aggregates deplete the limited pool of soluble protein and decrease the crystallization chances right from the beginning of the experiment. We have also demonstrated that variations in the oligomerization state caused by different solution parameters can be followed by DLS. A mild degree of protein oligomerization is beneficial in order to trigger the sequence of events leading to the heterogeneous nucleation of well differentiated crystals in the presence of a small number of metastable clusters.

4. Conclusions

Protein crystallization is known to be a poorly reproducible process (Luft *et al.*, 2011, 2014; Gorrec, 2013; Newman *et al.*, 2007) in which the success rates markedly decrease with incubation time (Ng *et al.*, 2016; Newman *et al.*, 2007). We studied conditions of unsteady-state nucleation of HEWL crystals and show that the erratic outcomes are the result of off-pathway aggregation involving the sequential formation of soluble oligomers and metastable clusters. The existence of parallel steps was first suggested by the non-monotonic influence of HEWL concentration on the apparent induction times and crystal nucleation rates. By comparing the measured nucleation kinetics with the oligomeric composition obtained for similar conditions using DLS, we identified a correlation between the nucleation-rate curves and the distribution of low-order HEWL oligomers. This correspondence was not affected by different rates of metastable cluster formation, suggesting that crystal nuclei are formed by the assembly of

protein monomers. Detailed analysis of time-resolved DLS data showed that both cluster and crystal growth occur at the expense of soluble protein; however, the formation of metastable clusters always preceded crystal nucleation, and the final number of crystals correlated with the number of precursor aggregates. These observations reveal a dual role of metastable clusters as both competitors and heterogeneous nucleants of protein crystals. Our findings help to understand the striking differences between small-molecule crystallization and protein crystallization, and why the latter remains a major obstacle in structural biology projects. They suggest that alternative high-throughput strategies can be rationally devised, focusing on windows of conditions that simultaneously favor protein stability and promote the formation of the right quantity of precursor protein aggregates.

References

Akella, S. V., Mowitz, A., Heymann, M. & Fraden, S. (2014). *Cryst. Growth Des.* **14**, 4487–4509.

Bolanos-Garcia, V. M. & Chayen, N. E. (2009). *Prog. Biophys. Mol. Biol.* **101**, 3–12.

Crespo, R., Villar-Alvarez, E., Taboada, P., Rocha, F. A., Damas, A. M. & Martins, P. M. (2016). *J. Biol. Chem.* **291**, 2018–2032.

Crespo, R., Villar-Alvarez, E., Taboada, P., Rocha, F. A., Damas, A. M. & Martins, P. M. (2017). *J. Phys. Chem. B*, **121**, 2288–2298.

Dierks, K., Meyer, A., Einspahr, H. & Betzel, C. (2008). *Cryst. Growth Des.* **8**, 1628–1634.

Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.

Fermani, S., Vettraino, C., Bonacini, I., Marcaccio, M., Falini, G., Gavira, J. A. & Garcia Ruiz, J. M. (2013). *Cryst. Growth Des.* **13**, 3110–3115.

Ferreira, C., Rocha, F. A., Damas, A. M. & Martins, P. M. (2012). *Cryst. Growth Des.* **12**, 40–43.

Ferreira, C., Rocha, F. A., Damas, A. M. & Martins, P. M. (2013). *J. Cryst. Growth*, **368**, 47–55.

Ferreira, C., Rocha, F. A., Damas, A. M. & Martins, P. M. (2016). *Cryst. Growth Des.* **16**, 4285–4291.

Galkin, O. & Vekilov, P. G. (1999). *J. Phys. Chem. B*, **103**, 10965–10971.

Galkin, O. & Vekilov, P. G. (2001). *J. Cryst. Growth*, **232**, 63–76.

García-Ruiz, J. M. (2003). *J. Struct. Biol.* **142**, 22–31.

García-Ruiz, J. M., Otálora, F. & García-Caballero, A. (2016). *Acta Cryst.* F**72**, 96–104.

Georgalis, Y., Schüler, J., Frank, J., Soumpasis, M. D. & Saenger, W. (1995). *Adv. Colloid Interface Sci.* **58**, 57–86.

Gorrec, F. (2013). *J. Appl. Cryst.* **46**, 795–797.

Ildefonso, M., Candoni, N. & Veesler, S. (2013). *Cryst. Growth Des.* **13**, 2107–2110.

Kovalchuk, M. V., Blagov, A. E., Dyakova, Y. A., Gruzinov, A. Y., Marchenkova, M. A., Peters, G. S., Pisarevsky, Y. V., Timofeev, V. I. & Volkov, V. V. (2016). *Cryst. Growth Des.* **16**, 1792–1797.

Li, M., Nadarajah, A. & Pusey, M. L. (1995). *J. Cryst. Growth*, **156**, 121–132.

Liu, Y., Wang, X. & Ching, C. B. (2010). *Cryst. Growth Des.* **10**, 548–558.

Luft, J. R., Newman, J. & Snell, E. H. (2014). *Acta Cryst.* F**70**, 835–853.

Luft, J. R., Wolfley, J. R. & Snell, E. H. (2011). *Cryst. Growth Des.* **11**, 651–663.

Martins, P., Pessoa, J., Sàrkàny, Z., Rocha, F. & Damas, A. (2008). *Cryst. Growth Des.* **8**, 4233–4243.

Muschol, M. & Rosenberger, F. (1997). *J. Chem. Phys.* **107**, 1953–1962.

Newman, J., Xu, J. & Willis, M. C. (2007). *Acta Cryst.* D**63**, 826–832.

Ng, J. T., Dekker, C., Reardon, P. & von Delft, F. (2016). *Acta Cryst.* D**72**, 224–235.

Onuma, K. & Kanzaki, N. (2007). *J. Cryst. Growth*, **304**, 452–459.

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

Pan, W., Vekilov, P. G. & Lubchenko, V. (2010). *J. Phys. Chem. B*, **114**, 7620–7630.

Parmar, A. S., Gottschall, P. E. & Muschol, M. (2007). *Biophys. Chem.* **129**, 224–234.

Provencher, S. W. (1982). *Comput. Phys. Commun.* **27**, 213–227.

Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. (2013). *Acta Cryst.* F**69**, 209–214.

Ristic, M., Rosa, N., Seabrook, S. A. & Newman, J. (2015). *Acta Cryst.* F**71**, 1359–1364.

Saridakis, E., Dierks, K., Moreno, A., Dieckmann, M. W. M. & Chayen, N. E. (2002). *Acta Cryst.* D**58**, 1597–1600.

Schubert, R., Meyer, A., Dierks, K., Kapis, S., Reimer, R., Einspahr, H., Perbandt, M. & Betzel, C. (2015). *J. Appl. Cryst.* **48**, 1476–1484.

Silver, B. R., Fülöp, V. & Unwin, P. R. (2011). *New J. Chem.* **35**, 602–606.

Tanaka, S., Ito, K., Hayakawa, R. & Ataka, M. (1999). *J. Chem. Phys.* **111**, 10330–10337.

Tyka, M. D., Keedy, D. A., André, I., Dimaio, F., Song, Y., Richardson, D. C., Richardson, J. S. & Baker, D. (2011). *J. Mol. Biol.* **405**, 607–618.

Veesler, S., Lafferrère, L., Garcia, E. & Hoff, C. (2003). *Org. Process Res. Dev.* **7**, 983–989.

Vekilov, P. G. (2016). *Prog. Cryst. Growth Charact. Mater.* **62**, 136–154.

Vekilov, P. G. & Vorontsova, M. A. (2014). *Acta Cryst.* F**70**, 271–282.

Vorontsova, M. A., Maes, D. & Vekilov, P. G. (2015). *Faraday Discuss.* **179**, 27–40.

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

Wilson, L., Adcock-Downey, L. & Pusey, M. (1996). *Biophys. J.* **71**, 2123–2129.

Wilson, W. W. (2003). *J. Struct. Biol.* **142**, 56–65.

Zulauf, M. & D'Arcy, A. (1992). *J. Cryst. Growth*, **122**, 102–106.

LEPABE - Laboratory for Process Engineering, Environment, Biotechnology and Energy