

Precrystallization clusters of holoferritin and apoferritin at low temperature

S. Boutet* and I. K. Robinson†

Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

(Received 29 September 2005; revised manuscript received 2 January 2007; published 21 February 2007)

The formation of small nanosized clusters of the proteins holoferritin and apoferritin at low temperature was studied using small angle x-ray scattering. A strikingly large temperature dependence for the average molecular spacing in the clusters was observed. Calculations of the scattered intensity for various cluster models were performed. Comparison of the data with the simulations revealed the presence of crystalline order in the clusters of size ranging from a few molecules to a few hundred molecules. The crystalline order was found to be preserved with the lattice spacing varying with temperature by up to 20%. The small clusters were observed to grow into large micron-sized crystals when they were annealed and under certain conditions, the small clusters were found to coexist with the large crystals. This suggests that these clusters are closely related to critical nucleation. The data are consistent with an isotropic nucleation pathway, but cannot completely rule out a smaller presence of planar nucleation.

DOI: [10.1103/PhysRevE.75.021913](https://doi.org/10.1103/PhysRevE.75.021913)

PACS number(s): 87.15.Nn, 61.10.Eq, 87.64.Bx, 81.10.-h

I. INTRODUCTION

The standard method to determine the atomic structure of biological macromolecules such as proteins is crystallography in which a crystal of the macromolecule is prepared using some standard techniques [1]. This crystal is irradiated by an x-ray beam, the diffraction pattern is measured and then inverted to obtain a three-dimensional atomic structure. Much of the difficulty in the process lies in obtaining good quality crystals which diffract well enough at atomic resolution. For many known proteins, attempts to find conditions which will yield crystals have failed. The search for crystallization conditions is generally carried out by trial and error resulting in painstaking and sometime unsuccessful work.

For this reason, it is worthwhile to study the protein crystallization process itself in order to understand what causes some proteins to crystallize while others do not. Also, there is hope that such studies can provide some kind of ability to predict which proteins will crystallize and under which conditions they are likely to do so [2,3].

There are two distinct stages in the crystal formation process. The first one is known as the nucleation process in which stable aggregates of proteins form from a saturated solution [4–6]. The second stage is called the growth process. It is during this latter process that the stable nuclei grow into large crystals, often by adding layers upon layers to the crystal [7,8].

The nucleation process is the less well understood of the two and the one generally thought to be responsible for the existence or nonexistence of crystals. Crystals cannot grow if there are no stable initial aggregates formed onto which other proteins can leave solution and attach to. Crystals are also prevented from growing when the nucleus is a noncrystalline aggregate. The stable nanocrystalline aggregates are called critical nuclei because when they reach a certain size, they become more likely to grow than they are likely to redissolve.

The size of the critical nuclei depends on the protein and also on the level of supersaturation of the solution [1]. They range from very small (a few molecules) at high saturation to infinite for solutions below saturation. In the usual case where one wants to produce a few large high quality crystals in a few days, the level of supersaturation needs to be adjusted in such a way that only a few critical nuclei are formed over this period. It is then very difficult to study critical nuclei because of their rare occurrence. Once formed, however, they should be of a size large enough to study with some standard techniques such as AFM [9]. However, those nuclei are floating around in solution which makes them visible to the AFM only when they attach to an existing substrate such as a large crystal. Observing isolated nuclei in solution as they form is very difficult.

One possible technique to study protein nucleation is coherent x-ray diffraction (CXD) [10,11], which allows the measurement of a complete diffraction pattern of a crystalline nanoparticle. This diffraction pattern can be inverted to yield the complete shape and internal structure of the particle. CXD should allow the study of isolated critical nuclei of size below 1 μm . It is currently difficult to perform such experiments since they are limited by x-ray flux, detector speed, and radiation damage.

During an attempt at performing one of those CXD experiments, a solution containing thousands of micron sized crystals of the protein ferritin in equilibrium with the solution phase was frozen in order to try to mitigate the effects of radiation damage as well as slow down the motions of the crystals in solution. A new aggregated state of ferritin was found to exist, as was previously reported by Kilcoyne *et al.* [12].

In this paper we investigate this freezing transition and relate the frozen state to nucleation. The samples and techniques used for the measurements are first detailed in Sec. II. The scattering results obtained with the protein holoferritin are then presented in Sec. III along with how the scattered intensity and structure factors vary with temperature and protein concentration of the solution. In Sec. IV, we present similar results for apoferritin solutions. Simulations of the scattering structure factors of small isotropic and planar clus-

*Electronic address: sboutet@slac.stanford.edu

†Electronic address: i.robinson@ucl.ac.uk

ters of various size are presented in Sec. V. Section VI discusses the fitting of the measured patterns to the simulated ones and finally Sec. VII discusses how the results relate to nucleation.

II. SAMPLE AND EXPERIMENTAL METHOD

The protein ferritin extracted from horse spleen was obtained from the Sigma-Aldrich Company. It is composed of 24 subunits arranged in a 432 symmetry forming a spherical shell of inner diameter ~ 80 Å, outer diameter ~ 130 Å, and $MW_t=474\,000$. It has the biological function of storing iron as ferrihydrite in its cavity and releasing it when it is needed by the organism [13]. It readily crystallizes into the face centered cubic (fcc) lattice with lattice parameter $a=183$ Å when cadmium salt is added. Two Cd^{2+} ions form salt bridges at every twofold axis of the protein [14]. Both the iron filled protein (holoferritin or sometimes called simply ferritin) and the empty shell (apoferritin) were studied. Holoferritin was obtained at a concentration of 100 mg/ml in 150 mM NaCl while apoferritin was obtained at a concentration of 50 mg/ml in 100 mM NaCl. Holoferritin is about 55% more massive than apoferritin due to the presence of iron. Solutions containing a higher concentration of proteins were obtained by filtering through an Amicon Ultra 10 000 MWCO centrifugal filter from Milipore.

Solutions of ferritin were placed, as bought, inside x-ray quartz capillaries of diameter 1 mm. The capillaries were glued using thermal paste to a Peltier cooling device which was mounted on a water-chilled copper block serving as a heat sink. This setup allowed the temperature of the sample to be varied from 15 to -35 °C. A K-type thermocouple was used to measure the temperature of the sample.

The samples were placed into the x-ray beam at Sector 34-ID-C at the Advanced Photon Source (APS) at Argonne National Laboratory. The small angle x-ray scattering (SAXS) patterns from the samples were collected using a direct-read CCD camera from Roper Scientific placed at various distances from the sample. The x-ray energy from the undulator was set to 9 keV and further monochromatized using a double crystal Si(111) monochromator. Beam-defining slits of $200 \times 200 \mu m^2$ were used 1 m upstream of the sample with cleanup slits 10 mm upstream of the sample.

III. HOLOFERRITIN RESULTS

The SAXS pattern observed from a solution of iron-loaded ferritin at room temperature shows a monotonic decrease in intensity versus momentum transfer [$q = 4\pi \sin(\theta/2)/\lambda$], where θ is the scattering angle and λ is the wavelength of the x rays. This is due to the fact that the iron core varies in size and shape from molecule to molecule. There are no observable features such as oscillations in the SAXS pattern due to the incoherent summation of diffraction patterns from effectively different molecules. On the other hand, apoferritin molecules are all identical and have a roughly spherically symmetric structure giving rise to Bessel-like function oscillations in the SAXS data [15]. The difference between holoferritin and apoferritin can be seen in

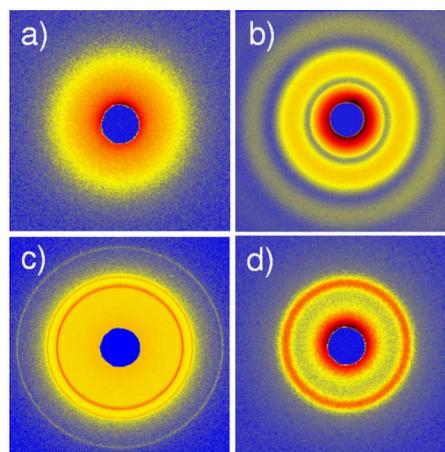


FIG. 1. (Color online) Two-dimensional SAXS data from ferritin (a) and apoferritin (b). The oscillations seen for apoferritin are washed away in holoferritin by the presence of the iron core. (c) SAXS pattern of ferritin to which Cd salt was added to produce many small crystals. (d) Solution of ferritin frozen to -20 °C.

Figs. 1(a) and 1(b). Also shown in Fig. 1(c) is the SAXS data from ferritin to which was added 90 mM $CdCl_2$ salt which was found to produce a shower of micron-sized crystals. Sharp rings of intensity can be seen as expected from a powder diffraction pattern. The Bragg peaks seen are the $\{111\}$, $\{200\}$, and $\{220\}$, the first three peaks expected from an fcc crystal of ferritin. Figure 1(d) shows what happens to a solution of holoferritin when it is frozen to -20 °C. The continuously decreasing intensity with q is replaced by a broad peak which is much broader than the powder peak from Fig. 1(c) but roughly at the same value of q as the $\{111\}$ Bragg peak.

A series of SAXS patterns from holoferritin were measured at many temperatures between 10 and -25 °C for a 200 mg/ml holoferritin sample in 150 mM NaCl. A significant change in the position of the broad peak as well as the width and height of this peak can be seen as the temperature is varied. This is shown in Fig. 2 where the two-dimensional data were integrated to yield the curves shown using the

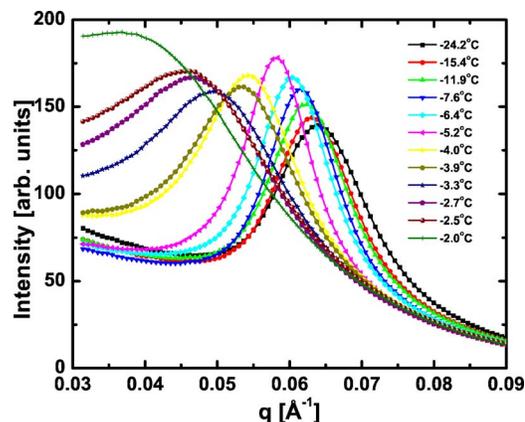


FIG. 2. (Color online) SAXS data at indicated temperatures showing the change in scattered intensity from holoferritin solution at 200 mg/ml upon freezing and as the temperature is varied.

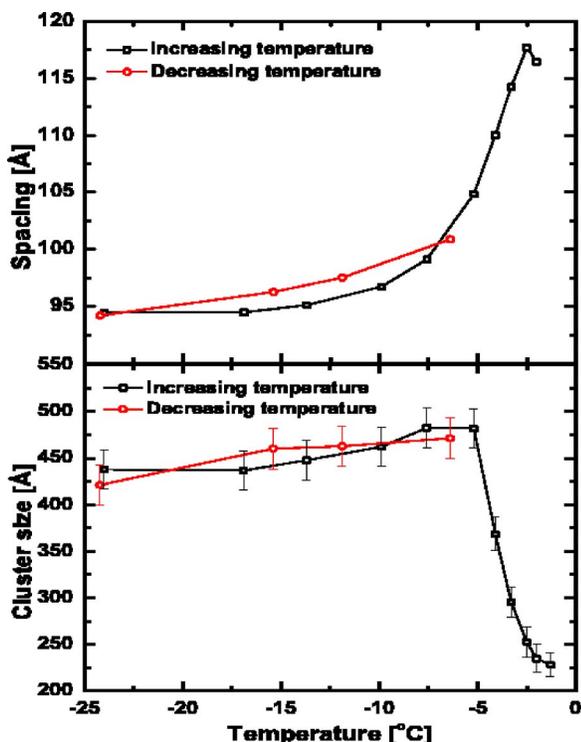


FIG. 3. (Color online) Top: Average spacing of the ferritin molecules (within the clusters) versus temperature. The {111} lattice spacing of fcc ferritin crystals is 106 Å. Thermal expansion of 30% is seen over the range of temperature. Bottom: Average size of the clusters obtained from inverting the width of the peaks plotted versus temperature.

program FIT2D [16]. The solution was found to freeze at ~ -7.5 °C as could be seen from the opaqueness of the sample but would only melt back at ~ -2 °C. Considering the relatively small solute concentration in the sample, this large depression of the freezing point appears to be an indication of supercooling. It is conceivable that the presence of protein molecules and salt make it easier to supercool the solution, allowing the lower temperatures to be reached without ice nucleation. Plenty of time was given to the temperature to equilibrate when it was changed and this difference between the freezing and melting points was reproducible. The data at temperatures between -2 and -7.5 °C was collected by heating the sample from a lower temperature, while the other data points were collected while cooling the sample from a higher temperature. There was significant hysteresis observed and the direction of the change in temperature was very much relevant.

The presence of the peak in the scattering data indicates that the proteins organize differently in the frozen solution than they do at room temperature. The peak is seen not to be due to the form factor of the molecules since it is not present at room temperature. Also, the value of q at which the peaks occur clearly indicates that it is the ferritin molecules and not ice crystals which cause the peak.

From the position of the peaks, we can extract a spacing of the molecules in the aggregates that are formed. This is shown on the top panel of Fig. 3 for all the temperatures studied. There is a strikingly large thermal expansion. The

spacing between the molecules changes $\sim 20\%$ over a range of only 25 °C. The spacing at low temperature is smaller than the spacing of the {111} lattice planes in the fcc crystal which is 106 Å.

The width of the peaks gives information about the average size of the individual aggregates using the Scherrer formula. The size is given by $2\pi/\text{FWHM}$ where FWHM is the full width at half the maximum of the peak. This value is plotted versus temperature on the bottom panel of Fig. 3. The values range between 200 and 600 Å which corresponds to clusters of 6 and 36 particles, respectively, assuming the clusters are roughly spherical.

We explain the observed behavior by a phase separation of the sample into domains which are rich in protein and domains almost devoid of proteins, a well-known liquid-liquid phase separation [17]. When in the liquid state, the sample can be cooled below the freezing point of water due to the high protein and salt content. Once it freezes, a network of ice crystals is formed. The protein molecules do not fit very well into the structure of ice and are pushed out to allow the formation of hydrogen bonds in the ice network. Voids are created in the ice and the protein molecules are trapped in them. As the temperature is increased, the protein content in the ice is actually very low due to the phase separation. The melting point is then much closer to the actual melting point of pure water. This explains the hysteresis observed in the freezing point temperature.

In the frozen solution, the protein molecules are much more concentrated because they are trapped into small volumes and are possibly under considerable hydrostatic pressure. The concentration in fcc crystals of ferritin is on the order of 650 mg/ml. At such high concentration, the molecules are very close to each other giving rise to a peak in the SAXS data. The peak position is inconsistent with simple liquidlike ordering. Such a situation would yield a peak in intensity at a q value corresponding to a spacing equal or slightly larger than the size of the protein, i.e., 130 Å corresponding to a peak at $q=0.048$ Å⁻¹. The presence of the peak at a higher q value indicates the presence of some level of crystalline ordering. Since ferritin crystallizes so readily, it becomes more thermodynamically favorable for the molecules to rearrange into a structure with some form of long range order after phase separation. In the fcc crystal of ferritin, the nearest neighbor distance is 130 Å in the $\langle 110 \rangle$ directions but the {110} reflections do not exist due to the exact cancellation in the unit cell. It is therefore clear from the data that there is some level of crystalline order present in the frozen samples. The position of the peak is generally near the expected {111} peak of the fcc crystal, at $q=0.059$ Å⁻¹ suggesting a structure similar to fcc is present.

The thermal expansion occurs because of changes in the ice network with changing temperature. As the temperature decreases, the voids in the ice where the proteins are trapped changes. Some of the remaining water in the voids freezes and forces more proteins into smaller voids. This has the effect of exerting pressure on the small clusters of proteins. It is known that protein crystals are very soft compared with inorganic crystals and also ice [18–20]. A small pressure applied to protein crystals by the ice will greatly affect their structure. This combined with the increased caging forces the

proteins closer together. As the temperature is increased, the pressure is relaxed and the protein molecules start to pull apart, as observed. Increasing the average volume of the voids in the ice by a factor of 2 would lead to an increase in molecular spacing of 20%.

The peak position, as mentioned above, corresponds to a spacing close to the $\{111\}$ fcc lattice spacing expected for ferritin crystals. However, the peak is seen to move to higher q than the value for the $\{111\}$ peak. This would seem to indicate that the molecules are spaced closer than the $\{111\}$ crystal planes. This is surprising since the fcc ferritin crystals are close-packed structures and therefore very little contraction should be possible without compressing the molecules themselves. This might be an indication that the clusters are crystalline but of a different lattice structure. It has been hypothesized that critical nuclei that lead to the formation of crystals could have a different lattice structure than the final crystals. This would at first glance appear to be the case here. However, closer inspection and simulations discussed later in this paper do not fully support the hypothesis that the clusters possess a different crystal lattice than fcc.

A. Structure factors

To better understand quantitatively the observed behavior, it is useful to examine the structure factors of the solutions measured rather than the intensity. The scattered intensity from a solution of identical particles can be separated in two factors, the form factor $F(q)$ and the structure factor $S(q)$ [21].

$$I(q) = [F(q)S(q)]^2. \quad (1)$$

The form factor is the Fourier transform of the electron density of one molecule. This contribution can be divided out from the measured intensity to obtain the structure factor, which is the Fourier transform of the spatial distribution of the molecules. In the dilute limit, the distribution is essentially a delta function because each particle scatters incoherently and this gives $S(q)=1$. This is roughly the case for the unfrozen sample in the q range measured. However, there is a large contribution to the intensity profile from $S(q)$ when the apoferritin solution is frozen.

The contribution to the scattered intensity from the electron density of the individual molecules, the form factor, can be factored out if the molecules are isotropic and identical. What is left is the structure factor of the sample which depends in principle solely on the arrangement of the molecules with no dependence on their actual shape. The form factor was measured using a dilute solution of the protein studied, making the explicit assumption that at low concentration the structure factor was exactly equal to 1 and the scattered intensity was therefore the square of the structure factor.

The factorization assumption in Eq. (1) that the protein molecules are isotropic and identical breaks down in the case of holoferritin due to the presence of the iron cores which are not all identical. Direct calculation of the 3D form factor of apoferritin from the known atomic structure has shown it to be nearly isotropic and well described by a spherical shell.

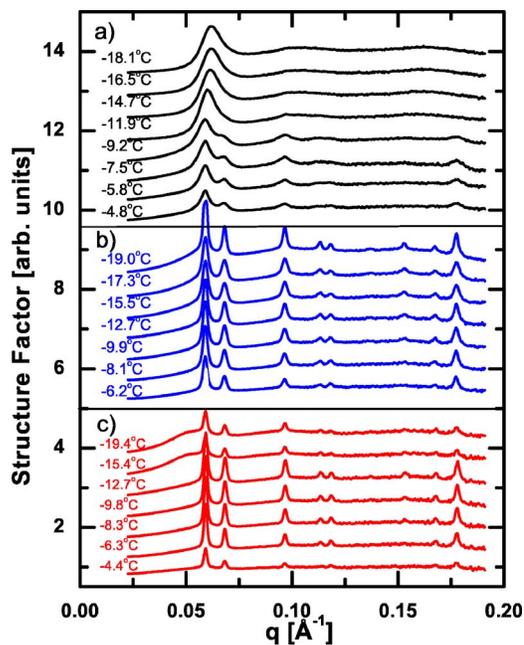


FIG. 4. (Color online) Structure factors at various temperature for holoferritin solutions at (a) 100 mg/ml, (b) 50 mg/ml, and (c) 35 mg/ml. In each case, the first temperature measured was the lowest, shown on top. As the temperature is increased, the peaks get narrower indicating an annealing process creating larger crystals. The overall cluster size is smaller as the concentration increases as seen by the broader peaks at 100 mg/ml.

However, the iron cores from different protein molecules are all different and therefore the intensity is no longer factorizable. Also, it is likely that the cores themselves are not isotropic. They are however not expected to have preferred orientation within the molecule even though the residues inside the shell are believed to play a role in the nucleation of the iron [22].

So in summary, the form factor of the iron-loaded proteins is not known and cannot be properly measured, leading to inaccurate measured structure factors. However, the low q part of the structure factor below $q=0.1 \text{ \AA}^{-1}$ is a good approximation. The analysis of the structure factors of holoferritin is therefore restricted to the low q part.

B. Concentration dependence

In Sec. III, we suggested that the cause of the formation of the clusters is a phase separation in which the proteins are excluded from the solution and trapped into a confined space. If this is so there should be a very significant difference in the behavior of solutions of different protein concentration when frozen. We therefore repeated the same experiment for a few holoferritin concentrations. Figure 4 shows the results for a series of temperatures on solutions of 100, 50, and 35 mg/ml of holoferritin in 150 mM NaCl. Each of these sets of data was collected by cooling down to the lowest temperature shown ($\sim -25 \text{ }^\circ\text{C}$) and then gradually heating the sample up until no peak could be seen any longer.

For a solution at 100 mg/ml shown in Fig. 4(a), a single broad peak is formed upon cooling. However, as the tem-

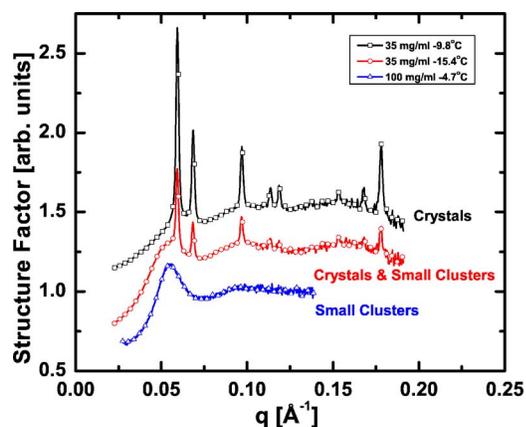


FIG. 5. (Color online) Comparison of the structure factor from large crystals (black), small clusters (blue), and the coexistence of the two (red). Sharp peaks arise from large crystals while broader features are associated with small nanoclusters.

perature is increased, this peak is seen to split into two narrower peaks. This indicates the formation of larger clusters as the temperature is increased. As simulations will show in Sec. V, the single broad peak seen at 200 mg/ml and the lowest temperature at 100 mg/ml is in fact the combination of the first two peaks of the face-centered cubic lattice and as the size of the cluster is increased, they get narrower and become distinct peaks. Therefore, the lattice spacing is roughly constant and at a value near the expected lattice spacing of 106 Å as the temperature is increased. The clusters, however, get larger and larger with increasing temperature.

A similar behavior is observed with a solution of 50 mg/ml in Fig. 4(b). The clusters at the lowest temperature are larger than at 100 mg/ml as indicated by the split peak and the smaller width. Again, however, as the temperature is increased, the peaks get narrower indicating larger crystalline clusters. Multiple Bragg peaks are seen and they correspond to the expected q positions of the expected fcc lattice with $a=183$ Å.

Figure 4(c) shows the temperature series at 35 mg/ml. When the sample is cooled to -19 °C, distinct Bragg peaks are already present. There is, however, a shoulder on the low q side which is reminiscent of the broad peak seen at 100 mg/ml. There is a coexistence of two separate states in the sample: small (~ 500 Å) clusters leading to the presence of a broad peak and large (~ 0.5 μm) crystals. The distinction between the characteristic structure factor of each state and the combination of both is made clear in Fig. 5

The experiment was repeated for more protein concentrations below 35 mg/ml. In all cases, large crystals giving rise to large Bragg peaks start to dominate the structure factor. The small clusters are, however, still present in small number and their contribution to the structure factor can still be seen. There is a coexistence of the two phases at many temperatures for certain concentrations. At 200 and 100 mg/ml, only one state was observed, with the small clusters growing into slightly large ones.

The constant position of the sharp peaks indicates that once large crystals form, the lattice spacing is fixed to a

value near the fcc value of 106 Å for the (111) spacing. The estimated crystal size from the Scherrer formula, i.e., the decrease in peak width with decreasing concentration shows that crystals get larger with decreasing protein concentration. This can be explained again by the trapping of proteins into voids in the ice after phase separation has occurred. After rapid cooling, the molecules are originally trapped into a certain conformation. At high concentration, there is a lot of crowding and the molecules cannot reorganize at any temperature before melting occurs. However, at lower concentrations, an increase in temperature provides enough energy to the system and increases the available space to allow the protein molecules to reorganize in such a way as to produce larger crystals. This is similar to an annealing process. The molecules have some room to move around and are not trapped in their original configuration. Therefore, the lower the concentration, the easier it is for large crystals to form, to the point that they are even seen immediately upon freezing. Also, in all cases the crystals are seen to get larger with increasing temperature, until a certain temperature is reached and then they are no longer seen. It seems that at lower concentrations, the proteins are only loosely trapped and heating the sample configuration provides enough room for diffusion for the proteins to get over the configurational barrier required to get out of the trapped state. There is enough space available for the proteins to explore new configurations and rearrange into larger aggregates with better defined crystalline order. The fact that small clusters can be annealed into large crystals might indicate that they are a precursor state to crystallization, possibly related to the nucleation process.

The spacing obtained from the broad structure factor peak is seen to be inversely proportional to the concentration. That is the peak moves to lower q with increasing concentration. This may be due to a crowding effect whereby more proteins get trapped per void for the more concentrated solution. Assuming that the size of the voids is the same for different concentrations, the 200 mg/ml solution will have twice as many proteins per void as the 100 mg/ml solution and 10 times more than the 20 mg/ml solution. If we assume the molecules space-fill voids of a fixed volume independent of the protein concentration, the more concentrated solution will have a smaller spacing. It is then not surprising that the spacing corresponding to the broad peak gets larger since there is less crowding. The small clusters are very compressible compared to the large crystals which have a fixed spacing. The size of the clusters obtained from the width of the peaks remains fairly constant for all concentrations below 100 mg/ml at roughly 300 Å. For the lowest concentrations, the spacing starts to approach the size of the molecules at 130 Å. The packing therefore resembles more a liquidlike structure than small crystalline clusters. Evidence of liquidlike packing is seen at low concentration as well as at higher temperature. At the highest concentrations measured, the lattice spacing increases with temperature.

IV. APOFERRITIN RESULTS

Similar experiments were performed using solutions of the empty protein shell apoferritin. The signal from these

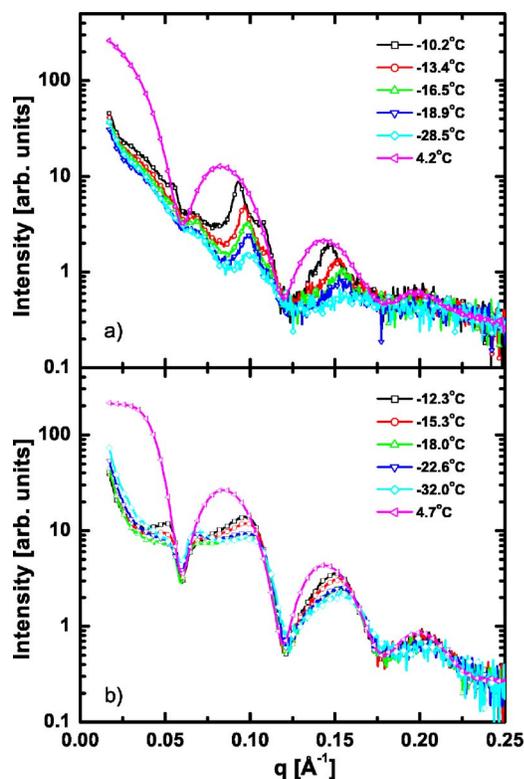


FIG. 6. (Color online) SAXS intensity from solutions containing (a) 50 mg/ml and (b) 150 mg/ml of apoferritin in 100 mM NaCl. The intensity changes significantly upon freezing indicating aggregation leading to a change in the structure factor.

samples was significantly weaker and made measurements of low concentrations difficult. Figure 6 shows the intensity measured from samples of apoferritin containing 50 and 150 mg/ml with 100 mM NaCl. The most intense curve in each case corresponds to the unfrozen sample. All have Bessel-function-like oscillations due to the spherical shape of the protein molecules. The form factor in the case of apoferritin is approximately given by the Fourier transform of a spherical shell with inner radius r_1 and outer radius r_2 :

$$F(q) = \frac{[\sin(qr_1) - qr_1 \cos(qr_1)]}{q^3(r_1^3 - r_2^3)} - \frac{[\sin(qr_2) - qr_2 \cos(qr_2)]}{q^3(r_1^3 - r_2^3)}. \quad (2)$$

A good fit to the intensity data can be obtained with Eq. (2) with a low level of polydispersity except at low q for the more concentrated solution. Comparing the two concentrations before freezing, one can see the appearance of a slight peak at roughly $q=0.035 \text{ \AA}^{-1}$ at 150 mg/ml. This peak could be attributed to liquidlike ordering but careful studies by other authors indicate the presence of paracrystalline ordering in solution [23].

As in the case of holoferritin solutions, as the temperature is decreased, the solution eventually freezes and a distinct change to the SAXS pattern can be seen. At both 50 and 150 mg/ml, the intensity near the origin drops and the oscillations from the form factor of the protein are distorted and even canceled off. The decrease in intensity at low q indi-

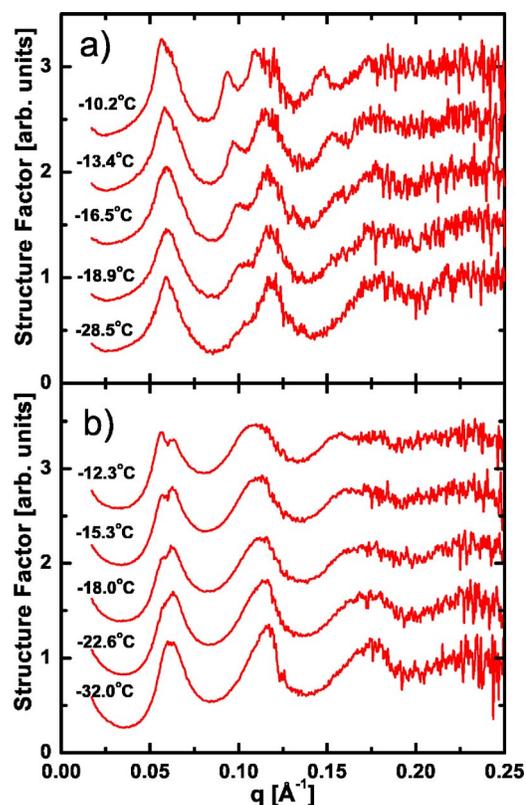


FIG. 7. (Color online) Structure factor measured from apoferritin at (a) 50 mg/ml and (b) 150 mg/ml. The structure factor changes with temperature.

cates the presence of aggregation. The total scattered intensity from the sample remains the same since the total number of electrons illuminated remains the same. However, cluster formation redistributes the intensity to different values of q . The maximum at the origin gets narrower as the cluster size increases and would become a delta function in the limit of an infinite crystal. New peaks in the intensity appear where peaks in the structure factor are present.

When compared with the case of holoferritin, the apoferritin intensity curves seem to be a very complicated function with no clearly identifiable peak which one could interpret as a typical spacing. There is, however, a great advantage in using the empty protein shells. They are very monodisperse and very closely isotropic. Therefore, the structure factor obtained is much more meaningful than in the iron-loaded protein case. In the latter, the structure factor measured was only a first order approximation valid over a small q range. In the apoferritin case, the entire structure factor curve obtained should be valid.

A. Structure factors

The unfrozen solution provides a good measure of the molecular form factor. The data for frozen solutions were therefore divided by this measured curve to obtain the structure factor using Eq. (1). The derived structure factors are shown in Figs. 7(a) and 7(b) for 50 and 150 mg/ml, respectively. The complicated intensity curves turn into smooth

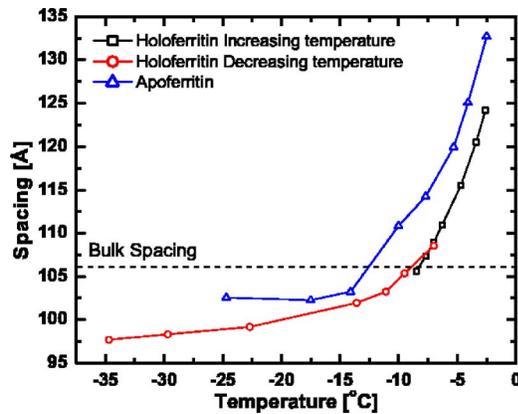


FIG. 8. (Color online) Spacing of the apoferritin molecules versus temperature overlaid with the spacing from iron loaded ferritin.

curves when divided by the proper form factor. Peaks in the structure factor again indicate the presence of some form of ordering. The presence of multiple peaks indicates a good level of long-range order.

At 150 mg/ml, the solution did not freeze until a temperature lower than -9°C was reached. The first three peaks in the structure factor are clearly visible. One thing to notice is the slight splitting of the first peak. The interesting aspect of it is the fact that the lower q part of the peak has a lower intensity than the higher q part. It is expected from the holo-ferritin data and from calculations of the diffraction from an fcc crystal that the second peak, the $\{200\}$ peak of fcc ferritin crystals is weaker than the first peak, the $\{111\}$ peak.

The structure factors for the solution at 50 mg/ml display more features. Most importantly, the second broad peak at $q \sim 0.12 \text{ \AA}^{-1}$ is split, with a sharp peak appearing at $q \sim 0.1 \text{ \AA}^{-1}$. The position, height and width of this peak changes significantly with temperature, moving to lower q with increasing temperature. Another small peak at $q \sim 0.15 \text{ \AA}^{-1}$ is seen for a few of the measured temperatures.

The thermal expansion measured is compared in Fig. 8 with that of holo-ferritin. The holo-ferritin sample was at 100 mg/ml with 150 mM NaCl while the apoferritin sample was at 50 mg/ml with 100 mM NaCl. The difference in mass concentration is due to the heavier protein with iron. The molar concentration of both samples is the same within 30%. As can be seen, the results for the two are very similar, indicating that the presence of the iron core does not play an important role, if at all, in the aggregation mechanism. The increase in the molecular spacing follows the same curve for both cases, with the absolute values differing only because the holo-ferritin sample is slightly more concentrated giving rise to a smaller spacing. As a reference, a dashed line was drawn at a value corresponding to the spacing of the $\{111\}$ planes in the fcc crystal of ferritin. The measured spacing appear to be smaller than the crystal spacing, but as mentioned before, this is due to the merging of the first two Bragg peaks into one broader peak.

V. CLUSTER SIMULATIONS OF STRUCTURE FACTOR

The size of the small clusters of ferritin we have observed ranges from 300 to 600 \AA roughly. Considering the size of

the ferritin molecules is 130 \AA , this means that a cluster of dimension 600 \AA has only 4–5 molecules across, containing a total of 100 molecules or less. This total is small enough that one can realistically calculate exactly the structure factor of a cluster by treating it as one “supermolecule.” If we assume that there is a central molecule located at the origin of real space, we can add molecules one by one and calculate the spherically averaged form factor directly using the Debye formula [21]

$$S^2(q) = \sum_i^N \sum_j^N \frac{\sin qr_{ij}}{qr_{ij}}, \quad (3)$$

where N is the total number of ferritin molecules and r_{ij} is the distance between molecules i and j . The value of N in our case would be limited to less than 200 making this calculation somewhat straightforward using a simple computer program. This formula gives a spherically symmetric scattering functions even for non symmetric distributions of particles.

A. Isotropic clusters

It is logical to start by assuming the small clusters possess a face-centered cubic structure since this is the most easily formed crystal structure of the protein. As a first guess of the structure of the clusters, one might assume a roughly spherical aggregation, with a central molecule to which new molecules attach forming an isotropic structure. Supermolecular clusters were built up starting with a central molecule located at the origin. Molecules were added one by one on fcc lattice sites. The structure of the dimer is unique. However, the trimer structure and most other oligomeric structure have degenerate states. Even if one limits the number of possible configurations to keep the overall supermolecule roughly spherical, there are still multiple possibilities. The first neighbors from the central molecule in the fcc structure are all the $\langle 110 \rangle$ lattice sites. It does not matter which one of the 12 possibilities is chosen first since a simple rotation transforms any cluster formed into any other one. However, different structures are formed depending on which of the remaining $\langle 110 \rangle$ sites is chosen next. In general, all supermolecules built this way are degenerate. However, clusters containing only filled shells of nearest neighbors are unique. If all the 12 $\langle 110 \rangle$ sites are occupied, the supermolecules containing 13 ferritin molecules is unique. The same goes for all other filled shells.

The structure factor was calculated using equation (3) for supermolecules of 1 to 177 ferritin molecules. The number 177 corresponds to a supermolecule containing the first 9 shells around the center molecule, that is adding every lattice point up to the $\langle 330 \rangle$ and $\langle 114 \rangle$ fcc points, which have the same distance to the center molecule.

In Fig. 9, the structure factor is seen to evolve from unity for the one central molecule to an oscillating curve for the 177 molecule cluster. Only filled shell cases are shown, containing 1, 2, 13, 19, 43, 55, 79, 87, 135, 141, and 177 molecules, corresponding to the monomer, the dimer, and filled shells up to $\langle 110 \rangle$, $\langle 200 \rangle$, $\langle 112 \rangle$, $\langle 220 \rangle$, $\langle 310 \rangle$, $\langle 222 \rangle$,

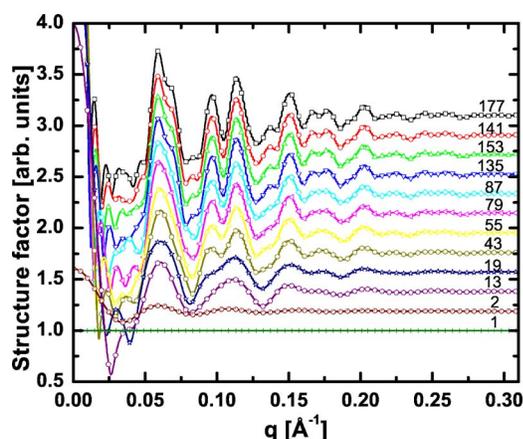


FIG. 9. (Color online) Simulated structure factors from fcc supermolecules made of 1, 2, 13, 19, 43, 55, 79, 87, 135, 141, and 177 molecules, corresponding to the monomer, the dimer, and filled shells up to $\{110\}$, $\{200\}$, $\{112\}$, $\{220\}$, $\{310\}$, $\{222\}$, $\{123\}$, $\{400\}$, and $\{330\}$ and $\{114\}$, respectively.

$\langle 123 \rangle$, $\langle 400 \rangle$, and $\langle 330 \rangle$ and $\langle 114 \rangle$, respectively. These represent roughly spherical clusters of diameter 260, 368, 450, 520, 581, 637, 688, 736, and 780 Å.

The structure factor from the monomer is exactly unity. The structure factor from the dimer is the sinc function. As the cluster size grows, broad peaks get more well defined. These peaks are at q values near the fcc Bragg peaks. Focusing our attention on the first peak at $q \sim 0.06 \text{ \AA}^{-1}$, one notices that as the number of protein molecules is increased, this peak gets narrower, then splits into two distinct peaks that also get progressively narrower. This is because there are two Bragg peaks near $q = 0.06 \text{ \AA}^{-1}$. The $\{111\}$ Bragg peak is at $q = 0.0591 \text{ \AA}^{-1}$ and the $\{002\}$ peak is at $q = 0.0686 \text{ \AA}^{-1}$. For a small cluster of 13 molecules, the diameter of the cluster is 390 Å corresponding to a width of the diffraction pattern of 0.016 \AA^{-1} . The two peaks are then indistinguishable. The same thing occurs at other q values. As the crystal grows, Bragg peaks become better defined, until they essentially become delta functions in the infinite limit.

The simulations shown in Fig. 9 show a good similarity with the structure factor data obtained for holoferritin at low q and also for apoferritin over the whole q range measured. The first peak was often measured to split with increasing temperature as seen also in the simulation. The very rapid oscillations in the very low q part of the simulation arises from the exact shape of the clusters. This part is very sensitive to the exact structure. It is not expected that every cluster in the sample will be identical. Instead, a weighted sum over many of the simulated clusters is a better description of the system. Such a summation over a small range of sizes will affect very little the overall structure factor at values of q above the first minimum. The very low q part, which is often not even measured in our experiments would then average out to a smooth function.

We next address the question of whether the small clusters can be composed of other plausible structures. One such possibility would be the formation of a hexagonal close-packed (hcp) structure. The hcp structure is similar to the fcc structure with identical layers in the $\{111\}$ plane but with a differ-

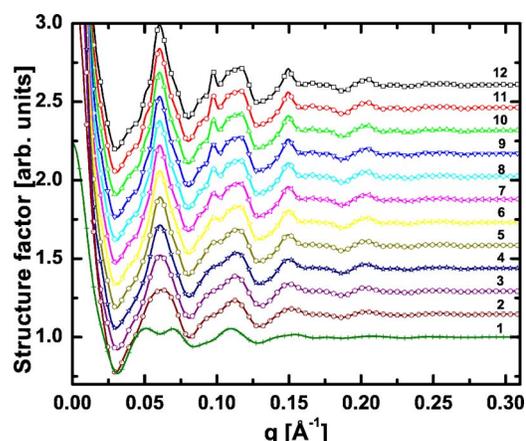


FIG. 10. (Color online) Simulated structure factors from planar supermolecules made of 5 $\langle 110 \rangle$ rods of 1 to 12 molecules each.

ent stacking of these layers. Similar simulations of the structure factor were performed using an hcp structure. The result (not shown) clearly showed the data is inconsistent with the hcp model. Most of the Bragg peaks of fcc and hcp are the same due to the similarity of the two structures. However, the hcp lattice gives rise to some peaks which are clearly not present in the data.

Also, it was mentioned above that the spacing measured by fitting the peak position of the iron-loaded ferritin corresponded at low temperature to a spacing smaller than the $\{111\}$ spacing. We mentioned that this might indicate a different structure than fcc. However, close inspection of the calculated structure factors indicates that the peak maximum from a small fcc supermolecule is not at the q value of the $\{111\}$ Bragg peak but rather at a value of q corresponding to a spacing of ~ 103 instead of 106 Å. This is still larger than what we measure indicating that the lattice spacing is indeed compressed from the expected bulk value.

B. Planar clusters

Work published by Yau and Vekilov presents evidence that the nucleation of apoferritin proceeds along a planar rather than an isotropic pathway [9]. They measured directly the structure of small clusters of apoferritin using atomic force microscopy (AFM). Their data indicate the formation of rod structures aggregating into a close to planar structure. Supermolecules of this type were simulated in order to determine if they are a good representation of our data.

Figure 10 shows simulated structure factors for those planar clusters which were observed by Yau *et al.* They observed using AFM that near critical nuclei of ferritin forming in solution and falling on a substrate were made of 4 to 8 rows of 4 to 7 molecules. The rows of molecules are along the $\langle 110 \rangle$ directions of the fcc crystal and are arranged in an accordion structure. These clusters were found to be nuclei near critical in size. Even though all the molecules in these clusters lie on fcc lattice spots, the cluster shapes are clearly different from the spherical fcc cluster simulated above and will therefore have a different scattering structure factor. Simulations were made for 1 to 12 rows of

1 to 12 molecules. Figure 10 shows the simulations for 5 rows of 1 to 12 molecules.

There are significant differences between the isotropic cluster structure factors of Fig. 9 and the structure factors from the planar structures. This indicates that SAXS is a good technique for distinguishing between the two even if they are randomly oriented. The peaks for both cases are and should be at the same q values and get narrower with increasing cluster size since the position of the peaks depends on the lattice, which is fcc in both cases. However, the planar structure has many planes of the fcc structure missing and therefore some characteristic distances either occur less often or not at all. This in turn leads to different widths of the peaks. Similarly, the size of the cluster is different in all directions. The directions where it is larger will give rise to narrower peaks while the peaks corresponding to the directions where the cluster is small will be much broader. This leads to the first peak getting much narrower before it starts to split into two peaks when comparing to the isotropic cluster case.

The planar clusters when compared to the fcc clusters display a much sharper first peak for the same number of molecules. For example, the cluster with 5 rows of 8 molecules, for a total of 48 ferritin molecules, has a much sharper peak at $q=0.06 \text{ \AA}^{-1}$ than the fcc cluster with 55 molecules. Also, the planar cluster simulation give rise to a sharper peak at $q=0.1 \text{ \AA}^{-1}$ without a splitting of the first maximum as seen for the isotropic clusters. A quick inspection of the data indicates that the first maximum is always narrower in the planar case than in the isotropic simulations. The only way an fcc cluster can yield a structure factor with a sharper first maximum is by increasing the size of the cluster, but this will in turn cause this first maximum to split into two. This may be an indication that sometimes planar clusters do form according to the way Yau *et al.* observed.

VI. FITTING OF STRUCTURE FACTORS

It is logical to assume that there is a cluster size which is more likely than others. There could be a Gaussian distribution, for example, around a central supermolecule size. In order to determine this central value, a fit was performed for each structure factor measured assuming only 1 type of supermolecule was present. A fit was performed for each model simulated, including spherical clusters up to 177 molecules and planar clusters containing anywhere between 1 to 11 rods of 1 to 12 molecules. The computer program kept track of the parameters obtained for each of these fits and at the end, the supermolecular model yielding the lowest value of $\chi^2 = \sum [S(\text{measured}) - S(\text{fit})]^2$ was output. Thus 309 fits were performed on each data set and the best fit was kept as the most likely supermolecule, among all the simulations performed.

The fits obtained at a few temperatures on the 200 mg/ml holoferritin sample are shown in Fig. 11(a). The fits for a 100 mg/ml sample is shown in Fig. 11(b). They are shown in the order at which they were measured from bottom to top. The width and position of the first peak is well captured and even the secondary features are fitted. The model which

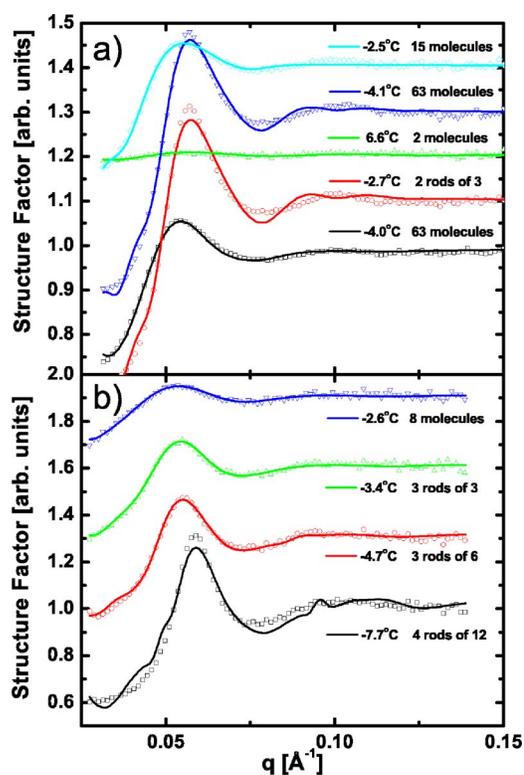


FIG. 11. (Color online) Fits (solid lines) to the measured structure factors (symbols) at multiple temperature for (a) a 200 mg/ml holoferritin sample and (b) a 100 mg/ml sample. The temperature as well as the model used for the fit is indicated directly next to the data curve. The curves are offset for clarity.

yielded the best fit shown is indicated directly next to the curve, as well as the temperature the data was collected at. The symbols are the data while the solid curves are the fits. Some of the best fits obtained were using the planar cluster model. This was only the case for fairly small clusters. This is not completely unexpected since Yau *et al.* only observed small ones and this planar structure is assumed to be only an intermediate crystallization state and is not the equilibrium large crystal shape. However, smaller clusters have a more generic structure factor with less well-defined features. The structure factors from the two models eventually converge as the size gets smaller, with the dimer models being identical. There are multiple models which can be used that give a decent fit for the small clusters. Nevertheless, the multiple occurrences where the best fitting model is a planar cluster indicates that these do indeed occur. They are, however, not the predominant form for most of the samples measured. The lattice spacing of the clusters fitted was found to match very well the estimated values from just the position of the first peak.

The fits to the structure factors for two different samples of apoferritin at 50 mg/ml are shown in Figs. 12(a) and 12(b). The data are shown in the order at which they were measured from bottom to top. As opposed to the holoferritin case, the structure factor of apoferritin samples is expected to be valid over the whole range of q due to the small polydispersity of the molecules. This allows us to fit more than just the first maximum and get more information about the sample.

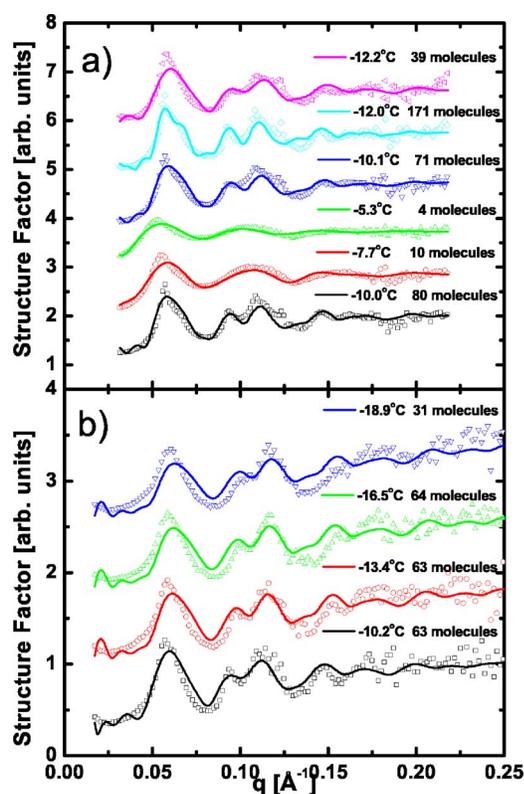


FIG. 12. (Color online) Fits (solid lines) to the measured structure factors (symbols) at multiple temperature for two separate 50 mg/ml apoferritin sample. The temperature as well as the model used for the fit is indicated directly next to the data curve. The curves are offset for clarity.

The fits obtained all indicate that the isotropic clusters are predominant. The size of the clusters varies with temperature, with the general trends noted previously, that small clusters at low temperature can grow larger when annealed near the melting point of the sample.

In order to obtain these fits, a few free parameters other than the lattice spacing of the clusters were included in the fitting programs. Those were a constant and a slope value to account for errors in the normalization and the background subtraction in the intensity data. Also, an overall scale multiplying the simulated structure factor was used to determine which fraction of the proteins in solution was included in the clusters. Finally, the last parameter was a damping term to the structure factor. This damping term is the Debye-Waller factor (DWF) used to simulate the presence of thermal vibrations in the sample as well as disorder.

The fraction of proteins included in clusters or crystallized fraction showed that when frozen, nearly 100% of the proteins were included in a cluster but as the temperature neared the melting point, this number started to drop, indicating the crystals were breaking apart. A typical value of the DWF corresponds to a root mean square displacement of roughly 4 \AA for holoferritin and 2 \AA for apoferritin. The different values are representative of the higher disorder due to the presence of the iron core.

VII. CONCLUSION

The aggregated state formed upon freezing solutions of ferritin, first seen by Kilcoyne [12], is found to be due to cluster formation. A phase separation process is envisaged in which protein depleted ice surrounds voids of concentrated protein, trapped in the ice network. The protein rich regions may or may not be frozen. They may still be in a liquid aqueous environment due to their high solute percentage. The phase separation explains the hysteresis of the system causing it to have different freezing and melting temperatures.

Clusters possessing face centered cubic symmetry are present at all temperatures. At the highest temperatures, they disintegrate into smaller fcc structures and eventually revert back to a solution. The evidence indicates the clusters break up into smaller ones before they revert to liquidlike highly concentrated structures though there is some evidence of liquidlike packing leading to peaks in the structure factors. The large lattice spacings observed at high temperatures are better explained by a concentrated liquid than a solid with an expanded lattice. However, there is clear evidence of a large thermal expansion of the fcc lattice with increasing temperature. The lattice spacing of holoferritin solutions producing large crystals is directly measured and shows a slight thermal expansion. It is however the smaller structures which show the larger thermal expansion for 100 and 200 mg/ml holoferritin. Fits to the structure factors show that the sample consists of many small fcc supermolecules and the average lattice spacing greatly increases. This is quite interesting because it indicates that the fcc structure is a stable one even when the protein molecules are far apart and no longer directly in contact. This is reminiscent of colloidal systems in which crystal structures can be formed even at very low volume fractions, below the fraction one would get if the colloids were in direct contact. In charge stabilized colloidal systems, the spacing between the particles in the crystal can be many times the diameter of the particles. It is well known that the salt concentration in colloidal suspensions changes the organization in the sample from a fluid to various crystalline structures by screening the Coulomb interactions [24,25]. The addition of salt to the ferritin solution changes the size of the fcc clusters formed and whether they occur at all. The ferritin proteins trapped in ice appear to behave similar to colloidal particles in the sense that they have crystalline order with a varying lattice spacing and without direct contact between nearest neighbors in the lattice.

The measured decrease in size and number of the clusters when approaching the melting point is consistent with previous measurements of Kilcoyne *et al.* using small angle neutron scattering [12]. Their analysis was based on the characteristics of the very low q range where the presence of large structures has a large effect. They did measure a few weak peaks in the intensity which they tentatively assigned to Bragg peaks from the hexagonal close-packed structure, but their data did not fully verify this. No measurement of the annealing process discussed above was made in their case.

The fits to the structure factors discussed were of only a single cluster model at a time. There is no reason to expect the sample to contain a single cluster structure. There should

be a range of different structures with different probabilities. Only the most likely structure was fitted for the following reasons. First, the structure factors simulated for a single type of clusters such as the spherical clusters is a slow varying function of the number of particles. There is not much difference between the structure factor calculated from a supermolecule containing 45 molecules and a supermolecule containing 50 molecules. Therefore, using multiple similar models and summing them together only slightly improved the fits. Furthermore, including too many models leads to many fit parameters which created too many degrees of freedom leading to multiple solutions. The real system probably includes a structure which is more likely to occur than others with the other structures within a range of sizes present with some probability, including isomorphic structures which were not all simulated. It proved impossible with the available data to determine this range of size with accuracy.

It is also possible that different habits of the clusters are present. Clusters could be forming with both planar and isotropic structures. It again becomes tricky to allow both to exist in the fitting routine, leading to ambivalent multiple solutions. Therefore, only the most likely structure of the supermolecules could be determined.

Some of the fits obtained indicate the most likely structure present is consistent with the planar nucleation structures measured by Yau and Vekilov [9]. Most of the data is, however, better explained by the isotropic simulations. We therefore conclude that isotropic nucleation is the main pathway but the planar pathway also occurs with a small probability.

The data clearly indicated the presence of very small clusters down to less than 10 molecules at early times shortly after the sample was cooled. As the temperature was raised, these small clusters grew into large crystals. There is clear evidence in Fig. 5 that both crystals and small clusters can coexist, with smaller clusters appearing first and then growing larger. The freezing process causes nucleation of the protein molecules. These small structures with fcc ordering are stable at certain temperatures. Raising the temperature causes a change in the level of saturation of the sample that allows these clusters to grow. Therefore, the small supermolecules

are identified as the nuclei of ferritin at these concentrations and temperatures. They are stable aggregates more likely to grow than to redissolve.

The fact that the peak in both the ferritin and apoferritin data shifts with temperature indicates that the contacts between the proteins are fairly loose. The molecules are not in direct contact as they are in the crystal. The large expansion of the spacing as the melting point of ice is approached from below while the peak remains present indicates that the structure is partially preserved but the proteins are no longer in contact. This is very reminiscent of colloidal crystals where the components are not directly in contact, yet are highly organized. The lattice spacing of a colloidal crystal can be changed without destroying the lattice by altering the conditions in the solution.

It is unclear how the results can be applied to other proteins. It is very likely that every protein behaves differently when frozen and most may not form crystals as readily as ferritin, but this is unknown. All proteins would likely show different thermal compressibility of the lattice if they crystallize at all in this way. More studies may reveal that it may be feasible to create small crystallites of many proteins using this technique. With new powerful free electron laser x-ray sources such as LCLS coming online in the next few years, it may be possible to use such small crystals for structure determination.

ACKNOWLEDGMENTS

This research was supported by the NSF under Grant No. DMR03-08660. The UNICAT facility at the Advanced Photon Source (APS) was supported by the University of Illinois at Urbana-Champaign, Materials Research Laboratory (U.S. DOE Contract No. DEFG02-91ER45439, the State of Illinois-IBHE-HECA, and the NSF), the Oak Ridge National Laboratory (U.S. DOE under contract with UT-Battelle LLC), and the National Institute of Standards and Technology (U.S. Department of Commerce). One of the authors (S. Boutet) wishes to thank the *Fonds québécois de la recherche sur la nature et les technologies* for its support.

-
- [1] A. McPherson, *Crystallization of Biological Macromolecules* (Cold Spring Harbor, Laboratory Press, Woodbury, NY, 1999).
- [2] A. George and W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 361 (1994).
- [3] A. Mirarefi and C. Zukoski, *J. Cryst. Growth* **265**, 274 (2004).
- [4] F. Rosenberger, P. Vekilov, M. Muschol, and B. Thomas, *J. Cryst. Growth* **168**, 1 (1996).
- [5] A. Chernov, *Modern Crystallography III Crystal Growth* (Springer-Verlag, Berlin, 1984).
- [6] A. Chernov and H. Komatsu, *Principles of Crystal Growth in Protein Crystallization* (Kluwer Academic Publishers, Dordrecht, 1995).
- [7] S.-T. Yau, D. Petsev, B. Thomas, and P. G. Vekilov, *J. Mol. Biol.* **303**, 667 (2000).
- [8] S.-T. Yau, B. R. Thomas, and P. G. Vekilov, *Phys. Rev. Lett.* **85**, 353 (2000).
- [9] S.-T. Yau and P. G. Vekilov, *J. Am. Chem. Soc.* **123**, 1080 (2001).
- [10] I. K. Robinson, I. A. Vartanyants, G. J. Williams, M. A. Pfeifer, and J. A. Pitney, *Phys. Rev. Lett.* **87**, 195505 (2001).
- [11] G. J. Williams, M. A. Pfeifer, I. A. Vartanyants, and I. K. Robinson, *Phys. Rev. Lett.* **90**, 175501 (2003).
- [12] S. Kilcoyne, G. Mitchell, and R. Cywinski, *Physica B* **180-181**, 767 (1992).
- [13] P. M. Harrison and P. Arosio, *Biochim. Biophys. Acta* **1275**, 161 (1996).
- [14] T. Granier, B. Gallois, A. Dautant, B. Destaintot, and G. Precigoux, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 580 (1997).
- [15] A. Guinier and G. Fournet, *Small-Angle Scattering of X-Rays*

- (Wiley, New York, 1955).
- [16] A. P. Hammersley, S. Svensson, M. Hanfland, A. Fitch, and D. Hausermann, *High Press. Res.* **14**, 235 (1996).
- [17] M. Muschol, and F. Rosenberger, *J. Chem. Phys.* **107**, 1953 (1997).
- [18] K. Gekko, *Water Relationships in Food* (Plenum Press, New York, 1991).
- [19] V. Morozov and T. Morozova, *Biopolymers* **20**, 451 (1981).
- [20] V. Morozov, T. Morozova, E. Myachin, and G. Kachalova, *Acta Crystallogr., Sect. B: Struct. Sci.* **41**, 202 (1985).
- [21] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics* (Wiley, New York, 2001).
- [22] F. Fischbach, P. Harrison, and T. Hoy, *J. Mol. Biol.* **39**, 235 (1969).
- [23] W. Haubler, A. Wilk, J. Gapinski, and A. Patkowski, *Biochim. Biophys. Acta* **1164**, 331 (1993).
- [24] T. Harada, H. Matsuoka, T. Ikeda, and H. Yamaoka, *Colloids Interfaces A* **174**, 79 (2000).
- [25] A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf, and P. Schurtenberger, *Nature (London)* **432**, 492 (2004).